



Cochrane
Library

Cochrane Database of Systematic Reviews

Search strategies (filters) to identify systematic reviews in MEDLINE and Embase (Review)

Escobar Liquitay CM, Garegnani L, Garrote V, Solà I, Franco JVA

Escobar Liquitay CM, Garegnani L, Garrote V, Solà I, Franco JVA.
Search strategies (filters) to identify systematic reviews in MEDLINE and Embase.
Cochrane Database of Systematic Reviews 2023, Issue 9. Art. No.: MR000054.
DOI: [10.1002/14651858.MR000054.pub2](https://doi.org/10.1002/14651858.MR000054.pub2).

www.cochranelibrary.com

TABLE OF CONTENTS

ABSTRACT	1
PLAIN LANGUAGE SUMMARY	2
SUMMARY OF FINDINGS	4
BACKGROUND	6
OBJECTIVES	7
METHODS	7
RESULTS	9
Figure 1.	10
DISCUSSION	13
AUTHORS' CONCLUSIONS	15
ACKNOWLEDGEMENTS	15
REFERENCES	16
CHARACTERISTICS OF STUDIES	21
ADDITIONAL TABLES	33
APPENDICES	69
HISTORY	74
CONTRIBUTIONS OF AUTHORS	74
DECLARATIONS OF INTEREST	74
SOURCES OF SUPPORT	75
DIFFERENCES BETWEEN PROTOCOL AND REVIEW	75
NOTES	75
INDEX TERMS	75

[Methodology Review]

Search strategies (filters) to identify systematic reviews in MEDLINE and Embase

Camila Micaela Escobar Liquitay¹, Luis Garegnani¹, Virginia Garrote², Ivan Solà³, Juan VA Franco⁴

¹Research Department, Instituto Universitario Hospital Italiano de Buenos Aires, Buenos Aires, Argentina. ²Central Library, Instituto Universitario Hospital Italiano de Buenos Aires, Buenos Aires, Argentina. ³Iberoamerican Cochrane Centre, Biomedical Research Institute Sant Pau (IIB Sant Pau), CIBER Epidemiología y Salud Pública (CIBERESP), Barcelona, Spain. ⁴Institute of General Practice, Medical Faculty of the Heinrich-Heine-University Düsseldorf, Düsseldorf, Germany

Contact: Juan VA Franco, juan.franco@med.uni-duesseldorf.de.

Editorial group: Cochrane Methodology Review Group.

Publication status and date: New, published in Issue 9, 2023.

Citation: Escobar Liquitay CM, Garegnani L, Garrote V, Solà I, Franco JVA. Search strategies (filters) to identify systematic reviews in MEDLINE and Embase. *Cochrane Database of Systematic Reviews* 2023, Issue 9. Art. No.: MR000054. DOI: [10.1002/14651858.MR000054.pub2](https://doi.org/10.1002/14651858.MR000054.pub2).

Copyright © 2023 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

ABSTRACT

Background

Bibliographic databases provide access to an international body of scientific literature in health and medical sciences. Systematic reviews are an important source of evidence for clinicians, researchers, consumers, and policymakers as they address a specific health-related question and use explicit methods to identify, appraise and synthesize evidence from which conclusions can be drawn and decisions made.

Methodological search filters help database end-users search the literature effectively with different levels of sensitivity and specificity. These filters have been developed for various study designs and have been found to be particularly useful for intervention studies. Other filters have been developed for finding systematic reviews. Considering the variety and number of available search filters for systematic reviews, there is a need for a review of them in order to provide evidence about their retrieval properties at the time they were developed.

Objectives

To review systematically empirical studies that report the development, evaluation, or comparison of search filters to retrieve reports of systematic reviews in MEDLINE and Embase.

Search methods

We searched the following databases from inception to January 2023: MEDLINE, Embase, PsycINFO; Library, Information Science & Technology Abstracts (LISTA) and Science Citation Index (Web of Science).

Selection criteria

We included studies if one of their primary objectives is the development, evaluation, or comparison of a search filter that could be used to retrieve systematic reviews on MEDLINE, Embase, or both.

Data collection and analysis

Two review authors independently extracted data using a pre-specified and piloted data extraction form using InterTASC Information Specialist Subgroup (ISSG) Search Filter Evaluation Checklist.

Main results

We identified eight studies that developed filters for MEDLINE and three studies that developed filters for Embase. Most studies are very old and some were limited to systematic reviews in specific clinical areas. Six included studies reported the sensitivity of their developed filter. Seven studies reported precision and six studies reported specificity. Only one study reported the number needed to read and positive predictive value. None of the filters were designed to differentiate systematic reviews on the basis of their methodological quality. For MEDLINE, all filters showed similar sensitivity and precision, and one filter showed higher levels of specificity. For Embase, filters showed variable sensitivity and precision, with limited study reports that may affect accuracy assessments. The report of these studies had some limitations, and the assessments of their accuracy may suffer from indirectness, considering that they were mostly developed before the release of the PRISMA 2009 statement or due to their limited scope in the selection of systematic review topics.

Search filters for MEDLINE

Three studies produced filters with sensitivity > 90% with variable degrees of precision, and only one of them was developed and validated in a gold-standard database, which allowed the calculation of specificity. The other two search filters had lower levels of sensitivity. One of these produced a filter with higher levels of specificity (> 90%). All filters showed similar sensitivity and precision in the external validation, except for one which was not externally validated and another one which was conceptually derived and only externally validated.

Search filters for Embase

We identified three studies that developed filters for this database. One of these studies developed filters with variable sensitivity and precision, including highly sensitive strategies (> 90%); however, it was not externally validated. The other study produced a filter with a lower sensitivity (72.7%) but high specificity (99.1%) with a similar performance in the external validation.

Authors' conclusions

Studies reporting the development, evaluation, or comparison of search filters to retrieve reports of systematic reviews in MEDLINE showed similar sensitivity and precision, with one filter showing higher levels of specificity. For Embase, filters showed variable sensitivity and precision, with limited information about how the filter was produced, which leaves us uncertain about their performance assessments. Newer filters had limitations in their methods or scope, including very focused subject topics for their gold standards, limiting their applicability across other topics. Our findings highlight that consensus guidance on the conduct of search filters and standardized reporting of search filters are needed, as we found highly heterogeneous development methods, accuracy assessments and outcome selection. New strategies adaptable across interfaces could enhance their usability. Moreover, the performance of existing filters needs to be evaluated in light of the impact of reporting guidelines, including the PRISMA 2009, on how systematic reviews are reported. Finally, future filter developments should also consider comparing the filters against a common reference set to establish comparative performance and assess the quality of systematic reviews retrieved by strategies.

PLAIN LANGUAGE SUMMARY

How can we best filter systematic reviews in MEDLINE and Embase?

Key Messages

A wide range of search filters to retrieve systematic reviews were evaluated. Although many had acceptable sensitivity (missed few relevant studies) and specificity (omitted irrelevant studies), no single filter can be recommended since most were derived from older sets of reviews that may not reflect current reporting characteristics and standards.

What are search filters for systematic reviews?

Search filters combine words and phrases to retrieve records with a common feature (e.g. study design, clinical topic) and are typically evaluated in terms of their sensitivity and precision. Systematic reviews summarise and synthesise scientific evidence and represent an important source of information for healthcare professionals. Databases provide access to them, and search filters can be used to retrieve systematic reviews pragmatically.

What did we want to find out?

We wanted to identify search filters for systematic reviews, assess their quality and retrieve data on their sensitivity, specificity and precision.

What did we do?

We searched for studies that developed, evaluated or compared a search filter that could be used to retrieve systematic reviews in MEDLINE, Embase, or both. We identified nine studies that developed filters for MEDLINE and three studies that developed filters for Embase.

What did we find?

For MEDLINE, all filters showed similar sensitivity and precision, and one filter showed higher levels of specificity. For Embase, filters showed variable sensitivity and precision, with limited study reports that may affect accuracy assessments.

What are the limitations of the evidence?

Some filters were developed for specific topics (e.g. public health), and most were developed using older studies, which may not reflect how systematic reviews are currently reported. Moreover, filters may not be able to discern between high- and low-quality reviews.

How up-to-date is the evidence?

The evidence is up-to-date to January 2023.

SUMMARY OF FINDINGS

Summary of findings 1. Summary of findings of the available filters

Range of measurements of accuracy from a) internal validation, b) external validation and c) independent evaluations							
Study	Database (interface)	Year of development/ Year of external evaluation	Sensitivity ^a	Specificity ^a	Precision ^a	NNR ^a	Comments
Filters with external evaluations and acceptable sensitivity and precision^b							
Shojania 2001	MEDLINE (PubMed)	1999 to 2000/2007, 2012 and 2021	b) 93 to 97% c) 62 to 90%	c) 97.2% to 99.1%	c) 1.7 to 33.2%	c) 3.01 to 57.8	Filter with independent evaluations with lower sensitivity.
Boynton 1998	MEDLINE (Ovid)	1992 to 1995/2001 and 2012	a) 39 to 98% c) 47.8 to 99.5%	c) 75.6 to 99.6%	a) 12 to 79% c) 0.1 to 2.1%	a) 2.04 to 8.33 c) 46.7 to 1395	Filter developed in an old dataset but with recent positive independent evaluations. High sensitivity but low precision.
Wilczynski 2007	Embase (Ovid)	2000/2012	a) 61.4 to 94.6% c) 63.4 to 96.3%	a) 63.7 to 99.3% c) 72.3 to 99.5%	a) 2 to 40.9% b) 0 to 0.9%	a) 2.44 to 50 b) 117.8 to 2709.5	Four filters with different sensitivity and specificity profiles with consistent independent evaluations.
	MEDLINE (Ovid)	2000/2012	a) 75.2 to 100% b) 71.2 to 99.9% c) 81.6 to 99.0%	a) 63.5 to 99.4% b) 52 to 99.2% c) 62 to 99.3%	a) 3.41 to 60.2% b) 3.14 to 57.1% c) 0 to 2%	a) 1.66 to 29.33 b) 1.75 to 31.84 c) 49.4 to 2191.2	Three filters with different sensitivity and specificity profiles with consistent independent evaluations.
Filters focus on specific topics							
Boluyt 2008	MEDLINE (PubMed)	1994 to 2004	b) 68 to 96%	N/A	b) 2 to 45%	2.22 to 50	Adaptation of existing search filters (Boynton 1998; Shojania 2001; White 2001; Wilczynski 2007) for systematic reviews on child health.

Lee 2012	MEDLINE (Ovid)	2004 to 2005	a) 86.8%	a) 99.2%	a) 1.1%	a) 91.6	Filter with a focus on public health topics without independent evaluations with suboptimal sensitivity.
			b) 89.9%	b) 98.9 %	b) 1.4%	b) 71.4	
	Embase (Ovid)		a) 72.7%	a) 99.1%	a) 0.6%	a) 171.6	
			b) 87.9%	b) 98.2%	b) 0.5%	b) 186.0	
Avau 2021	MEDLINE (PubMed)	2019-2020	N/A	b) 97%	b) 9.7%	b) 10	Filter with a focus on fist aid without independent evaluations.
	Embase (Elsevier)		N/A	b) 96%	b) 5.4%	b) 19	
Others							
White 2001	MEDLINE (Ovid)	1995 to 1997	a) 67.1 to 87.1% b) 84.2%	a) 89.2 to 99.4% b) 93%	N/A	N/A	Filter without independent evaluations with suboptimal sensitivity.
Salvador-Oliván 2021	MEDLINE (PubMed)	2020	N/A	N/A	b) 83.8%	b) 1.19	Filter without independent evaluation assessed to retrieve "possible systematic reviews"

Definition of outcome measures:

- Sensitivity: Proportion of systematic reviews that are correctly identified using the methodological filter.
- Specificity: Proportion of records that are not systematic reviews not identified using the methodological filter.
- Precision: Proportion of systematic reviews that are identified from all records retrieved using the methodological filter.
- Number needed to read (NNR): 1/precision

N/A: Not available. NR: not reported.

^a Range of point estimates across different reports, evaluations and subsets of filters.

^b >90% for sensitivity and >10% for precision

See [Table 1](#) for a more detailed summary of the number of filters developed and evaluated in each database.

BACKGROUND

Bibliographic databases, such as MEDLINE and Embase, provide access to an international body of scientific literature in health and medical sciences. They provide bibliographic citation information and, frequently, abstracts or links to full-text publications. These databases also provide controlled vocabulary (index terms) to make it easier to index, catalogue and search biomedical and health-related information and documents (Dhammi 2014; Leydesdorff 2016; Lipscomb 2000).

Systematic reviews of the literature are an important source of evidence for clinicians, researchers, consumers, and policymakers. They typically address a specific health-related question using explicit methods to identify, appraise and synthesize research-based evidence and present it in an accessible format, providing more reliable findings from which conclusions can be drawn and decisions made when reviews are well conducted (Chandler 2019). Systematic reviews can also be a useful starting point for researchers by identifying gaps in the evidence (Ioannidis 2016). Cochrane provides two main guidelines for the development and reporting of systematic reviews. These are the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2022b) and the *Methodological Expectations of Cochrane Intervention Reviews* (known as the MECIR standards) (Higgins 2022a). The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement (Moher 2009; Page 2021) provides guidance on how to report systematic reviews; there is evidence that adherence to these guidelines has improved over time (Page 2016); however, there is still scope for improvement. The PRISMA statement updates and adherence may have led to the improved reporting of systematic reviews, affecting filter performances.

Systematic reviews are also widely used to develop clinical practice guidelines (IOM 2011), overviews (Becker 2011), and other forms of evidence synthesis, such as evidence mappings (Bragge 2011). Consequently, the appropriate and prompt identification of systematic reviews is necessary for many important purposes.

Description of the methods being investigated

Search filters were originally defined by Wilczynski 1995 as a list of terms that can improve the detection of studies of high quality for clinical practice. The *Cochrane Handbook* defines search filters as "search strategies that are designed to retrieve specific types of records, such as those of a particular methodological design" (Lefebvre 2022). Alternative terms used include clinical queries, hedges, optimal search filters, optimal search strategies, quality filters, search filters, or search strategies (Jenkins 2004). At the same time, the original definition by Wilczynski and colleagues included study quality; methodological search filters are not necessarily designed to retrieve studies by their quality. Some search filters have been assessed to determine how effective they are at identifying relevant articles while avoiding the detection of irrelevant articles. For the purpose of this review, we will restrict the definition of "search filters" to search strategies with a formally published test of diagnostic performance (e.g. sensitivity, specificity, precision, etc.). We will refer to other methods for retrieving specific types of records as "search strategies" alone. As the amount of research evidence continues to increase rapidly in some areas and indexing is not consistent for all studies designs, the use of search filters has been advocated to assist the searching process, as it reduces the total number of records found and

increases the likelihood that they will be of interest. For that reason, the performance of a search filter is usually calculated according to its capacity to retrieve as many relevant citations as possible whilst also omitting irrelevant results (Wilczynski 2005), the aim being to reduce the number of irrelevant citations that may have to be screened to find a relevant systematic review (Bachmann 2002).

Currently, one of the ways that MEDLINE can be searched is via PubMed (www.pubmed.gov) and provides two related publication type descriptors for the retrieval of systematic reviews: "meta-analysis" (introduced in 1993), which might not be useful for those systematic reviews which do not include a meta-analysis; and "review" (introduced in 1966), which may not differentiate systematic reviews from narrative reviews. More recently, PubMed incorporated a filter to retrieve "systematic reviews" through the system interface (systematic review subset) (Shojania 2001). This filter was originally intended to retrieve citations identified as systematic reviews, meta-analyses, reviews of clinical trials, evidence-based medicine, consensus development conferences, guidelines, and citations to articles from journals specialising in reviews of value to clinicians. This filter has been updated periodically and now contains terms more specific to systematic reviews (the last update was in December 2018). These modifications were pragmatic and have not undergone testing for sensitivity, selectivity, precision, or accuracy in a formal validation process (Bradley 2010). Recently, the National Library of Medicine added new terminology to the Medical Subject Headings: "Systematic review as topic" and "Systematic review" [publication type], defined as "A review of primary literature in health and health policy that attempts to identify, appraise, and synthesise all the empirical evidence that meets specified eligibility criteria to answer a given research question ... aimed at minimising bias in order to produce more reliable findings regarding the effects of interventions for prevention, treatment, and rehabilitation that can be used to inform decision making." (NLM 2019a) The "systematic review as topic" heading will help with identifying studies that address systematic reviews as a method and the other heading will help to identify systematic reviews. However, the value of the index terms is dependent on the consistency with which the indexers apply the index terms appropriately (NLM 2019b; NLM 2019c).

Elsevier indexes Embase citations with the check tag "systematic review" for studies summarising systematically all the available evidence (Embase Indexing Guide 2020). Over the years, several organisations and individuals developed and evaluated search filters to retrieve systematic reviews in Embase (SIGN, Wilczynski 2007). Other filters have been developed for finding systematic reviews in MEDLINE via Ovid, such as one by Montori and colleagues, who created the filter by assessing index terms, text words and discussions with clinicians and biomedical librarians (Montori 2005).

How these methods might work

Methodological search filters are used to help end-users search the literature effectively (Jenkins 2004). Filters have been developed with different levels of sensitivity and specificity according to the requirements of the users (for instance, those with high sensitivity, high specificity, a balance between sensitivity and specificity or a balance between sensitivity and precision) (Brettelle 1998; Glanville 2000; Jenkins 2004).

Methodological search filters have been developed for various study designs and have been found to be particularly useful for intervention studies. Within the *Cochrane Handbook*, for example, a highly sensitive search strategy is proposed for identifying reports of randomised trials (Lefebvre 2022); and there are Cochrane Reviews of the evidence on filters for retrieving studies of diagnostic test accuracy (Beynon 2013) and observational studies (Li 2019). However, there is no guidance for finding the best filters for identifying systematic reviews (Becker 2011).

Why it is important to do this review

Systematic reviews provide core material for guidelines, overviews, health technology assessments and other forms of evidence synthesis, as well as being an invaluable tool for decision-making (Sprakel 2019). Authors of these documents are faced with a choice of different methods to retrieve systematic reviews for their research question or clinical scenario. Considering the variety and number of available search filters for systematic reviews, there is a need for a review of them in order to provide evidence about their retrieval properties at the time they were developed. We have restricted our review scope to MEDLINE and Embase since these are large and widely used bibliographic databases, often suggested as mandatory when conducting evidence syntheses in health care (Bramer 2017). We expect that the findings of our review will aid those who use systematic methods for information retrieval (e.g. researchers conducting overviews or evidence maps) who wish to use validated search filters with adequate sensitivity and specificity, whereas other stakeholders (e.g. clinicians and consumers) might use search filters that are built into search interfaces, such as the "systematic review" filter in PubMed/MEDLINE.

OBJECTIVES

To review systematically empirical studies that report the development, evaluation, or comparison of search filters to retrieve reports of systematic reviews in MEDLINE and Embase.

METHODS

Criteria for considering studies for this review

Types of studies

We included studies if one of their main objectives is the development, evaluation, or comparison of a search filter that

could be used to identify systematic reviews in MEDLINE, Embase, or both. A development study is one in which a filter was generated and tested for its ability to identify relevant articles while avoiding the detection of irrelevant articles. An evaluation study is one in which these properties of a developed filter are tested in a new reference set of relevant studies. A comparison study is one in which different search filters are tested in a reference set to compare their properties. We considered systematic reviews as defined by the study authors.

Types of data

We collected the following information from the included studies.

- Methodological filters: fully detailed search strategies
- Dates the searches were conducted
- Years covered by the searches
- Electronic bibliographic database (MEDLINE or Embase) and interface used (e.g. Ovid or PubMed)
- Healthcare topic
- Characteristics of the gold standard used to test the filter
- Outcome measures (e.g. sensitivity, specificity, or precision)

Types of methods

Search strategies for identifying reports of systematic reviews in MEDLINE and Embase.

Types of outcome measures

We included any of the following outcome measures (see **Table 1** below for more information).

Primary outcomes

- Sensitivity: proportion of systematic reviews that are correctly identified using the methodological filter.
- Specificity: proportion of records that are not systematic reviews not identified using the methodological filter.
- Precision: proportion of systematic reviews that are identified from all records retrieved using the methodological filter.

Table 1. Definition of outcome measures of this review

		Gold standard	
		Systematic review	Not systematic review
Searches with methodological filters	Detected by the filter	a	b
	Not detected by the filter	c	d

$$\text{Sensitivity} = a / (a + c)$$

$$\text{Specificity} = d / (b + d)$$

$$\text{Precision} = a / (a + b)$$

$$\text{Accuracy} = (a + d) / (a + b + c + d)$$

Number needed to read = $1 / \text{precision} = (a + b) / a$

Systematic reviews in the gold standard = $a + c$

Non-systematic reviews in the gold standard = $b + d$

We have defined a priori the levels of sensitivity (more than 90%) and precision (more than 10%) in external validation studies that would be an acceptable threshold for use when searching for systematic reviews (Beynon 2013).

Secondary outcomes

- Accuracy: proportion of records that were adequately classified using the methodological filter
- Number needed to read (NNR): $1/\text{precision}$
- Quality assessment of the systematic reviews retrieved and missed by the search strategy ("a" + "c" in **Table 1**, analyzed by the developers of the filter). This includes whether the authors of the included studies provide details on the quality of reviews that were retrieved or missed by the search filter. For instance, the proportion of high-quality reviews missed by the search filter.

Definitions adapted from Cooper 2018.

Search methods for identification of studies

Electronic searches

We searched the following databases from inception to January 2023.

- MEDLINE Ovid SP and Epub Ahead of Print, In-Process, In-Data-Review & Other Non-Indexed Citations, Daily and Versions (from 1946 to January 2023);
- Embase (Elsevier.com; from 1974 to January 2023);
- PsycINFO Ovid SP (from 1967 to January 2023);
- Library, Information Science & Technology Abstracts (LISTA) (EBSCO); (from inception to January 2023);
- Science Citation Index (Web of Science; Clarivate); (from 1964 to January 2023).

For detailed search strategies for each database, see [Appendix 1](#). We did not restrict searches by the language of publication.

Searching other resources

To identify additional published, unpublished and ongoing studies:

- relevant studies identified from the above sources were entered into PubMed, and the Related Articles feature was used; and
- reference lists of all relevant studies were assessed (Horsley 2011).

We also searched the websites of, among others:

- the InterTASC Information Specialists' Sub-Group (www.intertasc.org.uk/subgroups/issg);
- the Health Information Research Unit (McMaster) (<https://hiru.mcmaster.ca/hiru/>).

Data collection and analysis

Selection of studies

Four review authors (CMEL, VG, VV, JVAF) worked independently in pairs to screen the titles and abstracts of all retrieved records and assess papers for eligibility. Any disagreements were resolved by discussion or consultation with a third author (IS) to reach a consensus.

Full copies of the relevant reports were obtained for records possibly meeting the inclusion criteria. Each full report was assessed independently in pairs by four review authors (CMEL, VG, VV, JVAF) to determine if it met the inclusion criteria for the review. Any disagreements were resolved by discussion or by consultation with a third author (IS) to reach a consensus.

We presented results following a PRISMA 2020 flow diagram (Haddaway 2022; Page 2021).

Data extraction and management

Two review authors (LG and CMEL) independently extracted data, using a piloted pre-specified data extraction form. We extracted Information on the following.

- Citation details for the study
- Methodological filter used
- Dates the searches were conducted
- Years covered by the searches
- Search interface used (e.g. Ovid or PubMed)
- Healthcare topic
- Gold standard
- Outcome measures (e.g. sensitivity, specificity, or precision)

Any disagreements were resolved by discussion or by consultation with a third review author to reach a consensus. If there were studies with incomplete or missing data, we attempted to contact the corresponding author (Young 2011).

Assessment of risk of bias in included studies

A small number of critical appraisal tools have been developed to assess the quality of methodological search filters (Bak 2009; Glanville 2008; Jenkins 2004). The included studies were assessed against the search filter appraisal checklist proposed by the UK InterTASC Information Specialists' Sub-Group (Glanville 2008) and reported in the *Cochrane Handbook* (Lefebvre 2022).

