PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Considering the methodological limitations in the evidence base of antidepressants for depression: A reanalysis of a network meta-analysis
AUTHORS	Munkholm, Klaus; Paludan-Müller, Asger; Boesen, Kim

VERSION 1 - REVIEW

REVIEWER	Adrian Barnett Queensland University of Technology Australia
REVIEW RETURNED	18-Jul-2018

GENERAL COMMENTS	This is an interesting detailed examination of a published meta-analysis in an important area. A number of oversights are pointed out in the previous meta-analyses, and the authors examine some interesting additional analyses. The last line of the results in the abstract was that the "effect may even be negative". Although there were results in the main paper to back this up, there was nothing in the abstract on this and it is currently an isolated point that is not backed-up by any results. Also, the conclusion sentence felt too strong based on the abstract above. In my opinion the wording should be more around the doubt raised by these additional analysis. For the attrition bias, the cut-off used by Cipriani are labelled as "arbitrary". Whilst I agree, it's worth pointing out that arbitrary cut-off are frequently used in meta-analyses to stratify studies and look for potentially interesting patterns. These may be cut-offs in drop-out, calendar date, quality, etc. In the "Attrition bias" section you state, "This method is not in accordance with the Cochrane Handbook." It would be useful to know what the Cochrane method is. Not using LOCF is appropriate in my opinion (page 6), particularly given the many criticisms of it, see for example Kenward and Molenberghs doi: 10.1080/10543400903105406. The "vested interests" point on page 7 seems very important, but no analyses of this issue is presented. Could this not be another important comparison where the data are relatively easy to add?

	Similarly the issue of publication bias (page 7) seems very important. Could you not present an analysis on this too? "Misleading" on page 13 is a strong word that implies deliberate actions on behalf of the original authors. However, it is possible that the original authors had simply not considered the points made in this paper. Many meta-analyses are poorly done, and there has been lots of research on how meta-analyses are poorly executed, poorly reported, and key assumptions are often overlooked. Minor comments - Abstract, This sentence needs more explanation: "The certainty of evidence for placebo-controlled comparisons should be very low." Also, I would spell out the acronym ML in the next sentence. - New paragraph, page 6, line 20, starting "In total...". I was getting a little confused between the outcomes of the meta-analysis and the outcomes of the trials. - Page 6, line 40, in terms of the difference in attrition rates, was this because of a difference in the times examined, perhaps looking at early versus late follow-ups? - Can you share the data as a supplement in CSV or Excel format showing exactly how you categorised each study in terms of attrition? - Page 7, line 12, how was "moderate" bias defined? Also, although none of these methods are supported by "empirical evidence", it does have some face validity, in terms of examining a dose-response like relationship between greater bias and a difference mean treatment difference (if that's what Cipriani et al did). It's shocking that only one trial out of the hundreds was classified as "low risk of bias". - Page 7, line 43, add "likely" before "highly inflated"? - Page 10, line 30, is just one criteria enough to say that patients would prefer the placebo? - Page 12, first discussion paragraph, when you mention the potentially negative effect, it would be worth linking this back to the specific evidence you have for this.

REVIEWER	Jake Olivier School of Mathematics and Statistics University of New South Wales Sydney, Australia
REVIEW RETURNED	31-Jul-2018

GENERAL COMMENTS	The manuscript is a criticism of a recent network meta-analysis of studies of antidepressants used to treat depression. Some of the criticisms seem quite valid; however, I found it difficult to separate out objective criticisms from the authors' opinions. I think the general tone of the paper could be made from a more objective point of view and I think the authors should get to their main points earlier in the paper. Those who are agnostic to this topic, like myself, are likely to be more convinced of the validity of the criticisms when presented in a neutral tone. From my reading of the manuscript, the most convincing arguments are: (1) Previous meta-analysis found similar improvements in symptom scores, but concluded benefits were doubtful.

