Author’s response to reviews

Title: Risk of bias judgments for random sequence generation in Cochrane systematic reviews were not in line with Cochrane Handbook in more than one-quarter of analysed cases

Authors:

Ognjen Barcot (ognjen.barcot@gmail.com)
Matija Boric (matija.boric.st@gmail.com)
Tina Poklepovic Pericic (tinapoklepovic@gmail.com)
Marija Cavar (cavar_marija@yahoo.com)
Svjetlana Dosenovic (sdosenovic@yahoo.com)
Ivana Vuka (ivana.vuka@mefst.hr)
Livia Puljak (livia.puljak@gmail.com)

Version: 1 Date: 22 Nov 2018

Author’s response to reviews:

BMC Medical Research Methodology
November 22, 2018

Response to review for the manuscript BMRM-D-18-00253

Dear Editor,

We are very grateful to editors and reviewers for taking time to review our manuscript, number BMRM-D-18-00253. We appreciate the constructive feedback we received. Hereby you will find our point-by-point response to reviewers. All changes suggested by the reviewers were fully addressed, and marked with Track Changes in the manuscript. Along with copying revision in the online submission system, we are also attaching this Response-to-Review as a supplementary file, where it will be easier and more convenient to read due to nicer document formatting.
Melanie Bell, PhD (Reviewer 1):

This manuscript discusses an interesting topic of research in the area of misclassification of bias in Cochrane systematic reviews (CSR). The aim of this manuscript was to determine how common misclassification of risk of bias was in relation to randomized sequence generation. Overall the manuscript lacked clarity and there were many issues which require major revisions.

Major Concerns:

1. The data extraction paragraph on page 6 did not provide sufficient detail to understand what terms were used to filter through the raw data

Response: We did not use specific terms to filter through raw data. Cochrane Library webpage displays data sections uniformly so the RoB tables always have the same position in text. The title of this section of Cochrane reviews is always the same: Characteristics of included studies. After this title, each study included in a review is listed alphabetically, with its unique identifier containing usually last name of the first author and year of publication (for example: Almath 2012), followed by the table with characteristics of participants and then title Risk of bias, after which there is a RoB table for each study. After that, the descriptive package of the next study starts. Due to this uniform organization of data in each Cochrane review, it was easy to program data extraction that would not make a mistake in identifying studies and RoB tables for each study included in ‘Characteristics of included studies’ section.

We encountered the following difficulties while planning and analyzing data extractions:

- to control for the possibility that Cochrane authors will not create RoB table for each included study, we extracted number of included studies from the text too, to double-check for the number of RoB tables that were supposed to be present in the text.

- missing sections caused problems logged by the error handling subs. These were checked and extracted manually.

- some RoB tables did not address seven predefined domains of Cochrane RoB tool. Because of this, we did not program data extraction only for the seven predetermined domains. Instead, all the tables were extracted on “as is” basis and checked and sorted manually later.

Data extraction was time-consuming process that required programming and testing for every step of the way to produce neither errors nor missing data. Coded algorithms were applicable for one purpose only and generally cannot be reused without reprogramming. All of unique errors were checked manually. We simplified this section of methods because we find data extraction
absolutely fully done and without errors - checked while programming and rechecked in data analysis. Therefore we do not find it influencing the end result of the study.

Solution: We clarified our methods in the Data extraction section on pages 6 and 7.

2. I had trouble following your Categorization methods (starting on line 34, page 7). There were many mentions of different categories (15, 19, then 5) which were not immediately clear upon looking at the results tables.

Response: We revised the text to enable easier following of this section of text. This important part of methods describes logic behind data analysis. It has to be descriptive enough while depicting a complex recursive routine. The text was logically correct, although we corrected the subtotal count of categories – proving the reviewers point.

Solution: We corrected text for errors. We find necessary to support this section by adding flow diagram of categorization, sub categorization and aggregation, to make it easier for a reader to follow the text.

3. There are conflicting results throughout the manuscript. For example, authors mention 12% of trials were erroneously judged in line 6, page 11 and then on line 49, page 13 you mention that one in five judgment is erroneous.

Response: Thank you for your comment. This had to be corrected in text. We state 12% (or one in eight) of trials in total are erroneous and this representation is fine in text. However, in p13 of the discussion we used expression “one in five”. This information refers to our finding of 20 to 22% of errors when considering only studies with low-risk judgment. Thanks to this comment, we realized that this was not detailed enough in the manuscript.

Solution: Text was revised and clarified.

4. Why was block randomization considered as "Method of randomization was not described"? Did the excel macro used to trim raw data simply look for the words "block randomization" and then trim away any possible supporting information? Block randomization and stratification are both desired characteristics in a randomization scheme, although details need to be provided.