Two review authors (LIG and CMEL) completed the appraisal checklist ([Appendix 2](#)). Any disagreements were resolved by discussion or by consultation with the remaining group of authors (VV, JVAF, or IS) to reach a consensus.

Data synthesis

We synthesized filter performance measures separately for MEDLINE and Embase. We tabulated performance measures reported by development and evaluation studies grouped by individual filters, so that a comparison can be made between a filter's original reported performance and its performance in later evaluation studies. If sensitivity, specificity, or precision (along with 95% confidence intervals (CI)) were not reported in the original reports, they were calculated from 2 x 2 data tables, where possible. Data extraction tables were established that take the InterTASC Information Specialists Sub-Group (ISSG) [Glanville 2008](#) search filter evaluation checklist evaluation fields to evaluate the report on the studies of each of the filters, taking, for example, the methodological objective of the filter (search for systematic reviews/search interface), reference standard, Gold Standard preparation method, sensitivity and precision performance, external validation, among other elements.

Subgroup analysis and investigation of heterogeneity

Where sufficient data were available, we would have performed subgroup analyses based on the following characteristics.

- Dates the searches were conducted: searches conducted before the release of the PRISMA statement in 2009 versus those conducted after its release (because the PRISMA guidance may affect how systematic reviews are reported)

- Search interface used (e.g. PubMed or Ovid)
- Healthcare topic: searches conducted within a specific health topic (e.g. public health, cardiovascular disease, etc.) versus those conducted across the biomedical literature or within a core set of non-specialized biomedical journals (e.g. Core Clinical Journals)

RESULTS

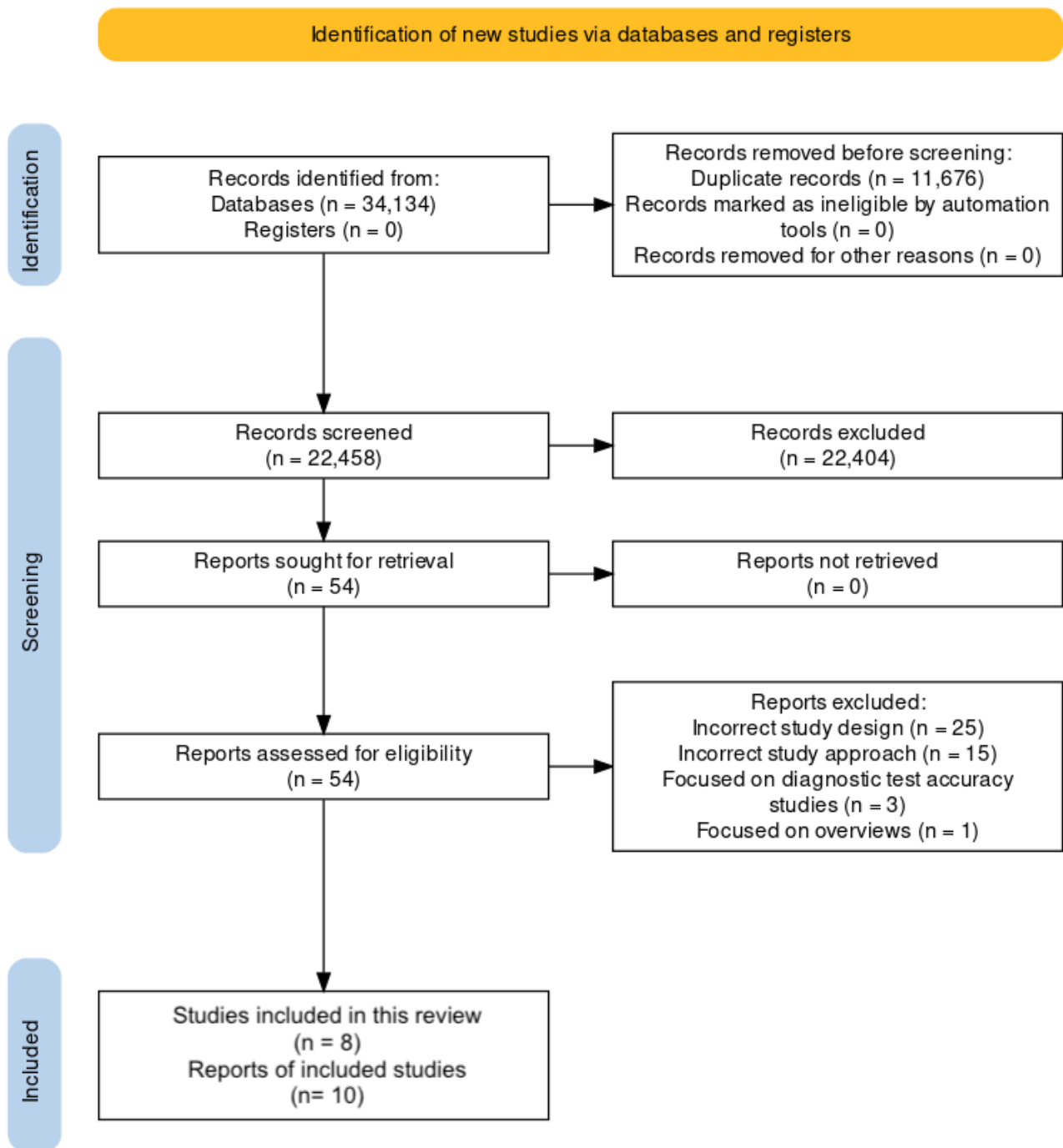
Description of studies

For study details, see [Characteristics of included studies](#); [Characteristics of excluded studies](#).

Results of the search

We retrieved 34,134 records after searching MEDLINE, Embase, PsycINFO, LISTA and Science Citation Index electronic databases. We did not identify studies using additional methods. After removing 11,676 duplicates, 22,458 records remained for the title and abstract screening. Of this total number of records, we eliminated 22,404 as they were considered irrelevant. We analyzed the full texts of the remaining 54 records to assess their inclusion. Finally, we included eight studies reported in 10 publications ([Figure 1](#)). See [Characteristics of included studies](#); [Characteristics of excluded studies](#). We did not classify any study as awaiting classification. We did not identify any ongoing studies.

Figure 1. PRISMA flow diagram



Included studies

The methodological filters of the included studies focused on developing a filter to retrieve systematic reviews. Four studies were development studies (Boynnton 1998; Lee 2012; White 2001; Wilczynski 2007). Seven studies were evaluation studies (Avau 2021; Boluyt 2008; Lee 2012; Salvador-Oliván 2021; Shojania 2001; White 2001; Wilczynski 2007). Seven studies compared their filter with other available filters, some of them included in this review (Avau 2021; Boluyt 2008; Boynnton 1998; Lee 2012; Salvador-Oliván 2021; White 2001; Wilczynski 2007).

The included studies developed filters for the following databases.

- MEDLINE (Ovid): [Boynnton 1998](#) ; [Lee 2012](#); [White 2001](#) ; [Wilczynski 2007](#)
- MEDLINE (PubMed): [Avau 2021](#); [Boluyt 2008](#); [Salvador-Oliván 2021](#); [Shojania 2001](#)
- Embase (Ovid): [Lee 2012](#) ; [Wilczynski 2007](#)
- Embase (via Elsevier): [Avau 2021](#)

The main objective of the studies that developed filters for MEDLINE was to improve the performance in the retrieval of systematic reviews by improving sensitivity, specificity and precision through different search approaches. These approaches included manual or electronic searches for selecting relevant terms.

The Ovid platform was used for the two Embase filters (Lee 2012; Wilczynski 2007). The declared objective was to recover systematic reviews, improving precision and sensitivity in the search strategy.

Each study may have developed and validated a variable number of filters. Moreover, a filter might have been tested in more than one dataset and compared with other filters. See Table 1 and Table 2 (below) for a summary of databases, the gold standards/quasi-gold standards and the types of validation and comparisons of the included studies.

Methods and platform for the development of the search filter

The following methods were used to develop the search filter.

- Analyses of terms frequency in MEDLINE (Boynton 1998) or a pre-defined set of citations using text mining software (White 2001)
- Expert input (Boynton 1998)
- Previously validated filters (Lee 2012)
- A combination of relevant publication types with title and text words typically found in systematic reviews (Shojania 2001)
- Using methodological search terms and phrases, including indexing terms and text words from clinical articles from a subset of journal articles (Wilczynski 2007)

The terms used in the search strategies were searched using various field labels, for example, abstract or subject heading.

Gold standard/Quasi-Gold Standard

Gold standards were constructed in various ways. Some of the gold standard sets of records were produced by different components such as indexing terms and keywords referring to systematic review methods (Lee 2012), handsearching in different journals (White

2001), or by guidance from clinicians and librarians (Wilczynski 2007).

Other studies used a *quasi-gold standard*, i.e. a database in which the authors identified systematic reviews, but did not classify the other records as non-systematic reviews. In these studies, specificity could not be assessed, only sensitivity and precision (Avau 2021; Boynton 1998; Shojania 2001).

Validation

One study (Lee 2012) performed external validation with a dataset of indexed articles. Another study (Wilczynski 2007), performed two external validations: One with a set of records that included the *Cochrane Database of Systematic Reviews* and one without it. Both search filters were validated internally, but only one (Lee 2012) was validated internally and externally. Three studies did not perform external validation (Boynton 1998; Shojania 2001; White 2001).

Comparisons to other filters

Most developed filters compared themselves with other available filters, such as the Hunt and K.A. McKibbin full and brief strategy (Hunt 1997), and the Centre for Reviews and Dissemination (CRD) brief and full strategy (NHS 1996). Some of the included filters in this review were compared with another filter also included in this review. Only one study (Shojania 2001) did not compare their filter with other commonly used filters.

Outcome measures

Six studies reported the sensitivity of their developed filter (Boluyt 2008; Boynton 1998; Lee 2012; Shojania 2001; White 2001; Wilczynski 2007). Seven studies reported precision (Avau 2021; Boluyt 2008; Boynton 1998; Lee 2012; Salvador-Oliván 2021; Shojania 2001; Wilczynski 2007), and six studies (Avau 2021; Boynton 1998; Lee 2012; Shojania 2001; White 2001; Wilczynski 2007) reported specificity. Only one study reported positive predictive value (Shojania 2001).

Table 2. Summary characteristics of search filters.

Author/year	Database (platform)	Gold Standard	Type of validation	Comparisons to other filters	Available outcome measures
Boynton 1998	MEDLINE (Ovid)	Quasi-gold standard handsearch and electronic search (n = 288 records)	Internal	Hunt and McKibbin full and brief (Hunt 1997) Centre for Reviews and Dissemination full and brief (NHS 1996)	Sensitivity, specificity and precision
Shojania 2001*	MEDLINE (PubMed)	Gold standard 1: systematic reviews from DARE (n = 100 records) Gold standard 2: handsearched from <i>AC J Club</i> (n=104 records)	External	None	Sensitivity, specificity, precision, positive predictive value

White 2001	MEDLINE (Ovid)	1 GS handsearch; QGS (SR): n = 110 records, Non SR: n = 110 records, Non review: n = 125 records)	Internal / External	Boynton 1998 NHS 1996	Sensitivity and specificity
Wilczynski 2007	Embase (Ovid)	Gold standard 1: handsearch 55 journals (n = 27,769 records)	Internal	None	Sensitivity, Specificity, Accuracy and Precision
	MEDLINE (Ovid)	Gold standard 1: handsearch journals (n = 27,769 records)	Internal / External	Hunt 1997 NHS 1996	Sensitivity, specificity and precision
		Gold standard 2: Validation dataset - CDSR (n = 10,446 records)		Shojania 2001	
		Gold standard 3: full validation database (n = 49,028 records)		Wilczynski 2010	
Boluyt 2008*	MEDLINE(PubMed)	GS. Was established by search for systematic reviews of children's health in DARE and by manually searching for various magazines for systematic reviews	External	Shojania 2001 , Boynton 1998 , White 2001 , Montori 2005 , Cochrane Child Health Field 2006	Sensitivity and Precision
Lee 2012*	MEDLINE (Ovid)	1 GS for each database Manual screening (n = 387 records)	Internal / External	Boynton 1998	Sensitivity, Specificity, Precision, NNR
	Embase (Ovid)			BMJ Clinical Evidence	
				Hunt 1997	
				NHS 1996	
				SIGN	
				Shojania 2001 Wilczynski 2007 Hunt 1997	
Avau 2021*	MEDLINE (PubMed)	1 Quasi GS for MEDLINE (PubMed) (n = 77 records);	External validation.	SIGN	Specificity, Precision, NNR
	Embase (via Elsevier)	1 Quasi GS for Embase (Embase.com) (n = 70 records)			
Salvador-Oliván 2021	MEDLINE (PubMed)	Does not develop GS	External validation.	Shojania 2001	Precision, NNR
			PubMed SR filter		

Footnote: QGS: Quasi-Gold Standard; CDSR: Cochrane Database of Systematic Reviews; DARE: Database of Abstracts of Reviews of Effect; NNR: number needed to read. * Filter developed for a specific clinical area (first aids in [Avau 2021](#), child health in [Boluyt 2008](#), public health in [Lee 2012](#) and multi-topic such as colorectal

cancer, thrombolytic therapy for venous thromboembolism, and treatment of dementia in [Shojania 2001](#))

Excluded studies

We excluded 44 records after full-text selection.

- We excluded 25 due to an incorrect study design that did not focus on the methodological development of the filters.
- We excluded 15 because the study approach was incorrect; they did not report search strategies to retrieve systematic reviews on PubMed or Embase.
- Three records focused on diagnostic test accuracy studies.
- One study focused on overviews.

See [Characteristics of excluded studies](#) for more details.

Risk of bias in included studies

All studies stated their objectives clearly, identifying the focus of their research and the database and interface used. The focus of the filters was also clearly reported. Four studies described additional specific topics for their filter's focus (for example, colorectal cancer, thrombolytic therapy for venous thromboembolism, and treatment of dementia in [Shojania 2001](#) or public health in [Lee 2012](#)).

The identification of the number of gold standards or quasi-gold standards of known relevant records was clearly reported, when applicable. The size of the gold standard was not reported in some studies ([Lee 2012](#); [Shojania 2001](#)). The size of gold standards was relatively small, ranging from 70 to 387 for most of the studies, with the exception of [Wilczynski 2007](#), which screened 27,769, 10,446 and 49,028 records for each gold standard to evaluate the search filters.

The identification of the search terms incorporated in the filters was highly variable across studies and also not clearly reported.

Internal validity testing was reported in four studies ([Boynton 1998](#); [Lee 2012](#); [White 2001](#); [Wilczynski 2007](#)). Most studies reported several tested strategies, with various performance measures for all strategies, although some studies only reported a single value for this performance measure, with no measure of variance reported.

External validity testing was performed in seven studies ([Avau 2021](#); [Boluyt 2008](#); [Lee 2012](#); [Salvador-Oliván 2021](#); [Shojania 2001](#); [White 2001](#); [Wilczynski 2007](#)). In these studies, various strategies were tested, and several performance measures were also reported.

Most studies addressed their limitation in the discussion sections, with no additional potential limitations identified when appraised by the review team. Most studies also compared their filters with other available filters, but the references for the compared filters were not consistently reported in the studies.

See Additional [Table 2](#); [Table 3](#); [Table 4](#); [Table 5](#); [Table 6](#); [Table 7](#); [Table 8](#); [Table 9](#) for a detailed assessment of the InterTASC appraisal tool.

Effect of methods

See [Summary of findings 1](#) for the main results of our review. See [Table 1](#) for a detail of the performance of each filter.

Search filters for MEDLINE

We identified eight studies that developed filters for this database ([Avau 2021](#); [Boluyt 2008](#); [Boynton 1998](#); [Lee 2012](#); [Salvador-Oliván 2021](#); [Shojania 2001](#); [White 2001](#); [Wilczynski 2007](#)). Most filters were developed between 1995 and 2008. Some studies did not

report dates of filter development and evaluation. Three studies produced filters with sensitivity >90%, with variable degrees of precision ([Boynton 1998](#); [Shojania 2001](#); [Wilczynski 2007](#)), and only one of them was developed and validated against a gold-standard dataset that included both class-of-interest and non-class-of-interest records, which allowed the calculation of specificity ([Wilczynski 2007](#)). The other two search filters had lower levels of sensitivity ([Lee 2012](#); [White 2001](#)). One of these produced a filter with higher levels of specificity (>90%, [White 2001](#)). All filters showed similar sensitivity and precision in the external validation, except for [Boynton 1998](#), which was not externally validated and [Shojania 2001](#), which was conceptually derived and only externally validated.

The report of these studies had some limitations (see [Risk of bias in included studies](#)), and the assessments of their accuracy may suffer from indirectness, as some filters were developed for retrieving systematic reviews of specific topics. (see [Overall completeness and applicability of evidence](#)).

Search filters for Embase

We identified three studies that developed filters for this database ([Avau 2021](#); [Lee 2012](#); [Wilczynski 2007](#)). Most filters were developed between 2000 and 2008. One of these studies developed filters with variable sensitivity and precision, including highly sensitive strategies (> 90%); however, it was not externally validated ([Wilczynski 2007](#)). The other study produced a filter with a lower sensitivity (72.7%) but high specificity (99.1%) with a similar performance in the external validation.

The report of these studies had some limitations (see [Risk of bias in included studies](#)), and the assessments of their accuracy may suffer from indirectness, as some filters were developed for retrieving systematic reviews of specific topics. (see [Overall completeness and applicability of evidence](#)).

DISCUSSION

Summary of main results

We identified eight studies that developed filters for MEDLINE and three studies that developed filters for Embase. Six included studies reported the sensitivity of their developed filter. Seven studies reported precision and six studies reported specificity. Only one study reported the number needed to read and positive predictive value. For MEDLINE, all filters showed similar sensitivity and precision, and one filter showed higher levels of specificity ([White 2001](#)). For Embase, filters showed variable sensitivity and precision, with limited study reports that may affect accuracy assessments. Three studies designed filters to retrieve systematic reviews on a specific topic (e.g. first aid, child health), and their performance and publication were limited to one context.

Overall completeness and applicability of evidence

The included studies used different search interfaces (Ovid, PubMed, embase.com). Most of them are over 15 years old, precluding being updated with current technology or databases and interfaces' features and raising concerns about their transferability and performance to the current platforms. Furthermore, most of the generic filters for systematic reviews (i.e. those not focused on a specific topic) were developed before the development of the PRISMA statement in 2009. The PRISMA

statement may have led to the improvement of the reporting of systematic reviews, including their identification as such in the title and abstract (Page 2016), which can affect the performance of the filters, improving sensitivity but perhaps losing precision due to the inclusion of non-specific terms in these search strategies for other types of non-systematic reviews.

Despite the difficulty of establishing an accurate comparison of the performance of a methodological filter, there is guidance to inform researchers, including information specialists, when deciding how to assess a filter and the available methods (Lefebvre 2017). Methods to search are evolving, and there is a rise in the use of machine-learning methods for the development of search strategies (Adam 2022); nonetheless, most of them were validated for the retrieval of randomised controlled trials, such as the RCT classifier (Thomas 2021).

We highlight three groups of filters that have been independently evaluated and reached our pre-defined threshold for sensitivity and precision, which users may prioritise when searching for systematic reviews in MEDLINE and Embase (Boynton 1998; Shojania 2001, Wilczynski 2007). However, many of the included filters were designed to retrieve systematic reviews on a specific topic, which may not be useful for generic approaches for researchers looking to retrieve systematic reviews outside these topics. Furthermore, the performance measures of these specific filters in generic fields are unknown and probably more limited.

Considering the uncertain performance of each of the filters in the current context, it is essential to highlight the tools that allow filtering and quick access to systematic reviews within the results obtained from a search, for example, "article type" in PubMed or "Evidence-Based Medicine/systematic reviews" for Embase.com. However, their performance in terms of sensitivity and precision in retrieving reports of systematic reviews is unknown. Epistemonikos (<https://www.epistemonikos.org/>) is a specialised information resource whose information is systematic reviews and clinical trials obtained from bibliographic databases such as Embase, MEDLINE and LILACS, amongst others (Rada 2020). Each tool provides good accessibility to systematic reviews, but their performance in terms of sensitivity and precision is unknown.

Quality of the evidence

Most studies had clearly stated their objectives and focus of the developed filter, in some cases including specific topics. The use of gold standards in the included studies was not clearly reported, with different numbers of records and no clear inclusion criteria. Furthermore, search terms included in the developed filters and their combination were not adequately reported or missing in some studies, without a clear development of the methods used to identify and combine search terms.

Regarding validity testing, the number of filters validated and sets of records used for the validation process were not clearly reported and rarely addressed in the limitations sections of the studies' discussion.

Most of the included studies compared their filters to other known and published filters, although outcome measures or performance measures were highly variable across studies. They were also inconsistently reported: some studies only reported a single value

for the selected performance measure, and other studies reported a range of values or a single value with a 95%CI.

Potential biases in the review process

As with most systematic reviews, there is always a possibility of missing important studies that may fulfil our inclusion criteria. However, we conducted a comprehensive search strategy in several databases, including specific databases for the topic of our research question. We also manually searched the reference list of included studies as well as the InterTASC Information Specialists' Sub-Group website.

The identification of relevant records for our review could have represented a barrier, as most studies' reports were highly variable, but we were cautious enough to involve independent screeners with experience in the systematic review process for the assessment of eligibility. We also conducted regular meetings to discuss and resolve disagreements.

As previously highlighted, the variability in studies' reports might have made the data extraction process cumbersome, as information to be extracted was not consistently reported across studies, if reported at all. We solved this issue by extracting relevant study data by two independent reviewers and by discussing all disagreements in our regular meetings.

Agreements and disagreements with other studies or reviews

Two Cochrane Reviews addressed search filters as a topic, aiming at systematically reviewing the development, evaluation and performance of several search strategies, but focusing on retrieving different study designs, like diagnostic test accuracy (DTA) or observational studies (Beynon 2013; Li 2019). Another non-Cochrane systematic review evaluated studies of electronic database search strategies designed to retrieve adverse effects data, identifying three studies with highly sensitive search strategies but with several methodological concerns related to the small number of records, the absence of a validation set of records for testing, and limited evaluation of precision or other performance measures (Golder 2009).

As in our review, most of the previously mentioned systematic reviews focused on search strategies developed for MEDLINE and Embase, limiting the applicability, implementation and generalizability of the identified search filters to be applied in other databases. Furthermore, most of the filters identified have not been adapted to the updated features of the databases they were developed for, raising concerns about their transferability and performance to the current platforms.

Nevertheless, the limited samples of articles retrieved and the unblinded classification of an article as a "systematic review" raise some concerns regarding the reported performance measures. Furthermore, the classification of this search strategy as a search filter does not fit into the definition applied in this review, nor the development, evaluation or comparison study types of interest for this review.

One study (Navarro-Ruan 2022) compared the recall of the PubMed/MEDLINE indexing term Systematic review as a publication type (SR[pt]) and the sensitive clinical query from one of our included studies (Wilczynski 2007). The authors found that SR[pt] retrieved a

smaller subset of records, of which 92% met the MeSH definition of systematic review, compared to 8% from the validated search filter. However, this study used an unblinded classification of an article as a "systematic review" and only verified these performance measures to a subset of the results (100/1000 for the indexing term and 50/253,613 for the search filter). Moreover, the authors highlight that the indexing term has limited coverage due to heterogeneous and delayed indexing processes, so this resource should be used with caution.

AUTHORS' CONCLUSIONS

Implication for systematic reviews and evaluations of healthcare

Studies reporting the development, evaluation, or comparison of search filters to retrieve reports of systematic reviews in MEDLINE showed similar sensitivity and precision, with one filter showing higher levels of specificity. For Embase, filters showed variable sensitivity and precision, with limited information about how the filter was produced, which leaves us uncertain about their performance assessments. Newer filters had limitations in their methods or scope, including very focused subject topics for their gold standards, limiting their applicability across other topics. None of the filters were designed to differentiate systematic reviews on the basis of their methodological quality. We highlight three groups of filters that have been independently evaluated and reached our pre-defined threshold for sensitivity and precision, which users may prioritise when searching for systematic reviews in MEDLINE and Embase ([Boynton 1998](#); [Shojania 2001](#), [Wilczynski 2007](#)), acknowledging their limitations in their development and lack of update or more recent evaluations of their performance.

