Evaluating and optimizing the search process for systematic reviews

Mariann Mathisen

OsloMet Avhandling 2020 nr 8

OSLO METROPOLITAN UNIVERSITY
STORBYUNIVERSITETET

Evaluating and optimizing the search process for systematic reviews

Mariann Mathisen

PhD Thesis Library and Information Science
Department of Archivistics, Library and Information Science
Faculty of Social Sciences
OsloMet - Oslo Metropolitan University

Spring 2020

OsloMet Avhandling 2020 nr 8

ISSN 2535-471X
ISBN 978-82-8364-234-6

OsloMet – storbyuniversitetet
Universitetsbiblioteket
Skriftserien
St. Olavs plass 4,
0130 Oslo,
Telefon (47) 64 84 90 00

Postadresse:
Postboks 4, St. Olavs plass
0130 Oslo

Trykket hos Byråservice

Trykket på Scandia 2000 white, 80 gram på materiesider/200 gram på coveret

Acknowledgements

Many people have contributed to this thesis during the last five years. I wish to express a sincere gratitude to all of them, and some of them deserve a special thank. Firstly, I would like to thank OsloMet for giving me the opportunity to undertake a PhD by financing the project for three years and for administrative support throughout the whole period. I am also very thankful to my supervisor Marianne Lykke for her guidance and support during the project. Special thanks goes to co-supervisor Louise Forsetlund and Marit Johansen. I am very grateful for their constructive criticism and encouragement, their constant availability, and not least, for the many discussions we have had over numerous dinners. I would also like to thank Karianne Hammerstrøm for her very helpful and constructive feedback at my final seminar and at a subsequent meeting. Thanks also to fellow student Inger Beate Nylund for discussions and mutual support.

Warm thanks to colleagues at the library at Vestfold Hospital Trust for being so supportive and also assisting in the project. More specifically, I would like to thank Zdravka Golub and Julie Skattebu for translating the published Ovid MEDLINE search strategies into PubMed syntax, to Zdravka Golub for running the conventional Boolean search strategies in PubMed and registering the hit numbers of the ranked result lists, to Peggy Wilson Andersen for counting the words in the search strategies, and to Hege Sletsjøe (HS) for contributing on Scoping review I and for categorizing the reviews into complex and non-complex. Also a great thank to Ingvild Kirkehei (IK) at The Norwegian Police University College for contributing on Scoping review III.

Finally, I would like to express my gratitude to friends and family for the support I have received during these years, particularly to my daughters Lene and Elise, and my husband Dag. Lene has proofread the whole thesis and given writing advice throughout the whole period. Elise has given support on data analysis and presentations, and both have been great listeners and discussion partners. Dag has been caring and supportive in every way. I have especially profited by his technical skills, and his great knowledge of Word and Excel has been invaluable to me. I am also extremely thankful to all three for never stop believing in me. Special thanks also to my three grandsons, Ferdinand, Vilhelm and Wilfred, for just being there and reminding me that life is so much more than hard work.

Abstract

Introduction: Systematic reviews aim to find and synthesize all relevant research as a basis for evidence based practice. The methods used in systematic reviewing, however, are extremely time and resource demanding. More efficient methods should be tested to make better use of scarce resources, and this study has examined three main approaches to simplify and optimize the search process for systematic reviews of healthcare interventions.

Methods: First, a cross sectional study was performed to investigate whether searching MEDLINE and Embase was enough for study identification in systematic reviews with respect to database coverage and search performance in a sample of 400 Cochrane reviews. Next, an exploratory design was used to find the optimal combination of search terms in a subsample of 254 of the sample reviews including 5 to 50 primary studies. Simpler search strategies containing the optimized search terms were developed and tested in Ovid MEDLINE and PubMed. Last, to test an automated information retrieval method for systematic reviews, the ranking function in PubMed was tested on conventional Boolean search strategies. Three scoping reviews were performed to determine what research exists on the main approaches and to inform the empirical studies.

Results: MEDLINE alone had a coverage of 84 % and MEDLINE and Embase combined 90 % in the sample of 400 reviews. The published Ovid MEDLINE search strategies of the sample reviews had 88 % recall and 2.8 % precision. In the subsample of 254 reviews, coverage was 85 % for MEDLINE alone and 90 % for MEDLINE and Embase combined. Recall was 87 % and precision 2.8 % for the search strategies of the subsample, which means that number of included studies in the sample reviews did not influence the total average values to any degree for this sample. An average of 7.2 search terms in each Ovid MEDLINE search strategy was necessary to retrieve the included studies of the subsample. That is approximately one tenth of the number of search terms used in the published search strategies. An average of 3.0 search terms per review occurred in the review title and 2.8 in the rest of the protocol. Average MAP for all reviews was 7.2 % for the ranked Boolean searches in PubMed, which were higher than for other sort options, but recall decreased at almost all cut-off values.

Conclusions: A high percentage of the included studies of the sample reviews in this study was indexed in MEDLINE and Embase, and the published Ovid MEDLINE search strategies had reasonably high performance in most cases. This means that it could be enough to search

only these two databases for some topics if other sources were searched in addition, and for some review groups, searching only MEDLINE could be enough. Based on the results of this study, conventional Boolean search strategies seem to be unnecessary long and complicated with an exaggerated use of search terms compared to what is necessary to retrieve relevant studies. For the sample of this study, most search terms could be found in the review title and/or rest of the protocol, except for reviews in Effective Practice and Organization of Care Group (EPOC) and other complex reviews. Ranking search results using “Best Match” in PubMed did not succeed in moving all or most relevant studies towards the top of the result lists and is therefore not recommended to simplify the search process for systematic reviews.

Sammendrag:

Innledning: Systematiske oversikter har som mål å finne og oppsummere all relevant forskning som grunnlag for en kunnskapsbasert praksis. Metodene som brukes i systematiske oversikter er imidlertid ekstremt tid- og ressurskrevende. Mer effektive metoder bør testes for å utnytte knappe ressurser bedre, og denne studien har undersøkt ulike tilnærminger for å forenkle og optimalisere søkeprosessen for systematiske oversikter i helsetjenesten.

Metoder: En tverrsnittstudie ble utført for å undersøke om det var nok å søke databasene MEDLINE og Embase for å identifisere de inkluderte studiene i et utvalg på 400 Cochrane-oversikter med hensyn til databasedekning og søkeprestasjon. Et eksplorativt design ble brukt for å finne den optimale kombinasjonen av søkeord i de oversiktene som inkluderte fra 5 til 50 primærstudier i utvalget (=254). Enklere søkestrategier som inneholdt de optimaliserte søkeordene, ble utviklet og testet i Ovid MEDLINE og PubMed. For å undersøke en automatisert metode for informasjonsgjenfinning til systematiske oversikter, ble rangeringsfunksjonen i PubMed testet på både konvensjonelle og enkle søkestrategier.

Resultater: Dekningen av inkluderte studier i MEDLINE alene var 84 % og MEDLINE og Embase 90 % i utvalget på 400 oversikter. De publiserte Ovid MEDLINE-søkestrategiene for de samme oversiktene hadde en fullstendighet på 88 % og 2,8 % presisjon. I utvalget på 254 oversikter var dekningen 85 % for MEDLINE alene og 90 % for MEDLINE og Embase kombinert. Fullstendighet var 87 % og presisjon 2,8 % for søkestrategiene i dette utvalget. Dette betyr at antall inkluderte studier i oversiktene ikke påvirket de totale gjennomsnittsverdiene i noen særlig grad for dette utvalget. Et gjennomsnitt på 7,2 søkeord i hver Ovid MEDLINE-søkestrategi var nødvendig for å finne de inkluderte studiene i utvalget på 254 oversikter. Det er omtrent en tidel av antallet søkeord som ble brukt i de publiserte søkestrategiene. I gjennomsnitt forekom 3,0 søketermer i tittelen og 2.8 i resten av protokollen. MAP-verdier for rangerte søk i PubMed var høyere enn for andre sorteringsalternativer, men fullstendigheten falt ved nesten alle cut-off-verdier.

Konklusjon: En høy prosentandel av de inkluderte studiene i oversiktene i denne studien var indeksert i MEDLINE og Embase, og de publiserte Ovid MEDLINE søkestrategiene presterte rimelig høyt i de fleste tilfeller. Det betyr at det kan være nok å søke bare i disse to databasene for noen oversikter hvis andre kilder blir søkt i tillegg, og for enkelte oversikter, kan det være nok å søke bare i MEDLINE. Basert på resultatene fra denne studien, ser konvensjonelle boolske søkestrategier ut til å være unødvendige lange og kompliserte med en overdreven

bruk av søkeord sammenlignet med hva som er nødvendig for å finne relevante studier. For utvalget i denne studien fantes de fleste søketermer i tittel og resten av protokollen, bortsett fra oversikter fra Effective Practice and Organization of Care Group (EPOC) og andre komplekse oversikter. Rangering av søkeresultater ved bruk av Best Match i PubMed lyktes ikke med å flytte alle eller de fleste relevante studier mot toppen av resultatlistene og anbefales derfor ikke for å forenkle søkeprosessen til systematiske oversikter.

Content

List of tables

List of figures

Abbreviations

AHRQ – Agency for Healthcare Research and Quality

ATR - Automatic Term Recognition

EBP – Evidence Based Practice

HTA – Health Technology Assessment

MAP – Mean Average Precision

MECIR - Methodological Expectations of Cochrane Intervention Reviews

MEDLINE - Medical Literature Analysis and Retrieval System Online

MeSH - Medical Subject Headings

NCBI - National Center for Biotechnology Information

NLM – National Library of Medicine

NNR – Number Needed to Read

PRISMA - Preferred Reporting Items for Systematic Reviews and Meta-Analyses

1 Introduction

Systematic reviews are increasingly used to inform practice and provide an evidence base for policy making. They use a transparent and systematic process to define a research question, search for studies, assess their quality and synthesize findings qualitatively or quantitatively (1). Systematic reviews are conducted through a robust, but sometimes very slow and resource-intensive process, and searching for research evidence is a laborious stage of this process (2, 3). New tools and methods of keeping up with this expanding workload are necessary to save resources and reduce the time it takes to produce a systematic review (3-6). The main work of this study is to evaluate and optimize the search process for systematic reviews in healthcare with the aim of identifying approaches that are more efficient than conventional search approaches.

1.1 Background

Before undertaking any new policy, practice or research, it is essential to find out what is already known about an issue in a fair and unbiased manner (7, 8). The amount of research information, however, is so great that it has become impossible for anyone to have the time and resources to find, appraise and critically evaluate this evidence, and incorporate it into healthcare decisions (9). Therefore, the continuous summarization of research evidence has long had high priority in healthcare to give users of health information access to high quality research. Summarized knowledge is important for politicians and health personnel, but also for lay people (10). Patients need easy access to high quality information to make informed choices concerning their treatment options (11).

1.1.1 Evidence Based Practice

To facilitate better access to summarized research evidence, Evidence Based Practice (EBP) was introduced in the health services in the beginning of the 1990s. Since then, EBP has been discussed and described in numerous books, articles, and web-sites both nationally and internationally (12-14). The goal of EBP is to provide high quality services by integrating expert opinion with current best scientific evidence and the client or user perspective. The step-by-step, circular process of EBP starts with reflection and identification of an information need. The next step is to formulate a precise research question before the best evidence with which to answer these questions is found and critically appraised. The results of the appraisal is integrated with clinical expertise, the preferences of the unique patient, and the context or setting in general. Finally, practice is evaluated and improvements are made when necessary.

1.1.2 Systematic reviews

The systematic review is considered the cornerstone of EBP and can be defined as a scientific method to identify, evaluate, summarize, and communicate the results of an otherwise unmanageable quantity of research (15, 16). The purpose of systematic reviews is to provide healthcare practitioners with an opportunity to stay up-to-date and is an important source for health professionals, policy-makers, and patients to make informed decisions. The results from individual studies may vary, and it is only when the entire knowledge base is examined that it is possible to make a more certain statement about what is known on a topic. Without systematic reviews of previous research, ineffective or even harmful interventions may be used because they are thought to be effective, and conversely, effective interventions may be considered ineffective and withheld (17). The aim of systematic reviewing is to arrive at a more comprehensive and trustworthy picture of the topic being studied than is possible from single studies (18).

The increasing use of EBP and systematic reviews has been described as marking a paradigm shift in clinical practice in healthcare (19). The development of a more systematic method for reviewing and synthesizing literature grew out of the recognition that traditional, non-systematic reviews lack scientific rigour. The same scientific principles that are applied to the design and conduct of primary research should also be applied to the process of reviewing that research (20). Reviews are more valuable if they are systematically developed according to internationally agreed standards or guidelines with transparent and verifiable methods. Then decisions can also be discussed and challenged by others (21). Systematic reviews should provide a comprehensive and balanced picture of the knowledge state and can contribute to new knowledge but can also discover knowledge gaps and provide way for further research (22).

There are several guidelines to assist in writing a systematic review describing the various steps of the process (23-25). The first step in the preparation of a systematic review is a well-formulated and focused research question. Furthermore, a systematic review is characterized by a clear title and a clear purpose before a comprehensive search strategy is developed and performed to find all relevant research results on a topic. Criteria for inclusion and exclusion of primary studies must be stated, and for transparency and reproducibility, the whole search process with at least one search strategy should be documented. The quality of the included studies must be assessed and characteristic features of all included studies described. Compilation of the results from included studies may consist of a descriptive summary or one or more meta-analyses. Good systematic reviews can support an evidence-based practice and serve as an evidence base in the development of guidelines and other practice sources (26). Systematic reviews are widely used to inform decision making but should also be used to evaluate the need for new research and thereby avoid waste of resources and unnecessary harm on study participants (8, 20). Whilst the systematic review is often associated with medical research and randomized controlled trials (RCTs), it is now a widely accepted research method conducted in an increasing number of academic, scientific and policy fields (27-32) to address any research question using any relevant type of research (25, 33).

1.1.3 The Cochrane Collaboration and Cochrane reviews

To enhance the development of systematic reviews, The Cochrane Collaboration was founded in 1993 and is a non-profit, independent research network dedicated to making up-to-date, accurate information about the effects of healthcare (34). The Cochrane Collaboration has been named after the British epidemiologist Archie Cochrane, who, ever since his student days in the 1930's, had strongly criticized the lack of reliable evidence behind many of the commonly accepted healthcare interventions at the time (35). He was concerned that most decisions about interventions in healthcare were based on an unstructured choice of information of varying quality (36). He called for a collection of systematic reviews organized by specialty or subspecialty and adapted periodically (37, 38), and today the Cochrane systematic reviews, published online in the Cochrane Library, are regarded as the gold standard for systematic reviews (9). They assess the benefits and harms of interventions used in healthcare and health policy (15). Cochrane reviews are prepared by health professionals worldwide and most of them work voluntarily for one of the 53 Cochrane Review Groups. In September 2019, a total of 8103 Cochrane reviews had been published by all review groups (39).

All Cochrane review groups have an editorial board with overall responsibility for developing and updating the reviews which ensures compliance with high quality standards. The quality standards to which all Cochrane protocols, reviews, and updates are expected to adhere are The Cochrane Handbook for Systematic Reviews of Interventions (the Cochrane Handbook) (23) and the Methodological Expectations of Cochrane Intervention Reviews (MECIR) standards (40). The Cochrane Handbook is the official guide that describes the whole process of preparing Cochrane systematic reviews. It provides a detailed description of each step in the process of developing and maintaining a review, and chapter six, "Searching for studies", describes the search requirements in detail. In 2013, the MECIR standards were introduced in order to ensure the highest possible quality of the Cochrane reviews (40). MECIR are standards for the conduct and reporting of new Cochrane Intervention Reviews. They are compliant with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement and are drawn from the Cochrane Handbook (41). PRISMA is an evidence-based minimum set of items for reporting in systematic reviews and meta-analyses (42).

Traditionally, systematic reviews have been conducted mainly to assess the effectiveness of health interventions (43), which applies to most Cochrane reviews. Intervention reviews differ in several aspects, including complexity. Complex interventions are usually described as interventions that contain different interacting components with several dimensions of complexity, but there is no sharp boundary between simple and complex interventions, that is, few interventions are truly simple, but there is a wide range of complexity (44). Moore et al. wrote in an editorial in Cochrane Database of Systematic Reviews on assessment of complexity that all interventions are complex, but some are more complex than others (45).

1.1.4 Systematic searching

With EBP and systematic reviews, there has been an increased recognition of the importance of information retrieval as part of the scientific process. To perform an exhaustive and comprehensive search is considered good practice in systematic reviewing (46). It is one of several stages in the systematic review process (3) and is referred to as a systematic search (47). A systematic search is expected to be performed according to special guidelines (23) and comprises (at least) three distinct stages: planning, performing and documenting the search. When planning a systematic search, the first to decide is which sources to use. Decisions concerning the choice of sources are for example which databases to search, if and how citations will be tracked, and what and how evidence will be identified in non-database sources, such as reference checking and contacting experts. In general, the traditional bibliographic databases give the highest yield of relevant studies and are therefore considered the most important sources to find research evidence (48). The next step in the planning process is to decide which words and concepts the search should contain and how they will be combined, that is, an exhaustive list of search terms are combined with the Boolean operators AND and OR. These often quite complex Boolean queries are iteratively developed and performed in several databases, and eventually, each reference of the search results is assessed for relevance (23). Finally, to ensure transparency and reproducibility, the whole search process should be reported with at least one search strategy attached (49).

Systematic reviews are produced by a team of systematic reviewers, and librarians or other information specialists are important members of the team (50). They manage the search process in its entirety, from designing the search strategy to conducting and documenting the search and managing the references (51). They must also be able to understand the complex information problems of health care professionals, the research scenario, and why a systematic review is required (52). The librarian's intellectual contribution to the review is central, and librarians should be included as authors in the published review (52). Librarians should also "engage in producing systematic reviews on topics of interest in the practice of librarianship and research in information science." (53)

A comprehensive systematic search is based on a broad search strategy performed in a wide selection of sources to retrieve as many relevant primary studies as possible. There is, however, a trade-off between exhaustiveness and comprehensiveness on the one hand and timeliness and use of resources on the other (54).

1.2 Systematic review challenges

There are different challenges in systematic reviewing, and one of the most fundamental is to identify all of the potentially relevant studies on a topic (20). Comprehensive searching for an unbiased, reproducible set of primary studies is a defining feature of every systematic review. It will generally increase the probability that all or at least a representative selection of relevant studies are identified and is a major factor distinguishing systematic from non-systematic reviews (55). Comprehensive searches, however, can be time-consuming and also costly to perform as access to many bibliographic databases often requires a paid subscription. Furthermore, using multiple databases complicates the logistics of the search process and requires expertise in using the interfaces and syntax of the different databases (56). The large and growing number of published studies in healthcare and their increasing rate of publication make the task of identifying relevant studies in an unbiased way even more time consuming (57). Producing a complex systematic review can take more than a year to complete with up to half of that time being spent searching and screening hits (7). This is problematic because policy-makers, clinicians, and other practitioners often need to know the state of the research evidence faster. The long and cumbersome process can lessen the likelihood that research evidence will be used at all, and policies and practices without a firm evidence base can have unintended, unwanted effects and occasionally do more harm than good (17). Unsurprisingly, there has been a growing interest in the development of simpler, more efficient search methods to reduce the burden of producing systematic reviews (5, 58).

1.3 Simplifying search approaches

The validity of systematic reviews is highly dependent on their including an unbiased sample of relevant studies (59). If a systematic review does not identify all or most available data on a topic, evidence selection bias can occur (60). This can arise from publication bias, where data from statistically significant studies are more likely to be published than those that are not statistically significant (60). To avoid this kind of bias, comprehensive searches are performed in a wide variety of sources. There is, however, a trade-off between timeliness, use of resources, and comprehensiveness. When time or financial resources are limited, simpler, more efficient methods have to be tried out (61, 62), and this study has investigated different approaches to simplify and optimize the search process as specified below.

In order not to miss any relevant research studies on a topic, several electronic databases are usually searched for each systematic review. This is a time consuming task, so reducing the number of databases to search can reduce the amount of work involved in searching and screening the search results and hence, be time and cost effective. In particular, costs will be reduced if only MEDLINE is searched, as this database is freely available through the PubMed interface (62). Several studies have been conducted earlier on the effect of reducing the number of databases to search in systematic reviewing. Some of them support a reduction (62, 63), while others claim that searching a wide selection of databases is necessary to find all relevant studies on a research question (64, 65). To contribute to this discussion, one approach of the present study is to investigate whether searching only MEDLINE and Embase is enough for a sample of Cochrane reviews.

Another important task when planning a systematic search is to decide which search terms to use. The prevailing method to develop a search strategy is for a librarian or other expert searcher to select a wide variety of search terms in close collaboration with the rest of the review team, often based on relevant words in the review protocol and/or a few relevant publications on the topic. Time constraints play a decisive role in the development of search strategies and especially in finding useful search terms, as the searchers often have no expert knowledge on many of the topics under investigation. This means that they must become acquainted with different topics within a short period of time to be able to develop valid search strategies (51). Some researchers find this approach too subjective and have tested statistical methods to find search terms and claim that this will both be simpler and give a more transparent and reproducible result (66). There is today, however, no common agreement on how to select the most useful terms for comprehensive searches. There is also little knowledge on the best sources for feasible search terms and how many terms are necessary to find relevant research studies. To investigate the need for search terms in systematic reviews, this study will find the optimal combination of search terms that retrieve the included studies in a sample of Cochrane reviews, use them in simple search strategies, and investigate whether the optimized search terms occur in the review title or rest of the protocol.

After deciding which sources to search and a list of useful search terms have been selected, a search must be performed. There are principally two main information retrieval techniques and that is exact match (Boolean retrieval) and partial match (ranked or automated retrieval). Boolean or exact match searching is the conventional retrieval method to find primary research studies to include in systematic reviews. Recently, however, more automated or partial match methods have been suggested for systematic review searches as a simpler, more objective alternative. Automated methods can make the searching easier and reduce the time it takes to conduct the search. Different text mining and ranking algorithms have been developed and evaluated for information retrieval in systematic reviews to increase effectiveness and efficiency of the search process (61, 67). Unfortunately, most of these new, more automated search systems are not available for public use yet (68), and some of them have been withdrawn (69). An existing, available database that offers the possibility to perform both partial and exact match searches is PubMed. PubMed is a collection of databases, including MEDLINE, the most important medical database, produced by the National Library of Medicine (NLM) in the USA. It is the database that clinicians generally use, due to free access and a simple search interface, but PubMed is used by systematic reviewers as well, since the database also allows for complicated Boolean queries and use of field codes. When performing searches in PubMed without field codes, an automatic term mapping function will help the users by expanding the search to include more search terms. To further automate the search process in PubMed, a ranking function has been introduced as a possibility to present the search results with the most relevant hits first. Using these automated search methods in PubMed for systematic reviews might save time and money and make the searching easier, and this study will test the ranking function in PubMed on conventional Boolean searches.

1.4 Aim and objectives

The main aim of this study is to investigate the need for comprehensiveness in systematic review searching and whether it is possible to improve efficiency of the search process by reducing the effort and resources used without noticeably degrading the effectiveness of the process. To achieve this aim, this study will evaluate both conventional and simple, more optimized search strategies.

1.5 Research questions

This research study addresses the following questions pertinent to comprehensiveness of the search process for systematic reviews in healthcare:

1. Searching only MEDLINE and Embase for study identification:
    a. Is searching the databases a) MEDLINE alone or b) MEDLINE and Embase combined sufficient for study identification in terms of database coverage and search performance?
2. Selection of search terms and development of simple, optimized search strategies:
    a. What methods have been used to select search terms and develop simple, more optimized and efficient search strategies?
    b. What are the characteristics of optimized search terms and what is the performance of simple, optimized search strategies?

3. Automated search methods:
    a. Which automated search methods have been tested and/or described to improve the search process, and how do they perform?
    b. How does extended Boolean information retrieval perform, more specifically, the relevance ranking function in PubMed on conventional Boolean search strategies?

1.6 Usefulness and relevance of the study

A study on simpler, more optimized search approaches for systematic reviews is important for several reasons and may have considerable implications for those involved in systematic review development. The results can reveal whether less comprehensive search methods can maintain effectiveness but be more efficient than conventional search methods. Simpler search methods can save time for researchers and their organizations producing systematic reviews and thereby lead to better use of scarce resources. Waste can be avoided, and more reviews can be produced and published faster (70). Simpler search methods can also have positive implications for anyone planning a research project. All research projects should start with an exhaustive search on existing literature (8). Failure to perform an adequate search for existing literature before starting a new research project is even included in a taxonomy of research misconduct (71), and simpler search methods will make it easier for researchers to fulfil this requirement.

