



*Canadian Agency for
Drugs and Technologies
in Health*

*Agence canadienne
des médicaments et des
technologies de la santé*

HTA

Development and Testing of Search
Filters to Identify Economic Evaluations
in MEDLINE and EMBASE

OCTOBER 2009

Supporting Informed Decisions

Until April 2006, the Canadian Agency for Drugs and Technologies in Health (CADTH) was known as the Canadian Coordinating Office for Health Technology Assessment (CCOHTA).

Publications can be requested from:

CADTH
600-865 Carling Avenue
Ottawa ON Canada K1S 5S8
Tel.: 613-226-2553
Fax: 613-226-5392
Email: pubs@cadth.ca

or downloaded from CADTH's website:
<http://www.cadth.ca>

Cite as: Glanville J, Fleetwood K, Yellowlees A, Kaunelis D, Mensinkai S. *Development and Testing of Search Filters to Identify Economic Evaluations in MEDLINE and EMBASE*. Ottawa: Canadian Agency for Drugs and Technologies in Health; 2009.

Production of this report is made possible by financial contributions from Health Canada and the governments of Alberta, British Columbia, Manitoba, New Brunswick, Newfoundland and Labrador, Northwest Territories, Nova Scotia, Nunavut, Prince Edward Island, Saskatchewan, and Yukon. The Canadian Agency for Drugs and Technologies in Health takes sole responsibility for the final form and content of this report. The views expressed herein do not necessarily represent the views of Health Canada or any provincial or territorial government.

Reproduction of this document for non-commercial purposes is permitted provided appropriate credit is given to CADTH.

CADTH is funded by Canadian federal, provincial, and territorial governments.

Legal Deposit — 2009
National Library of Canada
ISBN: 978-1-926680-10-1 (print)
ISBN: 978-1-926680-11-8 (online)
H0490 — October 2009

PUBLICATIONS MAIL AGREEMENT NO. 40026386
RETURN UNDELIVERABLE CANADIAN ADDRESSES TO
CANADIAN AGENCY FOR DRUGS AND TECHNOLOGIES IN HEALTH
600-865 CARLING AVENUE
OTTAWA ON K1S 5S8

Canadian Agency for Drugs and Technologies in Health

**Development and Testing of Search Filters to Identify
Economic Evaluations in MEDLINE and EMBASE**

Julie Glanville¹
Kelly Fleetwood²
Ann Yellowlees²
David Kaunelis³
Shaila Mensinkai³

October 2009

¹ York Health Economics Consortium, York, UK.

² Quantics Consulting, Tweed Horizons, UK

³ Canadian Agency for Drugs and Technologies in Health, Ottawa, Ontario

Disclaimer

This report is prepared by the Health Technology Assessment (HTA), a service of the Canadian Agency for Drugs and Technologies in Health.

The Research and Information Services unit is responsible for the provision of research information services to CADTH. Research information services staff retrieve and manage the scientific information required by the three CADTH science programs.

The report describes search filters development and testing to identify economic evaluations in the MEDLINE and EMBASE databases. It also reports on the relative performance of existing economic evaluation search filters.

This report has been peer reviewed and guided by input from experts in Information Science, Health Technology Assessment and Economic Evaluations. The information in this report is intended to assist Information Specialists and Researchers in the conduct of health technology assessments.

While CADTH has taken care in the preparation of the document to ensure that its contents are accurate, complete and up-to-date as at the date of publication, CADTH does not make any guarantee to that effect. CADTH does not guarantee and is not responsible for the quality, currency, propriety, accuracy or reasonableness of any statements, information or conclusions contained in the source documentation. CADTH is not responsible for any errors or omissions or injury, loss or damage arising from or relating to the use (or misuse) of any information, statements or conclusions contained in or implied by the information in this document or in any of the source documentation.

This document may contain links to other information available on the websites of third parties on the Internet. CADTH does not have control over the content of such sites. Use of third party sites is governed by the terms and conditions set out for such sites. CADTH does not make any guarantee with respect to any information contained on such third party sites and CADTH is not responsible for any injury, loss or damage suffered as a result of using such third party sites.

This document and the information and recommendations provided in this document are prepared and intended for use in the context of the Canadian health care system. Other health care systems are different; the issues, information and recommendations related to the subject matter of this document may be different in other jurisdictions and, if used outside of Canada, it is at the user's risk. This disclaimer and any questions or matters of any nature arising from or relating to the content or use (or misuse) of this document will be governed by and interpreted in accordance with the laws of the Province of Ontario and the laws of Canada applicable therein, and all proceedings shall be subject to the exclusive jurisdiction of the courts of the Province of Ontario, Canada

Copyright © CADTH October 2009. You are permitted to make copies of this document for non-commercial purposes provided it is not modified when reproduced and appropriate credit is given to CADTH. You may not otherwise copy, modify, translate, post on a web site, store electronically, republish or redistribute any material from the web site in any form or by any means without the prior written permission of CADTH.

CADTH takes sole responsibility for the final form and content of this document, subject to the limitations noted above. The statements and conclusions in this document are those of CADTH and not of its advisory committees and reviewers. The statements, conclusions and views expressed herein do not necessarily represent the views of Health Canada or any Canadian provincial or territorial government or pharmaceutical manufacturers.

Production of this document is made possible by financial contributions from Health Canada and the governments of Alberta, British Columbia, Manitoba, New Brunswick, Newfoundland and Labrador, Northwest Territories, Nova Scotia, Nunavut, Ontario, Prince Edward Island, Saskatchewan and Yukon.

Please contact CADTH's Vice President, Corporate Services at corporateservices@cadth.ca with any inquiries about this notice or other legal matters relating to CADTH's services.

Reviewers

These individuals kindly provided comments on this report.

External Reviewers

Ellen T. Crumley, MLIS AHIP PhD Student
University of Alberta
Edmonton, Alberta

Margaret Sampson, MLIS
Manager, Library Services
Children's Hospital of Eastern Ontario
Ottawa, Ontario

Anne McKibbin, MLS PhD FMLA
Assistant Professor
McMaster University
Hamilton, Ontario

Authorship

Julie Glanville was the research lead. She provided the conception and project design, planning and management of data acquisition, and the reformatting, analysis and testing of data and filters, including report writing and revisions.

Kelly Fleetwood completed the statistical analysis and critical reviews of the report.

Ann Yellowlees provided sample size calculations, concept for statistical methodology and supervision of analysis, and reviewed drafts of the report.

David Kaunelis participated in the design of the project, which included defining the objective of the filters, defining what is meant by economic evaluations, and determining desired levels of sensitivity and precision for the filters. He tested available and candidate filters in MEDLINE and EMBASE, managed the references within the report, and reviewed and contributed to report revisions.

Shaila Mensinkai participated in the protocol development and refining of the project scope, reviewed drafts, contributed to the responses to external reviewers' comments, and was the team liaison with the York Health Economics Consortium at the University of York.

All authors approved the final report.

Acknowledgements

The team would like to thank Alan Blighe for downloading, selecting, and editing of records from EMBASE and MEDLINE that formed the gold standard records; formatting the report; and general administration.

Conflicts of Interest

No conflicts of interest have been declared.

EXECUTIVE SUMMARY

Background

Health care decision-makers and health care professionals need efficient access to the best evidence. Health technology assessment (HTA) agencies and guideline developers need access to evidence from economic evaluations to identify the cost-effectiveness of a technology and to inform models.

Access to economic evaluations has improved with the development of the National Health Service Economic Evaluation Database (NHS EED) and the Health Economic Evaluations Database. Despite the availability of these resources, however, there are still issues with currency because of lags between the time when studies are identified and the time when full abstracts are published in the databases. This means that large biomedical databases such as EMBASE and MEDLINE still need to be searched for economic evaluations. Data are lacking on how well published search filters perform in finding economic evaluations in the databases.

Our objective was to develop search filters to identify economic evaluations in the MEDLINE and EMBASE databases that would maximize sensitivity and achieve levels of precision to meet the needs of health technology assessment researchers. This project was also designed to obtain data on the relative performance of new and published search filters.

Method

A gold standard set of economic evaluations was identified from the NHS EED. Records were selected from 2000, 2003, and 2006. These records were then identified in MEDLINE and EMBASE. A comparator set of non-economic evaluation records was obtained from MEDLINE and EMBASE using random selection from among those records that were retrieved using the NHS EED search filter. A 50% random sample of the gold standard records was analyzed to identify the occurrence of terms in the title, abstract, subject headings, and publication type fields. The terms that met a range of occurrence levels were identified. Data on the presence or absence of those terms in all gold standard and comparator records were analyzed using classification trees. Classification trees were developed for the records from the MEDLINE and EMBASE databases. For each database, 50% of the data were used to develop the tree (the training set), and 50% of the data were used to test the effectiveness of the tree (the validation set).

The candidate search filters that were identified from the search analysis, other published and unpublished filters, and combinations of some of these filters were tested in the Ovid interface to MEDLINE and EMBASE. Published and unpublished filters were identified from the InterTASC Information Specialists' Sub-Group search filter website, provided by the Canadian Agency for Drugs and Technologies in Health (CADTH), and provided by the reviewers of a first draft of this report. The filters that were tested were produced by CADTH, Emory University (the Grady EBM filter), McKinlay et al., NHS EED, National Health Service Quality Improvement Scotland (NHS QIS), Royle and Waugh, Sassi et al., the Scottish Intercollegiate Guidelines Network, and Wilczynski et al.

Results

A total of 2,070 full economic evaluations were identified from NHS EED for 2000, 2003, and 2006. Overall, 1,957 of these had corresponding records in MEDLINE and 1,876 had corresponding records in EMBASE. These two sets of records were the gold standards. The MEDLINE comparator set of non-economic evaluation records comprised 4,136 records and the EMBASE comparator set comprised 3,750 records.

After analysis of the occurrence of terms in the records, 347 terms (single words, subject headings, and publication types) were selected for analysis from the MEDLINE records, and 528 terms were selected for analysis from the EMBASE records. Eight candidate MEDLINE filters (MEDLINE A to MEDLINE H) and eight candidate EMBASE filters (EMBASE A to EMBASE H) emerged from the analysis. All candidate filters, and 13 published and unpublished filters were tested in the Ovid interface to the databases.

The MEDLINE filters with high sensitivity (more than 0.99) were the NHS QIS full and brief filters, NHS EED, and Royle and Waugh. The NHS EED filter had the highest precision (0.04) among these four filters. Achieving higher levels of precision thereby reduced sensitivity. Combining the NHS EED filter (using “AND”) with the MEDLINE D filter achieved more than 0.092 precision with more than 0.943 sensitivity. The Wilczynski best optimization of sensitivity and specificity filter achieved 0.093 precision with 0.923 sensitivity.

The MEDLINE G filter had the maximum precision (0.72 sensitivity and 0.257 precision). The Emory University Grady filter best met the objective of good precision with a sensitivity greater than 0.80 (0.845 sensitivity and 0.133 precision).

Four EMBASE filters (NHS QIS, CADTH, Royle and Waugh, and NHS EED) all had greater than 0.99 sensitivity (the precision ranged between 0.015 and 0.029). The highest precision, with more than 0.90 sensitivity (sensitivity 0.931 and precision 0.133) was achieved by the EMBASE G filter combined with the NHS EED filter.

The EMBASE H filter had the highest precision (0.716 sensitivity and 0.266 precision). The EMBASE G filter combined with the NHS EED filter best met the objective of good precision with a sensitivity of more than 0.80 (0.931 sensitivity and 0.133 precision).

Discussion

This research provides new performance data on published search filters used for the identification of economic evaluations in MEDLINE and EMBASE. Many publicly available filters have not been validated or compared.

Filters that are produced by NHS EED, NHS QIS, and Royle and Waugh continue to perform with high sensitivity in MEDLINE. None of the published or new filters could meet the objectives of high sensitivity (more than 0.95) with a precision of 0.20, or lower sensitivity (greater than 0.79) and enhanced precision (at least 0.50).

The search filters developed for this project were discriminating in the analysis. When tested in the Ovid interface, however, they did not perform as well. This indicates that the text words and

indexing terms in economic evaluation records do not discriminate evaluation records from other records that deal with economic issues in health care. The text words with high sensitivity for identifying economic evaluations (such as “cost” and “economics”) are used in many contexts other than economic evaluations. Therefore, maximizing precision remains a challenge.

This analysis shows that it is still difficult to efficiently identify economic evaluations using the indexing terms applied by the database producers. Even though there are indexing terms available that are specific to economic evaluations (such as “Cost-benefit Analysis/” in MEDLINE), they do not seem to be assigned to all relevant records. This information may help database producers make research evidence more accessible.

Further analysis of the data and gold standard records in this report may lead to the improved filters. Additional analysis could focus on the performance of lower frequency terms (below the selected cut-off frequency that was used in this project) and the analysis of phrases and terms in close proximity in the title and abstracts of records.

The filters have been tested in Ovid MEDLINE and EMBASE, but they require more extensive validation in other gold standard sets of records.

Conclusions

This research shows that searchers have several sensitive filters to use to identify economic evaluations in MEDLINE: NHS QIS (full and brief), NHS EED, and Royle and Waugh. Searchers may select filters based on highest precision (NHS EED) or conciseness (Royle and Waugh). Increased precision can be achieved by choosing filters with lower sensitivity. All the filters for EMBASE should be used without exclusion search lines that remove specific publication types and animal studies.

For searchers conducting scoping studies or rapid reviews, this research identified filters that offer higher levels of precision: the new MEDLINE G filter (more than 0.25 precision) and the new EMBASE H filter (more than 0.26 precision).

Searchers now have new comparative information on the performance of a range of filters, which can assist during the planning of the workload for technology assessments.

GLOSSARY

Cost-benefit analysis: A cost-benefit analysis values costs and outcomes in monetary terms. Values are usually obtained through using a willingness-to-pay approach, such as contingent valuation or conjoint analysis. (http://www.cadth.ca/media/pdf/186_EconomicGuidelines_e.pdf)

Cost-consequences analyses: Cost-consequences analyses present multiple outcome measures in a disaggregated format (e.g., intervention costs, hospital costs, clinical benefits, and adverse events). (http://www.cadth.ca/media/pdf/186_EconomicGuidelines_e.pdf)

Cost-effectiveness analysis: A cost-effectiveness analysis (CEA) is a specific type of economic evaluation where the health outcomes are measured in natural (health) units, such as life-years gained, lives saved, or clinical events avoided or achieved. (http://www.cadth.ca/media/pdf/186_EconomicGuidelines_e.pdf)

Cost-minimization analysis: In a cost-minimization analysis (CMA), alternatives are considered to be equivalent in terms of factors that are relevant to the decision (other than cost), and so, the lowest cost alternative is determined. A CMA can be regarded as an extension of a CEA or a CUA where the outcomes are demonstrated to be equivalent, and so only the costs of the alternatives are compared. (http://www.cadth.ca/media/pdf/186_EconomicGuidelines_e.pdf)

Cost-utility analysis: In a cost-utility analysis (CUA), outcomes are measured as health-related preferences, which are most often expressed as quality-adjusted life-years gained (i.e., a final outcome). A CUA uses a generic outcome measure that permits decision-makers to make broad comparisons across different conditions and interventions. (http://www.cadth.ca/media/pdf/186_EconomicGuidelines_e.pdf)

Economic evaluations: Economic evaluations are studies in which a comparison of two or more alternatives is undertaken, and costs and outcomes are examined for each alternative. Economic evaluations can be classified as cost-benefit analysis, cost-utility analysis, or cost-effectiveness analysis (including cost-consequences analysis).

Gold standard: A gold standard is a set of known relevant records that are determined by hand-searching or another method of sensitive selection such as relative recall. The gold standard is also known as the reference standard.

Hand-searching: Hand-searching is the scanning of publications cover to cover to identify relevant studies.

Health Economic Evaluations Database (HEED): This is a database of economic evaluations published by Wiley InterScience (available: <http://www3.interscience.wiley.com/cgi-bin/mrwhome/114130635/HOME>)

National Health Service Economic Evaluation Database (NHS EED): This is the database of economic evaluations compiled by the Centre for Reviews and Dissemination at the University of York (available: <http://www.crd.york.ac.uk/crdweb/>)

Partial economic evaluations: Partial economic evaluations consider costs or consequences. They neither involve a comparison between alternative interventions nor do they relate costs to

consequences; for example as cost comparison or cost-analysis studies and cost-of-illness studies do.

Precision: Precision is the percentage of records that are retrieved using a filter that are deemed to be relevant. Precision is calculated as the number of relevant records retrieved divided by the total number of records retrieved multiplied by 100.

Search filters: Search filters are collections of search terms that are used to retrieve selections of records. Search filters may be designed to retrieve records of research using a study design or by topic, or by another feature of the research question.
(<http://www.york.ac.uk/inst/crd/intertasc/about.htm>).

Sensitivity: Sensitivity is the percentage of known relevant records that are retrieved by a filter. Sensitivity is calculated as the number of relevant records retrieved divided by the total number of relevant records multiplied by 100.

Specificity: Specificity is the percentage of irrelevant records successfully not retrieved by using a filter. Specificity is calculated as the number of irrelevant records not retrieved divided by the total number of irrelevant records multiplied by 100.

ABBREVIATIONS

CADTH	Canadian Agency for Drugs and Technologies in Health
HEED	Health Economic Evaluations Database
HTA	health technology assessment
MeSH	Medical Subject Headings
NHS EED	National Health Service Economic Evaluation Database
NHS QIS	National Health Service Quality Improvement Scotland

TABLE OF CONTENTS

EXECUTIVE SUMMARY	iii
GLOSSARY	vi
ABBREVIATIONS	viii
1 INTRODUCTION	1
2 OBJECTIVES	2
3 METHOD	2
3.1 Definition of Economic Evaluations	3
3.2 Identification of Gold Standard	3
3.3 Creating Comparator Sets of Records	4
3.4 Word Occurrence Analysis	5
3.5 Discriminating Terms and Search Filters	6
3.6 Testing Search Filter Performance	7
4 RESULTS	10
4.1 Gold Standard and Comparator Records	10
4.2 Term Occurrence Analysis and Filter Development	11
4.3 Testing Search Filter Performance	14
4.4 Testing Search Filter Performance	18
5 DISCUSSION	26
5.1 Summary of Findings	26
5.2 Study Limitations	27
5.3 Generalizability and Recommendations	29
5.4 Conclusions	30
6 REFERENCES	31
APPENDIX 1: Search strategies that are used to identify comparator records	
APPENDIX 2: Stop word list	
APPENDIX 3: Candidate search terms	
APPENDIX 4: Published search filters used in comparative testing	
APPENDIX 5: Sensitivity-precision trade-offs between filters	

1 INTRODUCTION

Health care decision-makers and medical professionals who are practising evidence-based care need access to the best evidence. Health technology assessment (HTA) agencies and guideline developers who are interested in the cost-effectiveness of health care technologies need access to evidence from economic evaluations to identify the cost-effectiveness of a technology and to inform models.

Access to economic evaluations and other economic information such as costing studies has improved with the development of the National Health Service Economic Evaluation Database (NHS EED) and the Health Economic Evaluations Database (HEED). Despite these valuable resources, however, there are still identification issues that arise from several factors:

- There is a lag between the time when records are being identified from databases such as MEDLINE and full abstracts and the time when they are being added to NHS EED and HEED. NHS EED flags the records of studies being abstracted, but these efforts do not include the provision of a preliminary abstract. As a result, the search efficiency for those records is reduced while the full record is being produced.
- NHS EED focuses on economic evaluations. Although records reporting other economic information such as costing studies are included in the database, these records lack abstracts. Therefore, the search sensitivity for such studies may be reduced in comparison with MEDLINE, for example.
- The search filters that were developed by the Centre for Reviews and Dissemination and the Office of Health Economics to identify records from databases that were to be considered for inclusion in NHS EED and HEED are necessarily designed to be highly sensitive and tend to have low precision.¹ If these filters are used to identify the records of studies that have not yet been abstracted in NHS EED or HEED, they may be inadequately focused for researchers producing technology assessments and guidelines.
- The search filters that are used to identify records for consideration in NHS EED have not been revised since they were developed in the mid-1990s. Meanwhile, economic evaluation methods and terminology have evolved, particularly in the increased use of descriptions of utility values and modelling methods.
- Many of the search terms that are used in economics are also used in other contexts, for example, “energy costs,” “energy expenditure,” and “animal models.” These ambiguities may not be handled well when current filters are used. Recent innovations in filter design methods and text retrieval research, such as frequency analysis, discriminant analysis, and classification trees, may make these issues easier to tackle and thereby improve precision.