Search strategies designed for retrieving systematic reviews may facilitate identifying information for clinical practice guidelines development, satisfying different stakeholders' needs for supporting healthcare decision-making. However, there is a need to continuously assess the performance of existing high-quality filters to retrieve systematic reviews across topics in MEDLINE and Embase. Moreover, updates on the combination of terms informed by such assessments may improve their accuracy. Therefore, guidelines developers and authors of overviews need to consider the limitations of existing filters when applying them to their search strategies.

Implication for methodological research

Our findings highlight that consensus guidance on the conduct of search filters and standardized reporting of search filters are

needed, as we found highly heterogeneous development methods, accuracy assessments and outcome selection.

New strategies adaptable across interfaces could enhance their usability. Moreover, updates to the existing filters are needed considering the impact of reporting guidelines, including the PRISMA 2009, on how systematic reviews are reported. Finally, future filter developments should also consider comparing the filters against a common reference set to establish comparative performance and assess the quality of systematic reviews retrieved by strategies.

ACKNOWLEDGEMENTS

We thank Laura Martínez García, who provided input during the protocol development. We thank Valeria Vietto, who contributed with screening, data extraction and quality assessment and writing the protocol for this review.

Editorial and peer-reviewer contributions

The following people conducted the editorial process for this article:

- Sign-off Editor (final editorial decision): Mike Clarke, Cochrane Methodology, Queen's University Belfast, UK
- Managing Editor (selected peer reviewers, collated peer-reviewer comments, provided editorial guidance to authors, edited the article): Helen Wakeford, Central Editorial Service.
- Editorial Assistant (conducted editorial policy checks and supported editorial team): Leticia Rodrigues, Central Editorial Service
- Copy Editor (copy editing and production): Heather Maxwell, Cochrane Central Production Service
- Peer-reviewers (provided comments and recommended an editorial decision: Ina Monsef, Cochrane Haematology, Department I of Internal Medicine, Center for Integrated Oncology Aachen Bonn Cologne Duesseldorf, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany (content review), Steve McDonald, Cochrane Australia (content review), Julie Glanville, Glanville.info, York, UK (content review).

One consumer peer reviewer provided content peer review but chose not to be publicly acknowledged.

REFERENCES

References to studies included in this review

Avau 2021 {published data only}

Avau B, Van Remoortel H, De Buck E. Translation and validation of PubMed and Embase search filters for identification of systematic reviews, intervention studies, and observational studies in the field of first aid. *Journal of the Medical Library Association* 2021;**109**(4):599-608. [DOI: [10.5195/jmla.2021.1219](https://doi.org/10.5195/jmla.2021.1219)] [PMID: 34858089]

Boluyt 2008 {published data only}

Boluyt N, Tjosvold L, Lefebvre C, Klassen TP, Offringa M. Usefulness of systematic review search strategies in finding child health systematic reviews in MEDLINE. *Archives of Pediatrics & Adolescent Medicine* 2008;**162**(2):111-6. [DOI: [10.1001/archpediatrics.2007.40](https://doi.org/10.1001/archpediatrics.2007.40)]

Boynton 1998 {published data only}

Boynton J, Glanville J, McDaid D, Lefebvre C. Identifying systematic reviews in MEDLINE: developing an objective approach to search strategy design. *Journal of Information Science* 1998;**24**(3):137-54. [DOI: [10.1177/016555159802400301](https://doi.org/10.1177/016555159802400301)]

Lee 2012 {published data only}

Lee E, Dobbins M, Decorby K, McRae L, Tirilis D, Husson H. An optimal search filter for retrieving systematic reviews and meta-analyses. *BMC Medical Research Methodology* 2012;**12**:51.

Salvador-Oliván 2021 {published data only}

* Salvador-Oliván JA, Marco-Cuenca G, Arquero-Avilés R. Development of an efficient search filter to retrieve systematic reviews from PubMed. *Journal of the Medical Library Association* 2021;**109**(4):561-74. [DOI: [10.5195/jmla.2021.1223](https://doi.org/10.5195/jmla.2021.1223)]

Shojania 2001 {published data only}

Shojania KG, Bero LA. Taking advantage of the explosion of systematic reviews: an efficient MEDLINE search strategy. *Effective Clinical Practice* 2001;**4**(4):157-62.

White 2001 {published data only}

White VJ, Glanville JM, Lefebvre C, Sheldon TA. A statistical approach to designing search filters to find systematic reviews: objectivity enhances accuracy. *Journal of Information Science* 2001;**27**(6):357-70.

Wilczynski 2007 {published data only}

Montori VM, Wilczynski NL, Morgan D, Haynes RB, Hedges Team. Optimal search strategies for retrieving systematic reviews from Medline: analytical survey. *BMJ* 2005;**330**(7482):68.

* Wilczynski NL, Haynes RB, Hedges Team. EMBASE search strategies achieved high sensitivity and specificity for retrieving methodologically sound systematic reviews. *Journal of Clinical Epidemiology* 2007;**60**(1):29-33.

Wong SSL, Wilczynski NL, Haynes RB. Comparison of top-performing search strategies for detecting clinically sound treatment studies and systematic reviews in MEDLINE

and EMBASE. *Journal of the Medical Library Association* 2006;**94**(4):451-5.

References to studies excluded from this review

Alexanderson 2004 {published data only}

Alexanderson K, Norlund A. Chapter 2. Methods used for the systematic literature search and for the review of relevance, quality, and evidence of studies. *Scandinavian Journal of Public Health, Supplement* 2004;**32**(63):31-5.

Assendelft 2001 {published data only}

Assendelft WJ, Scholten RJ, Hoving JL, Offringa M, Bouter LM. [Systematic reviews in practice. VIII. Searching and assessing systematic reviews].. *Nederlands Tijdschrift voor Geneeskunde* 2001;**145**(34):1625-31.

Bayliss 2008 {published data only}

Bayliss Sue E, Davenport C. Locating systematic reviews of test accuracy studies: how five specialist review databases measure up. *International Journal of Technology Assessment in Health Care* 2008;**24**(4):403-11.

Berg 2006 {published data only}

Berg A, Fleischer S, Behrens J. Development of two search strategies for literature in MEDLINE-PubMed: nursing diagnoses in the context of evidence-based nursing. *International Journal of Nursing Terminologies and Classifications* 2006;**16**(2):26-32.

Bethel 2014 {published data only}

Bethel A, Rogers M. A checklist to assess database-hosting platforms for designing and running searches for systematic reviews.. *Health Information and Libraries Journal* 2014;**31**(1):43-53.

Bikbov 2018 {published data only}

Bikbov B, Perico N, Remuzzi G, GBD Genitourinary Diseases Expert Group. A comparison of metrics and performance characteristics of different search strategies for article retrieval for a systematic review of the global epidemiology of kidney and urinary diseases.. *BMC Medical Research Methodology* 2018;**18**(1):110.

Booth 2016 {published data only}

Booth A. Searching for qualitative research for inclusion in systematic reviews: a structured methodological review.. *Systematic Reviews* 2016;**5**(101580575):74.

Bradley 2010 {published data only}

Bradley SM. Examination of the clinical queries and systematic review "hedges" in EMBASE and MEDLINE.. *Journal of the Canadian Health Libraries Association (JCHLA)* 2010;**31**(2):27-37.

Bramer 2018 {published data only}

Bramer Wichor M, Rethlefsen ML, Mast F, Kleijnen J. Evaluation of a new method for librarian-mediated literature searches for systematic reviews.. *Research synthesis methods* 2018;**9**(4):510-20.

Cals 2016 {published data only}

Cals Jochen WL, Kotz D. Literature review in biomedical research: useful search engines beyond PubMed. *Journal of Clinical Epidemiology* 2016;**71**(jce, 8801383):115-7.

Damarell 2019 {published data only}

Damarell R, May N, Hammond S, Sladek RM, Tieman JJ. Topic search filters: a systematic scoping review.. *Health Information and Libraries Journal* 2019;**36**(1):4-40.

Dickersin 1990 {published data only}

Dickersin K, Higgins K, Meinert CL. Identification of meta-analyses. The need for standard terminology. *Controlled Clinical Trials* 1990;**11**(1):52-66.

Dunikowski 1998 {published data only}

Dunikowski L. Searching for reviews. *Canadian Family Physician Medecin de Famille Canadien* 1998;**44**(blo, 0120300):115-6.

Eady 2008 {published data only}

Eady AM, Wilczynski NL, Haynes RB. PsycINFO search strategies identified methodologically sound therapy studies and review articles for use by clinicians and researchers. *Journal of Clinical Epidemiology* 2008;**61**(1):34-40.

El Sherif 2016 {published data only}

El Sherif R, Pluye P, Gore G, Granikov V, Hong Quan N. Performance of a mixed filter to identify relevant studies for mixed studies reviews. *Journal of the Medical Library Association : JMLA* 2016;**104**(1):47-51.

Fenichel 1980 {published data only}

Fenichel CH. The process of searching online bibliographic databases: a review of research. *Library Research* 1980;**2**(2):107-27.

Fyfe 2012 {published data only}

Fyfe T, Dennett L. Building capacity in systematic review searching: a pilot program using virtual mentoring. *Journal of the Canadian Health Libraries Association (JCHLA)* 2012;**33**(1):12-6.

Glanville 2000 {published data only}

Glanville J, Lefebvre C. Identifying systematic reviews: key resources. *ACP Journal Club* 2000;**132**(3):A11-2.

Greyson 2019 {published data only}

Greyson D, Rafferty E, Slater L, MacDonald N, Bettinger JA, Dube E, et al. Systematic review searches must be systematic, comprehensive, and transparent: a critique of Perman et al. *BMC Public Health* 2019;**19**(1):153.

Grummich 2014 {published data only}

Grummich K, Jensen K, Obst O, Seiler CM, Diener MK. Evidence-based medicine in surgical practice - locating clinical studies and systematic reviews by searching the Medline database. *Zentralblatt fur Chirurgie* 2014;**139** Suppl 2(y5i, 0413645):e116-23.

Harari 2020 {published data only}

Harari MB, Parola HR, Hartwell CJ, Riegelman A. Literature searches in systematic reviews and meta-analyses: A review, evaluation, and recommendations. *Journal of Vocational Behavior* 2020;**118**:103377. [DOI: [10.1016/j.jvb.2020.103377](https://doi.org/10.1016/j.jvb.2020.103377)]

Harbour 2014 {published data only}

Harbour J, Fraser C, Lefebvre C, Glanville J, Beale S, Boachie C, et al. Reporting methodological search filter performance comparisons: a literature review. *Health Information and Libraries Journal* 2014;**31**(3):176-94.

Hausner 2015 {published data only}

Hausner E, Guddat C, Hermanns T, Lampert U, Waffenschmidt S. Development of search strategies for systematic reviews: validation showed the noninferiority of the objective approach. *Journal of Clinical Epidemiology* 2015;**68**(2):191-9.

Hausner 2016 {published data only}

Hausner E, Guddat C, Hermanns T, Lampert U, Waffenschmidt S. Prospective comparison of search strategies for systematic reviews: an objective approach yielded higher sensitivity than a conceptual one. *Journal of Clinical Epidemiology* 2016;**77**(jce, 8801383):118-24.

Huang 2016 {published data only}

Huang Y Yang Z, Wang J, Zhuo L, Li Z, Zhan S. Performance of search strategies to retrieve systematic reviews of diagnostic test accuracy from the Cochrane Library. *Journal of Evidence-Based Medicine* 2016;**9**(8):77-83. [DOI: [10.1111/jebm.12200](https://doi.org/10.1111/jebm.12200)]

Huang 2017 {published data only}

Huang Y, Yang Z, Wang J, Zhuo L, Li Z, Zhan S. Erratum to: Performance of search strategies to retrieve systematic reviews of diagnostic test accuracy from the Cochrane Library: Performance of search strategies. *Journal of Evidence-Based Medicine* 2017;**10**(4):332-3. [DOI: [10.1111/jebm.12200](https://doi.org/10.1111/jebm.12200)]

Jenkins 2004 {published data only}

Jenkins M. Evaluation of methodological search filters-- a review.. *Health Information and Libraries Journal* 2004;**21**(3):148-63.

Lefebvre 2017 {published data only}

Lefebvre C, Glanville J, Beale S, Boachie C, Duffy S, Fraser CY, et al. Assessing the performance of methodological search filters to improve the efficiency of evidence information retrieval: five literature reviews and a qualitative study. *Health technology assessment (Winchester, England)* 2017;**21**(69):1-148.

Littlewood 2019 {published data only}

Littlewood A, Kloukos D. Searching the literature for studies for a systematic review. Part 3: Using controlled vocabulary. *American journal of orthodontics and dentofacial orthopedics : official publication of the American Association of Orthodontists, its constituent societies, and the American Board of Orthodontics* 2019;**155**(4):604-5.

Lunny 2016 {published data only}

Lunny C, McKenzie JE, McDonald S. Retrieval of overviews of systematic reviews in MEDLINE was improved by the

development of an objectively derived and validated search strategy. *Journal of Clinical Epidemiology* 2016;**74**:107-18.

Methley 2014 {published data only}

Methley AM, Campbell S, Chew-Graham C, McNally R, Cheraghi-Sohi S. PICO, PICOS and SPIDER: a comparison study of specificity and sensitivity in three search tools for qualitative systematic reviews.. *BMC Health Services Research* 2014;**14**(101088677):579.

Murdoch 2004 {published data only}

Murdoch NH. Searching for systematic reviews. *Canadian Oncology nNrsing Journal = Revue Canadienne de Nursing Oncologique* 2004;**14**(3):152-9.

Neilson 2019 {published data only}

Neilson C, Le Me-L. A failed attempt at developing a search filter for systematic review methodology articles in Ovid Embase. *Journal of the Medical Library Association : JMLA* 2019;**107**(2):203-9.

Premji 2020 {published data only}

Premji Z, Ganshorn H. Can database-level MEDLINE exclusion filters in Embase and CINAHL be used to remove duplicate records without loss of relevant studies in systematic reviews? An exploratory study. *Journal of the Canadian Health Libraries Association* 2020;**41**(1):3-15.

Salvador-Oliván 2018 {published data only}

Salvador-Oliván JA, Marco-Cuenca G, Arquero-Avilés R. Systematic reviews in Library and Information Science: analysis and evaluation of the search process [Las revisiones sistemáticas en Biblioteconomía y Documentación: análisis y evaluación del proceso de búsqueda]. *Revista Española de Documentación Científica* 2018;**41**(2):1-19. [DOI: [10.3989/redc.2018.2.1491](https://doi.org/10.3989/redc.2018.2.1491)]

Sindhu 1997 {published data only}

Sindhu F, Dickson R. Literature searching for systematic reviews. *Nursing Standard* 1997;**11**(41):40-2.

Thompson 2014 {published data only}

Thompson J, Davis J, Mazerolle L. A systematic method for search term selection in systematic reviews. *Research Synthesis Methods* 2014;**5**(2):87-97.

Underwood 2020 {published data only}

Underwood PG. Book review: Paul Levay and Jenny Craven (eds), *Systematic searching: Practical ideas for improving results*. *Journal of Librarianship & Information Science* 2020;**52**(3):943-4.

Volpato 2014 {published data only}

Volpato Enilze S N, Betini M, El Dib R. Testing search strategies for systematic reviews in the Medline literature database through PubMed. *Journal of Evaluation in Clinical Practice* 2014;**20**(2):117-20.

Volpato 2018 {published data only}

Volpato ES, Betini M, Puga ME, Agarwal A, Cataneo AJ, Oliveira LD, et al. Strategies to optimize MEDLINE and EMBASE

search strategies for anesthesiology systematic reviews. An experimental study (vol 135, pg 1, 2018). *Sao Paulo Medical Journal* 2018;**136**(1):98-98.

Wilczynski 2011 {published data only}

Wilczynski NL, McKibbin KA, Haynes RB. Sensitive Clinical Queries retrieved relevant systematic reviews as well as primary studies: an analytic survey. *Journal of Clinical Epidemiology* 2011;**64**(12):1341-9.

Wong 2006 {published data only}

Wong SS, Wilczynski NL, Haynes RB. Optimal CINAHL strategies for identifying therapy studies and review articles. *Scholarsh* 2006;**38**(2):194-9.

Wong 2006a {published data only}

Wong Sharon S L, Wilczynski Nancy L, Haynes R Brian. Optimal CINAHL search strategies for identifying therapy studies and review articles.. *Journal of nursing scholarship : an official publication of Sigma Theta Tau International Honor Society of Nursing* 2006;**38**(2):194-9.

Wong 2006b {published data only}

Wong SS, Wilczynski NL, Haynes RB. Optimal CINAHL search strategies for identifying therapy studies and review articles: Health policy and systems. *Journal of Nursing Scholarship* 2006;**38**(2):194-9.

Additional references

ACP journal club

ACP journal club. Purpose and Procedure. <https://www.acpjournals.org/journal/aim/acpj/purpose-and-procedure> (accessed 15 February 2023).

Adam 2022

Adam GP, Paynter R. Development of literature search strategies for evidence syntheses: pros and cons of incorporating text mining tools and objective approaches. *BMJ Evidence Based Medicine* 2022;**Epub ahead of print**:bmjebm-2021-111892. [DOI: [10.1136/bmjebm-2021-111892](https://doi.org/10.1136/bmjebm-2021-111892)]

Bachmann 2002

Bachmann LM, Coray R, Estermann P, Ter Riet G. Identifying diagnostic studies in MEDLINE: reducing the number needed to read. *Journal of the American Medical Informatics Association: JAMIA* 2002;**9**(6):653-8. [PMID: 12386115]

Bak 2009

Bak G, Mierzwinski-Urban M, Fitzsimmons H, Morrison A, Maden-Jenkins M. A pragmatic critical appraisal instrument for search filters: introducing the CADTH CAL. *Health Information and Libraries Journal* 2009;**26**(3):211-9.

Becker 2011

Becker LA, Oxman AD. Chapter 22: Overviews of reviews. In: Higgins JPT, Green S (editors), *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0* (updated March 2011). Available from training.cochrane.org/handbook.

Beynon 2013

Beynon R, Leeftang MM, McDonald S, Eisinga A, Mitchell RL, Whiting P, et al. Search strategies to identify diagnostic accuracy studies in MEDLINE and EMBASE. *Cochrane Database of Systematic Reviews* 2013, Issue 9. Art. No: MR000022. [DOI: [10.1002/14651858.MR000022.pub3](https://doi.org/10.1002/14651858.MR000022.pub3)]

BMJ Clinical Evidence

BMJ Clinical Evidence. Default search strategies used for BMJ Clinical Evidence. http://clinicalevidence.bmj.com/cweb/about/search_filters.jsp (accessed 10 May 2022).

Bradley 2010

Bradley SM. Examination of the Clinical Queries and Systematic Review “hedges” in EMBASE and MEDLINE. *Journal of the Canadian Health Libraries Association / Journal de l'Association des bibliothèques de la santé du Canada* 2010;**31**:27-37.

Bragge 2011

Bragge P, Clavisi O, Turner T, Tavender E, Collie A, Gruen RL. The Global Evidence Mapping Initiative: scoping research in broad topic areas. *BMC Medical Research Methodology* 2011;**11**:92. [PMID: 21682870]

Bramer 2017

Bramer WM, Rethlefsen ML, Kleijnen J, Franco OH. Optimal database combinations for literature searches in systematic reviews: a prospective exploratory study. *Systematic Reviews* 2017;**6**(1):245. [PMID: 29208034]

Brettelle 1998

Brettelle AJ, Long AF, Grant MJ, Greenhalgh J. Searching for information on outcomes: do you need to be comprehensive? *Quality in Health Care* 1998;**7**(3):163-7.

Chandler 2019

Chandler J, Cumpston M, Thomas J, Higgins JP, Deeks JJ, Clarke MJ. Chapter I: Introduction. In: *Cochrane Handbook for Systematic Reviews of Interventions* version 6.0 (updated August 2019). Cochrane, 2019. Available from www.training.cochrane.org/handbook.

Cochrane Child Health Field 2006

Cochrane Child Health Field. Usefulness of systematic review search strategies in finding child health systematic reviews in MEDLINE. childhealth.cochrane.org (accessed prior to 15 February 2023).

Cooper 2018

Cooper C, Varley-Campbell J, Booth A, Britten N, Garside R. Systematic review identifies six metrics and one method for assessing literature search effectiveness but no consensus on appropriate use. *Journal of Clinical Epidemiology* 2018;**99**:53-63. [PMID: 29526555]

Dhammi 2014

Dhammi IK, Kumar S. Medical subject headings (MeSH) terms. *Indian Journal of Orthopaedics* 2014;**48**(5):443-4. [PMID: 25298548]

Embase Indexing Guide 2020

Embase Indexing Guide 2020. https://www.elsevier.com/_data/assets/pdf_file/0010/901693/Embase-indexing-guide-2020.pdf (accessed 1 July 2020).

Glanville 2000

Glanville J, Lefebvre C. Identifying systematic reviews: key resources. *ACP Journal Club* 2000;**132**(3):A11-2.

Glanville 2008

Glanville J, 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. *Journal of the Medical Library Association* 2008;**96**(4):356-61.

Golder 2009

Golder S, Loke Y. Search strategies to identify information on adverse effects: a systematic review. *Journal of the Medical Library Association* 2009;**97**(2):84-92.

Haddaway 2022

Haddaway NR, Page MJ, Pritchard CC, McGuinness LA. PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis. *Campbell Systematic Reviews* 2022;**18**(2):e1230.

Higgins 2022a

Higgins JP, Lasserson T, Chandler J, Tovey D, Thomas J, Flemyng E, Churchill R. *Methodological Expectations of Cochrane Intervention Reviews*. London: Cochrane, February 2022.

Higgins 2022b

Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions* version 6.3 (updated February 2022). Cochrane, 2022. Available from www.training.cochrane.org/handbook.

Horsley 2011

Horsley T, Dingwall O, Sampson M. Checking reference lists to find additional studies for systematic reviews. *Cochrane Database of Systematic Reviews* 2011, Issue 8. Art. No: MR000026. [DOI: [10.1002/14651858.MR000026.pub2](https://doi.org/10.1002/14651858.MR000026.pub2)]

Hunt 1997

Hunt DL, McKibbin KA. Locating and appraising systematic reviews. *Annals of Internal Medicine* 1997;**126**(7):532-8.

Ioannidis 2016

Ioannidis JPA. Why most clinical research is not useful. *PLOS Medicine* 2016;**13**(6):e1002049.

IOM 2011

IOM (Institute of Medicine). *Clinical Practice Guideline We Can Trust*. Washington, DC: The National Academies Press, 2011.

Jenkins 2004

Jenkins M. Evaluation of methodological search filters-- a review. *Health Information and Libraries Journal* 2004;**21**(3):148-63.

Lefebvre 2022

Lefebvre C, Glanville J, Briscoe S, Littlewood A, Marshall C, Metzendorf M-I, et al. Chapter 4: Searching for and selecting studies. In: *Cochrane Handbook for Systematic Reviews of Interventions* version 6.3 (updated February 2022). Cochrane, 2022. Available from www.training.cochrane.org/handbook.

Lefebvre 2017

Lefebvre C, Glanville J, Beale S, Boachie C, Duffy S, Fraser C, et al. Assessing the performance of methodological search filters to improve the efficiency of evidence information retrieval: five literature reviews and a qualitative study. *Health Technology Assessment* 2017;**21**(60):1-148. [DOI: [10.3310/hta21690](https://doi.org/10.3310/hta21690).]