1.7 Thesis structure

This thesis is about simplifying the search process for systematic reviews in healthcare and has evaluated both conventional and simpler, more optimized search approaches. In this first chapter, background information on EBP, systematic reviews, and systematic searching is outlined. The problem is introduced, aim, objectives, and research questions are presented, and the potential usefulness of the study is described. The second chapter presents theory related to the search process for systematic reviews relevant to this study. The third chapter presents scoping reviews on three approaches to simplify the search process for systematic reviews, and the fourth describes the different methods used in the empirical studies. The results of the scoping reviews in chapter three were used to develop and inform the empirical studies described in chapter four. The results of the empirical studies are presented in chapter five, and in chapter six, the discussion is outlined with a summary of the key findings and a presentation of conclusions.

1.8 Ethical considerations

As this is a study of published information and no human participants are involved, no ethics approval is required.

1.9 Summary

The production of new research increases every day, and it is impossible to keep up with the development in almost any field of interest (5). The use of systematic reviews is one way out to cope with this data deluge. To have high quality research results summarized in easily accessible sources is beneficial for clinicians, public authorities, and research communities but also for lay people. Systematic reviews can give a more reliable picture of the knowledge state of a specific question than reviews with a less rigorous and transparent approach. They must, however, be conducted faster than today to save resources and present up-to-date research results. Comprehensive searching is the conventional way to find research studies to include in systematic reviews, but there is no agreed-upon standard for how comprehensive the search process must be. In an effort to assess the need for comprehensiveness, this study will evaluate both conventional and simple, more optimized search approaches. The need for comprehensiveness will be estimated by evaluating a reduction in the number of databases to search and by developing and testing simplified, more optimized search strategies. To test an automated information retrieval method for systematic reviews, the “Best Match” ranking function in PubMed will be evaluated. Systematic reviews are the best sources for health personnel, politicians, and others to use in decision-making and staying up-to-date on a topic. They must, however, be produced faster than today to save time and resources.

2 Information retrieval in systematic reviews

The following chapter will describe theory that are important to the search process in systematic reviews.

2.1 The systematic search process

All systematic reviews should be based on a systematic search. A systematic search can be defined as a comprehensive, objective and reproducible search using a range of different sources to identify as many relevant studies as possible on a topic (72). The time required to conduct a search for a systematic review will vary. It is dependent on the review question, the breadth and complexity of the evidence base, and the scope of the proposed search as stated in the review protocol (73), but generally it is considered a time-consuming task (3, 74, 75). The systematic search process consists of three main stages: planning, performing and documenting the search, which will be further described in the next sections.

2.1.1 Planning the search

Planning a systematic search consists of all steps taken to prepare the search. Two important decisions at the planning stage are which sources to search and which search terms to include in a search strategy.

2.1.1.1 Selection of sources

One of the first decisions to make when planning a comprehensive search for systematic reviews, is which sources to search since this is likely to influence the amount and type of information retrieved (76). Studies can be identified from different sources, for example through citation searching, hand-searching journals, checking reference lists, or contacting experts. Electronic databases give the highest yield, however, and are regarded as the most important sources to find evidence for systematic reviews. The choice of databases depends on several factors. Most important is the topic under consideration and the content or topics covered by the databases. Another consideration involves the database platform. Different vendors or interfaces offer different indexing and searching possibilities. Searching several databases implies that expert searchers must have specific knowledge of different databases, each with its own search syntax, metadata, and vocabulary since interoperability among databases is rare. There are many databases that can be used to identify biomedical studies. As a minimum, the Cochrane Handbook recommends that for all Cochrane reviews, CENTRAL[1] and MEDLINE should be searched, together with Embase if it is available (72).

MEDLINE is the most important and commonly used electronic database in systematic reviews of biomedical research and is freely available through the PubMed interface (56, 77). PubMed contains several databases with MEDLINE as the most important and is developed and maintained by the National Center for Biotechnology Information (NCBI), at NLM. It is a bibliographic database that indexes from more than 5600 biomedical journals (78) and comprises more than 24 million records for biomedicine and health from 1946 onward (79). To ease the retrieval of information from MEDLINE, references are supplied with subject headings called MeSH, which is the NLM's controlled vocabulary thesaurus. It consists of sets of terms naming descriptors in a hierarchical structure that permits searching at various levels of specificity (80). Records in MEDLINE are assigned several MeSH-terms each, and a few of these are assigned status as Major Mesh, that is, MeSH terms tagged as the major focus of references retrieved (81). Access to MEDLINE is offered by different vendors, and one of the most well-known is Ovid, a software company specializing in text retrieval applications and search interfaces, which provides an easy to use interface for searching MEDLINE. One possibility available in Ovid MEDLINE and not in PubMed is the use of proximity operators, which makes the searching more flexible. A proximity operator searches for terms near each other with a specified number of words between the search terms.

Embase is a biomedical database published by Elsevier, which covers the most important international biomedical literature from 1947 to the present day. All records are indexed using Elsevier's thesaurus Emtree (82). As from 2010, MEDLINE records are included in Embase, whereas some Embase records are not covered by MEDLINE. Embase covers other journals, especially drug therapy journals, more European journals, and more non-English journals compared with MEDLINE (62).

Although MEDLINE and Embase cover most of the important medical and other health science journals, they may not report all relevant research. CENTRAL collects the studies found for producing the Cochrane systematic reviews and is the most complete source of clinical trials (9). Each Cochrane review group creates a specialized register containing bibliographic references of clinical trials retrieved by searching different databases worldwide

[1] The database CENTRAL is now called Trials. Usage of the old and new name varies throughout Cochrane documents, however, and will be referred to as CENTRAL in this thesis.

and also by hand-searching the literature on the topic covered by the group. CENTRAL comprises references from various sources, among others the specialized registers and MEDLINE and Embase (83).

Other databases that might also be searched for systematic reviews in health and medicine are Scopus, Web of Science, Biosis, and other more general databases offering a large coverage of health related research. In addition, specialized databases like PsycINFO (mental health), CINAHL (nursing), PEDro (physiotherapy), and regional databases like African Index Medicus and LILACS should be searched for some systematic reviews depending on the topic. Search engines like Google Scholar are also suggested as a source for research evidence (84). Due to publication bias, that is, the tendency only to publish studies with positive results, grey literature and not/not yet published literature are also important sources for systematic reviews (9).

There are both advantages and disadvantages to searching multiple databases. One disadvantage is that it is laborious for searchers to translate a search strategy into multiple interfaces and syntaxes, as field codes and proximity operators differ between databases. Differences in thesaurus terms between databases add another burden for translation. Furthermore, access to databases is often limited and only available on a subscription basis, and it is time-consuming for reviewers who have to screen more, and likely irrelevant titles and abstracts. The main advantage is that the probability of retrieving all relevant primary studies increases. Due to variations in search performance, articles indexed in several database must not necessarily be retrieved in all databases, but only identified by one (85). It is, however, difficult to defend using time and other resources if the benefit is marginal.

2.1.1.2 Selection of search terms and development of search strategies

Selecting feasible search terms is an essential stage in the systematic search process. When developing a comprehensive search strategy, different methods are used to identify as many relevant search terms as possible not to miss any relevant studies on a topic (73). The systematic review process starts by writing a review protocol with a clearly defined review question. The review question provides the foundation for the search strategy, and without a well-focused question, it can be very difficult and time-consuming to identify appropriate resources and develop the search strategy for relevant evidence. A preliminary search can help refine the research question and determine its feasibility and scope (86). Once the review question is determined, the key terms articulated in the question and the protocol need to be identified and reduced into searchable concepts, a process that is being referred to as the conceptualization process (87).

Organizing the search terms

There are different approaches to the development and organization of search terms. One of the most commonly used search styles in systematic searching is the “building blocks” approach. Dividing a query into facets with variants and synonyms and then adding these concepts together using the Boolean operator AND, has become particularly popular when used with the Patient-Intervention-Comparison-Outcome (PICO) formulation (88). PICO was developed for reviews on the effectiveness of interventions and is used to form the question and facilitate the literature search in bibliographic databases (89).

P	Population (Patient / Problem)
I	Intervention
C	Comparison
O	Outcome

When constructing a search strategy and searching for evidence for systematic reviews, a logic grid or concept map should be created once the review question is determined, to show how related concepts or synonyms will combine to construct the final search strategy (73). In this logic grid, also called a PICO table, each column represents a discrete concept that is generally aligned with each element of PICO (73). An example is shown in Table 1 below.

Table 1 PICO example

Research question: Will light therapy, compared to no treatment, improve sleep and decrease aggressive behaviour in patients with dementia?			
P **Patient/Problem**	**I** **Intervention**	**C** **Comparison**	**O** **Outcome**
Dementia Alzheimer	Light therapy Light treatment Bright light Phototherapy Photo-therapy	No treatment	Improved sleep Less aggression

Search term sources and selection methods

When selecting search terms manually, there are various ways to help find the most relevant terms. Words in title, objectives, and inclusion criteria of a review protocol are considered good starting points to find search terms for the PICO table, but may not be considered enough. The next step in the development of a feasible search strategy is therefore to determine any further alternative terms or synonyms for the identified concepts. Synonyms do also include acronyms, abbreviations, and variants (for example English/American spelling and singular/plural). Several sources can help identify synonyms and related terms, for example, medical dictionaries such as MEDLINEPlus or the entry terms of the MeSH database. The use of reference works can be useful to find synonyms, but depending on topic it can also introduce terms that are outdated or rarely used and might only increase the screening burden. Discussion with subject specialists can also be a useful guide to terminology and concepts (90).

“Citation pearl growing” is a well-known method to find relevant search terms and is the process of using the characteristics of a relevant and authoritative publication, called a “pearl”, to search for other relevant material (91). The technique involves starting with a very precise search to find one or a few relevant citations, which are examined for subject headings and text words. Any new terms are incorporated into the search strategy, and this iterative process continues until all additional relevant terms are identified and included in the strategy. “Comprehensive pearl growing” is an “extended” version of “citation pearl growing” described by Schlosser et al. (92) and starts with a compilation of studies from a relevant review or a topical bibliography that is used to select search terms. Thus, rather than beginning with only one “pearl”, “comprehensive pearl growing” requires that the searcher begins with several “pearls” in order to retrieve more of the same kind (92). “Citation snowballing” is another method often used where references in an article is examined for further relevant terms. These iterative methods are valuable strategies to find useful search terms. They could be used routinely as a systematic prelude to the building blocks approach and when updating systematic reviews and is appropriate for generating a comprehensive list of terms required for systematic review searching (88). The limitation of such methods is that they expand a review in favour of literature that uses the same language as the documents that have already been found (2).

Development of search strategies

A record in a bibliographic database consists of many different fields, and a search may limit to any of these fields, for example, title, author, or year. An important aspect of a systematic search strategy is that it must include both subject headings (unique concepts from a controlled vocabulary) and words in title and abstract (text words) to secure high recall in the search results and identify articles not yet indexed or indexed differently due to different interpretations given by the indexers (9). The free-text terms may be truncated and used in varying combinations of proximity, where proximity is possible, for maximum recall and precision (86). Search strategies must be constructed for each database as terminology will vary across disciplines as will also thesauri and subject headings.

Balancing a search strategy

An ideal search strategy is both sensitive and specific. A sensitive search will recall relevant studies while a specific search will exclude irrelevant. A search that is overly sensitive may capture all relevant studies but may require a labour-intensive examination of unnecessary studies at the stage of study selection, and a search that is overly specific will yield fewer results but is always subject to the risk that important studies may have been omitted (73). Finding a balance between a comprehensive, highly sensitive search with the aim of retrieving all relevant studies regarding the topic and a more precise search, running the risk of missing important information, is required (9). This balancing is based on the iterative addition and removal of parts of the search strategy until a strategy is as optimal as possible. The word “balance”, however, is not precisely defined and remains a vague concept in searching (93).

A challenge in balancing a search strategy is that there is usually a trade-off between precision and recall. Generally, either can only be increased at the cost of the other. Thus, in conventional search development it is difficult if not impossible to tell when a strategy is completed. Also, if the retrieval rate is high, a subsequent restriction of the search is required and is time consuming. It is difficult to determine whether the retrieval of a large number of records is because there are a many eligible studies for the research question or because there is an overwhelming proportion of irrelevant material (94). The searcher can, however, analyse the database sources and indexing of the relevant included studies, to see if the search can be refined and narrowed or broadened depending on the results. In conditions where it is difficult to achieve a precise database search, attention should shift to ensure that the searches in other databases and non-database methods are optimized (95).

Filter for study design

A specific contribution of EBP is to apply a methodological hedge or filter for study design as a fifth facet in the PICO(S) framework. Search filters are collections of search terms intended to capture frequently sought research methods, for both quantitative research, such as RCTs (96) and qualitative research (97). The aim of search filters is to limit a search and reduce the number of references to screen.

2.1.2 Performing the search

The next stage of the search process is to execute the search strategy developed at the planning stage in a database or other search system. There are different information retrieval techniques, and the two most important are exact match (Boolean searching) and partial match (ranked retrieval).

2.1.2.1 Exact match retrieval – Boolean searching

The conventional method to find research studies to include in systematic reviews is Boolean searching. In Boolean searching users search a database with a query that connects words and search lines with operators from George Boole’s mathematical logic. A query is divided into different facets or concepts with variants and synonyms, and these search terms are added together using Boolean operators. The Boolean operator AND is used to narrow a search and retrieve references containing *all* of the words or lines it separates. OR broadens a search and retrieves references containing *any* of the words or lines it separates. NOT narrows a search and retrieves references that do *not* contain the word or line following it but is normally not recommended for systematic searching as it can exclude relevant studies. Truncation, also called stemming, is a technique often used in Boolean searching in which a word ending is replaced by a symbol. The most frequently used symbol is the asterisk (*).

2.1.2.2 Partial match retrieval

In general, most information retrieval today is based on partial match searching. Partial match information retrieval is a complex process aimed at producing a ranking function. The ranking function assigns scores to documents that reflects the likelihood of the document being relevant to the query, and once such scores are computed, the document corpus can be ordered by decreasing score, and the ranking presented to the user (61). There are several ranking functions, but they make use of similar information, such as term frequency, both in a document and across a collection (61).

2.1.2.3 Extended Boolean information retrieval

Extended Boolean information retrieval is an intermediate between the Boolean system of query processing and the vector-processing model (98). It combines a weighting system with Boolean techniques in such a way that documents that fulfil all demands in the search profile obtain the highest ranking. The goal of the extended Boolean retrieval model is to overcome the weaknesses of the Boolean model. Ranking within the Boolean result set can improve the search performance by providing early indication of the quality of the results and thus, speeding up the iterative query-refinement process (61). This suggests that an interactive query-development process using a hybrid ranked and Boolean retrieval system has the potential for significant time-saving over the current search process in systematic reviewing (61).

2.1.2.4 Text mining

Text mining or text analysis are techniques that have recently been applied in the systematic review process to discover and extract information of value (57, 99). There are several definitions of text mining, but all encompass an automated process that assists the identification, extraction, and structuring of patterns in the text of individual documents and across multiple documents (99). Text mining aims to enable users to collect, maintain, interpret, and discover knowledge efficiently and systematically (100). Among the numerous text mining methods, currently most relevant for supporting different parts of the systematic review process, are automatic term recognition (ATR), the identification of relevant literature, its rapid description, document clustering, classification, and summarization (2). ATR identifies and extracts terms from text and alters or expands query terms to find the most relevant (101).

2.1.2.5 Searching PubMed

Most database vendors offer both an “advanced search” interface based on Boolean logic and an interface based on partial match techniques, often called “basic search”. In the default search interface in PubMed, a search can be performed either in exact match or partial match mode depending on search technique. When searching exact match in PubMed, Boolean operators and field codes have to be applied. In a partial match search, “Untagged terms that are entered into the search box in PubMed are matched (in this order) against a MeSH translation table, a Journals translation table, the Full Author translation table, Author index, the Full Investigator (Collaborator) translation table and an Investigator (Collaborator) index. When a match is found for a term or phrase in one translation table, the mapping process is complete and does not continue on to the next translation table”. (79) On October 22^{nd,} 2013, “Relevance Sort” was implemented in PubMed (102), today called “Best Match”. This relevance sort order for search results is based on an algorithm that analyses each PubMed reference that includes the search terms. For each search query, "weight" is calculated for references depending on search term frequency and in which fields they are found. In addition, recently-published articles are given a somewhat higher weight for sorting. Sort options are available from the "Display Settings" menu under the "Sort by" selections (102).

2.1.3 Reporting the search

The third main stage in the systematic search process is the reporting stage. For systematic reviews of intervention, the MECIR standards (41) and the PRISMA statement (103) are used as guidelines for reporting. PRISMA states that all systematic reviews should have the search process documented in the methods section according to specific standards and the final search strategy of at least one database attached (49). For Cochrane reviews, searches should be reported in enough detail so that they could be reproduced by someone else if necessary, and the full search strategy for each source must be provided (104). This should be included in an appendix at the end of the review and linked to from the search methods section. Each search strategy for each source must be given as it was run in the database (104).

The documentation is important to make the search process transparent and reproducible and is also a valuable source for research. The iterative search process, however, is difficult to capture in the review documentation, which affects the transparency of the review process (105). It has therefore been suggested to include a search narrative in addition to the search strategy, which would explain *why* a search was developed and performed the way it was, not only *how* (106, 107). A search narrative would aid the peer review of a search strategy, since it would contextualise and increase the transparency of any major decisions that shaped the development of the strategy (107).

2.2 Information retrieval evaluation

Search strategies and information retrieval systems are evaluated to measure performance. Traditionally in information retrieval, there has been a strong focus on measuring effectiveness, the ability of an information retrieval system or search strategy to discriminate between documents that are relevant or not relevant for a given user query (108) or more specifically, the ability to retrieve relevant documents while at the same time suppressing the retrieval of non-relevant documents (109). It is also important, however, to measure efficiency to accomplish the retrieval process with the least time and effort possible, to save resources.

2.2.1 Relevance

Relevance is a concept that is widely discussed and is of central concern to the evaluation process (110). An essential component in traditional evaluation is a test collection, a set of well-defined requests, and a set of binary relevance assessments identifying the documents that are topically relevant to each request (111). The requests in systematic reviews are usually well defined with statements of inclusion and exclusion criteria resulting in a finite or representative set of relevant studies to include in a review. The aim of systematic reviews is total recall, and the included studies in already produced systematic reviews can act as gold standard for relevance in an evaluation process (112). To test a developed search strategy against a defined set of relevant references from other systematic reviews on the same topic (a quasi-gold standard) was proposed by Sampson and McGowan (95) as a method of validation for the effectiveness of MEDLINE searches used in systematic reviews and called relative recall. Relative recall is the proportion that any specific system retrieves of the total or "pooled" relevant documents retrieved by all systems considered to be working as a composite (113). The thorough, comprehensive search methods used in systematic reviews make the included studies of published reviews a suitable gold standard. This is especially true for Cochrane reviews as they are known to be developed according to strict rules and have higher quality than other reviews (114).

2.2.2 Database coverage

In this study, database coverage is calculated as the total number of included references indexed in one database (in this study, MEDLINE and/or Embase) divided by the total number of included references retrieved by all sources.

$$\text{Coverage} = \frac{\text{Number of included studies indexed in one database}}{\text{Total number of included studies in a review}}$$

2.2.3 Search performance measures

A quantitative metric must systematically associate to the results produced by a search system or search strategy and should measure relevance of the results to the users (115). Both information retrieval algorithms and Boolean search strategies are evaluated for their ability to find relevant documents. The metrics described below are all useful to evaluate both search systems and search strategies.

Precision and recall

The basic and most popular performance measures used in evaluating the relevance of search results are recall and precision (116). Recall is the fraction of relevant documents that has been retrieved and is used to evaluate the ability of a search to identify (all) relevant studies (108). In this study, relevant documents are the included studies of the sample reviews.

$$\text{Recall} = \frac{\text{Number of included studies retrieved}}{\text{Total number of included studies in a review}}$$

Recall was calculated by taking the average and median of the recall scores for the individual systematic reviews.

Precision is the fraction of retrieved documents that is relevant and is calculated as the number of relevant retrieved records, that is, included studies, divided by the total number of retrieved documents.

$$\text{Precision} = \frac{\text{Number of included studies retrieved}}{\text{Number of documents retrieved}}$$

Precision was calculated by taking the average and median of the precision scores for the individual systematic reviews.

There are few published norms for recall and precision of systematic review searches against which the performance of test searches can be compared, although most guidelines state that searches should aim at high recall which often results in relatively low precision (72, 117). For a long time no published norms for what constitutes “relatively low precision” existed. To meet this need, Sampson et al. performed a study (117) with the objective to establish typical values for the precision of systematic review searches, and across 94 systematic reviews, precision was found to be 3.0 % (median=2.9 %). Repeated experiments have shown that there is an inverse relationship between precision and recall, and within any particular search system or search strategy, it is difficult to improve both precision and recall simultaneously (109).

Mean Average Precision (MAP)

MAP is used to generate a single value summary of a ranked result set by averaging the precision figures obtained after each new relevant document is observed (118). MAP is then calculated by taking the mean of the average precision values across all topics in the run (119).

$$\text{MAP} = \frac{1}{\text{N}} \sum_{\text{i}=1}^{\text{N}} \text{AP}_{\text{i}}$$

MAP has some problems, though, as the ranking will change with changes in the database. It will, however, provide an indication of the performance of a search (120).

Number needed to read

Number needed to read (NNR) (also called Number Needed to Screen or Number Needed to Retrieve) is an analogy to Number Needed to Treat (NNT) in diagnostic testing and is used to describe the number of irrelevant references that one has to screen in order to find one of relevance (121). NNR is the inverse of precision and represents screening efficiency or screening burden for the user (121).

$$\mathrm{NNR} = \frac{1}{\mathrm{Precision}}$$

2.3 Summary

EBP and systematic reviews are tools to increase quality in decision making in healthcare, and information retrieval is an important part of the systematic review process. To secure high quality, any systematic review should perform a systematic search to identify the best available research evidence and thereby minimize bias that might arise as a result of missed relevant literature. The conventional method to find relevant research studies is to perform a Boolean search, but more automated methods like text mining and ranked retrieval have been suggested to speed up the review process and increase objectivity. In an evaluation process for systematic reviews, the included studies in already produced systematic reviews can act as gold standard for relevance. Different measures are used when assessing information retrieval performance, with recall and precision as the most important for conventional Boolean search strategies. MAP is used to evaluate performance of ranked retrieval and NNR to evaluate efficiency.

3 Scoping reviews

3.1 Introduction

Three simpler, more optimized search approaches for systematic reviews were investigated in this study. The first approach investigated whether searching MEDLINE and Embase is enough for study identification, the second, the selection of optimized search terms and development of simpler search strategies, and the third, whether a more automated search method can support information retrieval in systematic reviews. It is unclear, however, what kind of information is available on these topics, and scoping reviews were therefore performed on the three approaches.

3.1.1 Comments

The aim of the scoping reviews was to get an overview of existing literature on the main search approaches investigated in this thesis and to inform the empirical studies. The last two scoping review topics can be difficult to separate and some studies discuss both, so they could have been considered together. There was, however, reasons to treat them separately as some of the approaches dealt with both automatic and manual approaches, and the main point of Scoping review III was to review only automatic approaches with the consequence that some studies were reviewed twice.

Scoping review I, which investigated whether searching only MEDLINE and Embase is enough for study identification in systematic reviews, found 28 relevant, quite similar, studies, which were summarized and compared. Scoping reviews II and III found 13 relevant studies each with various content and research design. Each of these studies were therefore described in more detail than the studies included in Scoping review I.

3.2 Scoping review methodology

Scoping reviews can be used as preliminary assessments of potential size and scope of available research literature. A scoping review is a form of knowledge synthesis that addresses an exploratory research question aimed at mapping key concepts, types of evidence, and gaps in research related to a defined area or field by systematically searching, selecting, and synthesizing existing knowledge (122). Scoping review methodologies are appropriate for reviews of broad topics and can map the distribution and characteristics of a particular topic or issue, summarize the state of knowledge, and identify research gaps (1). Scoping reviews will allow the clarification of working definitions and conceptual boundaries of the topic as working on the review. They can also be used to explore the extent of the literature in a particular domain without describing findings in detail, and for scoping reviews, completeness of searching is determined by time constraints (122-124). These are reasons why scoping review methodology was suitable for the main approaches of this study.

3.3 Scoping review I: Searching only MEDLINE and Embase for study identification

3.3.1 Introduction

The most important sources for identifying studies for systematic reviews are bibliographic databases as they give the highest yield in most cases (125). Searching many different databases is time and resource consuming, though, and requires special competence of the different interfaces since interoperability is rare. Every new database searched will increase the number of references to screen with only a small number ultimately being included in the review. This is an inefficient use of valuable resources, and a reduction in the number of databases searched will reduce the workload and simplify the search process in systematic reviewing.

