So many filters, so little time: The development of a Search Filter Appraisal Checklist

Julie Glanville, Sue Bayliss, Andrew Booth, Yenal Dundar, Nigel Fleeman, Louise Foster, Cynthia Fraser, Anne Fry-Smith, Su Golder, Carol Lefebvre, Rosalind McNally, Caroline Miller, Suzy Paisley, Liz Payne, Alison Price, Hasina Shaikh, Anthea Sutton, Karen Welch, Anna Wilkinson on behalf of the InterTASC Information Specialists' SubGroup

Julie Glanville MSc MCLIP

Project Director – Information Services, York Health Economics Consortium Ltd., Level 2, Market Square, University of York, York, YO10 5NH, UK.

jmg1@york.ac.uk

Sue Bayliss BA (Hons)

Information Specialist. ARIF/WMHTAC, Department of Public Health and Epidemiology, University of Birmingham, Edgbaston, B15 2TT, UK.

s.bayliss@bham.ac.uk

Andrew Booth MSc MCLIP

Director of Information Resources & Reader in Evidence Based Information Practice, School of Health and Related Research (ScHARR), University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA, UK.

A.Booth@sheffield.ac.uk

Yenal Dundar MD

Doctor, Department of Psychiatry, North Devon District Hospital, Raleigh Park, Barnstaple EX31 4JB, UK.

yenal@liverpool.ac.uk

Hasina Fernandes BA (Hons)

Information Specialist. National Institute for Health and Clinical Excellence, MidCity Place, 71 High Holborn, London WC1V 6NA, UK.

Hasina.Fernandes@nice.org.uk

Nigel David Fleeman MPH

Research Fellow. Liverpool Reviews and Implementation Group, School of Population, Community and Behavioural Sciences, University of Liverpool, Sherrington Buildings, Ashton Street, Liverpool, L69 3GE, UK.

Nigel.Fleeman@liverpool.ac.uk

Louise Foster MSc

Health Information Scientist. NHS Quality Improvement Scotland, Delta House, 50 West Nile Street, Glasgow, G1 2NP, UK.

louisefoster@nhs.net

Cynthia Fraser MA (Hons)

Information Officer. Health Services Research Unit, University of Aberdeen, Health Sciences Building, Foresterhill, Aberdeen AB25 2ZD, UK.

c.fraser@abdn.ac.uk

A. Fry-Smith BA (Hons)

Lead Information Specialist. West Midlands Health Technology Assessment Collaboration, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK.

A.S.FRY-SMITH@bham.ac.uk

Su Golder MSc

Information Officer. Centre for Reviews and Dissemination, University of York, York, YO10 5DD, UK.

spg3@york.ac.uk

Carol Lefebvre MSc HonFCLIP

Senior Information Specialist. UK Cochrane Centre, National Institute for Health Research,

Summertown Pavilion, Middle Way, Oxford, OX2 7LG, UK.

CLefebvre@cochrane.co.uk

Caroline Miller MA

Information Specialist. National Institute for Health and Clinical Excellence. MidCity Place, 71

High Holborn, London WC1V 6NA, UK.

Caroline.Miller@nice.org.uk

Suzy Paisley MA

Research Fellow. ScHARR, University of Sheffield, 30 Regent Street, Sheffield, S1 4DA, UK.

s.paisley@sheffield.ac.uk

Liz Payne PG Dip Lib MCLIP

Independent Information Specialist. *Note: Liz Payne would prefer not to give her address*

eapayne@go.com

AM Price MSc

Information Scientist. Wessex Institute for Health Research & Development, Mailpoint 728,

Boldrewood, University of Southampton, Southampton SO16 7PX, UK.

A.M.Price@soton.ac.uk

Karen Welch PG Dip Info Sci

Information Scientist. Wessex Institute for Health Research & Development, Mailpoint 728,

Boldrewood, University of Southampton, Southampton SO16 7PX, UK.

K.Welch@soton.ac.uk

Author for correspondence:

Julie Glanville, York Health Economics Consortium Ltd., Level 2, Market Square

University of York, York, YO10 5NH, UK. Email: jmg1@york.ac.uk

Abstract

Objectives

To develop a tool to assess the quality of search filters designed to retrieve records for studies with specific research designs.