Comment: Block randomization and stratification truly are both desired characteristics in a randomization scheme, although details need to be provided as you state. We clarified this in discussion on in p12:
Cochrane Handbook explicitly warns: "Sometimes trial authors provide some information, but they incompletely define their approach, and do not confirm some random component in the process. For example, authors may state that blocked randomization was used, but the process of selecting the blocks, such as a random number table or a computer random number generator, was not specified. The adequacy of sequence generation should then be classified as unclear”


We did not use Excel macro to trim any data in any part of this study. No automatic data deletion or trimming was undertaken in this study. Categorization was done manually, rechecked manually and eventually re-rechecked manually by search for characteristic strings.

5. Similarly, the authors present their findings as correct, how do they know they have not misclassified (as, I believe, the above example shows)? At the very least this should be acknowledged in the study limitations.

Comment: All of the above examples turn out to be a consequence of presumption of full automation of data analysis process. Once again, we state data extraction was full. Meaning every statement about randomization process was manually analyzed by human (co-authors). If there was a concern of missing or misleading data, data were checked by main author from both raw data workbook or from the CSR on webpage. Example in Q4 was already covered in text because it is a thing already explicitly identified by Cochrane Handbook authors in 2011 as a possible flaw in judging. However, we do acknowledge the possibility that human classification and comparison with instructions from Cochrane Handbook may have produced errors. To try to prevent this error, every classification made by one author was checked by the second author.

Solution: We acknowledged this as a study limitation.

6. There were many grammatical errors throughout the manuscript. Perhaps consider working with an editor.

Solution: We revised the full text of the manuscript after addressing all of the concerns from all reviewers.
Minor Concerns:

1. You've cited unpublished data in the paragraph starting on line 10, page 5. Is there any published data on this topic?

Solution: We added a reference to the recently published data on this subject; Propadalo et al. link: https://www.ncbi.nlm.nih.gov/pubmed/30312657, a study in which we analyzed allocation concealment domain of RoB table in Cochrane reviews, which was just published. We also have two other studies on our analyses of other domains under consideration in BMC Medical Research Methodology, but review process has been protracted; first of those studies was submitted in November of 2017, and the second one in May of 2018; both of them are still Under Review.

2. The addition of a flow diagram may assist in understanding reasons for inclusion and exclusion of RCTs.

Solution: We added a flow diagram for inclusion and exclusion of studies.

3. P4, L37. Perhaps the seven different domains should be introduced here, where they are first mentioned?

Solution: We introduced the domains at this point in text.

4. P6, L14. What do you mean by "advanced search"?

Comment: Advanced search is one of two options used to search through Cochrane Library. Simple search is available on the front page of the Cochrane Library, and it allows only for insertion of text and searching for this text in various sections of a Cochrane review. Advanced option gives opportunity to apply search limits such as content type, publication date or Cochrane group. Advanced option also allows searching for any review or protocol, or both, in a prespecified time frame, without being limited with searching certain text.

Solution: We revised the text of Methods to emphasize why this specific type of search was used.

5. P6. Perhaps the code used for data extraction could be included in an appendix? Also, this paragraph was unclear to me.
Comment: Paragraph about data extraction was expanded, in reply to major concern #1 for this reviewer. Initially, we did not provide a code because we did not consider it was relevant neither for outcomes nor for comprehension of this study. Based on this comment, we will supply additional text documents with most important VBA code for this study.

Solution: Macros in raw format were now presented in Additional files section.

6. P7, L2. The number of CSRs has been stated, but not the total number of RCTs, so there is no context for the number 1500. How many RCTs were included?

Comment: Thank you for this remark. We did not mention total number of analyzed RCTs in text to this point but only number of CSRs. This important issue is addressed in Results section. However this information is now presented in a flow diagram preceding this point in text as previously advised by the same reviewer.

Solution: We added this information in Results, and in newly added flow diagram.

7. P7, L20. "Supporting comments" have not been defined. They should probably be introduced in the introduction, for readers who are not familiar with CSR

Comment: We used the term accompanying comment instead of supporting as should be stated. This issue is now corrected and rephrased in background section.

Solution: The following was now added to Introduction: Cochrane authors should provide their RoB assessments in the RoB table, where assessment for each domain needs to contain two sections: a judgment (i.e. risk is low, unclear or high) and accompanying supporting comment that needs to justify the judgment.

8. P7, L46. Missing the word "bias"?

Solution: This was corrected.

9. P8, L46. "Same as above", "as above", etc. may be reasonable responses, but are not able to be classified through your method

Comment: This is true that ‘same as above’ or ‘as above’ may be reasonable in certain contexts, but not in this particular case. RoB domain for random sequence generation is the first domain in any RoB table. So, when authors write ‘as above’, this is meaningless to a reader because a
reader does not know what exactly is supposed to be ‘above’ that the Cochrane authors are alluding to. Information that is physically present ‘above’ in each RoB table is another table for the same study, with characteristics of a study; a so-called PICO table, which gives four rows of information about participants, interventions, comparisons, outcomes. The PICO table often has a fifth row called Notes, where Cochrane authors can put any other information that they want, that were not covered with the first four rows. So, when RoB table starts, and Cochrane authors simply write ‘as above’, a reader has no idea what exactly should mean ‘as above’.