The databases considered the most important for research studies in healthcare are MEDLINE, Embase and CENTRAL. The Cochrane Handbook states that searching these three databases should be mandatory, Embase only if it is available (126). It turns out, however, that reviewers often search more databases than this recommendation, and the mean number of databases searched in Cochrane reviews has increased over the years (127).

Research on how many and which databases to search for systematic reviews is important to increase efficiency and make expert searchers more confident in their database choices. Studies have been performed earlier on this topic, many of them with a focus on whether searching MEDLINE and Embase is enough. It is unclear, however, how many studies have been performed, what characterizes these studies and what their conclusions are. For these reasons, a scoping review was conducted in order to get an overview of the research done in this area and to inform the empirical study.

3.3.1.1 Aim and objectives

The main aim of this scoping review was to determine the extent, range and nature of existing literature on the performance of searching only MEDLINE and Embase for study identification in systematic reviews by gathering and presenting available research evidence on the topic.

3.3.1.2 Research question

The following research question was formulated: What are the characteristics and conclusions of previous research on searching MEDLINE alone or MEDLINE and Embase combined when conducting systematic reviews?

3.3.2 Methods

A scoping review of research on restricting a systematic search to MEDLINE only or MEDLINE and Embase combined was conducted. Literature was searched for and screened to identify, characterize, and summarize relevant studies to give a picture of existing evidence on the topic.

3.3.2.1 Information management

All records were uploaded to the reference manager program EndNote and further to Covidence for screening. EndNote is a commercial reference management software package, used to manage bibliographies and references. Covidence is a non-profit service working in partnership with Cochrane Collaboration to improve the production and use of systematic reviews for health and wellbeing and is a tool that makes the systematic reviewing faster and easier (128).

3.3.2.2 Search methods

Due to time constraints the only bibliographic database searched was MEDLINE, in the segment “Ovid MEDLINE(R) Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R) 1946 to Present”, and reference lists of relevant studies were read.

The search strategy consisted of three clusters of terms: one relating to systematic reviews, one to information retrieval, and the third to the databases Embase, MEDLINE and PubMed. The MEDLINE search strategy can be found in Appendix IV and a PRISMA flow diagram of the process in Appendix VII.

3.3.2.3 Eligibility criteria

Eligibility criteria used to select publications for review and extraction are listed below:

Inclusion criteria:

1. Must evaluate searching only MEDLINE and/or MEDLINE and Embase combined.
2. Must investigate study identification in systematic reviews.
3. Must be published after 2004.
4. Included languages were English and the Scandinavian languages.
5. All kinds of studies were included regardless of study design, quality and topic.

Exclusion criteria:

1. Development of search filters

3.3.2.4 Study selection process

Retrieved records were screened by two authors (HS and MM) in a two-stage screening process. First, all records were assessed against inclusion and exclusion criteria based on their titles and abstracts, and second, relevant records were assessed against the full text. Studies were selected for inclusion by both authors, independent of each other. The authors then met and reviewed every selection. Discrepancies were dissolved by consensus.

3.3.2.5 Data extraction

A data charting form was developed by one author (MM) to determine which variables to extract. The form was continuously updated during the extraction process. A successful systematic search is dependent both on the coverage of databases and whether a search strategy retrieves the studies that are indexed in a database. The iterative method used for charting the data of the included studies in this review revealed that some of the studies investigating whether searching MEDLINE and Embase is enough for systematic reviews had evaluated database coverage, some search performance, and some both. The studies included in this scoping review were grouped according to these two broad categories.

All studies were extracted and recorded for information on the following issues:

1. Bibliographic details
2. Study design
3. Topics
4. Aims/objectives
5. Other characteristics of studies, more specifically: sample size, number of included studies, and results

Effectiveness of the search results was assessed by recall and precision and efficiency by NNR.

3.3.3 Results

The results of the search and extraction process are presented in the following paragraphs.

3.3.3.1 Study selection

A search for studies on limiting the search to MEDLINE and Embase for research studies to include in systematic reviews was performed in Ovid MEDLINE on the 22nd of October, 2018. The result of the literature search was 482 unique publications and 6 from reference lists. After reading titles and abstracts, 52 were read in full text for further assessment. 28 of these were found relevant according to inclusion and exclusion criteria and included in the review. The flow diagram in Appendix VII shows the result of the search and the selection process.

3.3.3.2 Characteristics of included studies

All study characteristics charted are described in Table 2 below with a presentation of study design, topics, aims/objectives, sample size, number of included studies of the reviews, database coverage and/or search performance. The included studies showed that when investigating whether it is enough to search only MEDLINE and Embase, 22 studies assessed database coverage, 14 search performance, and 8 both.

Table 2 Characteristics of included studies - Scoping review I

	Author	Year	Study design	Topics	Aims/objectives	SR-sample(s)	# of included studies in SR-sample(s)	Recall – MED-LINE	Precision – MED-LINE	NNR – MED-LINE	Coverage MED-LINE	Coverage MED-LINE / Embase combined
1	Bayliss	2006	Case study	Social care - parent-training programs for the treatment of conduct disorders in children	Evaluate the success of search strategies in retrieving key documents for a technology assessment report, including the most appropriate sources.	1	32	44 %				
2	Betran	2005	Case study	Maternal morbidity and mortality (prevalence and incidence)	Evaluate the usefulness of different sources in identifying data for one systematic review.	1	2580	62 %	4.1 %			
3	Beyer	2013	Case study	Frozen shoulder management	Investigate the performance of bibliographic databases in identifying the included studies, the smallest combination of databases required to retrieve all included studies, and the performance of the searches themselves.	1	31	84 %	0.8 %	123	87 %	
4	Bramer	2013	Cross sectional study	Any medical topic, from therapeutic effectiveness and diagnostic accuracy to ethics and public health.	Discover whether the original authors would have found all included references by using Google Scholar only. (The results were compared to PubMed.)	21	541	68 %			91 %	
5	Bramer	2016	Cross sectional study	Any medical topic, from therapeutic effectiveness and diagnostic accuracy to ethics and public health.	Compare the coverage of MEDLINE, Embase and Google Scholar and their performance in terms of precision and recall for included references in systematic reviews.	120	4795	73 %	2.8 %		92 %	98 %
6	Bramer	2017	Cross sectional study	Any medical topic, from therapeutic effectiveness and diagnostic accuracy to ethics and public health.	Determine the optimal combination of databases needed to conduct efficient searches in systematic reviews and whether the current practice in published reviews is appropriate.	58	1746	79 %	2.4 %	41		

	Author	Year	Study design	Topics	Aims/objectives	SR-sample(s)	# of included studies in SR-sample(s)	Recall – MED-LINE	Precision – MED-LINE	NNR – MED-LINE	Coverage MED-LINE	Coverage MED-LINE / Embase combined
7	Day	2005	Cross sectional study	Pharmaceuticals and physical modalities (Cochrane reviews)	Assess the efficacy of simplified search strategies and identify the best electronic bibliographic database for clinical trials in the field of musculoskeletal disorders and pain.	10	122				90 %	94 %
8	Gargon	2015	Case study	Core outcome set development (COS)	Compare the contribution of databases towards identifying included studies and identify the best combination of methods to retrieve all included studies.	1	250	87 %	1.1 %	89	97 %	
9	Golder	2012	Case study	Fractures and bone mineral density (adverse effects data)	Determine the relative value and contribution of searching different sources to identify adverse effects data.	1	58	33 %	7.6 %	13		
10	Goossen	2017	Cross sectional study	Surgery	Determine which electronic databases contribute best to a literature search in surgical systematic reviews on randomized (RCT) and non-randomized studies (NRS).	10	147 (RCT) 380 (NRS)	88 % (RCT) 93 % (NRS)	1.9 % (RCT) 5.2 % (NRS)	53 (RCT) 19 (NRS)		
11	Halladay	2015	Cross sectional study	Therapeutic interventions (Cochrane reviews)	Investigate the impact of using sources beyond PubMed in systematic reviews of therapeutic interventions.	50	2700				84 %	
12	Hanneke	2017	Cross sectional study	Obesity prevention policy	Examine three questions pertaining to systematic reviews on obesity prevention policy in order to identify the most efficient search methods.	21	577				86 %	
13	Hartling	2016	Cross sectional study	Acute Respiratory Infections (ARI), Infectious Diseases (ID), Developmental Psychosocial and Learning Problems (DPLP) (Cochrane reviews)	Examine the potential impact of selective database searching on results of meta-analyses.	129	1892				84 %	

	Author	Year	Study design	Topics	Aims/objectives	SR-sample(s)	# of included studies in SR-sample(s)	Recall – MED-LINE	Precision – MED-LINE	NNR – MED-LINE	Coverage MED-LINE	Coverage MED-LINE / Embase combined
14	Kelley	2012	Case study	Exercise for Arthritis	Determine the database indexing of RCTs for a meta-analysis addressing the effects of exercise on pain and physical function in adults with arthritis and other rheumatic diseases.	1	36	75 %		9	94 %	
15	Kwon	2014	Case study	Ward closures as an infection control intervention	Determine the value and efficacy of searching biomedical databases beyond MEDLINE for systematic reviews.	1	97	86 %			96 %	
16	Lemeshow	2005	Case study	Alcohol consumption and risk of breast cancer and large bowel cancer	Address methodologic issues in searching for observational studies by presenting database search methods and results.	1	79	74 %				
17	Lorenzetti	2014	Cross sectional study	Clinical, economic and social areas	Explore the degree to which databases other than MEDLINE contribute studies relevant for inclusion in rapid HTAs.	25	2352				88 %	90 %
18	Michaleff	2011	Cross sectional study	Physical therapy interventions	Compare the completeness of indexing of reports of RCTs by eight bibliographic databases.	39	400				89 %	96 %
19	Moseley	2009	Cross sectional study	Physical therapy interventions (Cochrane reviews)	Compare the comprehensiveness of indexing the reports of RCTs by eight bibliographic databases.	30	281				91 %	
20	Pilkington	2007	Cross sectional study	Complementary therapies (CAM)	Explore the effectiveness of search strategies developed to identify trials of specific complementary therapies in a range of clinical conditions.	35	127				51 %	
21	Preston	2015	Exploratory study	Diagnostic test accuracy	Assess the viability of an approach restricting searches to MEDLINE, Embase and the reference lists of included studies.	9	302				91 %	95 %

	Author	Year	Study design	Topics	Aims/objectives	SR-sample(s)	# of included studies in SR-sample(s)	Recall – MED-LINE	Precision – MED-LINE	NNR – MED-LINE	Coverage MED-LINE	Coverage MED-LINE / Embase combined
22	Rice	2016	Cross sectional study	Diagnostic test accuracy	Evaluate meta-analyses on the diagnostic accuracy of depression screening tools.	16	398	92 %			94 %	
23	Rollin	2009	Cross sectional study	Occupational health (Cochrane reviews)	Discover whether limiting searches to MEDLINE would miss studies of high quality.	42	536				89 %	
24	Royle	2005	Cross sectional study	Diabetes	Analyse the effect on systematic reviews of including only trials that are indexed in MEDLINE, and to assess the impact of adding trials from other databases and the grey literature.	44	695				83 %	
25	Shariff	2012	Cross sectional study	Nephrology	Compare the availability of renal clinical studies in six major bibliographic databases.	151	2195				96 %	
26	Slobogean	2009	Cross sectional study	Orthopaedics	Determine the percentage of articles cited in meta-analyses that can be found in MEDLINE and Embase alone.	39	647				90 %	91 %
27	Sood	2005	Cross sectional study	Acupuncture (Cochrane reviews)	Assess the source of original literature contributing to Cochrane reviews on acupuncture.	10	108				69 %	
28	Whiting	2008	Cross sectional study	Diagnostic test accuracy	Estimate the yield from searching a range of bibliographic databases and additional sources to identify test accuracy studies for systematic reviews.	7	522	79 %		169	86 %	

3.3.3.3 References to studies included in this review

1.	Bayliss SE, Dretzke J. Health technology assessment in social care: a case study of randomized controlled trial retrieval. Int J Technol Assess Health Care. 2006;22(1):39-46.
2.	Betran AP, Say L, Gulmezoglu AM, Allen T, Hampson L. Effectiveness of different databases in identifying studies for systematic reviews: experience from the WHO systematic review of maternal morbidity and mortality. BMC Med Res Methodol. 2005;5(1):6.
3.	Beyer FR, Wright K. Can we prioritise which databases to search? A case study using a systematic review of frozen shoulder management. Health Info Libr J. 2013;30(1):49-58.
4.	Bramer WM, Giustini D, Kramer BM. Comparing the coverage, recall, and precision of searches for 120 systematic reviews in Embase, MEDLINE, and Google Scholar: a prospective study. Systems Review. 2016;5:39.
5.	Bramer WM, Giustini D, Kramer BM, Anderson P. The comparative recall of Google Scholar versus PubMed in identical searches for biomedical systematic reviews: a review of searches used in systematic reviews. Syst Rev. 2013;2:115.
6.	Bramer WM, Rethlefsen ML, Kleijnen J, Franco OH. Optimal database combinations for literature searches in systematic reviews: a prospective exploratory study. Syst Rev. 2017;6(1):245.
7.	Day D, Furlan A, Irvin E, Bombardier C. Simplified search strategies were effective in identifying clinical trials of pharmaceuticals and physical modalities. J Clin Epidemiol. 2005;58(9):874-81.
8.	Gargon E, Williamson PR, Clarke M. Collating the knowledge base for core outcome set development: developing and appraising the search strategy for a systematic review. BMC Med Res Methodol. 2015;15:26.
9.	Golder S, Loke YK. The contribution of different information sources for adverse effects data. Int J Technol Assess Health Care. 2012;28(2):133-7.
10.	Goossen K, Tenckhoff S, Probst P, Grummich K, Mihaljevic AL, Buchler MW, et al. Optimal literature search for systematic reviews in surgery. Langenbecks Arch Surg. 2018;403(1):119-29.
11.	Halladay CW, Trikalinos TA, Schmid IT, Schmid CH, Dahabreh IJ. Using data sources beyond PubMed has a modest impact on the results of systematic reviews of therapeutic interventions. J Clin Epidemiol. 2015;68(9):1076-84.
12.	Hanneke R, Young SK. Information sources for obesity prevention policy research: a review of systematic reviews. Syst Rev. 2017;6(1):156.
13.	Hartling L, Featherstone R, Nuspl M, Shave K, Dryden DM, Vandermeer B. The contribution of databases to the results of systematic reviews: a cross-sectional study. BMC Med Res Methodol. 2016;16(1):127.
14.	Kelley GA, Kelley KS. The Association between Databases for Indexing Studies Intended for an Exercise Meta-Analysis of Arthritis Randomized Controlled Trials. Arthritis. 2012;2012:624830.

15.	Kwon Y, Powelson SE, Wong H, Ghali WA, Conly JM. An assessment of the efficacy of searching in biomedical databases beyond MEDLINE in identifying studies for a systematic review on ward closures as an infection control intervention to control outbreaks. Syst Rev. 2014;3:135.
16.	Lemeshow AR, Blum RE, Berlin JA, Stoto MA, Colditz GA. Searching one or two databases was insufficient for meta-analysis of observational studies. J Clin Epidemiol. 2005;58(9):867-73.
17.	Lorenzetti DL, Topfer LA, Dennett L, Clement F. Value of databases other than medline for rapid health technology assessments. Int J Technol Assess Health Care. 2014;30(2):173-8.
18.	Michaleff ZA, Costa LO, Moseley AM, Maher CG, Elkins MR, Herbert RD, et al. CENTRAL, PEDro, PubMed, and EMBASE are the most comprehensive databases indexing randomized controlled trials of physical therapy interventions. Phys Ther. 2011;91(2):190-7.
19.	Moseley AM, Sherrington C, Elkins MR, Herbert RD, Maher CG. Indexing of randomised controlled trials of physiotherapy interventions: a comparison of AMED, CENTRAL, CINAHL, EMBASE, hooked on evidence, PEDro, PsycINFO and PubMed. Physiotherapy. 2009;95(3):151-6.
20.	Pilkington K. Searching for CAM evidence: an evaluation of therapy-specific search strategies. J Altern Complement Med. 2007;13(4):451-9.
21.	Preston L, Carroll C, Gardois P, Paisley S, Kaltenthaler E. Improving search efficiency for systematic reviews of diagnostic test accuracy: an exploratory study to assess the viability of limiting to MEDLINE, EMBASE and reference checking. Syst Rev. 2015;4(1):82.
22.	Rice DB, Kloda LA, Levis B, Qi B, Kingsland E, Thombs BD. Are MEDLINE searches sufficient for systematic reviews and meta-analyses of the diagnostic accuracy of depression screening tools? A review of meta-analyses. J Psychosom Res. 2016;87:7-13.
23.	Rollin L, Darmoni S, Caillard JF, Gehanno JF. Searching for high-quality articles about intervention studies in occupational health--what is really missed when using only the Medline database? Scand J Work Environ Health. 2010;36(6):484-7.
24.	Royle P, Waugh N. A simplified search strategy for identifying randomised controlled trials for systematic reviews of health care interventions: a comparison with more exhaustive strategies. BMC Med Res Methodol. 2005;5:23.
25.	Shariff SZ, Sontrop JM, Iansavichus AV, Haynes RB, Weir MA, Gandhi S, et al. Availability of renal literature in six bibliographic databases. NDT Plus. 2012;5(6):610-7.
26.	Slobogean GP, Verma A, Giustini D, Slobogean BL, Mulpuri K. MEDLINE, EMBASE, and Cochrane index most primary studies but not abstracts included in orthopedic meta-analyses. J Clin Epidemiol. 2009;62(12):1261-7.
27.	Sood A, Sood R, Bauer BA, Ebbert JO. Cochrane systematic reviews in acupuncture: methodological diversity in database searching. J Altern Complement Med. 2005;11(4):719-22.
28.	Whiting P, Westwood M, Burke M, Sterne J, Glanville J. Systematic reviews of test accuracy should search a range of databases to identify primary studies. J Clin Epidemiol. 2008;61(4):357-64.

3.3.3.4 Synthesis of results

This scoping review found that 28 earlier studies had evaluated whether it was enough to search only MEDLINE or MEDLINE and Embase combined for study identification in systematic reviews.

Aims and objectives of included studies

The main aim of all included studies was to investigate the effect of reducing the number of sources to search in systematic reviewing or finding the optimal number of databases for a research question (including restricting the search to only MEDLINE and Embase). Exactly what they evaluated differed, however, and is presented below:

- Usefulness of different sources of information for one or more systematic reviews
- Optimal combination of databases needed to retrieve a gold standard
- Impact of searching biomedical databases beyond MEDLINE
- Coverage of MEDLINE alone or MEDLINE and Embase combined
- Coverage of MEDLINE alone or MEDLINE and Embase combined + reference lists
- Search strategy performance
- Comparison of coverage and/or search performance of MEDLINE to other databases

Study design

The study designs of the 28 included studies were 1 exploratory study, 8 case studies (that is, studies describing a single systematic review) and 19 cross sectional studies (that is, studies describing a sample of systematic reviews).

Topics

The included studies investigated database coverage and search performance on a wide variety of topics. Five covered various topics in one study, collectively termed for example “any medical topic” or “therapeutic interventions”. Four assessed “diagnostic test accuracy” and two “physical therapy interventions”. The rest of the studies evaluated only one specific topic, for example, social care, surgery, acupuncture, CAM, or obesity prevention policy.

For more information on topics refer to Table 2.

Sample size / number of included studies

The sample sizes of each study and the number of included studies in each review varied considerably among the studies included in this review. The sample size of the cross sectional studies varied from 7 to 151 systematic reviews with an average of 40 (median=28). The total number of included studies in each cross sectional study varied from 108 to 4795 with an average of 998 (median=529). For all studies together, both case studies and cross sectional, the average number of included studies per review was 126. For the cross sectional studies, average number was 28 (median=21) and for the case studies 403 (median=78). The high average number of included studies in the case studies was mostly due to one review including 2580 primary studies, a case study by Betran et al. (64), on the very broad topic “Maternal morbidity and mortality (prevalence and incidence)”. All case studies were on quite broad topics and had high numbers of included studies, for example the study by Betran et al. mentioned above and a study on “core outcome set development” with 250 included studies. The lowest number of included studies in a case study was 31 on “frozen shoulder management”. This indicates that most case studies included in this scoping review were of a special character and quite unlike the average systematic review.

Database coverage

Independent of study design, 22 of the 28 included studies of this review evaluated coverage of MEDLINE alone and six coverage of MEDLINE and Embase combined. Coverage of MEDLINE alone varied between 69 % and 97 % with an average of 87 % (median=90 %). Coverage of MEDLINE and Embase combined varied between 90 % and 98 % with an average of 94 % (median=95 %). Based on the 17 cross sectional studies, MEDLINE alone had a coverage of 85 % (median=89 %). The four case studies evaluating coverage of MEDLINE alone had an average of 94 % (median=95 %). Coverage for MEDLINE and Embase combined was not evaluated in any of the case studies.

Search performance

There was a great variation in search performance of the MEDLINE search strategies evaluated in the reviewed studies. Independent of study design, 15 studies evaluated recall with an average of 78 % (median=79 %), ranging between 44 % and 93 %. Precision for MEDLINE searches was evaluated in six studies, ranging between 0.8 % and 5.2 % with an average of 2.9 % (median=2.6 %).

All case studies (=8) evaluated recall with an average of 68 % (median=75 %) ranging from 33 % to 87 %. Of these, three evaluated precision with 4.1 %, 1.1 % and 0.8 %, average=2.0 %. Eight of the cross sectional studies evaluated recall with an average of 81 % (median=79 %) ranging from 68 % to 93 %. Precision was evaluated in four of the cross sectional studies with an average of 3.1 % (median=2.6 %).

Efficiency was assessed by NNR and was evaluated in five of the included studies with an average of 59 (median=53) ranging from 9 to 123.

Conclusions of included studies

Seven of the eight case studies concluded that it was not enough to search only MEDLINE or MEDLINE and Embase combined but based on different assumptions. Four of them based this conclusion on search performance (64, 90, 129, 130) and three on both search performance and database coverage (85, 131, 132). The one case study that concluded it was enough to search only MEDLINE and Embase evaluated both coverage and search performance (94).

Of the 19 cross sectional studies, the 2 studies investigating only search performance (133, 134) and the 3 investigating both coverage and search performance (65, 84, 135) all concluded that it was not enough to search only MEDLINE or MEDLINE and Embase combined. Of the 14 cross sectional studies evaluating only coverage, 7 concluded that it was enough (56, 63, 136-140) and 7 that it was not enough to search only MEDLINE or MEDLINE and Embase combined (125, 141-146). The exploratory study concluded also that it was not enough to search only those two databases based on both coverage and search performance (147).

3.3.4 Discussion

This scoping review identified 28 studies published between 2005 and 2017 addressing the effect of searching only MEDLINE or MEDLINE and Embase combined for systematic reviews. Although MEDLINE and Embase are considered capable of identifying the majority of relevant research studies in health and medicine, searching only those two databases for the purpose of conducting a comprehensive literature search can be insufficient, and most of the studies included in this scoping review agreed that it was not enough to search only MEDLINE and Embase. What is important, however, is to consider the basis for the conclusions. The included studies were of two broad groups, those evaluating database coverage and those evaluating search performance, which could have influenced the conclusions. For some studies, coverage values were high, but the search strategies of the same reviews could have low performance, and vice versa, search strategy performance could be high but with very low coverage, and relevant studies could have been missed. Several other variables could also have influenced the results, for example, the age of the studies, study design, topics, and language but also the attitudes of the authors to how much searching is enough. These variables are discussed in the following sections.

3.3.4.1 Coverage and search performance

Coverage

Coverage values differed to a great extent between case studies and cross sectional studies in this scoping review. The four case studies evaluating coverage of MEDLINE alone, had an average value of 94 % (median=95 %), which probably is a bit higher than can be expected for the standard systematic review. Based on 19 cross sectional studies, average coverage in MEDLINE of included studies was 86 % (median=89 %), and MEDLINE and Embase combined had an average coverage of 94 % (median=95 %) based on 8 studies. These values will probably be more comparable to the results of the empirical part of this study than the results of the case studies and could be good estimates for typical expected coverage values as they were based on bigger samples and a greater variety of topics.

Search performance

Conclusions on database usefulness based on search performance alone are not very reliable. As is well known, a search strategy can fail to retrieve relevant studies that are indexed in a database as search strategies vary highly in effectiveness (26, 148). Nevertheless, 15 studies based their decision on whether it would have been enough to search only MEDLINE and Embase for study identification in systematic reviews on search performance. The average recall of these studies was 78 % (median=79 %), and ranged from 44 % to 93 %, which shows a great variation in results. Precision for MEDLINE searches was evaluated in six studies and ranged from 0.8 % to 5.2 % with an average of 2.9 % (median=2.6 %). As can be seen, the performance of the included studies varied widely and is probably not a reliable basis for predicting future search performance. Yet, the average precision value is consistent with the average of Sampson et al.'s cross sectional study to set a precision value for systematic reviews (117).