	(2) The methods for assessing bias used by Cipriani and colleagues have not been "supported by empirical evidence". That is, they made decisions that affected their analysis that are not widely accepted. This is problematic if different decisions of at least equal validity yield different results. (3) The effect sizes differed whether the trial had a placebo run-in or whether it was published. I find this argument to be the most convincing as publication bias is a well-known issue in drug trials. Other issues: What do the authors mean by pairwise meta-analysis? Was a meta-analysis performed for published studies and unpublished studies separately? Why not fit a meta-regression model with placebo run in (yes/no) and published (yes/no) as moderators? I think this better addresses the authors concerns. Can the authors define their standardised mean difference? Note that (mean1 - mean2)/sd is a biased estimator (I think this is in Hedges meta-analysis book). I think the R metafor package uses the unbiased estimator. Also, why not use maximum likelihood or restricted maximum likelihood for the meta-analysis model? The original study did a network meta-analysis, which I assume, is because there were 21 different antidepressants. They aren't all the same and would presumably have different levels of effectiveness and negative effects. It is unclear how this was accounted for in the re-analysis. A simple meta-analysis would assume all 21 antidepressants were essentially the same. I found it hard to follow the criticisms regarding risk of bias. Can the authors provide a summary table of what Cipriani reported, what the authors found using Cipriani's definitions, and how the authors believe the studies should have been rated. Some of this appears to be in the appendix, but I think it needs to be made clear towards the beginning of the paper. It's also unclear what type of bias is being assessed. For example, are assessments under the section "Selective outcome reporting" types of reporting bias? Although I have a lot of respect for the Cochrane Collaboration, not every journal or researcher adheres to their recommendations. I think it's a stretch to present the Cochrane Handbook as the "standard" for systematic reviews. It's clearly of very high quality and it's reasonable to compare, for example, GRADE versus what Cirpriani did. And, I think the authors have made their point that the Cipriani systematic review would not have met the Cochrane publication standards. But, I don't think it's correct to presume Cipriani should have followed the Cochrane Handbook or that The Lancet should have made them follow it. Is Table 1 really a table? It seems like subsections and text to me.

VERSION 1 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:

Reviewer: 1

Reviewer Name: Adrian Barnett

Institution and Country: Queensland University of Technology, Australia

Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below

This is an interesting detailed examination of a published meta-analysis in an important area. A number of oversights are pointed out in the previous meta-analyses, and the authors examine some interesting additional analyses.

- The last line of the results in the abstract was that the "effect may even be negative". Although there were results in the main paper to back this up, there was nothing in the abstract on this and it is currently an isolated point that is not backed-up by any results. Also, the conclusion sentence felt too strong based on the abstract above. In my opinion the wording should be more around the doubt raised by these additional analysis.

Response: Thank you for these valuable comments. We have deleted the mentioning of a possible negative effect of antidepressants from the abstract and also from the manuscript .

Regarding the abstract's conclusion, we have revised the wording to better reflect the uncertainty as pointed out by the reviewer. It now reads: "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."

- For the attrition bias, the cut-off used by Cipriani are labelled as "arbitrary". Whilst I agree, it's worth pointing out that arbitrary cut-off are frequently used in meta-analyses to stratify studies and look for potentially interesting patterns. These may be cut-offs in drop-out, calendar date, quality, etc.

Response: Thank you for this comment. We agree that arbitrary categories are (too) often used in research without any scientific or empirical justification, including in meta-analyses. Here, we are specifically wording a concern that their method of defining the "arbitrary cut-offs" in the attrition bias assessment was not in accordance with the Cochrane Handbook, although this was explicitly stated by the authors. We apologise for not clearly describing that one of the purposes of our evaluation of Cipriani et al.'s risk of bias assessment was to assess whether they had indeed followed the Cochrane Handbook as stated. We have added this information to the methods section, p. 5: "We evaluated whether Cipriani et al.'s risk of bias assessments were in accordance with the Cochrane Handbook,[3] as stated by the authors.[2] Where the approach differed we compared the risk of bias assessment by Cipriani et al.[2] with our reassessment following the Cochrane Handbook.[3]"

- In the "Attrition bias" section you state, "This method is not in accordance with the Cochrane Handbook." It would be useful to know what the Cochrane method is.

Response: We are thankful for this suggestion and we have now added more information regarding the method recommended in the Cochrane Handbook to the section, p. 6: "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.[3] Further, the authors did not consider the reasons for dropout, although this is also recommended by the Cochrane Handbook.[3]"

- Not using LOCF is appropriate in my opinion (page 6), particularly given the many criticisms of it, see for example Kenward and Molenberghs doi: 10.1080/10543400903105406.