HTA agencies and guideline producers have perceived that the identification of economic studies and other economic information is often problematic, and that the current economic evaluation and quality of life filters are not ideal.

To address this perceived problem, there is a need to develop search filters that can be used to identify economic evaluations, at least in MEDLINE and EMBASE. The sensitivity of these filters should be maximized, while a sufficient level of precision should be maintained to meet the needs of guidelines developers and HTA researchers. Sensitivity is the percentage of known relevant records that are retrieved by a filter. Precision is the percentage of records that are

retrieved using a filter and that are relevant. The achievement of high sensitivity often means a lowering of precision and vice versa. This trade-off is a challenge that all filter developers encounter.² Sensitivity and precision are determined by the terms that are selected for the filter, by the ways that the filter terms are combined using Boolean operators, and by the number of relevant records to be found in the database.

2 OBJECTIVES

The objectives of this research were to develop search filters to identify economic evaluations in the MEDLINE and EMBASE databases, and to obtain data on the relative performance of new and published search filters in those databases.

HTA researchers at the Canadian Agency for Drugs and Technologies in Health (CADTH) require high sensitivity in the identification of relevant studies (ideally more than 0.95). A high precision is ideal (more than 0.80), but given the trade-off between sensitivity and precision, researchers tend to accept the fact that precision will be lower than sensitivity and may change with the choice of different datasets where economic evaluations form a larger or smaller proportion of the dataset. In some circumstances, such as scoping exercises where a rapid overview of the evidence is needed, a reduction in sensitivity with an increase in precision might be acceptable. To create filters that would support these approaches, the following objectives were set for performance:

- a sensitivity-maximizing approach (sensitivity 0.95, precision 0.2)
- a precision-maximizing approach (sensitivity 0.8, precision 0.5)
- a balance between sensitivity and precision.

We sought to develop a range of filters to achieve different combinations of sensitivity, precision, and specificity to cater to differing search needs.^{1,3}

3 METHOD

There are many approaches to search filter development.^{1,2,4-7} This research used elements of the approaches that have been labelled “second generation” and “third generation” by Jenkins.² Second generation filter design, according to Jenkins, involves the derivation of search terms from a gold standard set of relevant records, the development of search filters based on the search terms, and the testing of the filters’ performance on the gold standard from which they were derived. Search filters are ideally validated on additional gold standards that have not been used to derive the filters, because the search filters may over-perform on the gold standard from which they were derived. In keeping with Jenkins’ categorization, this research also uses “third generation” techniques, including the use of (non-relevant) comparator records to assist with the creation of search filters that can discriminate (relevant) gold standard records from comparator records. In addition, this research builds on previous work by one of the authors (JG) in using a data-led approach to deriving the terms to be tested.^{8,9} This involves the use of terms that are in the records to determine which terms should be tested, rather than testing terms that have been derived from looking in thesauri, brainstorming, or other techniques. The best performing terms

were determined by using classification trees.¹⁰ The advantage of using this technique in this situation rather than the more usual additive models is that it allows for interactions between the predictive terms, particularly when there are many possible predictors and potential interactions.

3.1 Definition of Economic Evaluations

There are several ways in which researchers may define the term “economic evaluation,” and different research teams may classify economic evaluations in different ways. For this study, economic evaluations that are used by CADTH researchers were the focus. Relevant types of economic evaluations included the following terms, which are defined in the Glossary at the beginning of this document:

- Cost-effectiveness analyses
- Cost-utility analyses
- Cost-benefit analyses
- Cost-minimization studies
- Studies with multiple outcome measures (cost-consequences analyses).

Studies reporting utility estimates and partial economic evaluations were not included in the definition (see Glossary).

3.2 Identification of Gold Standard

The identification of a gold standard set of known database records meeting the criteria of an economic evaluation was required. Gold standard sets of records are ideally identified by hand-searching, but may also be developed by using other methods such as relative recall.^{2,7,9,11} The number of gold standard records that are needed for this research was determined statistically. The target total number of gold standard records was set at 2,000 so that the sensitivity could be estimated to within $\pm 2.5\%$, based on the 50% of records used as the validation set and assuming a minimum sensitivity of 80%.

There were insufficient resources to allow for extensive hand-searching, so the gold standard for this research was achieved using relative recall.¹¹ The identification of a gold standard set of known economic evaluations was achieved by searching NHS EED (<http://www.crd.york.ac.uk/crdweb/>). The NHS EED database of economic evaluations is populated after extensive sensitive searches of a range of databases and hand-searched journals (<http://www.crd.york.ac.uk/crdweb/html/help.htm>). Of the thousands of records that are identified each year, a small proportion is judged by health economists on the database team to be full economic evaluations according to the NHS EED definitions (<http://www.crd.york.ac.uk/crdweb/html/help.htm>). With the sensitive search approach and selection process that are used for NHS EED, collections of NHS EED records can form a relative recall gold standard, as an alternative to hand-searching a large number of journals.

Economic evaluations in NHS EED have full structured abstracts and are easy to identify in the database after searching on the coding in the TY field (see the NHS EED help pages). The gold standard was created by downloading all records that are coded as economic evaluations from NHS EED using the following strategy:

("economic evaluation":ty or "provisional abstract":ty) NOT ("partial":ty or "outcome":ty)

Using this strategy, we could retrieve all NHS EED records that have been coded as an economic evaluation and that had a full abstract ("economic evaluation":ty) or were getting a full abstract ("provisional abstract":ty). This search strategy excluded partial economic evaluations ("partial":ty) and outcome evaluation studies ("outcome":ty), which might otherwise have been retrieved because of the use of the first two search terms.

Records were downloaded for publications that had appeared in 2000, 2003, and 2006. These years were chosen so that adequate numbers of records were provided for analysis and so that as much of the decade as possible could be spanned. Different years were chosen to identify whether some search terms were more discriminating in earlier years than in later years. The year 2006 was chosen as the final year because it was the most recent year when all published articles were likely to have been identified and categorized. The records were downloaded from NHS EED in December 2008, and the database was unlikely to have complete sets of publications that were recorded for 2007 and 2008 at that time.

NHS EED records were downloaded into EndNote reference management software.¹² The records were then checked against MEDLINE and EMBASE. Where corresponding records were available in MEDLINE and EMBASE, they were downloaded to create two gold standard sets of records: the MEDLINE gold standard and the EMBASE gold standard. If a corresponding record was not identified, no record could be downloaded.

3.3 Creating Comparator Sets of Records

The search filters were developed by using a gold standard and a comparator set of economic publication records to identify those terms that best discriminate economic evaluations from other economic publications. This approach, which has been described by Jenkins, has been used in developing search filters for systematic reviews and randomized controlled trials.^{2,8,9}

The number of comparator records that was required was determined statistically: 4,000 records would be adequate to estimate the specificity to within $\pm 2.2\%$, based on the 50% of records used as the validation set and assuming a specificity of 50%. (This specificity is conservative in that it gives the maximum sample size.) For every gold standard record, two comparator records were required. The comparator record sets were created by finding records using the search filters to identify economic evaluations in MEDLINE and EMBASE used by the NHS EED database team (<http://www.crd.york.ac.uk/crdweb/html/help.htm>). These are sensitive filters that are designed so that they can be used to find a wider range of records that mention economics. These records are then assessed by the NHS EED team, which selects the records that are economic evaluations for abstracting.¹³ These filters produce low precision results. The comparator records in this

project were expected to include those that contribute to low precision in the NHS EED filters. The NHS EED filters for MEDLINE and EMBASE were used to identify a set of comparator records, and the records for 2000, 2003, and 2006 were downloaded (Appendix 1). Ovid MEDLINE was searched in the database segment covering 1950 to November week 3, 2008, and included in-process and non-indexed records. Non-indexed records and in-process records may feature in the comparator sets and will not have subject headings or publication types that have been assigned by MEDLINE indexers. Ovid EMBASE was searched in the database segment covering 1980 to 2009 Week 02.

The comparator records were selected randomly from all the records that were retrieved by using random numbers that were generated using Research Randomizer (<http://www.randomizer.org/form.htm>). Records that proved, upon checking, to be economic evaluations were removed from the comparator set, and substitute records were randomly selected and downloaded.

3.4 Word Occurrence Analysis

The filters were developed using a statistical data-led approach that was based on approaches described by Jenkins and previously developed in research involving one of the researchers on this project (JG).^{2,8,9}

A 50% random sample of the records from each gold standard database (and for each year in a database) was selected using random numbers that were generated by the Research Randomizer website (<http://www.randomizer.org/form.htm>). These records were then analyzed using Endnote's subject bibliography feature. This provided data on which words occurred in which records and allowed the compilation of tables of word occurrences in different parts of the records:

- Record title
- Record abstract
- Subject headings (Medical Subject Headings [MeSH] and Emtree)
- Single words in subject headings (MeSH and Emtree)
- Subheadings (MeSH and Emtree)
- Subject headings coordinated with subheadings (MeSH and Emtree)
- Type of article (Publication types: MeSH only).

Words, subject headings, subheadings, and publication types will all be described as “terms” in this report.

Common words or stop words (Appendix 2) such as “the,” “an,” and “in,” which occurred often but were unlikely to contribute to the uniqueness of records, were removed from the occurrence tables.

The terms were selected for analysis according to their occurrence in the 50% random sample of gold standard records. All terms (except stop words) that occurred above selected cut-off occurrence values (Table 1) were used in the analysis. The cut-off values were determined subjectively. The compilation of the occurrence data for gold standard and comparator records

was time consuming, so the cut-off values were based on the number of variables that could be processed during the time that was available. The terms in the record title were given a lower cut-off than the terms in the abstract and subject headings because it was expected that the title would contain a higher proportion of discriminating terms than the abstract.

Table 1: Term Occurrence Cut-Offs for Terms to Be Included in Analysis		
Term Location	MEDLINE 50% Gold Standard Sample	EMBASE 50% Gold Standard Sample
Record title	Terms occurring in at least 2% of records	Terms occurring in at least 1% of records
Record abstract	Terms occurring in at least 5% of records	Terms occurring in at least 5% of records
Subject headings and subheadings	Terms occurring in at least 5% of records	Terms occurring in at least 5% of records
Single words in subject headings and subheadings	Terms occurring in at least 5% of records	Terms occurring in at least 5% of records
Coordinated subject headings and subheadings	Terms occurring in at least 5% of records	Terms occurring in at least 5% of records
Publication type	All headings	Not applicable

3.5 Discriminating Terms and Search Filters

The occurrence or absence of the candidate terms in each gold standard and comparator record was compiled into a table. A sample encompassing 50% of the gold standard and comparator records was analyzed using S-Plus classification trees to identify those terms that best discriminated gold standard economic evaluations from non-economic evaluations in MEDLINE and EMBASE records. The analysis was conducted using the statistical software package S-Plus (Version 6.1).¹⁴

In the classification tree approach, a statistical algorithm (recursive partitioning) was used to identify terms that best discriminated between gold standard and comparator group citations.^{10,14} The result is a classification tree. Like a key that is used to identify species, a classification tree defines the differences between the records that allow the record type to be identified. Classification trees work well with large numbers of predictor variables. They are easy to interpret and to convert into a search filter.¹⁰

In the classification tree approach the records are divided until the terminal nodes of the tree are homogenous or until a further division would reduce the size of the terminal node to less than five records. This procedure produces trees that are often large and complex. It would be difficult to convert such a tree into a search filter, and the filter would likely be over-fitted to the training data set, in that it fits the training set closely but may not perform as well on a new set of records. To avoid this, the simplification of the initial tree is often achieved by removing less important splits that occur closest to the tree's terminal nodes (pruning of the classification tree).

In this project, we defined the number of terminal nodes and pruned the tree in such a way as to reduce the number of misclassifications. Using the S-Plus function `prune.misclass`, it is possible

to specify the size (the number of terminal nodes) of the pruned tree.¹⁴ The algorithm removes the least important splits according to the cost-complexity measure. Where it has been necessary to maximize sensitivity or precision, the pruned tree has been reviewed to select the branches that give the best sensitivities or precisions to define the number of terminal nodes. The tree is pruned in such a way as to reduce the number of misclassifications. The number of terminal nodes was kept low to build trees that produced simple search filters. A series of search filters with different numbers of terminal nodes and thus different levels of sensitivity, precision, and specificity were developed.

Classification trees were developed for the records from the MEDLINE and EMBASE databases. For each database, 50% of the data were used to develop the tree (the training set), and 50% of the data were used to evaluate the effectiveness of this tree (the validation set). For each search filter that was created, the parameters sensitivity, specificity, and precision were calculated based on the training and validation sets.

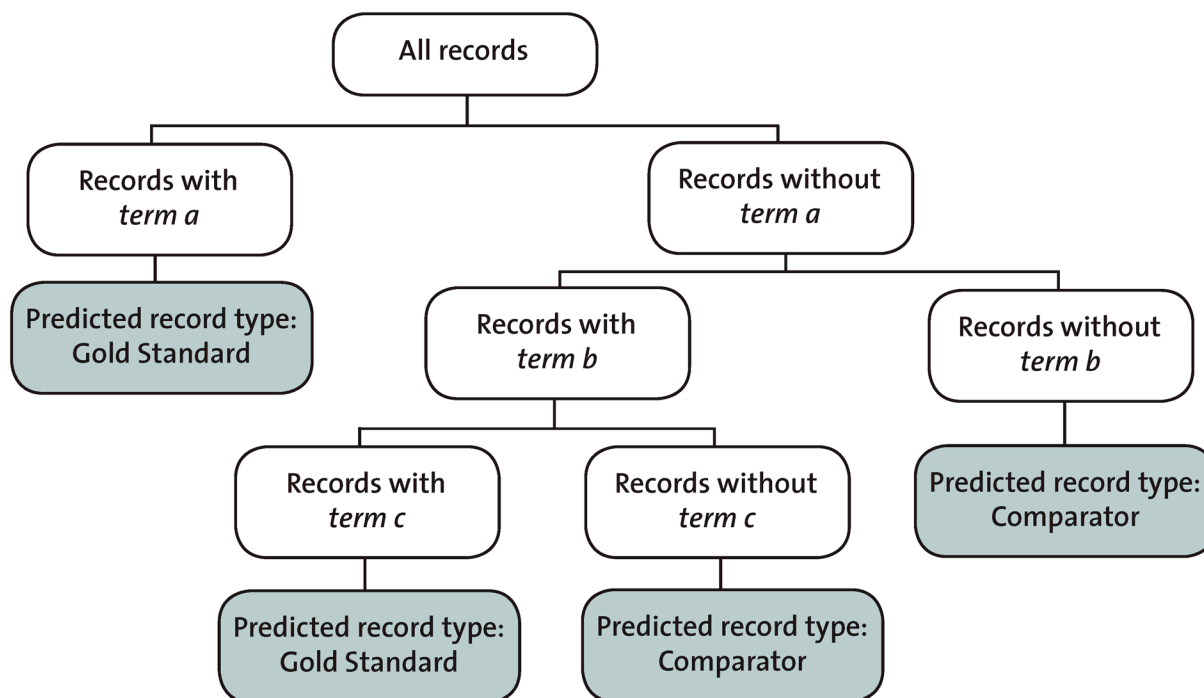
3.6 Testing Search Filter Performance

The best discriminating terms that were identified from the recursive partitioning analysis were compiled into search filters. The classification trees yielded a series of splits (each based on a term) that divided the records into those that were predicted to be gold standard records and those that were predicted to be comparator records. The first term in the tree defined the simplest strategy (a search for all words that included that term). More complex strategies were developed by including more splits. For a tree with several splits, some nodes would predict gold standard records and some nodes would predict comparator records. The search strategy would then be the series of terms combined using the Boolean operators (OR, AND, or NOT), such that the strategy described a path to the nodes where the probability of any record being a gold standard is above a certain value. Raising this value produced more precise trees; lowering it produced more sensitive trees. For the tree in Figure 1, the search strategy would be to find records with term “a” OR (term “b” AND term “c”).

The analysis retained the information about the type of term being tested, so the filters could reflect the type of term. For example, if the classification tree indicated that a specific subject heading was discriminating, the field limit “/” could be used for accurate retrieval in the Ovid interface. Similarly, title words could be identified from the classification tree and could be limited to the title field in Ovid using .ti.

The search filters resulting from the analysis were named MEDLINE A, MEDLINE B, and so forth, and were tested in the Ovid interface to MEDLINE. Analogously, EMBASE filters were named EMBASE A, EMBASE B, and so forth. To test the performance of the new filters in the Ovid interface, all the gold standard records were retrieved from the two databases after searching on their accession numbers (PubMed identifiers and EMBASE unique record numbers). The performance of the search filters was assessed in terms of the number of gold standard records that were identified (sensitivity) and the proportion of gold standard records to non-gold standard records that were retrieved (precision).

Figure 1: Example Classification Tree



The performance of the search filters was independently tested in Ovid by two researchers (JG, DK). The filter combinations appear in Table 2. The filter performance was evaluated based on records from 2000, 2003, or 2006 by restricting the search results using the search term (2000 or 2003 or 2006).yr.

Each filter was tested with and without exclusion. These exclusions consisted of publication types that are unlikely to yield detailed reports of economic evaluations and animal studies that are unlikely to be required for most HTA research. The exclusion searches appear in Table 2 (C. Exclusion strategy). The exclusion searches were performed differently in the two databases. No gold standard records were lost using the exclusion filter in MEDLINE, so it was retained and all results are based on its use. In EMBASE, the exclusion filter resulted in the loss of gold standard records, so the testing is reported without the use of the filter.

Table 2: Search Filter Testing Strategy for MEDLINE and EMBASE		
Description of Stage of Strategy	MEDLINE Testing Strategy	EMBASE Testing Strategy
A. Gold standard records set	Identify gold standard records using accession numbers	Identify gold standard records using accession numbers
B. Find relevant years	(2000 or 2003 or 2006).yr.	(2000 or 2003 or 2006).yr.
C. Exclusion strategy	1. Letter.pt. 2. Editorial.pt. 3. Historical article.pt. 4. Animals/ not humans/ 5. 1 or 2 or 3 or 4	1. Letter.pt. 2. Editorial.pt. 3. Note.pt. 4. Exp animal/ 5. Exp animal experiment/ 6. Nonhuman/ 7. (rat or rats or mouse or mice or hamster or hamsters or animal or animals or dog or dogs or cat or cats or bovine or sheep).ti,ab,sh. 8. exp human/ 9. exp human experiment/ 10. or/5-7 11. or/8-9 12. 10 not (10 and 11) 13. 1 or 2 or 3 or 12
D. Test filter terms	e.g., Term 1 or term 2	e.g., Term 1 or term 2
E. Sets are combined to find how many records are retrieved by test filter	(D AND B) NOT C	(D AND B) NOT C
F. Sets are combined to find how many gold standard records are retrieved by filter	E AND A	E AND A
.yr. Publication year .pt. Publication Type / Subject heading Exp Explode (subject heading) .ti. Title .ab. Abstract .sh. Subject heading Or/1-3 Combine sets 1 to 3 using OR		

The performance of other published and unpublished filters for MEDLINE and EMBASE were also tested. These filters were combined with the same exclusions from Table 2. The filters were identified from the InterTASC Information Specialists' Sub-Group search filter website (<http://www.york.ac.uk/inst/crd/intertasc/econ.htm>), from colleagues at CADTH, and from reviewers of an early draft of this report. The comparative filters that were tested were:

- CADTH filter¹⁵
- Emory University (Grady) filter¹⁶
- McKinlay et al. filters¹⁷

- NHS EED filter¹⁸
- NHS Quality Improvement Scotland (NHS QIS) filters¹⁹
- Royle and Waugh filters⁴
- Sassi et al. filters¹
- Scottish Intercollegiate Guidelines Network filter²⁰
- Wilczynski et al. filters.³

The search filters by McKinlay et al., Royle and Waugh, Sassi et al., and Wilczynski et al. have been developed using various methods and were reported in journals.^{1,3,4,17} The other filters were published, without details about the development methods, on websites and in reports.^{15,16,18-20}

The search terms in the comparative filters that were tested appear in Appendix 4.

