# Assessing the performance of methodological search filters to improve the efficiency of evidence information retrieval: five literature reviews and a qualitative study

Carol Lefebvre,[1,2]* Julie Glanville,[3] Sophie Beale,[3] Charles Boachie,[4] Steven Duffy,[3] Cynthia Fraser,[4] Jenny Harbour,[5] Rachael McCool[3] and Lynne Smith[5]

[1]UK Cochrane Centre, Oxford, UK
[2]Lefebvre Associates Ltd, Oxford, UK
[3]York Health Economics Consortium, York, UK
[4]Health Services Research Unit, University of Aberdeen, Aberdeen, UK
[5]Healthcare Improvement Scotland, Glasgow, UK

*Corresponding author Carol@LefebvreAssociates.org

**Declared competing interests of authors:** none

**Note to reader:** it is acknowledged that there has been a regrettable delay between carrying out the project, including the searches, and the publication of this report, because of serious illness of the principal investigator. The searches were carried out in 2010/11.

Published November 2017
DOI: 10.3310/hta21690

## Scientific summary

### Performance of methodological search filters

Health Technology Assessment 2017; Vol. 21: No. 69
DOI: 10.3310/hta21690

NIHR Journals Library www.journalslibrary.nihr.ac.uk

# Scientific summary

## Background

The effective retrieval of relevant evidence is essential in the development of clinical guidance or health policy, the conduct of health research and the support of health-care decision-making. Whether the purpose of the evidence retrieval is to find a representative set of results to inform the development of an economic model or to find extensive evidence on the clinical effectiveness or cost-effectiveness of a health-care intervention, retrieval methods need to be appropriate, efficient within time and cost restraints, consistent and reliable.

One tool that can be useful for effective retrieval is the search filter. Search filters are a combination of search terms designed to retrieve records about a specific concept, which may be a study design, such as randomised controlled trials (RCTs), outcomes such as adverse events, a population such as women or a disease or condition such as cardiovascular disease. A methodological search filter is designed to capture the records of studies that have used a specific study design. Effective search filters may seek to maximise sensitivity (the proportion of relevant records retrieved), maximise precision (the proportion of retrieved records that are relevant) or optimise retrieval using a balance between maximising sensitivity and achieving adequate precision. Search filters can offer a standard approach to study retrieval and release searcher time to focus on developing other sections of the search strategy such as the disease concept.

## Objectives

This project was funded to inform National Institute for Health and Care Excellence (NICE) methods development, but has wider application to efficient literature searching in support of evidence-based medicine in general. Its aim was to investigate the methods used to assess the performance of methodological search filters and explore what searchers require of search filters and what information searchers require to help them choose a search filter. We also explored systems and approaches for providing better access to relevant and useful performance data on methodological search filters, including developing suggested approaches to search filter performance measurement.

Our objectives were to identify and summarise:

- which performance measures for search filters are reported
- other performance measures reported in diagnostic test accuracy (DTA) studies and reviews
- different ways to present filter/test performance data to assist users in choosing which filters or tests to use
- evidence on how searchers choose search filters and what information they would like to receive to inform their choices
- evidence on how clinicians choose diagnostic tests.

The project website is at https://sites.google.com/a/york.ac.uk/search-filter-performance/ (accessed 22 August 2017).

## Methods

We conducted a series of five literature reviews in 2010/11 into various aspects of search filter reporting and use and analogous activity in the field of DTA studies. The reviews informed the development of an

interview schedule, to learn how search filters are used by information professionals working for NICE and organisations affiliated to NICE, and also the development of a web-based questionnaire aimed at a wider audience of search experts in the area of search filters.

The literature reviews explored:

- what performance measures are reported for single studies of search filters and how are they presented (review A)
- what performance measures are reported when comparing a range of search filters and how the performance measures are synthesised (review B)
- what performance measures are reported in DTA studies and DTA reviews (review C)
- how searchers choose search filters (review D)
- how filter/test performance data are presented to assist users in choosing which filters or tests to use (reviews A, B and C)
- how clinicians or organisations choose diagnostic tests (review E).

Information professionals working for NICE, the NICE Collaborating Centres and NICE Evidence Review Groups were interviewed using a semistructured interview protocol.

A web-based questionnaire survey was developed to obtain information on searchers' knowledge of and use of search filters. The questions were based on findings from the reviews and the interviews. The questionnaire was advertised to seven e-mail discussion lists aimed at health librarians.

The reviews, interviews and questionnaire informed the development of suggested approaches to gathering and reporting search filter performance.

