**Author’s response to reviews**

**Title:** Mapping the nomenclature, methodology and reporting of studies that review methods: A pilot methodological review

**Authors:**

Daeria Lawson (lawsod3@mcmaster.ca)

Alvin Leenus (leenusa@mcmaster.ca)

Lawrence Mbuagbaw (mbuagblc@mcmaster.ca)

**Version:** 2  **Date:** 01 Oct 2019

**Author’s response to reviews:**

Please note: all page and line numbers as mentioned in our responses below refer to the tracked changes version of the manuscript. This was submitted and is readily available and visible to reviewers, appearing right after the clean version of the manuscript. We want to stress this point so that it is not missed and to highlight that there should be no additional steps in order to obtain access to the tracked changes version of the manuscript. Below, our responses begin with ‘>’

Please return to the original reviews and spend time carefully addressing the comments and adding corresponding explanations to the manuscript.

>Thank you for this opportunity to revise our manuscript again. We have reviewed the original queries to ensure that all of the original points have been addressed in the revised manuscript, highlighting all revisions in yellow with tracked changes. We have also updated and provided additional sources in the ‘Background’ section that support the importance and need for this work (p5, lines 151-154).

In particular describe the sampling justification, gold standard for comparisons, calculations and explain the main results in Table 4.

>Sampling justification: some methodological studies use “all” the studies from their search, while others use a “sub-sample” of the identified studies. We examined the latter studies to determine if a formal sample size computation was done to determine how many studies should be included in the “sub-sample”. We have expanded on this in our response to the reviewer below.

>Gold standard for comparisons: after screening 236 articles, only 31 were methodological reviews. The ideal search strategy (gold standard) would identify all of these studies without including many of the non-eligible ones.

Using our database of identified studies (n = 236), we computed the proportion of studies identified with each search term that were actually methodological reviews (i.e. included studies; true positives), and the proportion identified that were not methodological reviews (i.e. excluded studies; false positives). We also computed the proportion of studies not captured by the search term that were not methodological reviews (i.e. excluded studies; true negatives) and the proportion of methodological reviews that were not captured by the search term (i.e. included studies; false negatives) We pooled the sensitivity and specificity estimates for multiple terms using the parallel testing approach in order to achieve an optimal combination of search terms with a good balance of sensitivity and specificity (p10, lines 264-273).
Table 4 is not easy to read in its current form and would be better turned around (columns become rows) so that each author/study is listed as a separate row in column 1 in alphabetical order of surname of first author (A-Z) and with the reference given (and not each author/paper in a separate column). Please highlight all changes so they are easy to see.

In addition to revising Table 4 for better legibility (i.e. turned around columns to become rows, listed authors in first column alphabetically, provided references), we have broken down and re-arranged the summary of the main results as presented in Table 4 under ‘Characteristics of included methodological reviews’ (p11, lines 301-307) in addition to ‘Nomenclature’ and ‘Methodological Features’ (p11-12).

Reviewer 1
These authors have not adequately addressed my queries. To be fair, I have not been able to access a version of the manuscript that highlights what changes have been made - so it has been time consuming trying to identify changes. However, I am not sure that many changes have been made.

We thank the reviewer for their suggestions to improve this manuscript, and we regret to learn that the reviewer felt that their queries were not fully addressed as this was not our intention. We have also confirmed that we did attach and submit a tracked changes version of the revised manuscript in our submission. This was available for download and review as a Supplementary File linked within the clean manuscript that we had previously submitted. We have included the revised, tracked changes manuscript with this submission.

A few examples:
1. I asked for clarification of the way articles end up in methodological reviews and then end up in guidelines. The authors have pointed me to figure 4 which was in the first version of the paper that I reviewed and describes where the current study fits into a process for developing guidelines. This has not addressed my comment.

We have clarified that we aim to develop guidelines for the “reporting” of methodological reviews (i.e. the studies that we have reviewed in this pilot) since there is currently no guidance for reporting them (p4, lines 104-105). We have also provided additional reasoning for the need for a standalone reporting guideline, developed especially for methodological reviews (p4-5, lines 106-154).

The articles in the methodological reviews do not end up in guidelines. Studies included in methodological reviews can be identified by search strategies, based on specific time periods and topics of interest for example. How authors of methodological reviews select the studies that they will use varies greatly, hence the need for guidelines for methodological reviews.