The effect of combining the high sensitivity published filters with the filters that were derived from this research was tested by one researcher (JG) to explore whether a two-step filter could improve precision. The new filters were combined with published filters by first running the published filter and then by running a new filter and combining the results using AND.

The sensitivity and precision scores were calculated for all filters. The sensitivity was calculated as the percentage of gold standard records that were retrieved using the filter:

(number of gold standard records retrieved / total number of gold standard records)

The precision was calculated as the percentage of relevant (gold standard) records that were retrieved from all the records that were retrieved using the filter:

(number of gold standard records retrieved / total number of records retrieved)

4 RESULTS

4.1 Gold Standard and Comparator Records

A total of 2,070 full economic evaluations were identified from NHS EED from 2000, 2003, and 2006. Across the years, the number of evaluations in the database increased from 612 to 789. Overall, 1,957 records of evaluations had corresponding records in MEDLINE, and 1,876 had records in EMBASE. A breakdown that is stratified by year appears in Table 3.

Table 3: Numbers of Economic Evaluation Records Downloaded from NHS EED, and Numbers of Those Records Subsequently Identified in MEDLINE and EMBASE

Year Published	Number of NHS EED Records Downloaded	Number of NHS EED Records with Matching MEDLINE Records: MEDLINE Gold Standard	Number of NHS EED Records with Matching EMBASE Records: EMBASE Gold Standard
2000	612	577	538
2003	669	618	595
2006	789	755	743
Total	2070	1,957	1,876

NHS EED = National Health Service Economic Evaluation Database.

The MEDLINE comparator set comprised 4,136 records and the EMBASE comparator set consisted of 3,750 records. Table 4 shows a breakdown by year.

Table 4: Numbers of Records Retrieved by Searches and Numbers of Records in Comparator Sets for MEDLINE and EMBASE

Year Published	Number of MEDLINE Records Retrieved by NHS EED Filter in MEDLINE	Number of Records in MEDLINE Comparator Set	Number of EMBASE Records Retrieved by NHS EED Filter in EMBASE	Number of Records in EMBASE Comparator Set
2000	17,567	1,226	13,300	1,076
2003	20,936	1,335	17,571	1,190
2006	24,171	1,575	20,735	1,484
Total	62,674	4,136	51,606	3,750

NHS EED = National Health Service Economic Evaluation Database.

4.2 Term Occurrence Analysis and Filter Development

Overall, 347 terms were analyzed for MEDLINE and 528 terms were analyzed for EMBASE. The list of candidate search terms appears in Appendix 3.

The first series of classification trees that were analyzed yielded filters with good precision performance, but lower than expected sensitivity, despite the search terms being intuitively sensitive. The filters were labelled MEDLINE A to MEDLINE D (Tables 5 and 6). The classification trees were reanalyzed with a focus on maximizing sensitivity to explore whether better performing filters would emerge when the following terms were excluded:

- Humans as a subject heading
- Hospitalization as a subject heading
- Female as a subject heading
- Male as a subject heading
- Middle aged as a subject heading.

The best performing filters for MEDLINE from the second series of analyses, appear in Tables 5 and 6 and are labelled MEDLINE E to MEDLINE H. Terms that were expected to be associated

with higher levels of sensitivity and precision based on their role as relevant index terms (for example, subject headings such as “Cost-benefit Analysis/” in MEDLINE) proved less discriminating than other terms that were expected to be associated with poor precision (for example, “costs.ab.” and “economics.fs.”).

Table 5: MEDLINE Test Search Filters Emerging from Classification Tree Analysis	
Filter Name	Filter Search Terms
MEDLINE A	Humans/ AND economics.fs.
MEDLINE B	1 Humans/ AND economics.fs. 2 (Humans/ AND therapy.fs.) NOT economics.fs. 3 1 OR 2
MEDLINE C	1 Humans/ AND economics.fs. 2 (Humans/ AND therapy.fs.) NOT economics.fs. 3 (Humans/ AND Randomized Controlled Trial.pt.) NOT (economics.fs. OR therapy.fs.) 4 1 OR 2 OR 3
MEDLINE D	1 Humans/ AND economics.fs. 2 (Humans/ AND therapy.fs.) NOT economics.fs. 3 (Humans/ AND Randomized Controlled Trial.pt.) NOT (economics.fs. OR therapy.fs.) 4 (Humans/ AND Dollars.ab.) NOT (economics.fs. OR therapy.fs. or randomized controlled trial.pt.) 5 1 OR 2 OR 3 OR 4
MEDLINE E	Economics.fs.
MEDLINE F	1 Economics.fs. 2 Randomized controlled trial.pt. 3 Costs.ab. 4 1 OR 2 OR 3
MEDLINE G	1 Economics.fs. 2 (cost-benefit analysis/ OR randomized controlled trial.pt.) 3 1 AND 2
MEDLINE H	1 Randomized Controlled Trial.pt. 2 Economics.fs. AND (Cost-benefit analysis/ OR cost.ab.) 3 (Costs.ab. AND (Cost adj Effectiveness.ab.)) NOT economics.fs. 4 1 OR 2 OR 3
/	Subject heading
.fs.	Floating subheading
.pt.	Publication Type
.ab.	Abstract
Adj	Adjacent terms

The best performing filters emerging from the classification tree analysis for EMBASE appear in Tables 7 and 8, and are labelled EMBASE A to EMBASE H.

Table 6: Sensitivity, Precision, and Specificity for Test Search Filters in MEDLINE Data						
Filter Name	Training Set			Validation Set		
	Sensitivity	Precision	Specificity	Sensitivity	Precision	Specificity
MEDLINE A	0.800	0.981	0.993	0.808	0.975	0.990
MEDLINE B	0.890	0.946	0.976	0.889	0.951	0.978
MEDLINE C	0.924	0.933	0.969	0.914	0.931	0.968
MEDLINE D	0.942	0.927	0.965	0.932	0.927	0.965
MEDLINE E	0.807	0.687	0.826	0.811	0.688	0.826
MEDLINE F	0.960	0.577	0.669	0.953	0.554	0.639
MEDLINE G	0.675	0.924	0.974	0.645	0.929	0.977
MEDLINE H	0.867	0.848	0.927	0.846	0.816	0.910

Table 7: EMBASE Test Search Filters Emerging from Classification Tree Analysis	
Filter Name	Filter Search Terms
EMBASE A	Cost adj effectiveness.ab.
EMBASE B	(Cost adj effectiveness.ab.) or costs.ab.
EMBASE C	1 (Cost adj effectiveness.ab.) or costs.ab. 2 Cost effectiveness analysis/ 3 1 OR 2
EMBASE D	Cost.hw.
EMBASE E	1 Cost adj effectiveness.ab. 2 Costs.ab. 3 Cost.ab. 4 1 OR 2 OR 3
EMBASE F	1 Cost adj effectiveness.ti. 2 (Costs.ab. and controlled study/ and cost.hw.) AND (effectiveness.hw. or randomized controlled trial/) 3 1 OR 2
EMBASE G	1 Cost adj effectiveness.ab. 2 Cost adj effectiveness.ti. 3 Life adj years.ab. 4 Life adj year.ab. 5 Qaly.ab. 6 (Cost or costs).ab. and Controlled Study/ 7 (Cost and costs).ab. 8 1 OR 2 OR 3 OR 4 OR 5 OR 6 OR 7
EMBASE H	1 Cost adj effectiveness.ti. 2 (cost adj effectiveness).ab. AND (costs or cost).ab. 3 (cost AND costs).ab. AND cost effectiveness analysis/ 4 (Costs.ab. AND controlled study/ and cost.hw.) AND (effectiveness.hw. OR randomized controlled trial/) 5 1 OR 2 OR 3 OR 4
Adj	Adjacent terms
.ab.	Abstract
/	Subject heading
.hw.	Subject heading word
.ti.	Title

Table 8: Sensitivity, Precision, and Specificity for Search Filters in EMBASE Data						
Filter Name	Training Set			Validation Set		
	Sensitivity	Precision	Specificity	Sensitivity	Precision	Specificity
EMBASE A	0.481	0.867	0.963	0.472	0.841	0.955
EMBASE B	0.833	0.665	0.790	0.824	0.655	0.782
EMBASE C	0.918	0.582	0.670	0.897	0.572	0.665
EMBASE D	0.969	0.473	0.460	0.975	0.476	0.462
EMBASE E	0.972	0.544	0.593	0.957	0.545	0.600
EMBASE F	0.519	0.929	0.980	0.552	0.945	0.984
EMBASE G	0.909	0.715	0.819	0.915	0.718	0.820
EMBASE H	0.674	0.898	0.962	0.684	0.888	0.957

4.3 Testing Search Filter Performance

The results of testing the eight new filters (MEDLINE A to MEDLINE H) in Ovid MEDLINE, the 13 published and unpublished comparative filters, and the nine combined new and published filters appear in Tables 9 and 10. Table 9 presents the results in order of highest sensitivity, and the results are presented in order of the highest precision in Table 10.

Because the search results were restricted to 2000, 2003, and 2006, two gold standard records were missed in MEDLINE. The publication years that were listed in MEDLINE for these records differed from those that were reported in NHS EED. Thus, the MEDLINE gold standard set was reduced to 1,955 retrievable records in Ovid MEDLINE.

In terms of sensitivity (more than 0.990) and with the exclusions (Table 2), the best performing filters in MEDLINE were NHS QIS (full and brief), NHS EED, and Royle and Waugh.^{4,18,19} Among these four filters, the best performing filter in terms of precision was the NHS EED filter (0.04). The achievement of higher levels of precision required a lowering of sensitivity. The combination of the NHS EED filter with the new MEDLINE D filter (using AND) achieved more than 0.092 precision with a corresponding sensitivity of 0.943. The Wilczynski best optimization filter, searched alone, achieved 0.093 precision with 0.923 sensitivity.³ The Wilczynski filter was the more concise of the two filters.

In terms of maximum precision, the best performing filter in MEDLINE, including the exclusion limits, was the new MEDLINE G filter (0.720 sensitivity and 0.257 precision). The filter that best met the project objective (good precision with a sensitivity of greater than or equal to 0.80) was the Emory University (Grady) filter (0.845 sensitivity and 0.133 precision).¹⁶

Table 9: Performance of Search Filters in Ovid MEDLINE in Order of Decreasing Sensitivity, Searched with Exclusion Limits

Filter	Filter Abbreviation	Number of MEDLINE Records Identified	Number of Gold Standard Records Identified (out of 1,955 total records)	Sensitivity	Precision
NHS QIS (full)	QISF	457,569	1,955	1.000	0.004
NHS QIS (brief)	QISB	111,551	1,955	1.000	0.018
NHS EED	NHS	48,917	1,953	0.999	0.040
Royle and Waugh	RW	67,521	1,948	0.996	0.029
CADTH	CADTH	74,746	1,912	0.978	0.026
MEDLINE F	MF	73,898	1,904	0.974	0.026
NHS QIS brief plus MEDLINE F	QISBMF	39,641	1,904	0.974	0.048
NHS EED plus MEDLINE F	NHSMF	24,214	1,903	0.973	0.079
Royle and Waugh plus MEDLINE F	RWMF	25,561	1,898	0.971	0.074
MEDLINE D	MD	162,886	1,845	0.944	0.011
NHS QIS brief plus MEDLINE D	QISBMD	37,964	1,845	0.944	0.049
NHS EED plus MEDLINE D	NHSMD	20,028	1,843	0.943	0.092
Royle and Waugh plus MEDLINE D	RWMD	25,098	1,838	0.940	0.073
MEDLINE C	MC	162,435	1,829	0.936	0.011
NHS QIS brief plus MEDLINE C	QISBMC	37,689	1,829	0.936	0.049
NHS EED plus MEDLINE C	NHSMC	19,786	1,827	0.935	0.092
Royle and Waugh plus MEDLINE C	RWMC	24,875	1,822	0.932	0.073
Wilczynski best optimization of sensitivity and specificity	WOPT	19,362	1,805	0.923	0.093
MEDLINE H	MH	48,051	1,804	0.923	0.038
Sassi extensive filter	SAS1	30,445	1,799	0.920	0.059
SIGN	SIGN	38,071	1,766	0.903	0.046
MEDLINE B	MB	130,054	1,754	0.897	0.014
Wilczynski best sensitivity	WSENS	24,391	1,726	0.883	0.071
MEDLINE E	ME	30,824	1,722	0.881	0.056
MEDLINE A	MA	25,675	1,712	0.876	0.067
Grady	GR	12,389	1,651	0.845	0.133
Sassi selective filter A	SASA	15,608	1,643	0.840	0.105

Table 9: Performance of Search Filters in Ovid MEDLINE in Order of Decreasing Sensitivity, Searched with Exclusion Limits

Filter	Filter Abbreviation	Number of MEDLINE Records Identified	Number of Gold Standard Records Identified (out of 1,955 total records)	Sensitivity	Precision
MEDLINE G	MG	5,479	1,408	0.720	0.257
Sassi selective filter C	SASC	6,316	1,393	0.713	0.221
Wilczynski best specificity	WSPEC	8,907	1,375	0.703	0.154
Sassi selective filter B	SASB	5,829	1,329	0.680	0.228

CADTH = Canadian Agency for Drugs and Technologies in Health; NHS EED = National Health Service Economic Evaluation Database; NHS QIS = National Health Service Quality Improvement Scotland; SIGN= Scottish Intercollegiate Guidelines Network

Table 10: Performance of Search Filters in Ovid MEDLINE, in Order of Decreasing Precision, Searched with Exclusion Limits

Filter	Filter Abbreviation	Number of MEDLINE Records Identified	Number of Gold Standard Records Identified (out of 1,955 total records)	Sensitivity	Precision
MEDLINE G	MG	5,479	1,408	0.720	0.257
Sassi selective filter B	SASB	5,829	1,329	0.680	0.228
Sassi selective filter C	SASC	6,316	1,393	0.713	0.221
Wilczynski best specificity	WSPEC	8,907	1,375	0.703	0.154
Grady	GR	12,389	1,651	0.845	0.133
Sassi selective filter A	SASA	15,608	1,643	0.840	0.105
Wilczynski best optimization of sensitivity and specificity	WOPT	19,362	1,805	0.923	0.093
NHS EED plus MEDLINE C	NHSMC	19,786	1,827	0.935	0.092
NHS EED plus MEDLINE D	NHSMD	20,028	1,843	0.943	0.092
NHS EED plus MEDLINE F	NHSMF	24,214	1,903	0.973	0.079
Royle and Waugh plus MEDLINE F	RWMF	25,561	1,898	0.971	0.074
Royle and Waugh plus MEDLINE D	RWMD	25,098	1,838	0.940	0.073
Royle and Waugh plus MEDLINE C	RWMC	24,875	1,822	0.932	0.073
Wilczynski best sensitivity	WSENS	24,391	1,726	0.883	0.071

Table 10: Performance of Search Filters in Ovid MEDLINE, in Order of Decreasing Precision, Searched with Exclusion Limits

Filter	Filter Abbreviation	Number of MEDLINE Records Identified	Number of Gold Standard Records Identified (out of 1,955 total records)	Sensitivity	Precision
MEDLINE A	MA	25,675	1,712	0.876	0.067
Sassi extensive filter	SAS1	30,445	1,799	0.920	0.059
MEDLINE E	ME	30,824	1,722	0.881	0.056
NHS QIS brief plus MEDLINE D	QISBMD	37,964	1,845	0.944	0.049
NHS QIS brief plus MEDLINE C	QISBMC	37,689	1,829	0.936	0.049
NHS QIS brief plus MEDLINE F	QISBMF	39,641	1,904	0.974	0.048
SIGN	SIGN	38,071	1,766	0.903	0.046
NHS EED	NHS	48,917	1,953	0.999	0.040
MEDLINE H	MH	48,051	1,804	0.923	0.038
Royle and Waugh	RW	67,521	1,948	0.996	0.029
MEDLINE F	MF	73,898	1,904	0.974	0.026
CADTH	CADTH	74,746	1,912	0.978	0.026
NHS QIS Brief	QISB	111,551	1,955	1.000	0.018
MEDLINE B	MB	130,054	1,754	0.897	0.014
MEDLINE D	MD	162,886	1,845	0.944	0.011
MEDLINE C	MC	162,435	1,829	0.936	0.011
NHS QIS Full	QISF	457,569	1,955	1.000	0.004

CADTH = Canadian Agency for Drugs and Technologies in Health; NHS EED = National Health Service Economic Evaluation Database; NHS QIS = National Health Service Quality Improvement Scotland; SIGN= Scottish Intercollegiate Guidelines Network.

The sensitivity-precision trade-offs for filters that achieve a sensitivity score greater than or equal to 0.900 and for filters with a sensitivity score below 0.900 appear in Appendix 5. The best performing filters at each precision value can be identified.

4.4 Testing Search Filter Performance

The results of testing the eight new filters (EMBASE A to EMBASE H) in Ovid EMBASE, the eight published and unpublished comparative filters, and nine combined new and published filters appear in Tables 11 and 12. These results reflect the performance of the search filters without the animal and publication type exclusions. Table 11 presents the results in order of highest sensitivity, and the results in order of highest precision appear in Table 12.

Because the search results were restricted to 2000, 2003, and 2006, two gold standard records were missed in EMBASE. The publication years that were listed in EMBASE for these two records differed from the years that were documented in NHS EED. Therefore, the EMBASE gold standard set was reduced to 1,873 retrievable records in Ovid EMBASE.

Filter	Filter Abbreviation	Number of EMBASE Records Identified	Number of Gold Standard Records Identified (out of 1,873 total records)	Sensitivity	Precision
NHS QIS	QIS	109,684	1,872	1.000	0.017
CADTH	CADTH	124,167	1,871	0.999	0.015
NHS EED	NHS	65,207	1,868	0.997	0.029
Royle and Waugh	RW	85,778	1,868	0.997	0.022
McKinlay best optimization of sensitivity and specificity	MCKOPT	29,068	1,847	0.986	0.064
EMBASE E	EE	27,442	1,825	0.974	0.067
NHS QIS plus EMBASE E	QISE	27,442	1,825	0.974	0.067
NHS EED plus EMBASE E	NHSE	27,117	1,825	0.974	0.067
Royle and Waugh plus EMBASE E	RWE	27,442	1,825	0.974	0.067
EMBASE D	ED	39,031	1,822	0.973	0.047
NHS QIS plus EMBASE D	QISD	37,955	1,822	0.973	0.048
Royle and Waugh plus EMBASE D	RWD	39,031	1,822	0.973	0.047
NHS EED plus EMBASE D	NHSD	37,749	1,821	0.972	0.048
SIGN	SIGN	52,190	1,772	0.946	0.034
EMBASE G	EG	13,396	1,743	0.931	0.130
NHS QIS plus EMBASE G	QISG	13,338	1,743	0.931	0.131
NHS EED plus	NHSG	13,091	1,743	0.931	0.133

Table 11: Performance of Search Filters in Ovid EMBASE in Order of Decreasing Sensitivity, Searched without Exclusion Limits

Filter	Filter Abbreviation	Number of EMBASE Records Identified	Number of Gold Standard Records Identified (out of 1,873 total records)	Sensitivity	Precision
EMBASE G					
Royle and Waugh plus EMBASE G	RWG	13,330	1,743	0.931	0.131
EMBASE C	EC	22,538	1,703	0.909	0.076
EMBASE B	EB	13,637	1,595	0.852	0.117
McKinlay best sensitivity	MCKSENS	54,972	1,555	0.830	0.028
EMBASE H	EH	5,047	1,341	0.716	0.266
McKinlay best specificity	MCKSPEC	4,946	1,174	0.627	0.237
EMBASE F	EF	2,162	1,067	0.570	0.494
EMBASE A	EA	3,553	1,028	0.549	0.289

NHS EED = National Health Service Economic Evaluation Database; NHS QIS = National Health Service Quality Improvement Scotland.