We acknowledge that there has been a regrettable delay between carrying out the project, including the searches, and the publication of this report, because of serious illness of the principal investigator.

## Results

### *Review A*

In total, 23 studies were identified in review A. In single studies reporting search filters:

- internal gold or reference standards were mostly derived by hand-searching journals
- filter validation was mostly carried out using internal validation
- sensitivity, precision and specificity were the most commonly used performance measures
- performance measures were most often presented in tables.

### *Review B*

In total, 18 studies were identified in review B. In filter comparison studies:

- sensitivity, precision and specificity were the most commonly reported performance measures
- the highest sensitivity, highest precision and optimal/balanced filter strategies were most frequently reported
- methods reporting was limited in papers reporting the development of new search filters and comparison with existing filters
- the most frequently used method for reporting the results of filter performance comparisons was in tables, although graphs might be more useful.

© Queen's Printer and Controller of HMSO 2017. This work was produced by Lefebvre *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

### Review C

In total, 47 studies were identified in review C. DTA studies and DTA reviews provided evidence that:

- studies should be carried out on a sample of patients who are representative of the target population and should use an appropriate reference standard
- sensitivity and specificity were the most commonly reported outcomes and are subject to spectrum bias
- predictive values are influenced by disease prevalence
- receiver operating characteristic curves present sensitivity and specificity pairs at different test thresholds
- the area under the curve gives an overall value of DTA
- health technology assessment organisations recommend that DTA studies should present 2 × 2 contingency tables, sensitivity and specificity pairs and likelihood ratio pairs
- several types of graphical presentation can be used to display DTA data but these had not been used extensively in the DTA literature
- poor-quality methods and reporting hinder the inferences that can be drawn from DTA studies.

### Review D

No studies were identified that reported how searchers chose search filters.

### Review E

Seven studies were identified that reported on factors that influenced clinicians' choice between diagnostic tests. They provided limited evidence suggesting that test performance is the main factor that informed choices. As a substantial proportion of clinicians have an inaccurate understanding of test performance parameters and how they should be applied, it might be the case that choices were being based on false assumptions.

### Interviews

A total of 12 interviews were conducted, capturing the views of 16 information professionals.

The interviews revealed the wide range of searching tasks that are undertaken in the NICE context and the various points at which search filters can be used. The use of search filters seemed to be linked predominantly to reducing the numbers of retrieved records, introducing focus and assisting with searches that are focused on a single study type.

The Cochrane RCT and McMaster Hedges team filters were cited most often. Various methods were used to identify filters, with the most frequently mentioned resource being the Information Specialists' Sub-Group (ISSG) Search Filters Resource [Glanville J, Lefebvre C, Wright K. *ISSG Search Filter Resource*. York: The InterTASC Information Specialists' Sub-Group; 2008 (updated 2017). URL: https://sites.google.com/a/york.ac.uk/issg-search-filters-resource/home (accessed 22 August 2017).].

Interviewees' practices when using, adapting and reporting search filters were not uniform, possibly indicating an absence of accepted published formal guidance. Interviewees found it difficult to keep informed about search filter developments. When choosing filters, interviewees tried to make judgements around the relative sensitivity, specificity and precision of search filters but were conscious of factors such as time constraints and knowledge gaps that impeded this. Some interviewees requested more guidance on the best filters to use or chose filters based on the authorship of the filter. Some desire for standardisation or guidance within the NICE family was also expressed.

### Questionnaire

In total, 90 individuals responded to the survey. About three-quarters of respondents said that they used search filters for extensive searches to inform guidelines or systematic reviews, with just over half saying that they would use them for rapid searches to answer brief questions and a similar number saying that they would use them for scoping searches to estimate the size of the literature on a topic.

The McMaster Hedges team was the most frequently reported source used to identify study design filters. Currently, respondents most frequently used search filters for RCTs and systematic reviews. The most frequently cited filters for a specific topic were the Cochrane RCT filters.

Just over half of the respondents reported that they generally use the in-built filters in database interfaces rather than typing in another filter. Once they had found a search filter, just over half of respondents reported that they sometimes amend the filter. Nearly all of those respondents who amended search filters tested the effect of the amendment by either comparing the results with and without the filter amendment or determining whether or not known relevant papers had been identified. Three-quarters of respondents documented their amendments when they wrote up the searches, using diverse approaches.

Information on search filter performance measures such as validation, sensitivity and precision, a description of the filter and the results of their own testing had helped respondents to choose between filters.

