PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

This paper was submitted to a another journal from BMJ but declined for publication following peer review. The authors addressed the reviewers' comments and submitted the revised paper to BMJ Open. The paper was subsequently accepted for publication at BMJ Open.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Investigation of bias in meta-analyses due to selective inclusion of trial effect estimates: empirical study
AUTHORS	Page, Matthew; Forbes, Andrew; Chau, Marisa; Green, Sally; McKenzie, Joanne

VERSION 1 - REVIEW

REVIEWER	Tendal, Britta The Nordic Cochrane Centre
REVIEW RETURNED	19-Mar-2015

GENERAL COMMENTS	Interesting study with results that are relevant for both readers and makers of systematic reviews. Abstract: Clear and the conclusion is in line with the results. 'Effects' maybe you should use the word 'results' instead Methods: Clear, mostly easy to follow Good with a reference to the protocol for the study. 'Screening ceased once 44 eligible reviews were included' Sorted by pub date? Explain a bit more about the handling of meta-analyses were the number of possible combinations was too high to calculate. This part can be a bit tricky to understand for many I think (also the related table 3) Results: Clear, mostly easy to follow 'Impact of potential selective…': The first part is a bit hard to read Table 3 see earlier comment. Discussion: Good discussion of strengths and limitations, addressing potential risk of bias in the study. It is not uncommon to include unpublished or incomplete data (by imputing for example missing SDs), might be an idea to look at this in future research, as this might increase the range of possible results to extract and thereby increase the risk of bias in selection of results in those cases.

REVIEWER	BOUTRON, Isabelle INSERM APHP University Paris Descartes I am member of the Cochrane methods executive with one of the author (McKensie Joanne)

REVIEW RETURNED	05-Apr-2015

GENERAL COMMENTS	This study explores an important issue: the selective inclusion of treatment effect estimates in meta-analysis. This issue has been neglected. Several studies explored publication bias, outcome reporting bias, but none the selective inclusion of effect estimates in meta-analysis. Nevertheless, researchers are frequently faced with multiple effect estimates reported in RCTs. The authors systematically tested for evidence of selective inclusion of effect estimates. They used a sample of 31 systematic reviews including 250 trials in the field of rheumatology and psychiatry. In their sample, they did not find any evidence of selective inclusion bias. The manuscript is very well written, the methods are straightforward and appropriate: a protocol has been published, all deviations from protocol are reported in an appendix; the interpretation of results is appropriate and all limitations are adequately reported. I only have few comments: 1) It was not clear how the decision rules reported in the systematic reviews were taken into account. Did the authors extract only effect estimates according to the decision rules reported in the protocol or methods section or did they extract all effect estimates? This should be clarified and discussed. 2) Another important result of this study is the lack of decision rules in most systematic reviews while multiplicity is frequent. This should be highlighted in the abstract. 3) the sample size was limited to 44 eligible systematic reviews of which 31 were included in the analysis. What was the rational for this sample size? Why not selecting all systematic reviews published in the 2-year period? Was the power to detect bias too limited? 4) Only one researcher extracted all data. Another researcher evaluated 14 reviews of which 8 were included in the analysis. The authors should be more explicit on the discrepancies between the 2 data extractions. 5) Only 2 systematic reviews were industry funded. Further, 61% evaluated non-pharmacologic treatments. It would be important to expand the discussion on this point as the risk of selective inclusion could be higher in industry funded systematic reviews.

REVIEWER	Reeves, Barnaby University of Bristol, Bristol Heart Institute I was a co-applicant on a grant awarded by the Cochrane Collaboration to develop a tool to assess the risk of bias in non-randomised studies of the effects of interventions. I coordinated a working group on the selective reporting bias domain. I did not gain financially from this award but the award paid for me to run workshops to disseminate the new tool.
REVIEW RETURNED	06-Apr-2015