3.3.4.2 Age of studies

Advances over the recent years, including the increased scope of MEDLINE and Embase, will probably improve coverage and search performance, and there was a small increase in both in the included studies published after 2010. Seventeen studies evaluated coverage of MEDLINE, and leaving out studies published before 2010 changed coverage from 86 % (median=89 %) to 90 % (median=90 %). A total of 15 studies evaluated recall of MEDLINE search strategies with an average of 78 % (median=79 %). Leaving out studies published before 2010 increased recall to 82 % (median=84 %). These increases could be due to mere chance or it could be that changes in indexing practices, and/or an increase in search quality affected the results. Guidelines on how to perform systematic searches are constantly updated and can lead to better quality of the search strategies. Thus, newer search strategies will probably find more relevant studies that are indexed in a database. There is also a continuous increase in the number of journals indexed in MEDLINE and Embase, and the indexing standards evolve, which can also improve search performance. It is therefore plausible that newer results are more pertinent to current systematic review practice (56).

3.3.4.3 Study designs and sample size

Study design and sample size are important factors concerning validity and reliability of research studies. The majority of studies included in this review were cross sectional, but eight were case studies. Seven of the 15 studies concluding that searching MEDLINE and Embase was not enough, were case studies, and accordingly, the conclusions are not as trustworthy as results based on a bigger sample. Five of the cross sectional studies had a sample of 50 reviews or more. Three of these had more than 100 reviews and are therefore probably more generalizable than the case studies.

3.3.4.4 Topics and language

Most topics covered by the included studies of this review, especially the case studies, were broad and some quite complex. Many of them varied to such an extent that it is difficult to compare the results or draw any reliable conclusions. Literature on broad topics, especially interdisciplinary, may not be described using uniform terminology or concentrated in one source (141). The study by Betran et al. (64) on “Maternal morbidity and mortality (prevalence and incidence)”, for example, including 2580 studies, could not be expected to have all included studies indexed in a few sources. Such a high number of studies would be impossible to retrieve from only one or two databases and is not comparable to the average systematic review. Some studies evaluating both coverage and search performance experienced that even the most sensitive search strategies did not retrieve all the relevant studies that were indexed in a database when working with a subject lacking standardized terms (129, 132).

3.3.4.5 Number of studies needed to conclude

An important question concerning database coverage of primary studies and search performance in systematic reviewing is whether it is necessary to retrieve all relevant studies on a topic. In one study, Halladay et al. (56) found that when additional relevant studies were identified from non-MEDLINE sources, they tended to not contribute substantial amount of information to meta-analyses, and their omission did not seem to bias the meta-analyses results. The authors concluded that decisions to search multiple databases beyond MEDLINE will generally depend on the particular goals of the systematic review, the context in which the review is conducted, and the available resources. They suggested that “When reviews are prepared under resource constraints and the expected number of relevant studies is not too small (e.g., when 10 or more studies are expected to be included in meta-analysis), systematic searching limited to PubMed can provide reliable inputs for subsequent decision and economic analyses.” In another study, Hartling et al. (63) conducted a cross sectional quantitative analysis to examine the potential impact of selective database searching on results of meta-analyses and found that meta-analyses based on the majority of studies did not differ in most cases. Results did not change in a systematic manner, suggesting that selective searching may not introduce bias in terms of effect estimates. The authors concluded that a majority of relevant studies can be found within a limited number of databases, but that the choice of databases is topic-specific. Both these studies claimed that leaving some studies out would not have affected the conclusions significantly, and limiting the search to MEDLINE alone or MEDLINE and Embase combined could be a viable solution for systematic reviews.

3.3.4.6 Authors’ attitudes

Conclusions on whether it is enough to search only MEDLINE and Embase were also dependent on the authors’ attitudes to what is enough. 100 % search recall or database coverage was not achieved by any of the included studies but was necessary for some authors to conclude that searching only MEDLINE and Embase is enough. Others claiming that searching MEDLINE and Embase is enough, however, argued that it is not necessary to retrieve all relevant studies on a topic, either because it is not possible to achieve full recall anyway and/or that the studies not retrieved by database searching will be found in other sources. On the other hand, for some reviews not retrieving all relevant studies it could have made a difference to the results, and even if some authors concluded that for their study it would have been enough to search only MEDLINE and Embase, they were reluctant to generalize their findings.

3.3.5 Conclusion

Based on the 28 included studies of this scoping review, there is no simple answer to whether searching only MEDLINE and Embase is enough for systematic reviews in general. The results indicate that to decide on the number of databases to search for systematic reviews is difficult and highly dependent on kind of topic, its broadness and complexity, and not least, the authors' attitudes to how much searching is enough. Limiting systematic reviews to searching only MEDLINE and Embase depends both on database coverage and search performance and thereby the ability of searchers to conduct highly sensitive searches for these databases on the topics of interest.

Most studies concluded that searching MEDLINE and Embase is not enough. Some of the studies with the largest samples, and the studies investigating the impact of losing a few studies, however, made the opposite conclusion. Thus, the findings suggest that for systematic reviews of therapeutic interventions, the gains from searching beyond MEDLINE and Embase may be more modest than commonly believed (56), and to save time and resources, it could be enough to search only MEDLINE and Embase for such reviews. For more specialized or complex topics, like social care, physiotherapy, and CAM, however, searching only MEDLINE and Embase will probably not be sufficient to find most relevant studies.

3.4 Scoping review II: Selection of search terms and development of simpler search strategies

3.4.1 Introduction

The selection of useful search terms and development of simpler search strategies are important considerations as they influence effectiveness and efficiency of a search. Simpler search approaches can contribute to save resources and reduce the time it takes to produce a systematic review. It is unclear, however, what kind of information is available on these topics. Therefore, a scoping review was conducted in order to map the research done in this area.

3.4.1.1 Aim and objectives

The aim of this scoping review was to determine the extent and nature of existing literature on how to select useful search terms and develop simpler search strategies for systematic reviews by gathering and presenting available research evidence on the topic.

3.4.1.2 Research question

The following research question was formulated: What previous research has been done regarding search term selection and the development and use of more optimized or simple search strategies when conducting systematic reviews, and what is its characteristics and conclusions?

3.4.2 Methods

A scoping review of research on how to select search terms and design simpler search strategies for information retrieval in systematic reviews was conducted. Literature on the two topics were searched for and screened to identify, characterize, and summarize existing evidence.

3.4.2.1 Search methods

Due to time constraints the only electronic database searched was MEDLINE, in the segment "Ovid MEDLINE(R) Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R) 1946 to Present", and reference lists of relevant studies were read.

The search strategy consisted of two clusters of terms: one relating to systematic reviews and the second to the development of simpler search strategies and selection of search terms. The MEDLINE search strategy used can be found in Appendix V, and a PRISMA flow diagram of the selection process in Appendix VIII.

3.4.2.2 Eligibility criteria

Eligibility criteria used to select publications for review and extraction are listed below:

Inclusion criteria:

1. Must be published after 2004.
2. Included languages were English and the Scandinavian languages.
3. All kinds of study designs were included.

Exclusion criteria:

1. Development of filters for study design.
2. Search methods primarily meant for updating existing systematic reviews.

3.4.2.3 Screening

Records were screened in a two-stage screening process. First, all retrieved records were assessed against inclusion and exclusion criteria based on their titles and abstracts, and second, relevant records were assessed against the full text. Records were selected for inclusion by only one author (MM).

3.4.2.4 Data extraction

A data charting form specifying which variables to extract, was developed. The form was continuously updated during the extraction process.

All studies were extracted and recorded for information on the following broad issues:

1. Bibliographic details
2. Study design
3. Topics
4. Characteristics of studies reviewed, more specifically: techniques/methods, tools/systems/databases, search performance and key findings/conclusions

3.4.3 Results

3.4.3.1 Study selection

A search for studies on methods for the selection of useful search terms and development of simpler search strategies for systematic reviews was performed in Ovid MEDLINE on the 30th of April, 2019. The result of the literature search was 403 unique publications. After reading titles and abstracts, 47 articles were read in full text for further assessment. Of these, 11 were found relevant according to inclusion and exclusion criteria. Two relevant studies were found in other sources. Thus, 13 studies were included in the review. The flow diagram in Appendix VIII shows the result of the search and the selection process.

3.4.3.2 Characteristics of included studies

Characteristics of included studies are presented in Table 3 below, including study design, topics, techniques/methods, tools/systems/databases used, and key findings/conclusions.

Twelve studies included in this scoping review investigated search term selection, nine automatic and three manual methods. One study investigated only the development and performance of a simpler search method, and three studies investigated both search term selection and development of simpler search strategies.

The 13 included studies were published between the years 2005 and 2018.

Table 3 Characteristics of included studies - Scoping review II

	Author	Year	Category	Topics	Study design	Aims / objectives	Techniques / methods	Tools / systems / databases used	Key findings / conclusions
1	Ananiadou	2009	Automatic search term selection	Social science	Descriptive	Describe the ASSERT project and how text mining has been used to enhance the production of systematic reviews.	Text mining: term extraction	Termine	Through use of semi-automated techniques reviews can be completed more quickly and systematically
2	Bradley	2011	Automatic search term selection	Borderline personality disorder	Experimental	Examine whether two free web data-mining tools provide reliable suggestions for searchers and to what extent they provide additional, useful text-words and MeSH terms.	Statistical analysis	GoPubMed, PubReMiner	GoPubMed and PubReMiner are two data-mining web tools that are a useful supplement for search planning but should be used with caution
3	Bramer	2018	Simpler search strategies	Various biomedical topics	Prospective, exploratory study	Examine how limiting searches to major thesaurus terms in Embase and MEDLINE as well as limiting to words in the title and abstract fields of those databases, affects the overall recall of systematic review searches.	The original search strategies used in Embase and MEDLINE were modified in four ways: (1) by searching Embase thesaurus terms as major descriptors, (2) by removing thesaurus terms from the Embase search such that only terms were searched in the title and/or abstract fields, (3) by searching both MEDLINE and Embase for major thesaurus terms, and (4) by searching both MEDLINE and Embase for terms in the title and/or abstract fields only.	Embase MEDLINE	Of the four search options, two options substantially reduced the overall search yield. This also resulted in a greater chance of losing relevant references and was therefore not recommended.
4	Day	2005	Simpler search strategies Manual search term selection	Pharmaceuticals and physical modalities	Experimental	Assess the efficacy of simplified search strategies.	Simple search strategies were developed manually that comprised of one line containing one term for the intervention, a second line containing one term for the	MEDLINE (Ovid and PubMed), Embase, Cinahl, CENTRAL	Simplified search strategies are an effective, efficient way to search for clinical trials. They work best when the intervention is a pharmaceutical or a well-

	Author	Year	Category	Topics	Study design	Aims / objectives	Techniques / methods	Tools / systems / databases used	Key findings / conclusions
							condition in question, a third line combining the first two, and a fourth line limiting the search to clinical trials. The super simple strategy consisted of one line for the intervention and a second line limiting the search to clinical trials.		defined physical treatment. Their sensitivity, however, is not adequate for conducting systematic reviews.
5	Hausner	2015	Automatic search term selection	Health and medicine	Experimental	Determine whether the objective approach for the development of search strategies was non-inferior to the conceptual approach commonly used in Cochrane reviews.	Text analysis	MEDLINE	The non-inferiority test showed that the use of text analysis was non-inferior to the conceptual approach used in conventional systematic reviews.
6	Hausner	2012	Automatic search term selection	Brachytherapy for treatment of prostate cancer	Descriptive / Case study	Use an empirically guided approach to the development of a search strategy to provide a way to increase transparency and efficiency.	Statistical analysis	MEDLINE	Transparency and efficiency were increased by developing an empirically (objective) guided approach in the development of search strategies.
7	Karimi	2010	Simpler search strategies Manual search term selection	Pharmaceuticals	Experimental	Explore the effectiveness of using ranked retrieval as compared to Boolean querying for the purpose of constructing a systematic review.	Formulated ranked queries by selecting search terms from the title, background, research question, and inclusion criteria of the reviews.	MEDLINE	In a ranked query, the more search terms used the better. None of the queries were suitable for systematic reviews, however.
8	O'Mara-Eves	2013	Automatic search term selection	Health and social care / Community engagement	Descriptive / Case study	Develop alternative systematic ways of identifying relevant evidence where the key concepts are generally not focal to the primary studies' aims and are	Used text mining to identify synonyms for hard-to-detect evidence.	Termine	Text mining helped to identify relevant search terms for a broad topic that was inconsistently referred to in the literature.

	Author	Year	Category	Topics	Study design	Aims / objectives	Techniques / methods	Tools / systems / databases used	Key findings / conclusions
						found across multiple disciplines.			
9	Simon	2010	Automatic search term selection	Nurse staffing	Experimental	Develop search strategies to identify primary publications on nurse staffing research in PubMed/MEDLINE.	Text mining	Tm package in R PubMed / MEDLINE	Empirically selected search terms can help to develop effective search strategies.
10	Stansfield	2017	Automatic search term selection	Social care and support of adults with intellectual disabilities as they get older	Review / Case study	I: Briefly review the literature on applications of text mining for search term development for systematic reviewing. II: Compare the use of individual text mining tools and techniques to increase sensitivity through identifying suitable search terms and to increase precision from examining preliminary outputs of a search for unwanted terms and concepts.	Text mining applications	Term frequency: BibExcel Endnote AntConc Voyant Tools Automatic term recognition: Termine Automatic clustering: EPPI-Reviewer (Lingo3G)	Text mining can aid the discovery of search terms for search strategies for diversely described topics to support an iterative search strategy development process. Using multiple tools appears to be particularly fruitful. Their usefulness is influenced by the varying functionality of the tools used, the way that they are used, and the text that is analysed.
11	Thomas	2011	Automatic search term selection	Not specified	Descriptive	Describe the application of four text mining technologies, namely, automatic term recognition, document clustering, classification and summarization, which support the identification of relevant studies in systematic reviews.	Text mining	TerMine	Text mining technologies have the potential to assist at various stages of the review process. They are relatively unknown in the systematic reviewing community, however, and substantial evaluation and methods development are required before their possible impact can be fully assessed.
12	Thompson	2013	Automatic search term selection	Transborder interventions for drug control	Descriptive / Case study	Present a replicable method for selecting terms in a systematic search using text analysis software.	Text analysis	Leximancer	The study demonstrated a method for selecting systematic search terms that was transparent, replicable, and generalizable across

	Author	Year	Category	Topics	Study design	Aims / objectives	Techniques / methods	Tools / systems / databases used	Key findings / conclusions
									disciplines and could be used to systematically generate search terms.
13	Waffenschmidt	2013	Manual search term selection Simpler search strategies	Pharmaceuticals	Retrospective analysis	Determine whether PubMed's Related Citations and/or a simple-structured Boolean search are efficient and reliable search techniques to assess the completeness of an evidence base consisting of published RCTs.	Simple-structured Boolean search strategies and a related articles function	PubMed's Related Citations	The findings indicated that the combination of the first 20 Related Citations and a simple-structured Boolean search in PubMed is an efficient and reliable method to assess the completeness of an evidence base consisting of published RCTs on new drugs. High reliability of the search is primarily achieved by the combination of the two different search techniques.

3.4.3.3 References to studies included in this review

1.	Ananiadou S, Rea B, Okazaki N, Procter R, Thomas J. Supporting Systematic Reviews Using Text Mining. Soc Sci Comput Rev. 2009;27(4):509-23.
2.	Bradley S, Giustini D, editors. GoPubMed versus PubReMiner for analyzing PubMed search results: a head to head comparison of two free web 'data mining' tools. CHLA/ABSC Conference; 2011.
3.	Bramer WM, Giustini D, Kleijnen J, Franco OH. Searching Embase and MEDLINE by using only major descriptors or title and abstract fields: a prospective exploratory study. Syst Rev. 2018;7(1):200.
4.	Day D, Furlan A, Irvin E, Bombardier C. Simplified search strategies were effective in identifying clinical trials of pharmaceuticals and physical modalities. J Clin Epidemiol. 2005;58(9):874-81.
5.	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. J Clin Epidemiol. 2015;68(2):191-9.
6.	Hausner E, Waffenschmidt S, Kaiser T, Simon M. Routine development of objectively derived search strategies. Systems Review. 2012;1:19.
7.	Karimi S, Pohl S, Scholer F, Cavedon L, Zobel J. Boolean versus ranked querying for biomedical systematic reviews. BMC Med Inf Decis Mak. 2010;10:58.
8.	O'Mara-Eves A, Brunton, Ginny, McDaid, David, Kavanagh, Josephine, Oliver, Sandy, Thomas, James. Techniques for identifying crossdisciplinary and 'hard-to-detect' evidence for systematic review. Res Synth Methods. 2013(5):50–9.
9.	Simon M, Hausner E, Klaus SF, Dunton NE. Identifying nurse staffing research in Medline: development and testing of empirically derived search strategies with the PubMed interface. BMC Med Res Methodol. 2010;10:76.
10.	Stansfield C, O'Mara-Eves A, Thomas J. Text mining for search term development in systematic reviewing: A discussion of some methods and challenges. Res Synth Methods. 2017;8(3):355-65.
11.	Thomas J, McNaught, John, Ananiadou, Sophia. Applications of text mining within systematic reviews. Res Synth Methods. 2011(2):1-14.
12.	Thompson J, Davis J, Mazerolle L. A systematic method for search term selection in systematic reviews. Res Synth Methods. 2013;5(2):87-97.
13.	Waffenschmidt S, Janzen T, Hausner E, Kaiser T. Simple search techniques in PubMed are potentially suitable for evaluating the completeness of systematic reviews. J Clin Epidemiol. 2013;66(6):660-5.

3.4.3.4 Synthesis of results

The 13 included studies showed that different methods or techniques, both manual and automatic, were used to select search terms and simplify the development of search strategies for systematic reviews. Refer to Table 3 for more information.

Study designs

All kinds of study designs were included in this review and resulted in two descriptive studies, three descriptive and case studies, one review and case study, one prospective exploratory study, five experiments, and one retrospective analysis.

Topics

The topics investigated were of different kinds. Some were specific or narrow, like "borderline personality disorder", "musculoskeletal disorders and pain" and "brachytherapy for treatment of prostate cancer". Others were of a broader character, like health and medicine in general.

Techniques/methods for search term selection

The techniques or methods used for search term selection were both manual and automatic or a combination of the two. The automatic methods used for the selection process were text mining and statistical analysis, for example the use of text analysis tools, like Termine and Leximancer.

Sources for search terms

The sources used for selecting search terms were mostly review title and occasionally, research question, background information, and inclusion criteria of a review. The included studies of other relevant reviews were also used as search term sources.

Simple search techniques

Four of the included studies investigated the development and use of simpler search strategies for systematic reviews. Some developed simpler searches based on fewer search terms, some simplified the searches by using major thesaurus terms or terms in title and/or abstract only and some used a combination of simple conventional searches and automated methods like the related articles function in PubMed.

Detailed description of studies

To better understand the contribution and content of the included studies, each study is described in more detail below.

To enhance the production of the whole systematic review process, Ananadiou et al. described the development of a text mining framework, the ASSERT project, in an article from 2009 (7). This project and the use of different text mining techniques were further described by Thomas et al. in 2011 (2). To facilitate search strategy development, different text mining technologies were suggested for different stages of the search process in both articles. The TerMine service was used to identify relevant search terms within a collection of documents. TerMine is a term extraction program that identifies the key terms and compound terms in a body of text and produces a list of terms in order of their frequency and significance of term occurrence. The authors claimed that the benefit of using a facility such as this is that the range of search terms can be expanded in a way that better describes the literature in a review but that the limitation is a function of its strength: it expands the review in favour of the literature that uses the same language as the documents that have already been found. This method will not identify cross-disciplinary research very well, the authors concluded, and that on its own it will not assist the reviewer in identifying literature that use different words to describe the same concepts.

Bradley et al. (149) compared the text mining tools GoPubMed and PubReMiner in their ability to perform statistical analyses of PubMed references. Both programs generate frequency tables from result sets outlining the number of records by text word, controlled vocabulary, year, etc. The study examined whether these two free text mining tools provided reliable search term suggestions and to what extent they provided additional, useful text words and MeSH terms. Problems and variances in the results were found, and the authors concluded that the two tools could be a useful supplement for search planning but should be used with caution.

To examine the impact of using search techniques aimed at higher precision, Bramer et al. (150) analysed previously completed systematic reviews and focused the original searches to major thesaurus terms or terms in title and/or abstract only in Embase.com or in Embase.com and Ovid MEDLINE combined. The total number of search results was examined in both Embase and MEDLINE, and it was checked whether included studies were retrieved by these more focused approaches. The authors concluded that if the number of search results retrieved is too high for the time and resources reviewers can dedicate to the screening process, search strategies in Embase alone or in Embase and MEDLINE combined can be focused by searching for thesaurus terms as major descriptors as long as thorough searches in other databases, such as Web of Science, was performed. Searching Embase and MEDLINE using terms in titles and abstracts alone resulted in too many relevant articles being missed and was therefore not recommended.

Day et al. (136) investigated the use of simplified search strategies, which they developed for Cochrane reviews from the “Back, Musculoskeletal Group” and the “Pain, Palliative and Supportive Care Group”. The simple search strategies were developed manually and comprised of one line containing one term for the intervention, a second line for the condition in question, a third line combining the first two, and a fourth line limiting the search to clinical trials. A super simple search consisted of one line for the intervention and a second line limiting the search to clinical trials. The search terms were derived from the titles of the reviews. Searches that yielded over 1000 hits were not considered, because the authors did not regard that a “simple” search. The searches were performed in both PubMed and Ovid MEDLINE. Ten systematic reviews including 122 clinical trials met the inclusion criteria. The results depended on the type of intervention being searched. Therefore, the reviews and results were grouped into three intervention-based categories: pharmaceutical interventions, physical interventions, and psychosocial interventions. PubMed found 57 % of the gold standard studies, while Ovid MEDLINE found 49 %. The authors suggested that this difference may be due to the fact that PubMed has an automatic term mapping function, resulting in PubMed providing a broader yield of results. The authors concluded that simple search strategies worked best when the intervention was a pharmaceutical or well-defined physical treatment, but that the recall values was not adequate for conducting systematic reviews for more complex interventions, when a more comprehensive search in at least MEDLINE and Embase was recommended.

In a study by Hausner et al. in 2012 (51), an empirically guided objective development process for search strategies applying text analytic procedures was described. A sample of three Cochrane reviews on brachytherapy was used as input for the search term selection process. After building a development set from a test set and importing the references into EndNote, a term frequency analysis was conducted using the Text Mining Package of the R statistical software. On the basis of information derived from the titles and abstracts of the downloaded references, terms were ranked by frequency. Terms that were present in at least 20 % of the references in the development set were selected for further examination. An experiment performed on 13 Cochrane reviews in 2015 indicated that the objective approach for the development of search strategies described in the previous article was non-inferior to the conventional subjective approach (66).

Karimi et al. (61) compared Boolean and ranked retrieval on a sample of 15 drug-related systematics reviews. The title, background, research question, and inclusion criteria of the reviews were used as search terms for various experiments. To construct the search strategies appropriate for the ranked searches, four search categories were formed. One consisted of the title of the review only, the second of title and background information, the third of title, research questions and inclusion criteria, and the fourth of the Boolean search strategies, as specified in the published sample reviews. Index terms were extracted, and Boolean operators removed. All search strategies used for the study were treated as “bag-of words”, that is, the entire text of the four different categories was copied and pasted and entered as a query without imposing any structure or word order. The results showed that for a ranked query, the more search terms used the better, but performance was low for all queries, so the authors concluded that none of them were suitable for systematic reviews.

In a case review study, O’Mara-Eves et al. (151) used TerMine to identify synonyms for “hard-to-detect” evidence on community engagement in public health interventions to reduce health inequalities. The purpose was to demonstrate alternative systematic ways of identifying relevant evidence where the key concepts that are of interest cut across various disciplines and are generally not central to the primary aims of the studies. The full-text documents of five relevant reviews and discussion papers on community engagement were ran through TerMine to ensure that cross-disciplinary terms were detected. The authors concluded that the use of TerMine to identify relevant search terms was effective and useful for reviews of broad research questions, especially where the key concepts were unlikely to be the main focus of the primary research or inconsistently referred to in the literature.