The main factors that would make choosing a filter easier were the availability of a critical appraisal or evaluation and more information on the effectiveness of the filter, what it does or what it provides, what it excludes, its limitations, when it was last updated, its advantages and disadvantages, its sensitivity and precision and what testing has been completed. Respondents wanted to be confident in the author/developer and the availability of the filter in a central location was important.

## Conclusions

Studies of search filter development and comparison studies reached similar conclusions. Internal gold or reference standards were mostly derived by hand-searching journals. Internal rather than more rigorous external validation was more usually undertaken. The most commonly reported performance measures were sensitivity/recall, precision and specificity.

Filter performance comparison studies most commonly reported the highest sensitivity, highest precision and optimal/balanced filter strategies. These measures were generally presented in tables, with little use of other graphical options that might be more useful methods of presentation. Limited details about methods were reported and guidance in this area could be improved.

Guidance available on conducting and analysing the results of DTA studies is applicable to several aspects of search filter research. The identification of a representative sample of records, of sufficient size and using a standardised approach, will assist in producing robust and generalisable results. The greater use of graphical presentation might facilitate the dissemination and interpretation of results.

We did not identify any published research on how searchers choose search filters and were unable to draw conclusions. Furthermore, limited evidence was identified in the review of clinicians' decision-making, resulting in few insights into how clinicians or organisations choose diagnostic tests, which might have been transferable to the challenges of choosing search filters. Diagnostic test performance was the most frequent factor mentioned and is the main factor that is readily applicable to search filter choice. The other message that we identified is that providing additional explanatory information when reporting search filter performance might be necessary to ensure that searchers make choices based on an accurate understanding of test performance parameters.

The interviews and the questionnaire survey indicated that search filters are not appropriate for all searching tasks but are used mainly for reducing large results sets and assisting with searches that are focused on a single study type. Searchers use several key resources to identify search filters but may find choosing between filters challenging. Choosing filters might be aided by making information about filters less technical, offering ratings and providing more details about filter validation strategies and filter provenance.

© Queen's Printer and Controller of HMSO 2017. This work was produced by Lefebvre *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

The responses to the questionnaire provide many messages for search filter designers. Filter performance measures need to be signposted more clearly and succinctly to help searchers make better use of the available filters. Filter and website designers should present less information and ensure that performance information can be clearly identified. The provenance of filters is clearly important to some searchers but there are no established parameters to measure this confidence. Clear authorship labelling and the provision of detailed information to show the robustness of the development methods would not only assist users of filters but also help filter designers to achieve recognition for their filters. The convenience of having filters from well-established producers available within database interfaces encourages their use. A convenient filter may, however, not always be the best one for the task. Searchers need to know how to choose between a range of filters and need information on whether filters have been validated and how.

## Recommendations for information retrieval practice

We recommend that:

- studies reporting search filter design and/or comparisons of search filter performance should explicitly report the methods and results to help searchers identify the most appropriate filter
- one or more gold or reference standards should be used for testing filter performance
- relative recall (RR) and hand-searching should be considered for the development of gold or reference standard(s) for filter development but caution should be exercised regarding the robustness of the original RR search
- search filters should be validated on gold or reference standards that are different from those from which they were developed (i.e. external validation)
- the size of the gold or reference standard(s) should be clearly stated and a sample size calculation presented to justify the size of the standard(s)
- when a filter has been translated for use in a different database and/or interface from that in which it was developed, this should be specifically reported
- results should be presented systematically, identifying clearly the best-performing filter for specific purposes (sensitive strategy, specific strategy, balanced strategy)
- tables of performance results should have a consistent format and order to enable information to be easily extracted
- additional reporting methods should be considered, including graphical options
- approaches such as those provided in this report should be considered regarding the use, adaptation and reporting of search filters.

## Recommendations for research

Further research might include:

- the development and testing of filters for a wider range of study designs and other topics
- the development and testing of translations of filters for different databases and interfaces
- the development and testing of filters that are independent of indexing language
- a review of the performance measures reported and the methods of presentation used in methodological filter performance comparisons for study designs not included in this review
- studies to explore alternative methods of displaying performance results from comparisons of multiple methodological search filters
- explorations of methods for the numerical synthesis of the results of several filter performance comparisons.

## Funding

The National Institute for Health Research (NIHR) Health Technology Assessment programme and Medical Research Council–NIHR Methodology Research Programme (grant number G0901496).

© Queen's Printer and Controller of HMSO 2017. This work was produced by Lefebvre *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

# Health Technology Assessment

HTA MRP–MRC

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Impact factor: 4.236

*Health Technology Assessment* is indexed in MEDLINE, CINAHL, EMBASE, The Cochrane Library and the Clarivate Analytics Science Citation Index.