Methods

The InterTASC Information Specialists' Subgroup (ISSG), a group of experienced healthcare information specialists, held consensus meetings to develop filter appraisal tools. The group developed a Search Filter Appraisal Checklist and a structured abstract using consensus methods and tested them using three published filters.

Results

A published appraisal checklist was assessed for suitability. The group developed a detailed Search Filter Appraisal Checklist that captured relevance criteria and methods used to develop and test search filters. The Checklist includes categorical and descriptive responses. A structured abstract accompanies the appraisal Checklist.

Discussion

The Checklist is a comprehensive appraisal tool to assist health sciences librarians and others choose between search filters. It reports filter design methods and search performance measures, such as sensitivity and precision. It also prompts filter developers by indicating which information on core methods needs to be reported to help librarians to assess filter suitability. The generalizability of the Checklist for non-methods filters remains to be explored.

Introduction

Search filters are developed to improve the efficiency and effectiveness of searching [1]. They are typically created by identifying and combining search terms to retrieve records with a common feature [1]. They can be expert-informed, research-based or a combination [1]. Information about the methods of filter development, along with the results of testing, is important to enable potential users to judge whether the filter may be relevant and reliable [1].

Over the last two decades research methods have been used increasingly to develop and test search filters, to make them more robust and reliable [1]. Research-based search filters feature in bibliographic databases such as PubMed and others have been developed to assist with international study identification exercises for databases such as CENTRAL and DARE [2-5]. Search filters are proliferating as librarians and researchers try to identify records reporting projects with specific designs to assist with evidence-based healthcare [2-5]. For example, there are at least 8 search filters available for retrieving diagnostic test accuracy studies from MEDLINE [6]. Even experienced health sciences librarians may be challenged to select appropriate filters and to advise researchers which, if any, to use for a particular search query.

In evidence-based health care many critical appraisal tools have been developed to assess the quality and relevance of research reports [7-9]. The UK InterTASC Information Specialists' SubGroup (ISSG), which supports the research groups providing technology assessments to the National Institute for Health and Clinical Excellence (NICE) in the UK, identified the need for such a tool to help its members select from the search filters on its website[10]. The ISSG decided to develop a tool to appraise search filters which would help their members, health sciences librarians and others to choose the most relevant filter for their needs. The ISSG was unable to identify an existing suitable tool and, therefore, decided to develop its own tool, the Search Filter Appraisal Checklist, described in this paper.

Methods

The Search Filter Appraisal Checklist was developed using consensus methods over three meetings of the ISSG during 2006 and 2007. ISSG members felt that, as a group of highly-skilled healthcare information specialists, they had the relevant skills to develop such a tool, having experience of publishing search filters, testing search filter performance, practising critical appraisal and developing checklists and structured abstracts.

Before the first meeting the ISSG members searched databases and their reference collections to identify existing tools. An ISSG draft checklist and a brief summary (see Figure 1) were designed. The checklists and the summary template were tested during the first meeting on a search filter by Zhang [11]. The strengths and weaknesses of the alternatives were discussed. Key checklist concepts were agreed to be:

- the focus and scope of the filter: limitations, generalizability and obsolescence;

- the quality of the methods used to develop the filter, specifically how gold standards of relevant records were identified, how search terms were identified, how the strategy was developed, and how the filter performance was tested on the test gold standard (internal validity) and on separate validation gold standards (external validity).

It was agreed that a checklist should contain both categorical and descriptive information. A checklist should avoid numerical quality scores, for individual elements and the overall tool, because of known difficulties in assigning scores to individual dimensions of a tool and in interpreting a final combined score [12].

The dilemma of reporting adequacy should be addressed by wording comments to indicate that the assessment of the quality of the design of the search filter has to be made from the (sometimes limited) information provided in the report. The format of a checklist should be flexible to cope with the variety of search filter design methods.

Following the first meeting a second draft of the ISSG Search Filter Appraisal Checklist was developed and then discussed by email.

The summary, complementing the checklist, was re-addressed. Two alternative templates were developed: a one hundred word summary and a longer structured abstract (see Figures 2 and 3). The structured abstract template was designed to describe the filter objective, the methods used to develop the filter, key validation data, any reported limitations of the filter design and additional comments as appropriate.