The identification of health services research is a cumbersome task, and Simon et al. (93) described the development process of performance-oriented search strategies on nurse staffing research by using a text mining approach. A set of relevant references in PubMed/MEDLINE was identified by means of three systematic reviews, and a text mining approach was used to identify potentially relevant free-text terms and MeSH terms from this development set. The frequency of these terms was compared to a random sample from PubMed/MEDLINE in order to identify terms specific to nurse staffing research. The terms were used to develop a sensitive, precise and balanced search strategy. The authors concluded that the described development process for an empirical search strategy was a useful, though technically demanding approach to building performance-oriented strategies.

Stansfield et al. (152) briefly reviewed the literature on applications of text mining for search term development for systematic reviewing and found that text mining tools can be used in five overarching ways:

1. Improving the precision of searches
2. Identifying search terms to improve search recall
3. Aiding the translation of search strategies across databases
4. Searching and screening within an integrated system
5. Developing objectively derived search strategies.

In a case study, the same authors reflected on the utility of certain text mining technologies in improving the precision and recall of searches. They found that text mining can aid the discovery of search terms for search strategies for diversely described topics to support an iterative search strategy development process. Using multiple tools appeared to be particularly fruitful. Their usefulness was influenced by the varying functionality of the tools, the way they were used, and the text that was analysed. The authors concluded that "An awareness of how the tools perform can help use them more efficiently and effectively, although the overriding challenge of finding efficient ways to identify an unknown body of literature for incorporation in systematic reviews still remains."

Leximancer is a text analytic tool and was used by Thompson et al. (153) to construct a set of search terms from a corpus of literature pertaining to transborder interventions for drug control. Leximancer examines a body of text to produce a ranked list of terms on the basis of frequency and co-occurrence usage. The search term selection process consisted of different stages. The first involved uploading relevant and randomly selected references into the Leximancer software. The second stage in the process generated a list of article omission criteria to reduce the initial search results to a more relevant corpus of text. This had to be done manually by the user by looking through titles of the results and comparing them with the omission criteria. A random fifty percent of the refined list was then used as the text corpus. The primary objective of the third stage was to provide a list of terms that encompassed the topic of the systematic review. To this list was added any user-generated concepts that did not appear in the list of concepts and terms generated by Leximancer. In stage five the results from the Leximancer output were used to generate a preliminary list of search terms. The final stage in this systematic approach for search term selection was a visual output, a concept cloud, provided by the Leximancer software, which enabled the user to easily view the most frequent and relevant concepts and could be part of the search process documentation in a review. The method aimed to contribute a more systematic approach for selecting terms in a manner that should be replicable for any user. The authors argued that traditional iterative methods, like "citation pearl growing" and "citation snowballing" are valuable strategies to find useful search terms, but to find a more complete set of search terms on transborder interventions for drug control, the use of Leximancer proved successful to select search terms from a body of literature.

In a study on simple search techniques in PubMed, Waffenschmidt et al. (154) aimed to determine whether use of the “Similar articles” function in PubMed and a simple-structured Boolean search, as proposed by Sampson in her PhD thesis as one method for updating systematic reviews (155), were efficient and reliable search techniques for assessing the completeness of an evidence base consisting of published RCTs on pharmaceutical interventions. The two techniques were tested alone or in combination to determine whether they were able to retrieve all studies considered in the review. The simple Boolean searches were applied manually selected search terms without restriction to search fields to use PubMed’s automatic term mapping function. For each search strategy, search terms were selected for the therapeutic indication and intervention, linked via AND and combined with PubMed’s Clinical Query feature. The components of the research question in 19 systematic reviews were used to select search terms and classified as either belonging to the indication or to the intervention. The search terms were generally the active ingredient and the disease. The single techniques yielded either insufficient completeness, reliability, or efficiency. The authors concluded that their findings indicated that the combination of the first 20 references in the list generated by the “Related Articles” function and a simple-structured Boolean search in PubMed was an efficient and reliable method to assess the completeness of the evidence base consisting of published RCTs on new drugs. Depending on the desired completeness, reliability, and efficiency of a search, such a combination could also be used for preliminary searches, the validation of search strategies, or as a routine component in any systematic search.

3.4.4 Discussion

Different alternative methods for search term selection and development of simpler search strategies have been tried out to make the process simpler to perform, more objective, and transparent. This scoping review found 13 relevant studies on the two topics. Twelve of them described how to select search terms for systematic reviews. Of these, three studies described a manual and nine more automated methods to select search terms. Three studies investigated the use of simpler Boolean search strategies, and one study explored ranked searches without the use of Boolean operators and extended Boolean retrieval.

3.4.4.1 Selection of search terms

Manual selection methods

A manual selection method is the conventional way to find feasible search terms for information retrieval in systematic reviews. The terms are selected by the search expert in close collaboration with the rest of the review team. The sources most often used for manual selection of search terms in the studies included in this review, were the title, background, research question, or inclusion criteria of the review. The title is a reasonable source for selecting search terms since titles of academic publications should be informative (156) and indicate the subject and content of the publication (157). For some of the reviewed articles, search terms derived from the title of a review was helpful to retrieve relevant studies. For systematic reviews on pharmaceutical interventions for example, terms belonging to the therapeutic condition and intervention, derived from the titles of the reviews, were used to develop search strategies with high performance (136).

Automatic selection methods

Most of the studies included in this review found benefit in automating search term selection for systematic reviews. This was especially true for reviews comprising large unfocused topics that were inconsistently referred to in the literature. A systematic review may address outcomes that are latent constructs, that is, factors that cannot be directly measured or observed, and referred to by different terms within a single body of literature (153). A review may also address a question that spans multiple disciplines with varying terminology, or standard terminology is not used by the authors for some reason. Given the complex nature of many topics, it is difficult to know when a strategy is complete, and the use of automated methods in search term selection may help searchers decide when enough search terms have been found.

Automatic search term selection could be used to ease the selection process, especially on broad or multidisciplinary topics. There are some challenges concerning automatic term selection, though, and one concerns subjectivity/objectivity. Statistical methods are claimed to be more objective than the conventional, non-statistical method. Certain stages of the statistical methods are, however, based on user input. Thompson et al. (153) described a method for search term selection in systematic reviews consisting of six different stages, and the fifth, forming a list of terms from the machine-generated results, was user-generated. The overall approaches of text mining for search term selection are similar. The first step is to create a developmental set to train the text mining application, and the most common method to generate such sets is to create a corpus of included references from completed systematic reviews on the topic of interest. Using a set of relevant references as starting point will favour the retrieval towards a similar vocabulary, though (2). The term extraction algorithm will always depend on the content of the documents supplied to it and should therefore never be used on its own but rather in conjunction with the expertise and usual processes that are followed when developing a search strategy (151).

Some of the included studies in this review concluded that by help of automated search term selection, systematic reviews can be completed more quickly and thereby increase efficiency. Others were more reluctant and concluded that text mining tools could be a useful supplement to manual search term selection but should be used with caution as the only approach. A review by AHRQ on the use of text mining tools for systematic reviews concluded that mostly, such tools seem promising, but further research is needed on various variables, like costs, workload, and when such tools are best suited and beneficial (105).

A key difference between a systematic and a non-systematic review lies in the detailed, transparent, and reproducible documentation of the search process. The selection of search terms, however, is one part of the systematic search process that is less transparent (153). The use of search term selection methods that are transparent, reproducible, and generalizable across disciplines would give the whole systematic review process more credibility, improve the overall methodical nature, and make the findings more defensible (153), and more automated methods could be one way to increase transparency of the search process. It is more debatable, however, whether it simplifies the process. The search term selection process using Leximancer, for example, was described as meticulous and time-consuming by Thompson et al. (153).

3.4.4.2 Development of simpler search strategies

Only a few studies included in this review had investigated approaches to simplify the development of search strategies for systematic reviews. The main reason to test simpler search strategies is to increase efficiency. Simpler search strategies will most likely reduce the number of references to screen and thereby reduce workload. Such strategies are also easier to develop and will not require advanced search competence.

In a systematic search, both subject headings and text words, that is, words in title and abstract of the references, should be searched. Approaches to simplify search strategies could be to search fewer fields, for example only subject headings or words in the titles of the references. Another way to simplify a search strategy is to focus the search, for example by limiting to major thesaurus terms. Bramer et al. (150) found that a search could restrict to major thesaurus terms in MEDLINE and Embase if other databases were searched in addition. This could, however, result in an equally high amount of references to screen and therefore not increase efficiency.

Guidelines encourage searchers to use few concepts in a systematic search strategy but many synonyms and word variants for each concept. It should be tested out, however, whether it is necessary to use many search terms in every search strategy. Another option to simplify a search strategy could therefore be to reduce the number of synonyms and/or word variants. Day et al. developed simple search strategies manually that comprised only two search terms limited to clinical trials. They also developed a super simple strategy consisting of one line for the intervention and a second line limiting the search to clinical trials. The sensitivity of both kinds of searches, however, was not adequate for systematic reviews. Simple-structured Boolean search strategies was also tested by Waffenschmidt et al. but had to be combined with PubMed's "Similar Articles" to be effective. The two methods combined proved an efficient and reliable search technique to assess the completeness of an evidence base consisting of published RCTs but not the simple-structured search strategies alone.

A reduction in search terms will possibly reduce the number of references to screen and certainly make the development of search strategies much easier, but it will also most likely lose relevant studies. This will, however, differ according to research question. It could be that the need for search terms is not as comprehensive as is usually thought at least for non-complex and narrow topics. For more complex and broad questions, simple search strategies will probably not be adequate.

3.4.5 Conclusion

It seems like a combination of manual and automatic methods will provide the best solution on search term selection for many research questions based on the studies included in this review. Automatic selection of search terms proved especially useful for topics on diffuse and broad questions. When language is standardized, like on drugs and physical interventions, and the titles of the reviews describe the content of the publication, titles are good sources for selecting search terms and could in some cases probably be the only source.

Based on the studies included in this review, the use of simpler search strategies might be an alternative to conventional, comprehensive strategies on some occasions. The simpler methods described proved most successful for research questions of non-complex interventions, like the effect of drugs and physical interventions. It is difficult to know beforehand, though, what will be an optimal search strategy. Stansfield et al. concluded in their study on search term selection that the overriding challenge of finding efficient ways to identify an unknown body of literature for incorporation in systematic reviews remains (152).

3.5 Scoping review III: Automated retrieval methods for systematic reviews

3.5.1 Introduction

The conventional method to find research studies to include in systematic reviews is comprehensive Boolean searching. Such searches, however, are time-consuming and difficult to perform, and human effort is a scarce and valuable resource which should be expended only where automation is impossible, impractical, or undesirable (158). Automated retrieval methods have therefore been suggested as an alternative to Boolean searching and great efforts have been invested over the last years to develop and evaluate more automated retrieval methods for systematic reviews (159, 160). The question is whether they can compete with conventional Boolean searches. To present an overview of research on this topic, a scoping review was performed to investigate which automated information retrieval methods have been tried out for systematic reviews and how they performed. The included studies were also used to inform the empirical study.

3.5.1.1 Aim and objectives

The main aim of this scoping review was to determine the extent and nature of existing evidence on the development and use of automated retrieval methods for systematic reviews by gathering and presenting available research evidence on existing research on the topic.

3.5.1.2 Research question

The following research question was formulated: What previous research has been done regarding automated retrieval methods when conducting systematic reviews and what are its characteristics and conclusions?

3.5.1.3 Comments

The search process for information retrieval in systematic reviews consists of several stages, and this review focuses on the performing or executing stage of a search. Scoping reviews I and II focused on the planning stage of the systematic search process.

When searching for relevant studies for systematic reviews, extensive searches are carried out to locate as many relevant research studies as possible on a research question. Such a process usually includes using several sources, like searching electronic databases, scanning reference lists, and asking authors. This study will only investigate automated methods used for searching electronic databases.

AHRQ has written a review on the use of text mining software in systematic reviews (105). The present scoping review aimed to review all automated retrieval methods for the execution of searches in systematic reviewing.

3.5.2 Methods

A scoping review of research on the application of automated methods for information retrieval in systematic reviews was conducted. To give a picture of existing evidence on the topic, literature on different automated search techniques was searched for and screened to identify, characterize, and summarize any potential alternative automated information retrieval methods to Boolean searching.

3.5.2.1 Information management

All records were uploaded to the reference manager program EndNote for deduplication and further to Covidence for screening.

3.5.2.2 Search methods

The research question is of an interdisciplinary character, so the following databases and registers were searched:

1. Ovid MEDLINE In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily, Ovid MEDLINE(R) and Ovid OLDMEDLINE(R) 1946 to Present
2. Web of Science - Clarivate Analytics
3. Library Information Science & Technology Abstracts (LISTA) - Ebsco
4. Library & Information Science Source (LISS) - Ebsco
5. Cochrane Methodology Register

Publication lists of relevant websites and organizations and reference lists of relevant studies were also browsed.

The search strategy used in the database searches was tested in Ovid MEDLINE and then adapted for the other databases. A sensitive search strategy consisting of three clusters of terms was used: one relating to systematic reviews, another to information retrieval, and the third to different automated methods (like text mining and ranked retrieval). The MEDLINE search strategy can be found in Appendix VI and a PRISMA flow diagram of the process in Appendix IX.

Database and website searches were conducted in November 2015. An update search was performed in Ovid MEDLINE by one of the authors on 13th of December, 2018.

3.5.2.3 Eligibility criteria

Eligibility criteria used to select publications for review and extraction are listed below:

Reviews were included based on the following inclusion and exclusion criteria:

Inclusion criteria:

1. Published after 2004.
2. Concerning any automated or partial match information retrieval method, more specifically, to the execution of searches.
3. The aim of the search is high recall, that is, included studies have to be relevant for information retrieval suited for systematic reviews.
4. Published in English or any Scandinavian language.
5. All study designs, except reviews, were included.

Exclusion criteria:

1. Topic too narrow, such as genetics, proteins, chemical, or molecular compounds where a more specific language is used.
2. Literature concerning the screening of references to include in systematic reviews.

3.5.2.4 Screening

Studies were screened in a two-stage screening process. First, records were assessed independently against inclusion and exclusion criteria based on their titles and abstracts, and second, relevant records were assessed based on the full text. Studies were selected for inclusion by two authors (IK and MM), independent of each other. The authors then met and reviewed every selection. Discrepancies were dissolved by consensus.

3.5.2.5 Data extraction

A data charting form was developed to determine which variables to extract by one author (MM). The form was continuously revised and updated during the extraction process.

All studies were extracted and recorded for information on the following issues:

1. Bibliographic details
2. Research fields/topics
3. Study designs
4. Aims and objectives
5. Methods/techniques used
6. Tools or systems developed or used
7. Key findings/conclusions

3.5.2.6 Limitations

As this is a broad topic investigated in different fields, studies may have been missed due to differences in language used between the areas.

3.5.3 Results

3.5.3.1 Study selection

The result of the literature search was 1442 unique publications after de-duplication. After reading titles and abstracts, 74 were read in full text for further assessment. Eleven of these were found relevant according to inclusion and exclusion criteria. Two relevant studies were found in other sources. Thus, 13 studies were found eligible. The flow diagram in Appendix IX shows the result of the searches and the selection process.

3.5.3.2 Characteristics of included studies

Characteristics of included studies are presented in Table 4, more specifically, fields/topics, study design, aims/objectives, techniques/methods, tools/systems/databases used, and key findings/conclusions.

Table 4 Characteristics of included studies - Scoping review III

	Author	Year	Field / topic	Study design	Aim and/or objective	Technique / method	Tools / systems / databases used	Key findings / conclusions
1	Ananiadou	2009	Social science	Descriptive	The aim of the project (ASSERT) was to develop a text mining framework to support and enhance systematic reviews. The search stage was supported by using query expansion.	Text mining: Document clustering Document classification Query expansion	Carrot2 TinySVM Lucene 2.2 TerMine	Through use of semi-automated techniques reviews can be completed more quickly and more systematically.
2	Bowes	2012	Software engineering	Descriptive	Develop an open source web-enabled database. Support the search stage by applying predefined search terms to online databases automatically.		SLuRp (Systematic Literature unified Review program)	Specific tools for systematic reviewing, like SLuRp, can reduce time and errors.
3	Bui	2015	Health and medicine	Experimental	Improve the traditional search approach for clinical guidelines by developing a reference retrieval system integrated over PubMed composed of query expansion and citation ranking.	Unsupervised query expansion and ranking techniques	PubMed	Unsupervised query expansion and ranking techniques were more flexible and effective than PubMed's default search engine behaviour and a machine learning classifier. Compared with the baseline PubMed expansion, the query expansion algorithm improved recall (80.2% vs. 51.5%) with a small loss on precision (0.4% vs. 0.6%).
4	Cohen	2010	Health and medicine	Descriptive	Support literature collection, collation, and the triage step in the production and management of systematic reviews by developing a text mining based pipeline for accelerating the systematic review process. Support the search stage by developing a meta-search engine	Text mining		The text mining-based pipeline will decrease the manual burden of systematic reviewers during the literature collection and review process. Ultimately, this should lead to better and more cost-effective healthcare.
5	Fernandez-Saez	2010	Software Engineering	Descriptive	Support and enhance the production of systematic reviews by developing a tool for performing systematic literature reviews. Support the search stage by applying text mining techniques to cluster the review documents by using the similarities among them.	Text mining	Lucene SLR-Tool	The main contribution of the article was to present the SLR-Tool.
6	Karimi	2010	Health and medicine	Experimental	Improve information retrieval for systematic reviews by exploring and comparing the effectiveness of Boolean and ranked querying.	Ranked retrieval Extended Boolean retrieval	MEDLINE (Ovid interface) Zettair Okapi BM25 Lucene	The best performance was obtained by the queries containing the most search terms, with recall score of 15.5% when considering the top 1,000 documents, and 46.1% when considering the top 10,000 documents

	Author	Year	Field / topic	Study design	Aim and/or objective	Technique / method	Tools / systems / databases used	Key findings / conclusions
								The experiments showed that ranked retrieval by itself is not viable for search tasks requiring high recall. However, extended Boolean retrieval, a combination of less complicated Boolean queries and ranked retrieval, outperformed either of them individually, leading to possible time savings over the current process.
7	Pohl	2010	Health and medicine	Experimental	Improve information retrieval for systematic reviews by exploring ranked querying versus Boolean and extended Boolean retrieval.	Ranked retrieval Extended Boolean retrieval	Ovid MEDLINE Zettair Okapi BM25 Lucene	Extended Boolean retrieval models are able to increase the fraction of relevant documents found after inspecting the usual 500 to 2000 documents, by loosening the strictness of conjunctive operators and introducing some elements of ranking.
8	Ramampiaro	2010	Software engineering	Descriptive	Present an effective information retrieval tool to support evidence-based software engineering on performing systematic reviews. Support the search stage by developing a meta-search engine.	Text mining (clustering) Information retrieval techniques (indexing)	EBSE (Evidence Software Engineering)	The use of techniques from information retrieval, as well as text mining, can support systematic reviews and improve the creation of repositories of software engineering empirical evidence.
9	Smalheiser	2014	Health and medicine	Descriptive	Support the search stage by developing a meta-search engine. Presentation of Metta, a meta-search engine, one piece of a multi-tool pipeline that will assist systematic reviewers in retrieving, filtering, and assessing publications.		PubMed, Embase, CINAHL, PsycINFO, CENTRAL Metta	Metta can play a valuable role in speeding up the process of retrieving the initial set of records during the preparation of a systematic review, as part of an overall reengineering of the process.
10	Sturm	2015	Information Systems (IS) field	Descriptive	Support the search stage by developing a meta-search engine. Presentation of LitSonar, a tool to enhance the efficiency of the search process.		LitSonar	The proposed system has the potential to enhance systematic literature searches.
11	Thomas	2011	Health and social care	Descriptive	Automate the identification of relevant literature to reduce the time it takes to conduct systematic reviews.	Text mining Machine learning	MEDLINE TerMine	Text mining technologies have the potential to assist at various stages of the review process.
12	Waffenschmidt	2013	Health and medicine	Experimental	Assess the completeness of an evidence base consisting of published RCTs by using simple search techniques in PubMed. The aim was to determine whether PubMed's Related Citations and/or a simple-structured Boolean search were efficient and reliable search techniques.	PubMed's related citations	PubMed	Combining the first 20 "Related articles" in PubMed and a simple search strategy was a suitable method to assess the completeness of an evidence base of published RCTs.
13	Young	2011	Health and medicine	Experimental	To develop and test a learning algorithm to effectively utilize MeSH terms and refine search strategies.	Machine learning algorithm	MEDLINE	The use of automatic and learning-based algorithms for searching seemed to be both plausible and desirable.

3.5.3.3 *References to studies included in this review*

1.	Ananiadou S, Procter, Rob, Rea, Brian, Sasaki, Yutaka, Thomas, James. Supporting Systematic Reviews Using Text Mining [cited 2014 Apr 22]. Available from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.186.3095&rep=rep1&type=pdf.
2.	Bowes D, Hall T, Beecham S, editors. SLuRp: a tool to help large complex systematic literature reviews deliver valid and rigorous results. Proceedings of the 2nd international workshop on Evidential assessment of software technologies; 2012: ACM.
3.	Bui DD, Jonnalagadda S, Del Fiol G. Automatically finding relevant citations for clinical guideline development. J Biomed Inform. 2015;57:436-45.
4.	Cohen AM, Adams CE, Davis JM, Yu C, Yu PS, Meng W, et al., editors. Evidence-based medicine, the essential role of systematic reviews, and the need for automated text mining tools. IHI'10 - Proceedings of the 1st ACM International Health Informatics Symposium; 2010.
5.	Fernández-Sáez AM, Bocco MG, Romero FP, editors. SLR-Tool a tool for performing systematic literature reviews. ICSOFT 2010 - Proceedings of the 5th International Conference on Software and Data Technologies; 2010.
6.	Karimi S, Pohl S, Scholer F, Cavedon L, Zobel J. Boolean versus ranked querying for biomedical systematic reviews. BMC Med Inf Decis Mak. 2010;10:58.
7.	Pohl S, Zobel J, Moffat A, editors. Extended Boolean retrieval for systematic biomedical reviews. Conferences in Research and Practice in Information Technology Series; 2010.
8.	Ramampiaro H, Cruzes D, Conradi R, Mendona M, editors. Supporting evidence-based Software Engineering with collaborative information retrieval. Proceedings of the 6th International Conference on Collaborative Computing: Networking, Applications and Worksharing, CollaborateCom 2010.
9.	Smalheiser NR, Lin C, Jia L, Jiang Y, Cohen AM, Yu C, et al. Design and implementation of Metta, a metasearch engine for biomedical literature retrieval intended for systematic reviewers. Health Inf Sci Syst. 2014;2:1.
10.	Sturm B, Schneider S, Sunyaev A, editors. Leave no stone unturned: introducing a revolutionary meta-search tool for rigorous and efficient systematic literature searches. ECIS - European Conference on Information Systems 2015; Münster, Germany: Association for Information Systems.
11.	Thomas J, McNaught, John, Ananiadou, Sophia. Applications of text mining within systematic reviews. Res Synth Methods. 2011(2):1-14.
12.	Waffenschmidt S, Janzen T, Hausner E, Kaiser T. Simple search techniques in PubMed are potentially suitable for evaluating the completeness of systematic reviews. J Clin Epidemiol. 2013;66(6):660-5.
13.	Young S, Duffull SB. A learning-based approach for performing an in-depth literature search using MEDLINE. J Clin Pharm Ther. 2011;36(4):504-12.

3.5.3.4 Synthesis of results

The 13 included studies were published between the years 2008 and 2015.

The included documents were from three different fields. Eight articles were from healthcare and medicine, four from software engineering and one from the social sciences.

All kinds of study designs, except reviews, were included and among the included there were five experimental and eight descriptive studies.

Aims

The aim of all the included studies was to improve the information retrieval process for systematic reviews by developing and/or testing various automated methods using different algorithms, tools and systems. Some of the systems were old and some were developed for new projects. MEDLINE was the electronic database most often used for testing new technology to improve search performance. Other databases were also used, but mostly as part of meta-search engines.

Text mining

The use of text mining techniques for systematic reviews was described and/or tested in several of the included studies. The text mining techniques most relevant for performing searches were query or term expansion and document clustering or classification. The ASSERT project was described both by Ananiadou et al. (7) and Thomas et al. (2). ASSERT is a collaborative project between the UK National Centre for Text Mining, the Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre), and the National Centre for e-Social Science (NCeSS) on how to apply text mining techniques for accelerating the process of systematic reviewing. It was demonstrated to systematic reviewers how they can work alongside computer scientists to develop tools and methods that will help them bring robust evidence to bear on decisions in a timelier manner than is possible at present. Query expansion and document clustering were the text mining techniques described in these articles to facilitate identification of relevant literature. The authors claimed that text mining might contribute to different aspects of conventional systematic review methods.

