Author's response to reviews

Title: Glomerular disease search filters for PubMed, Ovid Medline, and Embase: A Development and Validation Study

Authors:

Ainslie M Hildebrand (ainslie.hildebrand@londonhospitals.ca)
Arthur V Iansavichus (Arthur.Iansavitchous@lhsc.on.ca)
Christopher WC Lee (clee2011@meds.uwo.ca)
R. Brian Haynes (bhaynes@mcmaster.ca)
Nancy L Wilczynski (wilczyn@mcmaster.ca)
K. Ann McKibbon (mckib@mcmaster.ca)
Michelle A Hladunewich (Michelle.Hladunewich@sunnybrook.ca)
William F Clark (William.Clark@LHSC.ON.CA)
Daniel C Cattran (Dr.Daniel.Cattran@uhn.on.ca)
Amit X Garg (Amit.Garg@lhsc.on.ca)

Version: 2 Date: 21 February 2012

Author's response to reviews: see over
February 20, 2012

RE: Research article revisions
Glomerular disease search filters for PubMed, Ovid Medline, and Embase: A Development and Validation Study

Dear Drs. Glanville and Sladek,

We are pleased to submit our revised manuscript entitled “Glomerular disease search filters for PubMed, Ovid Medline, and Embase: a development and validation study”, for consideration for publication. We have considered each suggestion for revision and have addressed them point-by-point in the attached document. The two primary concerns brought up on your review were the lack of acknowledgement of search filter work other than our own and the unclear methodology used for creation and reporting of proof of concept searches. We have carefully considered each of these issues and have modified the text of the manuscript considerably to reflect these concerns.

We anticipate these filters will contribute to more efficient evidence retrieval and more effective evidence-based decision making for all physicians dealing with glomerular disease. Thank you for considering our manuscript for publication in BMC Medical Informatics and Decision Making. We are open to any further comments or suggestions you may have to strengthen this research.

Kind regards,

Ainslie Hildebrand MD, FRCPC
Nephrology Post-Graduate Year 5
MSc Candidate Epidemiology & Biostatistics
University of Western Ontario
London, Ontario, Canada
Response to suggested revisions for manuscript entitled ‘Glomerular disease search filters for PubMed, Ovid Medline, and Embase: a development and validation study’.

Ainslie M. Hildebrand MD, Arthur V. Iansavichus MLIS, Christopher W.C. Lee MD, R. Brian Haynes MD, PhD, Nancy L. Wilczynski PhD, K. Ann McKibbon PhD, Michelle A. Hladunewich MD, William F. Clark MD, Daniel C. Cattran MD, Amit X. Garg MD, PhD

Reviewer 1: Julie Glanville

Response to major essential revisions:

1. Define ‘glomerular disease’ in the text and appendix.

   • We have clarified the definition of glomerular disease used for this study in the methods section as well as in Appendix B. Glomerular disease was defined as any disease in which the glomerulus of the kidney is affected, resulting in hyperplasia, atrophy, necrosis, scarring, or deposits in the glomeruli.

2. Clarify the methods used for proof of concept searches.

   • We have expanded on the methods used for proof of concept searches in the manuscript. Essentially, six independent and randomly selected nephrologists from across Canada were asked to specify the search terms that they would use to retrieve relevant studies for a clinical question that we generated from recent systematic reviews in glomerular disease. These search queries were created without knowledge of the search filter or database in use. Each search query was then applied to the entire PubMed database with and without the best performing filters developed as part of this study. Article retrieval was compared with the reference standard, which in this case was the set of relevant articles as determined by each systematic review. However, it is important to note that proof of concept searches were only used to illustrate the functionality of these search filters with real physician searches and were not used to develop or validate our search filters.

3. What is ‘utility’ when referring to proof of concept searches?

   • The term utility was initially referring to the effectiveness of the search filter, or the functionality of this filter with real physician searches in the PubMed database at large. We have clarified this concept throughout the manuscript and have reworded the paragraph on proof of concept searches in the methods section.