ISSG members pilot-tested the checklist and abstracts at the second meeting using three different filters which had different methods of filter design [13-15]. Two of the filters were from published articles with detailed methods sections and one was published on a website which reported little about its development. During the meeting the ISSG members discussed the usability, clarity, practicality and reproducibility of the draft Checklist and the two summary formats (Figures 2 and 3).

Following discussions the Checklist was revised and underwent a final round of feedback. The final ISSG Search Filter Appraisal Checklist and structured abstract format were agreed by the ISSG at a third meeting in April 2007.

Results

One checklist was identified by searches: Michelle Jenkins' search filter appraisal checklist [1].

The consensus of the first ISSG meeting was that the Jenkins' checklist was helpful but not entirely suitable. It focused on determining generally whether filter design methods were reported. The checklist did not focus on filter design details or offer opportunities to extract data from the study. For example, the Jenkins' checklist asked "Do the authors report clearly how the filter performance was tested?", but did not ask what performance testing was undertaken or prompt the assessor to report performance data. The Jenkins' checklist asked some highly technical questions that might be difficult for some assessors to answer, for example, "Does the gold standard have sufficient power to allow statistically significant results?" Some of the questions, such as "Are the methods of search term derivation clearly

described, and are they reasonable and likely to be effective?", were difficult to answer because they contained several elements. ISSG members agreed that the Jenkins' checklist was a helpful prompt, however, a more detailed checklist was required.

ISSG members agreed that the draft ISSG Checklist required further development. The stages of search filter design and purpose should be broken down into focused questions and the checklist should include data extracted from the publication. The members agreed that the ISSG Checklist should allow for narrative comment and that the brief summary (see Figure 1) was too brief to be a useful decision aid.

At the second meeting, the group agreed that the revised ISSG Checklist, tested against three filters, was an improvement. It captured relevance information more effectively. Members felt that the revision addressed the issue that the assessor can only assess what the filter author reports. It achieved this by recording the reporting of the design but also by including prompts which reflect issues of design quality. These prompts should alert assessors to consider whether (unreported) alternative approaches might have been more suitable.

The ISSG felt that the revised checklist was flexible enough to capture the growing variety of methods reported in search filter design. It could capture information about multiple gold standards and validation testing activity. It also allowed an assessor to report performance comparisons against other filters, which strengthened the information available for deciding between filters. The checklist, however, still required work to capture information on how strategies were derived from the selected search terms.

At the second meeting the ISSG members also chose between summaries. The structured abstract was agreed to be more helpful than the one hundred word summary because it captured the filter objective, the main methods used to develop the filter, any key validation data, and any major limitations to the filter design. It provided space to summarise the strengths and weakness of the filter design. ISSG members agreed that the abstract was

suitable for quick relevance assessment, with the Checklist offering the essential detail required for informed decision-making.

The final ISSG Search Filter Appraisal Checklist is shown in Table 1. Examples of completed Checklists are published on the ISSG website [16]. An example structured abstract is shown in Figure 3.

Discussion

The Search Filter Appraisal Checklist is being used by ISSG members to appraise published search filters. Checklists are completed by an information professional, checked by an independent assessor, and edited by the website editor for consistency. Copies of Checklists will be sent to the original authors of the filters and feedback received will be published. Completed checklists are published on the ISSG website [10].

The ISSG Search Filter Appraisal Checklist is designed to be comprehensive. Its structure follows the life-cycle of the process involved in developing a search filter from gold standard identification, search term selection, strategy development, testing and validation through to comparison with other filters. It may take time to complete but should provide clearer insight into the quality and suitability of a filter.

Health sciences librarians trying to decide between filters now have several tools. They can use the ISSG website to find appraisals of filters in the form of structured abstracts and Checklists. The abstract offers a rapid assessment of relevance and the Checklist offers more detailed information to assist with deciding whether a filter is useful. Alternatively, librarians can complete the blank Checklist themselves, to assess a filter of interest. The website and Checklist are also resources that librarians can recommend to relevant enquirers.

The Checklist is not exclusive. It does not 'reject' search filters which have been designed informally or which have not been tested or validated. It does, however, allow librarians and

others to differentiate easily between evidence-based, validated filters and those of a less rigorous design.

