Reviewer's report

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

Version: 1 Date: 1 August 2011

Reviewer: Julie Glanville

Reviewer's report:

This is an interesting paper which would benefit from the provision of additional detail on the methods employed to develop the search filter and some further clarifications to help the reader understand the processes and perspectives adopted in this paper.

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.

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.”

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.

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.

5. The authors mention that they adapted the PUBMED 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.

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.

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.

8. P6. “remains limited in comparison to primary studies” – this should probably be qualified, i.e. limited in number/volume.

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.

10. P8 “Prior to 2008, the Public Health (PH) search filter we used was a content-specific search filter. Our content specific filter, developed in collaboration with health sciences librarians, consisted of two distinct components: indexing terms referring to systematic review methods, and indexing terms referring to public health content areas.” The authors call the filter ‘content-specific’ and then say it was ‘partly’ content-specific – they need to choose how to describe it. It is also not clear whether the two components were combined using AND or OR. It would be best just to show the filter.

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

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.

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

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

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.

15. p15 heading: Performance of SR Search Filter to Others. Suggest adding in a word such as ‘compared’ and clarifying ‘Others’.

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.

17. p21. Typo: ‘thereby eliminating’

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.

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

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?

p.23 The authors cite the 5 line SR filter on PubMed here as well – this is perplexing because the filter does not have 5 lines.

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.

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?

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

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Acceptable

Statistical review: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests:

I declare that I have no competing financial interests. I am co-author on a number of search filter papers: some of the filters in those papers have been tested by the authors in this paper.