4. How were nephrologists identified in the proof of concept searches?

   • Six independent nephrologists were selected at random from the directory of Canadian nephrologists provided by the Royal College of Physicians and Surgeons of Canada. We have clarified this in the proof of concept section of the methods.

5. Was a relative recall approach used in proof of concept searches?

   • Yes, a relative recall approach was used for the proof of concept searches, however this method was not used in the development or validation of our filters as described
in the statistical analysis section of the methods. The methods used for defining the reference standard based on articles used in only a few systematic reviews of variable quality is indirect and has not been compared with one derived from hand searching. For this reason, the proof of concept searches should be viewed as illustrative examples, not as evidence of further filter validation. We have discussed this limitation in reference to proof of concept searches in the discussion section.

6. Who developed the searches for proof of concept searches?

- Searches were devised by six independent and randomly selected nephrologists from across Canada based on clinical questions we generated from recent systematic reviews in glomerular disease. We have clarified this point in the proof of concept searches section of the methods.

7. For the proof of concept searches to remain in the paper, the authors should present the detailed results with figures in the results section. In particular, the low precision scores should be presented so that clinicians can appreciate the tradeoffs between sensitivity and precision

- Given that proof of concept searches were used only for illustration purposes, we are reluctant to include performance measures for the searches done within this portion of the study. Search filters were developed and validated as described in the methods section of the study and performance measures for the best performing filters are provided for these filters in both the development and validation set of articles in Table 2. There is too much variation in the quality of search terms entered by the users and accuracy of the reference standard (systematic reviews) for performance measures to be appropriately interpreted in the context of proof of concept searches. We have clarified the purpose of these proof of concept searches throughout the manuscript and hope that this portion of the study may still be included.

8. Discuss the strengths and weaknesses of the approach used for proof of concept searches.

- We have clarified the purpose of proof of concept searches in the methods section as described in point 2 of major essential revisions above and have elaborated on the strengths and weaknesses of using this approach in the discussion. Although the methodology used for proof of concept searches is not strong enough to use to validate our search filters as described in point number 5 of major essential revisions above, they do provide illustrative examples that we hope will lead some readers to try out these filters for their own searching.

Response to minor essential revisions:

1. The first sentence of the background section of the abstract should be more conditional as direct results are unlikely to be traceable.

- We have reworded this sentence in the background section of the abstract to reflect this concern, as filters have not yet been shown to improve patient care.
2. Standardize on ‘physicians’ or ‘clinicians’ in the abstract.
   • We have standardized on ‘physicians’ throughout the manuscript.

3. Clarify what is meant by proof of concept searches in the abstract.
   • As this is not the focus of the study and was only used for illustrative purposes, we have eliminated mention of proof of concept searches from the abstract.

4. Qualify the word, ‘physicians’ in the background section as references refer to primary care/family practitioners.
   • We have reworded this sentence to refer to ‘users’ as these barriers to information seeking have been shown to apply not only to family practitioners, but also to consultant physicians, residents, and other experienced users. We have added an additional citation to reflect this.

5. Illustrate and cite some evidence to support the phrase ‘indexing of articles is often inconsistent’ in the background section.
   • Variable terminology used for similar clinical entities or histologic diagnoses is a well-known phenomenon among glomerular diseases and therefore it is plausible that this would result in indexing inconsistencies. We certainly noticed this during our filter development. For example, an article on membranous nephropathy may be indexed using the term ‘membranous’ with any of the following additional synonyms used to define the glomerular disease: ‘nephropathy’, ‘glomerulopathy’, or ‘glomerulonephritis’. Although we are not aware of any literature published to confirm the indexing inconsistencies that relate specifically to glomerular disease, we have cited the evidence on indexing inconsistencies that have been shown to be present in Medline. We have also provided an example later in the background section to illustrate how search filters can be used to reduce the effect of these indexing inconsistencies.

6. Consider search filters published by other groups.
   • We have expanded on the references used to support our filter development and have modified the text to reflect this concern. Many of these studies are included in the InterTASC Information Specialists’ Sub-Group Search Filter Resource online.

7. Define what is meant by reliability in the background section of the manuscript.
   • By ‘reliability’, we are referring to precision of the filter to consistently capture articles relevant to glomerular disease. We have reworded this sentence in the background section to clarify this point.