Response: We agree with the peer-reviewer that the LOCF method is inappropriate. Cipriani et al. also considered the LOCF method inappropriate and, in keeping with their own arbitrary cut-offs for dropout, the use of LOCF should have led to many more trials being rated at high risk of attrition bias, than actually ended up being rated at high risk. We have added information to this section to make this apparent discrepancy more clear, p. 6: "Cipriani et al. characterised the last observation carried forward (LOCF) method as inappropriate,[4] 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."

- The "vested interests" point on page 7 seems very important, but no analysis of this issue is presented. Could this not be another important comparison where the data are relatively easy to add?

Response: Thank you for this suggestion. We very much agree that such an analysis could be interesting. We have conducted the suggested analysis and included the results in the section of "Other bias domain",

p. 7: "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)."

We have added a discussion of these results to the discussion section, p. 13: "There are also some limitations to our sponsorship subgroup analysis: In contrast to previous findings on the impact of sponsorship,[5] our analysis showed that industry-sponsored trials reported a lower effect estimate of antidepressants compared to 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 (S1 Figure): Non-industry sponsored trials were smaller and older than industry sponsored trials and almost all of the non-industry sponsored trials were published." We have also added a table (Table 1) with the results of our meta-analyses.

- Similarly, the issue of publication bias (page 7) seems very important. Could you not present an analysis on this too?

Response: The issue of publication is indeed very important. We conducted a meta-analysis on this and apologise that this was not entirely clear from the manuscript. We have now rephrased the particular section in the manuscript in order to better describe this, p. 8: "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)."

- "Misleading" on page 13 is a strong word that implies deliberate actions on behalf of the original authors.

However, it is possible that the original authors had simply not considered the points made in this paper. Many meta-analyses are poorly done, and there has been lots of research on how meta-analyses are poorly executed, poorly reported, and key assumptions are often overlooked.

Response: Thank you for this consideration of the wording of this part of the manuscript. We revised this part of the discussion section and deleted the word 'misleading'. The section now reads, p. 13: "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,[6] 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."

Minor comments

- Abstract, This sentence needs more explanation: "The certainty of evidence for placebo-controlled comparisons should be very low." Also, I would spell out the acronym ML in the next sentence.

Response: Thank you for this comment. We have taken the opportunity to provide more explanation in the abstract: "The certainty of the evidence for the placebo-controlled comparisons according to GRADE should be very low due to a high risk of bias, indirectness, and publication bias." Additionally, we have spelled out all the acronyms in the abstract.

- New paragraph, page 6, line 20, starting "In total...". I was getting a little confused between the outcomes of the meta-analysis and the outcomes of the trials.

Response: We apologise that this distinction was not clear from the manuscript. We have rephrased the two relevant sentences in the section, which now read, p. 7: "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.[3]"

- Page 6, line 40, in terms of the difference in attrition rates, was this because of a difference in the times examined, perhaps looking at early versus late follow-ups?

Response: Thank you for raising this issue regarding different attrition rates between the treatment arms. It would indeed be relevant to investigate when the dropout occurs; however, these data were not available in the Cipriani et al. data files or in their appendices. This information was most probably not reported in the original trial publications either and it would likely require access to the clinical study reports or the individual patient data to obtain this information.

- Can you share the data as a supplement in CSV or Excel format showing exactly how you categorised each study in terms of attrition?

Response: Because Cipriani et al. did not provide information on the imputation method used in each trial as part of their dataset, we were not able to apply their criteria for attrition bias assessments in our own categorisation. We have now clarified this, p. 6: “Cipriani et al. characterised the last observation carried forward (LOCF) method as inappropriate,[4] 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.”

The absolute dropout rates that we refer to in our own calculation of the “unbalanced” dropout rates between the arms, using the cut-offs defined by Cipriani et al., are part of the online data we make available with the article on the Open Science Framework database (please see data sharing section of the manuscript for a link to the data).