This paper is a pilot study with two aims:
1) To determine the feasibility of conducting a methodological review of methodological reviews (given the disparate nomenclature and absence of indexing terms). This is contingent on if we can identify accurate search terms.
2) To determine the need for reporting guidelines. This is contingent on if we identify large differences in design, analysis and reporting practices.

In this paper we have demonstrated the feasibility of conducting a larger methodological review of methodological reviews, and we have identified the need for reporting guidelines. This work is outlined in #1a in Figure 4. It will be followed by a larger methodological review (#1b), then a consensus study (#2) and finally, a guideline document (#3).

[Original review: There are places in the manuscript where I cannot tell what part of this process you are referring to. For example, on line 92 you mention 'studies of interest' - are these MRs or the original
As per our last revision and response to reviewers, we have updated ‘studies of interest’ to ‘MRs’. We have also revised the manuscript to make sure that it is clearer which studies we are assessing throughout.

2. I had a series of questions about Table 1. Table 1 has not been changed in any way since I read the original manuscript. The table ought to be self explanatory and, in my view, it is not.

We have updated Table 1 (p6-7) in line with our explanations addressing the original queries (we have also listed the responses that resulted in revisions to the table below for reference). Table 1 is a list of objectives for this pilot methodological review and how the information will be objectively interpreted.

[Original review: Many questions are raised for me by Table 1. (a) 'Determine the need for guidance'. Is this the MR authors, the original authors or the guideline writers?]

(a) ‘Determine the need for guidance’ is referring to the objective of this paper. We have added a statement (p6, lines 179-182) to further clarify this point.

[Original review: (b) Why have you used a coefficient of variation to look at the number of databases searched? It's not clear to me what its meaning is here.]

(b) The coefficient of variation (CV) is the ratio of the standard deviation (SD) to the mean (SD/mean), where a CV of 1 suggests that the spread in the results are relatively equal to the mean. For example, if we find that authors of methodological reviews search 1 database on average and the CV is 5 (i.e. standard deviation is greater than the mean), this indicates that there is large variability in the number of databases that are searched. If that is the case, it may be worthwhile investigating the basis for deciding how many databases should be searched when conducting a methodological review. Does this depend on the scope of the research question? Does this matter for the generalizability of methodological reviews? For systematic reviews, a search of at least 2 databases is recommended. Is this the case for methodological reviews?

[Original review: (c) Justification and calculation of a samples size. These points read as if you are checking the original papers for this, not the MRs. But are reviews of the original papers part of this review process? I thought it was a review of the MRs.]

(c) On p5, line 162 we state that this pilot is a review of methodological reviews, and so the research questions in Table 1 are referring to properties of methodological reviews. This is further outlined under the Methods ‘Study Selection’ and ‘Data Extraction’ sections. Original papers were not checked as part of this pilot. We added in ‘MRs’ on p8, line 231 to clarify that data is being extracted only from methodological reviews. Methodological reviews use different sampling strategies, including comprehensive searches. We have elaborated on this further, with examples below.

3. I queried the simple random sampling of papers for a review. I have not heard of this before and cannot see how it makes any sense. The response has been that it is common to do this with a citation to their own publication where they have sampled papers in the context of registered clinical trials. This is not evidence that samples are commonly taken in methodological reviews.

[Original review: (d) 'Is a simple random sample of studies used?'. I don't understand where this fits into this work. I have never seen a review done on a simple random sample of studies - could you explain what this is about and why you would expect to see a simple random sample? Isn't it common to use all the studies in a review - not just a sample of them.]

The research questions in methodological reviews are different from those in systematic reviews, where the goal is the summarize all of the available evidence. Methodological reviews can use sub-samples of their search, and often do if the number of retrieved articles is prohibitively large. Likewise,
some methodological studies are designed as before-after studies, longitudinal studies or cross-sectional studies and use the sampling strategies of these designs. A search is conducted only to create a sampling frame. In addition to the previously cited reference, we provide the following list of an additional eight methodological studies (from a variety of authors/research groups) that further attest to the use of this sampling strategy:


We have added in some of these examples to the Discussion section (p15, lines 438-441). We have also revised Tables 1-4 to clarify that we assessed this only in studies that did not take a systematic approach (i.e. aim to capture all available studies). We have added a description of these findings in the Results section (p12, lines 350-353).