8. What does comprehensiveness refer to in the background section of the manuscript?
   • Comprehensiveness was used here to refer to the extent to which the search retrieves all articles relevant to glomerular disease. However, we have reworded this section of
the background to include an example of how this search filter may be used to improve retrieval of articles without using this term.

9. How were articles designated to the development and validation sets?

- Assignment of articles to the development and validation sets is described in detail in the sample of articles section in the methods. This method has also been used in our prior search filter studies, which have resulted in filters that generalize well over publication years and journal types. We have cited these studies to support our methods.

10. How were ‘best performing’ filters defined?

- Best performing filters were defined by cut off points based on what we deemed reasonable to accept on a practical level. Generally, we aimed to keep the sensitivity and specificity above 90%, and maximize either sensitivity or specificity and precision for high sensitivity filters and high specificity filters respectively. Given that we had not established a cut-off point a priori, we have not included this in the study methods.

11. The presentation of the methods is not in a particularly logical order. It would help to have all the detail presented within each step. Diagrams may also be helpful for clinicians.

- We have reorganized the methods section to reflect this concern and have clarified each section of the methods. We tried a diagram, but it seemed a little onerous for this particular paper. We do cite our BMJ paper which used a near identical process for filter development [1]. In that paper there is a flow diagram (presented below) that readers can refer to if they are really interested in more details about the methods.

12. How was the list of journals that had published at least one article relevant to renal care from 1961 to 2005 obtained?

- We have cited the paper in which these detailed methods have been published [2]. Essentially, we identified 195 renal systematic reviews from a detailed search of
Medline. Two nephrologists used a standardized checklist to independently confirm whether each review was pertinent to renal care. Pairs of reviewers also independently confirmed that each review met specified methodological criteria for quality. All primary studies that were published in journals were compiled and primary studies that were cited in multiple systematic reviews were only counted once. A list of 466 journals had at least one article relevant to renal care from 1961 to 2005. These journals were then ranked according to the number of articles with renal information.

13. How was relevance to renal care determined and how were the number of articles identified (in reference to ranking the journals for creation of the development and validation sets of journals)?

- We have cited the paper in which these detailed methods have been published and have summarized above in point number 12 of minor essential revisions.

14. How were the journals divided into the development and validation sets?

- This is described in detail in the sample of articles section of the methods. From a list of 466 journals created as described above, journals were ranked according to the number of articles with renal information and the top 20 journals were selected. Nineteen more journals were selected at random from the remaining 446 journals. These 39 journals were then randomly divided into the development and validation sets at a ratio of two to one respectively. Division was at the journal level, not the article level.

15. Provide detail on which Boolean operators where used and how for filter development.

- We tested 736,043 multiple terms that used different combinations of search terms with a variety of Boolean operators. In order to maximize sensitivity, the primary search terms were frequently linked with ‘OR’, however ‘AND’ and ‘NOT’ were used sparingly to maximize specificity when necessary. It is not practical for us to describe the detailed process for development of each different filter. High performance filters have been provided, with Boolean operators, in Table 2.

16. What is the rationale for calculating performance measures of each filter? Why do the clinicians need to know specificity and accuracy?

- Performance measures such as sensitivity, specificity, precision, and accuracy were used in the development and validation of filters to measure how well each filter was able to retrieve articles relevant to glomerular disease in the development and validation phases and to provide rationale for selection of high performance filters. These details are provided in the manuscript as evidence of how robust our filters are for retrieval of articles relevant to glomerular disease in an ideal environment. This information will be easily understood by physicians, as it is similar to the evaluation of a diagnostic test (and physicians are accustomed to reviewing articles that describe these measures in the evaluation of a diagnostic test). Providing these measures will help guide physicians as to what filter to use for a given purpose. For example, high
sensitivity filters are optimal for physicians who need to be sure they have maximized their identification of all potentially relevant articles (which is important for systematic reviews and clinical practice guideline development). Physicians should be aware of the loss of precision when using the high sensitivity filters and the potential for loss of relevant articles when using the highly specific and precise filters to minimize non-relevant articles. Accuracy was high in all cases and was only included for completeness.