Leydesdorff 2016

Leydesdorff L, Comins JA, Sorensen AA, Bornmann L, Hellsten I. Cited references and Medical Subject Headings (MeSH) as two different knowledge representations: clustering and mappings at the paper level. *Scientometrics* 2016;**109**(3):2077-91. [PMID: 27942085]

Li 2019

Li L, Smith HE, Atun R, Tudor Car L. Search strategies to identify observational studies in MEDLINE and Embase. *Cochrane Database of Systematic Reviews* 2019, Issue 3. Art. No: MR000041. [DOI: [10.1002/14651858.MR000041.pub2](https://doi.org/10.1002/14651858.MR000041.pub2)]

Lipscomb 2000

Lipscomb CE. Medical Subject Headings (MeSH). *Bulletin of the Medical Library Association* 2000;**88**(3):265-6. [PMID: 10928714]

Moher 2009

Moher D, Liberati A, Tetzlaff J, Altman DG, the PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: the PRISMA Statement. *PLOS Medicine* 2009;**6**(7):e1000097.

Montori 2005

Montori VM, Wilczynski NL, Morgan D, Haynes RB, Hedges Team. Optimal search strategies for retrieving systematic reviews from Medline: analytical survey. *BMJ* 2005;**330**(7482):68.

Navarro-Ruan 2022

Navarro-Ruan T, Haynes RB. Preliminary comparison of the performance of the National Library of Medicine's systematic review publication type and the sensitive clinical queries filter for systematic reviews in PubMed. *Journal of the Medical Library Association* 2022;**110**(1):43-6.

NHS 1996

NHS Centre for Reviews and Dissemination. Undertaking systematic reviews of research on effectiveness: CRD guidelines for those carrying out or commissioning reviews. York: NHS CRD, 1996.

NLM 2019a

NLM - MeSH. Systematic Review [Publication Type]. <https://www.ncbi.nlm.nih.gov/mesh/2028176> (accessed 1 July 2020).

NLM 2019b

National Library of Medicine. PubMed updates February 2019. National Library of Medicine Technical Bulletin 2019;(427):b6.

NLM 2019c

National Library of Medicine. Support for Systematic Reviews. National Library of Medicine Technical Bulletin 2019;(427):b6.

Page 2016

Page MJ, Shamseer L, Altman DG, Tetzlaff J, Sampson M, Tricco AC, et al. Epidemiology and reporting characteristics of systematic reviews of biomedical research: a cross-sectional study. *PLOS Medicine* 2016;**13**(5):e1002028.

Page 2021

Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;**372**:n71. [DOI: [10.1136/bmj.n71](https://doi.org/10.1136/bmj.n71)]

Rada 2020

Rada G, Pérez D, Araya-Quintanilla F, Ávila C, Bravo-Soto G, Bravo-Jeria R, et al. Epistemonikos: a comprehensive database of systematic reviews for health decision-making. *BMC Medical Research Methodology* 2020;**20**(1):286. [DOI: [10.1186/s12874-020-01157-x](https://doi.org/10.1186/s12874-020-01157-x)] [PMID: 33256642]

Shojania 2001

Shojania KG, Bero LA. Taking advantage of the explosion of systematic reviews: an efficient MEDLINE search strategy. *Effective Clinical Practice* 2001;**4**(4):157-62. [PMID: 11525102]

SIGN

Scottish Intercollegiate Guidelines Network. SIGN: search filters. www.sign.ac.uk/search-filters.html (accessed 1 July 2020).

Sprakel 2019

Sprakel J, Carrara H, Manzer BM, Fedorowicz Z. A mapping study and recommendations for a joint NGO (Think Pink) and Bahrain Government Breast Cancer project. *Journal of Evidence-Based Medicine* 2019;**12**(3):209-17. [DOI: [10.1111/jebm.12357](https://doi.org/10.1111/jebm.12357)]

Thomas 2021

Thomas J, McDonald S, Noel-Storr A, Shemilt I, Elliott J, Mavergames C, Marshall IJ. Machine learning reduced workload with minimal risk of missing studies: development and evaluation of a randomized controlled trial classifier for Cochrane Reviews. *Journal of Clinical Epidemiology* 2021;**133**:140-51. [DOI: [10.1016/j.jclinepi.2020.11.003](https://doi.org/10.1016/j.jclinepi.2020.11.003)]

Wilczynski 1995

Wilczynski NL, Walker CJ, McKibbin KA, Haynes RB. Reasons for the loss of sensitivity and specificity of methodologic MeSH terms and textwords in MEDLINE. Proceedings. Symposium on Computer Applications in Medical Care 1995:436-40. [PMID: 8563319]

Wilczynski 2005

Wilczynski NL, Morgan D, Haynes RB. An overview of the design and methods for retrieving high-quality studies for clinical care.

BMC Medical Informatics and Decision Making 2005;**5**:20. [PMID: 15969765]

Wilczynski 2007

Wilczynski NL, Haynes RB, Hedges Team. EMBASE search strategies achieved high sensitivity and specificity for retrieving methodologically sound systematic reviews. *Journal of Clinical Epidemiology* 2007;**60**(1):29-33. [PMID: 17161751]

Wilczynski 2010

Wilczynski NL, Haynes RB, for the QI Hedges Team. Optimal search filters for detecting quality improvement studies in Medline. *Quality & Safety in Health Care* 2010;**19**(6):e31-5.

Wong 2006

Wong SS, Wilczynski NL, Haynes RB. Comparison of top-performing search strategies for detecting clinically sound treatment studies and systematic reviews in MEDLINE

and EMBASE. *Journal of the Medical Library Association* 2006;**94**(4):451-5.

Young 2011

Young T, Hopewell S. Methods for obtaining unpublished data. *Cochrane Database of Systematic Reviews* 2011, Issue 11. Art. No: MR000027. [DOI: [10.1002/14651858.MR000027.pub2](https://doi.org/10.1002/14651858.MR000027.pub2)]

References to other published versions of this review

Franco 2020

Franco JV, Garrote V, Vietto V, Escobar Liquitay CM, Solà I. Search strategies (filters) to identify systematic reviews in MEDLINE and Embase. *Cochrane Database of Systematic Reviews* 2020, Issue 7. Art. No: MR000054. [DOI: [10.1002/14651858.MR000054](https://doi.org/10.1002/14651858.MR000054)]

* Indicates the major publication for the study

CHARACTERISTICS OF STUDIES

Characteristics of included studies [ordered by study ID]

Avau 2021

Study characteristics	
Methods	<p>Objective</p> <p>To validate search filters for systematic reviews, intervention and observational studies translated from Ovid MEDLINE and Embase syntax to PubMed and Embase.com (Elsevier)</p>
Data	<p>A Quasi Gold Standard is developed for MEDLINE (PubMed) and a Gold Standard for Embase (Embase.com):</p> <ul style="list-style-type: none"> - Obtained from searches for the 2019 Sub-Saharan Africa Advanced First Aid Manual or 2020 updates to the Flanders, Belgium, or Sub-Saharan Africa Basic First Aid Guidelines. - Identified as a relevant systematic review, interventional study or observational study in the opinion of the evidence summary reviewer according to the predefined study selection criteria described in the CEBaP's methodological letter. - Originally retrieved without using a methodological search filter. Records from different searches were accrued until a minimum of 70 relevant publications of a specific study design were included in a gold standard.
Comparisons	<p>Internal Validation: this filter was not internally validated.</p> <p>External Validation:</p> <p>PubMed: the reference GS consisted of 77 systematic review references, collected in 33 evidence summaries on different first aid topics.</p> <p>Embase: The reference GS consisted of 70 systematic review references, collected in 35 evidence summaries</p>
Outcomes	<p>Precision and Specificity</p>
Notes	<p>Methodological search filters tested for systematic reviews</p> <p>MEDLINE (PubMed)</p>

Avau 2021 (Continued)

((("Meta-Analysis as Topic"[MeSH] OR meta analy*[TIAB] OR metaanaly*[TIAB] OR "Meta-Analysis"[PT] OR "Systematic Review"[PT] OR "Systematic Reviews as Topic"[MeSH] OR systematic review*[TIAB] OR systematic overview*[TIAB] OR "Review Literature as Topic"[MeSH]) OR (cochrane[TIAB] OR embase[TIAB] OR psychlit[TIAB] OR psyclit[TIAB] OR psychinfo[TIAB] OR psycinfo[TIAB] OR cinahl[TIAB] OR cinhal[TIAB] OR "science citation index"[TIAB] OR bids[TIAB] OR cancerlit[TIAB]) OR (reference list*[TIAB] OR bibliograph*[TIAB] OR hand-search*[TIAB] OR "relevant journals"[TIAB] OR manual search*[TIAB]) OR ((("selection criteria"[TIAB] OR "data extraction"[TIAB]) AND "Review"[PT])) NOT ("Comment"[PT] OR "Letter"[PT] OR "Editorial"[PT] OR ("Animals"[MeSH] NOT ("Animals"[MeSH] AND "Humans"[MeSH])))

PubMed search filter for systematic reviews, which resulted in a recall of 90%, specificity of 97%, and precision of 9.7%

Embase (Embase.com)

((('meta analysis (topic)'/exp OR 'meta analysis'/exp OR (meta NEXT/1 analy*):ab,ti OR metaanaly*:ab,ti OR 'systematic review (topic)'/exp OR 'systematic review'/exp OR (systematic NEXT/1 review*):ab,ti OR (systematic NEXT/1 overview*):ab,ti) OR (cancerlit:ab,ti OR cochrane:ab,ti OR embase:ab,ti OR psychlit:ab,ti OR psyclit:ab,ti OR psychinfo:ab,ti OR psycinfo:ab,ti OR cinahl:ab,ti OR cinhal:ab,ti OR 'science citation index':ab,ti OR bids:ab,ti) OR ((reference NEXT/1 list*):ab,ti OR bibliograph*:ab,ti OR hand-search*:ab,ti OR (manual NEXT/1 search*):ab,ti OR 'relevant journals':ab,ti) OR (('data extraction':ab,ti OR 'selection criteria':ab,ti) AND review/it)) NOT (letter/it OR editorial/it OR ('animal'/exp NOT ('animal'/exp AND 'human'/exp)))

Recall 91%, specificity 96% and precision 5,4%.

Boluyt 2008
Study characteristics

Methods	Objective
	<p>To determine the sensitivity and precision of existing search strategies for retrieving child health systematic reviews (SRs) in MEDLINE using PubMed.</p> <p>A GOLD standard was developed, and the standard reference set of SRs was established by searching children's health SRs in DARE and manually searching SRs in various journals.</p>
Data	<p>All titles and abstracts in DARE (Cochrane Library, Issue 2, 2004) were searched for SRs of child health also indexed in MEDLINE.</p> <p>Using a hand search of 7 MEDLINE-indexed pediatric journals All issues of each journal were searched for the following 5 years: 1994, 1997, 2000, 2002, and 2004</p>
Comparisons	Shojania 2001, Boynton 1998, White 2001, Montori 2005, PubMed Plus
Outcomes	Sensitivity-maximising Precision-maximising
Notes	Systematic Review search strategies Shojania ((meta-analysis [pt] OR meta-analysis [tw] OR metaanalysis [tw]) OR ((review [pt] OR guideline [pt] OR consensus [ti] OR guideline* [ti] OR literature [ti] OR overview [ti] OR review [ti]) AND ((Cochrane [tw] OR Medline [tw] OR CINAHL [tw] OR (National [tw] AND Library [tw])) OR (handsearch* [tw] OR search* [tw] OR searching [tw]) AND (hand [tw] OR manual [tw] OR electronic [tw] OR bibliographi* [tw] OR database* OR (Cochrane [tw] OR Medline [tw] OR CINAHL [tw] OR (National [tw] AND Library [tw]))) OR ((synthesis [ti] OR overview [ti] OR review [ti] OR survey [ti]) AND (systematic [ti] OR critical

Search strategies (filters) to identify systematic reviews in MEDLINE and Embase (Review)

Boluyt 2008 (Continued)

[ti] OR methodologic [ti] OR quantitative [ti] OR qualitative [ti] OR literature [ti] OR evidence [ti] OR evidence-based [ti])) BUTNOT (case* [ti] OR report [ti] OR editorial [pt] OR comment [pt] OR letter [pt])

Boynton (Most Sensitive Strategy)

((Meta[tiab] NOT meta[ti]) OR (synthesis[tiab] NOT synthesis[ti]) OR (literature[tiab] NOT literature[ti]) OR (randomized[tw] NOT randomized[tiab]) OR published[tiab] OR Meta-Analysis[ptyp] OR extraction[tiab] OR (Trials[tw] NOT Trials[ti]) OR (controlled[tw] NOT controlled[tiab]) OR (search[tiab] NOT search[ti]) OR (medline[tiab] NOT medline[ti]) OR (selection[tiab] NOT selection[ti]) OR (sources[tiab] NOT sources[ti]) OR (review[tiab] NOT review[ti]) OR review[ptyp] OR articles[tiab] OR (reviewed[tiab] NOT reviewed[ti]) OR (english[tiab] NOT english[ti]) OR (language[tiab] NOT language[ti])) NOT (Letter[ptyp] OR comment[ptyp] OR editorial[ptyp])

White 1 (Most Sensitive Strategy)

((Controlled[tiab] NOT controlled[ti]) OR (design[tiab] NOT design[ti]) OR (evidence[tiab] NOT evidence[ti]) OR (extraction[tiab] NOT extraction[ti]) OR “randomized controlled trials”[MeSH] OR Meta-Analysis[ptyp] OR Review[ptyp] OR (sources[tiab] NOT sources[ti]) OR (studies[tiab] NOT studies[ti])) NOT (Letter[ptyp] OR comment[ptyp] OR editorial[ptyp]) White 2 (Most Precise Strategy) ((Review[tiab] NOT review[ti]) OR Review[ptyp] OR meta-analysis[tiab] OR Meta-Analysis[ptyp]) NOT (Letter[ptyp] OR comment[ptyp] OR editorial[ptyp]) Montori 1 (Most Sensitive Strategy) search*[tiab] OR meta-analysis[ptyp] OR meta-analysis[tiab] OR meta analysis[MeSH] OR review[ptyp] OR diagnosis[MeSH Sub-heading] OR associated[tiab]

Montori 2 (Most Precise Strategy)

Medline[tiab] OR (systematic[tiab] AND review[tiab]) OR meta-analysis[ptyp]

Montori 3 (Minimizing the Difference Between Sensitivity and Specificity)

Meta-analysis[ptyp] OR meta-analysis[tiab] OR meta-analysis[MeSH] OR review[ptyp] OR search*[tiab]

Montori 4 (Combining Most Precise Term With Most Sensitive Terms) Cochrane Database Syst Rev [ta] OR search[tiab] OR meta-analysis[ptyp] OR; medline[tiab] OR (systematic[tiab] AND review[tiab])

PubMed

(Systematic review* [tiab] OR systematic literature review* OR meta-analysis [ptyp] OR meta-analysis [ti] OR metaanalysis [ti] OR meta-analyses [ti] OR evidence-based medicine OR (evidence-based AND (guideline [tiab] OR guidelines [tiab] OR recommendations)) OR (evidenced-based AND (guideline [tiab] OR guidelines [tiab] OR recommendation*)) OR consensus development conference [ptyp] OR health planning guidelines OR guideline[ptyp] OR cochrane database syst rev OR acp journal club OR health technol assess OR evid rep technol assess summ OR evid based dent OR evid based nurs OR evid based ment health OR clin evid) OR ((systematic [tiab] OR systematically OR critical [tiab] OR (study [tiab] AND selection [tiab]) OR (predetermined OR inclusion AND criteri*) OR exclusion criteri* OR “main outcome measures” OR “standard of care” OR “standards of care”) AND (survey [tiab] OR surveys [tiab] OR overview* OR review [tiab] OR reviews [tiab] OR search* OR handsearch OR analysis [tiab] OR critique [tiab] OR appraisal OR (reduction AND risk AND (death OR recurrence))) AND (literature [tiab] OR articles [tiab] OR publications [tiab] OR publication [tiab] OR bibliography [tiab] OR bibliographies [tiab] OR published [tiab] OR unpublished OR citation OR citations OR database [tiab] OR internet [tiab] OR textbooks [tiab] OR references OR trials [tiab] OR metaanalysis [mh] OR (clinical [tiab] AND studies [tiab]) OR treatment outcome)) NOT (case report [ti] OR editorial [ti] OR editorial [ptyp] OR letter [ptyp] OR newspaper article [ptyp])

Child Search Strategy

Infant[MeSH] OR Infant* OR infancy OR Newborn* OR Baby* OR Babies OR Neonat* OR Preterm* OR Prematur* OR Postmatur* OR Child[MeSH] OR Child* OR Schoolchild* OR School age* OR Preschool* OR Kid or kids OR Toddler* OR Adolescent[MeSH] OR Adoles* OR Teen* OR Boy* OR Girl* OR Minors[MeSH] OR Minors* OR Puberty[MeSH] OR Pubert* OR Pubescen* OR Prepubescen* OR Pediatrics[MeSH] OR Paediatric* OR Paediatric* OR Peadiatric* OR Schools[MeSH] OR Nurs

Boynton 1998
Study characteristics

Methods	<p>Objective</p> <p>To develop a search filter using word frequency analysis ('objective approach') tested on the MEDLINE(Ovid). This method was validated with the use of expert input using the word frequency word list produced by ListIndex.</p>
Data	<p>Quasi-gold standard: produced by using the Institute for Scientific Information (ISI) citation index impact factors to identify six journals from the top ten high-impact factor journals in the 'Medicine general and internal' category:</p> <ul style="list-style-type: none"> • Annals of Internal Medicine • Archives of Internal Medicine • BMJ • JAMA • Lancet • New England Journal of Medicine <p>Hand search from 1992 to 1995 to identify systematic reviews and search via MEDLINE(Ovid).</p>
Comparisons	<p>External validation: this filter was not externally validated</p> <p>Comparisons to other filters (non-validated): D.L. Hunt and K.A. McKibbin full and brief strategy (Hunt 1997 and Centre for Reviews and Dissemination (CRD) brief and full strategy (NHS 1996)</p>
Outcomes	<p>Sensitivity and precision (specificity could not be assessed due to the use of a quasi-gold standard)</p>
Notes	<p><i>Objectively derived filters</i></p> <p>Strategy A:</p> <p>review.pt.</p> <p>Strategy B:</p> <p>review.pt. data.ab. review tutorial.pt.* studies.ab. review.ab. clinical.ab. trials.ab. medline.ab. sources.ab. selection.ab.</p> <p>Strategy C:</p> <p>review.pt. review.ab.</p> <p>trials.ab. medline.ab. sources.ab. selection.ab.</p> <p>Strategy D:</p> <p>medline.ab.</p>

Boynton 1998 (Continued)

sources.ab.
selection.ab.

Strategy E:

medline.ab

Strategy F:

meta.ti.
meta-analysis.pt.*
medline.ab.
meta.ab.
extraction.ab.
english.ab.
synthesis.ab.
articles.ab.
search.ab.
language.ab.
selection.ab.
literature.ab.
review academic.pt.*
sources.ab.

Strategy H:

meta.ti.
meta-analysis.pt.
medline.ab.
meta.ab.
extraction.ab.
english.ab. review.pt.
synthesis.ab.
articles.ab.
search.ab.
language.ab.
selection.ab.
literature.ab.
review academic.pt.
sources.ab.
randomized.hw.
controlled.hw.
published.ab.
trials.ab.
analysis.ti.
relevant.ab.
trials.hw.
reviewed.ab.
review.ab.
review.pt.

Strategy J:

meta.ab.
synthesis.ab.
literature.ab.
randomized.hw.
published.ab.
meta-analysis.pt.
extraction.ab.
trials.hw.
controlled.hw.

Boynton 1998 (Continued)

search.ab.

medline.ab.

selection.ab.

sources.ab.

trials.ab.

review.ab.

review.pt.

articles.ab.

reviewed.ab.

english.ab.

language.ab.

*Filters developed with the input of experts***Strategy K:**

systematic adj

review\$.tw.

data adj synthesis.tw.

published adj studies.ab.

data adj extraction.ab.

meta-analysis/

meta-analysis.ti.

Strategy L:

systematic adj review\$.tw.

data adj synthesis.tw.

published adj studies.ab.

data adj extraction.ab.

meta-analysis/

meta-analysis.ti.

review.pt.

Strategy M:

meta-analy\$.ab.

data adj synthesis.ab.

literature adj3 review.ab.

randomized controlled trials/

published.ab.

meta-analysis.pt.

data adj extraction.ab.

articles.ab.

reviewed.ab.

search.ab.

Boynton 1998 (Continued)

english adj language.ab.

All strategies excluded comments, letters, editorials and animal studies using the following strategy:

- 1 comment.pt.
- 2 letter.pt.
- 3 editorial.pt.
- 4 animal/
- 5 human/
- 6 4 not (4 and 5)
- 7 {SELECTED SEARCH STRATEGY}
- 8 7 not (1 or 2 or 3 or 6)

Lee 2012
Study characteristics

Methods	<p>Objective</p> <p>Development of a search filter using a previously validated filter, which included the terms: MEDLINE.tw, systematic review.tw, meta-analysis.pt, combined with the Boolean OR operator. To customize this filter to retrieve only those systematic reviews of interventions, the term 'intervention' was added as an indexing term. This method was validated with the use of a gold standard and compared to other published systematic review filters.</p> <p>The authors modified the filter for use in Embase and CINAHL due to differences in indexing terms between the various databases.</p>
Data	<p>Gold standard: produced by using two distinct components: 1) indexing terms and keywords referring to systematic review methods, combined with the Boolean 'OR' operator (systematic, meta analysis, review); and 2) indexing terms and keywords referring to public health content areas, combined with the Boolean 'OR' operator (community health services, education, health education, health promotion, prevention, preventive). The content and methods components were then combined using the Boolean 'AND' operator.</p>
Comparisons	<p>External validation: data set of 219 articles indexed between January 2004 and December 2005.</p> <p>Comparisons to other filters: Boynton 1998 ; BMJ Clinical Evidence ; Hunt 1997 ; NHS 1996 ; SIGN ; Shojania 2001 ; Wilczynski 2007</p>
Outcomes	<p>Sensitivity, specificity, precision, number needed to read</p>
Notes	<p>MEDLINE</p> <ol style="list-style-type: none"> 1. MEDLINE.tw. 2. systematic review.tw. 3. meta-analysis.pt. 4. intervention\$.ti 5. or/1-4 <p>EMBASE</p> <ol style="list-style-type: none"> 1. MEDLINE.tw. 2. exp systematic review/ or systematic review.tw 3. meta-analysis/

Lee 2012 (Continued)

4. intervention\$.ti
5. or/1-4

Salvador-Oliván 2021
Study characteristics

Methods	Objective
	Develop a search filter to retrieve systematic reviews, using terms extracted from the titles of articles indexed as systematic review [pt] and different from those already in the PubMed SR filter.
Data	Terms extracted from the titles of articles indexed as systematic review [pt] and differing from those already in the PubMed SR filter.
Comparisons	The validation was compared with the Pubmed SR filter.
Outcomes	Sensitivity-maximising and precision-maximising
Notes	<p>#1 systematic[<i>sb</i>]</p> <p>#2 LETTER (PT) OR EDITORIAL (PT) OR COMMENT [PT] OR CASE REPORTS (PT) OR HISTORICAL ARTICLE (PT) OR REPORT (TI) OR PROTOCOL (TI) OR PROTOCOLS [TI] OR WITHDRAWN (TI) OR RETRACTION OF PUBLICATION [PT] OR RETRACTION OF PUBLICATION AS TOPIC (MESH) OR RETRACTED PUBLICATION [PT] OR REPLY [TI] OR PUBLISHED ERRATUM (PT)</p> <p>#3 (systematic* [ti] AND review [ti]) OR Systematic overview* [ti] OR Cochrane review* [ti] OR systemic review* [ti] OR scoping review [ti] OR scoping literature review [ti] OR mapping review [ti] OR Umbrella review* [ti] OR (review of reviews [ti] OR overview of reviews [ti]) OR meta-review [ti] OR (integrative review [ti] OR integrated review [ti] OR integrative overview [ti] OR meta synthesis [ti] OR metasynthesis [ti] OR quantitative review [ti] OR quantitative synthesis [ti] OR research synthesis (ti) OR meta-ethnography (ti) OR Systematic literature search [ti] OR Systematic literature research sti] OR meta-analyses [ti] OR metaanalyses [ti] OR metaanalysis (ti) OR meta-analysis (ti) OR meta-analytic review sti] OR meta-analytical review [ti] OR meta-analysis [pt] OR ((search* [tiab] OR medline [tiab] OR pubmed [tiab] OR embase [tiab] OR Cochrane (tiab) OR scopus (tiab) OR web of science stiab] OR sources of information (tiab) OR data sources [tiab] OR following databases [tiab]) AND (study selection (tiab) OR selection criteria [tiab] OR eligibility criteria (tiab) OR inclusion criteria (tiab) OR exclusion criteria [tiab]))</p> <p>#4 #1 NOT #2</p> <p>#5 #3 NOT #2</p> <p>#6 #5 NOT #4</p> <p>#7 #4 AND #5</p> <p>#8 #4 OR #5</p> <p>#9 (#3 OR systematic review [pt]) NOT #2</p> <p>#10 Systematic review [pt]</p> <p>#11 #10 NOT #5</p> <p>#12 #4 NOT #10</p> <p>#13 #5 NOT #10</p>