- Page 7, line 12, how was "moderate" bias defined? Also, although none of these methods are supported by "empirical evidence", it does have some face validity, in terms of examining a dose-response like relationship between greater bias and a difference mean treatment difference (if that's what Cipriani et al did). It's shocking that only one trial out of the hundreds was classified as "low risk of bias".

Response: We agree with the reviewer that the lack of trials that are of low risk of bias is a very important point and it emphasises the importance of using transparent and valid methods for the assessment of the overall risk of bias. In response to the reviewer’s suggestion, we have added a description of the criteria for the overall risk of bias assessment used by Cipriani et al., including their definition of “moderate” risk of bias, p. 8: “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.[2] 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.[3] The Handbook instead recommends an overall qualitative assessment considering the relative importance of different domains.[3]”

- Page 7, line 43, add "likely" before "highly inflated"?

Response: Thank you for this suggestion. We have changed the sentence, which now reads, p. 8: “This indicates that the reported effect sizes by Cipriani et al.[2] are likely inflated due to publication bias.”

- Page 10, line 30, is just one criteria enough to say that patients would prefer the placebo?

Response: Thank you for this comment. The measure of total dropouts is a pragmatic measure of the balance between benefits and harms by a drug as weighed by the participants. However, during the revision we have deleted the sentence entirely.

- Page 12, first discussion paragraph, when you mention the potentially negative effect, it would be worth linking this back to the specific evidence you have for this.

Response: Thank you for this comment. We have deleted the mentioning of a negative effect in the manuscript. The concluding sentence of the first paragraph of the discussion section therefore now reads, p. 12: “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.”

Reviewer: 2

Reviewer Name: Jake Olivier

Institution and Country: School of Mathematics and Statistics, University of New South Wales, Sydney, Australia

Please state any competing interests or state ‘None declared’: None declared

Please leave your comments for the authors below

- The manuscript is a criticism of a recent network meta-analysis of studies of antidepressants used to treat depression. Some of the criticisms seem quite valid; however, I found it difficult to separate out objective criticisms from the authors' opinions. I think the general tone of the paper could be made from a more objective point of view and I think the authors should get to their main points earlier in the paper. Those who are agnostic to this topic, like myself, are likely to be more convinced of the validity of the criticisms when presented in a neutral tone.

Response: Thank you for these valuable comments. With this in mind, we have taken the opportunity to make alterations to the manuscript throughout in order to clearly communicate in which instances our statements are subjective and also to communicate our findings in a neutral and balanced tone.

- From my reading of the manuscript, the most convincing arguments are:

(1) Previous meta-analysis found similar improvements in symptom scores, but concluded benefits were doubtful.

(2) The methods for assessing bias used by Cipriani and colleagues have not been "supported by empirical evidence". That is, they made decisions that affected their analysis that are not widely accepted. This is problematic if different decisions of at least equal validity yield different results.

(3) The effect sizes differed whether the trial had a placebo run-in or whether it was published. I find this argument to be the most convincing as publication bias is a well-known issue in drug trials.

Response: Thank you for these comments. We agree that these are some of the most important findings.

Other issues:

- What do the authors mean by pairwise meta-analysis? Was a meta-analysis performed for published studies and unpublished studies separately? Why not fit a meta-regression model with placebo run in (yes/no) and published (yes/no) as moderators? I think this better addresses the authors concerns.

Response: Thank you for these comments and suggestions. We refer to pairwise meta-analysis as metaanalyses that involve direct comparisons, in contrast to network meta-analysis that involves indirect comparisons of multiple agents. Regarding our subgroup analysis according to publication status, a subgroup analysis may be more intuitively communicated for categorical moderators with only two or three variables than a meta-regression model and we have therefore chosen this method. Furthermore, a meta-regression model, as proposed by the peer-reviewer, yields identical results, as seen from the output below (R output of meta-regression model with publication status as a moderator variable).

