BMJ Open

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

Klaus Munkholm, Asger Sand Paludan-Müller, Kim Boesen

To cite: Munkholm K, Paludan-Müller AS, Boesen K. Considering the methodological limitations in the evidence base of antidepressants for depression: a reanalysis of a network meta-analysis. *BMJ Open* 2019;**9**:e024886. doi:10.1136/bmjopen-2018-024886

► Prepublication history and additional material for this paper are available online. To view, please visit the journal (http://dx.doi.org/10.1136/bmjopen-2018-024886).

Received 19 June 2018
Revised 18 October 2018
Accepted 8 May 2019

Nordic Cochrane Centre, Rigshospitalet, Copenhagen, Denmark

Correspondence to
Dr Klaus Munkholm;
km@cochrane.dk

ABSTRACT

Objectives To investigate whether the conclusion of a recent systematic review and network meta-analysis (Cipriani *et al*) 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* 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 pairwise 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*. 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 with trials without a placebo run-in design (p=0.05). The effect size of antidepressants was higher in published trials compared with unpublished trials (p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 0–52 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* 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* 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 USA, 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 UK 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

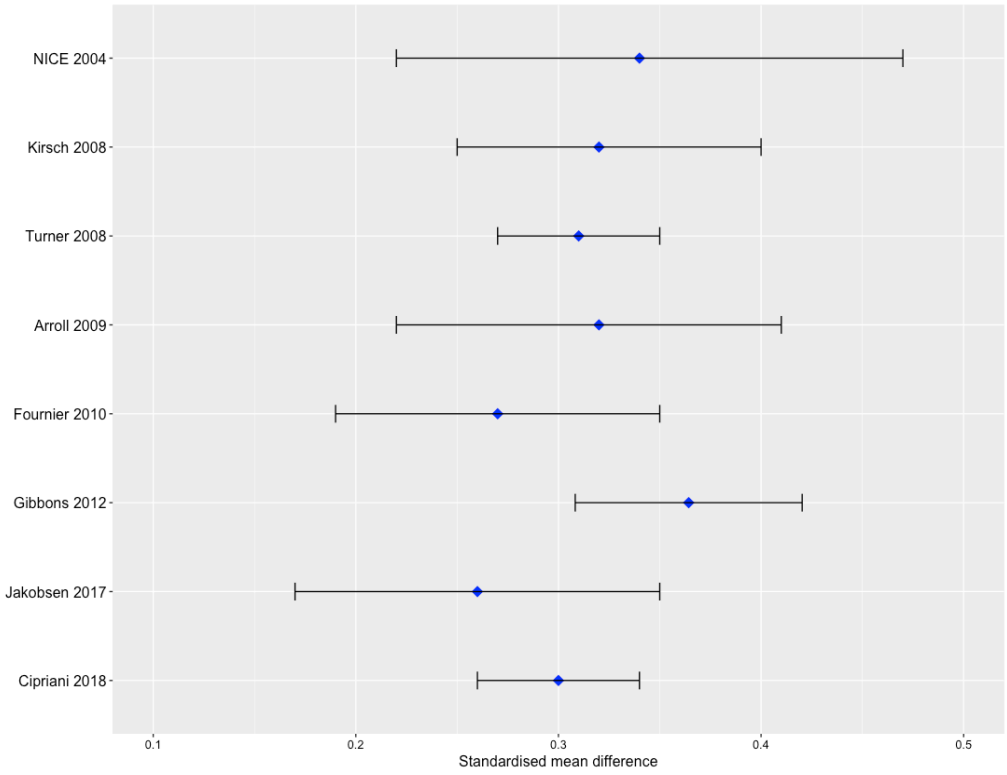

Figure 1 Previous meta-analyses reporting effect sizes for antidepressants versus placebo in adults. Data are reported as standardised mean differences with 95% CIs. NICE 2004[6]: SSRIs. Kirsch 2008[8]: 'new generation' antidepressants. Turner 2008[9]: all antidepressants. Arroll 2009[41]: antidepressants for depression in primary care. Data represent a pooled estimate of tricyclic antidepressants and SSRIs versus placebo, fixed-effects model. Fournier 2010[42]: all antidepressants. Data represent pooled estimate from three groups of severity (mild to moderate, severe, very severe), fixed-effects model. Gibbons 2012[43]: fluoxetine and venlafaxine. Jakobsen 2017[7]: SSRIs. The effect size of mean change scores. Cipriani 2018[5]: all antidepressants. SSRIs, selective serotonin reuptake inhibitors.

