Author's response to reviews

Title: Optimizing Search Strategies to Identify Randomized Controlled Trials in Medline

Authors:

Li Zhang (li.zhang@usask.ca)
Isola Ajiferuke (iajiferu@uwo.ca)
Margaret Sampson (msampson@cheo.on.ca)

Version: 3 Date: 30 November 2005

Author's response to reviews: see over
Dear editors,

Thank you for the opportunity to revise this manuscript to take into account the reviewers comments. We have adopted most and it has strengthened the paper.

We do believe that some of the reviewer’s comments reflect a lack of clarity on our part in previous versions regarding the objective and scope of this manuscript. We have made changes throughout in an attempt to make those clearer. With these changes we hope we have conveyed that we are not recommending using a two-phase strategy over a more sensitive three-phrase strategy. Our main objective is to provide evidence to help readers choose between various two-phase strategies that have been proposed. The reader must also decide between this approach and other searches, including those published by some of our reviewers. However, the other major searches are published with their own supporting evidence, albeit each uses somewhat different methods and comparison may pose some difficulties for the reader. However, comparing between those is beyond the scope of this manuscript.

We look forward to your decision regarding acceptance,

Respectfully yours,

Li Zhang
The article is much improved, and I can now understand how the authors did their study. I think it is a valuable contribution, for testing the proposed terms for improving the precision of Cochrane searches.

That said, I believe the authors’ interpretation is incorrect: they have shown what you gain and lose, but they haven’t shown what you lose is acceptable, in terms of the conclusions of the reviews. I think that this is because they are making too much of the minuscule number of articles retrieved by the SSversus strategy (which is not statistically distinguishable from the other search strategies they tested), and they are likely undervaluing the 56 articles it misses. They should redress this in their comments.

The primary objective of this study is simply to validate four proposed variants of the first two phases of the HSSS to an accepted search strategy. We admit that not assessing the effects of the 56 missing articles is a key limitation, and have addressed this in the Discussion section (see details later). But we think an important contribution of this study is that we provided evidences to four recommended search strategies with no prior supporting data. Our purpose is to inform readers the results, so they can weight the trade offs and make informed decisions when selecting a search filter.

We changed the interpretation. The new conclusion is that, compared to SS_{12}, three of the proposed search strategies do not work at all, and SS_{versus} is only slightly better. However, we still think it is worthwhile to use SS_{versus} if reviewers do decide to use only the first two phases of the HSSS for various reasons because the workload is comparable to SS_{12} but it may retrieve more relevant items though the differences are very small.

Specific comments:

From Responses to Reviewers

“Although a few studies [13,14] have explored different search strategies to identify RCTs in MEDLINE, they are all focused on improving the comprehensiveness of the search strategies. As an increasing number of systematic reviews have to be completed within tight budgets and timelines, it is necessary to strike a balance between the comprehensiveness and the precision.”

Accepted. We changed the sentence to the following (p.4-5):

A few studies [13,14] have explored different search strategies to identify RCTs in MEDLINE recently. Haynes and colleagues developed separate strategies for different purposes: strategies with high sensitivity for comprehensive searching and strategies with high precision for more focused searching.

We would also like to clarify that it is possible to device separate search strategies for high sensitivity and high precision but controlling for both at the same time is a little bit tricky. In your 2005 paper, you did device strategies for optimizing sensitivity and specificity, but ended up achieving high values for both measures but low values for precision (e.g. for the first strategy in Table 6 of the paper, sensitivity was 95.8%, Specificity 95.0%, and Precision 38.5%).

“We added this study to the second paragraph of Discussion sections: Haynes and colleagues {63} recently developed a search strategy to identify RCTs in the Medline that has a sensitivity of 95.8%.”

Thanks; 95.8% should be 99.3% (page 10 in the revised Ms)

Accepted. We changed the number to 99.3%

From the revised paper:

P. 8.

“A closer examination of the data found that, across the 94 reviews, SSversus was able to find 3 more relevant articles than SSCrossover, SSCROSS-OVER STUDIES, SSvolunteer, or SS12, but SS123 found 56 more relevant articles than SSversus.”

This is a key point and must be included in the abstract – the abstract makes it sound as if SSversus is better than the alternatives, but the difference likely isn’t statistically significant and systematic reviewers will need to know that using SSversus will cause them to miss 56 articles.

Accepted. We have included this in the abstract:

Across the 61 reviews, the search term versus combined with the top two phases of the HSSS was able to find 3 more included studies than the top two phases of the HSSS alone, or in combination with any of the other proposed search terms, but at the expense of missing 56 relevant articles that would be found if all three phases of the HSSS were used.

“Across the 61 reviews, SS123 retrieved 508,625 articles in total, while SSversus retrieved 171,032 articles”

This is not a helpful way to report the data, as no reviewer is simultaneously trying to do 61 reviews. It would be helpful to have a number needed to read based on, say, the median no. of retrieved articles, divided by the median no. of included studies retrieved, comparing this for SS123 and SSversus.
Accepted. We deleted the sentence and added the Article Read Ratio both in the text and in Table 4 according to your suggestion. We also calculated the estimated time needed to finish a review based on the equation developed by Allen and Olkin. The new description is as follows (P 9):

Table 4 also shows the Article Read Ratio (ARR), which is defined as the median of articles initially retrieved divided by the median of included studies retrieved. The ARR of SS_{123} (182) is significantly higher than that of SS_{versus} (54), and the ARRs of SS_{12}, SS_{crossover}, SS_{CROSS-OVER STUDIES}, and SS_{volunteer} are the same (51). We calculated the estimated time to finish a review for each search strategy based on the regression equation developed by Allen and Olkin: time = 721 + 0.243x – 0.000123x^2, where x denotes the number of articles initially retrieved, as shown in Table 4. The time needed to finish a review is 1086 hours for SS_{123}, 823 hours for SS_{versus}, 818 hours for SS_{12}, SS_{crossover}, SS_{CROSS-OVER STUDIES}, or SS_{volunteer}.

This is also added to the abstract because it gives readers important information.

P 10
“When searching this database, reviewers can retrieve relevant studies by using relatively simple subject search, thus avoiding the problem of selecting search filters.”

This is an untested statement! Subject searchers are not inherently simple – eg, for mental health studies, we’ve been able to get up to only about 90% sensitivity in MEDLINE (unpublished). Further, the Cochrane indexing and search engines are not likely to be as good as NLMS (indeed, they are widely thought to be bad).

Accepted. We changed the statement to this:
When searching this database, reviewers only need to develop subject search strategies, thus avoiding the problem of selecting search filters

Discussion
A key limitation of the study not mentioned is that there is no assessment of the effect of the missing 56 relevant trials on the conclusions of their reviews. Until this is done, we won’t know whether we can “afford” the increased efficiency of skipping Phase 3 of the HHHS Cochrane strategy.

Accepted. We added the following to the conclusion section (P 12):
First, we tested only how many included studies could be retrieved by each of the six search strategies, not whether the 56 relevant studies missed by SS_{versus} would change the outcomes of the 61 reviews. Therefore, we can not judge whether skipping Phase Three of the HSSS would result in bias in systematic reviews.
Version: 2 Date: 15 September 2005
Reviewer: Ellen Crumley
Reviewer's report:
Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

- There are still grammatical errors and missing "the's" in the article. I would recommend another proof-read to correct these.

Thanks! We had it proof-read.
Reviewers report:

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

1. I don't feel that the added text fully addresses the issue raised in my previous review (#5) nor is it, I think, entirely correct. It is not clear to me on what basis other studies concerning the development of optimal search strategies are dismissed as being focused on comprehensiveness alone. For instance, precision was considered in my paper referenced, including in the testing of the addition of CROSS-OVER-STUDIES. The authors provide the Lefebvre and Clarke chapter as a rationale for testing the four terms but do not address how this study fits with other work. I think there needs to be a stronger context set for this study in the introduction -- Why these four terms? Have they been tested previously (yes)? Why were other terms that have been tested not included?

We do not set out to dismiss other studies, rather they are out of scope for this paper. We will try to express that more clearly. We set out merely to validate these variants proposed without supporting evidence in Lefebvre and Clarke’s chapter. Robinson implies in this comment that the variants have been previously tested. We are not aware of published results of that testing. If we have overlooked published results then our manuscript may be redundant and not make enough of a contribution to the literature to warrant publication.

Unlike the Lefebvre and Clark paper, Robinson’s paper with Dickersin does support its proposed amendments with empirical evidence. It is beyond the scope of this manuscript to compare terms already tested. For instance, in addressing Robinson’s comment, we examined the strategy presented by Robinson and Dickersin. If we were to consider their changes to HSSS$_{12}$ alongside those proposed but not validated by Lefebvre and Clarke, we would have to first replicate Robinson and Dickersin’s search. We immediately are faced with disagreement between the text and the figure as to exactly how the “latin square” term was used. Would we render it as “latin[tw] AND square[tw]” or “(latin adj square).tw.”? We encourage Robinson and Dickersin to clarify in an erratum whether the text or figure is correct. We also do not know if Robinson and Dickersin are reporting gains relative to the HSSS run with .ti,.ab., as Robinson asked us to do in a previous major compulsory revisions, or with .tw. as they present in their revised strategy (they do not present the version of the original strategy that they used).

We have tried to alter the text throughout to make clearer what the modest goals and scope of this paper are. The new description of the relevant Introduction section is (p.4-5):

A few recent studies [13,14] have explored different search strategies to identify RCTs in MEDLINE. Haynes and colleagues [13] developed
separate strategies for different purposes: strategies with high sensitivity for comprehensive searching and strategies with high precision for more focused searching. Robinson and Dickersin [14] recommend that reviewers searching PubMed use a revision of all three phases of the HSSS, which has a better performance than the original HSSS. Because an increasing number of systematic reviews have to be completed within tight budgets and timelines, it is sometimes necessary to strike a balance between comprehensiveness and precision. Several proposals to refine that balance by making minor modifications to the first two phases of the HSSS were put forth by Lefebvre and Clarke in 2001 [11]. A comprehensive literature search reveals that there are no published data that evaluate the performances of the four search strategies proposed. Because balancing the initial retrieval size greatly improves the efficiency of a systematic review, we tested the performances of these four proposed revisions of the HSSS: combining the top two phases of HSSS with the free-text terms *volunteer*, *crossover*, *versus*, and the MeSH term **CROSS-OVER STUDIES**, respectively [11].

2. (Last sentence) How do the limits in using precision and sensitivity influence the study results and interpretation? At the moment it seems as thought that sentence was simply dropped in to the text.

The limitations in using precision and sensitivity to evaluate information retrieval are well discussed by Kagolovsky and Moehr, and we have added that reference. These limitations are related to all studies using the two measurements to test information retrieval, not particularly to this study. Therefore, we do not think it is necessary to list the limits. If readers are interested in these, they may read the original article.

3. The conclusions may be strengthened if there was more specific discussion of the cost-benefit balance of using the suggested strategy. For instance, the conclusion in the abstract is that adding ‘versus’ is able to balance the sensitivity and precision of the two phases of the HSSS. Is it worthwhile to use the strategy that found 3 more studies? How many additional reports had to be read to identify those 3 studies? In general, I feel that the interpretation and conclusions by the authors are not supported by the results.

Thanks! We added two measurements to better illustrate our points:

1) We added Article Read Ratio, defined as the median of articles initially retrieved divided by the median of included studies retrieved.

2) We calculated the estimated time to finish a review for each search strategy based on the equation developed by Allen and Olkin.

More specifically, the following statements were added to the Results section (p.9):

Table 4 also shows the Article Read Ratio (ARR), which is defined as the median of articles initially retrieved divided by the median of included studies retrieved.
The ARR of SS_{123} (182) is significantly higher than that of SS_{versus} (54), and the ARRs of SS_{12}, SS_{crossover}, SS_{CROSS-OVER STUDIES}, and SS_{volunteer} are the same (51). We calculated the estimated time to finish a review for each search strategy based on the regression equation developed by Allen and Olkin [7]: time \(= 721 + 0.243x - 0.0000123x^2\), where \(x\) denotes the number of articles initially retrieved, as shown in Table 4. The time needed to finish a review is 1086 hours for SS_{123}, 823 hours for SS_{versus}, 818 hours for SS_{12}, SS_{crossover}, SS_{CROSS-OVER STUDIES}, or SS_{volunteer}.

The following statements were added to the Discussion section (p.10):

Across the 61 reviews, the performance of SS_{crossover}, SS_{CROSS-OVER STUDIES}, and SS_{volunteer} are the same as SS_{12}. SS_{123} found 56 more relevant articles than SS_{versus}, and SS_{versus} found 3 more relevant articles than SS_{12}, SS_{crossover}, SS_{CROSS-OVER STUDIES}, or SS_{volunteer}. The number of articles needed to read per review when SS_{versus} is used is about 1/4 of that when SS_{123} is used, and the estimated time to finish a review for SS_{123} is 32% higher than that for SS_{versus}. On the other hand, the number of articles needed to read when SS_{versus} is used is only 5% (22 additional articles) more than that when SS_{12} is used, and the estimated time to finish a review for SS_{versus} is 0.6% (5 hours) more than that for SS_{12}. The result shows that, compared to SS_{123}, SS_{versus} will reduce the number of articles needed to read significantly, thus reducing the reviewers’ work in assessing citations for eligibility and the total time to complete a review, while still maintaining a workload comparable to SS_{12} but a slightly better sensitivity than SS_{12}. Although the other three proposed search strategies also have a lower initial retrieval size than SS_{123}, their sensitivity is the same as SS_{12}.

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

1. There are some issues with the text that need to be addressed. For example, page 10, sentence about work to make CENTRAL comprehensive. Also, need to add ‘the’ before HSSS in a consistent manner.

Thanks! We corrected these.

Discretionary Revisions (which the author can choose to ignore)

1. Thank you to the authors for clarifying that in the current OVID interface [tw] searches only the title and abstract information. However, [tw] is generally defined as an "alias for all fields in the database" and, in other interfaces such as PubMed, [tw] does produce this broader less precise search. There is potential for confusion, especially for those wanting to translate the authors’ strategy into PubMed. As another source of confusion, the Cochrane HSSS in OVID format uses [ti,ab] not [tw] (See Appendix 1). I would suggest that the authors consider using the [ti,ab] format in their search strategy to avoid these potential areas for confusion.
Accepted. We changed these in Appendix 1.

2. Methods - I would suggest clarifying whether the excluded articles list that had to be included in the systematic reviews was a listing based on all screening levels or full-text only.

The Cochrane Handbook for Systematic Reviews of Interventions 4.2.5 (http://www.cochrane.org/resources/handbook/handbook.pdf) provides the following guidance to reviewers:

**Excluded studies:** Studies that specifically do not meet the inclusion criteria and are not included in the review should be listed here.

We cannot know the basis used by the individual review teams to list a study there, we suspect that the excluded articles list in the reviews was a listing based on the full-text screening level.

3. (First paragraph, page 11) I would appreciate clarification about how or in what way this study has more practical significance for people conducting systematic reviews.

We added more clarification (p12-13):

We replicated the search strategies used in real systematic reviews, therefore, our method and results may have more practical significance for those conducting systematic reviews because they provide quantified data, including initial retrieval size and time needed to finish a review; to describe the cost-effectiveness of each search strategy. Systematic reviews should use evidence-based methods, and the validation of search filters is important in that context. Our contribution is to provide methodological rigor to previous non-validated recommendations. Thus reviewers can make informed decision based on this evidence.