Mixed-Effects Model (k = 390; tau^2 estimator: SJ)

tau^2 (estimated amount of residual heterogeneity): 0.03 (SE = 0.00)

tau (square root of estimated tau^2 value): 0.19 I^2 (residual heterogeneity / unaccounted variability): 57.33%

H^2 (unaccounted variability / sampling variability): 2.34

R^2 (amount of heterogeneity accounted for): 10.24%

Test for Residual Heterogeneity:

QE(df = 388) = 581.79, p-val < .01

Test of Moderators (coefficient(s) 2):

F(df1 = 1, df2 = 388) = 46.55, p-val < .01

Model Results:

estimate se tval pval ci.lb ci.ub intrcpt -0.33 0.01 -26.36 <.01 -0.35 -0.30 *** publication_statusunpublished 0.18 0.03 6.82 <.01 0.13 0.23 *** ---

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

- Can the authors define their standardised mean difference? Note that (mean1 - mean2)/sd is a biased estimator (I think this is in Hedges meta-analysis book). I think the R metafor package uses the unbiased estimator. Also, why not use maximum likelihood or restricted maximum likelihood for the meta-analysis model?

Response: We apologise for not specifying our estimator of standardised mean difference. We did indeed use the bias corrected Hedges' g and have added this information to the methods section, p. 5: "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) as Hedges' g with corresponding 95% confidence intervals (95% CI)."

With regards to the estimator of between-study variance, we used the Hartung-Knapp-Sidik-Jonkman (HKSJ) approach because it results in fewer type I errors than the DerSimonian and Laird approach. There are a number of possible estimators, among which is the REML approach; however, given previous recommendations we used the HKSJ method.[7]

- The original study did a network meta-analysis, which I assume, is because there were 21 different antidepressants. They aren't all the same and would presumably have different levels of effectiveness and negative effects. It is unclear how this was accounted for in the re-analysis. A simple meta-analysis would assume all 21 antidepressants were essentially the same.

Response: Thank you for noting this issue of analysing the combined data for all antidepressants. There may indeed be differences in the benefits and harms of the included drugs. We did provide an overall effect size for the combined group of antidepressants versus placebo as an SMD and as a mean difference measured on the 17-item Hamilton depression rating scale, as Cipriani et al.[2] similarly did for their continuous outcome measure (SMD). Our analyses showed that the heterogeneity for those analyses was relatively low, with an I^2 of 40% and 27%, respectively. However, our focus was to investigate how flaws impact the apparent effect size for antidepressants and these are likely relevant to all of the antidepressants. Most importantly though, the outcomes that could help identifying the clinically relevant differences between the drugs were not assessed by Cipriani et al, such as the drugs' actual harms, serious adverse events, participant-rated symptom scales, and quality of life. Furthermore, the basic underlying assumptions for including studies in a single pair-wise meta-analysis are not different than in a network meta-analysis.

- I found it hard to follow the criticisms regarding risk of bias. Can the authors provide a summary table of what Cipriani reported, what the authors found using Cipriani's definitions, and how the authors believe the studies should have been rated. Some of this appears to be in the appendix, but I think it needs to be made clear towards the beginning of the paper. It's also unclear what type of bias is being assessed. For example, are assessments under the section "Selective outcome reporting" types of reporting bias?

Response: Thank you for raising these important issues. We apologise for not making our results regarding the risk of bias assessments sufficiently clear. We have restructured the section, so that it follows the order of the risk of bias domains outlined in the Cochrane Handbook and we have provided an introductory description of the types of bias assessed within the individual domains, p. 5: "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.[3]"

We have tried to describe more clearly which method was used by Cipriani et al., which method is outlined in the Cochrane Handbook, and where there is room for interpretation. We considered making a summary table as suggested by the peer-reviewer but we have opted not to do so because it would not provide the necessary room for detail. Instead, we have tried to separate the assessments made by Cipriani et al. and their methods, and the corresponding recommendations in the Cochrane Handbook to improve the overall readability.

- Although I have a lot of respect for the Cochrane Collaboration, not every journal or researcher adheres to their recommendations. I think it's a stretch to present the Cochrane

Handbook as the "standard" for systematic reviews. It's clearly of very high quality and it's reasonable to compare, for example, GRADE versus what Cipriani did. And, I think the authors have made their point that the Cipriani systematic review would not have met the Cochrane publication standards. But, I don't think it's correct to presume Cipriani should have followed the Cochrane Handbook or that The Lancet should have made them follow it.