The best performing EMBASE filters in terms of sensitivity (more than 0.99 sensitivity) were NHS QIS, CADTH, Royle and Waugh, and NHS EED.^{4,15,18,19} The precision ranged from 0.015 to 0.029. The new EMBASE G filter combined (using AND) with NHS EED had the maximum precision with more than 0.900 sensitivity (sensitivity 0.931 and precision 0.133). EMBASE G used alone was a more concise filter with the same level of sensitivity, but with a drop in precision to 0.130.

The new EMBASE H filter had maximum precision with high levels of sensitivity (0.716 sensitivity and 0.266 precision). The new EMBASE G filter combined (using AND) with the NHS EED filter best met the project objective for good precision with a sensitivity of more than 0.799 (0.931 sensitivity and 0.133 precision).

The sensitivity-precision trade-offs for filters achieving a sensitivity score greater than or equal to 0.900 and for filters with a sensitivity score below 0.900 appear in Appendix 5. The best performing filters at each precision value can be identified.

Table 12: Performance of Search Filters in Ovid EMBASE in Order of Decreasing Precision, Searched without Exclusion Limits

Filter	Filter Abbreviation	Number of EMBASE Records Identified	Number of Gold Standard Records Identified (out of 1,873 total records)	Sensitivity	Precision
EMBASE F	EF	2,162	1,067	0.570	0.494
EMBASE A	EA	3,553	1,028	0.549	0.289
EMBASE H	EH	5,047	1,341	0.716	0.266
McKinlay best specificity	MCKSPEC	4,946	1,174	0.627	0.237
NHS EED plus EMBASE G	NHSG	13,091	1,743	0.931	0.133
Royle and Waugh plus EMBASE G	RWG	13,330	1,743	0.931	0.131
NHS QIS plus EMBASE G	QISG	13,338	1,743	0.931	0.131
EMBASE G	EG	13,396	1,743	0.931	0.130
EMBASE B	EB	13,637	1,595	0.852	0.117
EMBASE C	EC	22,538	1,703	0.909	0.076
NHS EED plus EMBASE E	NHSE	27,117	1,825	0.974	0.067
EMBASE E	EE	27,442	1,825	0.974	0.067
NHS QIS plus EMBASE E	QISE	27,442	1,825	0.974	0.067
Royle plus EMBASE E	RWE	27,442	1,825	0.974	0.067
McKinlay best optimization of sensitivity/specificity	MCKOPT	29,068	1,847	0.986	0.064
NHS EED plus EMBASE D	NHSD	37,749	1,821	0.972	0.048
NHS QIS plus EMBASE D	QISD	37,955	1,822	0.973	0.048
EMBASE D	ED	39,031	1,822	0.973	0.047
Royle plus EMBASE D	RWD	39,031	1,822	0.973	0.047
SIGN	SIGN	52,190	1,772	0.946	0.034
NHS EED	NHS	65,207	1,868	0.997	0.029
McKinlay best sensitivity	MCKSENS	54,972	1,555	0.830	0.028
Royle and Waugh	RW	85,778	1,868	0.997	0.022
NHS QIS	QIS	109,684	1,872	1.000	0.017
CADTH	CADTH	124,167	1,871	0.999	0.015

CADTH = Canadian Agency for Drugs and Technologies in Health; NHS EED = National Health Service Economic Evaluation Database; NHS QIS = National Health Service Quality Improvement Scotland; SIGN = Scottish Intercollegiate Guidelines Network.

The search terms that are used in the best performing filters for MEDLINE and EMBASE, in terms of sensitivity and precision, appear in Tables 13 to 16.

Table 13: MEDLINE Search Filters — Best Sensitivity

Filter Name	Filter Details
NHS QIS brief ¹⁹	1. exp Economics/ 2. quality of life/ 3. value of life/ 4. Quality-adjusted life years/ 5. models, economic/ 6. markov chains/ 7. monte carlo method/ 8. decision tree/ 9. ec.fs. 10. economic\$.tw. 11. (cost? or costing? or costly or costed).tw. 12. (price? or pricing?).tw. 13. (pharmacoeconomic? or (pharmaco adj economic?)).tw. 14. budget\$.tw. 15. expenditure\$.tw. 16. (value adj1 (money or monetary)).tw. 17. (fee or fees).tw. 18. "quality of life".tw. 19. qol\$.tw. 20. hrqol\$.tw. 21. "Quality adjusted life year\$".tw. 22. qaly\$.tw. 23. cba.tw. 24. cea.tw. 25. cua.tw. 26. utilit\$.tw. 27. markov\$.tw. 28. monte carlo.tw. 29. (decision adj2 (tree\$ or analys\$ or model\$)).tw. 30. ((clinical or critical or patient) adj (path? or pathway?)).tw. 31. (managed adj2 (care or network?)).tw. 32. or/1-31 33. Letter.pt. 34. Editorial.pt. 35. Historical article.pt. 36. Animals/ not humans/ 37. 33 or 34 or 35 or 36 38. 32 not 37
NHS EED ¹⁸	1. economics/ 2. exp "costs and cost analysis"/ 3. economics, dental/ 4. exp "economics, hospital"/ 5. economics, medical/ 6. economics, nursing/ 7. economics, pharmaceutical/ 8. (economic\$ or cost or costs or costly or costing or price or prices or pricing or pharmacoeconomic\$.ti,ab. 9. (expenditure\$ not energy).ti,ab. 10. (value adj1 money).ti,ab. 11. budget\$.ti,ab.

Table 13: MEDLINE Search Filters — Best Sensitivity

Filter Name	Filter Details
	12. or/1-11 13. ((energy or oxygen) adj cost).ti,ab. 14. (metabolic adj cost).ti,ab. 15. ((energy or oxygen) adj expenditure).ti,ab. 16. or/13-15 17. 12 not 16 18. Letter.pt. 19. Editorial.pt. 20. Historical article.pt. 21. Animals/ not humans/ 22. 18 or 19 or 20 or 21 23. 17 not 22
Royle and Waugh ⁴	1. (cost* or economic* or (quality adj2 life)).ti,ab,hw. 2. Letter.pt. 3. Editorial.pt. 4. Historical article.pt. 5. Animals/ not humans/ 6. 2 or 3 or 4 or 5 7. 1 not 6
Wilczynski et al. ³	1. Cost-benefit analysis/ or costs.tw. or cost effective.tw. 2. Letter.pt. 3. Editorial.pt. 4. Historical article.pt. 5. Animals/ not humans/ 6. 2 or 3 or 4 or 5 7. 1 not 6
\$	Truncation symbol
*	Indicates a major heading when preceding a subject heading (e.g. *economics/)
*	Truncation symbol when following a term (e.g. economic*)
.ab.	Abstract
.fs.	Floating subheading
.hw.	Subject heading word
.mp.	Title, abstract and indexing
.pt.	Publication Type
.sh.	Subject heading
.ti.	Title.
tw,.	Textword
.yr.	Publication year
/	Subject heading
:	Truncation symbol
?	Indicates there can be zero or one additional characters (e.g. cost? Identifies cost and also costs)
Adj	Adjacent terms
Adj1	Terms within one word of each other
Exp	Explode (subject heading)
Or/1-3	Combine sets 1 to 3 using OR

NHS EED = National Health Service Economic Evaluation Database; NHS QIS = National Health Service Quality Improvement Scotland.

Table 14: EMBASE Search Filters — Best Sensitivity

Filter Name	Filter Details
NHS QIS ¹⁹	1. exp health economics/ 2. exp health care cost/ 3. exp quality of life/ 4. economic\$.tw. 5. (cost? or costing? or costly or costed).tw. 6. (price? or pricing?).tw. 7. (pharmacoeconomic? or (pharmaco adj economic?)).tw. 8. budget\$.tw. 9. expenditure\$.tw. 10. (value adj1 (money or monetary)).tw. 11. (fee or fees).tw. 12. "quality of life".tw. 13. qol\$.tw. 14. hrqol\$.tw. 15. "quality adjusted life year\$.tw. 16. qaly\$.tw. 17. cba.tw. 18. cea.tw. 19. cua.tw. 20. utilit\$.tw. 21. markov\$.tw. 22. monte carlo.tw. 23. (decision adj2 (tree\$ or analys\$ or model\$)).tw. 24. ((clinical or critical or patient) adj (path? or pathway?)).tw. 25. (managed adj2 (care or network?)).tw. 26. or/1-25
CADTH ¹⁵	1. *Economics/ 2. *Economics, Medical/ 3. *Economics, Pharmaceutical/ 4. exp "Costs and Cost Analysis"/ 5. exp Health Care Costs/ 6. exp decision support techniques/ 7. economic value of life.sh. 8. exp models, economic/ 9. markov chains.sh. 10. monte carlo method.sh. 11. uncertainty.sh. 12. quality of life.sh. 13. quality-adjusted life years.sh. 14. exp health economics/ 15. exp economic evaluation/ 16. exp pharmacoeconomics/ 17. exp economic aspect/ 18. quality adjusted life year/ 19. quality of life/ 20. exp "costs and cost analyses"/ 21. cost containment.sh. 22. (economic impact or economic value or pharmacoeconomics or health care cost or

Table 14: EMBASE Search Filters — Best Sensitivity

Filter Name	Filter Details
	economic factors or cost analysis Or economic analysis or cost or cost-effectiveness or cost effectiveness or costs or health care cost or cost savings or cost-benefit analysis or hospital costs or medical costs or quality-of-life).sh. 23. health resource allocation.sh. 24. (econom\$ or cost or costly or costing or costed or price or prices or pricing or priced or discount or discounts or discounted or discounting or expenditure or expenditures or budget\$ or afford\$ or pharmacoeconomic or pharmaco-economic\$).ti,ab. 25. (cost\$ adj1 (util\$ or effective\$ or efficac\$ or benefit\$ or consequence\$ or analy\$ or minimi\$ or saving\$ or breakdown or lowering or estimate\$ or variable\$ or allocation or control or illness or sharing or life or lives or affordabl\$ or instrument\$ or technolog\$ or day\$ or fee or fees or charge or charges)).ti,ab. 26. (decision adj1 (tree\$ or analy\$ or model\$)).ti,ab. 27. ((value or values or valuation) adj2 (money or monetary or life or lives or costs or cost)).ti,ab. 28. (qol or qoly or qolys or hrqol or qaly or qalys or qale or qales).ti,ab. 29. (sensitivity analys\$ or "willingness to pay" or quality-adjusted life year\$ or quality adjusted life year\$ or quality-adjusted life expectanc\$ or quality adjusted life expectanc\$).ti,ab. 30. (unit cost or unit-cost or unit-costs or unit costs or drug cost or drug costs or hospital costs or health-care costs or health care cost or medical cost or medical costs).ti,ab. 31. (decision adj1 (tree\$ or analy\$ or model\$)).ti,ab. 32. or/1-31
Royle and Waugh ⁴	1. (cost* or economic* or (quality adj2 life)).ti,ab,hw.
NHS EED ¹⁸	1. health-economics/ 2. exp economic-evaluation/ 3. exp health-care-cost/ 4. exp pharmacoeconomics/ 5. 1 or 2 or 3 or 4 6. (econom\$ or cost or costs or costly or costing or price or prices or pricing or pharmacoeconomic\$).ti,ab 7. (expenditure\$ not energy).ti,ab 8. (value adj2 money).ti,ab 9. budget\$.ti,ab 10. 6 or 7 or 8 or 9 11. 5 or 10 12. (metabolic adj cost).ti,ab 13. ((energy or oxygen) adj cost).ti,ab 14. ((energy or oxygen) adj expenditure).ti,ab 15. 12 or 13 or 14 16. 11 not 15
EMBASE G	1. 1 Cost adj effectiveness.ab. 2. 2 Cost adj effectiveness.ti. 3. 3 Life adj years.ab. 4. 4 Life adj year.ab. 5. 5 Qaly.ab.

Table 14: EMBASE Search Filters — Best Sensitivity

Filter Name	Filter Details
	6. (Cost or costs).ab. and Controlled Study/ 7. (Cost and costs).ab. 8. 1 or 2 or 3 or 4 or 5 or 6 or 7
\$	Truncation symbol
*	Indicates a major heading when preceding a subject heading (e.g. *economics/)
*	Truncation symbol when following a term (e.g. economic*)
.ab.	Abstract
.fs.	Floating subheading
.hw.	Subject heading word
.mp.	Title, abstract and indexing
.pt.	Publication Type
.sh.	Subject heading
.ti.	Title
.tw.,.	Textword
.yr.	Publication year
/	Subject heading
:	Truncation symbol
?	Indicates there can be zero or one additional characters (e.g. cost? Identifies cost and also costs)
Adj	Adjacent terms
Adj1	Terms within one word of each other
Exp	Explode (subject heading)
Or/1-3	Combine sets 1 to 3 using OR

CADTH = Canadian Agency for Drugs and Technologies in Health; NHS EED = National Health Service Economic Evaluation Database; NHS QIS = National Health Service Quality Improvement Scotland.

Table 15: MEDLINE Search Filter — Precision Maximizing

Filter Name	Filter Details
MEDLINE G	1. Economics.fs. 2. (cost-benefit analysis/ or randomized controlled trial.pt.) 3. 1 and 2 4. Letter.pt. 5. Editorial.pt. 6. Historical article.pt. 7. Animals/ not humans/ 8. 4 or 5 or 6 or 7 9. 3 not 8
/	Subject heading
.fs.	Floating subheading
.pt.	Publication Type
.ab.	Abstract
Adj	Adjacent terms

Table 16: EMBASE Search Filter — Precision Maximizing	
Filter Name	Filter Details
EMBASE H	1. Cost adj effectiveness.ti. 2. (cost adj effectiveness).ab. and (costs or cost).ab. 3. (cost and costs).ab. and cost effectiveness analysis/ 4. (Costs.ab. and controlled study/ and cost.hw.) and (effectiveness.hw. or randomized controlled trial/) 5. 1 or 2 or 3 or 4
Adj	Adjacent terms
.ab.	Abstract
/	Subject heading
.hw.	Subject heading word
.ti.	Title

5 DISCUSSION

5.1 Summary of Findings

This research has provided new performance data on published and newly developed search filters for the identification of economic evaluations in MEDLINE and EMBASE. Many publicly available filters have not been previously validated.^{15,16,18-20} There are little comparative performance data available on formally published search filters, so this research adds to the performance picture of those filters.^{1,3,4,17} Based on the findings from this analysis, long-standing filters such as those developed by NHS EED, NHS QIS, Royle and Waugh, and CADTH continue to perform with high sensitivity in MEDLINE. None of the published or newly derived filters met the project objectives of high sensitivity (defined as sensitivity of more than 0.950) with a level of precision (0.200) that would be valued by researchers. The analysis failed to identify filters with more moderate levels of sensitivity and higher levels of precision (defined as sensitivity 0.800, precision 0.500).

Although search filters proved to be discriminating in the analysis, they did not perform at the same levels when used in the Ovid interface. This suggests that the text words and indexing terms in economic evaluation records do not sufficiently discriminate economic evaluation records from the other records that deal with economic issues in health care. The title and abstract terms with high sensitivity for the identification of economic evaluations are used in several ways other than those in this context, thereby complicating the maximization of precision. Given recent trends in health care economic evaluation, terms pertaining to economic modelling and the incorporation of utility were hypothesized a priori to feature more prominently in the term analysis. These terms, however, did not emerge as being discriminating from the data analysis based on classification trees, thereby suggesting a continued reliance on terms with more sensitivity, such as “costs” and “cost” and “economic,” which are used in many contexts.

Although none of the new or published filters met the project’s desired performance levels in terms of sensitivity and precision, this analysis does show which filters performed best in finding a large gold standard set of relevant records. Tables 9 to 12 allow searchers to select filters

depending on their preference for levels of sensitivity and precision. EMBASE G, for example, has high sensitivity and a level of precision that means approximately 13 in every 100 records may be relevant. Many of the MEDLINE filters reached 0.090 precision with a sensitivity of 0.900 or greater. Searchers wishing to use concise filters (such as those by Royle and Waugh) to identify economic evaluations, rather than using lengthy filters, can also compare performance data in the tables.

The comparator records for this study were selected to ensure that the emerging filters would be robust at discriminating reports of economic evaluations from records mentioning other economic issues or methods. The use of comparator records that were not focused on economic issues might have produced different filters because the frequency of economic terms in such records is likely to be lower. This suggested that the filters developed in this project could be used as “second pass” filters (filters that could be used after economic filters with more sensitivity have been used in a prior search phase). The filters provided here may then offer a further focus in the subset of retrieved records. The use of this approach led to some improvement in the precision of the filters with the highest sensitivity. For example, running MEDLINE F on the results of the NHS QIS brief filter improved precision from 0.017 to 0.048, with a drop in sensitivity from 1.000 to 0.974. An analogous implementation on the results of NHS EED improved precision from 0.040 to 0.079 and reduced sensitivity from 0.999 to 0.973. Such gains in precision and losses in sensitivity may be less desirable, however, when search filters are used to search for economic evaluations of specific interventions and fewer records are returned for assessment. The use of a non-economic focused comparator set and the value of combined or two-stage filters remain to be explored.

This analysis shows that it remains difficult to efficiently identify economic evaluations by using the single indexing terms that are applied by database creators. The use of indexing terms that are specific for economic evaluations (such as “Cost-benefit Analysis/” in MEDLINE and “Cost effectiveness analysis/” in EMBASE) did not result in high levels of specificity and precision in this study. This may be the result of several factors, including poor reporting by authors, indexer uncertainty, and indexing lapses. As a result, researchers cannot rely on a few precise search terms to identify economic evaluations in large general biomedical databases, as is the case with randomized controlled trials.⁹ Database producers can use this information to make research evidence more accessible to researchers. It also falls to authors of papers to report economic evaluations clearly and use economic evaluation terminology as consistently as possible to help indexers and searchers reliably identify reports of economic evaluations. This research also provides evidence about the value of international databases such as NHS EED and HEED, which identify economic evaluations from a range of resources to improve access to this type of study.

5.2 Study Limitations

This study is based on a gold standard set of records that was obtained from NHS EED. There is a risk that the NHS EED search filter (and other filters developed from the NHS EED filter such as the one produced by the Scottish Intercollegiate Guidelines Network) that was used to build the database over-performs in comparison with other filters. This raises questions as to whether the database searches of CINAHL and PsycINFO and the hand-searching of 60 journals dilute

this over-performance by identifying additional studies that were indexed in MEDLINE or EMBASE, but that were not identified by the NHS EED MEDLINE or EMBASE filters. Evidence from a systematic review indicates that additional studies are likely to be identified when hand-searching for randomized controlled trials is compared with database searches alone, but it is unclear whether this also applies to economic evaluations.²¹ Without a retrospective analysis, it is difficult to judge if over-performance is an issue.