Most of the projects described in this review were developed in health and medicine or social science. A few, however, were developed in software engineering, for example the SLR-tool, a free tool to support and enhance the production of systematic reviews, described by Fernandez-Saez, Bocco and Romero (69). Text mining techniques were used by the SLR-Tool to refine searches within the documents. The SLR-Tool not only supported the process of performing systematic reviews but also provided additional functionalities, such as refining searches within the documents by applying text mining techniques and exporting the references from the primary studies to the formats used in bibliographic packages such as EndNote. Users' perception of the SLR-Tool was that it was both highly necessary and useful.

Ranked retrieval

Some of the included studies investigated different machine learning and ranking algorithms. Bui et al. (67) aimed to improve the traditional search approach for clinical guidelines by developing a retrieval system composed of both query expansion and citation ranking integrated over PubMed. Query expansion was used to improve recall while the document ranking aimed to improve precision on top-ranked documents. To validate the system, a gold standard was developed consisting of references that were systematically searched for and screened to support the development of cardiovascular clinical practice guidelines. The expansion and ranking methods were evaluated separately and compared with PubMed expansion and ranking. The conclusion was that the query expansion algorithm tested by the authors improved recall with a small loss on precision compared to baseline PubMed.

Different experiments were performed on the same 15 reviews from AHRQ exploring the effectiveness of using ranked retrieval compared to Boolean querying in the search task for biomedical systematic reviews (61). This study was also reviewed in Scoping review II but then with focus on search term selection. The aim of the study was to develop specialized search tools to improve information retrieval performance in systematic reviews. A series of experiments was conducted, and the results of the ranked searches were compared to baseline Boolean search results obtained using Ovid MEDLINE. Precision and recall were used as evaluation for search performance, and the included studies of the reviews were used as gold standard. For comparability, precision and recall were reported at cut-off levels of 1000 and 10 000. The authors stated that the thresholds were chosen to match the typical size of Boolean query output sets when searching for documents to include in systematic reviews. The effectiveness of the ranked queries showed low recall values for all four types of ranked searches. The queries which incorporated more information, performed best, with a recall score of 15.5 % when considering the top 1000 documents, and 46.1 % when considering the top 10 000 documents. Ranked queries which were derived by taking all terms from the original Boolean queries showed performance that was lower than for the other queries. The authors also investigated the extended Boolean retrieval model, which provides ranked output to conventional Boolean searches. In this hybrid approach, Boolean search strategies were used to identify candidate documents, and ranking was then applied to order the result set. The extended Boolean retrieval model outperformed both conventional Boolean searches and the different ranked queries. The findings showed that although Boolean searches have limitations, a combination of simpler Boolean queries[2] and ranked retrieval outperformed either of them individually, leading to possible time savings over the current process.

Young et al. (161) investigated a machine-learning-based approach for performing in-depth literature searches using MEDLINE. The purpose of the study was to develop and test a learning-based approach for conducting a comprehensive literature search in MEDLINE. A learning algorithm to effectively utilize MeSH terms was presented. The algorithm created combinations of available MeSH terms from which a search was conducted. The authors concluded that the use of automatic and learning-based algorithms for searching seemed to be both plausible and desirable.

Related/similar articles

Sampson et al. (162) developed a simple search method in 2008 to identify new evidence eligible for updating systematic reviews based on PubMed's "Similar articles" search in combination with simple Boolean searches limited to PubMed's "Clinical Queries". This method was used by Waffenschmidt et al. (154) to evaluate the completeness of the evidence base presented by pharmaceutical companies. The aim was to determine whether the "Similar articles" function in PubMed, a simple-structured Boolean search (reviewed in Scoping review II) and the Clinical query filter in PubMed were efficient and reliable search techniques, alone or in combination, to find all studies included in a sample of 19 drug-related systematic reviews. The results were used to determine completeness (recall), reliability (the range of recall), and efficiency (precision and NNR). High recall and reliability with an acceptable NNR and sufficient precision was only achieved by the combination of the search techniques. The method pairing the "Similar articles" function and the clinical query filter with simple Boolean searches, gave excellent recall of new material. A great advantage with this method is that it is easy to use and therefore requires little previous knowledge of search techniques. The authors concluded that it seemed possible to use this method for other purposes as well, for example to validate search strategies or for preliminary searches

[2] Simpler Boolean queries in this study meant modified queries, that is, some of the keyword matchings were generalized, and limits on for example publication type and dates were removed.

conducted to gain a rapid overview of the published evidence at the start of a systematic review.

Meta-search engines

Four of the included studies described the development of meta-search engines. Cohen et al. (163) described how they developed a text mining based pipeline including a meta-search engine to support the creation and updating of systematic reviews to increase the rate of production. The development of the meta-search engine, called Metta, was further described in an article by Smalheiser et al. in 2014 (68). Metta was envisioned as one component of the pipeline of informatics tools. The aim was to simplify the process of submitting queries so that users did not need to concern themselves with complex queries, the use of search tags, or other specialized commands. On the other hand, as much flexibility as possible was retained, so that users could adjust the pre-set options for each database if they desired. The authors concluded that Metta may find wide utility for anyone who is carrying out a comprehensive search of the biomedical literature.

A project in software engineering to support the production of systematic reviews was the development of a meta-search engine, EBSE (164). The authors argued that the use of techniques from information retrieval as well as text mining, could support the production of systematic reviews and improve the creation of repositories of empirical evidence in software engineering. They wrote further that an explicit search strategy had to be developed and systematically applied to a range of resources. Their approach was intended as an interface on top of existing repositories with the capability of filtering out records not in the software engineering field to restrict the results to a more manageable size. The authors concluded that the need for such a tool is widely recognized but that they met a number of challenges in the development process.

Sturm et al. (165) described the development of the meta-search engine, LitSonar, as a revolutionary meta-search engine for academic literature in software engineering which consolidated search results from several literature databases. LitSonar aimed to improve the quality of literature reviews by enhancing rigour and efficiency of literature searches. The authors concluded that the proposed system had the potential to enhance systematic literature searches and reviews in several ways, but that the current prototype implementation had some limitations.

The development of SLuRp (Systematic Literature unified Review program) was described by Bowes et al. (166) as a tool to help support all phases in the management of systematic reviews in software engineering. Concerning the search process, SLuRp applied pre-defined search terms to online databases and semi-automatically extracted references from databases and saved these. The authors concluded that if we are to have confidence in the outcomes of systematic reviews, it is essential that such automated systems are used.

See Table 4 for further information.

3.5.4 Discussion and conclusion

In a systematic review, searching involves balancing the requirement to identify all relevant research with the constraints of the resources of the project (2). The aim of all the articles included in this scoping review was to enhance the systematic search process by describing and/or evaluating different automated search methods. The enhancements were to reduce the time it takes to perform a search for systematic reviews and make the process easier. Existing systems were used in new ways, and new systems and algorithms were developed.

Five of the studies described the development of meta-search engines especially suited for systematic reviews in their field. Metta was developed for health and medicine (68, 163) and EBSE, LitSonar and SLuRp for software engineering (164-166). The problem with some of the meta-search engines and other search tools developed to improve information retrieval for systematic reviews is that they are still not available for public use, or they do not exist anymore. The SLR-tool, for example, is not available today (167), which also seems to be the destiny of SLuRp and some of the other projects.

Previous studies have tested various automated search methods, like text mining, machine learning, and ranking algorithms for information retrieval in systematic reviews. Some of the included studies claimed that more automated search systems had the potential to enhance information retrieval in systematic reviews, and one of the studies concluded that a combination of less complicated Boolean queries and ranked retrieval outperformed either of them individually. Based on the reviewed studies, however, it seems plausible to say that no automated search method can replace the conventional Boolean search at the moment. More automated search methods can, however, assist in routine searching, for updates, and for evaluating the completeness of an evidence base. In the creation and maintenance phase of the development of a good search strategy, human specialist skills is still required (158).

3.6 Limitations / Comments

Relevant studies could have been missed due to searching only MEDLINE for Scoping reviews I and II.

3.7 Summary of scoping reviews

In this section, summary of all three scoping reviews will be shortly presented and how key findings relate to the empirical studies of this thesis. Based on the results of the three reviews, there are reasons to believe that for non-complex interventions, a less comprehensive search process is enough to retrieve relevant research studies to include in systematic reviews.

Scoping review I found that searching only MEDLINE or MEDLINE and Embase combined could be enough for some research questions if other sources were searched in addition. The 28 included studies varied to such an extent in conclusions, topics, study design, and other variables, however, that more research is required to decide whether it is enough to search only MEDLINE and Embase for systematic reviews. Scoping review I also found that both database coverage and search performance were used to support conclusions on whether searching the two databases is enough for study identification in systematic reviews, and both measures were therefore used in the empirical part of this study.

Research evidence on the development of simple search strategies is scarce. Only a few studies were found on this topic in Scoping review II. Search term selection has been investigated somewhat more widely, with a few studies using manual selection, but most investigating various automated methods. More research is needed on both topics, however, since only 13 studies were found altogether, and some of them were only descriptive.

Scoping review III found that several more automated information retrieval systems had been developed to enhance the systematic search process. Unfortunately, many of these systems have never been available for public use, and some of them are withdrawn. None of them has replaced Boolean searching, which is still the preferred method to find research studies to include in systematic reviews. One study found, however, that ranking Boolean search results, extended Boolean retrieval, can enhance the retrieval process in systematic reviews. To test an easy to use ranking possibility on conventional, Boolean search strategies, PubMed's sort option "Best Match" was used in the empirical part of this study.

4 Methods

4.1 Introduction

This study has investigated various approaches to simplify and optimize the search process for systematic reviews with the aim of identifying approaches that are more efficient than conventional, comprehensive approaches. The first approach was to perform a cross sectional study evaluating the effect of searching only MEDLINE and Embase for retrieval of included studies in a sample of 400 Cochrane reviews. The Ovid MEDLINE search strategies of a subsample of 254 reviews, including 5 to 50 research studies, was translated into PubMed syntax to compare the two interfaces. Next, an exploratory design was used to find the optimal combination of search terms for each review of the subsample. Then, simplified search strategies using the optimized search terms were developed and performed in Ovid MEDLINE and PubMed, and last, to evaluate a more automated search method, the ranking function “Best Match” in PubMed was tested on conventional, Boolean search strategies in the subsample.

4.2 Scoping reviews

To put the main empirical approaches into context, three scoping reviews were conducted to determine the extent and nature of previous research, as well as to inform the empirical studies. Furthermore, the scoping reviews were used to answer the research questions, alongside the empirical studies. For a presentation of the scoping reviews, see Chapter three.

4.3 Creating the sample

The research population of the empirical part of this study was systematic reviews from The Cochrane Collaboration since the aim was to evaluate the search process of systematic reviews that are known to have high quality (66). Cochrane reviews must meet specific standards and are reported to be of higher methodological quality than other reviews (114, 168) and are therefore also likely to have identified all or most relevant studies (95). Information on the studies included in the Cochrane reviews and the published search process were used as data sources.

4.3.1 Inclusion and exclusion criteria

The reviews included in the sample had to report the search process in detail and present a list of studies included in the review. They also had to include at least one primary study indexed in MEDLINE for the cross sectional studies and 5 to 50 studies for the exploratory and experimental studies and have a reproducible Ovid MEDLINE search strategy attached. MEDLINE was chosen as main search tool and literature database for the test searches since it is the most well-known medical database and the one most often used. Published search strategies performed in Ovid MEDLINE were chosen since most Cochrane review groups use the Ovid interface, and the results will be more consistent if the same interface is used for the whole sample. Overviews, review of reviews, and withdrawn reviews were excluded.

4.3.2 Sampling procedures

A purposive sample consisting of the newest 400 systematic reviews representing 16 different Cochrane Review Groups that fulfilled the inclusion criteria as of May 2015, were used as sample for the cross sectional studies. The sample size and different review groups were chosen to be representative of the population. The newest Cochrane reviews were chosen since search conditions change, for example the indexing and interfaces of the databases, which can make it difficult to reconstruct the searches. In addition, the search methods will probably improve, which can also influence the results. Cochrane has its own Information Retrieval Methods Group (169) to develop methodology, provide advice and support, conduct research, and facilitate information exchange regarding methods to support the information retrieval activities of the Cochrane Collaboration (170). All 53 Cochrane review groups were examined to find 16 that fulfilled the inclusion criteria. It was necessary to examine 497 reviews to find 400 that fulfilled the inclusion criteria. Reasons for exclusion were:

- A total of 55 reviews had zero studies included in the review or zero of the included studies were indexed in MEDLINE.
- In 25 reviews the MEDLINE search strategy was not performed in the Ovid interface.
- 17 search strategies were not reproducible.

For further information on this process, see Appendix I.

The newest 25 systematic reviews from each of the following 16 Cochrane review groups were used as sample for the cross sectional study:

1. Anaesthesia, Critical and Emergency Care Group (ACECG)
2. Dementia and Cognitive Impairment Group (DCIG)
3. Developmental, Psychosocial and Learning Problems Group (DPLPG)
4. Epilepsy group (EG)
5. Effective Practice and Organization of Care Group (EPOC)
6. Eyes and Vision Group (EVG)
7. Gynaecology and Fertility Group (GFG)
8. Gynaecological, Neuro-oncology and Orphan Cancer Group (GNOCG)
9. Heart Group (HG)
10. Injuries Group (IG)
11. Kidney and Transplant Group (KTG)
12. Musculoskeletal Group (MG)
13. Oral Health Group (OHG)
14. Pain, Palliative and Supportive Group (PPSG)
15. Stroke Group (SG)
16. Wounds Group (WG)

The reviews from these groups were used as the study sample as they fulfilled the inclusion criteria and had the newest 25 reviews of all groups at the time of inclusion.

A subsample containing reviews with less than 5 and more than 50 included studies was also evaluated for the cross sectional studies, as the consequences of missing one study in reviews with few studies included can affect the overall performance more than reviews with a higher total of included studies (171), and the few reviews with more than 50 included studies could also skew the results as some of these had extreme values. This resulted in a sample of 254 reviews, which were also used for the exploratory study and the experiment. The subsample of 254 gave the following distribution of reviews across the 16 groups:

	Review group	**# of reviews**
1	Anaesthesia, Critical and Emergency Care Group (ACECG)	20
2	Dementia and Cognitive Impairment Group (DCIG)	16
3	Developmental, Psychosocial and Learning Problems Group (DPLPG)	16
4	Epilepsy group (EG)	12
5	Effective Practice and Organization of Care Group (EPOC)	18

6	Eyes and Vision Group (EVG)	13
7	Gynaecology and Fertility Group (GFG)	17
8	Gynaecological, Neuro-oncology and Orphan Cancer Group (GNOCG)	12
9	Heart Group (HG)	17
10	Injuries Group (IG)	13
11	Kidney and Transplant Group (KTG)	17
12	Musculoskeletal Group (MG)	20
13	Oral Health Group (OHG)	16
14	Pain, Palliative and Supportive Group (PPSG)	17
15	Stroke Group (SG)	16
16	Wounds Group (WG)	14

For further information on this process, see Appendix II.

The results of the full (=400) and reduced sample (=254) were compared for the cross sectional study to investigate whether removing the extreme values affected the results.

4.3.3 Number of included studies in reviews

Before evaluating coverage and search performance, the number of included studies in the sample reviews had to be recorded.

Full sample:

The 400 systematic reviews included a total of 6222 primary studies with an average of 15.6 (median=9.5) per review. The average number of included studies per review in each group varied considerably and ranged from 7.2 in the Epilepsy Group (EG) and the Eyes and Vision Group (EVG) to 25.0 in the Heart Group (HG). There were vast differences between number of included studies across all reviews with 1 included study as the least and 296 as the highest in the review “Lipid- lowering efficacy of atorvastatin” from the HG (172).

Subsample:

The subsample of 254 systematic reviews, including from 5 to 50 studies, included a total of 4722 primary studies with an average of 18.6 (median=13.0) per review. In this sample, the average number of included studies per review in each group ranged from 10.8 in the Epilepsy group (EG) to 24.1 in the Kidney and Transplant Group (KTG).

As could be expected, both the median and average number of included studies per review was higher in the subsample since the number of reviews with less than 5 included studies (=135) in the full sample was much bigger than the number of studies with more than 50 (=9).

4.4 Data analysis

Excel was used to keep track of data and analyse study results.

Different measures were used for the different studies and are described in the paragraphs below.

4.4.1 Descriptive statistics

Average and median were used to measure central tendency of the results.

4.4.2 Relevance assessment

To measure coverage of MEDLINE and Embase for primary studies, the total number of included studies of each of the sample reviews was used as gold standard.

To measure search performance, all MEDLINE-indexed included studies of the sample reviews were used as gold standard for research studies to be found by the searches (95).

4.4.3 Database coverage

Coverage was calculated for relevant references to be indexed in MEDLINE and if a reference was not indexed in MEDLINE, whether it was indexed in Embase.

4.4.4 Search performance measures

Precision and recall were used to measure effectiveness of all searches.

Mean average precision (MAP) was used to measure effectiveness of the ranked searches.

Number Needed to Read (NNR) was used to measure efficiency of all searches.

4.5 Searching only MEDLINE and Embase for study identification

The aim of the cross-sectional study was to evaluate the performance of searching MEDLINE alone or MEDLINE and Embase combined for the sample reviews. Effectiveness was measured by investigating the coverage of MEDLINE and Embase for studies included in the sample reviews and testing the performance of the published Ovid MEDLINE search strategies of the same reviews. To test the performance of conventional search strategies, both PubMed and Ovid MEDLINE were evaluated to compare the two interfaces. Efficiency of the search process was measured by counting electronic databases and registers searched in each review and the number of search terms used in the published Ovid MEDLINE search strategies, and also by calculating NNR of the search strategies. Information on each of the sample reviews was extracted including search strategies, number of electronic databases and registers searched, and bibliographic data of each included study.

4.5.1 Comment

The MECIR standards recommend that MEDLINE, Embase, and CENTRAL should be searched as a minimum in the development of Cochrane intervention reviews (41). This study has only investigated the coverage of MEDLINE and Embase and the performance of MEDLINE search strategies. CENTRAL, one of the most important sources for clinical trials, was left out since most included studies in Cochrane reviews will eventually be indexed in CENTRAL, which would bias the results in favour of this database (63). It could not be assumed that the included studies would have been located in this source prior to the publication of the Cochrane reviews in the sample of this study (63).

4.5.2 Database coverage

To determine the coverage of MEDLINE alone or MEDLINE and Embase combined for included studies in the sample of 400 Cochrane reviews, the following procedures were used:

First, the list of included studies, found in the References section of the published reviews, was extracted from every review in the sample. Next, it was determined whether a study was indexed in MEDLINE or Embase through manual searches using various combinations of extracted bibliographic information. More specifically, a known-item search was undertaken in Ovid MEDLINE for each included study in all reviews to investigate whether they were indexed in MEDLINE and then, if not indexed in MEDLINE, whether they were indexed in Embase. A known-item search implies a user who is looking for one particular study (62). The two databases were searched using the “Find Citation” option in Ovid with different information of each reference, for example, first author’s surname, year of publication, and the first page of the reference. Studies not found in the first search were looked up for a second time using different search terms, for example words in title, until it was certain that the article was not indexed in MEDLINE or Embase. If a reference was difficult to verify, it was also checked whether the journal was indexed in MEDLINE using PubMed’s “Journals in NCBI Databases”. All known-item searches were conducted in the period from May to October, 2015.

4.5.3 Search performance

The second approach of the cross sectional study was to measure performance of the published Ovid MEDLINE search strategies of the 400 sample reviews. The following procedures were used: information on the search strategies was found in the section “Appendices – Search strategy for MEDLINE (Ovid SP)” of the reviews. If more than one MEDLINE search strategy was published in the review, the newest was used. If it was an update search, the search strategy was rerun without time limits used for the update. In the period from May to October 2015, the search strategies were typed and rerun one by one to investigate whether they found the included studies that were indexed in the database in the following Ovid MEDLINE segment: Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily, Ovid MEDLINE(R) and Ovid OLDMEDLINE(R) 1946 to Present.

4.5.3.1 Filter for study design

Filter for study design is frequently used for systematic reviews to decrease the number of retrieved records. Cochrane reviews of interventions typically limit eligible studies to those using robust design, such as RCTs. To restrict the search accordingly, a search filter for identifying randomized trials is often applied. Some empirical evidence, however, has found that even the most sensitive methodological filters miss relevant studies (173, 174). To contribute additional data to this area of research, the number of hits found by the published search strategies were recorded with and without filter for study design (where relevant), and the performance of both were investigated.

Below follows an example for one of the published Ovid MEDLINE search strategies with filter for RCTs on the topic dementia and aromatherapy (175):

1. exp Dementia/
2. Delirium/
3. Wernicke Encephalopathy/
4. Delirium, Dementia, Amnestic, Cognitive Disorders/
5. dement*.mp.
6. alzheimer*.mp.
7. (lewy* adj2 bod*).mp.
8. deliri*.mp.
9. (chronic adj2 cerebrovascular).mp.
10. ("organic brain disease" or "organic brain syndrome").mp.
11. ("normal pressure hydrocephalus" and "shunt*").mp.
12. "benign senescent forgetfulness".mp.
13. (cerebr* adj2 deteriorat*).mp.
14. (cerebral* adj2 insufficient*).mp.
15. (pick* adj2 disease).mp.
16. (creutzfeldt or jcd or cjd).mp.
17. huntington*.mp.
18. binswanger*.mp.
19. korsako*.mp.
20. ((cognit* or memory* or mental*) adj3 (declin* or impair* or los* or deteriorat*)).mp.
21. or/1-20
22. "aroma therap*".mp.
23. exp *Aromatherapy/
24. aromatherapy.mp.
25. "complementary therap*".mp.
26. exp Complementary Therapies/

27. "alternative therap*".mp.
28. exp Complementary Therapies/
29. "essential oil*".mp.
30. aroma*.ti,ab.
31. ("lemon balm" or "rose* oil*" or "lavender oil*").mp.
32. or/22-31
33. 21 and 32
34. randomized controlled trial.pt.
35. controlled clinical trial.pt.
36. placebo.ab.
37. random*.ab.
38. trial.ab.
39. groups.ab.
40. or/34-39
41. (animals not (humans and animals)).sh.
42. 40 not 41
43. 42 and 33

4.5.3.2 Number of electronic databases searched

The average number of electronic databases searched was calculated for each review as one way to measure efficiency. The documentation on how many databases that are searched for each review is not fully documented in all Cochrane reviews. There is, however, a section in every review called “Methods - Search methods for identification of studies - Electronic searches”, and the databases mentioned in this section and documented in an appendix with a search strategy were used as data.

4.5.3.3 Number of search terms in published Ovid MEDLINE search strategies

Another approach to evaluate efficiency of the published Ovid MEDLINE search strategies, was to calculate the average number of search terms used in each strategy. Search terms were counted by copying and pasting each search strategy into a word document. Boolean operators and numbers were removed from the strategy, and the number of terms in the word count command in Word was registered in an Excel file. Words in study design filters were also left out.

4.5.4 Translation of Ovid MEDLINE search strategies into PubMed syntax

The 254 of the 400 published Ovid MEDLINE search strategies containing 5 to 50 included studies were translated into PubMed syntax to investigate whether there was a difference in effectiveness and efficiency between the two search systems. The greatest challenge of a translation between the Ovid MEDLINE and PubMed interfaces is that use of proximity operators is not an option in PubMed but is commonly used in Ovid MEDLINE search strategies. In PubMed, two or more terms forming an expression or a phrase must either be put in quotation marks or combined with the Boolean operator AND, both of which will influence the search results. Phrase searching will narrow the search, and the use of AND will broaden the search. Another challenge when searching PubMed is the use of truncation. Truncation in PubMed turns off the automatic mapping function and can result in so many term variants that only the first 600 is searched (176), which can influence the search result. Consequently, in most cases, truncation was avoided and relevant word variants were searched one by one. The published search strategies translated into PubMed syntax were compared to the original Ovid MEDLINE search strategies.

4.6 Selection of search terms and development of simple, optimized search strategies

To avoid or at least reduce bias as much as possible, the conventional approach to search for research studies to include in systematic reviews is comprehensive searching. Comprehensive searches are designed to find as many relevant studies on a topic as possible based on the assumption that the more explicit and meticulous the search strategy, the more likely a systematic review is to find all relevant studies (177). There is, however, little information on how comprehensive search strategies need to be to retrieve relevant research studies to include in systematic reviews. Therefore, the next approach of this study was to shed light on this issue by using an exploratory study design to find the optimal combination of search terms needed to retrieve the included studies of the subsample of 254 Cochrane reviews.

4.6.1 Selection of search terms

The choice of search terms is essential in identifying relevant research studies for systematic reviews (152). When designing search strategies, according to The Cochrane Handbook, in order to be as comprehensive as possible, too many different search concepts should be avoided, but "in order to identify as many relevant records as possible searches should comprise a combination of subject terms selected from the controlled vocabulary or thesaurus ("exploded" where appropriate) with a wide range of free-text terms" (178). The phrase "a wide range" indicates that a high number of search terms should be used for each concept. Little is known, however, about how many synonyms and word variants are needed to retrieve relevant studies to include in a review and where to find them.