GENERAL COMMENTS	Investigation of bias in meta-analyses due to selective inclusion of trial effect estimates: empirical study This paper describes a very carefully conducted methodological

	systematic review. There is now quite good evidence about how the validity of findings from RCTs is threatened by primary authors selecting the results to report on the basis of their direction, magnitude or statistical significance. The paper reports a methodological systematic review to investigate whether the same happens in systematic reviews, when review authors have a range of effect estimates to choose from when extracting data for a meta-analysis. If this were the case, then meta-analyses would be generate biased estimates of the benefits (or harms) of the intervention(s) being reviewed. This threat should already be addressed by specifying in the protocol in sufficient detail how a treatment effect will be chosen when multiple estimates are available. However, the protocols often do not specify sufficient details or anticipate all of the opportunities for multiplicity or a protocol for a review does not exist. I think this topic, i.e. selective reporting by review authors, is extremely topical and important since reviews form a body of evidence that has privileged status and the opportunity for introducing bias in the reviewing process may be overlooked. General comment: 1. The paper investigates selective inclusion of treatment effects in just two (three) disease areas: arthritis (osteo or rheumatoid) and depression/anxiety. The authors justify these choices in terms of their interest in some subgroup comparisons but I have a general reservation relates about the extent to which readers should be reassured about selective inclusion of treatment effects in meta-analyses (since the paper reports that, on average, review authors do not appear to be selecting treatment effects of inclusion in a meta-analysis). Set against my reservation are the facts that: (a) it is the first empirical evidence about whether selective inclusion of treatment effects in meta-analyses is something that readers should worry about; (b) despite focusing on a limited number of diseases areas, the trials covered diverse research questions (e.g. pharmacological vs. non-pharmacological interventions); (c) the authors mention that the findings may have limited generalisability (both disease area and types of outcome included). However, (c) is rather brief (4 lines) and doesn't give me any 'feel' about how I should interpret the next meta-analysis I pick up. I would have been particularly interested in an additional (post-hoc) description/subgroup comparison relating to pharmacological vs. non-pharmacological interventions. I agree about the importance of extending research on this topic to binary outcomes and to other disease areas. If additional research were to be done, it would be ideal if it could follow the same (or very similar protocol) so that multiple methodological reviews could be meta-meta-analysed! Minor comments/criticisms: I have only minor comments and criticisms on the manuscript. 2. For setting the scene to the reader, I was quite surprised that the authors did not draw the analogy with selective reporting [inclusion] of treatment effects in RCTs when primary researchers carry out multiple analyses of the same outcome. 3. Page 5, first para. It was not clear to me (at this point) whether

	treatment effects were included when primary researchers merely reported sufficient data for intervention and control groups to allow a treatment effect to be derived but did not report a treatment. (I think this was one of the sensitivity analyses.) 4. Page 5, end of middle para. I think “timing of measurement” should be included in the list (in parentheses) of reasons for multiplicity. 5. Page 6, second para (Data extraction). I accept that double data extraction would have been impracticable for this review. Nevertheless, I wanted more detail of the discrepancies for the subsample of data that were extracted twice. It’s not clear what resolving discrepancies for this particular sample achieves – potentially better ‘final’ included data for the subsample but, apparently, no impact at all on the majority of the data. Some of the discrepancies are explained later but not attributed to the main data extractor or the second data extractor. For example, did the action of resolving discrepancies have any impact on the first reviewer’s extraction of data for other studies (either because he was midway in the data extraction or because some of the other data were rechecked)? 6. Page 7, first line. Just in terms of clarification, did the authors actually extract “outcome data” or treatment effects? I think this is probably the answer to my query 3 above (yes, outcome data) – but does this mean that the treatment effects in this review may not have matched the treatment effects reported in the primary papers? Is the argument that the original review authors are likely to have extracted outcome data? I am not sure that this is the case for continuous outcomes, potentially reported for different instruments (and potentially requiring specification of treatment effects and variances for the meta-analysis). 7. Page 9, second para. I can see that the two pre-specified subgroups are probably the most important but was disappointed not to see other subgroup comparisons just for descriptive / hypothesis generating purposes (potentially to help me with my question about generalisability). These might have included Cochrane vs. non-Cochrane (presumably a lot of overlap with protocol vs. no protocol) and drug vs. non-drug intervention. 8. Page 10. Table 1. Are there any issues around ceiling/floor effects that need to be discussed? This table is the first time I began to get a feel for the nature of the outcomes reported in the RCTs – and they made me wonder whether the continuous outcomes were actually discrete and treatment effects constrained by end-of-scale problems… and whether these factors might have limited the opportunity to identify selective inclusion of estimates. 9. Page 10. Table 1. Couldn’t the authors work out whether the meta-analyses were fixed or random effects where the information was not reported? 10.