Another concern about basing the research on a gold standard derived from NHS EED is whether it provides an adequate gold standard. The NHS EED gold standard has benefited from the use of hand-searching and relative recall methods of gold standard development. The search approaches that were used to identify candidate records for inclusion in NHS EED involve searches of MEDLINE, EMBASE, CINAHL, and PsycINFO, and hand-searches of 60 journals. High levels of duplication in the searches should reduce the risk of missing relevant economic evaluations. The value of this gold standard stems from its clear inclusion definitions, wide sampling of journals, international coverage (which means that there are no language restrictions), and the extraction of records for different years. A limitation to the gold standard may be its reliance on the NHS EED definition of an economic evaluation, which may differ from the definitions that are used by other research teams. The gold standard also relies on the consistency of the NHS EED production process over time. If economic evaluations have been missed or miscategorized by NHS EED researchers, the gold standard may not reflect the true number of economic evaluations that are available to be retrieved.

Some records that were obtained from NHS EED had different publication years when checked in MEDLINE and EMBASE. In such cases, the MEDLINE or EMBASE publication year was used. This affected two records in each database. Records that had authors with multiple surnames and corporate authors (organizations) were sometimes difficult to identify. Some NHS EED gold standard records may not have been successfully identified in MEDLINE and EMBASE because the correct match was not made after several searches for author variants.

By using the search filters, records of economic evaluations that had been missed by the NHS EED identification process may have been identified. It was not possible, with the available resources, to assess the search filter results for additional studies or to compare how well each filter performed in identifying any economic evaluations missed by NHS EED.

The MEDLINE gold standard set included some non-indexed records. Most of the MEDLINE records that were analyzed, however, were indexed with MeSH and publication type headings. As a result, the search filters are likely to perform best in MEDLINE records that have been indexed, and their performance in finding in process and non-indexed records has not been optimized. This is an issue if the search filters are being used to identify current studies where indexing has not yet been assigned. It is a potential topic for further research that would be possible by using the text and abstract word variables from this gold standard.

Resource constraints meant that pragmatic and subjective cut-offs were applied to select which terms should be analyzed. Classification trees can be used to handle any number of variables, so the data could be analyzed in greater depth using terms that occurred in fewer records (below the cut-off used in this project), which might result in improved filter performance. The analysis of

phrases and co-occurring terms in the title and abstracts of records was also not possible in this project. The EMBASE filters that were developed during this analysis seem to indicate that the exploration of term co-occurrence and possibly term frequency in records may enhance search filters by improving precision. This remains a topic for future study.

The exclusion strategy was efficient in MEDLINE (by reducing the number of irrelevant records while retaining sensitivity) but unhelpful in EMBASE where improvements in precision resulted in the loss of relevant records. Exclusion approaches tend to be based on the limits that are provided by database indexers to remove specific publication types that are unlikely to contain reports of research, such as editorials. Such an approach works best with indexed records and does not assist with excluding non-indexed records. However, the exclusion of non-indexed records based on searching titles and abstracts for terms such as “letter” are likely to increase the risk of excluding relevant records where those terms are used in other contexts. For example, the cost-effectiveness of sending referral letters by email rather than mail might be the topic of a relevant paper. That record, however, would be excluded from the result set if the exclusion strategy was used to exclude records that contained the word “letter” in the title or abstract. This type of issue makes exclusion strategies challenging to develop. The exclusion strategy that is used in this research is typical of the limiting strategies used in searches to identify studies for HTAs, but there are no standards for these types of strategies. Other exclusion strategies might perform differently and this could be tested.

5.3 Generalizability and Recommendations

The filters have been tested in Ovid MEDLINE and EMBASE, but require more extensive validation through one or more of the following tests:

- After the completion of a review of economic evaluations, a “relative recall” exercise could be undertaken. This would involve the identification of the evaluations that were reviewed in the research, the identification of those study records in MEDLINE and EMBASE, and then the testing of how well the search filters in this report perform in combination with the subject searches that were used to identify the studies originally.^{11,22}
- Many journals are hand-searched to identify economic evaluations (this occurs with the creation of the NHS EED, HEED, and other databases). The performance of the filters in finding the records of economic evaluations (in MEDLINE and EMBASE) that are identified by such hand-searches could be tested.
- As further years of NHS EED abstracts are completed, those records can be identified in MEDLINE and EMBASE, and filter performance assessed.
- Additional gold standards might be revealed by searches of other economic evaluation databases such as the HEED (<http://www3.interscience.wiley.com/cgi-bin/mrwhome/114130635/HOME>) or the Paediatric Economic Database Evaluation (<http://pede.bioinfo.sickkids.on.ca/pede/index.jsp>). The records could be created as gold standard sets in MEDLINE and EMBASE and used to test search filter performance. Those databases’ definitions of economic evaluations would need to be assessed to ensure that they are similar to those used in this project.

5.4 Conclusions

This research has shown that searchers can choose to use several sensitive filters to help them to identify economic evaluations in MEDLINE (Table 13). Those filters with the highest sensitivity have low precision: NHS QIS (using the brief filter rather than full filter to enhance precision), NHS EED, and Royle and Waugh.^{4,18,19} Searchers who are choosing between these filters may make their decision based on factors such as highest precision (choosing NHS EED) or low numbers of search lines in the filter (choosing Royle and Waugh). Increased precision can be achieved after the lowering of sensitivity. The Wilczynski best optimization of sensitivity and precision filter offers more than 0.900 sensitivity with a precision of more than 0.090 in a one line strategy.³ We recommend that all the filters for MEDLINE be performed with exclusion search lines that remove specific publication types and animal studies. This gives enhanced precision without losing gold standard records.

Searchers who need high sensitivity filters to identify economic evaluations in EMBASE can choose between the filters provided by NHS QIS, CADTH, Royle and Waugh, and NHS EED (Table 14).^{4,18,19} These filters have low precision. A precision of more than 0.100 with lower sensitivity (more than 0.900) may be achieved with the brief EMBASE G filter that was developed during this research. We recommend that the filters for EMBASE be performed without an exclusion strategy that removes specific publication types and animal studies to avoid losing relevant economic evaluations.

For searchers who are conducting scoping studies or rapid reviews and who wish to gain a rapid, but necessarily incomplete set of the available economic evaluations, this research has identified filters that offer higher levels of precision. The new MEDLINE G filter that was produced during this research achieved more than 0.250 precision and the new EMBASE H filter achieved more than 0.260 precision. Searchers who are using these filters should be aware that their sensitivity is approximately 0.700 (Tables 15 and 16).

Searchers now have new comparative information on the performance of various filters. The findings of this project are useful for HTA organizations in terms of planning the workload for technology assessments. The filter performance data suggest that, if using high sensitivity filters, at best four or five records out of every 100 records scanned will be relevant. These data are also useful for database producers who are populating economic evaluation databases because the filters with the highest sensitivity are indicated. Data from this project continue to encourage researchers to report their methods as clearly as possible and may assist the producers of biomedical databases because the data show how indexing terms are being applied.

6 REFERENCES

1. Sassi F, Archard L, McDaid D. Searching literature databases for health care economic evaluations: how systematic can we afford to be? *Med Care* 2002;40(5):387-94.
2. Jenkins M. Evaluation of methodological search filters: a review. *Health Info Libr J* 2004;21(3):148-63.
3. Wilczynski NL, Haynes RB, Lavis JN, Ramkissoonsingh R, Arnold-Oatley AE, HSR Hedges team. Optimal search strategies for detecting health services research studies in MEDLINE. *CMAJ* 2004;171(10):1179-85.
4. Royle P, Waugh N. Literature searching for clinical and cost-effectiveness studies used in health technology assessment reports carried out for the National Institute for Clinical Excellence appraisal system. *Health Technol Assess* 2003;7(34):iii, ix-51. Available: <http://www.hta.ac.uk/fullmono/mon734.pdf> (accessed 2009 Feb 15).
5. Bak G, Mierzwinski-Urban M, Fitzsimmons H, Morrison A, Maden-Jenkins M. A pragmatic critical appraisal instrument for search filters: introducing the CADTH CAI [Epub ahead of print]. *Health Info Libr J* 2008.
6. Glanville JM, Bayliss S, Booth A, Dundar Y, Fernandes H, Fleeman ND, et al. So many filters, so little time: the development of a search filter appraisal checklist. *J Med Libr Assoc* 2008;96(4):356-61. Available: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2568852> (accessed 2009 Mar 10).
7. Wilczynski NL, Morgan D, Haynes RB, Hedges Team. An overview of the design and methods for retrieving high-quality studies for clinical care. *BMC Med Inform Decis Mak* 2005;5:20. Available: <http://www.biomedcentral.com/content/pdf/1472-6947-5-20.pdf> (accessed 2009 Mar 10).
8. White VJ, Glanville JM, Lefebvre C, Sheldon TA. A statistical approach to designing search filters to find systematic reviews: objectivity enhances accuracy. *J Inf Sci* 2001;27(6):357-70.
9. Glanville JM, Lefebvre C, Miles JN, Camosso-Stefinovic J. How to identify randomized controlled trials in MEDLINE: ten years on. *J Med Libr Assoc* 2006;94(2):130-6.
10. Breiman L, Friedman JH, Olshen RA, Stone CJ. *Classification and regression trees*. Pacific Grove (CA): Wadsworth; 1984.
11. Sampson M, Zhang L, Morrison A, Barrowman NJ, Clifford TJ, Platt RW, et al. An alternative to the hand searching gold standard: validating methodological search filters using relative recall. *BMC Med Res Methodol* 2006;6(33). Available: <http://www.biomedcentral.com/1471-2288/6/33> (accessed 2009 Feb 15).
12. *EndNote XI.1.01*. Carlsbad (CA): Thomson Reuters; 2009.
13. Centre for Reviews and Dissemination. *NHS Economic Evaluation Database handbook*. York (UK): Centre for Reviews and Dissemination; 2007.
14. *S-PLUS 6 for Windows guide to statistics. Vol 2*. Seattle: Insightful Corporation; 2001.
15. Ho C, Li H, Noorani H, Cimon K, Campbell K, Tang A, et al. Appendix 2: literature search strategy for cost-effectiveness studies. In: *Implantable cardiac defibrillators for primary prevention of sudden cardiac death in high risk patients: a meta-analysis of clinical efficacy, and a review of cost-effectiveness and psychosocial issues* [Technology report no 81]. Ottawa: Canadian Agency for Drugs and Technologies in Health (CADTH); 2007. Available: http://www.cadth.ca/media/pdf/332_ICD_tr_appendices_e.pdf (accessed 2009 Mar 5).

16. Woodworth K. *Grady EBM filters for Ovid*. Atlanta (GA): Emory University; 2007. Available: <http://www.emory.edu/WHSCS/grady/inetgrp/ovidfilter.doc> (accessed 2009 Feb 15).
17. McKinlay RJ, Wilczynski NL, Haynes RB, Hedges Team. Optimal search strategies for detecting cost and economic studies in EMBASE. *BMC Health Serv Res* 2006;6:67. Available: <http://www.biomedcentral.com/1472-6963/6/67> (accessed 2009 Mar 10).
18. Centre for Reviews and Dissemination. *How are studies identified for NHS EED* [Internet]. York (UK): Centre for Reviews and Dissemination; 2009. Available: <http://www.crd.york.ac.uk/crdweb/html/help.htm> (accessed 2009 Feb 15).
19. Macpherson K, Boynton J. *Economics filter* [Internet]. York (UK): InterTASC Information Specialists' Sub-Group; 2009. Available: <http://www.york.ac.uk/inst/crd/intertasc/econ3.htm> (accessed 2009 Feb 15).
20. Scottish Intercollegiate Guidelines Network. *Search filters* [Internet]. Edinburgh: Scottish Intercollegiate Guidelines Network; 2009. Available: <http://www.sign.ac.uk/methodology/filters.html#econ> (accessed 2009 Feb 15).
21. Hopewell S, Clarke MJ, Lefebvre C, Scherer RW. Handsearching versus electronic searching to identify reports of randomized trials [Cochrane review]. *Cochrane Database Syst Rev* 2007;(2):MR000001.
22. Ritchie G, Glanville J, Lefebvre C. Do published search filters to identify diagnostic test accuracy studies perform adequately? *Health Info Libr J* 2007;24(3):188-92.

APPENDIX 1: SEARCH STRATEGIES THAT ARE USED TO IDENTIFY COMPARATOR RECORDS

NHS EED filter that was used to identify comparator records in MEDLINE for records with a publication date of 2003.

Ovid MEDLINE(R) <1950 to November Week 3 2008>	
1	ec.fs. (251600)
2	exp "costs and cost analysis"/ (142628)
3	economic value of life/ (5091)
4	economics dental/ (1800)
5	exp economics hospital/ (15962)
6	economics medical/ (7383)
7	economics nursing/ (3861)
8	economics pharmaceutical/ (2012)
9	or/1-8 (311207)
10	(econom* or cost or costs or costly or costing or price or prices or pricing or pharmaco-economic*).ti,ab. (285130)
11	(expenditure* not energy).ti,ab. (12011)
12	(value adj3 money).ti,ab. (572)
13	budget*.ti,ab. (12042)
14	or/10-13 (298034)
15	9 or 14 (491960)
16	letter.pt. (654713)
17	editorial.pt. (234908)
18	historical article.pt. (258893)
19	or/16-18 (1137542)
20	15 not 19 (451790)
21	animals/ not humans/ (3311256)
22	20 not 21 (430615)
23	(metabolic adj3 cost).ti,ab. (599)
24	((energy or oxygen) adj3 cost).ti,ab. (2287)
25	22 not (23 or 24) (428443)
26	25 and "2003".yr. (20936)
27	from 26 keep 19,27,32,43,54,64,83,92,123,132,146-147,164,186-187,221,225,235-236,257,275,293-
There follow hundreds of numbers representing the record numbers selected according to the random number generator.	

NHS EED filter that was used to identify comparator records in EMBASE for records with a publication date of 2000.

EMBASE <1980 to 2009 Week 02>	
1	health-economics/ (10416)
2	exp economic-evaluation/ (101519)
3	exp health-care-cost/ (104137)
4	exp pharmacoeconomics/ (55277)
5	(econom\$ or cost or costs or costly or costing or price or prices or pricing or pharmacoeconomic\$).ti,ab. (228467)
6	(expenditure\$ not energy).ti,ab. (9649)
7	(value adj2 money).ti,ab. (443)
8	(value adj2 money).ti,ab. (443)
9	budget\$.ti,ab. (8701)
10	or/1-9 (338602)
11	letter.pt. (434885)
12	editorial.pt. (222426)
13	note.pt. (241715)
14	or/11-13 (899026)
15	10 not 14 (293089)
16	(metabolic adj cost).ti,ab. (383)
17	((energy or oxygen) adj cost).ti,ab. (1694)
18	((energy or oxygen) adj3 expenditure).ti,ab. (10160)
19	or/16-18 (11744)
20	15 not 19 (290538)
21	exp animal/ (18265)
22	exp animal-experiment/ (1274169)
23	nonhuman/ (3165159)
24	(rat or rats or mouse or mice or hamster or hamsters or animal or animals or dog or dogs or cat or cats or bovine or sheep).ti,ab,sh. (2036149)
25	or/21-24 (3582783)
26	exp human/ (6419235)
27	exp human-experiment/ (253962)
28	or/26-27 (6420101)
29	25 not (25 and 28) (2940809)
30	20 not 29 (267240)
31	30 and "2000".yr. (13300)
32	from 31 keep 7,15-16,29,33,41,52,54,68-
There follow hundreds of numbers representing the record numbers selected according to the random number generator	

APPENDIX 2: STOP WORD LIST

Common words that were removed from the term occurrence analysis.

a	been	end	good	just
about	before	ended	goods	
above	began	ending	got	k
across	behind	ends	great	keep
after	being	enough	greater	keeps
again	beings	even	greatest	kind
against	best	evenly	group	knew
all	better	ever	grouped	know
almost	between	every	grouping	known
alone	big	everybody	groups	knows
along	both	everyone		
already	but	everything	h	l
also	by	everywhere	had	large
although			has	largely
always	c	f	have	last
among	came	face	having	later
an	can	faces	he	latest
and	cannot	fact	her	least
another	case	facts	here	let
any	cases	far	herself	lets
anybody	certain	felt	high	like
anyone	certainly	few	high	likely
anything	clear	find	high	long
anywhere	clearly	finds	higher	longer
are	come	first	highest	longest
area	could	for	him	
areas		four	himself	m
around	d	from	his	made
as	did	full	how	make
ask	differ	fully	however	making
asked	different	further		man
asking	differently	furthered	i	many
asks	do	furthering	if	may
at	does	furtherers	important	me
away	done		in	member
	down	g	interest	members
b	down	gave	interested	men
back	downed	general	interesting	might
backed	downing	generally	interests	more
backing	downs	get	into	most
backs	during	gets	is	mostly
be		give	it	mr
became	e	given	its	mrs
because	each	gives	itself	much
become	early	go		must
becomes	either	going	j	my

myself	p	should	together	worked
n	part	show	too	working
necessary	parted	showed	took	works
need	parting	showing	toward	would
needed	parts	shows	turn	
needing	per	side	turned	x
needs	perhaps	sides	turning	y
never	place	since	turns	
new	places	small	two	year
new	point	smaller		years
newer	pointed	smallest	u	yet
newest	pointing	so	under	you
next	points	some	until	young
no	possible	somebody	up	younger
nobody	present	someone	upon	youngest
non	presented	something	us	your
noone	presenting	somewhere	use	yours
not	presents	state	used	
nothing	problem	states	uses	z
now	problems	still		
nowhere	put	still	v	
number	puts	such	very	
numbers		sure		
	q		w	
o	quite	t	want	
of		take	wanted	
off	r	taken	wanting	
often	rather	than	wants	
old	really	that	was	
older	right	the	way	
oldest	right	their	ways	
on	room	them	we	
once	rooms	then	well	
one		there	wells	
only	s	therefore	went	
open	said	these	were	
opened	same	they	what	
opening	saw	thing	when	
opens	say	things	where	
or	says	think	whether	
order	second	thinks	which	
ordered	seconds	this	while	
ordering	see	those	who	
orders	seem	though	whole	
other	seemed	thought	whose	
others	seeming	thoughts	why	
our	seems	three	will	
out	sees	through	with	
over	several	thus	within	
	shall	to	without	
	she	today	work	

APPENDIX 3: CANDIDATE SEARCH TERMS

MEDLINE

The following tables show the words that were used in the classification tree analysis. The terms were selected by looking at the frequency of their occurrence in the records in the MEDLINE gold standard 50% sample. These tables show the variables remaining after frequency cut offs were applied. The tables exclude commonly occurring words such as “the,” “and,” and “for” (Appendix 2).