Shojania 2001

Study characteristics

Methods	<p>Objective</p> <p>Development of a search filter to retrieve systematic reviews (broad definition) in MEDLINE(PubMed). It was then tested in a subset of three clinical topics to assess the predictive properties.</p>
Data	<p>Quasi-gold standard: DARE database and ACP Journal Club (1999-2000)</p>
Comparisons	<p>External validation: this filter was developed by the authors (not derived from a database) and was tested on a gold standard.</p> <p>Comparisons to other filters: none</p>
Outcomes	<p>Sensitivity and positive predictive value (specificity could not be assessed due to the use of a quasi-gold standard)</p>
Notes	<p>Search strategy</p> <p>Sensitivity - DARE-indexed systematic reviews: 93% (95% CI 86% to 97%)</p> <p>Sensitivity - ACP Journal Club: 97% (95% CI 91% to 99%)</p> <p>1 meta-analysis[pt] OR meta-analysis[tw] OR metanalysis[tw]</p> <p>2 Cochrane[tw] OR Medline[tw] OR CINAHL[tw] OR (National[tw] AND Library[tw])</p> <p>3 (handsearch*[tw] OR search*[tw] OR searching[tw]) AND (hand[tw] OR manual[tw] OR electronic[tw] OR bibliographi*[tw] OR database* OR #2)</p> <p>4 (review[pt] OR guideline[pt] OR consensus[ti] OR guideline*[ti] OR literature[ti] OR overview[ti] OR review[ti]) AND (#2 OR #3)</p> <p>5 (synthesis[ti] OR overview[ti] OR review[ti] OR survey[ti]) AND (systematic[ti] OR critical[ti] OR methodologic[ti] OR quantitative[ti] OR qualitative[ti] OR literature[ti] OR evidence[ti] OR evidence-based[ti])</p> <p>6 (#1 OR #4 OR #5) NOT (case*[ti] OR report[ti] OR editorial[pt] OR comment[pt] OR letter[pt])</p> <p>The original search strategy mentioned the use of "BUTNOT", however the Pubmed platform does not currently have that operator (only "NOT")</p>

White 2001

Study characteristics

Methods	<p>Methods and platform for the development of the search filter</p> <p>Search strategies developed by three groups of records from 1995 and 1997 generated from five medical journals indexed in MEDLINE, using Wordstat and Simstat textual analysis software.</p>
Data	<p>Gold standard: hansearch from five journals <i>Annals of Internal Medicine</i>, <i>Archives of Internal Medicine</i>, <i>BMJ</i>, <i>JAMA</i>, and the <i>Lancet</i> . This included 110 systematic reviews, 110 reviews (not systematic), 125 non-review articles.</p>
Comparisons	<p>External validation: this filter was not externally validated</p> <p>Comparisons to other filters (non-validated): Boynton, Centre for Reviews and Dissemination.</p>

White 2001 (Continued)

Outcomes	Sensitivity and precision
Notes	<p><i>Model A and B (model A based on the frequency of the terms, model B based on the presence/absence of the terms)</i></p> <p>Controlled/AB Design/AB Evidence/AB Extraction/AB Randomized controlled trials/DE Meta-analysis/PT Review/PT Sources/AB Studies/AB</p> <p><i>Model C</i></p> <p>Review/AB Review/PT Meta-analysis/AB Meta-analysis/PT Meta-analysis/TI</p> <p><i>Model D</i></p> <p>Controlled/AB Extraction/AB Meta-analysis/PT Randomized controlled trials/DE Review/PT Selection/AB Studies/AB Study/AB</p>

Wilczynski 2007

Study characteristics	
Methods	<p>Objective</p> <p>MEDLINE search strategies developed by using methodological search terms and phrases in a subset of MEDLINE records matched with a handsearch of the contents of 161 journal titles for 2000.</p> <p>Embase search strategies were developed using a 55-journal subset chosen based on journals with the highest number of methodologically sound original studies and systematic reviews. An initial list of 5,385 search terms, including indexing terms and textwords from clinical articles were compiled.</p>
Data	<p>Gold standard: produced by guidance from clinicians and librarians, taking into account the Science Citation Index impact factors, as well as an iterative process of evaluation of over 400 journals for yield of studies and reviews of scientific merit and clinical relevance, a set of 161 clinical journals indexed in MEDLINE was defined.</p> <p>For Embase, a hand search of each article for each issue of 55 journal titles for the year 2000, chosen based on recommendations of clinicians and librarians, Science Citation Index Impact Factors provided by the Institute for Scientific Information, and ongoing assessment of their yield of studies and reviews of scientific merit and clinical relevance for different disciplines was developed.</p>
Comparisons	<p>External validation: With 2 databases of records including Cochrane Database of Systematic Reviews and without Cochrane Database of Systematic Reviews.</p>

Search strategies (filters) to identify systematic reviews in MEDLINE and Embase (Review)

Wilczynski 2007 (Continued)

Comparisons to other filters: [Hunt 1997](#) ; [NHS 1996](#) ; [Shojania 2001](#)

Outcomes	Sensitivity, specificity and precision
Notes	<p>Embase</p> <p>Best sensitivity: exp methodology OR search:.tw. OR review.pt.</p> <p>Best specificity: meta-analysis.tw. OR systematic review.tw.</p> <p>Small drop in specificity with a substantive gain in sensitivity: meta-analysis.tw. OR systematic review.tw. OR MEDLINE.tw.</p> <p>Best optimization of sensitivity and specificity: meta-analys:.mp. OR search:.tw. OR review.pt.</p> <p>MEDLINE</p> <p>Top sensitivity: search:.tw. or meta-analysis.mp,pt. or review.pt. or di.xs. or associated.tw.</p> <p>Top strategy minimising the difference between sensitivity and specificity: meta-analysis.mp,pt. or review.pt or search:.tw.</p> <p>Top precision: Medline.tw. or systematic review.tw. or meta-analysis.pt.</p>

CDSR : Cochrane Database of Systematic Reviews

Characteristics of excluded studies [ordered by study ID]

Study	Reason for exclusion
Alexanderson 2004	This paper focuses on methods for the retrieval of primary studies in systematic reviews.
Assendelft 2001	General guidance as to how to retrieve systematic reviews.
Bayliss 2008	This paper focuses on the location of systematic reviews of test accuracy in five specialist review databases
Berg 2006	This paper focuses on the retrieval of studies related to evidence-based nursing
Bethel 2014	This study does not develop a search strategy
Bikbov 2018	This paper focuses on methods for the retrieval of primary studies.
Booth 2016	This paper focuses on retrieving qualitative research.
Bradley 2010	This paper focuses on the examination of Clinical Queries hedges.
Bramer 2018	This paper describes the description of a novel method for designing search strategies.
Cals 2016	This paper is a narrative review describing several search engines.
Damarell 2019	This paper is a scoping review of topic search filters.
Dickersin 1990	This paper focuses on identifying terms related to meta-analyses.
Dunikowski 1998	This paper is a narrative review focusing on systematic reviews' searching.
Eady 2008	This paper focuses on retrieving primary studies and systematic reviews in PsycINFO.

Study	Reason for exclusion
El Sherif 2016	This paper focuses on retrieving primary mixed studies.
Fenichel 1980	This paper focuses on solutions to user interface problems in online search systems.
Fyfe 2012	This paper describes the results of a pilot mentoring program for systematic review searching.
Glanville 2000	This narrative review describes resources for identifying systematic reviews.
Greyson 2019	This paper is a correspondence response to another article.
Grummich 2014	This paper describes step-by-step instructions for locating clinical studies and systematic reviews in MEDLINE.
Harari 2020	This paper summarises studies assessing literature searches in systematic reviews and meta-analyses.
Harbour 2014	This paper focuses on performance measures reported in search filter development studies.
Hausner 2015	This paper compares different approaches for search strategies development.
Hausner 2016	This paper compares different approaches for search strategies development.
Huang 2016	This paper describes the performance of search strategies to retrieve systematic reviews of diagnostic test accuracy from the Cochrane Library.
Huang 2017	Focus on diagnostic accuracy reviews from the Cochrane Library
Jenkins 2004	This narrative review summarises different approaches to methodological search filter development.
Lefebvre 2017	This paper reports the methods used to assess the performance of methodological search filters.
Littlewood 2019	This paper describes the use of controlled vocabulary for searching studies for systematic reviews.
Lunny 2016	This study focused on the retrieval of overviews of systematic reviews.
Methley 2014	This paper describes a tool for retrieving qualitative research.
Murdoch 2004	This paper describes the literature search process in evidence-based medicine.
Neilson 2019	This paper describes the development of a search strategy for retrieving methodology articles.
Premji 2020	This paper focuses on strategies for diminishing duplicate records in systematic review searches.
Salvador-Oliván 2018	This paper focuses on methods for the retrieval of primary studies in systematic reviews.
Sindhu 1997	This paper focuses on methods for the retrieval of primary studies
Thompson 2014	This paper describes a novel method for selecting terms in systematic searches.
Underwood 2020	This paper describes approaches for improving search results in systematic reviews.
Volpato 2014	This paper assesses the effect of truncation on the number of results in MEDLINE.

Study	Reason for exclusion
Volpato 2018	This paper explores the use search terms in MEDLINE and Embase search strategies for anesthesiology systematic reviews.
Wilczynski 2011	This paper assesses the performance of Clinical Queries in retrieving primary studies.
Wong 2006	This paper focuses on methods for the retrieval of primary studies in systematic reviews.
Wong 2006a	This paper focuses on methods for the retrieval of primary studies in systematic reviews in CINAHL.
Wong 2006b	This paper focuses on methods for the retrieval of primary studies in systematic reviews in CINAHL.

ADDITIONAL TABLES

Table 1. Performance of each search filter

Author/year	Reference standard/database (interface)	Internal validity (with 95% confidence interval when available)	External validity (with 95% confidence interval when available)	External evaluations
Boynnton 1998	<p>Quasi-gold standard: handsearch in high impact journals and electronic searches (1992-1995)</p> <p>MEDLINE (Ovid)</p>	<p>8 filters using objective measures (A, B, C, D, E, F, H, J)</p> <p>3 filters using expert input (K, L, M)</p> <p>Sensitivity:</p> <p>A: 66%; B: 95%; C: 92%; D: 39%; E: 29%; F: 61%; H: 98%; J: 98%</p> <p>K: 55%; L: 89%; M: 58%</p> <p>Precision:</p> <p>A: 26%; B: 12%; C: 23%; D: 49%; E: 79%; F: 42%; H: 19%; J: 20%</p> <p>K: 71%; L: 31%; M: 37%</p> <p>Specificity: not reported</p> <p>Number needed to read^(a):</p> <p>A: 3.85; B: 8.33; C: 4.35; D: 2.04; E: 1.27; F: 2.38; H: 5.26; J: 5; K: 1.43; L: 3.23; M: 2.7</p>	No external validation	<p>By Lee 2012</p> <p>Sensitivity maximiser</p> <p>Sensitivity 99.5% (97.3 to 99.9)</p> <p>Specificity 75.6% (75.6 to 75.6)</p> <p>Precision 0.1% (0.1 to 0.1)</p> <p>Number needed to read 1395.1 (1387.7 to 1437.2)</p> <p>Precision query (> 70%)</p> <p>Sensitivity 47.8% (41.2 to 54.6)</p> <p>Specificity 99.6% (99.6 to 99.6)</p> <p>Precision 2.1% (1.8 to 2.5)</p> <p>Number needed to read 46.7 (40.9 to 54.4)</p> <p>By White 2001</p> <p>Most sensitive strategy</p> <p>Sensitivity 93.6%</p> <p>Precision 11.3%</p>
Shojania 2001	Quasi-gold standard 1 : system-	No internal validation	Gold standard 1:	<p>By Wilczynski 2007</p> <p>Sensitivity</p>

Table 1. Performance of each search filter (Continued)

	atic reviews from DARE (2000)			
	Quasi-gold standard 1 : handsearch ACP journal club (1999-2000)			
	MEDLINE (PubMed)			
		Sensitivity 93% (86 to 97)		90.0% (87.9 to 92.2)
			Specificity	
		Gold standard 2:		97.2% (97.0 to 97.4)
		Sensitivity 97% (91 to 99)		Precision
		Specificity : not reported		33.2% (31.2 to 35.2)
				Number needed to read ^(a) 3.01 (2.84 to 3.21)
				By Lee 2012
				Sensitivity 85.5% (80.1 to 89.7)
				Specificity 99.1% (99.1 to 99.1)
				Precision 1.7% (1.6 to 1.8)
				Number needed to read 57.8 (55.1 to 61.8)
				By Salvador-Oliván 2021
				Sensitivity 62% ^(b)
White 2001	Gold standard (1995 to 1997) handsearch from five journals <i>Annals of Internal Medicine</i> , <i>Archives of Internal Medicine</i> , <i>BMJ</i> , <i>JAMA</i> , and <i>The Lancet</i> . This included 110 systematic reviews, 110 reviews (not systematic), and 125 non-review articles.	5 filters (A, B, C, D, E) Sensitivity: A: 73.4%; B: 67.1%; C: 81.9%; D: 87.1%; E: 77.2% Specificity: A: 93.3%; B: 94.9%; C: 99.4%; D: 89.2%; E: 94.9%	Filter A: Sensitivity: 84.2% Specificity: 93.0%	None
MEDLINE (Ovid)				
Wilczynski 2007	Gold standard 1 : Handsearch 55 journals (2000) Embase (Ovid)	Best sensitivity Sensitivity 94.6% (91.5, 97.6) Specificity 63.7% (63.2, 64.3) Precision 2.0% (1.8, 2.3) Accuracy 64.0% (63.4, 64.5) NNR ^(a) : 50 Best specificity Sensitivity 61.4% (54.9, 67.8) Specificity 99.3% (99.2, 99.4) Precision 40.9% (35.6, 46.2)	No external validation	By Lee 2012 Best sensitivity Sensitivity 96.3% (90.8 to 98.5) Specificity 72.3% (72.3 to 72.3) Precision 0 (0 to 0) Number needed to read 2709.5 (2622.5 to 2945.2)

Table 1. Performance of each search filter *(Continued)*

Accuracy 99.0% (98.9, 99.1) Number needed to read ^(a) : 2.44 Small drop in specificity with a substantive gain in sensitivity Sensitivity 75.0% (69.3, 80.7) Specificity 98.5% (98.4, 98.7) Precision 29.2% (25.4, 32.9) Accuracy 98.4% (98.2, 98.5) Number needed to read ^(a) : 3.42 Best optimization of sensitivity and specificity Sensitivity 92.3% (88.7, 95.8) Specificity 87.7% (87.3, 88.1) Precision 5.6% (4.9, 6.4) Accuracy 87.7 (87.3, 88.1) Number needed to read ^(a) : 17.86	Small drop in specificity with a substantive gain in sensitivity Sensitivity 75.7% (66.7 to 82.8) Specificity 99.3% (99.3 to 99.3) Precision 1.1% (1 to 1.2) Number needed to read 88.2 (80.5 to 100.1) Best optimization of sensitivity and specificity Sensitivity 96.3% (90.8 to 98.5) Specificity 85.5% (85.5 to 85.5) Precision 0.1% (0.1 to 0.1) Number needed to read 1403.4 (1363.4 to 1502.0) Best specificity Sensitivity 63.4% (28.0 to 45.9) Specificity 99.5% (99.5 to 99.5) Precision 0.9% (0.7 to 1.1) Number needed to read 117.8 (93.4, 154.2)
---	--

Gold standard 1: Handsearch 55 journals (2000) Gold standard 2: Validation DS - CDST Gold standard 3: Full validation DB MEDLINE (Ovid)	Top sensitivity strategies Sensitivity 100% (97.3 to 100) Specificity 63.5% (62.5 to 64.4) Precision 3.41% (2.86 to 4.03) Number needed to read ^(a) : 29.33 Top strategy minimising the difference between sensitivity and specificity Sensitivity 92.5% (86.6 to 96.3) Specificity 93.0% (92.5 to 93.5) Precision 14.6% (12.3 to 17.2) Number needed to read ^(a) : 6.85 Top precision performer Sensitivity 75.2% (67.0 to 82.3)	Top sensitivity strategies Sensitivity 99.9 (99.6 to 100) Specificity 52.0% (51.6 to 52.5) Precision 3.14% (2.92 to 3.37) Number needed to read ^(a) : 31.84 Top strategy minimising the difference between sensitivity and specificity Sensitivity 98.0% (97.0 to 99.0)	By Lee 2012 Top sensitivity strategies Sensitivity 99.0% (96.5 to 99.7) Specificity 62.0% (62.0 to 62.0) Precision 0% (0 to 0) Number needed to read 2191.2 (2166.3 to 2284.3) Balanced query (sensitivity > specificity) Sensitivity 99.0% (96.5 to 99.7) Specificity 87.6% (87.6 to 87.6)
--	---	--	--

Table 1. Performance of each search filter *(Continued)*

Specificity 99.4% (99.2 to 99.5) Precision 60.2% (52.4 to 67.7) Number needed to read ^(a) : 1.66	Specificity 90.8% (90.5 to 91.1) Precision 14.2% (13.3 to 15.2) Number needed to read ^(a) : 7.04 Top precision performer Sensitivity 71.2% (68.0 to 74.4) Specificity 99.2% (99.1 to 99.3) Precision 57.1 (53.9 to 60.3) Number needed to read ^(a) : 1.75	Precision 0.1% (0.1 to 0.1) Number needed to read 712.4 (706.7 to 733.4) Balanced query (specificity > sensitivity) Sensitivity 87.9% (82.8 to 91.7) Specificity 98.5% (98.5 to 98.5) Precision 1.1% (1.0 to 1.1) Number needed to read 94.9% (90.9 to 100.9) Specific query Sensitivity 81.6% (75.8 to 86.3) Specificity 99.3% (99.3 to 99.3) Precision 2.0% (1.9 to 2.3) Number needed to read 49.4 (46.7 to 53.2)
---	---	--

Boluyt 2008
Gold standard.

Was established by searching for RS of children's health in DARE and by manually searching for various magazines for RS (1994 to 2004)

MEDLINE (Pubmed)

No internal validation

Sensitivity

1. Shojania+child 74% (69 to 78)
2. Boynton+child 95% (92 to 97)
3. White 1+child 93% (91 to 96)
4. White 2+child 94% (91 to 96)
5. Montori 1+child 96% (93 to 97)
6. Montori 2+child 68% (64 to 73)
7. Montori 3+child 94% (91 to 96)
8. Montori 4+child 72% (67 to 76)
9. Pubmed +child 76% (72 to 80)

Precision

1. Shojania+child 45% (36 to 55)

None

Table 1. Performance of each search filter (Continued)

					2. Boynton+child 3% (1 to 9) 3. White 2+child 2% (1 to 7) 4. Montori 2+child 45% (36 to 55) 5. Montori 3+child 3% (1 to 9) 6. Pubmed +child 32% (24 to 42) Number needed to read^(a) 1. Shojania+child 2.22 2. Boynton+child 33.3 3. White 2+child 50 4. Montori 2+child 2.22 5. Montori 3+child 33.33 6. Pubmed +child 3.13
Lee 2012	1 Gold standard for each database: Manual screening (2004-2005) MEDLINE (Ovid) Embase (Ovid)	MEDLINE Sensitivity 86.8% (75.2 to 93.5) Specificity 99.2% (99.2 to 99.2) Precision 1.1% (0.9 to 1.2) Number needed to read 91.6 (85.0 to 105.9)	MEDLINE Sensitivity 89.9% (85.0 to 93.3) Specificity 98.9% (98.9 to 98.9) Precision 1.4% (1.3 to 1.5) Number needed to read 71.4 (68.7 to 75.5)	None	
		EMBASE Sensitivity 72.7% (55.8 to 84.9) Specificity 99.1% (99.1 to 99.1) Precision 0.6% (0.4 to 0.7) Number needed to read 171.6 (146.7 to 224.6)	EMBASE Sensitivity 87.9% (80.2 to 92.8) Specificity 98.2% (98.2 to 98.2) Precision 0.5% (0.5 to 0.6) Number needed to read 186.0 (176.0 to 208.9)		
Avau 2021	1 Quasi-gold standard for MEDLINE (PubMed):	No internal validation	Specificity: 97% Precision: 9.7% Number needed to read: 10	None	

Table 1. Performance of each search filter (Continued)

	(2019-2020) The reference GS consisted of 77 SR references, collected in 33 evidence summaries on different first aid topics.			
	1 Quasi-gold standard for Embase (Embase.com):			
	(2019-2020) The reference GS consisted of 70 SR references, collected in 35 evidence summaries.		Specificity: 96%	
			Precision: 5.4%	
			Number needed to read: 19	
Salvador-Oliván 2021	No gold standard. A set of probable systematic reviews was Precisions identified using the PubMed search filter for systematic reviews. MEDLINE (Pubmed)	No internal validation	Precision* 83.8% (range 72.3 to 96.7%)	None
			Number needed to read^(a) 1.19	

Definition of outcome measures:

- Sensitivity: Proportion of systematic reviews that are correctly identified using the methodological filter.
- Specificity: Proportion of records that are not systematic reviews not identified using the methodological filter.
- Precision: Proportion of systematic reviews that are identified from all records retrieved using the methodological filter.
- Number needed to read (NNR): 1/precision

(a): Calculated from precision. *for "probable systematic reviews"; SR: systematic review. (b) This version of the filter was modified by PubMed and this number indicates the proportion of potential systematic reviews retrieved by the PubMed filter.

Table 2. InterTASC - Avau 2021

A. Information	
A.1 State the author's objective	Validate search filters for systematic reviews, intervention and observational studies translated from Ovid MEDLINE and Embase syntax to PubMed and Embase.com (Elsevier)
A.2 State the focus of the research.	Balance of sensitivity "number needed to read" (NNR)
A.3 Database(s) and search interface(s) .	MEDLINE (PubMed) and MEDLINE (Ovid) Embase (Embase.com)
A.4 Describe the methodological focus of the filter (e.g. RCTs).	Systematic reviews,

Table 2. InterTASC - Avau 2021 (Continued)

	Intervention studies, Observational studies
A.5 Describe any other topic that forms an additional focus of the filter (e.g. clinical topics such as breast cancer, geographic location such as Asia or population grouping such as paediatrics).	Yes, First aids
A.6 Other observations.	
B. Identification of a gold standard (GS) of known relevant records	
B.1 Did the authors identify one or more gold standards (GSs)?	1 GSs for MEDLINE (PubMed) 1 GSs for Embase (Embase.com)
B.2 How did the authors identify the records in each GS?	1) Obtained from searches developed for an evidence summary informing our 2019 Sub-Saharan Africa Advanced First Aid Manual or 2020 updates to our Flanders, Belgium, or Sub-Saharan Africa Basic First Aid Guidelines on PubMed or Embase. 2) Identified as a relevant systematic review, intervention study, or observational study as judged by the reviewer of the evidence summary according to predefined study selection criteria described in CEBaP's methodological charter 3) Originally retrieved without using a methodological search filter Records from different searches were accrued until a minimum of 70 relevant publications of a specific study design were included in a gold standard. For PubMed, the reference gold standard consisted of 77 systematic review references, collected in 33 evidence summaries on different first aid topics. For Embase, The reference gold standard consisted of 70 systematic review references, collected in 35 evidence summaries.
B.3 Report the dates of the records in each GS.	Not reported
B.4 What are the inclusion criteria for each GS?	Yes, detailed in Appendix B.2
B.5 Describe the size of each GS and the authors' justification, if provided (for example the size of the gold standard may have been determined by a power calculation)	- PubMed: The reference GS consisted of 77 SR references, collected in 33 evidence summaries on different first aid topics. - Embase: The reference GS consisted of 70 SR references, collected in 35 evidence summaries.
B.6 Are there limitations to the gold standard(s)?	Search dates limited and the filters evaluated in this paper were specifically tested on searches of evidence summaries used for first aid guideline projects.
B.7 How was each gold standard used?	To test external validity.
B.8 Other observations.	
C. How did the researchers identify the search terms in their filter(s) (select all that apply)?	
C.1 Adapted a published search strategy.	The systematic review filters tested were translated from existing filters from SIGN, designed for Ovid MEDLINE and Ovid Embase.