Response: Thank you for these valuable comments. It is correct that there are alternative options for risk of bias assessments and for evaluating the certainty of evidence. However, we found it necessary to evaluate whether Cipriani et al. used the method as described in the Cochrane Handbook because the authors explicitly stated to do so. As this is an important point, we have made this clearer from our description of our aims, highlighting that we sought to investigate whether their methods followed the Cochrane Handbook recommendations, p. 4: “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 have also added information to the beginning of the discussion section, where we outline our main findings, p. 12: “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.”

- Is Table 1 really a table? It seems like subsections and text to me.

Response: Thank you for pointing this out. While Table 1 is perhaps not a standard table, we consider it a meaningful way of providing background information on some of the important methodological issues. Furthermore it increases the readability of the manuscript.

VERSION 2 – REVIEW

REVIEWER	Adrian Barnett Queensland University of Technology Australia
REVIEW RETURNED	17-Sep-2018

GENERAL COMMENTS	This paper highlights frequent problems with randomised trials that examine the benefits of anti-depressants. The authors highlight several problems that mean that the published results likely over-estimate the benefits of treatment. This is troubling given the widespread use of these drugs, and this is an important paper that should be published. The authors have had a letter accepted in the same journal as the meta-analyses critiqued in this paper. They state that this paper is more detailed, which is likely given that this is a full paper and the other is a letter. However, it would have been useful to have included their letter as a supplementary file. Minor comments - Section "Blinding of participants, personnel and outcome assessment": "Given these issues, all placebo-controlled trials should arguably be categorised as at least unclear, or perhaps

	even at high risk of bias." Should say "trials of antidepressants", otherwise this sentence could be taken out of context. - "According to our knowledge, none of their methods are supported by empirical evidence." I can't grasp the meaning of this sentence, and I'm not sure it's needed given that the rest of the paragraph well describes the approaches taken, the problems with those approaches, and some updated results. - "However, that figure was misleading because only four of these 12 trials contained". I would use 'incorrect' rather than 'misleading'.

VERSION 2 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:

Reviewer: 1

Reviewer Name: Adrian Barnett

Institution and Country: Queensland University of Technology, Australia

Please state any competing interests or state ‘None declared’: None declared Please leave your comments for the authors below

This paper highlights frequent problems with randomised trials that examine the benefits of antidepressants. The authors highlight several problems that mean that the published results likely overestimate the benefits of treatment. This is troubling given the widespread use of these drugs, and this is an important paper that should be published.

The authors have had a letter accepted in the same journal as the meta-analyses critiqued in this paper. They state that this paper is more detailed, which is likely given that this is a full paper and the other is a letter. However, it would have been useful to have included their letter as a supplementary file.

Response: We apologise for not enclosing the accepted letter with our revised manuscript. The letter is now published in The Lancet1 and is freely available online on The Lancet’s website (https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(18)31783-5/fulltext).

Minor comments

- Section "Blinding of participants, personnel and outcome assessment": "Given these issues, all placebocontrolled trials should arguably be categorised as at least unclear, or perhaps even at high risk of bias." Should say "trials of antidepressants", otherwise this sentence could be taken out of context.

Response: Thank you for this comment. We have changed the sentence accordingly, and it now reads, p. 6: “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.”

- "According to our knowledge, none of their methods are supported by empirical evidence." I can't grasp the meaning of this sentence, and I'm not sure it's needed given that the rest of the paragraph well describes the approaches taken, the problems with those approaches, and some updated results.

Response: Thank you for this comment. We agree the sentence is not needed and have now deleted the sentence entirely.

- "However, that figure was misleading because only four of these 12 trials contained". I would use 'incorrect' rather than 'misleading'.

Response: Based on the reviewer's comment we have changed the wording of the sentence, removing the word "misleading" and the judgement it implied. The sentence now reads, p. 9: "However, we found that only four of these 12 trials contained an uninterrupted double-blind, placebo-controlled phase of more than 12 weeks (S2 Appendix)."

References

1. 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]

VERSION 3 - REVIEW

REVIEWER	Adrian Barnett Queensland University of Technology Australia
REVIEW RETURNED	22-Oct-2018

GENERAL COMMENTS	Thanks for these final few changes. I can see that the letter in the Lancet is very short, and this paper goes into far more detail.