Table A1: Occurrence of Publication Type terms in 50% random sample of MEDLINE gold standard records

Publication Type term	Number of records with term (total=974)	Percentage of 50% gold standard records (%)
Research Support, Non-U.S. Govt	477	48.97
Comparative Study	352	36.14
Randomized Controlled Trial	258	26.49
Clinical Trial	187	19.20
Multicenter Study	118	12.11
Research Support, U.S. Govt, P.H.S.	63	6.47
Evaluation Studies	58	5.95
Journal Article	57	5.85
Review	40	4.11
Research Support, U.S. Govt, Non-P.H.S.	32	3.29
English Abstract	21	2.16
Controlled Clinical Trial	20	2.05
Meta-Analysis	18	1.85
Clinical Trial, Phase III	7	0.72
Case Reports	3	0.31
Validation Studies	3	0.31
Comment	2	0.21
Clinical Trial, Phase II	2	0.21

Table A2: Occurrence of terms in record title in 50% random sample of MEDLINE gold standard records

Terms in title	Number of records with term (total=974)	Percentage of 50% gold standard sample (%)
Cost-effectiveness	330	33.88
Analysis	217	22.28
Patients	156	16.02
Treatment	149	15.30
Comment	139	14.27
Cost	135	13.86
Economic	123	12.63
Trial	97	9.96
Versus	92	9.45
Clinical	89	9.14
Cancer	78	8.01
Screening	77	7.91
Randomized	68	6.98
Costs	63	6.47
Impact	47	4.83
Disease	44	4.52
Surgery	42	4.31
Primary	38	3.90
Management	37	3.80
Health	37	3.80
Results	35	3.59
Prevention	34	3.49
Program	34	3.49
Chronic	34	3.49
Model	33	3.39
Breast	32	3.29
Randomised	32	3.29
Compared	32	3.29
Strategies	31	3.18
Hepatitis	30	3.08
Cost-utility	28	2.87
Prospective	26	2.67
Hospital	26	2.67
Vaccination	26	2.67
Prophylaxis	25	2.57
Cost-effective	25	2.57
Blood	24	2.46
Using	21	2.16
Efficacy	21	2.16

Terms in title	Number of records with term (total=974)	Percentage of 50% gold standard sample (%)
Pain	20	2.05
Comparing	20	2.05
Pharmacoeconomic *	18	1.85
Cost-benefit *	16	1.64
Benefits *	14	1.44

*These terms were below the cut-off levels chosen but were tested to gain data on their performance in case of later queries from searchers and researchers who might query their effect.

Table A3: Occurrence of terms in record abstract in 50% random sample of MEDLINE gold standard records

Terms in abstract	Number of records with term (total=974)	Percentage of 50% gold standard sample (%)
Results	801	82.24
Cost	763	78.34
Costs	700	71.87
Patients	672	68.99
Methods	544	55.85
Study	519	53.29
Compared	514	52.77
Cost-effectiveness	479	49.18
Analysis	461	47.33
Treatment	456	46.82
Conclusions	423	43.43
Using	389	39.94
Data	377	38.71
Patient	366	37.58
Clinical	364	37.37
Care	349	35.83
Objective	345	35.42
Health	343	35.22
Model	338	34.70
Years	326	33.47
Cost-effective	324	33.26
Conclusion	307	31.52
Background	302	31.01
Group	301	30.90
Total	280	28.75
Life	276	28.34
Outcomes	267	27.41
Hospital	262	26.90
Incremental	261	26.80
Outcome	259	26.59
Effectiveness	256	26.28

Terms in abstract	Number of records with term (total=974)	Percentage of 50% gold standard sample (%)
Economic	247	25.36
Therapy	247	25.36
Sensitivity	245	25.15
Significant	244	25.05
Design	242	24.85
Effective	241	24.74
Trial	241	24.74
Less	236	24.22
Associated	233	23.92
Dollars	232	23.82
Year	229	23.51
Groups	224	23.00
Time	217	22.28
Based	215	22.07
Medical	210	21.56
Randomized	205	21.05
Higher	204	20.94
Ratio	204	20.94
Respectively	204	20.94
Significantly	201	20.64
Performed	200	20.53
Months	199	20.43
Disease	195	20.02
Lower	194	19.92
Mean	193	19.82
Rate	193	19.82
Perspective	190	19.51
Included	189	19.40
Setting	188	19.30
Days	184	18.89
Rates	184	18.89
Additional	183	18.79
Estimated	182	18.69
Compare	179	18.38
Quality-adjusted	174	17.86
Measures	172	17.66
Number	172	17.66
Risk	172	17.66
Strategy	172	17.66
Analyses	170	17.45
Decision	168	17.25
Intervention	165	16.94
Savings	158	16.22
Average	157	16.12
Difference	154	15.81
Reduced	154	15.81

Terms in abstract	Number of records with term (total=974)	Percentage of 50% gold standard sample (%)
Treated	153	15.71
Primary	151	15.50
Evaluate	150	15.40
Quality	150	15.40
Reduction	149	15.30
Vs	149	15.30
Direct	147	15.09
Calculated	146	14.99
High	146	14.99
Benefits	144	14.78
Main	144	14.78
Versus	144	14.78
Efficacy	143	14.68
Gained	143	14.68
Similar	142	14.58
Strategies	141	14.48
Period	139	14.27
Follow-up	138	14.17
Population	138	14.17
Overall	136	13.96
Management	134	13.76
Qaly	134	13.76
Case	132	13.55
Interventions	132	13.55
Published	132	13.55
Differences	131	13.45
Estimates	128	13.14
Increased	128	13.14
Conducted	127	13.04
Determine	127	13.04
Different	126	12.94
Assess	124	12.73
Cohort	123	12.63
Impact	123	12.63
Increase	123	12.63
Literature	123	12.63
Evaluated	121	12.42
Estimate	119	12.22
Benefit	118	12.11
Cases	118	12.11
Control	118	12.11
Standard	118	12.11
Complications	117	12.01
Including	116	11.91
Obtained	116	11.91
Considered	114	11.70

Terms in abstract	Number of records with term (total=974)	Percentage of 50% gold standard sample (%)
Controlled	114	11.70
Evaluation	113	11.60
Drug	112	11.50
Mortality	112	11.50
Effects	111	11.40
Stay	111	11.40
Studies	111	11.40
Survival	111	11.40
Objectives	110	11.29
Showed	109	11.19
Greater	108	11.09
Prospective	107	10.99
Developed	106	10.88
Incidence	106	10.88
Range	106	10.88
Resource	105	10.78
System	105	10.78
Screening	104	10.68
Cancer	102	10.47
Baseline	101	10.37
Initial	101	10.37
Relative	101	10.37
Surgery	101	10.37
Found	100	10.27
Current	99	10.16
Markov	99	10.16
Ratios	90	9.24
Saved	90	9.24
Life-year	87	8.93
Qalys	87	8.93
Sensitive	84	8.62
Societal	82	8.42
Assumptions	77	7.91
Costly	76	7.80
Values	75	7.70
Value	70	7.19
Saving	67	6.88
Million	66	6.78
Life-years	64	6.57
Pounds	62	6.37
Resources	62	6.37
Lifetime	60	6.16
Discounted	59	6.06
Adjusted	54	5.54
Small	54	5.54
Utility	50	5.13

Terms in abstract	Number of records with term (total=974)	Percentage of 50% gold standard sample (%)
Modeling *	24	2.46

*This term was below the cut-off levels chosen but was tested to gain data on its performance in case of later queries from searchers and researchers who might query its effect.

Table A4: Occurrence of Medical Subject Headings (MeSH) and subheadings in 50% random sample of MEDLINE gold standard records

Subject headings and subheadings	Number of records with term (total=974)	Percentage of 50% gold standard sample (%)
Humans	955	98.05
ec [Economics]	856	87.88
Cost-Benefit Analysis	676	69.40
Female	591	60.68
Male	478	49.08
Middle Aged	441	45.28
Adult	369	37.89
Aged	350	35.93
tu [Therapeutic Use]	310	31.83
dt [Drug Therapy]	258	26.49
Treatment Outcome	234	24.02
pc [Prevention & Control]	223	22.90
mt [Methods]	213	21.87
Health Care Costs	192	19.71
th [Therapy]	149	15.30
Quality-Adjusted Life Years	148	15.20
ep [Epidemiology]	145	14.89
sn [Statistics & Numerical Data]	136	13.96
di [Diagnosis]	126	12.94
ad [Administration & Dosage]	126	12.94
Aged, 80 and over	122	12.53
Adolescent	115	11.81
Prospective Studies	113	11.60
ae [Adverse Effects]	112	11.50
Sensitivity and Specificity	109	11.19
su [Surgery]	105	10.78
Retrospective Studies	103	10.57
Markov Chains	89	9.14
Length of Stay	85	8.73
Decision Support Techniques	83	8.52
Follow-Up Studies	81	8.32
Costs and Cost Analysis	77	7.91
Drug Costs	76	7.80
Quality of Life	75	7.70

Subject headings and subheadings	Number of records with term (total=974)	Percentage of 50% gold standard sample (%)
Mass Screening	74	7.60
mo [Mortality]	73	7.49
Child	70	7.19
Cohort Studies	66	6.78
Hospital Costs	62	6.37
Models, Economic	62	6.37
Time Factors	59	6.06
ut [Utilization]	59	6.06
et [Etiology]	57	5.85
Risk Factors	55	5.65
st [Standards]	55	5.65
Randomized Controlled Trials as Topic	55	5.65
Child, Preschool	55	5.65
co [Complications]	51	5.24
Infant	51	5.24
Hospitalization	51	5.24
Outcome Assessment Health Care	49	5.03
Decision Trees *	48	4.92
Cost of Illness *	47	4.82
Cost Savings *	31	3.18
Models, Theoretical *	23	2.36
Monte Carlo Method *	21	2.16
Value of Life *	10	1.03
Cost Control *	10	1.03

*These terms were below the cut-off levels chosen but were tested to gain data on their performance in case of later queries from searchers and researchers who might query their effect.

Table A5: Occurrence of single words from multi-word Medical Subject Headings (MeSH) and subheadings in 50% random sample of MEDLINE gold standard records

Single words in multi-word Medical Subject Headings and subheadings	Number of records with term (total=974)	Percentage of 50% gold standard sample (%)
Analysis	749	76.90
Costs	364	37.37
Health	346	35.52
Drug	343	35.22
Studies	333	34.19
Care	322	33.06
Outcome	293	30.08
Treatment	242	24.85
Prevention	240	24.64
Agents	233	23.92

Single words in multi-word Medical Subject Headings and subheadings	Number of records with term (total=974)	Percentage of 50% gold standard sample (%)
Life	223	22.90
Control]	223	22.90
Statistics	165	16.94
Administration	162	16.63
Cost	158	16.22
Numerical	154	15.81
Therapy	149	15.30
Years	148	15.19
Quality-adjusted	148	15.19
Factors	142	14.58
Decision	141	14.48
Models	140	14.37
Adverse	123	12.63
Disease	114	11.70
Prospective	114	11.70
Sensitivity	111	11.40
Specificity	109	11.19
Patient	106	10.88
Retrospective	103	10.57
Assessment	100	10.27
Techniques	98	10.06
Survival	92	9.45
Services	91	9.34
Risk	90	9.24
Hospital	89	9.14
Quality	89	9.14
Screening	88	9.03
Support	88	9.03
Mass	85	8.73
Economics	84	8.62
Infections	81	8.32
Follow-Up	81	8.32
Diseases	67	6.88
Cohort	66	6.78
Postoperative	66	6.78
Method	66	6.78
Blood	65	6.67
Complications	65	6.67
Illness	63	6.47
Time	62	6.37
Economic	62	6.37
Trials	59	6.06
Controlled	55	5.65

Single words in multi-word Medical Subject Headings and subheadings	Number of records with term (total=974)	Percentage of 50% gold standard sample (%)
Randomized	55	5.65
Preschool	55	5.65
Combination	54	5.54
Surgical	52	5.34
Chronic	52	5.34
Hospitalization	51	5.24
Evaluation	50	5.13
Antineoplastic	49	5.03

Table A6: Occurrence of Medical Subject Headings (MeSH) coordinated with subheadings that appear in more than 5% of 50% random sample of MEDLINE gold standard records

Medical Subject Headings coordinated with subheadings	Number of records with term (total=974)	Percentage of 50% gold standard sample (%)
Mass Screening/ec [Economics]	60	6.16

EMBASE

The following tables show the terms that were used in the classification tree analysis. The terms were selected by looking at the frequency of their occurrence in the records in the EMBASE gold standard 50% sample. These tables show the variables remaining after frequency cut offs were applied. The tables exclude commonly occurring words such as “the,” “and,” and “for” (Appendix 2).

Table A7: Occurrence of terms in title in 50% random sample of EMBASE gold standard records.

Title word term	Number of records with term (total=927)	Percentage of 50% gold standard sample (%)
Cost-effectiveness	318	34.30
Analysis	208	22.44
Patients	151	16.29
Treatment	134	14.46
Cost	133	14.35
Economic	109	11.76
Trial	97	10.46
Evaluation	94	10.14
Effectiveness	90	9.71
Versus	88	9.49
Care	82	8.85
Study	73	7.87

Title word term	Number of records with term (total=927)	Percentage of 50% gold standard sample (%)
Clinical	72	7.77
Cancer	70	7.55
Screening	67	7.23
Therapy	65	7.01
Randomized	62	6.69
Costs	61	6.58
Disease	56	6.04
Controlled	49	5.29
Management	48	5.18
Comparison	47	5.07
Chronic	42	4.53
Impact	42	4.53
Surgery	41	4.42
Primary	40	4.31
Prevention	38	4.10
Outcomes	37	3.99
Strategies	35	3.78
Randomised	34	3.67
Acute	33	3.56
Coronary	31	3.34
Health	31	3.34
Breast	30	3.24
Intervention	30	3.24
Model	28	3.02
Results	28	3.02
Women	28	3.02
Cost-utility	27	2.91
Using	27	2.91
Prospective	26	2.80
Based	25	2.70
Compared	25	2.70
Children	24	2.59
Hepatitis	24	2.59
Program	24	2.59
Prophylaxis	24	2.59
Vs	24	2.59
Diabetes	23	2.48
Early	22	2.37
Hospital	22	2.37
Decision	21	2.27
Cost-effective	20	2.16
Tuberculosis	20	2.16
Bypass	19	2.05

Title word term	Number of records with term (total=927)	Percentage of 50% gold standard sample (%)
Comparing	19	2.05
Diagnosis	19	2.05
Heart	19	2.05
Laparoscopic	19	2.05
Pain	19	2.05
Diagnostic	17	1.83
Pharmacoeconomic	16	1.73
Cost-benefit	15	1.62
Utility	12	1.29

Table A8: Occurrence of terms in abstract in 50% random sample of EMBASE gold standard records.

Term in abstract	Number of records with term (total=927)	Percentage of 50% gold standard sample (%)
Results	762	82.20
Cost	709	76.48
Costs	677	73.03
Patients	637	68.72
Methods	542	58.47
Compared	494	53.29
Study	487	52.54
Cost-effectiveness	454	48.98
Treatment	441	47.57
Conclusions	436	47.03
Analysis	425	45.85
Data	391	42.18
Using	365	39.37
Patient	335	36.14
Clinical	328	35.38
Objective	327	35.28
Years	320	34.52
Care	319	34.41
Cost-effective	317	34.20
Health	317	34.20
Model	316	34.09
Group	296	31.93
Background	292	31.50
Total	283	30.53
Conclusion	281	30.31
Life	277	29.88
Effectiveness	254	27.40
Incremental	253	27.29
Hospital	251	27.08

Term in abstract	Number of records with term (total=927)	Percentage of 50% gold standard sample (%)
Outcomes	240	25.89
Trial	236	25.46
Effective	232	25.03
Sensitivity	230	24.81
Outcome	228	24.60
Associated	225	24.27
Therapy	225	24.27
Significant	219	23.62
Based	216	23.30
Groups	216	23.30
Economic	214	23.09
Significantly	212	22.87
Design	211	22.76
Time	207	22.33
Mean	206	22.22
Disease	202	21.79
Performed	200	21.57
Medical	197	21.25
Lower	194	20.93
Risk	192	20.71
Respectively	191	20.60
Estimated	186	20.06
Randomized	186	20.06
Ratio	185	19.96
Rate	179	19.31
Strategy	179	19.31
Rates	177	19.09
Months	175	18.88
Additional	172	18.55
Days	171	18.45
Quality-adjusted	167	18.02
Perspective	163	17.58
Included	162	17.48
Compare	160	17.26
Intervention	160	17.26
Setting	156	16.83
Analyses	155	16.72
Decision	155	16.72
Period	155	16.72
Quality	155	16.72
Strategies	155	16.72
Average	152	16.40
Reduced	152	16.40
Increased	147	15.86
Savings	147	15.86

Term in abstract	Number of records with term (total=927)	Percentage of 50% gold standard sample (%)
Evaluate	146	15.75
Primary	146	15.75
Versus	146	15.75
Gained	144	15.53
Similar	144	15.53
Vs	142	15.32
Measures	141	15.21
Population	141	15.21
Reduction	139	14.99
Treated	135	14.56
Benefits	134	14.46
Published	133	14.35
Efficacy	132	14.24
Management	131	14.13
Qaly	131	14.13
Difference	130	14.02
Assess	129	13.92
Determine	129	13.92
Estimate	128	13.81
Main	126	13.59
Case	125	13.48
Control	124	13.38
Direct	124	13.38
Impact	124	13.38
Follow-up	121	13.05
Different	120	12.94
Evaluated	118	12.73
Overall	118	12.73
Evaluation	117	12.62
Differences	116	12.51
Conducted	115	12.41
Controlled	114	12.30
Obtained	114	12.30
Cohort	113	12.19
Ratios	113	12.19
Considered	112	12.08
Cases	111	11.97
Interventions	111	11.97
Markov	110	11.87
Incidence	108	11.65
Stay	108	11.65
Complications	107	11.54
Estimates	107	11.54
Calculated	106	11.43
Mortality	106	11.43

Term in abstract	Number of records with term (total=927)	Percentage of 50% gold standard sample (%)
Reserved	106	11.43
Women	106	11.43
Drug	105	11.33
Literature	105	11.33
Surgery	105	11.33
Increase	104	11.22
Rights	102	11.00
Benefit	101	10.90
Cancer	101	10.90
Objectives	101	10.90
Test	101	10.90
Including	100	10.79
Resulted	100	10.79
Followed	99	10.68
Screening	98	10.57
Developed	97	10.46
Purpose	97	10.46
Relative	97	10.46
Survival	97	10.46
Standard	96	10.36
Assessed	94	10.14
Improved	93	10.03
Studies	93	10.03
System	93	10.03
Available	92	9.92
Range	92	9.92
Effects	91	9.82
Expensive	91	9.82
Initial	91	9.82
Prospective	91	9.82
Equal	90	9.71
Method	90	9.71
Low	89	9.60
Following	88	9.49
Received	88	9.49
Society	88	9.49
Acute	87	9.39
American	87	9.39
Baseline	87	9.39
Length	87	9.39
Alternative	86	9.28
Life-year	86	9.28
Reduce	86	9.28
Resource	85	9.17
Long-term	84	9.06

Term in abstract	Number of records with term (total=927)	Percentage of 50% gold standard sample (%)
National	84	9.06
Current	83	8.95
Ci	81	8.74
Day	81	8.74
Expected	81	8.74
Program	81	8.74
Qalys	80	8.63
Testing	80	8.63
Trials	80	8.63
Underwent	80	8.63
Values	80	8.63
Aim	79	8.52
Found	79	8.52
Identified	79	8.52
Randomised	79	8.52
Comparing	78	8.41
Death	78	8.41
Infection	78	8.41
Interval	78	8.41
Result	78	8.41
Conventional	77	8.31
Practice	76	8.20
Sensitive	76	8.20
Societal	76	8.20
Potential	75	8.09
Randomly	75	8.09
Children	74	7.98
Events	74	7.98
Saved	74	7.98
Surgical	74	7.98
Comparison	73	7.87
Unit	73	7.87
Confidence	72	7.77
Costly	72	7.77
Prevention	72	7.77
Probability	72	7.77
Routine	72	7.77
Statistically	72	7.77
Measure	71	7.66
Required	71	7.66
Review	71	7.66
Service	71	7.66
Chronic	70	7.55
Collected	70	7.55
Assigned	69	7.44

Term in abstract	Number of records with term (total=927)	Percentage of 50% gold standard sample (%)
Expectancy	69	7.44
Measured	69	7.44
Mg	69	7.44
Saving	69	7.44
Decreased	68	7.34
Diagnostic	68	7.34
Fewer	68	7.34
Lifetime	68	7.34
Net	68	7.34
Participants	68	7.34
Weeks	67	7.23
Assumptions	66	7.12
Blood	66	7.12
Derived	66	7.12
Diagnosis	66	7.12
Morbidity	66	7.12
Observed	66	7.12
Positive	66	7.12
Terms	66	7.12
Undergoing	66	7.12
Effect	65	7.01
Hospitalization	65	7.01
Information	65	7.01
Procedures	65	7.01
Reported	65	7.01
Services	65	7.01
Treating	65	7.01
Annual	64	6.90
Demonstrated	64	6.90
Retrospective	64	6.90
Symptoms	64	6.90
Treatments	64	6.90
Adverse	63	6.80
Discounted	63	6.80
Findings	63	6.80
Receiving	63	6.80
Severe	62	6.69
Value	62	6.69
Aged	61	6.58
Drugs	61	6.58
Due	61	6.58
Provide	61	6.58
Evidence	60	6.47
Procedure	60	6.47
Secondary	60	6.47

Term in abstract	Number of records with term (total=927)	Percentage of 50% gold standard sample (%)
Receive	59	6.36
Sources	59	6.36
Approximately	58	6.26
Combination	58	6.26
Major	58	6.26
Addition	57	6.15
Duration	57	6.15
Highly	57	6.15
Improve	57	6.15
Research	57	6.15
Subjects	57	6.15
Coronary	56	6.04
Life-years	56	6.04
Provided	56	6.04
Shown	56	6.04
Plus	55	5.93
Resulting	55	5.93
Utility	55	5.93
Base	54	5.83
Changes	54	5.83
Intensive	54	5.83
Resources	54	5.83
Median	53	5.72
Pain	53	5.72
Safety	53	5.72
Single	53	5.72
Tests	53	5.72
Achieved	52	5.61
Healthcare	52	5.61
Prospectively	52	5.61
Reducing	52	5.61
According	51	5.50
Aims	51	5.50
Basis	51	5.50
Dollars	51	5.50
Prevalence	51	5.50
Prevented	51	5.50
Simulation	51	5.50
Daily	50	5.39
Defined	50	5.39
Option	50	5.39
Postoperative	50	5.39
Related	50	5.39
Times	50	5.39
Analyzed	49	5.29

Term in abstract	Number of records with term (total=927)	Percentage of 50% gold standard sample (%)
Decrease	49	5.29
Despite	49	5.29
Factors	49	5.29
Improvement	49	5.29
Levels	49	5.29
Million	49	5.29
Applied	48	5.18
Community	48	5.18
Consecutive	48	5.18
Limited	48	5.18
Shorter	48	5.18
Usual	48	5.18
Visits	48	5.18
Change	47	5.07
Combined	47	5.07
Dose	47	5.07
Failure	47	5.07
Gain	47	5.07
Hundred	47	5.07
Oral	47	5.07
Recommended	47	5.07

Table A9: Occurrence of EMTREE subject headings and subheadings in 50% random sample of EMBASE gold standard records.