Table 2. InterTASC - Avau 2021 (Continued)

	<p>“The SR filters tested were translated from existing SIGN filters, designed for Ovid MEDLINE and Ovid Embase, into PubMed and Embase.com syntax.</p> <p>Adaptations were made to accommodate the SR-related index terms that were added to the PubMed MeSH tree and Embase Emtree after the development of the SIGN filters. For PubMed, "Systematic Review"[PT] and "Systematic Reviews as Topic"[MeSH] were included in the filter. For Embase, we included 'meta-analysis (topic)/exp, 'systematic review (topic)/exp, and 'systematic review'/exp in the filter."</p>
C.2 Asked experts for suggestions of relevant terms.	Not reported
C.3 Used a database thesaurus.	MEDLINE: MeSH Embase: Emtree
C.4 Statistical analysis of terms in a gold standard set of records (see B above).	Not reported
C.5 Extracted terms from the gold standard set of records (see B above).	Not reported
C.6 Extracted terms from some relevant records (but not a gold standard).	Not reported
C.7 Tick all types of search terms tested.	Subject headings Text words (e.g. in title, abstract) Publication types Subheadings
C.8 Include the citation of any adapted strategies.	Scottish Intercollegiate Guidelines Network. Search filters [Internet]. Edinburgh, UK: Healthcare Improvement Scotland [cited Nov 25 2020]. Available from: < https://www.sign.ac.uk/what-we-do/methodology/search-filters/ >.
C.9 How were the (final) combination(s) of search terms selected?	Not reported
C.10 Were the search terms combined (using Boolean logic) in a way that is likely to retrieve the studies of interest?	<p><i>Pubmed filter:</i></p> <p>((“Meta-Analysis as Topic”[MeSH] OR metaanaly*[TIAB] OR metaanaly*[TIAB] OR “Meta-Analysis”[PT] OR “Systematic Review”[PT] OR “Systematic Reviews as Topic”[MeSH] OR systematic review*[TIAB] OR systematic overview*[TIAB] OR “Review Literature as Topic”[MeSH]) OR (cochrane[TIAB] OR embase[TIAB] OR psychlit[TIAB] OR psyclit[TIAB] OR psychinfo[TIAB] OR psycinfo[TIAB] OR cinahl[TIAB] OR cinhal[TIAB] OR “science citation index”[TIAB] OR bids[TIAB] OR cancerlit[TIAB]) OR (reference list*[TIAB] OR bibliograph*[TIAB] OR hand-search*[TIAB] OR “relevant journals”[TIAB] OR manual search*[TIAB]) OR ((“selection criteria”[TIAB] OR “data extraction”[TIAB]) AND “Review”[PT])) NOT (“Comment”[PT] OR “Letter”[PT] OR “Editorial”[PT] OR (“Animals”[MeSH] NOT (“Animals”[MeSH] AND “Humans”[MeSH])))</p> <p><i>Embase Filter:</i></p>

Table 2. InterTASC - Avau 2021 (Continued)

	(('meta analysis (topic)/exp OR 'metaanalysis'/exp OR (meta NEXT/1 analy*):ab,ti OR metaanaly*:ab,ti OR 'systematic review (topic)/exp OR 'systematic review'/exp OR (systematic NEXT/1 review*):ab,ti OR (systematic NEXT/1 overview*):ab,ti) OR (cancerlit:ab,ti OR cochrane:ab,ti OR embase:ab,ti OR psychlit:ab,ti OR psyclit:ab,ti OR psychinfo:ab,ti OR psycinfo:ab,ti OR cinahl:ab,ti OR cinhal:ab,ti OR 'science citation index':ab,ti OR bids:ab,ti) OR ((reference NEXT/1 list*):ab,ti OR bibliograph*:ab,ti OR hand-search*:ab,ti OR (manual NEXT/1 search*):ab,ti OR 'relevant journals':ab,ti) OR (('data extraction':ab,ti OR 'selection criteria':ab,ti) AND review/it)) NOT (letter/it OR editorial/it OR ('animal'/exp NOT ('animal'/exp AND 'human'/exp)))
C.11 Other observations.	Not reported
D. Internal validity testing (This type of testing is possible when the search filter terms were developed from a known gold standard set of records).	
D.1 How many filters were tested for internal validity?	Not applicable
For each filter report the following information	
D.2 Was the performance of the search filter tested on the gold standard from which it was derived?	Not applicable
D.3 Report sensitivity data (a single value, a range, 'Unclear'* or 'not reported', as appropriate). *Please describe.	Not applicable
D.4 Report precision data (a single value, a range, 'Unclear'* or 'not reported' as appropriate). *Please describe.	Not applicable
D.5 Report specificity data (a single value, a range, 'Unclear'* or 'not reported' as appropriate). *Please describe.	Not applicable
D.6 Other performance measures reported.	Not applicable
D.7 Other observations.	Not applicable
E. External validity testing (This section relates to testing the search filter on records that are different from the records used to identify the search terms)	
E.1 How many filters were tested for external validity on records different from those used to identify the search terms?	2 filters. Reported Appendix 4
E.2 Describe the validation set(s) of records, including the interface.	PubMed: the reference GS consisted of 77 systematic review references, collected in 33 evidence summaries on different first aid topics. Embase: The reference GS consisted of 70 systematic review references, collected in 35 evidence summaries
E.3 On which validation set(s) was the filter tested?	Not reported.

Table 2. InterTASC - Avau 2021 (Continued)

E.4 Report sensitivity data for each validation set (a single value, a range or 'Unclear' or 'not reported', as appropriate).	Not reported
E.5 Report precision data for each validation set (report a single value, a range or 'Unclear' or 'not reported', as appropriate).	Single value: Pubmed 9.7% Embase 5,4%
E.6 Report specificity data for each validation set (a single value, a range or 'Unclear' or 'not reported', as appropriate).	Single value: Pubmed 97% Embase 96%
E.7 Other performance measures reported.	NNR – Single value: 10
E.8 Other observations.	
F. Limitations and comparisons.	
F.1 Did the authors discuss any limitations to their research?	Yes: <i>First:</i> the filters evaluated in this paper were specifically tested on searches of evidence summaries used for first aid guideline projects. <i>Second:</i> we use the relative recall technique which is one more "real world" technique application of the use of search filters that compose a reference gold standard by hand searching journals. <i>Third:</i> systematic searches for guidelines CEBaP projects are more pragmatic than searches for RS s in the sense that they try to balance methodological rigor with the time constraints associated with guide production.
F.2 Are there other potential limitations to this research that you have noticed?	Not reported
F.3 Report any comparisons of the performance of the filter against other relevant published filters (sensitivity, precision, specificity or other measures).	Not reported
F.4 Include the citations of any compared filters.	Not reported
F.5 Other observations and / or comments.	Not reported
G. Other comments. This section can be used to provide any other comments. Selected prompts for issues to bear in mind are given below.	
G.1 Have you noticed any errors in the document that might impact on the usability of the filter?	Not reported
G.2 Are there any published errata or comments (for example in the MEDLINE record)?	Not reported
G.3 Is there public access to pre-publication history and / or correspondence?	Not reported

Table 2. InterTASC - Avau 2021 (Continued)

G.4 Are further data available on a linked site or from the authors?	Not reported
G.5 Include references to related papers and/or other relevant material.	Not reported
G.6 Other comments.	Not reported

Table 3. InterTASC - Boynton 1998

A. Information	
A.1 State the author's objective	"To develop a highly sensitive search strategy to identify systematic reviews of intervention, including meta-analyses, indexed in MEDLINE and to evaluate this strategy in terms of its sensitivity and precision."
A.2 State the focus of the research.	Sensitivity-maximising Precision-maximising
A.3 Database(s) and search interface(s) .	MEDLINE(Ovid).
A.4 Describe the methodological focus of the filter (e.g. RCTs).	Systematic reviews
A.5 Describe any other topic that forms an additional focus of the filter (e.g. clinical topics such as breast cancer, geographic location such as Asia or population grouping such as paediatrics).	N/A
A.6 Other observations.	N/A
B. Identification of a gold standard (GS) of known relevant records	
B.1 Did the authors identify one or more gold standards (GSs)?	1 GSs To test the sensitivity and precision of various search terms.a 'quasi-gold standard' of systematic reviews was established by a combination of hand and electronic searching.
B.2 How did the authors identify the records in each GS?	The 'quasi-gold standard' was produced by using:Institute for Scientific Information (ISI) citation index impact factors to identify six journals from the top ten high-impact factor journals in the 'Medicine general and internal' category produced by ISI, searched by hand for the year.Search strategy used to identify candidate systematic reviews in MEDLINE (Ovid)
B.3 Report the dates of the records in each GS.	1992 and 1995
B.4 What are the inclusion criteria for each GS?	Decisions on whether to include studies in the 'quasi-gold standard' by two authors (JB and JG) independently, using a CRD quick-assessment sifting process.
B.5 Describe the size of each GS and the authors' justification, if provided (for example the size of the gold standard may have been determined by a power calculation)	288 papers. Ninety per cent of these systematic reviews were identified by hand searching and a further 10% by searching MEDLINE.
B.6 Are there limitations to the gold standard(s)?	Search dates limited (1992-1995)

Table 3. InterTASC - Boynton 1998 (Continued)

B.7 How was each gold standard used?	- To identify potential search terms - To derive potential strategies (groups of terms) - To test internal validity To test external validity
B.8 Other observations.	No
C. How did the researchers identify the search terms in their filter(s) (select all that apply)?	
C.1 Adapted a published search strategy.	No
C.2 Asked experts for suggestions of relevant terms.	No
C.3 Used a database thesaurus.	Mesh, Subject Headings.
C.4 Statistical analysis of terms in a gold standard set of records (see B above).	The list of candidate terms for search strategies was derived by analysing the MEDLINE records of the 288 articles in the 'quasi-gold standard' for word frequency.
C.5 Extracted terms from the gold standard set of records (see B above).	No
C.6 Extracted terms from some relevant records (but not a gold standard).	No
C.7 Tick all types of search terms tested.	- Subject headings - Text words (e.g. in title, abstract) - Publication types
C.8 Include the citation of any adapted strategies.	N/A
C.9 How were the (final) combination(s) of search terms selected?	The list of terms was sorted according to the best results for sensitivity and precision.
C.10 Were the search terms combined (using Boolean logic) in a way that is likely to retrieve the studies of interest?	Yes, Only MEDLINE search appendix 1
C.11 Other observations.	No
D. Internal validity testing (This type of testing is possible when the search filter terms were developed from a known gold standard set of records).	
D.1 How many filters were tested for internal validity?	8 filters: Strategy A: Sensitivity \geq 50, Precision \geq 25%, Strategy B: Sensitivity \geq 25%, Precision \geq 10%, Strategy C: Sensitivity \geq 25%, Precision \geq 25%, Strategy D: Sensitivity \geq 25%, Precision \geq 50%, Strategy E: Sensitivity \geq 25%, Precision \geq 70%, Strategy F: Sensitivity \geq 50%, Precision \geq 10%, Strategy H: Sensitivity \geq 25%, Precision \geq 10%, Strategy J: Sensitivity \geq 15%, Precision \geq 25% 3 additional filters were tested with expert input.

For each filter report the following information

Table 3. InterTASC - Boynton 1998 (Continued)

D.2 Was the performance of the search filter tested on the gold standard from which it was derived?	Yes, the search strategies were compared against the performance of other published search strategies in relation to the 'quasi-gold standard' and the OVID MEDLINE interface.
D.3 Report sensitivity data (a single value, a range, 'Unclear'* or 'not reported', as appropriate). *Please describe.	Yes, Strategy A: 66%, Strategy B: 95%, Strategy C: 92%, Strategy D: 39%, Strategy E: 29%, Strategy F: 61%, Strategy H: 98%, Strategy J: 98%
D.4 Report precision data (a single value, a range, 'Unclear'* or 'not reported' as appropriate). *Please describe.	A single value: Strategy A: 26%, Strategy B: 12%, Strategy C: 23%, Strategy D: 49%, Strategy E: 79%, Strategy F: 42%, Strategy H: 19%, Strategy J: 20%
D.5 Report specificity data (a single value, a range, 'Unclear'* or 'not reported' as appropriate). *Please describe.	No
D.6 Other performance measures reported.	No
D.7 Other observations.	No
E. External validity testing (This section relates to testing the search filter on records that are different from the records used to identify the search terms)	
E.1 How many filters were tested for external validity on records different from those used to identify the search terms?	No external validity test performed
E.2 Describe the validation set(s) of records, including the interface.	N/A
E.3 On which validation set(s) was the filter tested?	N/A
E.4 Report sensitivity data for each validation set (a single value, a range or 'Unclear' or 'not reported', as appropriate).	N/A
E.5 Report precision data for each validation set (report a single value, a range or 'Unclear' or 'not reported', as appropriate).	N/A
E.6 Report specificity data for each validation set (a single value, a range or 'Unclear' or 'not reported', as appropriate).	N/A
E.6 Other performance measures reported.	N/A
E.7 Other observations.	No
F. Limitations and comparisons.	
F.1 Did the authors discuss any limitations to their research?	Yes, "This study is based on six 'high-impact' factor English-language journals which may not be representative of health care journals as a whole""The years chosen for the creation of the 'quasi-gold standard' may present a limitation of this study""The CRD definition of systematic reviews and the associated criteria used by CRD to identify articles for inclusion in DARE, as outlined in Section 1.1 above, may be criticised for being too demanding"
F.2 Are there other potential limitations to this research that you have noticed?	NO

Table 3. InterTASC - Boynton 1998 (Continued)

F.3 Report any comparisons of the performance of the filter against other relevant published filters (sensitivity, precision, specificity or other measures).	Yes: Hunt and McKibbon full: Sensitivity 41%, Precision 75%. Hunt and McKibbon brief: Sensitivity 40%, Precision 75%. CRD full: Sensitivity 84%, Precision 31%. CRD brief: Sensitivity 41%, Precision 64%.
F.4 Include the citations of any compared filters.	Yes. (Hunt 1997 ; NHS 1996)
F.5 Other observations and / or comments.	No
G. Other comments. This section can be used to provide any other comments. Selected prompts for issues to bear in mind are given below.	
G.1 Have you noticed any errors in the document that might impact on the usability of the filter?	No
G.2 Are there any published errata or comments (for example in the MEDLINE record)?	No
G.3 Is there public access to pre-publication history and / or correspondence?	No
G.4 Are further data available on a linked site or from the authors?	No
G.5 Include references to related papers and/or other relevant material.	No
G.6 Other comments.	No

Table 4. InterTASC - Boluyt 2008

A. Information	
A.1 State the author's objective	To determine the sensitivity and precision of existing search strategies for retrieving child health systematic reviews in MEDLINE using PubMed.
A.2 State the focus of the research.	Sensitivity-maximising Precision-maximising
A.3 Database(s) and search interface(s) .	MEDLINE(PubMed)
A.4 Describe the methodological focus of the filter (e.g. RCTs).	Systematic reviews and meta-analyses.
A.5 Describe any other topic that forms an additional focus of the filter (e.g. clinical topics such as breast cancer, geographic location such as Asia or population grouping such as paediatrics).	Child health
A.6 Other observations.	NO
B. Identification of a gold standard (GS) of known relevant records	

Table 4. InterTASC - Boluyt 2008 (Continued)

B.1 Did the authors identify one or more gold standards (GSs)?	1 GSs. To measure search sensitivity strategies, a reference standard set of RS was established by search for RS of children's health in DARE and by manually searching for various magazines for RS.
B.2 How did the authors identify the records in each GS?	All titles and abstracts in DARE (Cochrane Library, Issue 2, 2004) were searched for SRs of child health also indexed in MEDLINE. We hand-searched 7 MEDLINE-indexed pediatric journals with a variety of impact factors and for which full-text electronic copies were available in our medical library. All issues of each journal were searched for the following 5 years: 1994, 1997, 2000, 2002 and 2004
B.3 Report the dates of the records in each GS.	Yes, 1994, 1997, 2000, 2002, and 2004
B.4 What are the inclusion criteria for each GS?	"Any literature review, meta-analysis, or other article that explicitly indicates the use of a strategy for locating evidence by mentioning at least the databases that were searched and reviewing the empirical evidence on children"
B.5 Describe the size of each GS and the authors' justification, if provided (for example the size of the gold standard may have been determined by a power calculation)	Total Reference Standard: 387 (298 by DARE + 115 found by hand search) - 26 overlap = 387
B.6 Are there limitations to the gold standard(s)?	Subset of the MEDLINE database
B.7 How was each gold standard used?	To test external validity
B.8 Other observations.	
C. How did the researchers identify the search terms in their filter(s) (select all that apply)?	
C.1 Adapted a published search strategy.	To identify articles that report on the development and validation of systems review the search filters in MEDLINE, searched MEDLINE from January 1995 to January 2006 with the following MeSH terms: MEDLINE, information storage and Retrieval/Methods, and Review, Literature. In addition, reference lists of relevant articles were reviewed and content experts were contacted to find further studies. To improve accuracy, © InterTASC Information Specialist Subgroup (ISSG) March 2008 We combined the systematic review filters with a sensitive child filter developed by the Cochrane Field of Child Health to retrieve only studies in children.
C.2 Asked experts for suggestions of relevant terms.	Contact with experts
C.3 Used a database thesaurus.	MeSH terms
C.4 Statistical analysis of terms in a gold standard set of records (see B above).	Not reported
C.5 Extracted terms from the gold standard set of records (see B above).	Not reported

Table 4. InterTASC - Boluyt 2008 (Continued)

C.6 Extracted terms from some relevant records (but not a gold standard).	Not reported
C.7 Tick all types of search terms tested.	Subject headings Text words (e.g. in title, abstract) Publication types
C.8 Include the citation of any adapted strategies.	Yes: Shojania, Boyton, White, Montori, PubMed plus child
C.9 How were the (final) combination(s) of search terms selected?	Not applicable.
C.10 Were the search terms combined (using Boolean logic) in a way that is likely to retrieve the studies of interest?	Not reported.
C.11 Other observations.	
D. Internal validity testing (This type of testing is possible when the search filter terms were developed from a known gold standard set of records).	
D.1 How many filters were tested for internal validity?	Not applicable.
For each filter report the following information	
D.2 Was the performance of the search filter tested on the gold standard from which it was derived?	Not applicable.
D.3 Report sensitivity data (a single value, a range, 'Unclear'* or 'not reported', as appropriate). *Please describe.	Not applicable.
D.4 Report precision data (a single value, a range, 'Unclear'* or 'not reported' as appropriate). *Please describe.	Not applicable.
D.5 Report specificity data (a single value, a range, 'Unclear'* or 'not reported' as appropriate). *Please describe.	Not applicable.
D.6 Other performance measures reported.	Not applicable.
D.7 Other observations.	
E. External validity testing (This section relates to testing the search filter on records that are different from the records used to identify the search terms)	
E.1 How many filters were tested for external validity on records different from those used to identify the search terms?	1. Shojania Plus Child 2. Boynton plus child 3. White 1 plus child 4. White 2 plus child 5. Montori 1 plus child 6. Montori 2 plus child 7. Montori 3 plus child 8. Montori 4 plus child

Table 4. InterTASC - Boluyt 2008 (Continued)

	9. PubMed plus child
E.2 Describe the validation set(s) of records, including the interface.	Total Reference Standard: 387 True child health systematic reviews found in MEDLINE (limits 1990-2006).
E.3 On which validation set(s) was the filter tested?	Reference standard all Pubmed records
E.4 Report sensitivity data for each validation set (a single value, a range or 'Unclear' or 'not reported', as appropriate).	Shojania+child 74 (95%CI 69-78) Boynton+child 95 (95%CI 92-97) White 1+child 93 (95%CI 91-96) White 2+child 94 (95%CI 91-96) Montori 1+child 96 (95%CI 93-97) Montori 2+child 68 (95%CI 64-73) Montori 3+child 94 (95%CI 91-96) Montori 4+child 72 (95%CI 67-76) Pubmed +child 76 (95%CI 72-80)
E.5 Report precision data for each validation set (report a single value, a range or 'Unclear' or 'not reported', as appropriate).	Shojania+child 45 (95%CI 36-55) Boynton+child 3 (95%CI 1-9) White 2+child 2 (95%CI 1-7) Montori 2+child 45 (95%CI 36-55) Montori 3+child 3 (95%CI 1-9) Pubmed +child 32 (95%CI 24-42)
E.6 Report specificity data for each validation set (a single value, a range or 'Unclear' or 'not reported', as appropriate).	Not reported
E.6 Other performance measures reported.	Not reported
E.7 Other observations.	
F. Limitations and comparisons.	
F.1 Did the authors discuss any limitations to their research?	Broad definition of systematic reviews
F.2 Are there other potential limitations to this research that you have noticed?	Not reported
F.3 Report any comparisons of the performance of the filter against other relevant published filters (sensitivity, precision, specificity or other measures).	Not reported
F.4 Include the citations of any compared filters.	Not reported
F.5 Other observations and / or comments.	Not reported

Table 4. InterTASC - Boluyt 2008 (Continued)

G. Other comments. This section can be used to provide any other comments. Selected prompts for issues to bear in mind are given below.

G.1 Have you noticed any errors in the document that might impact on the usability of the filter?	Not reported
G.2 Are there any published errata or comments (for example in the MEDLINE record)?	Not reported
G.3 Is there public access to pre-publication history and / or correspondence?	Not reported
G.4 Are further data available on a linked site or from the authors?	Not reported
G.5 Include references to related papers and/or other relevant material.	Not reported
G.6 Other comments.	

