Reviewer's report

Title: Optimal database combinations for literature searches in systematic reviews: a prospective exploratory study

Version: 0 Date: 10 Sep 2017

Reviewer: Steve McDonald

Reviewer's report:

This is a substantial study that provides an interesting perspective, through an analysis of real-life prospective searches, on how retrieving studies for SRs is affected by the choice of database. It throws up some interesting and sometimes surprising findings. Because of this, it's incumbent on the authors to provide as full an explanation as possible. My biggest criticism is that we are not provided with more context around the differential performance of databases (and combinations of databases) as applied to individual review types (intervention, intervention RCT-only, DTAs, etc.). If there are differences according to the type, scope and criteria of reviews then this is information that will improve the usefulness and generalisability of the findings from a user perspective.

ABSTRACT

Background: The sentence containing the aim/objective should emphasise the different approach this study takes, namely assessing coverage based on prospective real-life searches as opposed to theoretical coverage.

Methods (line 9/10): implies that the search results of all the information specialists at EMC are part of the study, when in fact it's only the author's searches that are included. This can be clarified in the sentence "PubMed was used..."

Results (line 20/21): the finding reported in the main paper that 84 references out of the 1830 references from these 58 reviews were not identified by the SR searches is not insignificant (about 4.5% of all the references). Encourage giving consideration to incorporating this finding into the Results section.
BACKGROUND

Para 1: an even bigger disadvantage to using multiple databases is having to translate thesaurus terms - I don't think this is captured in what is written now.

METHODS

Section on development of search strategies

Para 1: consider adding a footnote to explain why searches of PubMed were restricted to the publisher[sb] only. [I know this is commented on in the Discussion, but helpful for readers to know of this at this point.]

Para 1: you say hits from GS were restricted to 100 when number of references from other databases was low - is this correct or do you mean 1000? (Seems counter-intuitive to restrict when other sources are low, but perhaps the logic is that if it's low in the databases, it will be low in GS.)

Section on Determining relevant references of published reviews

Para 1: the word 'Retrospectively' to start the paragraph is unnecessary

Section of Data analysis

- general comment: the descriptions and definitions of the outcome measures can be difficult to read and understand; it's important that terms are used unambiguously and consistently.

- with so many search performance measures included in the analysis, strongly consider including a table in this section (performance measure + definition) to help with understanding and navigate through the section

- consider including a sentence in the first paragraph that summarises your primary performance measure of interest - i.e. what is it the data analysis seeking to uncover and how is this to be achieved.
- Para 1 (last sentence): consider saying 'unique included references' rather than 'relevant' references; clarify what is meant by 'rest of the research'

- Para 2: implies that the study is calculating a single statistic to convey 'Efficiency' but I don't think this is the case - in the results overall efficiency is the equivalent to sensitivity. Please clarify.

- Para 3: first sentence is a bit confusing, do you mean to say "Sensitivity was defined as the number of included references retrieved by each database (or combination of databases) divided by the total number of included references retrieved by all databases. ??

- Para 4: first sentence, the word 'results' is used twice, for consistency, shouldn't this be 'references'?

RESULTS

- Para 1: I think it's essential to include a description of the SRs included in the sample of 58 to give added context. Were these mostly intervention effectiveness reviews? If so were these mostly RCT-only reviews? How many DTA reviews? This may give some explanation to the surprising data for CENTRAL.

- Para 1: consider adding "(4.6%)" after 84 references. (I know it's probably beyond the scope of this study, but that's a higher figure than I would have expected given the databases searched. Curious to know more about the sources of these missing included references.)

Section on Unique references per database

- Para 1: "A total of 292" rather a 'number'

- Para 1: "... Cochrane CENTRAL, is absent from the table, as it did not add any UNIQUE included references..."
The following sections of the Results refer to the figures. As they stand at the moment, the figures are difficult to read and interpret - they require titles, ideally a legend, axis labels and an explanation included as footnote.