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: journals.library@nihr.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nihr.ac.uk

### Criteria for inclusion in the *Health Technology Assessment* journal

Reports are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme or, commissioned/managed through the Methodology research programme (MRP), and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

### HTA programme

The HTA programme, part of the National Institute for Health Research (NIHR), was set up in 1993. It produces high-quality research information on the effectiveness, costs and broader impact of health technologies for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined as all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

For more information about the HTA programme please visit the website: http://www.nets.nihr.ac.uk/programmes/hta

### This report

This issue of the Health Technology Assessment journal series contains a project commissioned/managed by the Methodology research programme (MRP). The Medical Research Council (MRC) is working with NIHR to deliver the single joint health strategy and the MRP was launched in 2008 as part of the delivery model. MRC is lead funding partner for MRP and part of this programme is the joint MRC–NIHR funding panel 'The Methodology Research Programme Panel'.

To strengthen the evidence base for health research, the MRP oversees and implements the evolving strategy for high-quality methodological research. In addition to the MRC and NIHR funding partners, the MRP takes into account the needs of other stakeholders including the devolved administrations, industry R&D, and regulatory/advisory agencies and other public bodies. The MRP funds investigator-led and needs-led research proposals from across the UK. In addition to the standard MRC and RCUK terms and conditions, projects commissioned/managed by the MRP are expected to provide a detailed report on the research findings and may publish the findings in the HTA journal, if supported by NIHR funds.

The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

This report presents independent research funded under a MRC–NIHR partnership. The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, the MRC, NETSCC, the HTA programme or the Department of Health. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, the MRC, NETSCC, the HTA programme or the Department of Health.

**© Queen's Printer and Controller of HMSO 2017. This work was produced by Lefebvre *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.**

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).

## *Health Technology Assessment* Editor-in-Chief

**Professor Hywel Williams** Director, HTA Programme, UK and Foundation Professor and Co-Director of the Centre of Evidence-Based Dermatology, University of Nottingham, UK

## NIHR Journals Library Editor-in-Chief

**Professor Tom Walley** Director, NIHR Evaluation, Trials and Studies and Director of the EME Programme, UK

## NIHR Journals Library Editors

**Professor Ken Stein** Chair of HTA and EME Editorial Board and Professor of Public Health, University of Exeter Medical School, UK

**Professor Andrée Le May** Chair of NIHR Journals Library Editorial Group (HS&DR, PGfAR, PHR journals)

**Dr Martin Ashton-Key** Consultant in Public Health Medicine/Consultant Advisor, NETSCC, UK

**Professor Matthias Beck** Chair in Public Sector Management and Subject Leader (Management Group), Queen's University Management School, Queen's University Belfast, UK

**Dr Tessa Crilly** Director, Crystal Blue Consulting Ltd, UK

**Dr Eugenia Cronin** Senior Scientific Advisor, Wessex Institute, UK

**Dr Peter Davidson** Director of the NIHR Dissemination Centre, University of Southampton, UK

**Ms Tara Lamont** Scientific Advisor, NETSCC, UK

**Dr Catriona McDaid** Senior Research Fellow, York Trials Unit, Department of Health Sciences, University of York, UK

**Professor William McGuire** Professor of Child Health, Hull York Medical School, University of York, UK

**Professor Geoffrey Meads** Professor of Wellbeing Research, University of Winchester, UK

**Professor John Norrie** Chair in Medical Statistics, University of Edinburgh, UK

**Professor John Powell** Consultant Clinical Adviser, National Institute for Health and Care Excellence (NICE), UK

**Professor James Raftery** Professor of Health Technology Assessment, Wessex Institute, Faculty of Medicine, University of Southampton, UK

**Dr Rob Riemsma** Reviews Manager, Kleijnen Systematic Reviews Ltd, UK

**Professor Helen Roberts** Professor of Child Health Research, UCL Institute of Child Health, UK

**Professor Jonathan Ross** Professor of Sexual Health and HIV, University Hospital Birmingham, UK

**Professor Helen Snooks** Professor of Health Services Research, Institute of Life Science, College of Medicine, Swansea University, UK

**Professor Jim Thornton** Professor of Obstetrics and Gynaecology, Faculty of Medicine and Health Sciences, University of Nottingham, UK

**Professor Martin Underwood** Director, Warwick Clinical Trials Unit, Warwick Medical School, University of Warwick, UK

Please visit the website for a list of members of the NIHR Journals Library Board:
www.journalslibrary.nihr.ac.uk/about/editors

**Editorial contact:** journals.library@nihr.ac.uk