Table 5. InterTASC - Lee 2012
A. Information

A.1 State the author's objective	"The objective is to describe the development and validation of the health-evidence.ca systematic review search filter and to compare its performance with other available systematic review filters."
A.2 State the focus of the research.	Balance of sensitivity
A.3 Database(s) and search interface(s) .	MEDLINE, EMBASE, and CINAHL (Ovid)
A.4 Describe the methodological focus of the filter (e.g. RCTs).	Systematic reviews
A.5 Describe any other topic that forms an additional focus of the filter (e.g. clinical topics such as breast cancer, geographic location such as Asia or population grouping such as paediatrics).	Included reviews should meet the criteria for relevance and should be systematic reviews that focus on public health, provide outcome data on the effectiveness of interventions, and include a documented search strategy.
A.6 Other observations.	No

B. Identification of a gold standard (GS) of known relevant records

B.1 Did the authors identify one or more gold standards (GSs)?	1 GSs. "We considered this set (the electronic database searches plus additional search strategies), the 'gold standard' for health-evidence.ca."
B.2 How did the authors identify the records in each GS?	"Our PH search filter typically yielded a very high volume of results with very low precision. For example, between January 2006 and December 2007, of the 136,427 titles screened, 409 were relevant for the health evidence.ca registry, or in other words, precision was 0.3%. In addition to using the PH search filter, more than 40 public health-relevant journals were hand searched annually, as well as the reference lists of all relevant reviews. Given this systematic search of the published review literature, we were reasonably confi-

Table 5. InterTASC - Lee 2012 (Continued)

	dent that our retrieval methods were capturing a near complete set of relevant articles.”
B.3 Report the dates of the records in each GS.	Jan 2006 to Dec 2007
B.4 What are the inclusion criteria for each GS?	Systematic reviews that focus on public health, provide outcome data on the effectiveness of interventions, and include a documented search strategy.
B.5 Describe the size of each GS and the authors' justification, if provided (for example the size of the gold standard may have been determined by a power calculation)	No reported
B.6 Are there limitations to the gold standard(s)?	No
B.7 How was each gold standard used?	To test internal validity
B.8 Other observations.	No
C. How did the researchers identify the search terms in their filter(s) (select all that apply)?	
C.1 Adapted a published search strategy.	Yes, The health-evidence.ca Systematic Review (SR) search filter we developed in 2008 was adapted from a previously validated filter
C.2 Asked experts for suggestions of relevant terms.	No
C.3 Used a database thesaurus.	Yes, only Medline Ovid, and adapted to Embase and CINAHL.
C.4 Statistical analysis of terms in a gold standard set of records (see B above).	No
C.5 Extracted terms from the gold standard set of records (see B above).	No
C.6 Extracted terms from some relevant records (but not a gold standard).	No
C.7 Tick all types of search terms tested.	- Text words (e.g. in title, abstract) - Publication types
C.8 Include the citation of any adapted strategies.	Default search strategies used for BMJ Clinical Evidence
C.9 How were the (final) combination(s) of search terms selected?	No
C.10 Were the search terms combined (using Boolean logic) in a way that is likely to retrieve the studies of interest?	The health-evidence.ca Systematic Review (SR) search filter we developed in 2008 was adapted from a previously-validated filter, which included the terms: MEDLINE.tw, systematic review.tw, meta-analysis.pt, combined with the Boolean OR operator.
C.11 Other observations.	No
D. Internal validity testing (This type of testing is possible when the search filter terms were developed from a known gold standard set of records).	
D.1 How many filters were tested for internal validity?	1 filters

Table 5. InterTASC - Lee 2012 (Continued)

For each filter report the following information

D.2 Was the performance of the search filter tested on the gold standard from which it was derived?	No
D.3 Report sensitivity data (a single value, a range, 'Unclear'* or 'not reported', as appropriate). *Please describe.	- MEDLINE: 86.8% (75.2 to 93.5) - Embase: 72.7% (55.8 to 84.9) - CINAHL: 86.1% (71.4 to 93.9)
D.4 Report precision data (a single value, a range, 'Unclear'* or 'not reported' as appropriate). *Please describe.	- MEDLINE: 1.1% (0.9 to 1.2) - Embase: 0.6% (0.4 to 0.7) - CINAHL: 1.6% (1.4 to 1.8)
D.5 Report specificity data (a single value, a range, 'Unclear'* or 'not reported' as appropriate). *Please describe.	- MEDLINE: 99.2% (99.2 to 99.2) - Embase: 99.1% (99.1 to 99.1) - CINAHL: 98.2% (98.2 to 98.2)
D.6 Other performance measures reported.	Number needed to read - MEDLINE: 91.6 (85.0 to 105.9) - Embase: 171.6 (146.7 to 224.6) - CINAHL: 61.3 (56.1 to 74.3)
D.7 Other observations.	No
E. External validity testing (This section relates to testing the search filter on records that are different from the records used to identify the search terms)	
E.1 How many filters were tested for external validity on records different from those used to identify the search terms?	1 filters
E.2 Describe the validation set(s) of records, including the interface.	No
E.3 On which validation set(s) was the filter tested?	No
E.4 Report sensitivity data for each validation set (a single value, a range or 'Unclear' or 'not reported', as appropriate).	- MEDLINE: 89.9 (85.0, 93.3) - Embase: 87.9 (80.3, 92.8) - CINAHL: 89.9 (93.5, 94.0)
E.5 Report precision data for each validation set (report a single value, a range or 'Unclear' or 'not reported', as appropriate).	- MEDLINE: 1.4 (1.3, 1.5) - Embase: 0.5 (0.5, 0.6) - CINAHL: 1.8 (1.6, 1.8)
E.6 Report specificity data for each validation set (a single value, a range or 'Unclear' or 'not reported', as appropriate).	- MEDLINE: 98.9 (98.9, 98.9) - Embase: 98.2 (98.2, 98.2) - CINAHL: 97.6 (97.6, 97.6)

Table 5. InterTASC - Lee 2012 (Continued)

E.7 Other performance measures reported.	No
F. Limitations and comparisons.	
F.1 Did the authors discuss any limitations to their research?	Yes: - "Searching was conducted in OVID's search interface for all three databases; other search interfaces for these databases (e.g. PubMed) may handle the searches somewhat differently" - "Precision and NNR scores were calculated specifically for public health content and cannot be generalized to topic areas outside of public health" -
F.2 Are there other potential limitations to this research that you have noticed?	No
F.3 Report any comparisons of the performance of the filter against other relevant published filters (sensitivity, precision, specificity or other measures).	Yes: Performance of the health-evidence.ca SR search filter compared to the PH search filter in retrieving systematic reviews in MEDLINE. 1. health-evidence.ca SR search filter†: Development: sensitivity 86.8 (75.2, 93.5) Specificity 86.8 (75.2, 93.5); Precision 1.1 (0.9, 1.2) Number needed to read; 91.6 (85.0, 105.9) Validation: sensitivity 89.9 (85.0, 93.3); Specificity 98.9 (98.9, 98.9); Precision 1.4 (1.3, 1.5); 2. PH search filter Development sensitivity 86.8 (75.2, 93.5) Specificity 86.8 (75.2, 93.5); Precision 1.1 (0.9, 1.2) Number needed to read; 91.6 (85.0, 105.9) Validation: sensitivity 89.9 (85.0, 93.3); Specificity 98.9 (98.9, 98.9); Precision 1.4 (1.3, 1.5)
F.4 Include the citations of any compared filters.	Yes, Montori 2005 ; Shojania 2001 ; Hunt 1997 ; Boynton 1998 ; BMJ Clinical Evidence ; SIGN ; Wilczynski 2007 ; Wong 2006 ; Wilczynski 1995
F.5 Other observations and / or comments.	No
G. Other comments. This section can be used to provide any other comments. Selected prompts for issues to bear in mind are given below.	
G.1 Have you noticed any errors in the document that might impact on the usability of the filter?	No
G.2 Are there any published errata or comments (for example in the MEDLINE record)?	No

Table 5. InterTASC - Lee 2012 (Continued)

G.3 Is there public access to pre-publication history and / or correspondence?	No
G.4 Are further data available on a linked site or from the authors?	Yes, additional files 1, 2 and 3 with the details on the search strategies and their performance.
G.5 Include references to related papers and/or other relevant material.	No
G.6 Other comments.	No

Table 6. InterTASC - Salvador-Oliván 2021

A. Information	
A.1 State the author's objective	Develop a search filter for retrieving systematic reviews.
A.2 State the focus of the research.	<ul style="list-style-type: none"> • Sensitivity-maximising • Precision-maximising
A.3 Database(s) and search interface(s) .	MEDLINE via PubMed
A.4 Describe the methodological focus of the filter (e.g. RCTs).	Systematic reviews.
A.5 Describe any other topic that forms an additional focus of the filter (e.g. clinical topics such as breast cancer, geographic location such as Asia or population grouping such as paediatrics).	None
A.6 Other observations.	None
B. Identification of a gold standard (GS) of known relevant records	
B.1 Did the authors identify one or more gold standards (GSs)?	None
B.2 How did the authors identify the records in each GS?	None
B.3 Report the dates of the records in each GS.	None
B.4 What are the inclusion criteria for each GS?	None
B.5 Describe the size of each GS and the authors' justification, if provided (for example the size of the gold standard may have been determined by a power calculation)	None
B.6 Are there limitations to the gold standard(s)?	None
B.7 How was each gold standard used?	None
B.8 Other observations.	None
C. How did the researchers identify the search terms in their filter(s) (select all that apply)?	
C.1 Adapted a published search strategy.	Terms extracted from the titles of articles indexed as systematic review [pt] and differ-

Table 6. InterTASC - Salvador-Oliván 2021 (Continued)

	ing from those already in the PubMed SR filter.
C.2 Asked experts for suggestions of relevant terms.	None
C.3 Used a database thesaurus.	Yes, MeSH
C.4 Statistical analysis of terms in a gold standard set of records (see B above).	None
C.5 Extracted terms from the gold standard set of records (see B above).	None
C.6 Extracted terms from some relevant records (but not a gold standard).	Yes, terms extracted from the titles of articles indexed as systematic review [pt] and differing from those already in the PubMed SR filter.
C.7 Tick all types of search terms tested.	<ul style="list-style-type: none"> • Subject headings • Text words (e.g. in title, abstract) • publication types
C.8 Include the citation of any adapted strategies.	Yes
C.9 How were the (final) combination(s) of search terms selected?	The list of terms was sorted according to the best results for recall and precision.
C.10 Were the search terms combined (using Boolean logic) in a way that is likely to retrieve the studies of interest?	yes
C.11 Other observations.	None
D. Internal validity testing (This type of testing is possible when the search filter terms were developed from a known gold standard set of records).	
D.1 How many filters were tested for internal validity?	None
For each filter report the following information	
D.2 Was the performance of the search filter tested on the gold standard from which it was derived?	None
D.3 Report sensitivity data (a single value, a range, 'Unclear'* or 'not reported', as appropriate). *Please describe.	None
D.4 Report precision data (a single value, a range, 'Unclear'* or 'not reported' as appropriate). *Please describe.	None
D.5 Report specificity data (a single value, a range, 'Unclear'* or 'not reported' as appropriate). *Please describe.	None
D.6 Other performance measures reported.	None
D.7 Other observations.	None
E. External validity testing (This section relates to testing the search filter on records that are different from the records used to identify the search terms)	

Table 6. InterTASC - Salvador-Oliván 2021 (Continued)

E.1 How many filters were tested for external validity on records different from those used to identify the search terms?	1 filters
E.2 Describe the validation set(s) of records, including the interface.	The validation was compared with the Pubmed SR filter.
E.3 On which validation set(s) was the filter tested?	Does not report a validation set
E.4 Report sensitivity data for each validation set (a single value, a range or 'Unclear' or 'not reported', as appropriate).	Not reported
E.5 Report precision data for each validation set (report a single value, a range or 'Unclear' or 'not reported', as appropriate).	Between 72.3 and 96.7%, with a weighted mean precision of 83.8%.
E.6 Report specificity data for each validation set (a single value, a range or 'Unclear' or 'not reported', as appropriate).	Not reported
E.6 Other performance measures reported.	Recall – Single value. 91.6%
E.7 Other observations.	None
F. Limitations and comparisons.	
F.1 Did the authors discuss any limitations to their research?	Broad definition of SR. Not using a gold standard set of records.
F.2 Are there other potential limitations to this research that you have noticed?	None
F.3 Report any comparisons of the performance of the filter against other relevant published filters (sensitivity, precision, specificity or other measures).	“The PubMed SR filter retrieved 62.0% (168,677/272,048) of the articles of our final filter, which means that it is likely to have missed a large number of potential systematic reviews.”
F.4 Include the citations of any compared filters.	Yes, Shojania 2001
F.5 Other observations and / or comments.	None
G. Other comments. This section can be used to provide any other comments. Selected prompts for issues to bear in mind are given below.	
G.1 Have you noticed any errors in the document that might impact on the usability of the filter?	None
G.2 Are there any published errata or comments (for example in the MEDLINE record)?	None
G.3 Is there public access to pre-publication history and / or correspondence?	None
G.4 Are further data available on a linked site or from the authors?	Yes, "SUPPLEMENTAL FILES"
G.5 Include references to related papers and/or other relevant material.	Yes
G.6 Other comments.	None

Table 7. InterTASC - Shojania 2001

A. Information	
A.1 State the author's objective.	"To develop and evaluate a search strategy for identifying systematic reviews by using a publicly available MEDLINE interface (PubMed)."
A.2 State the focus of the research.	<ul style="list-style-type: none"> • Sensitivity-maximising • Precision-maximising • Balance of sensitivity and specificity / precision
A.3 Database(s) and search interface(s).	MEDLINE (PubMed)
A.4 Describe the methodological focus of the filter (e.g. RCTs).	Systematic reviews
A.5 Describe any other topic that forms an additional focus of the filter (e.g. clinical topics such as breast cancer, geographic location such as Asia or population grouping such as paediatrics).	Screening for colorectal cancer, thrombolytic therapy for venous thromboembolism, and treatment of dementia (for calculating predictive values)
A.6 Other observations.	No
B. Identification of a gold standard (GS) of known relevant records	
B.1 Did the authors identify one or more gold standards (GSs)?	MEDLINE (PubMed)
B.2 How did the authors identify the records in each GS?	Systematic reviews
B.3 Report the dates of the records in each GS.	
B.4 What are the inclusion criteria for each GS?	No reported
B.5 Describe the size of each GS and the authors' justification, if provided (for example the size of the gold standard may have been determined by a power calculation)	"Included reviews must meet at least four methodologic criteria, and all included articles must display "evidence of a substantial effort to search for all relevant research." "Review articles included in ACP Journal Club must contain "an identifiable description of the methods indicating the sources and methods for searching for articles and state the clinical topic and the inclusion and exclusion criteria for selecting articles for detailed review."
B.6 Are there limitations to the gold standard(s)?	Quasi-Gold Standard (only systematic reviews) DARE-indexed systematic reviews: first 100 records of the database. ACP Journal Club: handsearch of 104.
B.7 How was each gold standard used?	Broad definition of Systematic Reviews. Methodological inclusion criteria for gold standards not reported.
B.8 Other observations.	To test internal validity

Table 7. InterTASC - Shojanian 2001 (Continued)

C. How did the researchers identify the search terms in their filter(s) (select all that apply)?

C.1 Adapted a published search strategy.	Not mentioned
C.2 Asked experts for suggestions of relevant terms.	Not mentioned
C.3 Used a database thesaurus.	Not mentioned
C.4 Statistical analysis of terms in a gold standard set of records (see B above).	Not mentioned
C.5 Extracted terms from the gold standard set of records (see B above).	Not mentioned
C.6 Extracted terms from some relevant records (but not a gold standard).	Not mentioned
C.7 Tick all types of search terms tested.	<ul style="list-style-type: none"> • Text words (e.g. in title, abstract) • Publication types
C.8 Include the citation of any adapted strategies.	Not mentioned
C.9 How were the (final) combination(s) of search terms selected?	Not reported
C.10 Were the search terms combined (using Boolean logic) in a way that is likely to retrieve the studies of interest?	Yes
C.11 Other observations.	No

D. Internal validity testing (This type of testing is possible when the search filter terms were developed from a known gold standard set of records).

D.1 How many filters were tested for internal validity?	None
---	------

For each filter report the following information

D.2 Was the performance of the search filter tested on the gold standard from which it was derived?	No
D.3 Report sensitivity data (a single value, a range, 'Unclear'* or 'not reported', as appropriate). *Please describe.	Not applicable
D.4 Report precision data (a single value, a range, 'Unclear'* or 'not reported' as appropriate). *Please describe.	Not applicable
D.5 Report specificity data (a single value, a range, 'Unclear'* or 'not reported' as appropriate). *Please describe.	Not applicable
D.6 Other performance measures reported.	Not applicable
D.7 Other observations.	No

E. External validity testing (This section relates to testing the search filter on records that are different from the records used to identify the search terms)

E.1 How many filters were tested for external validity on records different from those used to identify the search terms?	One
---	-----

Table 7. InterTASC - Shojania 2001 (Continued)

E.2 Describe the validation set(s) of records, including the interface.	The filter was combined with words on clinical topics to retrieve systematic reviews in MEDLINE(Pubmed)
E.3 On which validation set(s) was the filter tested?	2
E.4 Report sensitivity data for each validation set (a single value, a range or 'Unclear' or 'not reported', as appropriate).	Sensitivity (95%CI) DARE-indexed systematic reviews: 93% (86% to 97%) ACP Journal Club: 97% (91% to 99%)
E.5 Report precision data for each validation set (report a single value, a range or 'Unclear' or 'not reported', as appropriate).	"Approximately 50% of the retrieved articles met the criteria for true positive in each of the three sample searches."screening for colorectal cancer: 53% (30%–75%) thrombolytic therapy for venous thromboembolism: 54% (24%–82%) treatment of dementia: 50% (41%–60%)
E.6 Report specificity data for each validation set (a single value, a range or 'Unclear' or 'not reported', as appropriate).	Not reported
E.7 Other observations.	No
F. Limitations and comparisons.	
F.1 Did the authors discuss any limitations to their research?	Yes, broad definition of systematic review (including consensus and guidelines).
F.2 Are there other potential limitations to this research that you have noticed?	Yes, Precision measures are not reported correctly. It is not clear if it was tested in the Gold Standard or in different set of articles.
F.3 Report any comparisons of the performance of the filter against other relevant published filters (sensitivity, precision, specificity or other measures).	No
F.4 Include the citations of any compared filters.	No
F.5 Other observations and / or comments.	No
G. Other comments. This section can be used to provide any other comments. Selected prompts for issues to bear in mind are given below.	
G.1 Have you noticed any errors in the document that might impact on the usability of the filter?	No
G.2 Are there any published errata or comments (for example in the MEDLINE record)?	No
G.3 Is there public access to pre-publication history and / or correspondence?	No
G.4 Are further data available on a linked site or from the authors?	Yes, There is an Appendix cited in the manuscript, but there is no link available and the appendix is not accesible.
G.5 Include references to related papers and/or other relevant material.	Not Included.

Table 7. InterTASC - Shojanian 2001 (Continued)

G.6 Other comments.	No
---------------------	----

Table 8. InterTASC - White 2001

Information	
A.1 State the author's objective.	To improve previously developed methods to derive a more objective search strategy to identify systematic reviews in MEDLINE.
A.2 State the focus of the research.	Sensitivity-maximising Precision-maximising Specificity-maximising
A.3 Database(s) and search interface(s).	MEDLINE (Ovid)
A.4 Describe the methodological focus of the filter (e.g. RCTs).	Systematic reviews
A.5 Describe any other topic that forms an additional focus of the filter (e.g. clinical topics such as breast cancer, geographic location such as Asia or population grouping such as paediatrics).	No reported
A.6 Other observations.	No
B. Identification of a gold standard (GS) of known relevant records	
B.1 Did the authors identify one or more gold standards (GSs)?	1 GS
B.2 How did the authors identify the records in each GS?	"A 'quasi-gold standard' [we reclassified this as gold standard as it includes systematic reviews and non-systematic reviews] set of known systematic reviews from key general medical journals was identified." "Three groups of records were generated from five medical journals indexed in MEDLINE: Annals of Internal Medicine, Archives of Internal Medicine, BMJ, JAMA, and the Lancet." "Every article in each available issue of the five chosen journals was scanned independently by two information staff"
B.3 Report the dates of the records in each GS.	1995 - 1997
B.4 What are the inclusion criteria for each GS?	yes, "Two questions were asked to determine whether an article was a 'review'. In order to be classed as a review the answer to both questions must be 'yes'. Firstly, does the article attempt to summarize the findings of one or more studies?" "Secondly, is the article a review of the side effects or effectiveness of an intervention?"
B.5 Describe the size of each GS and the authors' justification, if provided (for example the size of the gold standard may have been determined by a power calculation)	110 systematic reviews 110 reviews (not systematic) 125 non-review articles (mainly primary studies)

Table 8. InterTASC - White 2001 (Continued)

B.6 Are there limitations to the gold standard(s)?	Yes, The GS is limited to two years search in only five journals.
B.7 How was each gold standard used?	To identify potential search terms To derive potential strategies (groups of terms) To test internal validity To test external validity
B.8 Other observations.	No
C. How did the researchers identify the search terms in their filter(s) (select all that apply)?	
C.1 Adapted a published search strategy.	No
C.2 Asked experts for suggestions of relevant terms.	No
C.3 Used a database thesaurus.	No
C.4 Statistical analysis of terms in a gold standard set of records (see B above).	Yes, Statistical analysis was used to determine which of the selected terms or phrases would best distinguish between types of records.
C.5 Extracted terms from the gold standard set of records (see B above).	Yes
C.6 Extracted terms from some relevant records (but not a gold standard).	Yes, "Wordstat and Simstat textual analysis software were used to identify frequently occurring terms in the titles and abstracts of the systematic review test set"
C.7 Tick all types of search terms tested.	<ul style="list-style-type: none"> • subject headings • text words (e.g. in title, abstract) • publication types
C.8 Include the citation of any adapted strategies.	No
C.9 How were the (final) combination(s) of search terms selected?	"Two optimal models were developed, one which allows the number of times a term appears to be a variable, and another in which only whether a term appears at all is considered. The performance of these was noted."
C.10 Were the search terms combined (using Boolean logic) in a way that is likely to retrieve the studies of interest?	No
C.11 Other observations.	No
D. Internal validity testing (This type of testing is possible when the search filter terms were developed from a known gold standard set of records).	
D.1 How many filters were tested for internal validity?	Unclear
For each filter report the following information	
D.2 Was the performance of the search filter tested on the gold standard from which it was derived?	1 GS Handsearch 5 journals 1995-1997; QGS (SR): n=110; Non SR: n= 110, Non review: n= 125

Table 8. InterTASC - White 2001 (Continued)

	Total: 5 filters (A, B, C, D, E)
D.3 Report sensitivity data (a single value, a range, 'Unclear'* or 'not reported', as appropriate). *Please describe.	A: 73.4% B: 67.1% C: 81.9% D: 87.1% E: 77.2%
D.4 Report precision data (a single value, a range, 'Unclear'* or 'not reported' as appropriate). *Please describe.	Not reported
D.5 Report specificity data (a single value, a range, 'Unclear'* or 'not reported' as appropriate). *Please describe.	A: 93.3% B: 94.9% C: 99.4% D: 89.2% E: 94.9%
D.6 Other performance measures reported.	No
D.7 Other observations.	No
E. External validity testing (This section relates to testing the search filter on records that are different from the records used to identify the search terms)	
E.1 How many filters were tested for external validity on records different from those used to identify the search terms?	5
E.2 Describe the validation set(s) of records, including the interface.	The sensitivity and precision of each model and two previously published strategies [5] were tested in a 'real world' scenario (described above) using OVID MEDLINE on CD-ROM (1995 to September 1998 issue; Table 12).
For each filter report the following information.	
E.3 On which validation set(s) was the filter tested?	Not reported
E.4 Report sensitivity data for each validation set (a single value, a range or 'Unclear' or 'not reported', as appropriate).	A: 100%B: 100%C: 93.6%D: 99%E: 91.8%
E.5 Report precision data for each validation set (report a single value, a range or 'Unclear' or 'not reported', as appropriate).	A: 4.4% B: 4.4% C: 10% D: 4.7% E: 9.7%
E.6 Report specificity data for each validation set (a single value, a range or 'Unclear' or 'not reported', as appropriate).	A: 98.3% B: 94.9% C: 99.4%

Table 8. InterTASC - White 2001 (Continued)

	D: 89.2%
	E: 94.9%
E.7 Other performance measures reported.	No
E.8 Other observations.	No
F. Limitations and comparisons.	
F.1 Did the authors discuss any limitations to their research?	Yes
F.2 Are there other potential limitations to this research that you have noticed?	No
F.3 Report any comparisons of the performance of the filter against other relevant published filters (sensitivity, precision, specificity or other measures).	"Comparison against most sensitive strategy from Boynton et al and CRD strategy to identify systematic reviews. Most sensitive strategy from Boynton et al: - sensitivity: 99 - precision: 7.5 - CRD strategy to identify systematic reviews: sensitivity: 93.6 - precision: 11.3"
F.4 Include the citations of any compared filters.	No
F.5 Other observations and / or comments.	No
G. Other comments. This section can be used to provide any other comments. Selected prompts for issues to bear in mind are given below.	
G.1 Have you noticed any errors in the document that might impact on the usability of the filter?	No
G.2 Are there any published errata or comments (for example in the MEDLINE record)?	No
G.3 Is there public access to pre-publication history and / or correspondence?	No
G.4 Are further data available on a linked site or from the authors?	No
G.5 Include references to related papers and/or other relevant material.	No
G.6 Other comments.	No