	Page 11. Para under Table 1. I found it quite difficult to understand the information reported here. I expect it's all correct – but, for example, I found myself trying to resolve the facts that 47% of trials had multiplicity and that a median of 50% of trials per meta-analysis had multiplicity. 11. Page 12, second para. I think the first sentence should be at the start of the Discussion (I think it is as well), not here. 12. Page 13, Table 2. I assume that the statistical tests of "subgroup difference" are tests of interaction. If yes, perhaps this could be made clearer. 13. Page 13, Table 2. The subgroup comparison with respect to core outcome set is completely confounded with topic (OA/RA vs. anxiety/depression). Can the authors please comment about this? 14. Page 16, para 1. The authors do not provide any data to support their assertion that "our results are unlikely to have been notably affected by data extraction errors." 15. Page 17, para 1. I view the post-hoc analysis of 'extent of multiplicity' as another subgroup comparison. The test result is interesting but does not appear to show a trend to the null (i.e. to PBI=0.5). I note that the amount of data diminishes with increasing degrees of multiplicity. Do the authors feel confident about this result? 16. Page 17, para 2. I think that this paragraph is perhaps more important than it appears – since I think it is very unlikely that a review author will select a result on the basis of its weight in a meta-analysis. 17. Page 17, para 3. Please revise the first sentence so that the sentence affirms a more positive position about the need to pre-define methods for selecting effect estimates when multiple estimates from a trial are reported. (e.g. "Systematic reviewers should continue to pre-define …") 18. Page 18, para 1, line 2. I would prefer "defined/described" to "reported". 18. Page 18, para 2. This recommendation is excellent in principle – but I worry that it means a lot more work for review authors facing every increasing pressures to carry out reviews more efficiently … and longer reviews. What is the authors response? Figures No comments on the main figures. But I would like the supplementary figure 1 to include labels on/near each blob describing the number of trials. I am still not completely sure that I

	understand what the data represent, e.g. trial effect estimates per trial report, per meta-analysis, etc.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Interesting study with results that are relevant for both readers and makers of systematic reviews.
Thank you. No response required.

Abstract:
Clear and the conclusion is in line with the results.
Thank you. No response required.

'Effects' maybe you should use the word 'results' instead
We believe "results" is ambiguous so have used "effect estimates" or "effects" throughout the manuscript.

Methods:
Clear, mostly easy to follow. Good with a reference to the protocol for the study.
Thank you. No response required.

'Screening ceased once 44 eligible reviews were included' Sorted by pub date?
No, screening was not sorted by publication date.

Explain a bit more about the handling of meta-analyses were the number of possible combinations was too high to calculate. This part can be a bit tricky to understand for many I think (also the related table 3)
We have added to page 9, paragraph 1 an explanation of how we handled the meta-analyses where the number of possible combinations was too high to calculate. This text is briefly repeated as a footnote to Table 3.

Results:
Clear, mostly easy to follow
Thank you. No response required.

'Impact of potential selective…': The first part is a bit hard to read
We have edited this section to increase readability (see page 14, paragraph 1).

Table 3 see earlier comment.
As stated above, we have added to page 9, paragraph 1 an explanation of how we handled the meta-analyses where the number of possible combinations was too high to calculate. This text is briefly repeated as a footnote to Table 3.

Discussion:
Good discussion of strengths and limitations, addressing potential risk of bias in the study.
Thank you. No response required.

It is not uncommon to include unpublished or incomplete data (by imputing for example missing SDs), might be an idea to look at this in future research, as this might increase the range of possible results to extract and thereby increase the risk of bias in selection of results in those cases.
We agree, and have added this suggestion for further research to page 19, paragraph 3.

Reviewer: 2

This study explores an important issue: the selective inclusion of treatment effect estimates in meta-analysis.
This issue has been neglected. Several studies explored publication bias, outcome reporting bias, but

none the selective inclusion of effect estimates in meta-analysis. Nevertheless, researchers are frequently faced with multiple effect estimates reported in RCTs.
The authors systematically tested for evidence of selective inclusion of effect estimates. They used a sample of 31 systematic reviews including 250 trials in the field of rheumatology and psychiatry. In their sample, they did not find any evidence of selective inclusion bias.
The manuscript is very well written, the methods are straightforward and appropriate: a protocol has been published, all deviations from protocol are reported in an appendix; the interpretation of results is appropriate and all limitations are adequately reported.
Thank you. No response required.

