Reviewer's report

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

Version: 0 Date: 10 Sep 2017

Reviewer: Elizabeth Stovold

Reviewer's report:

General comments

This is an important question to try and answer as searching multiple databases, then downloading and de-duplicating large record sets can be very time-consuming. This paper will be of interest to information specialists and librarians involved in searching for systematic reviews, and more generally for researchers involved in systematic reviews.

Background/aims

1. The stated aims of paper are to determine which single or multiple database combination retrieves most relevant references. There is a second part to the study which examines current practice in recent published systematic reviews, but this is not explicitly mentioned as an aim of the paper. This section just needs a line to describe the aim of this part of the work.

Methods

2. Development of search strategies:

It would be good to have a bit more detail here about how the search strategies are developed and then translated for different databases, especially for readers who may not be familiar with how a search strategy is developed.

Data analysis:

3. Did you also look at the overlap between databases, as well as the number of unique studies? What was the overall coverage of Embase and MEDLINE for example, in terms of how many included studies were retrieved by each?
Results

4. Of the 58 reviews analysed, what study designs were included? In the methods you say review projects cover a wide range from effectiveness, to diagnostic, ethics etc. so what were the review types of this set? This is relevant as you comment on the lack of unique studies found in CENTRAL, which is a database of RCTs.

5. As far as I understand, the subject specific databases were not searched for all the systematic reviews under study, but only for those where the topic was relevant, so it's not surprising that they only contribute unique references for a relevant topic. The fact that they did find unique studies is very useful to know, but we don't know either way if they would have contributed unique studies in other topics, or not.

6. You haven't reported the NNR in the results text. This is of interest - for example I think it's notable that the NNR of the recommended combination of databases is 73. By searching a combination of only 3 databases (ML-EM-GS) the NNR reduces to 64, with only a small drop in the overall sensitivity.

Discussion

7. "The highest scoring database combination without Embase is a combination of MEDLINE, Web of Science, and Google Scholar, but that reaches satisfactory recall for only 39% of all investigated systematic reviews, while still requiring a paid subscription to Web of Science."

In this situation the addition of CENTRAL might be very beneficial, if the systematic review is focussing on RCTs, as CENTRAL contains reports of trials identified from Embase.

8. "Subject-specific databases like PsycINFO only added unique references to a small percentage of systematic reviews in our research"

As per my previous comments, my understanding from reading the methods is that databases like this were only searched for a sub-set of the reviews?

9. "Our research shows that the Cochrane Handbook's recommendation to search MEDLINE, Cochrane CENTRAL, and Embase is not sufficient for systematic reviews"
This is the minimum recommended for reviews of RCTs. I think your analysis included other review types.

10. "To our surprise, Cochrane CENTRAL did not identify any unique references for the reviews in this research, neither in general, nor compared with only those found in Embase and MEDLINE, even for the three reviews focusing entirely on RCTs."

CENTRAL is a database of probable and possible RCTs. If only three of the reviews under study focussed solely on RCTs then it is not that surprising that CENTRAL did not contribute unique studies to the other reviews, or in fact in the RCT reviews, as reports of RCTs from both PubMed & Embase are fed into CENTRAL. I think it is worth commenting on this as at the moment it is an unfair comparison.

11. "Whether a reference is available in a database is important, but whether the article can be found in a reasonably sensitive, precise search is not only impacted by the database's coverage. It is impacted far more by the ability of the searcher, followed closely by the accuracy of indexing and the complexity of terminology in a particular field."

Please can you provide a reference to support this statement?

12. "Because these studies based on retrospective analysis of database coverage do not account for the searchers' abilities, the actual findings from the searches performed, and the indexing for particular articles, their conclusions are tenuous at best and misguided at worst"

I agree that a strength of the current study is it's prospective design, but I think it's a bit strong to say the conclusions of previous studies are misguided, especially if the authors of those studies have acknowledged and discussed the limitations of their retrospective design.

13. It would be helpful to know how the findings of this study compare with the previous studies - does it support or refute them?
Conclusions

14. "Special topics databases such as CINAHL and PsycINFO should be added if the topic of the review is directly focused on nursing or psychiatry"

Is it possible to draw a conclusion about this as these databases were only searched for those specific topics, and not all 58 in the set?

15. Finally - what are the authors' recommendations for future research in this topic area?

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:

Acceptable

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 am employed by a Cochrane Review Group and as part of my role I contribute records to CENTRAL, which is one of the databases discussed in this paper.

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