10. "Mechanic method" should probably be "mechanical method"
Comment-solution: Thank you for your suggestion. This was corrected.

11. P10, L5-6; L24-26. Unclear
Comment-solution: Due to other suggestions in this review, we revised this part of the text and we added supporting flow diagram 1 and table 1. Errors in text were corrected as well.

Comment-solution: Text was rephrased and clarified.

Reviewer 2 (Reviewer 2):

PEER REVIEWER COMMENTS:

1. GENERAL COMMENTS: Whilst the study focuses on one area in a fairly narrow category of research in that Cochrane systematic reviews are quite a formulaic template that suits assessment of some kinds of research, the inference is that secondary research which is conducted with the very purpose that it is to be trusted is frequently riddled with mistakes. I support the author's investigation and attempt to publish this finding as I believe it to be a true finding and a potential underestimation of a much bigger problem.
Comment: We are grateful to the reviewer for kind words about our study.
2. The conclusion, as it is phrased, may be a little bit inflammatory, and I'm not particularly a Cochrane advocate. But I think the point is that reviewer's judgements when conducting RoB cannot necessarily be trusted, and that if this occurs in a sample of Cochrane reviews, which are regarded by many as gold standard systematic reviews, then what are the implications for all the other systematic reviews being published? Moreover as the finding that judgements were erroneous for sequence generation and this was the only RoB domain that was assessed, what are the implications for all domains of quality assessment (some unpublished data is referred to page 5 so it would be great if this were published by the time this goes to press), and indeed all phases of the review process where those conducting systematic reviews are trusted to make objective, transparent decisions, a proportion of which should be verified for accuracy by a second reviewer (i.e., study selection and data extraction)? Therefore I think the conclusions and implications could be broadened to highlight that authors of systematic reviews may often be making incorrect judgements which compromise the reliability of the systematic reviews.

Comment: We can concur with the reviewer that it would be probably better to write more diplomatic and less inflammatory conclusion. For this reason, we made our conclusions more neutral, to frame them in a way that our study can serve to improve reporting and reduce mistakes, and ultimately make Cochrane reviews better.

3. The authors use of the CSR is interesting, but its unclear whether/how many of the Cochrane reviews themselves used the CSR or just the published paper. There's an ongoing issue in RoB with the selection of either "unclear" or "high" risk of bias where a method is not described properly depending on individual reviewer's tendency to give trials the "benefit of the doubt". A more generous judgement is to give "unclear" in this instance but is often because they are relying on the published paper and the reviewer may assume it was described in more detail in the CSR, but never check. This issue could be mentioned in the introduction by way of clarifying whether this issue was encountered or dealt with in this research.

Comment: In our methods we indicated that we used only Cochrane reviews that included randomized controlled trials. We excluded overviews in which included studies were other reviews. We agree with the reviewer that it is possible that Cochrane authors could be more generous, and that they make assumptions based on the partial data available, and we have indicated this in the manuscript. This sentence says: According to the Cochrane Handbook, Cochrane authors can make assumptions, but need to elaborate them.

4. Line 6-8 of page 10 and Table 2, the N/A category isn't clearly explained. Since only reviews with RCTs or those with both RCTs and non-RCTs were included, what kind of trials would
be labelled by either the Cochrane reviewer, automated process or a reviewer as "N/A" and why? Could the Cochrane author's labelling of this domain as N/A also be erroneous?

Comment-solution: This part of the text was revised. Additionally, we added supporting flow diagram 1 and table 1. Errors in text were corrected as well. N/A was used and clarified as “not-available” only as a part of the supporting judgment of the RoB table; therefore Cochrane author’s labelling was not found erroneous.

5. Page 6, "screening for study eligibility" I would prefer to know what proportion of titles/abstracts were verified or if all were verified. Equally what proportion of the calibrations of categorisations were verified by OB?

Comment-solution: All of the titles for exclusion of CSRs were rechecked by second author, and verified by last. Pilot calibration/categorization by OB was verified in total by LP. All of the categorizations from coauthors were rechecked by the first author. Everything about this issue is now covered in detailed representation in left part of Flow diagram 1.

6. The English is broadly clear and understandable but there are frequently connecting words missing such as "[The] First author analysed" page 7 and "According to the Cochrane Handbook, [an] additional two categories" page 8.

Comment-solution: Thank you for your remarks. We corrected these errors and we revised the whole text of the manuscript to accommodate English grammar.

7. As discussed above, whilst I appreciate the author's attempt to limit their conclusions to those studies (only the one domain and to Cochrane reviews) I think the interpretation needs to be widened. From this example it is clear that systematic reviews are being conducted and published with errors, which means that as products which are regarded as trustworthy, high-quality evidence they, and the people that conduct them, need to be scrutinised more.

Comment: We agree with the author and we added a note about this issue in Discussion.

We hope that our revised manuscript will be satisfactory.

Sincerely,

Livia Puljak and colleagues