I only have few comments:
1) It was not clear how the decision rules reported in the systematic reviews were taken into account. Did the authors extract only effect estimates according to the decision rules reported in the protocol or methods section or did they extract all effect estimates? This should be clarified and discussed.
We have expanded our description of how decision rules were used to select effect estimates on page 7, paragraph 2.

2) Another important result of this study is the lack of decision rules in most systematic reviews while multiplicity is frequent. This should be highlighted in the abstract.
We agree that this result is important but have not included it in the abstract because of the 300 word limit of the journal.

3) The sample size was limited to 44 eligible systematic reviews of which 31 were included in the analysis. What was the rational for this sample size? Why not selecting all systematic reviews published in the 2-year period? Was the power to detect bias too limited?
As this is the first study of its kind, no calculations are available to estimate the power required to detect evidence of selective inclusion bias. We acknowledge this as a limitation of the study on page 16, paragraph 4 to page 17, paragraph 1. The sample size estimation was calculated for a companion study which estimated the extent of multiplicity in trials (http://www.ncbi.nlm.nih.gov/pubmed/25841706), and was pre-specified in our protocol.

4) Only one researcher extracted all data. Another researcher evaluated 14 reviews of which 8 were included in the analysis. The authors should be more explicit on the discrepancies between the 2 data extractions.
We have added a description of the data extraction discrepancies to page 10, paragraph 1.

5) Only 2 systematic reviews were industry funded. Further, 61% evaluated non-pharmacologic treatments. It would be important to expand the discussion on this point as the risk of selective inclusion could be higher in industry funded systematic reviews.
We have added to page 19, paragraph 3 a suggestion that it is worth exploring in future research whether selective inclusion is more common in reviews funded by the sponsor of the product under investigation.

Reviewer: 3

This paper describes a very carefully conducted methodological systematic review. There is now quite good evidence about how the validity of findings from RCTs is threatened by primary authors selecting the results to report on the basis of their direction, magnitude or statistical significance. The paper reports a methodological systematic review to investigate whether the same happens in systematic reviews, when review authors have a range of effect estimates to choose from when extracting data for a meta-analysis. If this were the case, then meta-analyses would be generate biased estimates of the benefits (or harms) of the intervention(s) being reviewed. This threat should already be addressed by specifying in the protocol in sufficient detail how a treatment effect will be chosen when multiple estimates are available. However, the protocols often do not specify sufficient details or anticipate all of the opportunities for multiplicity or a protocol for a review does not exist. I think this topic, i.e. selective reporting by review authors, is extremely topical and important since reviews form a body of evidence that has privileged status and the opportunity for introducing bias in the reviewing process may be overlooked.

Thank you. No response required.

General comment:
1.
The paper investigates selective inclusion of treatment effects in just two (three) disease areas: arthritis (osteo or rheumatoid) and depression/anxiety. The authors justify these choices in terms of their interest in some subgroup comparisons but I have a general reservation relates about the extent to which readers should be reassured about selective inclusion of treatment effects in meta-analyses (since the paper reports that, on average, review authors do not appear to be selecting treatment effects of inclusion in a meta-analysis). Set against my reservation are the facts that: (a) it is the first empirical evidence about whether selective inclusion of treatment effects in meta-analyses is something that readers should worry about; (b) despite focusing on a limited number of diseases areas, the trials covered diverse research questions (e.g. pharmacological vs. non-pharmacological interventions); (c) the authors mention that the findings may have limited generalisability (both disease area and types of outcome included). However, (c) is rather brief (4 lines) and doesn't give me any 'feel' about how I should interpret the next meta-analysis I pick up. I would have been particularly interested in an additional (post-hoc) description/subgroup comparison relating to pharmacological vs. non-pharmacological interventions. I agree about the importance of extending research on this topic to binary outcomes and to other disease areas. If additional research were to be done, it would be ideal if it could follow the same (or very similar protocol) so that multiple methodological reviews could be meta-meta-analysed!
Thank you. We have expanded our recommendations for future research in line with the suggestions made above.

Minor comments/criticisms:

I have only minor comments and criticisms on the manuscript.

2.
For setting the scene to the reader, I was quite surprised that the authors did not draw the analogy with selective reporting [inclusion] of treatment effects in RCTs when primary researchers carry out multiple analyses of the same outcome.
We have added this analogy to page 4, paragraph 1.