17. What were the cut off levels for high sensitivity and specificity used to select the high performance filters. What level of precision was being (ideally) sought?

- This was addressed above in point number 10 of minor essential revisions.

18. How many of the records were relevant in the sample of articles.

- Of the 22,992 articles from 39 journals reviewed between 2004 and 2008, 640 articles were relevant to glomerular disease. We have provided this information in Appendix A.

19. What is the threshold for high sensitivity and high specificity for selection of multiple term filters?

- This was addressed above in point number 10 of minor essential revisions.

20. The performance in the validation set should be reported in numbers.

- We have reported the operating characteristics of high performance filters in the validation set (and development set) in the results section in addition to Table 2.

21. It is not clear why the topic is ‘glomerular disease’ and the search terms discussed are about kidney disease.

- Each of the search terms discussed are specific to glomerular disease. Given that glomerular diseases are a subset of conditions within the broad category of kidney disease, they will obviously be closely related. We have provided the definition used for glomerular disease in the manuscript and appendix as described above.

22. The readers need a more detailed exploration of what filters achieve in the introduction to better understand the results.

- We have expanded on the purpose and benefit of search filters in the background section and have provided an example of how this search filter may be used to find articles relevant to glomerular disease. Hopefully this will help readers better understand the results.

23. The discussion might become clearer with the addition of a diagram – as it is I don’t think a clinician will struggle to understand the point being made. There are issues of the use of the filter, plus subject terms to convey.
• We have clarified the discussion to help readers understand the benefits and limitations of this search filter and have purposefully used examples from the proof of concept searches that can be referred to in Table 3. Although proof of concept searches were not used in the development or validation of the search filters, they do provide illustrative examples of the functionality of these filters with real physician searches and provide a good reference point for discussing the benefits and limitations of the filter. We are not aware of a diagram that would help clarify the discussion further, but have directed readers to Table 3 when appropriate. We would encourage users to appreciate the search phrases entered by physicians, the number of relevant articles from the systematic review, and the number of non-relevant and relevant articles retrieved in each of the three cases (physician search alone, with high-sensitivity filter, and with high-specificity filter) in order to better understand the discussion. Also provided in the discussion is a link to a website where readers can try out the filters themselves. Any reader who is interested can follow this link, and can appreciate how these filters can be used in practice.

24. Some of the information about the clinician searches could be put in the results section.
• We have elaborated on the results of proof of concept searches in the results section and have referenced Table 3. Again, the development and validation of the filters was a very arduous task. For the proof of concept searches we simply presented some illustrative searches to show how the filters can operate in practice. We believed this important for clinical readers. Additional studies are needed to determine the utility of these search filters for physician information retrieval, physician knowledge, physician decision-making, and patient outcomes.

25. Stating that you expect the filter to be ‘equally robust’ to address questions that potentially relate to all glomerular disease is speculation without supporting evidence.
• Unfortunately we are unable to provide supporting evidence for the use of the filter in this way, but do want to highlight the potential other uses of the filter. We have reworded this paragraph in the discussion to make fewer assumptions.

26. How was high impact determined when referring to selection of journals?
• We did not take into account impact factor when selecting journals for our development and validation sets. We have reworded this sentence in the discussion to reflect the enrichment of the journals with leading nephrology journals.

27. Are all three measures of sensitivity, specificity, and precision being taken into account? This needs to be clearer.
• This was addressed above in point 10 of minor essential revisions. In all cases, all four performance measures were taken into account in selecting high performance filters. We have clarified this in the statistical analysis section of the methods.

28. Discuss the pros and cons of the suggestion to use multiple filters in combination.
Members of our group recently published a study on using filters in combination, specifically using a renal topic-based filter in combination with a methods-based filter on therapy in PubMed and reported a marked improvement in efficiency when using filters in combination [3]. We have cited this study in the discussion to support our suggestion of using our glomerular disease filter in combination with methods based filters and have acknowledged that this approach has not been formally tested. This will be a further area of research for our group.