Critical appraisal checklists serve several purposes. The clear breakdown of the reported methods in the Search Filter Appraisal Checklist is designed in the hope, shared with designers of other critical appraisal tools, that it will encourage filter authors, many of whom are librarians, to report detailed methods [17]. In highlighting which methods to report to help readers assess the quality and relevance of a filter, librarians can also assist authors to achieve more transparent research reporting.

The Checklist focuses on search filters designed to retrieve studies with specific research methods (such as systematic reviews) or study type focus (such as diagnostic tests). Some of the Checklist's elements are likely to be applicable to search filters in other areas. Health sciences librarians may wish to explore the applicability of the Checklist beyond methods search filters.

There is scope to evaluate the performance of the Checklist, using independent assessors and a range of filters. Evaluations could assess ease of use, clarity, comprehensiveness and consistency. Since the Checklist was finalised the CADTH critical appraisal and ranking tool for search filters has been developed. It is less detailed than the ISSG Checklist and incorporates a score [18]. A comparison of the two tools is a topic for further research. Meanwhile, the ISSG Checklist is being used to assess filters on the ISSG website and is offered for evaluation. The ISSG would be grateful for feedback from health sciences librarians on its use.

Conflict of interest statement

Acknowledgements

References

1.   Jenkins M. Evaluation of methodological search filters - a review. Health Info Libr J. 2004 Sep;21(3):148-63.

2.   Glanville JM, Lefebvre C, Miles JNV, Camosso-Stefinovic J. How to identify randomized controlled trials in MEDLINE: ten years on. J Med Libr Assoc. 2006 Apr;94:130-6.

3.   National Library of Medicine. PubMed Clinical Queries [internet]. Bethesda, MD: National Library of Medicine;2007. [cited 20 May 2008]. <http://www.ncbi.nlm.nih.gov/entrez/query/static/clinical.shtml>.

4.   White VJ, Glanville J, Lefebvre C, Sheldon TA. A statistical approach to designing search filters to find systematic reviews: objectivity enhances accuracy. J Inf Sci. 2001 Jun;27:357-70.

5.   Lefebvre C, Manheimer E, Glanville J. Chapter 6: Searching for studies [internet]. In: Higgins JP, Green S, eds. Cochrane handbook for systematic reviews of interventions. Version 5.0.0 (updated February 2008). The Cochrane Collaboration, 2008. [rev. Feb 2008; cited 20 May 2008]. <www.cochrane-handbook.org>

6.   InterTASC Information Specialists' Sub-Group. Search filter resource: diagnostic studies [internet]. York: InterTASC Information Specialists' Sub-Group, 2008. [rev. 13 May 2008; cited 20 May 2008]. <http://www.york.ac.uk/inst/crd/intertasc/diag.htm>.

7.   Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of reporting of meta-analyses. Lancet 1999 Nov;354:1896–900.

8.   The CONSORT Group. The CONSORT statement [internet]. The CONSORT Group, 2001. [rev. 22 Oct 2007; cited 20 May 2008]. <http://www.consort-statement.org/index.aspx?o=1011>.

9.   Whiting P, Rutjes AWS, Reitsma JB, Bossuyt PMM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol. 2003 Nov;3(25). doi: 10.1186/1471-2288-3-25.

10.   InterTASC Information Specialists' Sub-Group. Search filter resource: home page [internet]. York: InterTASC Information Specialists' Sub-Group, 2008. [rev. 13 May 2008; cited 20 May 2008]. <http://www.york.ac.uk/inst/crd/intertasc/index.htm>.

11.   Zhang L, Ajiferuke I, Sampson M. Optimizing search strategies to identify randomized controlled trials in MEDLINE. BMC Med Res Methodol. 2006;6. doi: 10.1186/1471-2288-6-23.

12.   Whiting P, Harbord R, Kleijnen J. No role for quality scores in systematic reviews of diagnostic accuracy studies. BMC Med Res Methodol. 2005;5. doi: 10.1186/1471-2288-5-19.

13.   Bachmann LM, Coray R, Estermann P, Ter Riet G. Identifying diagnostic studies in MEDLINE: reducing the number needed to read. J Am Med Inform Assoc. 2002 Nov–Dec;9(6):653-8.