The aim of the exploratory study was to contribute empirical evidence to the debate on search term selection in systematic reviews by exploring the optimal combination of search terms in simple search strategies. An exploratory design was used since there is little previous research on the topic, and the goal was to learn more about this phenomenon (179). Simple, optimized search strategies were tested and developed for the reviews using the conventional, manual, iterative method that is common when searching for research studies to include in systematic reviews. By optimized search strategies are meant strategies that will retrieve the included studies of the sample reviews as efficiently as possible, that is, strategies that contain the least number of search terms needed to retrieve the included studies of the sample reviews. By simple are meant strategies without proximity operators and only one field code.

The aim of the simple, optimized search strategies was high recall, but at the same time the number of retrieved records was kept as low as possible, requiring a balancing between recall and precision. This balancing was based on the iterative addition and removal of parts of the search strategy until an optimal strategy was obtained. More specifically, search terms were selected and tested against the included studies of a review in an iterative process until all included studies were found. Different combinations of search terms that were possible candidates according to the research questions, starting with the health problem and intervention words, were combined in order to develop optimized search strategies.

The process described above identified effective and efficient candidate terms for the simple, optimized searches. The method used for selecting the terms is the same method used in conventional searching when relevant references are known and is similar to the "comprehensive pearl growing" method described by Schlosser (92). The main purpose of the exploratory study was to shed light on the actual need for search terms and whether the optimized search terms occurred in the review title or rest of the protocol.

4.6.2 Search term sources

As part of the exploratory study, it was registered whether the optimized search terms occurred in the title of the review, the rest of the review protocol, or merely in the references of the included studies. This was done to investigate whether the title and protocol of the reviews would have been sufficient search term sources for this sample. The titles of Cochrane reviews all follow a similar logic (180), and new titles should fit one of the following three prescribed formats:

1. [Intervention] for [health problem], or

2. [Intervention A] versus [Intervention B/Control] for [health problem], or

3. [Intervention] for [health problem] in [participant group/location]

As can be seen, intervention and health problem should be part of any Cochrane review title and are therefore suitable starting points to select feasible search terms. The review protocols contain background information and inclusion criteria which can also serve as good sources for search terms. The protocol was not accessible at the time of the investigation, but it was assumed that the background, objectives, and methods sections of the reviews had been part of the protocol and were therefore used as “substitutes” for the protocol.

4.6.3 Development of simple search strategies

For each Cochrane review of the subsample, two simpler search strategies were developed, one for Ovid MEDLINE and one for PubMed. All the simple searches used the optimized search terms selected in the exploratory study.

The biggest difference between the simple searches in Ovid MEDLINE and PubMed is that PubMed has a mapping function that broadens the search.

To make the conditions as equal as possible, all searches were limited to the date each sample review was added to PubMed (Create Date).

Ovid MEDLINE

In “Advanced search” in Ovid MEDLINE, simple Boolean searches were performed without proximity operators using only the “multi-purpose” (.mp) field code for all search terms to make the searching simpler. By default, Ovid looks for search terms in the .mp field when no field codes are attached. In a multi-purpose search, the following fields are searched: [mp=title, abstract, original title, name of substance word, subject heading word, key word heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms] (181). The fields searched in the .mp field code may be compared to searching PubMed without field codes.

PubMed

In PubMed, the same simple searches were performed without any field codes. The fields searched when field codes are not specified in PubMed are MeSH and All Fields. “All Fields” in PubMed means “Untagged terms and terms tagged with [all fields] are processed using Automatic Term Mapping. Terms that do not map are searched in all search fields except for Place of Publication, Transliterated Title, Create Date, Completion Date, Entrez Date, MeSH Date, and Modification Date.” (182). AND is the default Boolean operator between search terms in partial match mode in PubMed and can therefore be left out in a search strategy, but OR was used to combine synonyms. The main concepts represented by different synonyms were enclosed by parentheses. Truncation was used in a few cases to improve recall of a search.

4.6.4 Filter for study design

Filter for study design was also used in the simple search strategies to increase efficiency. In this exploratory study, a filter for RCTs was used in 174 of the 254 simple search strategies. A filter for study design was only used when there was a filter for RCTs in the original search strategy. No filter was used if a review included different kinds of study designs or study designs other than RCTs. The filters were not developed and adjusted for each search strategy, as the subject terms were, but were standardized. Two different filters were used, one “narrow” and one “broad”, depending on how they performed, that is, the narrow was tried out first and used if all included studies were found. If not, the broad filter was used. Both are slight variants of the Cochrane Highly Sensitive Search Strategy for identifying randomized trials in MEDLINE: sensitivity-maximizing version (2008 revision), PubMed format (183):

(randomized controlled trial [pt] OR controlled clinical trial [pt] OR randomized [tiab] OR placebo [tiab] OR drug therapy [sh] OR randomly [tiab] OR trial [tiab] OR groups [tiab]) NOT (animals [mh] NOT humans[mh]))

The narrow search filters used in this study were as follows:

<u>PubMed:</u>

((randomized controlled trial[pt] or controlled clinical trial[pt] or randomized[tiab] or randomised[tiab] or placebo[tiab] or randomly[tiab] or “clinical trials as topic”[mh]) not ("Animals"[mh] not ("Animals"[mh] and "Humans"[mh])))

<u>Ovid:</u>

((randomized controlled trial or controlled clinical trial).pt. or randomized.tw. or randomised.tw. or placebo.tw. or randomly.tw. or "clinical trials as topic"/) not ("Animals"/ not ("Animals"/ and "Humans"/))

The broad search filters used in this study were as follows:

<u>PubMed:</u>

((randomized controlled trial[pt] or controlled clinical trial[pt] or randomized[tiab] or randomised[tiab] or placebo[tiab] or drug therapy[sh] or randomly[tiab] or trial[tiab] or groups[tiab] or “clinical trials as topic”[mh]) not ("Animals"[mh] not ("Animals"[mh] and "Humans"[mh])))

<u>Ovid:</u>

((randomized controlled trial or controlled clinical trial).pt. or randomized.tw. or randomised.tw. or placebo.tw. or drug therapy.fs. or randomly.tw. or trial.tw. or groups.tw. or "clinical trials as topic"/) not ("Animals"/ not ("Animals"/ and "Humans"/))

Below follows an example of a simple, optimized search strategy for the Cochrane review “Aromatherapy for dementia” (175) with narrow filter for study design, PubMed syntax:

(dementia or Alzheimer) aromatherapy ((randomized controlled trial[pt] or controlled clinical trial[pt] or randomized[tiab] or randomised[tiab] or placebo[tiab] or randomly[tiab] or “clinical trials as topic”[mh]) not ("Animals"[mh] not ("Animals"[mh] and "Humans"[mh])))

4.6.5 Review complexity

The subsample of 254 Cochrane reviews of this study were categorized into complex and non-complex to investigate whether the results differed depending on review complexity. The Cochrane Handbook makes a distinction between “non-complex interventions such as drugs, physical interventions, routes of delivery, doses, timing or length of delivery” and “complex interventions, such as those that evaluate psychotherapy, behavioural and educational topics or healthcare delivery strategies” (184). The sample reviews were grouped according to these two main categories. Diagnostic reviews and reviews about dietary supplements, for example vitamins, were treated as non-complex interventions, and reviews about rehabilitation and screening as complex. Of the 254 reviews, 16 were diagnostic reviews. The categorization was done by two authors (HS and MM). The reviews were categorized by both authors, independent of each other. The authors then met and reviewed every selection. Discrepancies were dissolved by consensus.

Reviews on drugs are one category of non-complex reviews and assumed easier to retrieve since the language used is more standardized. The sample reviews were therefore also categorized as drug-related or not drug-related to investigate whether this was true for the sample of this study. Vitamins and food supplements were categorized as drug-related. Reviews on several interventions, including pharmacological, were treated as not drug-related. Drug categorization was performed by only one author (MM).

The different categories were evaluated for effectiveness and efficiency using precision, recall, NNR, and number of search terms.

The reason for categorizing the reviews according to complexity was that most likely, research evidence for non-complex reviews is easier to retrieve than evidence for complex reviews, but little research has been done in this area.

4.7 Use of an automated search method – extended Boolean retrieval

The need to automate the whole or part of the systematic review process has been discussed for several years (3, 160, 163, 185), also the information retrieval stage (67, 68, 161). As was seen in Scoping review III, several attempts were made to develop more automated search systems for systematic reviews. None of these techniques or systems, however, is open for public use yet, or they have been withdrawn. One study of the included studies, however, found that extended Boolean information retrieval performed best of several ranked approaches for information retrieval in systematic reviews (61). Extended Boolean retrieval means that the search results of conventional Boolean search strategies are ranked to get the most relevant hits first and was tested in the empirical part of this study for conventional Boolean searches in PubMed. The “Best Match” option under the “Sort by” function in PubMed was used as ranking function. The ranking numbers of the included studies were recorded in an Excel spread sheet, and MAP was used to measure performance. To investigate whether ranking the search results would move the most relevant hits, that is, the included studies, towards the top of the result lists, cut-off values were set at 500, 1000, 1500 and 2000 for the conventional Boolean search strategies. The ranked results were compared to the sort options “First Author” and “Most Recent”.

4.8 Summary

The main objective of this study was to investigate simpler, more optimized search approaches to retrieve research evidence for systematic reviews in healthcare than the present comprehensive approaches. In this chapter, sample considerations and the different sub-studies were described. Research methods used were cross-sectional, exploratory and experimental design. For the cross-sectional study, the newest 400 systematic reviews representing 16 different Cochrane Review Groups were used as sample. For the exploratory and experimental designs, a subsample of 254 reviews containing from 5 to 50 included studies was used. To investigate automated information retrieval for systematic reviews, the relevance ranking of PubMed was tested on the subsample. Scoping reviews were conducted to map previous research on the main approaches and to inform the empirical studies.

5 Results

5.1 Searching only MEDLINE and Embase for study identification

To investigate whether searching only MEDLINE and Embase is enough for study identification in systematic reviews, both database coverage and search performance of MEDLINE search strategies were evaluated.

5.1.1 Coverage of MEDLINE and Embase

The first approach of the cross sectional study was to evaluate coverage of MEDLINE alone or MEDLINE and Embase combined of primary studies included in a sample of 400 Cochrane reviews and a subsample of 254 containing 5 to 50 included studies.

The reasons references were not found in MEDLINE or Embase were typically that they were conference abstracts, theses, dissertations, personal communications, or published in supplements or journals not indexed in the two databases.

5.1.1.1 MEDLINE alone

Of a total of 6222 included studies, 5223 were indexed in MEDLINE across the 400 systematic reviews which gives a coverage of 84 %. The average number of included studies indexed in MEDLINE per review across all groups was 13.1 (median=8.0) (total average=15.6 (median=9.5)). Thus, an average of 2.5 (median=1.0) studies per review was not indexed in MEDLINE. In 158 of the 400 reviews, all included studies were indexed in MEDLINE, and in 94, only one study was not indexed in MEDLINE.

Coverage of included studies in MEDLINE in the sample of 400 is shown in Figure 1 below.

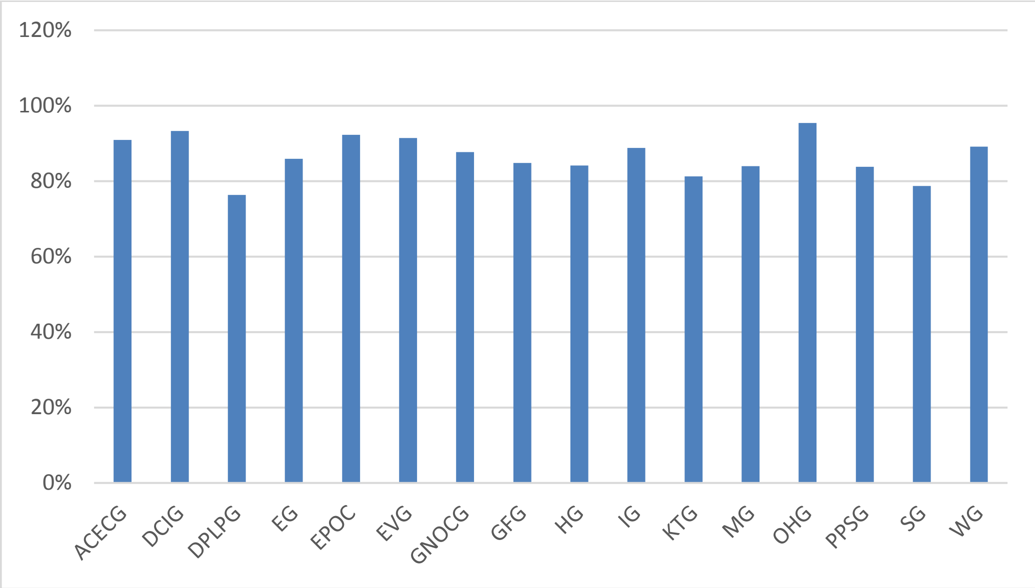


Figure 1 Coverage in MEDLINE of included studies in a sample of 400 Cochrane reviews

Group differences

As can be seen from Figure 1, the average number of included studies indexed in MEDLINE differed between the review groups. Effective Practice and Organization of Care Group (EPOC) had the highest number with 93 % (433 of 464) and Developmental, Psychosocial and Learning Problems Group (DPLPG) the least with 67 % (355 of 533). The highest number of included studies not indexed in MEDLINE for a single review was 76 of 102 in the review “Kinship care for the safety, permanency, and well-being of children removed from the home for maltreatment” in DPLPG (186). Average number of MEDLINE-indexed studies per review varied between the groups from 5.9 of 7.2 in the Epilepsy group (EG) to 21.6 of 25.0 in the Heart Group (HG). EPOC had a total average of 18.6 included studies per review with 17.3 of them indexed in MEDLINE.

Subsample

Of a total of 4722 included studies, 3999 were indexed in MEDLINE across the subsample of 254 systematic reviews which gives a coverage of 85 %. The average number of included studies indexed in MEDLINE per review was 15.7 (median=12.0) of 18.6 (median=13.0). Thus, an average of 2.9 (median=1.0) studies per review was not indexed in MEDLINE. In 73 of the 254 reviews all included studies were indexed in MEDLINE, and in 67 reviews only one study was not indexed in MEDLINE.

5.1.1.2 MEDLINE and Embase combined

In 158 reviews of the sample of 400, all included studies were indexed in MEDLINE. Thus, 242 reviews were investigated for included studies indexed in Embase. A total number of 332 included studies in the 242 reviews were indexed in Embase and not in MEDLINE with an average of 1.4 (median=1.0) per review. Of the total of 6222 included studies, 5571 were indexed in MEDLINE and Embase combined. This is 90 % and gives an average of 13.9 (median=8.0) studies per review (total average=15.6 (median=9.5)). An average of 1.6 (median=0.0) of the included studies was neither indexed in MEDLINE nor Embase. In 221 reviews, all studies were indexed in MEDLINE and Embase combined, and in 79 reviews, only one study was not indexed in MEDLINE or Embase. Thus, in 100 of the 400 reviews more than one study was not indexed in any of the two databases.

Group differences

The number of included studies indexed in Embase and not in MEDLINE varied between the groups with the Effective Practice and Organization of Care Group (EPOC) having the lowest amount. Only 4 of 464 included studies in EPOC were indexed in Embase and not in MEDLINE. The Stroke Group (SG) had the highest number of included studies indexed in Embase alone with 54 of 403, which amounts to 2.8 per review. In no group were all included studies indexed in the two databases, but in 6 of the 16 groups, 94 % or more was indexed in MEDLINE and Embase combined.

Average number of included studies per review, included studies indexed in MEDLINE, and included studies indexed in MEDLINE and Embase combined is shown in Figure 2 below.

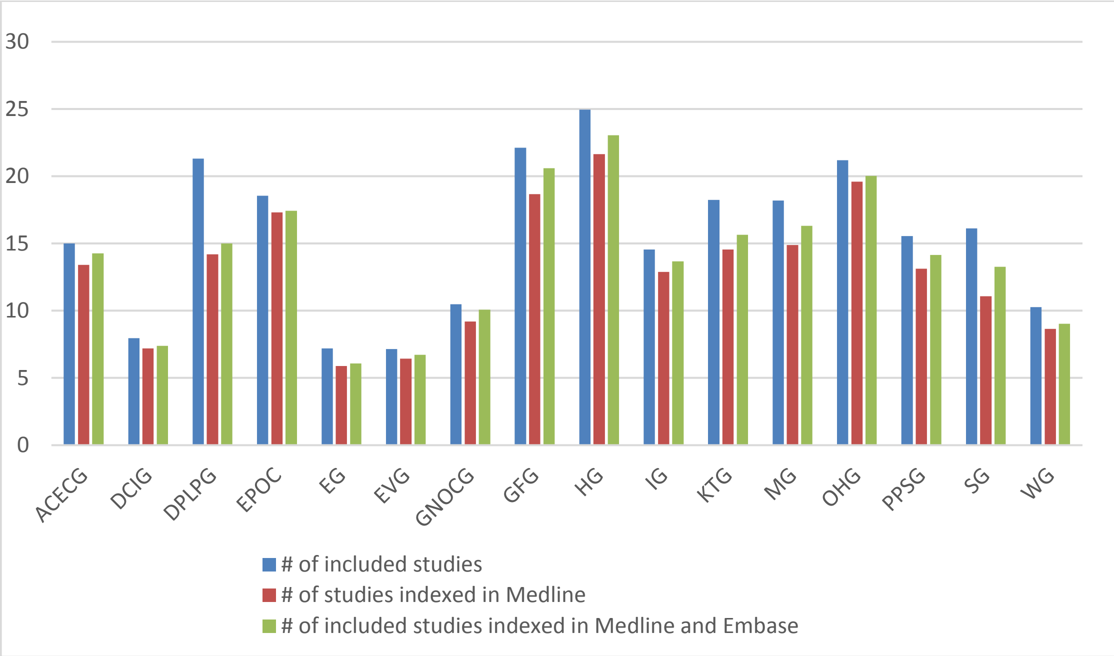


Figure 2 Average number of included studies per review, included studies indexed in MEDLINE only and included studies indexed in MEDLINE and Embase combined

Subsample

In 73 reviews in the subsample of 254 (containing 5 to 50 included studies), all included studies were indexed in MEDLINE. Thus, 181 reviews of this sample were investigated for included studies indexed in Embase. A total number of 247 of the 4722 included studies were indexed in Embase and not in MEDLINE with an average of 1.4 (median=1.0) per review. Of a total of 4722 included studies, 4250 were indexed in MEDLINE and Embase combined. Coverage of both databases for all reviews was 90 %. Average number of included studies per review was 16.7 (median=13.0). An average of 1.8 (median=1.0) of the included studies was neither indexed in MEDLINE nor Embase. In 125 reviews, all studies were indexed in MEDLINE and Embase combined, and in 56 reviews only one study was not indexed in any of the two databases.

5.1.2 Number of electronic databases searched

The number of electronic databases searched for each review ranged from 2 to 25 with an average of 7.8 (median=7.0) per review in the sample of 400 reviews. Average number of databases searched for each review group is shown in Figure 3 below.

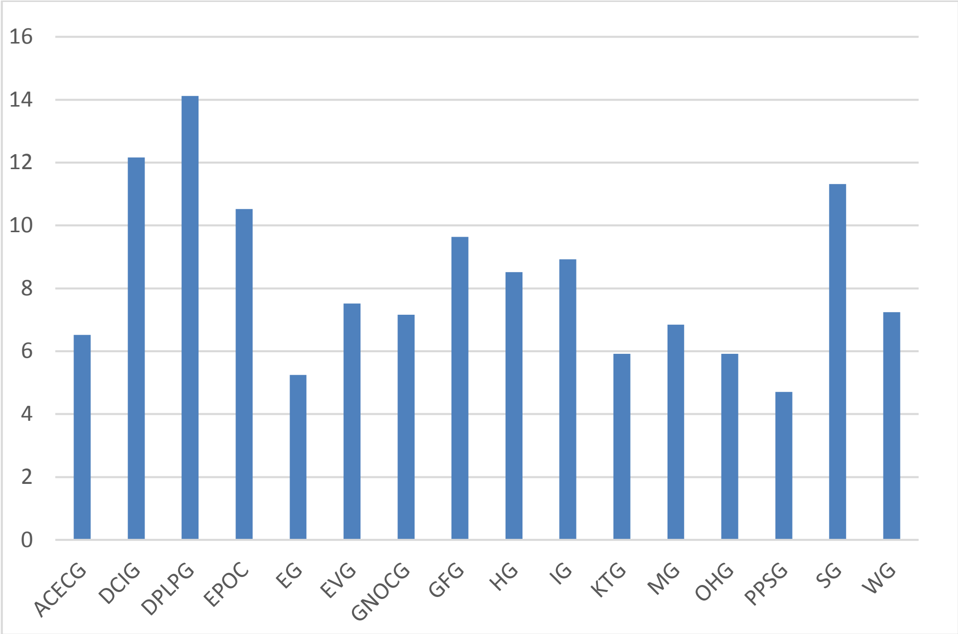


Figure 3 Average number of databases searched for each review group (sample=400)

Group differences

As can be seen from Figure 3, there were great differences in the average number of electronic databases searched in the 16 groups with Developmental, Psychosocial and Learning Problems Group (DPLPG) searching the most (=14) and Pain, Palliative and Supportive Group (PPSG) searching the least (=5).

5.1.3 Performance of published searches - Ovid MEDLINE

The next approach of the cross sectional study was to measure performance of the published Ovid MEDLINE search strategies of the sample reviews.

There was a total of 940 183 retrieved records in the 400 Ovid MEDLINE search strategies with an average of 2350 (median=801) per review.

The average number of included MEDLINE-indexed studies per review found by the published Ovid MEDLINE search strategies was 11.3 (median=7.0) of 13.1 (median=8.0).

Average number of included studies indexed in MEDLINE and average number of included studies found by the published Ovid MEDLINE search strategies for each review group is shown in Figure 4 below.

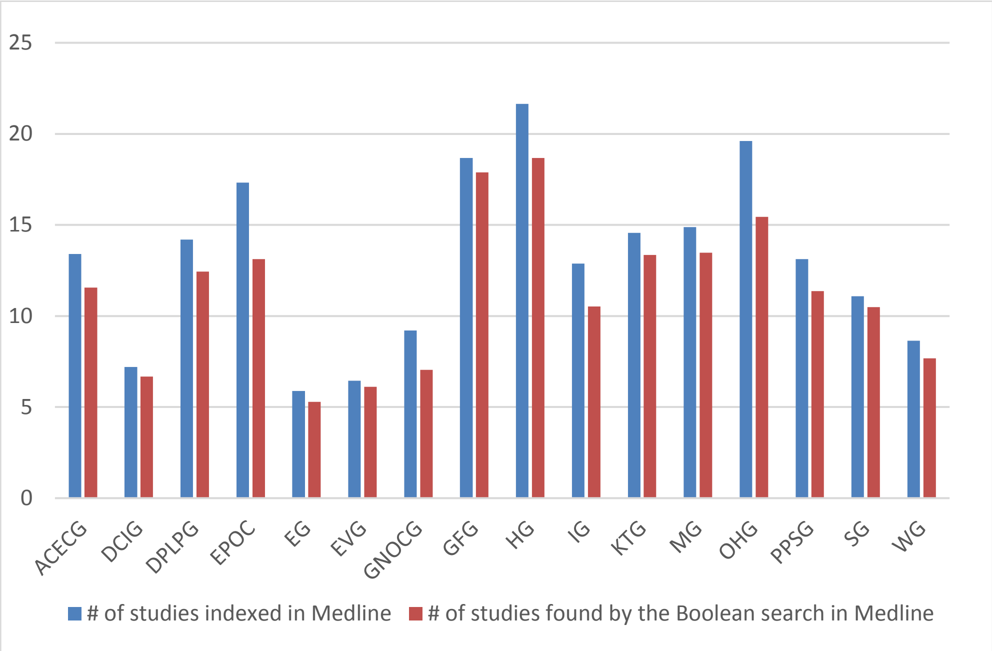


Figure 4 Average number of included studies indexed in MEDLINE and found by the published Boolean search strategies

As can be seen from Figure 4, in many of the groups the published search strategies found almost all MEDLINE-indexed included studies. For 228 reviews, all included studies were found by the search strategies.

5.1.3.1 Effect of using filter for study design

Of the 400 published Ovid MEDLINE search strategies, 297 used filter for study design. In 229 of these, none of the included studies was lost, and in 39 reviews, only one was lost. That leaves 29 reviews with more than one study lost when using filter for study design. The maximum loss for one search strategy was 13 out of 33.