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, they[5] 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 limitations 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. We have already addressed some of the limitations in the risk of bias assessment in the Cipriani *et al* review.[14] However, given the potential implications of Cipriani *et al*'s review,[5] 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*.[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 dataset[5] 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 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 V.3.4.3 for random-effects meta-analyses based on the inverse variance method and calculated effect sizes as standardised mean differences (SMD) as Hedges' g with corresponding 95% CI. The extent of variation among the intervention effects observed in different studies was calculated as Tau^2 and the percentage of the variability in effect estimates that was due to heterogeneity was calculated as I^2. 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 SD were unchanged.[17] We did subgroup analyses based on the use of a 'placebo run-in' study design, sponsorship and publication status, according to the trial characteristics published by Cipriani *et al*.[5]

Quality assessments

We evaluated whether Cipriani *et al*'s risk of bias assessments were in accordance with the Cochrane Handbook,[17] as stated by the authors.[5] Where the approach differed we compared the risk of bias assessment by Cipriani *et al*[5] with our reassessment following the Cochrane Handbook.[17] The specific domains (and type of bias) assessed were sequence generation (selection bias), allocation sequence concealment (selection bias), blinding of participants and personnel (performance bias), blinding of outcome assessment (detection bias), incomplete outcome data (attrition bias), selective outcome reporting (reporting bias) and other potential sources of bias.[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 CIs, and publication bias.[18]

Patient and public 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

Randomisation sequence generation and allocation sequence 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] 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.[17]

Blinding of participants, personnel and outcome assessment

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 assessment[5] 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*[5] to be at low risk of bias in the blinding of participants' domain had tested the blinding integrity (online 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,[20] and smaller effects have been observed when the trials were better blinded by adding atropine to the placebo.[21] 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.[5] 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

>20% they were rated unclear, and if the total dropout rate was <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.[17] Further, the authors did not consider the reasons for dropout, although this is also recommended by the Cochrane Handbook.[17]

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 >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 trials (39%) and according to the methods described by Cipriani *et al*[5] they should have been rated as high risk of bias unless an 'appropriate imputation method' was used. Cipriani *et al* characterised the last observation carried forward (LOCF) method as inappropriate,[22] but they did not provide data on the used imputation method in the included trials. We were therefore not able to apply Cipriani *et al*'s categorisations in our reassessment of the attrition bias. Most antidepressant trials use the LOCF imputation method,[23] which may lead to an underestimation of the variability, a falsely low p value and an overestimation of treatment effects.[24]

Selective outcome reporting

Cipriani *et al*[5] 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.[17] 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.[17] Selective outcome reporting leads to overestimation of the benefits and underestimation of the harms of interventions.[25]

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.[17] Relevant biases included in this domain were baseline imbalances and design-specific risks of bias for crossover and cluster randomised trials, which were eligible according to the Cipriani *et al* protocol,[22] although the trial designs were not specified in the review.[5] 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.[26] 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* (online 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).

Summary risk of bias assessment

The authors deviated from Cochrane's overall risk of bias categorisation of low, unclear or high risk of bias,[17] 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.[5] 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.[17] The Handbook instead recommends an overall qualitative assessment considering the relative importance of different domains.[17] 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.[17] Applying these criteria (Cochrane Handbook, table 8.7.a)[17] 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 (ie, 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 (online 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 (online S1 Appendix).