14.   SIGN. Search filters [internet]. Edinburgh: SIGN, 2006. [rev. 14 Mar 2008 cited 20 May 2008]. <http://www.sign.ac.uk/methodology/filters.html#econ>.

15.   Wong SSL, Wilczynski NL, Haynes RB. Developing optimal search strategies for detecting clinically relevant qualitative studies in Medline. Medinfo. 2004;11(1):311-4.

16.   InterTASC Information Specialists' Sub-Group. Search filter resource: qualitative research [internet]. York: InterTASC Information Specialists' Sub-Group, 2008. [rev. 11 Mar 2008; cited 20 May 2008]. <http://www.york.ac.uk/inst/crd/intertasc/qualitat.htm>.

17.   The CONSORT Group. Welcome to the CONSORT statement website [internet]. The CONSORT Group, 2007. [cited 20 May 2008]. <http://www.consort-statement.org/>.

18.   Bak G. CADTH CAI and ranking tool for search filters [workshop paper]. Health Technology Assessment International. Barcelona, 2007.

Table 1. ISSG Search Filter Appraisal Checklist

Information and methodological issues	Categorisation options	Detailed information, as appropriate
A. Information		
A.1 State the author's objective.		
A.2 State the focus of the research.	[ ] Sensitivity-maximising [ ] Precision-maximising [ ] Specificity-maximising [ ] Balance of sensitivity and specificity / precision [ ] Other	
A.3 Database(s) and search interface(s).		
A.4 Describe the methodological focus of the filter (e.g. RCTs).		
A.5 Describe any other topic that forms an additional focus of the filter (e.g. clinical topics such as breast cancer, geographic location such as Asia or population grouping such as paediatrics).		
A.6 Other observations.		
B. Identification of a gold standard (GS) of known relevant records		
B.1 Did the authors identify one or more gold standards (GSs)?	none/1/2/3/4/5/more than 5	
B.2 How did the authors identify the records in each GS?		
B.3 Report the dates of the records in each GS.		

B.4 What are the inclusion criteria for each GS?		
B.5 Describe the size of each GS and the authors' justification, if provided (for example the size of the gold standard may have been determined by a power calculation)		
B.6 Are there limitations to the gold standard(s)?	Yes/No/Unclear	
B.7 How was each gold standard used?	[ ] to identify potential search terms [ ] to derive potential strategies (groups of terms) [ ] to test internal validity [ ] to test external validity [ ] other, please specify	
B.8 Other observations.		
C. How did the researchers identify the search terms in their filter(s) (select all that apply)?		
C.1 Adapted a published search strategy.	Yes/No/Unclear (please describe)	
C.2 Asked experts for suggestions of relevant terms.	Yes/No/Unclear (please describe)	
C.3 Used a database thesaurus.	Yes/No/Unclear (please describe)	
C.4 Statistical analysis of terms in a gold standard set of records (see B above).	Yes/No/Unclear (please describe)	

C.5 Extracted terms from the gold standard set of records (see B above).	Yes/No/Unclear (please describe)	
C.6 Extracted terms from some relevant records (but not a gold standard).	Yes/No/Unclear (please describe)	
C.7 Tick all types of search terms tested.	[ ] subject headings [ ] text words (e.g. in title, abstract) [ ] publication types [ ] subheadings [ ] check tags [ ] other, please specify	
C.8 Include the citation of any adapted strategies.		
C.9 How were the (final) combination(s) of search terms selected?		
C.10 Were the search terms combined (using Boolean logic) in a way that is likely to retrieve the studies of interest?		
C.11 Other observations.		
D. Internal validity testing (This type of testing is possible when the search filter terms were developed from a known gold standard set of records).		
D.1 How many filters were tested for internal validity?		
For each filter report the following information		

D.2 Was the performance of the search filter tested on the gold standard from which it was derived?	Yes/No/Unclear (please describe)	
D.3 Report sensitivity data (a single value, a range, 'Unclear'* or 'not reported', as appropriate). *Please describe.		
D.4 Report precision data (a single value, a range, 'Unclear'* or 'not reported' as appropriate). *Please describe.		
D.5 Report specificity data (a single value, a range, 'Unclear'* or 'not reported' as appropriate). *Please describe.		
D.6 Other performance measures reported.		
D.7 Other observations.		
E. External validity testing (This section relates to testing the search filter on records that are different from the records used to identify the search terms).		
E.1 How many filters were tested for external validity on records different from those used to identify the search terms?		
E.2 Describe the validation set(s) of records, including the interface.		
For each filter report the following information.		
E.3 On which validation set(s) was the filter tested?		