4. I queried the calculation of specificity and sensitivity. The authors have reiterated what they said in the manuscript and directed me to look at the Data Analysis section. This section appears to be unchanged from the previous version. Since I reviewed the paper in March, I have read this section and it is not clear that the calculations make sense. The sections either needs to be written more carefully or they need to reconsider their calculation. Neither of these things appear to have occurred.

[Original review: There is a measure of sensitivity and specificity - a requirement that both should be >= 70% for the identification of MRs in your searches. (a) As far as I can tell there is no establishment of a 'gold standard' with which to compare your results (and, hence, calculate sensitivity and specificity). Usually, this is achieved by a hand search of the literature (over a period shorter than a year, presumably) - so you know which are MRs and which are not. Then you run your search and see what it brings up compared to what you know as an accurate measure. Could you explain your method?]

We have provided further clarification to elaborate on the calculation of sensitivity and specificity (p10, lines 264-273). The reviewer is correct in pointing out that there is no gold standard, but after screening 236 studies, we know which studies are methodological reviews and which are not. This was our “gold standard”. Using various search terms to try to identify these studies will give us estimates of sensitivity and specificity for those search terms. After we had identified all eligible studies, we applied our original search terms (individually and in combinations) into the EndNote search fields, to the library of all of the retrieved articles from our search (n = 236). In this way we were able to run a
search in EndNote and determine what combinations of the original search terms were most accurate (balance of sensitivity and specificity) in capturing the eligible studies. We combined estimates using the parallel testing calculation for sensitivity and specificity (Cebul RD, Hershey JC, Williams SV (1982) Using multiple tests: series and parallel approaches. Clin Lab Med. 2(4): 871–90). We have added these details to the Data Analysis section in the Methods (p10, lines 264-273).

Original response: (a) We outline our approach under the Methods ‘Data Analysis’ section. As the reviewer has correctly pointed out, we computed the sensitivity and specificity of search terms after we completed a search of PubMed restricted to a year (2017). From this search we identified studies that were methodological reviews. Running this same search, we then applied combinations of terms to assess the proportion of true methodological reviews that were identified with that term (or combination of terms).

[Original review: (b) It appears that you have done your calculations purely on the basis of the outcome of, what seems to be, a very successful search strategy so your sensitivity is 100% and your specificity is 0.99%. Actually, I'm not sure what you have done here - but it looks like you are only using your own search results for these calculations which is not the correct thing to do. (c) The overall results seem inconsistent with the individual search terms. Could you explain how you combined these estimates?]

>Original response: (b & c) We computed sensitivity and specificity based on a list of search terms that correctly identified methodological reviews in our initial search. These results refer to a combination of six terms that correctly identified methodological reviews in our initial search. That is, ‘literature survey’ + ‘meta-epidemiological’ + ‘systematic review’ + ‘systematic survey’ + ‘methodological review’ + ‘methodological overview’ which differs from our original search strategy. This is what the search accuracy would have been if we had used those terms. The next step will be to apply these terms in the full review and additional databases to determine their sensitivity and specificity beyond this pilot work. This pilot is in place to identify the best combination of terms that would have the most relevant yield for a larger study. In other words, we have used the results of our search to identify a better combination of terms with higher sensitivity and specificity, by retrospectively applying the terms with the best hits to the search results. This accuracy is not meant to describe the current work, but rather to inform future searches.

[Original review: (d) Doing a duplicate screen for the MRs will introduce bias (when you have not done the same for the non-MR studies).]

>The duplicate screen was done for all studies (methodological reviews and non-methodological reviews) from the original search. Once the methodological reviews were identified, only they (n = 31) were processed further.

>After reviewing the other original queries and our responses to these, we believe that we have adequately addressed the remaining points raised.

Review 2 (no comments)

>No further comments were raised by the reviewer and after reviewing the original queries and our responses to these, we believe that we have adequately addressed all of the original points. We have also revised some of the discussion about the search strategy and generalizability – which were previously raised by this reviewer – for clarity (p13, lines 380-383; p15, lines 442-447).