Publication bias

Publication bias of antidepressant trials is pervasive and distorts the evidence base.[9] Many industry funded

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 to 0.31	0.038	40.1%
Overall (mean difference on the HAMD17)						
Overall	109	166	1.97**	1.74 to 2.21	1.896	27.6%
Publication status						
Published	196	294	0.33*	0.30 to 0.35	0.037	40.0%
Unpublished	57	96	0.15*	0.11 to 0.19	0.020	0.0%
'Placebo run-in'						
Yes	142	221	0.31*	0.28 to 0.34	0.043	35.0%
Unclear	79	120	0.29*	0.25 to 0.33	0.032	47.6%
No	30	46	0.22*	0.16 to 0.29	0.032	35.5%
Sponsorship						
Sponsored	207	341	0.27*	0.25 to 0.30	0.033	35.4%
Unclear	10	12	0.39*	0.25 to 0.52	0.026	33.0%
Not sponsored	36	37	0.41*	0.31 to 0.52	0.075	55.7%

*SMD
**Mean difference on 17-item Hamilton depression rating scale.
ES, effect size; HAMD17, 17-item Hamilton depression rating scale; I^2, inconsistency; Tau^2, estimate of overall heterogeneity; SMD, standardised mean difference.

antidepressant trials remain unpublished or are inadequately reported.[9] Cipriani *et al*[5] included 436 published and 86 unpublished studies, but as many as a thousand antidepressant studies may have been conducted.[13] 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*[9] in 2008 of published versus unpublished antidepressant trials registered by the United States Food and Drug Administration (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*[5] are likely inflated due to publication bias. They correctly downgraded their confidence in the evidence due to the risk of publication bias, but it would also have been appropriate to estimate the impact of publication bias on their effect estimate.

Trial duration and long-term effects

Cipriani *et al*[5] extracted outcome data as close to 8 weeks follow-up as possible within an interval of 4–12 weeks,[5] but they did not provide a rationale for this decision.[22] The common clinical practice is to prescribe antidepressants for much longer periods. In the Netherlands, 43% of SSRI (Selective serotonin reuptake inhibitor) users receive treatment for 15 months or more,[27] while 68% of those who use antidepressants in the USA take them for 2 years or more, and 25% take them for >10 years.[3] Although the 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'. A more appropriate method would have been to extract outcome data according to length of treatment and follow-up to assess any change in the treatment effect over time. According to the trial characteristics reported by Cipriani *et al*,[5] 12 (4%) of the 304 placebo-controlled trials lasted >12 weeks. However, we found that only four of these 12 trials contained an uninterrupted double-blind, placebo-controlled phase of >12 weeks (online S2 Appendix). The two placebo-controlled trials with the longest follow-up included 81 participants at 36 weeks (online S2 Appendix). A further consequence of a short follow-up period is an underestimation of serious and non-serious adverse events.[28]

Placebo run-in and inclusion of already treated patients

The placebo run-in study design distorts the estimates of benefits and harms (box 1). Cipriani *et al* did not provide a clear definition of a placebo run-in,[22] 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] We performed random-effects meta-analyses of the placebo-controlled trials according to the use of a placebo run-in design and 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

Box 1 '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*[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 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 with 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.[44] Withdrawal symptoms typically occur within a few days after discontinuation but there is great clinical variation[44] (*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.[44] 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 >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 <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 with placebo.

C. Minimal clinically relevant difference.

Cipriani *et al* reported an overall effect estimate measured as a standardised mean difference (SMD) of 0.3 between antidepressants and placebo.[5] 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 SMD of 0.5, as a clinically significant change.[6] However, this difference was arbitrary and not based on empirical data.[45] 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.[46] Others have interpreted the same data and suggested that changes of seven points or more on the Hamilton scale, corresponding to a SMD of at least 0.875, are necessary for a clinician to detect a minimal clinical improvement.[47] We found that the mean difference between antidepressants and placebo on the 17-item Hamilton depression rating scale (range 0–52 points), based on Cipriani *et al*'s data,[5] was 1.97 points.

of 0.29 (95% CI 0.25 to 0.33, 120 comparisons, 79 trials) where the use of a placebo run-in was unclear and an SMD of 0.22 (95% CI 0.16 to 0.29, 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, eight comparisons, five trials). The use of the placebo run-in design and its implications were not discussed by Cipriani *et al*.[5]