EMTREE headings and subheadings	Number of records with term (total=927)	Percentage of 50% gold standard sample (%)
Human	883	90.66
Article	867	89.01
Cost effectiveness analysis	615	63.14
dm [disease management]	602	61.81
Priority journal	596	61.19
Major clinical study	538	55.24
Controlled study	533	54.72
Female	476	48.87
Adult	468	48.05
dt [drug therapy]	448	46.00
Male	411	42.20
pe [pharmacoeconomics]	408	41.89
Health care cost	367	37.68
Clinical trial	334	34.29
Aged	280	28.75
Drug cost	224	23.00
Randomized controlled trial	218	22.38

EMTREE headings and subheadings	Number of records with term (total=927)	Percentage of 50% gold standard sample (%)
Treatment outcome	199	20.43
Cost benefit analysis	196	20.12
di [diagnosis]	180	18.48
ct [clinical trial]	179	18.38
pc [prevention]	160	16.43
cm [drug comparison]	154	15.81
quality adjusted life year	152	15.61
co [complication]	143	14.68
drug efficacy	143	14.68
ae [adverse drug reaction]	134	13.76
Quality of life	134	13.76
si [side effect]	134	13.76
su [surgery]	131	13.45
th [therapy]	121	12.42
do [drug dose]	119	12.22
Intermethod comparison	117	12.01
Cost control	113	11.60
Follow up	111	11.40
Hospitalization	104	10.68
cb [drug combination]	103	10.57
ep [epidemiology]	96	9.86
Hospital cost	86	8.83
Sensitivity analysis	84	8.62
ec [endogenous compound]	77	7.91
Outcomes research	77	7.91
Mortality	76	7.80
Probability	74	7.60
po [oral drug administration]	73	7.49
Adolescent	70	7.19
Length of stay	69	7.08
Clinical article	68	6.98
iv [intravenous drug administration]	67	6.88
Multicenter study	67	6.88
Cost utility analysis	65	6.67
pd [pharmacology]	63	6.47
Disease severity	61	6.26
Economic evaluation	61	6.26
Cost of illness	60	6.16
Life expectancy	59	6.06
et [etiology]	58	5.95
Prospective study	58	5.95
Child	57	5.85
Cohort analysis	57	5.85
Medical decision making	56	5.75
Health care utilization	55	5.65

EMTREE headings and subheadings	Number of records with term (total=927)	Percentage of 50% gold standard sample (%)
Economic aspect	52	5.34
Outcome assessment	51	5.24
Health program	50	5.13
Double blind procedure	49	5.03

Table A10: Occurrence of single words in EMTREE subject headings and subheadings in a 50% random sample of EMBASE gold standard records.

Single words in EMTREE heading and subheadings	Number of records with term (total=927)	Percentage of 50% gold standard sample (%)
Cost	901	92.51
Human	886	90.97
Analysis	767	78.75
Study	706	72.48
Clinical	692	71.05
Disease	663	68.07
Effectiveness	617	63.35
Controlled	541	55.54
Care	533	54.72
Health	525	53.90
therapy]	507	52.05
Life	282	28.95
Quality	274	28.13
Treatment	253	25.98
Outcome	245	25.15
Benefit	205	21.05
Patient	198	20.33
Hospital	181	18.58
Agent	180	18.48
Risk	176	18.07
Medical	164	16.84
Adjusted	155	15.91
Control	139	14.27
Therapy	128	13.14
Cancer	127	13.04
Infection	127	13.04
Decision	121	12.42
Comparison	119	12.22
Assessment	116	11.91
Heart	111	11.40
Sensitivity	111	11.40
Economic	110	11.29

Single words in EMTREE heading and subheadings	Number of records with term (total=927)	Percentage of 50% gold standard sample (%)
Hospitalization	108	11.09
Mortality	104	10.68
Evaluation	101	10.37
Making	97	9.96
Model	97	9.96
Screening	94	9.65
Survival	94	9.65
Research	93	9.55
Surgery	93	9.55
Statistical	87	8.93
Blood	86	8.83
Test	84	8.62
Practice	83	8.52
Procedure	75	7.70
System	75	7.70
Diagnostic	74	7.60
Postoperative	70	7.19
Utilization	70	7.19
Acid	66	6.78
Utility	66	6.78
Effect	64	6.57
Pain	64	6.57
Severity	64	6.57
Computer	63	6.47
Inhibitor	63	6.47
Lung	62	6.37
Surgical	62	6.37
Prospective	60	6.16
Aspect	59	6.06
Dose	59	6.06
Service	59	6.06
Virus	59	6.06
Program	58	5.95
Blind	57	5.85
Data	57	5.85
High	57	5.85
Review	56	5.75
Safety	53	5.44
Artery	52	5.34
Accuracy	51	5.24
Technique	51	5.24
Tissue	51	5.24
Cell	50	5.13

Single words in Emtree heading and subheadings	Number of records with term (total=927)	Percentage of 50% gold standard sample (%)
Population	50	5.13
Antibiotic	49	5.03
Primary	49	5.03

The 50% EMBASE gold standard was also analysed for the occurrences of subject headings combined with specific subheadings. No combinations appeared in more than 5% of the sample records. The highest combination was Side Effect/si [Side effect] which appeared in 40 records (4.32% of records).

APPENDIX 4: PUBLISHED SEARCH FILTERS USED IN COMPARATIVE TESTING

The filters that are listed below were obtained from websites and publications. Some small changes were made to some filters, and these are indicated in the text. Any animal study limits used in the original were removed to allow a standardised comparison of all the filters using the same animal limits. All the filters were combined with the three publication years (2000, 2003 and 2006). They were also tested with the publication type and animal exclusions described in the methods section, and without.

\$	Truncation symbol
*	Indicates a major heading when preceding a subject heading (e.g. *economics/)
*	Truncation symbol when following a term (e.g. economic*)
.ab.	Abstract
.fs.	Floating subheading
.hw.	Subject heading word
.mp.	Title, abstract and indexing
.pt.	Publication Type
.sh.	Subject heading
.ti.	Title
.tw.,.	Textword
.yr.	Publication year
/	Subject heading
:	Truncation symbol
?	Indicates there can be zero or one additional characters (e.g. cost? Identifies cost and also costs)
Adj	Adjacent terms
Adj1	Terms within one word of each other
Exp	Explode (subject heading)
Or/1-3	Combine sets 1 to 3 using OR

MEDLINE

Canadian Agency for Drugs and Technologies in Health (CADTH) sensitive filter

1. *Economics/
2. *Economics, Medical/
3. *Economics, Pharmaceutical/
4. exp "Costs and Cost Analysis"/
5. exp Health Care Costs/
6. exp decision support techniques/
7. economic value of life.sh.
8. exp models, economic/
9. markov chains.sh.
10. monte carlo method.sh.
11. uncertainty.sh.
12. quality of life.sh.
13. quality-adjusted life years.sh.

14. exp health economics/
15. exp economic evaluation/
16. exp pharmacoeconomics/
17. exp economic aspect/
18. quality adjusted life year/
19. quality of life/
20. exp "costs and cost analyses"/
21. cost containment.sh.
22. (economic impact or economic value or pharmacoeconomics or health care cost or economic factors or cost analysis or economic analysis or cost or cost-effectiveness or cost effectiveness or costs or health care cost or cost savings or cost-benefit analysis or hospital costs or medical costs or quality-of-life).sh.
23. health resource allocation.sh.
24. (econom\$ or cost or costly or costing or costed or price or prices or pricing or priced or discount or discounts or discounted or discounting or expenditure or expenditures or budget\$ or afford\$ or pharmacoeconomic or pharmaco-economic\$).ti,ab.
25. (cost\$ adj1 (util\$ or effective\$ or efficac\$ or benefit\$ or consequence\$ or analy\$ or minimi\$ or saving\$ or breakdown or lowering or estimate\$ or variable\$ or allocation or control or illness or sharing or life or lives or affordabl\$ or instrument\$ or technolog\$ or day\$ or fee or fees or charge or charges)).ti,ab.
26. (decision adj1 (tree\$ or analy\$ or model\$)).ti,ab.
27. ((value or values or valuation) adj2 (money or monetary or life or lives or costs or cost)).ti,ab.
28. (qol or qoly or qolys or hrqol or qaly or qalys or qale or qales).ti,ab.
29. (sensitivity analys\$s or "willingness to pay" or quality-adjusted life year\$ or quality adjusted life year\$ or quality-adjusted life expectanc\$ or quality adjusted life expectanc\$).ti,ab.
30. (unit cost or unit-cost or unit-costs or unit costs or drug cost or drug costs or hospital costs or health-care costs or health care cost or medical cost or medical costs).ti,ab.
31. (decision adj1 (tree\$ or analy\$ or model\$)).ti,ab.
32. or/1-31

This filter is used for searching EMBASE and MEDLINE, so some of the subject headings will not work in MEDLINE.

Emory University (Grady)

((economic\$ or cost\$).ti. or cost benefit analysis/ or (treatment outcome/ and ec.fs.)) not ((animals/ not humans/) or letter.pt.))

NHS Economic Evaluation Database (NHS EED)

1. economics/
2. exp "costs and cost analysis"/
3. economics, dental/
4. exp "economics, hospital"/
5. economics, medical/

6. economics, nursing/
7. economics, pharmaceutical/
8. (economic\$ or cost or costs or costly or costing or price or prices or pricing or pharmacoeconomic\$).ti,ab.
9. (expenditure\$ not energy).ti,ab.
10. (value adj1 money).ti,ab.
11. budget\$.ti,ab.
12. or/1-11
13. ((energy or oxygen) adj cost).ti,ab.
14. (metabolic adj cost).ti,ab.
15. ((energy or oxygen) adj expenditure).ti,ab.
16. or/13-15
17. 12 not 16

NHS Quality Improvement Scotland

Full filter

1. exp economics/
2. exp "fees and charges"/
3. exp hospitalization/
4. consumer satisfaction/
5. patient acceptance of health care/
6. disease management/
7. physician's practice patterns/
8. exp "patient care planning"/
9. health care rationing/
10. quality of life/
11. value of life/
12. quality-adjusted life years/
13. "Outcome and Process Assessment (Health Care)"/
14. "outcome assessment (health care)"/
15. models, economic/
16. markov chains/
17. monte carlo method/
18. decision tree/
19. economic\$.tw.
20. ec.fs.
21. (cost? or costing? or costly or costed).tw.
22. (price? or pricing?).tw.
23. (pharmacoeconomic? or (pharmaco adj economic?)).tw.
24. budget\$.tw.
25. (value adj1 money).tw.
26. (value adj1 monetary).tw.
27. charge?.tw.
28. (fee or fees).tw.

29. saving?.tw.
30. preference?.tw.
31. (satisfaction or satisfied).tw.
32. ration\$.tw.
33. "quality of life".tw.
34. qol?.tw.
35. hrqol?.tw.
36. "quality adjusted life year?".tw.
37. qaly?.tw.
38. cba.tw.
39. cea.tw.
40. cua.tw.
41. value?.tw.
42. model\$.tw.
43. markov\$.tw.
44. (monte adj carlo).tw.
45. (decision adj2 (tree? or analys\$)).tw.
46. outcome?.tw.
47. utilit\$.tw.
48. pathway?.tw.
49. protocol?.tw.
50. ((clinical or critical or patient) adj path?).tw.
51. (managed adj2 (care or clinical or network)).tw.
52. (resource? adj1 allocat\$).tw.
53. or/1-52 **

**In the original this line read 'or/1-49' and we have amended this to read 'or/1-52'

Brief filter

1. exp Economics/
2. quality of life/
3. value of life/
4. Quality-adjusted life years/
5. models, economic/
6. markov chains/
7. monte carlo method/
8. decision tree/
9. ec.fs.
10. economic\$.tw.
11. (cost? or costing? or costly or costed).tw.
12. (price? or pricing?).tw.
13. (pharmacoeconomic? or (pharmaco adj economic?)).tw.
14. budget\$.tw.
15. expenditure\$.tw.
16. (value adj1 (money or monetary)).tw.

17. (fee or fees).tw.
18. "quality of life".tw.
19. qol\$.tw.
20. hrqol\$.tw.
21. "Quality adjusted life year\$.tw.
22. qaly\$.tw.
23. cba.tw.
24. cea.tw.
25. cua.tw.
26. utilit\$.tw.
27. markov\$.tw.
28. monte carlo.tw.
29. (decision adj2 (tree\$ or analys\$ or model\$)).tw.
30. ((clinical or critical or patient) adj (path? or pathway?)).tw.
31. (managed adj2 (care or network?)).tw.
32. or/1-31

Royle and Waugh

The filter in the original publication is presented as:

cost* OR economic* OR (quality NEAR life)

We have interpreted this to mean the following in Ovid syntax:

(cost* or economic* or (quality adj2 life)).ti,ab,hw.

Sassi et al.

Extensive filter

1. Exp "Costs and cost analysis"/
2. Economics.fs.
3. OR 2
4. Journal article.pt.
5. 3 AND 4

We have made the following conversions:

"Economics" (subheading) became economics.fs.

Selective filter A

1. Exp "Costs and cost analysis"/
2. Journal article.pt.
3. 1 AND 2

Selective filter B

1. Cost-benefit analysis/
2. Journal article.pt.
3. 1 AND 2

‘Cost-benefit analysis/’ has narrower terms in Medline. We have copied Sassi’s approach and have not exploded the term.

Selective filter C

1. Cost-benefit analysis/
2. Exp "Costs and cost analysis"/
3. Cost-effective\$.ti,ab.
4. 1 OR (2 AND 3)
5. Journal article.pt.
6. 4 AND 5

‘Cost-benefit analysis/’ has narrower terms in Medline. We have copied Sassi’s approach and have not exploded the term.

Scottish Intercollegiate Guidelines Network (SIGN)

Economics/

1. "costs and cost analysis"/
2. Cost allocation/
3. Cost-benefit analysis/
4. Cost control/
5. Cost savings/
6. Cost of illness/
7. Cost sharing/
8. "deductibles and coinsurance"/
9. Medical savings accounts/
10. Health care costs/
11. Direct service costs/
12. Drug costs/
13. Employer health costs/
14. Hospital costs/
15. Health expenditures/
16. Capital expenditures/
17. Value of life/
18. Exp economics, hospital/
19. Exp economics, medical/
20. Economics, nursing/
21. Economics, pharmaceutical/
22. Exp "fees and charges"/
23. Exp budgets/
24. (low adj cost).mp.
25. (high adj cost).mp.
26. (health?care adj cost\$).mp.
27. (fiscal or funding or financial or finance).tw.
28. (cost adj estimate\$).mp.
29. (cost adj variable).mp.
30. (unit adj cost\$).mp.

31. (economic\$ or pharmacoeconomic\$ or price\$ or pricing).tw.
32. Or/1-32

Wilczynski et al.

Best sensitivity filter

Costs.tw. or cost effective.tw. or economic.tw.

Best specificity filter

The original publication presents the filter as:

Cost effective.tw. or sensitivity analys:.tw. or cost effectiveness.tw.

We have corrected the typo in cost effectiveness, and used the ‘*’ truncation because it is easier to see than the colon:

(cost effective or sensitivity analys* or cost effectiveness).tw.

Best optimising of sensitivity and specificity filter

Cost-benefit analysis/ or costs.tw. or cost effective.tw.

EMBASE

Canadian Agency for Drugs and Technologies in Health (CADTH) sensitive filter

1. *Economics/
2. *Economics, Medical/
3. *Economics, Pharmaceutical/
4. exp "Costs and Cost Analysis"/
5. exp Health Care Costs/
6. exp decision support techniques/
7. economic value of life.sh.
8. exp models, economic/
9. markov chains.sh.
10. monte carlo method.sh.
11. uncertainty.sh.
12. quality of life.sh.
13. quality-adjusted life years.sh.
14. exp health economics/
15. exp economic evaluation/
16. exp pharmacoeconomics/
17. exp economic aspect/
18. quality adjusted life year/
19. quality of life/
20. exp "costs and cost analyses"/
21. cost containment.sh.
22. (economic impact or economic value or pharmacoeconomics or health care cost or economic factors or cost analysis or economic analysis or cost or cost-effectiveness or cost

effectiveness or costs or health care cost or cost savings or cost-benefit analysis or hospital costs or medical costs or quality-of-life).sh.

23. health resource allocation.sh.

24. (econom\$ or cost or costly or costing or costed or price or prices or pricing or priced or discount or discounts or discounted or discounting or expenditure or expenditures or budget\$ or afford\$ or pharmacoeconomic or pharmaco-economic\$).ti,ab.

25. (cost\$ adj1 (util\$ or effective\$ or efficac\$ or benefit\$ or consequence\$ or analy\$ or minimi\$ or saving\$ or breakdown or lowering or estimate\$ or variable\$ or allocation or control or illness or sharing or life or lives or affordabl\$ or instrument\$ or technolog\$ or day\$ or fee or fees or charge or charges)).ti,ab.