3.
Page 5, first para. It was not clear to me (at this point) whether treatment effects were included when primary researchers merely reported sufficient data for intervention and control groups to allow a treatment effect to be derived but did not report a treatment. (I think this was one of the sensitivity analyses.)
See response to comment 6 below.

4.
Page 5, end of middle para. I think "timing of measurement" should be included in the list (in parentheses) of reasons for multiplicity.
We have not included "timing of measurement" in this list, because multiple time points may be available in trials of dichotomous outcomes; the examples in the parentheses are examples of multiplicity that are unique to continuous outcomes.

5.
Page 6, second para (Data extraction). I accept that double data extraction would have been impracticable for this review. Nevertheless, I wanted more detail of the discrepancies for the subsample of data that were extracted twice. It's not clear what resolving discrepancies for this particular sample achieves – potentially better 'final' included data for the subsample but, apparently, no impact at all on the majority of the data. Some of the discrepancies are explained later but not attributed to the main data extractor or the second data extractor. For example, did the action of resolving discrepancies have any impact on the first reviewer's extraction of data for other studies (either because he was midway in the data extraction or because some of the other data were rechecked)?
We have added a description of the data extraction discrepancies to page 10, paragraph 1.

6.
Page 7, first line. Just in terms of clarification, did the authors actually extract "outcome data" or treatment effects? I think this is probably the answer to my query 3 above (yes, outcome data) – but does this mean that the treatment effects in this review may not have matched the treatment effects reported in the primary papers? Is the argument that the original review authors are likely to have extracted outcome data? I am not sure that this is the case for continuous outcomes, potentially reported for different instruments (and potentially requiring specification of treatment effects and variances for the meta-analysis).
We have clarified on page 7, paragraph 2 that we extracted outcome data (i.e. means and standard deviations) as well as treatment effects if both were reported.

7.
Page 9, second para. I can see that the two pre-specified subgroups are probably the most important but was disappointed not to see other subgroup comparisons just for descriptive / hypothesis generating purposes (potentially to help me with my question about generalisability). These might have included Cochrane vs. non-Cochrane (presumably a lot of overlap with protocol vs. no protocol) and drug vs. non-drug intervention.
We have not conducted a subgroup analysis of Cochrane versus non-Cochrane reviews, as this overlaps almost completely with the protocol versus no protocol subgroup analysis. Also, we cannot think of any reason why selective inclusion should be more/less common in drug versus non-drug reviews; we believe it is more to do with the type of outcome rather than the type of intervention. Therefore we have decided not to perform this post-hoc analysis.

8.
Page 10. Table 1. Are there any issues around ceiling/floor effects that need to be discussed? This table is the first time I began to get a feel for the nature of the outcomes reported in the RCTs – and they made me wonder whether the continuous outcomes were actually discrete and treatment effects constrained by end-of-scale problems… and whether these factors might have limited the opportunity to identify selective inclusion of estimates.
We think ceiling/floor effects would only hamper our ability to identify selective inclusion of estimates if all of the available estimates for an outcome were highly correlated (i.e. all had ceiling effects). This would mean there would be little difference in the magnitude of each of the effects, so ultimately it would not matter which estimate the systematic reviewer selected. We have discussed this issue on page 18, paragraph 2.

9.
Page 10. Table 1. Couldn't the authors work out whether the meta-analyses were fixed or random effects where the information was not reported?
We cannot work out which model was used, because when we analysed the data using both a fixed-effect and random-effects model, each model produced the same SMD and 95% CI (we have noted this in a footnote under Table 1).

10.
Page 11. Para under Table 1. I found it quite difficult to understand the information reported here. I expect it's all correct – but, for example, I found myself trying to resolve the facts that 47% of trials had multiplicity and that a median of 50% of trials per meta-analysis had multiplicity.
We have edited this paragraph (page 12, paragraph 1) to enhance readability.

11.
Page 12, second para. I think the first sentence should be at the start of the Discussion (I think it is as well), not here.
We have removed the first sentence as suggested (see page 13, paragraph 1).

12.
Page 13, Table 2. I assume that the statistical tests of "subgroup difference" are tests of interaction. If yes, perhaps this could be made clearer.
We have clarified in Table 2 that P-values are from a test of interaction.

13.