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 (box 1). Furthermore, by not including a careful analysis of the serious harms, which include aggression, suicide and death,[29] 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 with 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 data[7] and as serious harms, including death, of antidepressants are often not reported in published papers.[30]

Except for two drugs none of the included antidepressants had statistically significant lower total dropout rates than placebo.[5] 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.[31]

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 (online 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 data[7] and even clinical study reports are unreliable because of selective reporting of this outcome.[31] The inability to cope with daily activities and the drugs' side-effects may be more important to patients than their depressed mood[32] and the exclusion of patient-relevant outcomes in the protocol[22] 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 (box 1). It is a problematic outcome because it lacks clinical relevance and it may create an illusion of clinical effectiveness.[33] Dichotomisation of outcomes measured on rating scales leads to loss of statistical power, and it increases the risk of false positive results[34] and spuriously inflated effect sizes.[33] Therefore, methodologists discourage the use of such dichotomised outcomes and they should generally be avoided when rating scale data are available.[34] 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 ORs and not the trials' absolute response rates has been criticised.[35] However, even the 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] also reported the SMD on symptom rating scales, which is more meaningful than the dichotomised outcomes.[33 34] 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 pairwise comparisons of 0.29 (95% CI 0.27 to 0.31, 390 comparisons, 253 studies) (table 1). These effect estimates are statistically significant, but likely below what could be considered a clinically relevant effect (box 1). We also calculated an overall mean difference for the trials that reported endpoint or change scores on the 17-item Hamilton depression rating scale, which was the most commonly used scale in the included trials (online 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 17-item Hamilton depression rating scale (range 0–52) (table 1). This mean difference on the Hamilton scale is likely also below what could be considered a clinically relevant effect (box 1). Cipriani *et al* did not discuss the clinical significance of their reported effect size.[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.[36] The majority of patients in a clinical setting would not be eligible to enter randomised trials due to such exclusion criteria,[37] and the evidence coming from these trials is therefore of limited relevance. Furthermore, the exclusion of previous 'non-responders' and inclusion of those who are expected to respond more favourably to treatment may bias the trials (box 1). These issues were not considered by Cipriani *et al*[5] but should arguably have resulted in downgrading of the confidence in the evidence in the GRADE domain of indirectness.[18]

The certainty of 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 issues related to the quality of the evidence, the certainty of evidence for the placebo comparisons should arguably be downgraded two levels due to a 'high risk' of bias, two levels in the domain of indirectness due to short trial lengths, strict inclusion criteria and the use of placebo run-in, in addition to downgrading one level due to publication bias as acknowledged by Cipriani *et al*.[5] Downgrading due to the indirectness of the network meta-analysis' methodology should also be considered.[38] Taken together, the certainty of evidence should be 'very low'.[18]

DISCUSSION

We have identified several important biases that were not taken into account in the systematic review by Cipriani *et al*.[5] We showed that the reported effect of antidepressants over placebo measured on depression rating scales was small and likely inflated by several methodological limitations in the trials. For the first time, we showed that the placebo run-in study design appears to work towards producing inflated effect sizes, in addition to publication bias and other methodological limitations. Further, we showed that the outcome data reported by Cipriani *et al* differed from the clinical study reports and that their risk of bias assessment did not follow the methods outlined in the Cochrane Handbook. Finally, we found

that the certainty of evidence for antidepressants versus placebo for all outcomes assessed should be very low. Taken together, the evidence does not support definitive conclusions regarding the efficacy of antidepressants for depression in adults, including whether they are more efficacious than placebo for depression.

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

We found that the evidence base consists of mainly short-term trials (12 weeks or less) with no evidence for treatment beyond 36 weeks, although most patients are treated for years.[3 27] Further, the apparent effect of antidepressants reported in the review by Cipriani *et al*[5] measured on investigator-rated symptom scales was small and likely not clinically relevant. Observational studies also indicate that the effectiveness of antidepressants in practice is very low: in the large, publicly funded, Sequenced Treatment Alternatives to Relieve Depression study, only 3% of the 4041 enrolled patients were considered 'in remission' after 1 year.[39] The recent finding based on clinical study reports of randomised trials that more participants drop out on antidepressants than on placebo,[31] further suggests that the benefits of antidepressants may not outweigh the harms.