Table 9. InterTASC - Wylczynski 2007

Information	
A.1 State the author's objective.	The objective of this study was to develop search strategies that optimize the retrieval of methodologically sound systematic reviews from Embase and MEDLINE (Ovid)

Table 9. InterTASC - Wylczynski 2007 (Continued)

A.2 State the focus of the re-search.	<ul style="list-style-type: none"> • Sensitivity-maximising • Specificity-maximising • Precision-maximising • Balance of sensitivity/specificity
A.3 Database(s) and search inter-face(s).	Embase and MEDLINE (via Ovid)
A.4 Describe the methodological focus of the filter (e.g. RCTs).	Methodologically sound systematic reviews.
A.5 Describe any other topic that forms an additional focus of the filter (e.g. clinical topics such as breast cancer, geographic location such as Asia or population grouping such as paediatrics).	No
A.6 Other observations.	No
B. Identification of a gold standard (GS) of known relevant records	
B.1 Did the authors identify one or more gold standards (GSs)?	One for Embase and two for MEDLINE
B.2 How did the authors identify the records in each GS?	<p>The authors used a handsearch to retrieve the records.</p> <p><i>Embase filter</i></p> <p>"Manual review (hand search) of each article for each issue of 55 journal titles for the year 2000. Overall, research staff hand searched 170 English-language journal titles. These journals were chosen based on recommendations of clinicians and librarians, Science Citation Index Impact Factors provided by the Institute for Scientific Information, and ongoing assessment of their yield of studies and reviews of scientific merit and clinical relevance for the disciplines of internal medicine, general medical practice, mental health, and general nursing practice"</p> <p><i>MEDLINE filter</i></p> <p>"We defined a set of 161 clinical journals for the fields of general internal medicine, family practice, nursing, and mental health that were indexed in Medline [...] Then, each article was searched in Medline for 2000 (using the Ovid interface) and the full Medline record captured and linked with the handsearch data."</p>
B.3 Report the dates of the records in each GS.	2000
B.4 What are the inclusion criteria for each GS?	<p><i>Embase filter</i></p> <p>Methodologically sound reviews described as: "statement of the clinical topic; explicit statement of the inclusion and exclusion criteria for original articles; description of the methods of review; and at least one article included in the review had to meet minimum methods standards."</p> <p><i>For the MEDLINE filter</i></p> <p>A similar criteria was described: "For an article to be considered a systematic review, the authors had to clearly state the clinical topic of the review and how the evidence was retrieved and from what sources, and they had to provide explicit inclusion and exclusion criteria and include at least one study that passed methodological criteria for the purpose category. For example, re-</p>

Table 9. InterTASC - Wylczynski 2007 (Continued)

	views of interventions had to have at least one study with random allocation of participants to comparison groups and assessment of at least one clinical outcome."
B.5 Describe the size of each GS and the authors' justification, if provided (for example the size of the gold standard may have been determined by a power calculation)	<p><i>Embase filter</i></p> <p>Database: 27.769 articles; 1,354 were classified as reviews, of which 220 (16.2%) were methodologically sound.</p> <p><i>MEDLINE filter</i></p> <p>Derivation database: 10.446 records, of which 133 (1.3%) were systematic reviews</p> <p>Validation database (including the Cochrane Database of Systematic Reviews): 49.028 records, of which 753 (1.5%) were systematic reviews.</p>
B.6 Are there limitations to the gold standard(s)?	Yes, limited to year 2000 and by handsearch.
B.7 How was each gold standard used?	<p>To identify potential search terms</p> <p>To test internal validity (derive the search strategy)</p> <p>To test external validity</p>
B.8 Other observations.	No
C. How did the researchers identify the search terms in their filter(s) (select all that apply)?	
C.1 Adapted a published search strategy.	Yes, Authors' state that "We compiled an initial list of search terms, including indexing terms and textwords from clinical articles."
C.2 Asked experts for suggestions of relevant terms.	Yes, "Input was then sought from clinicians and librarians in the United States and Canada."
C.3 Used a database thesaurus.	Yes, "Examples of the search terms tested are "meta-analysis," "selection criteria," "research review," and "overview," all as textwords; "review," the index term, and the index term 'clinical trial,' exploded"
C.4 Statistical analysis of terms in a gold standard set of records (see B above).	No
C.5 Extracted terms from the gold standard set of records (see B above).	No
C.6 Extracted terms from some relevant records (but not a gold standard).	No mentioned
C.7 Tick all types of search terms tested.	<ul style="list-style-type: none"> • subject headings • text words (e.g. in title, abstract) • publication types • subheadings
C.8 Include the citation of any adapted strategies.	<p><i>Embase filter:</i></p> <p>"Individual search terms with sensitivity >25% and specificity >75% for a given format type (i.e., original study or review) or purpose category (e.g., treatment, diagnosis, prognosis) were incorporated into the development of search strategies that included two or more terms."</p>

Table 9. InterTASC - Wylczynski 2007 (Continued)

MEDLINE filter:

"Individual search terms with a sensitivity of more than 50% (to develop strategies that optimised sensitivity) and a specificity of more than 75% (to develop strategies that optimised specificity) for identifying systematic reviews were incorporated into the development of two term strategies."

C.9 How were the (final) combination(s) of search terms selected?	Using the Boolean OR
C.10 Were the search terms combined (using Boolean logic) in a way that is likely to retrieve the studies of interest?	Not reported
C.11 Other observations.	No

D. Internal validity testing (This type of testing is possible when the search filter terms were developed from a known gold standard set of records).

D.1 How many filters were tested for internal validity?	<p><i>Embase filter:</i></p> <p>4 filters. Index terms and text words from clinical studies and advice sought from clinicians and librarians. Terms with individual sensitivity of >25% and specificity of >75% were incorporated into the development of filters. Tested out combining terms with OR</p> <p><i>MEDLINE filter:</i></p> <p>4 Filters. Index terms and text words from clinical studies and advice sought from clinicians and librarians. Terms with individual sensitivity of >50% and specificity of >75% were incorporated into the development of the filters. Tested out combining terms with OR</p>
---	--

For each filter report the following information

D.2 Was the performance of the search filter tested on the gold standard from which it was derived?	Yes
D.3 Report sensitivity data (a single value, a range, 'Unclear'* or 'not reported', as appropriate). *Please describe.	<p><i>Embase filter:</i></p> <p>Best sensitivity: 94.6% (95% CI 91.5, 97.6)</p> <p>Best specificity: 61.4% (95% CI 54.9, 67.8)</p> <p>Small drop in specificity with a substantive gain in sensitivity: 75.0% (95% CI 69.3, 80.7)</p> <p>Best optimization of sensitivity and specificity: 92.3% (95% CI 88.7, 95.8)</p> <p><i>MEDLINE filter:</i></p> <p>Best sensitivity: Sensitivity: 100% (97.3 to 100);</p> <p>Top strategy minimising the difference between sensitivity and specificity: sensitivity:Specificity: 93.0% (92.5 to 93.5)</p> <p>Top precision performer: Specificity: 99.4% (99.2 to 99.5);</p>
D.4 Report precision data (a single value, a range, 'Unclear'* or	<p><i>Embase filter:</i></p> <p>Best sensitivity and best optimization of sensitivity & specificity: 2.1% (1.8, 2.5)</p>

Table 9. InterTASC - Wylczynski 2007 (Continued)

'not reported' as appropriate). *Please describe.	Best specificity: 23.9 % (20.0, 27.8) Small drop in specificity with a substantive gain in sensitivity: 29.2% (25.4, 32.9) Best optimization of sensitivity and specificity: 5.6% (4.9, 6.4) MEDLINE filter: Top sensitivity strategies‡: 3.41% (2.86 to 4.03) Top strategy minimising the difference between sensitivity and specificity§: 14.6% (12.3 to 17.2) Top precision performer‡: 60.2% (52.4 to 67.7)
D.5 Report specificity data (a single value, a range, 'Unclear'* or 'not reported' as appropriate). *Please describe.	Embase filter: Best sensitivity and best optimization of sensitivity & specificity: 72.7% (72.2, 73.3) Best specificity: 98.7% (98.6, 98.9) Small drop in specificity with a substantive gain in sensitivity: 98.5% (98.4, 98.7) Best optimization of sensitivity and specificity: 87.7% (87.3, 88.1) MEDLINE filter: Top sensitivity strategies‡: 63.5% (62.5 to 64.4) Top strategy minimising the difference between sensitivity and specificity§: 93.0% (92.5 to 93.5) Top precision performer‡: 99.4% (99.2 to 99.5)
D.6 Other performance measures reported.	Not reported
D.7 Other observations.	No
E. External validity testing (This section relates to testing the search filter on records that are different from the records used to identify the search terms)	
E.1 How many filters were tested for external validity on records different from those used to identify the search terms?	Only MEDLINE 4 filters. 1- Best sensitivity 2- Best specificity 3- Small drop in specificity with a substantive gain in sensitivity 4- Best optimization of sensitivity and specificity
E.2 Describe the validation set(s) of records, including the interface.	"Indexing information was downloaded from EMBASE for 27,769 articles from the 55 journals hand searched. Of these, 1,354 were classified as reviews, of which 220 (16.2%) were methodologically sound (i.e., a systematic review with at least one included study meeting minimum methods standards). Search strategies were developed using all 27,769 articles."
For each filter report the following information.	
E.3 On which validation set(s) was the filter tested?	"The strategies were tested for their ability to retrieve high-quality review articles from all other articles, including both low-quality review articles and all nonreview studies."
E.4 Report sensitivity data for each validation set (a single val-	Best sensitivity 99.9% (IC 95% 99.6 to 100) Top strategy minimising the difference between sensitivity and specificity: 92.5% (86.6 to 96.3)

Table 9. InterTASC - Wylczynski 2007 (Continued)

ue, a range or 'Unclear' or 'not reported', as appropriate).	Top precision performer: 71.2% (68.0 to 74.4)
E.5 Report precision data for each validation set (report a single value, a range or 'Unclear' or 'not reported', as appropriate).	Best sensitivity: 3.14% (2.92 to 3.37) Top strategy minimising the difference between sensitivity and specificity: 14.2% (13.3 to 15.2) Top precision performer. 57.1% (53.9 to 60.3)
E.6 Report specificity data for each validation set (a single value, a range or 'Unclear' or 'not reported', as appropriate).	Best sensitivity: 52.0% (51.6 to 52.5) Top strategy minimising the difference between sensitivity and specificity: 93.0% (92.5 to 93.5) Top precision performer: 99.2% (99.1 to 99.3)
E.7 Other performance measures reported.	No
E.8 Other observations .	No

F. Limitations and comparisons.

F.1 Did the authors discuss any limitations to their research?	Yes, "Low values for precision should not be over-interpreted because we did not limit the searches by clinical content, topic terms. In absolute, if not relative, terms, precision will be enhanced by combining the search strategies in these tables with content-specific terms using the Boolean "AND." It may also be possible to increase precision by combining search strategies with methodologic terms using the Boolean "AND NOT."
F.2 Are there other potential limitations to this research that you have noticed?	Limitation of the gold standard to one year search and the wide definition for "methodologically sound review".
F.3 Report any comparisons of the performance of the filter against other relevant published filters (sensitivity, precision, specificity or other measures).	<p>Yes, comparative tables of Performance from published strategies to identify systematic reviews in MEDLINE tested in our full validation database. Values are percentages (95% confidence intervals)</p> <p><i>1. Centre for Reviews and Dissemination:</i></p> <ul style="list-style-type: none"> - High sensitivity (16 terms): Sensitivity*: 97.6 (96.5 to 98.7); Specificity**: 69.6 (69.2 to 70.0); Precision: 4.77 (4.43 to 5.11). - Intermediate sensitivity and precision (29 terms): Sensitivity*: 96.7 (95.4 to 98.0); Specificity**: 79.7 (79.3 to 80.0); Precision: 6.91 (6.42 to 7.39) <p><i>2. Hunt and McKibbon:</i></p> <ul style="list-style-type: none"> - Simple query (4 terms): Sensitivity: 68.8 (65.5 to 72.1); Specificity: 99.2 (99.1 to 99.3); Precision: 56.7 (53.5 to 59.9) - Sensitive query (8 terms): Sensitivity*: 73.4 (70.3 to 76.6); Specificity**: 99.1 (99.0 to 99.2); Precision: 55.1 (52.0 to 58.2) <p><i>3. Shojania and Bero:</i></p> <ul style="list-style-type: none"> - PubMed based query (71 terms) : Sensitivity*: 90.0 (87.9 to 92.2); Specificity**: 97.2 (97.0 to 97.4) ; Precision: 33.2 (31.2 to 35.2) <p><i>4. Hedges (this report):</i></p> <ul style="list-style-type: none"> - Sensitive query (5 terms) ‡: Sensitivity*: 99.9 (99.6 to 100); Specificity**: 52.0 (51.6 to 52.5) ; Precision: 3.14 (2.92 to 3.37)

Table 9. InterTASC - Wylczynski 2007 (Continued)

- sensitivity>specificity (3 terms)§: Sensitivity*: 98.0 (97.0 to 99.0); Specificity**: 90.8 (90.5 to 91.1); Precision:14.2 (13.3 to 15.2)

- Balanced query, specificity>sensitivity (5 terms)¶: Sensitivity*: 90.2 (88.1 to 92.3); Specificity**: 98.4 (98.3 to 98.5); precision: 46.5 (43.9 to 49.0)

- Specific query (3 terms)††: Sensitivity*: 71.2 (68.0 to 74.4); Specificity**: 99.2 (99.1 to 99.3); Precision:57.1 (53.9 to 60.3)

*Validation (n=753). ** Validation (n=48 275). †Numbers vary by row. ‡search:.tw. or meta-analysis.mp.pt. or review.pt. or di.xs. or associated.tw. §meta-analysis.mp.pt. or review.pt. or search:.tw. ¶Cochrane database of systematic reviews.jn. or search.tw. or meta-analysis.pt. or Medline.tw. or systematic review.tw. ††Medline.tw. or systematic review.tw. or meta-analysis.pt.

F.4 Include the citations of any compared filters. Yes ([Shojania 2001](#) ; [Hunt 1997](#))

F.5 Other observations and / or comments. No

G. Other comments. This section can be used to provide any other comments. Selected prompts for issues to bear in mind are given below.

G.1 Have you noticed any errors in the document that might impact on the usability of the filter? No

G.2 Are there any published errata or comments (for example in the MEDLINE record)? No

G.3 Is there public access to pre-publication history and / or correspondence? No

G.4 Are further data available on a linked site or from the authors? No

G.5 Include references to related papers and/or other relevant material. No

G.6 Other comments. No

APPENDICES

Appendix 1. Search strategy

Medline (Ovid)

- 1 search*.ti.
- 2 strateg*.ti.
- 3 filter.ti,ab.
- 4 filters.ti,ab.
- 5 retriev*.ti.
- 6 identif*.ti.
- 7 locat*.ti.

8 find*.ti.
9 (indexing or indexed).ti.
10 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9
11 systematic review*.ti.
12 meta analys*.ti.
13 metaanalys*.ti.
14 11 or 12 or 13
15 10 and 14
16 review*.ti.
17 filter*.ti.
18 filters.ti.
19 1 or 17 or 18
20 16 and 19
21 15 or 20

Embase (Elsevier.com)

#1. search*:ti
#2. strateg*:ti
#3. filter:ti,ab
#4. filters:ti,ab
#5. retriev*:ti
#6. identif*:ti
#7. indexing:ti OR indexed:ti
#8. locat*:ti
#9. find*:ti
#10. #1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9
#11. systematic AND review*:ti
#12. meta AND analys*:ti
#13. metaanalys*:ti
#14. #11 OR #12 OR #13
#15. #10 AND #14
#16. review*:ti
#17. filter:ti
#18. filters:ti
#19. #1 OR #17 OR #18
#20. #16 AND #19
#21. #15 OR #20

PsycINFO (Ovid)

1 search*.ti.
2 strateg*.ti.
3 filter.ti,ab.
4 filters.ti,ab.

5 retriev*.ti,ab.

6 identif*.ti.

7 (indexing or indexed).ti.

8 locat*.ti.

9 find*.ti.

10 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9

11 systematic review*.ti.

12 meta analys*.ti.

13 metaanalys*.ti.

14 11 or 12 or 13

15 10 and 14

16 review*.ti.

17 filter.ti.

18 filters.ti.

19 1 or 17 or 18

20 16 and 19

21 15 or 20

LISTA (EBSCO)

S1 TI search*

S2 TI strateg*

S3 TI filter OR AB filter

S4 TI filters OR AB filters

S5 TI retriev*

S6 TI identif*

S7 TI indexing OR indexed

S8 TI Locat*

S9 TI Find*

S10 S1 OR S2 OR S3 OR S4 OR S5 OR S6 OR S7 OR S8 OR S9

S11 TI systematic review*

S12 TI meta analys*

S13 TI metaanalys*

S14 S11 OR S12 OR S13

S15 S10 AND S14

S16 TI review*

S17 TI filter

S18 TI filters

S19 S1 OR S17 OR S18

S20 S16 AND S19

S21 S15 OR S20

Science Citation Index (Web of Science; Clarivate)

#1 TI=(search* OR strateg* OR filter OR filters OR retriev* OR identif* OR indexing OR indexed) OR AB=(filter OR filters)

#2 TI=(systematic review* OR meta analys* OR metaanalys*)

#3 #1 AND #2

#4 TI=(review*)

#5 TI=(filter OR filters)

#6 TI=(search*)

#7 #5 OR #6

#8 #4 AND #7

#9 #3 OR #8

Appendix 2. UK InterTASC Information Specialists' Sub-Group (ISSG) Search Filter Appraisal Checklist

A. Information

A.1. State the author's objective.

A.2. State the focus of the research.

- Sensitivity-maximizing
- Precision-maximizing
- Specificity-maximizing
- Balance of sensitivity and specificity/precision
- Other

A.3. Database(s) and search interface(s).

A.4. Describe the methodological focus of the filter (e.g. RCTs).

A.5. Describe any other topic that forms an additional focus of the filter (e.g. clinical topics, such as breast cancer; geographic location, such as Asia; or population grouping, such as paediatrics).

A.6. Other observations.

B. Identification of a gold standard (GS) of known relevant records

B.1. Did the authors identify 1 or more gold standards (GSs)? None/1/2/3/4/5/More than 5

B.2. How did the authors identify the records in each GS?

B.3. Report the dates of the records in each GS.

B.4. What are the inclusion criteria for each GS?

B.5. Describe the size of each GS and the authors' justification, if provided (e.g. the size of the GS may have been determined by a power calculation).

B.6. Are there limitations to the gold standard(s)? Yes/No/Unclear

B.7. How was each GS used?

- To identify potential search terms
- To derive potential strategies (groups of terms)
- To test internal validity
- To test external validity
- Other, please specify

B.8. Other observations.

C. How did the researchers identify the search terms in their filter(s)? (Select all that apply)

C.1. Adapted a published search strategy. Yes/No/Unclear (please describe)

C.2. Asked experts for suggestions of relevant terms. Yes/No/Unclear (please describe)

C.3. Used a database thesaurus. Yes/No/Unclear (please describe)

C.4. Performed statistical analysis of terms in a GS set of records (see B above). Yes/No/Unclear (please describe)

C.5. Extracted terms from the GS set of records (see B above). Yes/No/Unclear (please describe)

C.6. Extracted terms from some relevant records (but not a GS). Yes/No/Unclear (please describe)

C.7. Tick all types of search terms tested.

- Subject headings
- Text words (e.g. in title, abstract)
- Publication types
- Subheadings
- Check tags
- Other, please specify

C.8. Include the citation of any adapted strategies.

C.9. How were the (final) combination(s) of search terms selected?

C.10. Were the search terms combined (using Boolean logic) in a way that is likely to retrieve the studies of interest?

C.11. Other observations.

D. Internal validity testing (this type of testing is possible when the search filter terms were developed from a known GS set of records.)

D.1. How many filters were tested for internal validity?

For each filter report the following information.

D.2. Was the performance of the search filter tested on the GS from which it was derived? Yes/No/Unclear (please describe)

D.3. Report sensitivity data (a single value, a range, "Unclear"* or "Not reported", as appropriate).

D.4. Report precision data (a single value, a range, "Unclear"* or "Not reported", as appropriate).

D.5. Report specificity data (a single value, a range, "Unclear"* or "Not reported", as appropriate).

D.6. Other performance measures reported.

D.7. Other observations.

E. External validity testing (this section relates to testing the search filter on records that are different from the records used to identify the search terms.)

E.1. How many filters were tested for external validity on records different from those used to identify the search terms?

E.2. Describe the validation set(s) of records, including the interface.

For each filter report the following information.

Search strategies (filters) to identify systematic reviews in MEDLINE and Embase (Review)

- E.3. On which validation set(s) was the filter tested?
- E.4. Report sensitivity data for each validation set (a single value, a range, "Unclear" or "Not reported", as appropriate).
- E.5. Report precision data for each validation set (report a single value, a range, "Unclear" or "Not reported", as appropriate).
- E.6. Report specificity data for each validation set (a single value, a range, "Unclear" or "Not reported", as appropriate).
- E.6. Other performance measures reported.
- E.7. Other observations.

F. Limitations and comparisons

- F.1. Did the authors discuss any limitations to their research?
- F.2. Are there other potential limitations to this research that you have noticed?
- F.3. Report any comparisons of the performance of the filter against other relevant published filters (sensitivity, precision, specificity, or other measures).
- F.4. Include the citations of any compared filters.
- F.5. Other observations and/or comments.

G. Other comments (this section can be used to provide any other comments. Selected prompts for issues to bear in mind are given below.)

- G.1. Have you noticed any errors in the document that might impact on the usability of the filter?
- G.2. Are there any published errata or comments (e.g. in the MEDLINE record)?
- G.3. Is there public access to prepublication history and/or correspondence?
- G.4. Are further data available on a linked site or from the authors?
- G.5. Include references to related papers and/or other relevant material.
- G.6. Other comments.

HISTORY

Protocol first published: Issue 7, 2020

CONTRIBUTIONS OF AUTHORS

Camila Micaela Escobar Liquitay: participated in screening, data extraction and quality assessment. Wrote the final draft of the review.

Luis Garegnani: participated in screening, data extraction and quality assessment. Wrote the final draft of the review.

Virginia Garrote: participated in screening, data extraction and quality assessment. Provided feedback on the manuscript.

Ivan Solà: wrote the initial protocol and provided feedback and oversight of the review process.

Juan VA Franco: participated in screening, data extraction and quality assessment. Wrote the final draft of the review.

All authors approved the final version of the manuscript.

DECLARATIONS OF INTEREST

Camila Micaela Escobar Liquitay: none known.

Luis Garegnani: none known.

Virginia Garrote: none known.

Ivan Solà: none known.

Juan VA Franco: is a Contact Editor for the Cochrane Urology Group and Managing Editor for the Cochrane Metabolic and Endocrine Disorders Group. He was not involved in the editorial processing of this review.

SOURCES OF SUPPORT

Internal sources

- Instituto Universitario Hospital Italiano, Argentina
Provides funding for Camila Micaela Escobar Liquitay and Luis Garegnani
- Heinrich-Heine-University-Düsseldorf, Germany
Provides financial support to Juan VA Franco

External sources

- No sources of support provided

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

We were unable to assess the performance of filters considering the secondary outcomes and subgroup analysis due to the scarcity of data in the primary studies. We presented the results from different interfaces and topics separately.

We deleted the secondary outcome "Number of unique systematic reviews retrieved by each search strategy" as it was suggested to be of low importance during the review process.

NOTES

The methods section of this protocol has been adapted from a previous review on a similar topic ([Beynon 2013](#)).

INDEX TERMS

Medical Subject Headings (MeSH)

*Checklist; Databases, Bibliographic; MEDLINE; *Systematic Reviews as Topic

MeSH check words

Humans