26. (decision adj1 (tree\$ or analy\$ or model\$)).ti,ab.

27. ((value or values or valuation) adj2 (money or monetary or life or lives or costs or cost)).ti,ab.

28. (qol or qoly or qolys or hrqol or qaly or qalys or qale or qales).ti,ab.

29. (sensitivity analys\$ or "willingness to pay" or quality-adjusted life year\$ or quality adjusted life year\$ or quality-adjusted life expectanc\$ or quality adjusted life expectanc\$).ti,ab.

30. (unit cost or unit-cost or unit-costs or unit costs or drug cost or drug costs or hospital costs or health-care costs or health care cost or medical cost or medical costs).ti,ab.

31. (decision adj1 (tree\$ or analy\$ or model\$)).ti,ab.

32. or/1-31

This filter is used for searching EMBASE and MEDLINE, so some of the subject headings will not work in EMBASE.

McKinlay et al.

Best sensitivity filter

Cost effectiveness analysis.sh. OR randomized.tw. OR economic.tw.

Best specificity filter

Cost effectiveness.tw. OR sensitivity analys:.tw.

Best optimization of sensitivity and specificity

Cost.tw. OR costs.tw.

NHS Economic Evaluation Database

1. health-economics/
2. exp economic-evaluation/
3. exp health-care-cost/
4. exp pharmacoeconomics/
5. 1 or 2 or 3 or 4

6. (econom\$ or cost or costs or costly or costing or price or prices or pricing or pharmaco-economic\$).ti,ab
7. (expenditure\$ not energy).ti,ab
8. (value adj2 money).ti,ab
9. budget\$.ti,ab
10. 6 or 7 or 8 or 9
11. 5 or 10
12. (metabolic adj cost).ti,ab
13. ((energy or oxygen) adj cost).ti,ab
14. ((energy or oxygen) adj expenditure).ti,ab
15. 12 or 13 or 14
16. 11 not 15

Line 14 has been changed from the original. The original had 'near' (which is unavailable in Ovid) and this has been replaced with 'adj'.

NHS Quality Improvement Scotland

1. exp health economics/
2. exp health care cost/
3. exp quality of life/
4. economic\$.tw.
5. (cost? or costing? or costly or costed).tw.
6. (price? or pricing?).tw.
7. (pharmaco-economic? or (pharmaco adj economic?)).tw.
8. budget\$.tw.
9. expenditure\$.tw.
10. (value adj1 (money or monetary)).tw.
11. (fee or fees).tw.
12. "quality of life".tw.
13. qol\$.tw.
14. hrqol\$.tw.
15. "quality adjusted life year\$".tw.
16. qaly\$.tw.
17. cba.tw.
18. cea.tw.
19. cua.tw.
20. utilit\$.tw.
21. markov\$.tw.
22. monte carlo.tw.
23. (decision adj2 (tree\$ or analys\$ or model\$)).tw.
24. ((clinical or critical or patient) adj (path? or pathway?)).tw.
25. (managed adj2 (care or network?)).tw.
26. or/1-25

Royle and Waugh

The original publication presents the filter as follows:

cost* OR economic* OR (quality NEAR life)

We have interpreted this to mean the following in Ovid syntax:

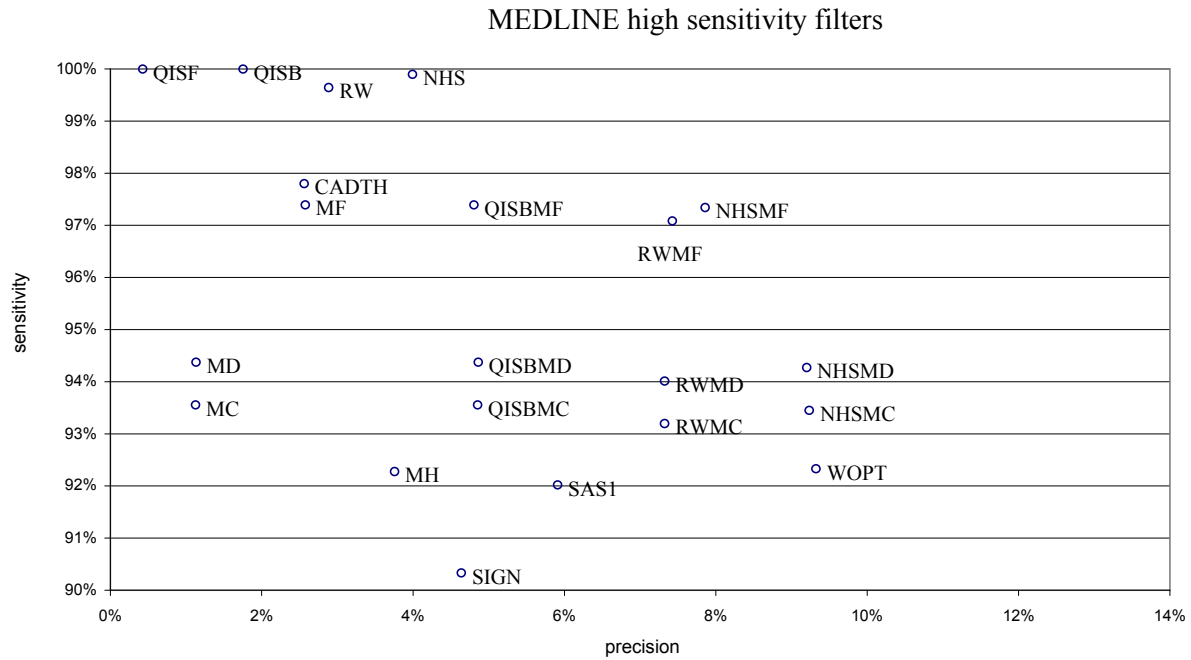
(cost* or economic* or (quality adj2 life)).ti,ab,hw.

Scottish Intercollegiate Guidelines Network (SIGN)

1. Socioeconomics/
2. Cost benefit analysis/
3. Cost effectiveness analysis/
4. Cost of illness/
5. Cost control/
6. Economic aspect/
7. Financial management/
8. Health care cost/
9. Health care financing/
10. Health economics/
11. Hospital cost/
12. (fiscal or financial or finance or funding).tw.
13. Cost minimization analysis/
14. (cost adj estimate\$).mp.
15. (cost adj variable\$).mp.
16. (unit adj cost\$).mp.
17. Or/1-16

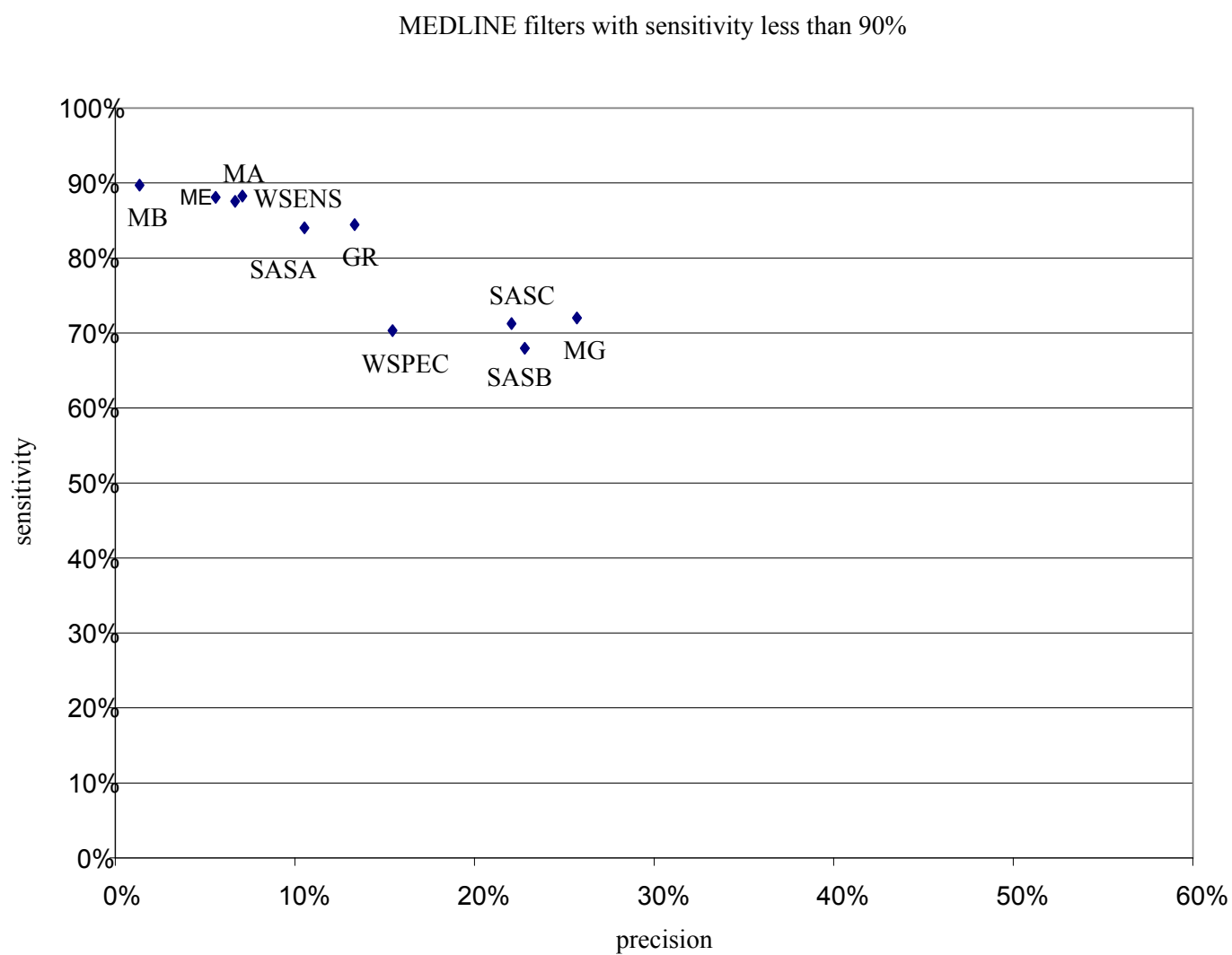
APPENDIX 5: SENSITIVITY-PRECISION TRADE-OFFS BETWEEN FILTERS

Figure A1: MEDLINE filters with high sensitivity (90% or above)



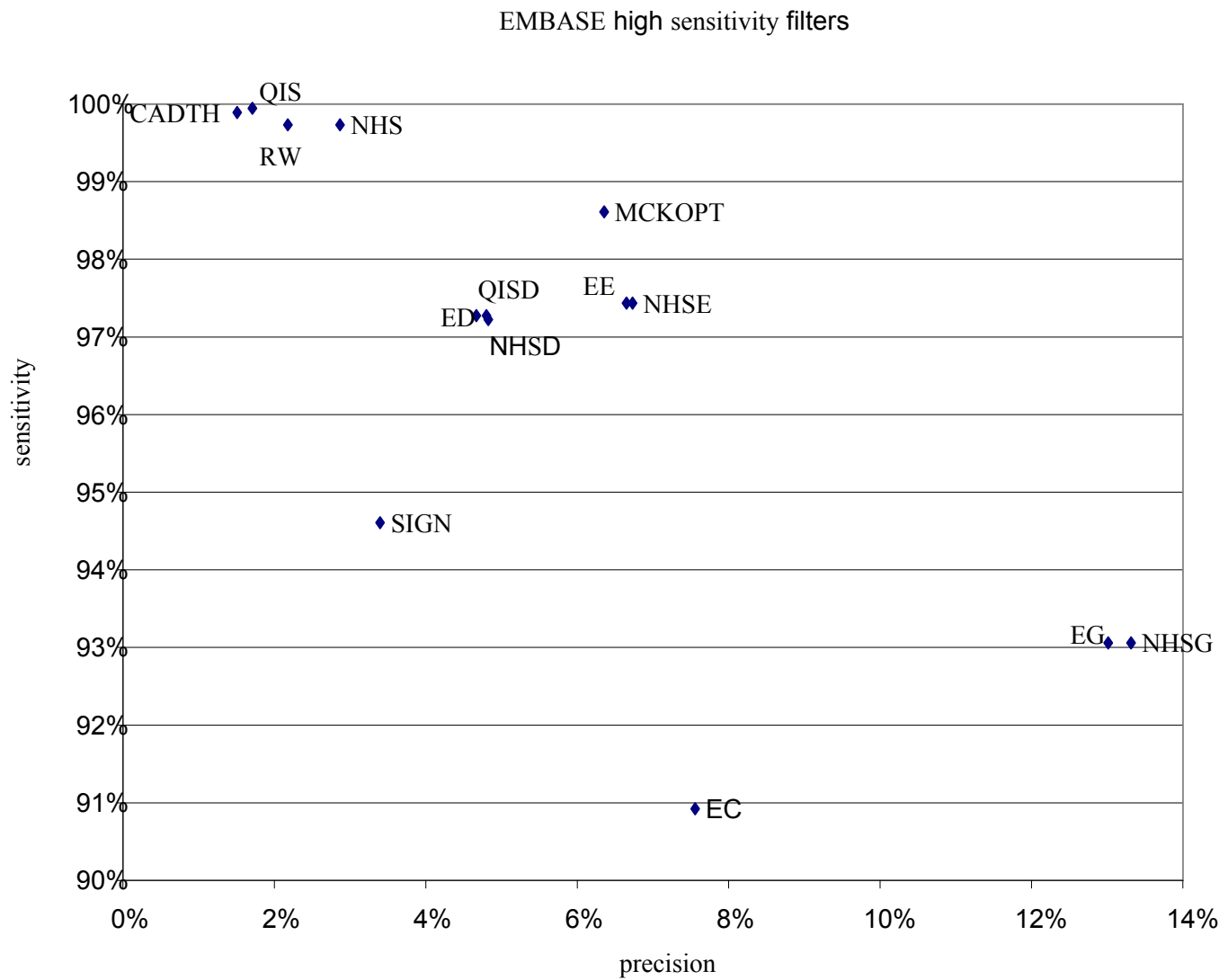
CADTH	CADTH
MC	MEDLINE C
MD	MEDLINE D
MF	MEDLINE F
MH	MEDLINE H
NHS	NHS EED
NHSMC	NHS EED plus Medline C
NHSMD	NHS EED plus Medline D
NHSMF	NHS EED plus Medline F
QISB	NHS Quality Improvement Scotland Brief
QISBMC	NHS QIS brief plus Medline C
QISBMD	NHS QIS brief plus Medline D
QISBMF	NHS QIS brief plus Medline F
QISF	NHS Quality Improvement Scotland Full
RW	Royle and Waugh
RWMC	Royle plus Medline C
RWMD	Royle plus Medline D
RWMF	Royle plus Medline F
SAS1	Sassi LSE extensive filter
SIGN	SIGN
WOPT	Wilczynski best optimisation of sensitivity and precision

Figure A2: MEDLINE filters with sensitivity less than 90%



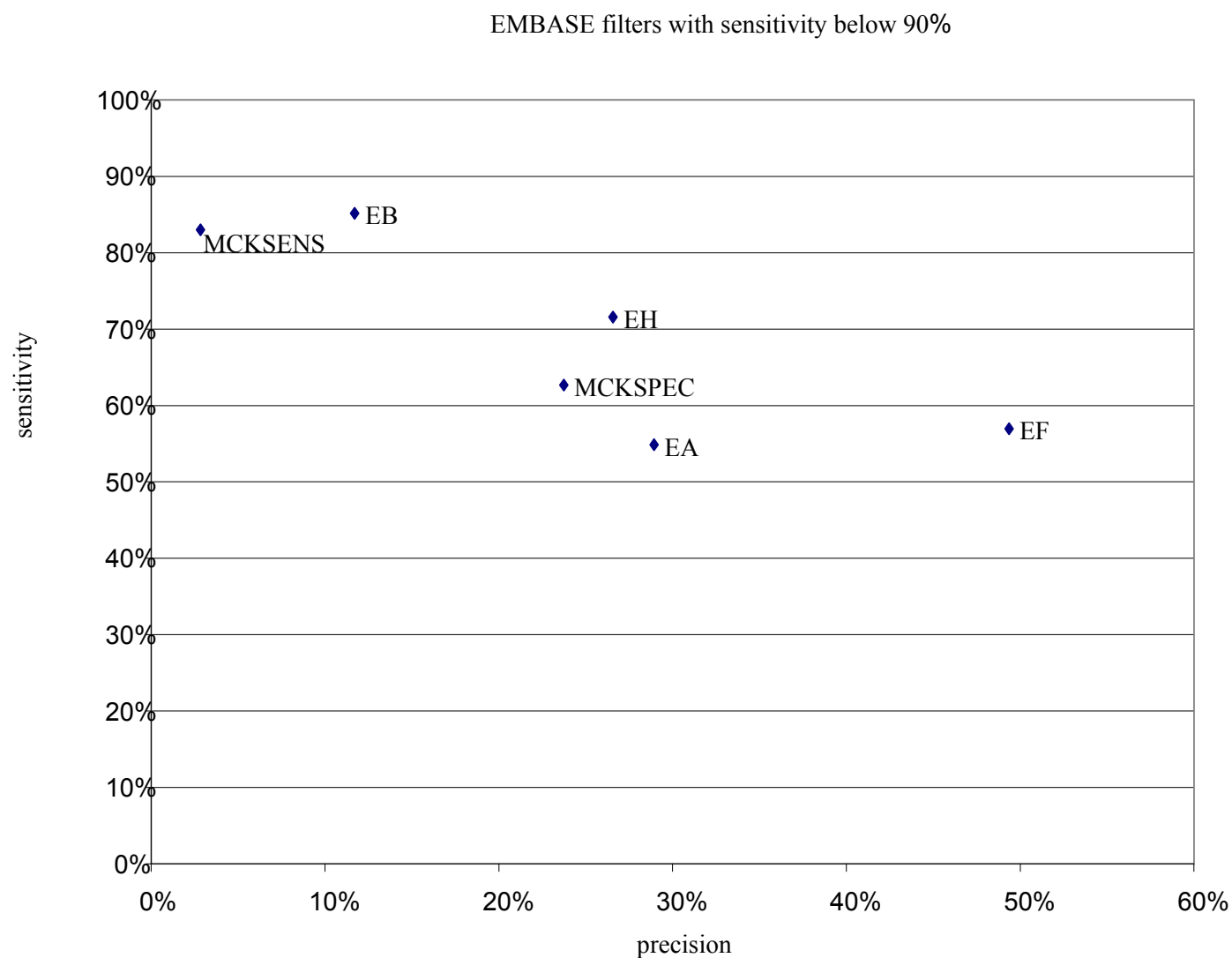
GR	Grady
MA	MEDLINE A
MB	MEDLINE B
ME	MEDLINE E
MG	MEDLINE G
SASA	Sassi LSE selective filter A
SASB	Sassi LSE selective filter B
SASC	Sassi LSE selective filter C
WSENS	Wilczynski best sensitivity
WSPEC	Wilczynski best specificity

Figure A3: EMBASE filters with high sensitivity (90% and above)



CADTH	CADTH
EC	EMBASE C
ED	EMBASE D
EE	EMBASE E
EG	EMBASE G
MCKOPT	McKinlay best optimization of sensitivity/specificity
NHS	NHS EED
NHSD	NHS EED plus EMBASE D
NHSE	NHS EED plus EMBASE E
NHSG	NHS EED plus EMBASE G
QIS	NHS Quality Improvement Scotland
QISD	NHS QIS plus EMBASE D
RW	Royle and Waugh
SIGN	SIGN

Figure A4: EMBASE filters with sensitivity less than 90%



EA	EMBASE A
EB	EMBASE B
EF	EMBASE F
EH	EMBASE H
MCKSENS	McKinlay best sensitivity
MCKSPEC	McKinlay best specificity