Our findings showed that Cipriani *et al*'s data[5] were inaccurate, and their estimates may therefore be incorrect because they relied on published data. It may be perceived as a limitation that we relied on the data by Cipriani *et al* and did not perform our own separate systematic literature search and data extraction. Considering the multiple methodological limitations we have identified, it would be necessary to analyse data based on clinical study reports and individual patient data to make a reliable assessment of the benefits and harms of antidepressants because they are the most reliable source of trial data.[40] There are also some limitations to our sponsorship subgroup analysis: while industry sponsored studies have been found to report favourable efficacy results more often than non-industry sponsored studies,[26] our analysis showed that industry sponsored trials reported a lower effect estimate of antidepressants compared with placebo than non-industry sponsored trials on investigator-rated depression symptom scales. However, there were important differences between the two subgroups that likely contributed to the observed difference (online S1 figure). Non-industry sponsored trials were smaller and older than industry sponsored trials and almost all of the non-industry sponsored trials included by Cipriani *et al* were published.

Our results highlight that the many hundreds of placebo-controlled trials of antidepressants have not addressed the most important, patient-relevant questions regarding antidepressants' benefits and harms. Although this has been known for years,[13] it has not led to changes in research practice. Erroneous conclusions that antidepressants are efficacious for depression have the effect that they may prevent people suffering from depression from seeking other solutions to alleviate their condition, such as psychotherapy and dealing with psychosocial stressors, and they may stall funding and research of such treatment modalities. Importantly, such conclusions may also lead to a loss of interest in providing a better evidence base to determine the true clinical value of antidepressants.

Our review has two implications. First, the review by Cipriani *et al*[5] and its conclusion should be carefully revisited. In the light of our findings, the review should not inform clinical practice. Second, our reanalysis has highlighted the need for a radical change in the way antidepressant trials are being conducted, reported and interpreted. We hope that doctors, patients, peers and politicians will consider the limitations of the current evidence of antidepressants for depression that we have presented and collectively act accordingly. This involves informing the patients about the limitations of the current evidence, thus providing a basis for a true informed consent, and working towards a better evidence base for the use of antidepressants in the treatment of depression. To get reliable answers about the antidepressants' benefits and harms in adults with depression, we need large-scale, industry-independent and better blinded, long-term trials of drug naïve participants, with patient-relevant outcomes rather than ranking scales.

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

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.

Funding The study was funded by the Nordic Cochrane Centre. The research was designed, conducted, analysed and interpreted by the authors entirely independently of the funding source.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement All data files and the code for the statistical analyses are available from the Open Science Framework database (https://osf.io/3prz9/?view_only=0012d96af77e435f90e77755f728551a).