29. Which filters are being used in combination in the conclusion?

- We are referring to the use of our filters developed as part of this study with previously developed methods filters. We have reworded this sentence in the conclusion to clarify this statement.

30. Were there any differences between performance in particular journal subsets?

- We did not analyze our results based on individual journals or based on high volume nephrology journals compared with randomly selected ones. Performance measures of each filter were only determined for the development set of articles (from multiple journals) and validation set of articles (from multiple journals) independently. There was a drop in precision of many of the filters when applied to the validation set of articles, which was expected as it was a smaller database of articles by design and had lower proportion of articles relevant to glomerular disease as shown in Appendix A. This was mentioned in the discussion of limitations.

31. In Table 1, the definition of precision reads the same as the definition of sensitivity.

- The definitions of sensitivity and precision provided were correct and distinctly different. However, we have re-worded these definitions in Table 1 to clarify the terms further. The 2x2 table is also provided with formulae for clarification.

32. Why does precision decline sharply in the validation set testing?

- This has been addressed in the discussion. Precision of each filter was expected to decline in the validation set of articles as it was a smaller database by design with a lower proportion of articles relevant to glomerular disease as shown in Appendix A.

33. In Table 3, the data are here to calculate all the performance measures – they should be provided so we don’t have to calculate this ourselves. This exercise shows high sensitivity, but very poor precision. Clinicians will find this disappointing, but need to know.

- This has been addressed in point number 7 under major essential revisions and has been clarified in the entire manuscript. Proof of concept searches were not used to measure filter performance and therefore it is not appropriate to provide detailed performance measures for each of the searches.
Reviewer 2: Ruth Sladek

Response to major essential revisions:

1. The authors do not place their research within the context of any broader body of literature around filters research other than their own. References and text should reflect the substantial body of UK search filter work.
   
   - Thank you for this feedback. We have revised the text and references to address this concern. Many of these studies are included in the InterTASC Information Specialists’ Sub-Group Search Filter Resource online.

2. The description of the methods for proof of concept searches is not clear.
   
   - We have expanded on the methods used for proof of concept searches in the manuscript. Essentially, six independent and randomly selected nephrologists from across Canada were asked to specify the search terms that they would use to retrieve relevant studies for a clinical question we generated from recent systematic reviews in glomerular disease. These search queries were then applied to the entire PubMed database with and without the best performing filters developed as part of this study. Article retrieval was compared with the reference standard, which in this case was the set of relevant articles as determined by each systematic review. However, it is important to note that this part of the study was simply used to illustrate the functionality of these search filters with real physician searches, and was not used to develop or validate our search filters.

3. Include precision and accuracy in the results and report the full range of sensitivities.
   
   - We have reported all performance characteristics in the results section with precision and accuracy and have provided ranges when appropriate.

Response to minor essential revisions:

1. In the study overview section of the methods, was this a ‘new’ reference standard process for this filter or is it capitalizing on an initial corpus of review work reported elsewhere? Was it dual review?
   
   - We have cited a prior study performed by members of our group that describe the methods of article sampling in detail [2]. However, manual review of each of the 22,992 articles for relevance to glomerular disease was performed specifically for this study. Three readers used a standardized checklist of qualifications and terms to classify articles as relevant to glomerular disease and were calibrated against a nephrologist in their application of checklist criteria. This has been explained in the article review section of the methods.

2. In the last sentence of paragraph 3 of the results, the meaning of the last sentence is unclear.
• We have included detailed performance characteristics of each filter in the validation set to clarify this point.

3. Can search dates be provided?

• We have not provided search dates for the proof of concept searches, as the search dates were restricted to the date on which the review was updated.

Response to discretionary revisions:

1. Should paragraph 2 of the discussion be separated as two paragraphs as it introduces a new focus for discussion?

• Yes. We have reformatted the discussion to reflect this concern.

2. In the discussion, the drop in precision in the validation set needs to be justified.

• Precision of each filter was expected to decline in the validation set of articles as it was a smaller database by design with a lower proportion of articles relevant to glomerular disease as shown in Appendix A. We have addressed this issue in the discussion.