E.4 Report sensitivity data for each validation set (a single value, a range or 'Unclear' or 'not reported', as appropriate).		
E.5 Report precision data for each validation set (report a single value, a range or 'Unclear' or 'not reported', as appropriate).		
E.6 Report specificity data for each validation set (a single value, a range or 'Unclear' or 'not reported', as appropriate).		
E.6 Other performance measures reported.		
E.7 Other observations.		
F. Limitations and comparisons.		
F.1 Did the authors discuss any limitations to their research?		
F.2 Are there other potential limitations to this research that you have noticed?		
F.3 Report any comparisons of the performance of the filter against other relevant published filters (sensitivity, precision, specificity or other measures).		
F.4 Include the citations of any compared filters.		
F.5 Other observations and / or comments.		

G. Other comments. This section can be used to provide any other comments. Selected prompts for issues to bear in mind are given below.		
G.1 Have you noticed any errors in the document that might impact on the usability of the filter?		
G.2 Are there any published errata or comments (for example in the MEDLINE record)?		
G.3 Is there public access to pre-publication history and / or correspondence?		
G.4 Are further data available on a linked site or from the authors?		
G.5 Include references to related papers and/or other relevant material.		
G.6 Other comments.		

Figure 1. Two brief summary formats tabled at the first consensus meeting

Example 1

A gold standard set of relevant records was selected by hand searching six key journals for two years. Discriminating search terms were identified from gold standard records by consulting a group of experts. The performance of the strategies, in terms of sensitivity and precision, was tested on a subset of the original gold standard. The authors do not describe any other testing.

Example 2

The authors employed a pragmatic filter design approach. A set of candidate records was identified by searching MEDLINE and the authors derived terms from those records and tested out the precision of the strategy in MEDLINE.

Figure 2. Example brief abstract tabled and rejected at the second consensus meeting

Wong S S L, Wilczynski N L, Haynes R B. Optimal search strategies for qualitative studies in MEDLINE. Medinfo 2004;11(1):311-316.

Presents filters for clinically relevant qualitative research in MEDLINE. Search terms were collected by consulting widely with experts, and from relevant records. A gold standard (GS) of 366 records was derived by hand searching 161 core health care journals in 2000. 60% of the GS was used to derive the search terms and filters. The remaining 40% of the GS was used to validate the filters. Offers highly sensitive (92.47%), highly specific (99.36%) and best sensitivity and specificity compromise filters. The scale of the hand search is impressive, but the focus on one year may affect the future currency of the filter.

Figure 3. Example structured abstract tabled and accepted at the second consensus meeting.

Wong S S L, Wilczynski N L, Haynes R B. Optimal search strategies for qualitative studies in MEDLINE. Medinfo 2004;11(1):311-316.

Objective:

This filter is designed to identify qualitative research in Medline.

Methods:

The authors identified a gold standard (GS) of 366 records by hand searching 161 core health care journals in 2000. 60% of the GS was used to derive the filters and the remaining 40% was used to validate the filters. The search terms for the filters were collected from consulting widely with experts, and from relevant records.

Results:

Several filters are offered. In the validation set the most sensitive filter scored 92.47%, the most precise scored 39.59%, the most specific scored 99.36% and the best compromise between sensitivity and specificity scored 86.99% sensitivity and 92% specificity (7.53% precision). The authors reported that the differences in performance between the development and validation sets were not statistically significant.

Discussion:

The authors note that, unlike their other filters, the records were not assessed for methodological quality, and that further testing of the filters is required to provide more performance data.

ISSG commentary:

The scale of the hand search is impressive, but the focus on one year may affect the future currency of the filter, especially if the authors' desired changes to reporting and indexing of

qualitative research are realised. More detail of the number of qualitative terms tested and

test methods would have been helpful.