REFERENCES

1. World Health Organisation, Media Centre. Depression: Fact Sheet. 2017 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 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. 2017. Antidepressant Use Among Persons Aged 12 and Over: United States, 2011–2014. Available from: accessed 1 May 2018.
4. The Guardian. NHS prescribed record number of antidepressants last year [updated 29 June 2017]. Available from: www.theguardian.com/society/2017/jun/29/nhs-prescribed-record-number-of-antidepressants-last-year.
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;391:1357–66.
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. (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;17:58.
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:e45.
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;358:252–60.
10. The Guardian. The drugs do work: Antidepressants are effective, study shows: The Guardian. 2018 https://www.theguardian.com/science/2018/feb/21/the-drugs-do-work-antidepressants-are-effective-study-shows (accessed 1 May 2018).
11. Time Magazine. These antidepressants are most effective, study says Time Magazine. 2018 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.
13. Ioannidis JP. Effectiveness of antidepressants: an evidence myth constructed from a thousand randomized trials? *Philos Ethics Humanit Med* 2008;3:14.
14. Boesen K, Paludan-Müller AS, Munkholm K. Network meta-analysis of antidepressants. *Lancet* 2018;392:1011.
15. Gøtzsche PC, Jørgensen AW. Opening up data at the European Medicines Agency. *BMJ* 2011;342:d2686.
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 Med Res Methodol* 2014;14:25.
17. Higgins JPT, Green S, eds. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0*: The Cochrane Collaboration, 2011.
18. Schünemann H, Brożek J, Guyatt G, *et al*. GRADE handbook for grading quality of evidence and strength of recommendations. 2013. 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:e0159267.
20. Perlis RH, Ostacher M, Fava M, *et al*. Assuring that double-blind is blind. *Am J Psychiatry* 2010;167:250–2.
21. Moncrieff J, Wessely S, Hardy R. Active placebos versus antidepressants for depression. *Cochrane Database Syst Rev* 2004;1:CD003012.
22. Furukawa TA, Salanti G, Atkinson LZ, *et al*. Comparative efficacy and acceptability of first-generation and second-generation antidepressants in the acute treatment of major depression: protocol for a network meta-analysis. *BMJ Open* 2016;6:e010919.
23. Woolley SB, Cardoni AA, Goethe JW. Last-observation-carried-forward imputation method in clinical efficacy trials: review of 352 antidepressant studies. *Pharmacotherapy* 2009;29:1408–16.
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;86:343–58.
25. Chan AW, Hróbjartsson A, Haahr MT, *et al*. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA* 2004;291:2457–65.
26. Lundh A, Lexchin J, Mintzes B, *et al*. Industry sponsorship and research outcome. *Cochrane Database Syst Rev* 2017;2:MR000033.
27. 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;86:362–9.
28. Singh S, Loke YK. Drug safety assessment in clinical trials: methodological challenges and opportunities. *Trials* 2012;13:138.
29. 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.
30. 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:e005535.
31. Sharma T. *Effects of selective serotonin reuptake inhibitors (SSRIs) and serotonin - norepinephrine reuptake inhibitors (SNRIs) on suicidality, violence, and quality of life (PhD Thesis)*: University of Copenhagen, 2018.
32. Eiring Ø, 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:e007848.
33. Kirsch I, Moncrieff J. Clinical trials and the response rate illusion. *Contemp Clin Trials* 2007;28:348–51.
34. Altman DG, Royston P. The cost of dichotomising continuous variables. *BMJ* 2006;332:1080.1.
35. McCormack J, Korownyk C. Effectiveness of antidepressants. *BMJ* 2018;360:k1073.
36. Posternak MA, Zimmerman M, Keitner GI, *et al*. A reevaluation of the exclusion criteria used in antidepressant efficacy trials. *Am J Psychiatry* 2002;159:191–200.
37. Zetin M, Hoepner CT. Relevance of exclusion criteria in antidepressant clinical trials: a replication study. *J Clin Psychopharmacol* 2007;27:295–301.
38. Guyatt GH, Oxman AD, Kunz R, *et al*. GRADE guidelines: 8. Rating the quality of evidence--indirectness. *J Clin Epidemiol* 2011;64:1303–10.
39. 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;60:9–13.
40. Doshi P, Jefferson T, Del Mar C. The imperative to share clinical study reports: recommendations from the Tamiflu experience. *PLoS Med* 2012;9:e1001201.
41. Arroll B, Elley CR, Fishman T, *et al*. Antidepressants versus placebo for depression in primary care. *Cochrane Database Syst Rev* 2009;3:CD007954.
42. Fournier JC, DeRubeis RJ, Hollon SD, *et al*. Antidepressant drug effects and depression severity: a patient-level meta-analysis. *JAMA* 2010;303:47–53.
43. 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;69:572–9.
44. Fava GA, Gatti A, Belaise C, *et al*. Withdrawal Symptoms after Selective Serotonin Reuptake Inhibitor Discontinuation: A Systematic Review. *Psychother Psychosom* 2015;84:72–81.
45. Turner EH, Rosenthal R. Efficacy of antidepressants. *BMJ* 2008;336:516–7.
46. Leucht S, Fennema H, Engel R, *et al*. What does the HAMD mean? *J Affect Disord* 2013;148(2-3):243–8.
47. Moncrieff J, Kirsch I. Empirically derived criteria cast doubt on the clinical significance of antidepressant-placebo differences. *Contemp Clin Trials* 2015;43:60–2.