Page 13, Table 2. The subgroup comparison with respect to core outcome set is completely confounded with topic (OA/RA vs. anxiety/depression). Can the authors please comment about this?
We have acknowledged this as a limitation on page 16, paragraph 4.

14.
Page 16, para 1. The authors do not provide any data to support their assertion that "our results are unlikely to have been notably affected by data extraction errors."
We have clarified on page 17, paragraph 1, that "…since the systematic reviews that were extracted by two authors were randomly selected, we expect that the data extraction error rate in the random sample (which was very low) is representative of that of the entire sample."

15.
Page 17, para 1. I view the post-hoc analysis of 'extent of multiplicity' as another subgroup comparison. The test result is interesting but does not appear to show a trend to the null (i.e. to PBI=0.5). I note that the amount of data diminishes with increasing degrees of multiplicity. Do the authors feel confident about this result?
Yes, the amount of data diminishes with increasing degrees of multiplicity, or in other words, most trials only have 2 available effect estimates per trial, while very few trials had a large number of available effect estimates (e.g. 8 or 12). We failed to mention that the regression line on the scatterplot is weighted by the number of observations available (we have clarified this in the Figure legend in Supplementary Figure S1). Therefore, we are confident about this result.

16.
Page 17, para 2. I think that this paragraph is perhaps more important than it appears – since I think it is very unlikely that a review author will select a result on the basis of its weight in a meta-analysis.
To clarify, we are not suggesting in this paragraph that review authors may select a result on the basis of its weight in the meta-analysis. Rather, we are suggesting that regardless of whether systematic reviewers selectively include an estimate, if the study contributes little weight to the meta-analysis, then the impact of such selective inclusion on the meta-analysis is likely to be negligible.

17.
Page 17, para 3. Please revise the first sentence so that the sentence affirms a more positive position about the need to pre-define methods for selecting effect estimates when multiple estimates from a trial are reported. (e.g. "Systematic reviewers should continue to pre-define …")
We have revised the sentence as suggested (see page 18, paragraph 3).

18.
Page 18, para 1, line 2. I would prefer "defined/described" to "reported".
We have changed "reported" to "described" (see page 19, paragraph 1).

18.
Page 18, para 2. This recommendation is excellent in principle – but I worry that it means a lot more work for review authors facing every increasing pressures to carry out reviews more efficiently … and longer reviews. What is the authors response?
We recognise that adopting our recommendations would involve more work for review authors. However, we believe the benefits to users of the review (in terms of being able to assess the risk of selective inclusion bias) outweigh this concern. We have noted on page 19, paragraph 2 that we recommend that a standardised table which facilitates reporting of information on multiplicity and selection methods be developed.

Figures
No comments on the main figures. But I would like the supplementary figure 1 to include labels on/near each blob describing the number of trials. I am still not completely sure that I understand what the data represent, e.g. trial effect estimates per trial report, per meta-analysis, etc.
We have added a label describing the number of trials above each blob. Each blob represents the PBI for a given number of trial effect estimates that were available per meta-analysis. So for example, the first blob indicates that when 2 effects estimates are available for inclusion in

an index meta-analysis, the PBI is approximately 0.60. This PBI was estimated based on 59 trials which had only 2 available effect estimates.

Appendix 6: really interesting information, thank you.
Thank you. No response required.

Typographic errors
Page 4, Introduction: missing word “multiple [abstracts]”
We think you mean “multiple estimates”, which we have added to page 1, paragraph 2.

Page 8, middle para last sentence: sentence construction is confusing should be edited, moving the clause about bootstrap resampling closer to the text about the statistical test.
We have divided this sentence in two: one describing the statistical test, and the following describing that confidence limits for the PBI were obtained by bootstrap resampling (see page 8, paragraph 2).

VERSION 2 – REVIEW

REVIEWER	Britta Tendal Danish Health Authority Evidence, Education and Emergency Management
REVIEW RETURNED	01-Apr-2016

GENERAL COMMENTS	The paper has become easier to read after revision and I have no further comments

REVIEWER	Barney Reeves University of Bristol, UK
REVIEW RETURNED	01-Apr-2016

GENERAL COMMENTS	I am happy with the edits. My only substantive comment is to the Editors - could you waive the abstract word limit to include the information about the lack of decision rules about multiplicity (reviewer 2, point 2). I think this is a really important point and I applaud the recommendation of reviewer 2. The information may be overlooked by readers if not in the abstract.