Section on Status of current practice

- I wondered if this was essential to include in the study (I'm not convinced it is.). However, I think if the implications were framed a little differently, this would help. In the results for the MEDLINE/Embase combination, you could describe more narratively: "Around a third of the reviews (37%) relied on the combination of MEDLINE and Embase. Based on our findings, this combination achieves acceptable recall about half the time (47%), implying that less than a fifth (17%) of the reviews in the PubMed sample would have achieved an acceptable recall of 95%.

TABLE 1

- consider combining results column 2 and 3, e.g. 29 (50%)
- column 4 totals 293, should be 292?
- column 5 'Percentage of unique references' - this could be combined with column 4. Not clear how these data have been calculated; should the total percentage = 100 rather than 110?

TABLE 2

- column 2 #includes - consider adding (N=1746) in column header
- it would be helpful to add footnotes, especially for median sensitivity, minimum sensitivity and percentage 100% sensitivity, to aid interpretation of the data.

FIGURE 1

- not clear why the relative recall data for 'All databases' is worse than for the combinations of fewer databases?
DISCUSSION

Para 2: the opening sentence should be framed in the context of the study findings (e.g. "Our study found that...") otherwise it reads like an opinion.

Para 2: second sentence: To ensure instead of To permit

Para 2: to what extent is the recommendation to search the four databases (Embase, Medline, WOS and GS) generalisable across all types of review (DTA, intervention, prognosis, etc.)? Data are collected at the individual review level, so do the stats vary according to the type of review? This seems an important point to reflect on and to inform readers about.

Para 2: PsycINFO is broader than psychiatry, perhaps say 'mental health-related reviews'

Para 3: the Cochrane Handbook recommendation refers to intervention reviews that have RCT as the inclusion criterion, not all SRs - this should be checked and clarified. It seems that a relatively small proportion of reviews were RCT-only.

Para 3: the sentence beginning "The combination of Embase and MEDLINE resulted in reasonably..." doesn't seem to fit with this paragraph.

Para 7 (final para before strengths/limitations): it says "...probability of missing more than 5% of included ... is 33% and 30% respectively." Clarify what does 'respectively' refers to here.

Strengths and limitations

Para 1 - could also reflect on other limitations that result from the sample of reviews included in this study being broad scope and topic, especially in terms of generalisability across review types.

Para 2: the last sentence reads "This separate set had retrieved..." I think this refers to the subset as supplied by the publisher, but is ambiguous because of the sentence immediately before it. Please clarify what 'separate set' refers to.

Para 3: second sentence, clarify what is meant by the phrase, "... and database needs..."
Para 3: sentence beginning "We do not suggest that if anyone..." the double negative confused me initially, could consider re-phrasing something like: "We are not implying that a combined search of the four recommended databases will never result in relevant references being missed, rather that failure to search any one of these four databases will lead to relevant references being missed."

Para 3: last sentence: might want to add "Our experience in this study shows that additional..."

Para 4: first sentence "... a random sample of..."

Do the authors want to comment on what effect knowledge that these prospective searches would be later analysed for their sensitivity etc had on the first author's search performance? CONCLUSIONS

- we can't verify the opening statement that "all" SRs need searches of the four databases because that information (the extent to which these results apply equally across all types of SRs is not provided to us)

- PsycINFO: use of word psychiatry vs mental health

- last sentence: consider adding to the beginning: "This study also highlights once more that searching databases alone is not enough...

Final comment, the word probability is used to describe databases or combinations of databases missing relevant studies - I don't know if 'probability' is being in the correct statistical sense here or as a synonym for chances/likelihood, or whether this is even important.
Level of interest
Please indicate how interesting you found the manuscript:

An article whose findings are important to those with closely related research interests

Quality of written English
Please indicate the quality of language in the manuscript:

Needs some language corrections before being published

Declaration of competing interests
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests.

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal