Author's response to reviews

Title: An optimal search filter for retrieving systematic reviews and meta-analyses that evaluate the effectiveness of public health interventions

Authors:

Edwin Lee (elee2011@meds.uwo.ca)
Maureen Dobbins (dobbinsm@mcmaster.ca)
Kara DeCorby (kdecorby@health-evidence.ca)
Lyndsey McRae (lmcrae@health-evidence.ca)
Daiva Tirilis (dtirilis@health-evidence.ca)
Heather Husson (hhusson@health-evidence.ca)

Version: 3 Date: 30 November 2011

Author's response to reviews: see over
Dear Mr. Aldcroft:

I am pleased to submit the enclosed revised manuscript entitled: An optimal search filter for retrieving systematic reviews and meta-analyses that evaluate the effectiveness of public health interventions, for publication in *BMC Medical Research Methodology*. This paper reports on the performance of the health-evidence.ca Systematic Review search filter for retrieving systematic reviews and meta-analyses that evaluate the effectiveness of public health interventions. We believe this manuscript will contribute to search filter literature, specifically methodological filters, since the health-evidence.ca Systematic Review filter provides a high degree of sensitivity as well as precision, saving considerable time and resources for those conducting literature searches in the field of public health.

We appreciate the reviewers’ feedback and have made revisions to the manuscript to reflect them below.

**REVIEWER 1**

Comment 1
This paper presents a new search filter to identify systematic reviews of public health in three databases. However, it does not describe how the search strategy was developed. The authors provide lots of details of testing, but the process by which search terms were identified, chosen and combined into the SR strategies is not described. The authors say they tested and developed their new strategy on a test gold standard but do not describe the process. The use of two different gold standards suggests to me that the authors used their test set of 53 MEDLINE records to provide terms for the strategy – which is why they then needed to demonstrate performance on a validation set – but this is not stated in the paper. If the test set was not used to develop the strategies then it can possibly count as an additional validation set. Providing more information on the identification and selection of search terms will enhance the transparency of the methods and clarify the use of the test gold standard.

We have addressed these comments in the background section: we have qualified our use of ‘gold standard’ and elaborated on our processes for developing our SR search filter, re-positioned a portion of the methods section to the background section where we provide more details on the development and validation data sets and testing of the filter.

Comment 2
The development of the gold standards is described in more detail but could be clearer- my understanding is that the authors’ PH search filter was used to compile a relevant set of records for two time periods (2 gold standards), and then the performance of the authors new strategy (or strategies?) and other published strategies were tested against those two gold standards. If the gold standard was developed using the authors’ PH strategy then it is unlikely that the new filter would “potentially capture additional relevant systematic reviews that our previous search strategy may have missed.”
We have provided additional details on the development of the gold standard set in the background section. To clarify the reviewer’s comment re: capturing additional relevant SRs, the PH search filter is one component of three search strategies that contribute to the gold standard – we also handsearch over 40 journals and search the reference lists of all relevant reviews included in the health-evidence.ca registry. The gold standard represents the screened sum of all of these searches.

Comment 3
p.9 One key issue when comparing the performance of a new filter with other published filters is that we have to be sure that our definitions of the methods being sought are similar: the authors do not discuss whether the systematic review hedge being used in PUBMED and the other published filters tested were designed to find systematic reviews which met the authors’ own definitions of a SR. The definition of SR used for this research and to select documents for inclusion in the authors’ resource should be included.

The goal of the comparison of the health-evidence.ca SR filter with other published filters, along with health-evidence.ca’s definition of a relevant review has been added to the last paragraph of Background.

Comment 4
The authors have translated what they describe as the systematic review hedge being used in PUBMED into 4 search terms in the Ovid syntax. However, the PUBMED SR filter is lengthy (http://www.nlm.nih.gov/bsd/pubmed_subsets/sysreviews_strategy.html) and it would be helpful to know if the authors have used a cut down version of the original.

This portion of the methods section (pg. 11) has been clarified. We mistakenly referred to the PubMed SR filter in this section and that reference has been removed.

Comment 5
The authors mention that they adapted the SR strategy to run in CINAHL and EMBASE, but do not discuss the detailed issues involved with such as translation. It is important to know how this was achieved in the interests of fair comparison and transparency.

We have added details of how we translated the SR filter to EMBASE and CINAHL in the first paragraph of the methods section (pg. 10).

Comment 6
P5. – ‘all systematic reviews’ – this claim needs to be contextualised – SRs for a specific issue and according to the authors’ definitions of a SR.

We have removed ‘all’ and added context to our statement of the type of SRs available on health-evidence.ca.

Comment 7
P6 – the authors define ‘public health’ – this is helpful but might be even more useful earlier in the paper. Some punctuation is missing from the definition.

We have moved this definition to the first paragraph in the background section.

Comment 8
P6. “remains limited in comparison to primary studies” – this should probably be qualified, i.e. limited in number/volume.
Comment 9
P6. “one-tenth were indexed as reviews, with only 1.5% of those indexed as systematic reviews [28,29]. Currently, there is no MEDLINE subject heading term used for “systematic review.”” This is contradictory and needs revisiting. In the next sentence the authors talk about a ‘consistent indexing term’ – the term can’t be consistently used if it is not available.

This statement has been revised and clarified (pg. 7).

Comment 10
P8 “Prior to 2008, the Public Health (PH) search filter….”
It is also not clear whether the two components were combined using AND or OR. It would be best just to show the filter.

Revised as per reviewer feedback; elaborated on description of PH search filter (pg 8 & 9).

Comment 11
P8. “(0.3% capture rate)” This ‘capture rate’ is the precision of the search.

Revised as per reviewer feedback; replaced ‘capture rate’ with ‘search precision’.

Comment 12
p8 “Thus, we were reasonably confident that our retrieval methods were capturing a near complete set of relevant articles, and could be considered the gold standard.” I suggest that the authors qualify this to indicate that this one possible gold standard for the types of systematic reviews in which they were interested obtained from the resources they searched. It is unlikely to represent a complete gold standard of all PH systematic reviews.

Revised as per reviewer feedback; we have further explained and qualified our use of ‘gold standard set’.

p.10 The authors should define specificity in the same detailed way as they define sensitivity.

Revised as per reviewer feedback; Definition of specificity detailed similarly to sensitivity.

Comment 13
The authors should present the PH strategy in the paper – we need to see if it is very different to the SR strategy.

Revised as per reviewer feedback; Included in second-last paragraph of background section.

Comment 14
p.13 sensitivity: The authors have 13260 in Table1 and 13259 in the text. The sensitivity is only presented for the SR filter (89.9%) – it is better for PH, but not presented in the text.

The in-text number has been updated (to 13260), and written results for the PH filter’s sensitivity has been included.

Comment 15
p15 heading: Performance of SR Search Filter to Others. Suggest adding in a word such as ‘compared’ and clarifying ‘Others’.
Revised to ‘Comparison of the health-evidence.ca SR Search Filter to Other Published SR Filters’

**Comment 16**
In the results the authors seem to have a desire to improve the performance of their PH filter in terms of increasing precision while maintaining sensitivity, but the desired scale of improvement is never established and the actual objective is not clearly specified. Was any improvement helpful, or were specific levels sought? In addition the comparisons with other filters are indiscriminate: if the objective of the authors’ filter is to be highly sensitive then comparisons against precision maximising or best compromise filters (balanced queries) should be justified – those filters are unlikely to perform as well as the authors’ filters because those filters were not designed to be sensitivity maximising. Conversely the authors’ filters can’t match the precision of filters designed to be precision maximising.

The reviewer’s comment has been addressed by describing the overall goal and how it related to each of the four variables examined in the paper (sensitivity, specificity, precision, and number needed to read) in the introduction of the paper. This new content appears in the last paragraph before the methods section and is revisited in terms of how the goal was or was not achieved in the second paragraph of the discussion section.

**Comment 17**
p21. Typo: ‘therebyeliminating’
Corrected.

**Comment 18**
p.21. “Although a filter may perform exceptionally well on any single outcome, it is the balance of performance across these four domains – sensitivity, specificity, precision, NNR – that distinguishes the best filter.” I suggest that the ‘best filter’ depends on the user’s priorities. Precision and NNR are different views on the same measure so can’t really be balanced. The relationship between the four performance measures is usually seen as a tradeoff because usually one measure is being favoured against the others because of the focus of the filter user (sensitivity or precision). ‘Best balance’ is also mentioned in the last paragraph on this page, but the authors really seem to be valuing sensitivity over the other performance measures. This perspective does not imply achieving balance across the measures but more a willingness to trade-off low precision to achieve the high sensitivity needed for the purposes of their work. The authors go on to be more explicit about their perspective at the end of page 21 and on page 22 where sensitivity greater than 85% seems to be established as a benchmark. I suggest it would be best to be clear about objectives in terms of acceptable levels of sensitivity etc early on in the paper and then the discussion can be more focused. It also allows comparative terms such as ‘best’, and ‘right balance’ to be contextualised.

The reviewer’s point is well-taken and is addressed by acknowledging in the background (last paragraph prior to methods) that a trade off was anticipated and that high precision and sensitivity were prioritized. Additionally, the discussion was enhanced to address how the SR filter offered the desired combination of priority features (second paragraph of discussion).

**Comment 19**
p21. ‘Other authors’ might be best replaced with ‘One study’.

Revised as per reviewer feedback.

**Comment 20**
p.21 ‘Saving time both in search strategy development and screening of results was the most common benefit of search filters reported by librarians [41].’ I think this reflects the desired benefit of filters rather than the benefit necessarily achieved in practice?

Revised as per reviewer feedback; this statement is qualified with ‘the desired benefit’.

Comment 21
The authors cite the 5 line SR filter on PubMed here as well – this is perplexing because the filter does not have 5 lines.

This reference has been removed from the discussion section.

Comment 22
p.23 Limitations. The adaptation issues involved in translating a strategy into EBSCO should be described in more detail as it may impact on the filters’ performance.

Revised as per reviewer feedback; we acknowledge the impact and limitation of previously validated search filters when platforms change (limitations section, first paragraph).

Comment 23
p.24 The authors state ‘The SR search filter may not perform as well for topics outside of public health and health promotion’. However, the strategy only has SR terms so shouldn’t it perform similarly in any topic?

This statement has been removed from the limitations section.

Comment 24
p.24 In the conclusion the authors use the phrase ‘significantly reduces’ – is this statistical significance?

Revised as per reviewer feedback; revised conclusion to remove ‘significantly’.

REVIEWER 2

Comment 1
The authors use an indirect style which makes it difficult to follow their line of argument.

Revised as per reviewer feedback; accepted suggested re-wording.

Comment 2
"700,00" - missing digit(s)

Corrected.

Comment 3
"Furthermore, while descriptions of these search filters had been published, databases such as MEDLINE had not incorporated the use of these filters within their search functions”. - I am not sure that this is true given that PubMed has the Clinical Queries feature, including a systematic review filter, doesn't Ovid also have a built-in reviews filter?
Effective retrieval of relevant and appropriate numbers of articles is challenging. The aim of this study is “to report on the performance of a methodological filter – the health evidence.ca Systematic Review search filter – for retrieving systematic reviews and meta-analyses that evaluate the effectiveness of public health interventions.” However, the authors provide little basis for understanding the development of their initial search which they describe as a search filter. Indeed, it is difficult to assess whether it would meet the criteria for search filter development articulated by Glanville, Bayliss et al in 2008 and further developed by Bak, Mierzwinski-Urban et al in 2009. The search in question appears to have been adapted from an earlier filter designed to retrieve systematic reviews, which has been combined with their expert content search for public health content. No information on the basis or testing of the adaption is provided or on the development of the content component. In essence this study attempts to validate the “search filter” by comparing it to the previous content search and to other systematic review search filters. However the latter comparison involves a search that combines 2 potential searches (i.e. systematic review search and a public health search) against search filters only searching for systematic reviews. This introduces an unknown set of biases into the comparative data. For example, were the characteristics of public health systematic reviews somehow atypical of systematic reviews across all clinical conditions or health circumstances. It may have been more useful to test the effectiveness of the systematic review component of the search by comparing the reported search against other filters that were combined with the content search for public health. There is also no detail on how the searches uses in the first comparison were “modified’ for use in the various databases. The net effect of these concerns is that as a reader I was:

Comment 1
Unable to discern if the proposed “search filter” had been developed experimentally and following an appropriate methodology

Comment 2
Uncertain of how to interpret the relative returns of the public health versus the systematic review search given the nature of the gold standard, that is already pre-selected as both a systematic review and public health, and

Comment 3
Unable to assess the value of comparing a systematic review and content search against systematic review only searches as a validation of the effectiveness of the proposed search

I would suggest that these areas need to be more carefully considered, more clearly explained and more effectively analysed to claim that an optimal search filter has been developed.
In response to this paragraph leading into itemized Comments, we have further provided basis for understanding the development of the initial search filter, as this was a Comment of Reviewer 1 as well. Secondly, the paper has been revised so that the CADTH critical appraisal instrument (Bak, Mierzwinski-Urban et al, 2009) can be applied to the SR search filter and its development (ie. Information about its intended purpose, filter design, gold standard, filter validation, limitations and ‘killer’ question). Next, the reviewer’s comment that “this study attempts to validate the “search filter” by comparing it to the previous content search and to other systematic review search filters. However the latter comparison involves a search that combines 2 potential searches (i.e. systematic review search and a public health search) against search filters only searching for systematic reviews” has been addressed by better delineating the comparisons made in the paper (second paragraph of methods). We shifted away from incorporating the content search and have explained the rationale for this shift in the limitations section. We felt that the combining of the methods portion of the PH search filter with the content portion of the PH search filter was too limiting in that it could have potentially excluded PH reviews that were not well-indexed. In comparing the PH search filter to the Health-evidence.ca SR search filter, it became apparent that removing the content portion of the filter was appropriate. The reviewer’s comment that there is no detail on how the search strategies were adapted for use in the various databases is described further in the first paragraph of the methods section and each search is detailed in the accompanying Tables 1-3.

Response to Reviewer 3’s Comments 1-3:
The reviewer’s comments have been addressed by an overall edit of multiple sections of the paper in an attempt to provide more clarity around methods used, additional context to offer support for methods, and further detail as noted by the reviewer. It is hoped these edits will be well received.

Comment 4
It would be helpful to provide a definition of a systematic review as is included in health.ca. (e.g. according to PRISMA, quantitative only, qualitative,) as this may affect the effectiveness of the individual search filter.

Addressed in response to Reviewer 1, Comment 3; health-evidence.ca’s definition of a relevant review has been added to the last paragraph of Background.

Comment 5
P5 Sentence 5: It is a bold claim that decision makers have access to all systematic reviews published since 1985.

This sentence has been revised ‘Using health-evidence.ca, decision makers have easy access to public health-relevant, quality-appraised systematic reviews evaluating the effectiveness of public health interventions.’

Comment 6
Use of significance/significant: Please do not use a significant reduction unless there has been a statistical test for significance applied.

Revised as per reviewer feedback; reviewed manuscript and rephrased where inappropriate to use significance/significant.

Comment 7
P21 L6: I suspect the best filter is determined by the needs of its purpose not necessarily a “balance”. If, for example, I wish to use a search filter to identify articles reporting studies on a
particular clinical outcome for a systematic review, the choice would be for sensitivity as the potential costs of non-identification are high.

The reviewer’s Comment 7 was addressed along with a similar comment made by Reviewer 1 (Comment 3). Additional clarification was added to the background section (last paragraph before methods) as well as to the Discussion (second paragraph).

Comment 8
P24 L3 Low precision is surely not a function of low prevalence as precision is a measure or relevant records retrieved by the search /all records retrieved by the search. The search may only retrieve all relevant articles from an extremely large database if the characteristic is distinct enough still providing high precision.

Reviewer feedback was considered and the authors have removed the sentence that was unclear as there were other cited works that indicated the point.

Comment 9
P24 L7: I think a greater disclaimer about the potential impact of the gold standard selection is required as there is a real possibility of bias. The reader for example does not know if the searches to generate the dataset for the 2004-2005 validation set were based on the content search embedded in the systematic review.

Reviewer feedback incorporated; disclaimer enhanced in last sentence of limitations paragraph.

I hope you will find these revisions acceptable. We look forward to hearing from you.
Please do not hesitate to contact me at dobbinsm@mcmaster.ca or (905) 525-9140 extension 22481, should you have any questions or concerns.

Sincerely,

Maureen Dobbins, RN, PhD
Assistant Professor
McMaster University School of Nursing
Tel: 905 525-9140 ext 22481
Fax: 905 524-5199
E-mail: dobbinsm@mcmaster.ca