The 297 search strategies with filter for study design had a total of 503 381 retrieved records with an average of 1695 (median=726) per review. Without filter for study design the total sum was 2 379 656 retrieved records with an average of 8012 (median=3485) per review.

The average number of retrieved records per review in the published Ovid MEDLINE search strategies for each review group with and without filter for study design is shown in Figure 5 below.

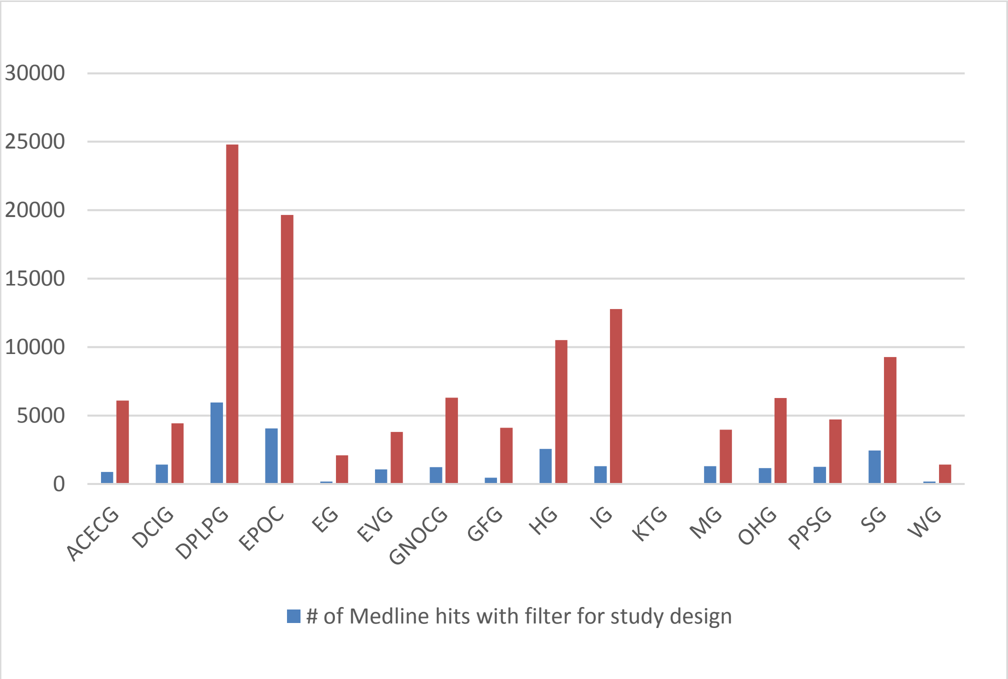


Figure 5 Average number of hits in published Ovid MEDLINE search strategies for each review group with and without filter for study design

Filter for study design was not used by the Kidney and Transplant Group (KTG). Gynaecology and Fertility Group (GFG) was the only group that used filter in all reviews. Anaesthesia, Critical And Emergency Care Group (ACECG) had only one review without filter.

In the 297 reviews using filter for study design, the average number of studies found by the published search strategies was 10.5 (median=6.0) with filter for study design and 11.1 (median=7.0) without filter of the 13.1 (median=8.0) indexed in MEDLINE. Average number of included studies found per review by the published Ovid MEDLINE search strategies with and without filter for study design for each review group is shown in Figure 6 below.

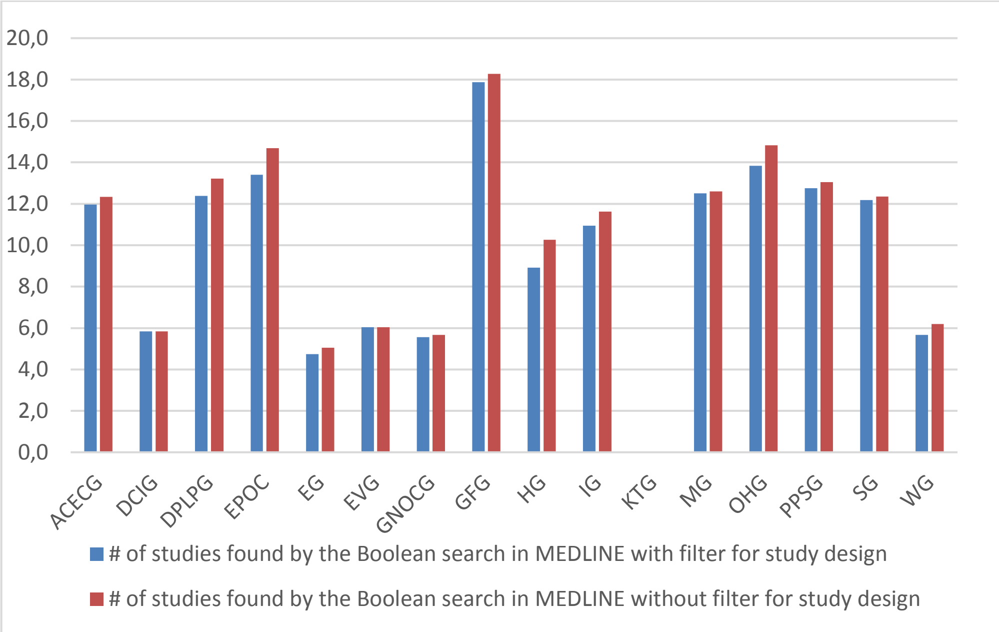


Figure 6 Average number of included studies per review found by the published Boolean search strategies in Ovid MEDLINE with and without filter for study design

As can be seen from Figure 6, the number of included studies that was not retrieved in the reviews using filter for study design, was very small.

5.1.3.2 Recall and precision

Recall was calculated as number of included studies indexed in MEDLINE that was found by the published Ovid MEDLINE search strategies. Average recall for all groups was 88 % (median=100 %), and average precision was 2.8 % (median=0.8 %) in the sample of 400 reviews.

Precision and recall for the published Ovid MEDLINE search strategies are shown in Figure 7 below.

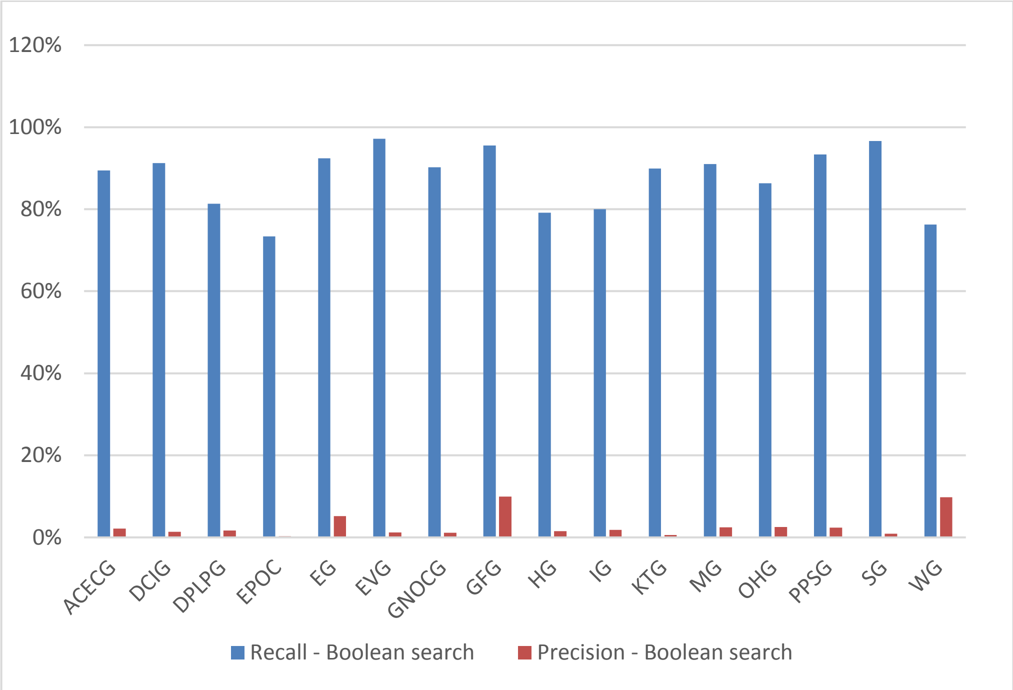


Figure 7 Recall and precision for published Boolean search strategies in Ovid MEDLINE

Group differences

The group with the highest performance was Gynaecology and Fertility Group (GFG) with a recall of 96 % and precision of 9.9 %. Wounds Group (WG) also had a high precision with 9.8 % but at the expense of recall, which was 76 %. Removing GFG and WG reduced average precision to 1.8 % for the rest of the groups. Removing Effective Practice and Organization of Care Group (EPOC) and Kidney and Transplant Group (KTG), the two groups with the lowest precision values, on the other hand, increased average precision for the rest of the groups to 3.1 %.

Search performance in subsample and comparison between the two samples

For the sample of 254 reviews, average recall for the published Ovid MEDLINE searches for all groups was 87 % (median=94 %). Average precision for all groups was 2.8 % (median=1.1 %). For the sample of 400, average recall for all groups was 88 % (median=100 %), and average precision was 2.8 % (median=0.8 %). Thus, average precision and recall values across all groups were not much influenced by the number of included studies in each review for this sample. The average values of some of the groups were highly influenced by the number of included studies, though, depending on how many and which reviews that were lost in each group in the subsample. Wounds Group (WG), for example, had an increase in recall from 76 % in the full sample to 93 % in the subsample. Four reviews in the full sample of this group had a recall of 0 %, because they had only one or two included studies, and none of them was retrieved by the published search strategy.

Recall and precision for the two samples are shown in Figure 8 and Figure 9 below.

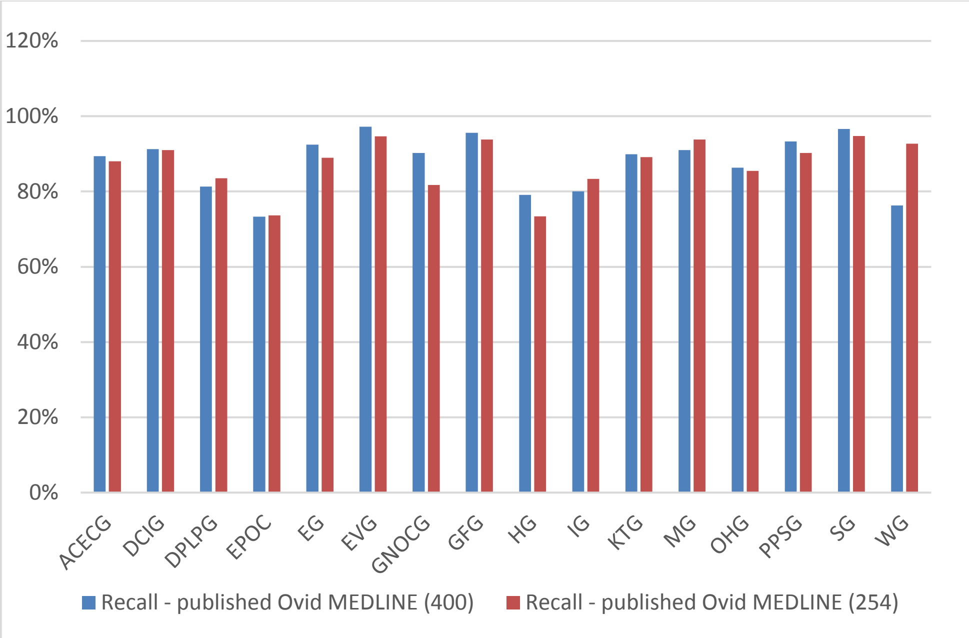


Figure 8 Recall for total (=400) and reduced sample (=254) in the published Ovid MEDLINE search strategies

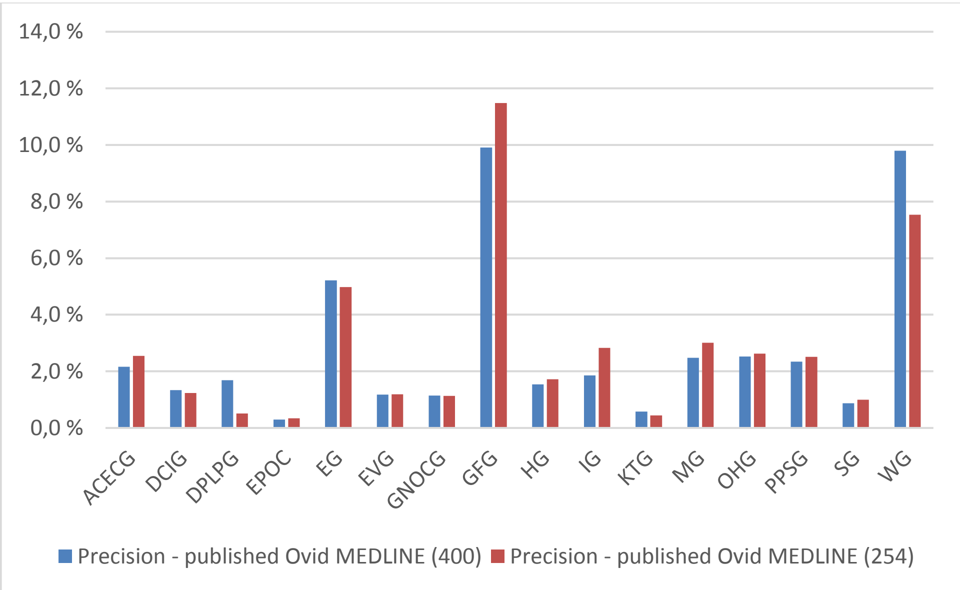


Figure 9 Precision for total (=400) and reduced sample (=254) in the published Ovid MEDLINE search strategies

5.1.3.3 NNR

Average NNR for each review in all groups was 426 (median=121) in the sample of 400.

For the sample of 254 reviews, average NNR was 249 (median=95) per review.

Average NNR for the published Ovid MEDLINE searches per review for each review group in both samples is shown in Figure 10 below.

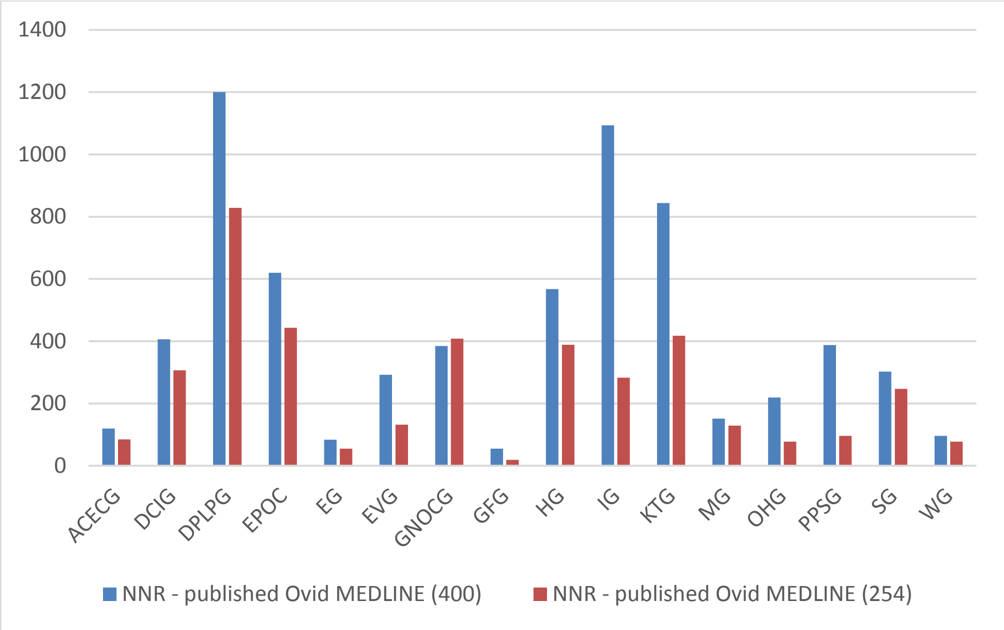


Figure 10 Average NNR per review in Ovid MEDLINE search strategies for each group for total sample (=400) and subsample (=254)

Group differences

As can be seen from Figure 10, NNR varied considerably between the groups and between full and reduced sample. Developmental, Psychosocial and Learning Problems Group (DPLPG) had the highest NNR for both samples with 1201 (median=317) for the sample of 400 and 828 (median=345) for the sample of 254. This was mostly due to two reviews with extremely high numbers of hits (15 368 and 31 038). Kidney and Transplant Group (KTG) also had a high average NNR. This was probably because the group did not use filter for study design for any of the sample reviews. Gynaecology and Fertility Group (GFG) had the lowest NNR with only 19 (median=12) references to read to find one of the included studies in the subsample and 55 (median=14) in the full sample.

5.1.4 Performance of published searches - PubMed

The Ovid MEDLINE search strategies of the subsample of 254 reviews including from 5 to 50 primary studies was translated into PubMed syntax to compare the two MEDLINE interfaces.

For the translated PubMed search strategies, there was a total of 803 761 retrieved records across the 254 reviews with an average of 3215 (median=1082) per review.

5.1.4.1 Recall and precision

Average recall for conventional search strategies in PubMed for all groups was 87 % (median=95 %) and average precision 2.7 % (median=1.0 %). For comparison, recall for the published Ovid MEDLINE search strategies was 87 % (94 %) and precision 2.8 % (median=1.1 %). Of the 254 Ovid MEDLINE strategies translated into PubMed syntax, 118 had 100 % recall compared to 115 for the published Ovid MEDLINE search strategies. The effectiveness of the searches performed in the two databases was thus approximately the same.

5.1.4.2 NNR

Average NNR for conventional search strategies in PubMed was 307 (median=96) as compared to 249 (median=95) for the same search strategies performed in Ovid MEDLINE. The efficiency of conventional search strategies was, thus, slightly lower in PubMed than in Ovid MEDLINE.

5.1.5 Number of search terms in published search strategies

Number of search terms was only counted for the MEDLINE search strategies in the subsample (=254), and the average number was 72 (median=53) search terms per strategy for the published Ovid MEDLINE search strategies. For the same strategies translated into PubMed syntax, the average number of search terms was 83 (median=57) for all groups.

Number of search terms per review in published Ovid MEDLINE search strategies is shown in Figure 11 below.

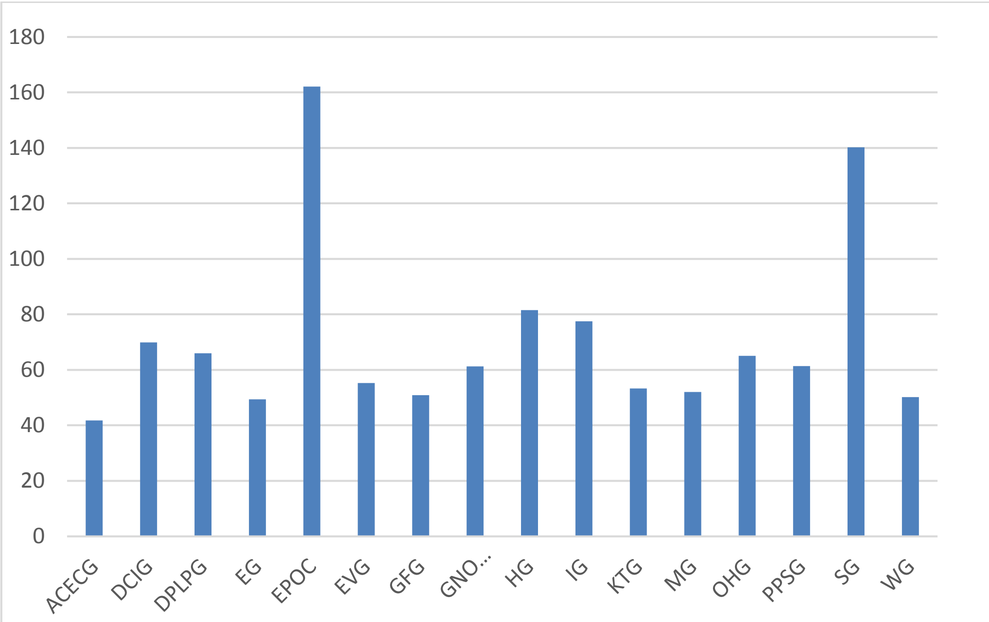


Figure 11 Average number of search terms in published Ovid MEDLINE search strategies

Group differences

As can be seen from Figure 11, two of the groups used considerably more search terms than the other groups and that was Effective Practice and Organization of Care Group (EPOC) and Stroke Group (SG) with an average of 163 and 140 per search strategy in the Ovid MEDLINE search strategies, respectively. Leaving EPOC and SG out, the average number of search terms was reduced to 50 (median=45) for the rest of the groups.

For further information on search results of the cross sectional study, see Appendix III.

5.2 Selection of search terms and development of simple, optimized search strategies

One approach of this study to simplify and optimize the search process for systematic reviews was to investigate how many search terms were necessary to retrieve the included studies in the subsample of 254 Cochrane reviews. This was done by manually testing feasible search terms one by one until an optimal search strategy was obtained in an attempt to assess the actual need for search terms and investigate where to find the most useful terms.

5.2.1 Selection of search terms

In the sections below, there is first a description and some examples of the search term selection process and how the optimized search terms were allocated according to PICO. Next, the number of necessary search terms were counted and compared to the number used in the published Ovid MEDLINE search strategies. Finally, it was investigated whether the optimal search terms could be found in the review title or rest of the review protocol.

5.2.1.1 Allocation of search terms

For most of the simple search strategies, the optimized search terms were allocated to the two main PICO groups: health problem (P) and intervention (I) as recommended by The Cochrane Handbook (187). For some strategies, however, search terms for only one of the groups were enough. In a few reviews, search terms for outcome (O) were also used. There were also some terms called “questionable” that did not belong to the PICO framework. Below follows examples of different kinds of simple, optimized search strategies that were the results of the exploratory study:

Example I (one P-word and one I-word):

A review fitting the first example was “Allopurinol for chronic gout” (188). When testing suitable search terms for this review, a combination of the intervention word “allopurinol” and the health problem word “gout” was tried first. It turned out that the combination of these two words was enough to find the 10 included studies, and thus, the search strategy was complete, and no more testing was necessary. For comparison, the published Ovid MEDLINE search strategy comprised 166 search terms. NNR was 124 for the published search and 15 for the optimized.

Simple search strategies for the review “Allopurinol for chronic gout”:

<u>Ovid MEDLINE:</u>

(allopurinol and gout).mp

<u>PubMed:</u>

allopurinol gout

Example II (more search terms needed)

When one word for patient group (P) and one for the intervention (I) were not enough, synonyms for these two word groups had to be searched.

For several reviews, one search term was not enough for P, and one example was the review "Aromatherapy for dementia" (175), where "aromatherapy" was the intervention and "dementia" the health problem. A search for "aromatherapy" found all the included studies indexed in MEDLINE, so no more search terms were needed to retrieve the six included studies for the intervention part of the search strategy. The term "dementia", however, retrieved only five of the six included studies. To retrieve the last study, "Alzheimer" had to be searched as well. Searching "Alzheimer" alone would only have retrieved two of the six studies. Thus, both "dementia" and "Alzheimer" had to be searched to find the included studies in this review and resulted in the following simple, optimized search strategies:

Ovid MEDLINE:
((dementia or Alzheimer) and aromatherapy).mp
PubMed:
(dementia or Alzheimer) aromatherapy
For comparison, the published Ovid MEDLINE search strategy comprised 69 search terms.

Example III (only P-word(s) or I-word(s)):

For some reviews, if it was sufficient, only health problem or intervention terms were used. For the review "Blood pressure control for diabetic retinopathy" (189), for example, only the health problem terms (diabetic retinopathy) were necessary to find the included studies in an effective way. In the review "Selenium supplementation for critically ill adults" (190), on the other hand, only the intervention terms (selenium OR ebselen) were necessary to find the included studies effectively. Simple search strategies for the reviews "Blood pressure control for diabetic retinopathy" and "Selenium supplementation for critically ill adults":

Ovid MEDLINE:

(diabetic and retinopathy).mp

(selenium or ebselen).mp

PubMed:

diabetic retinopathy

selenium or ebselen

Example IV (“questionable terms” added):

There was also a fourth group, “questionable terms” (51), that did not directly represent the health condition or intervention. These search terms were used when terms for health condition and/or intervention did not identify the included studies in an efficient way, for example for the review “Drug therapy for treating post-dural puncture headache” (191) from Pain, Palliative and Supportive Group (PPSG) the following search strategy was used:

<u>Ovid MEDLINE:</u>
(puncture and headache and (post or postdural) and therapy).mp
<u>PubMed:</u>
puncture headache (post or postdural) therapy

The term “therapy” was not necessary to retrieve the included studies. It was only used to balance the search strategy to keep the number of hits down.

Example V (several search terms necessary):

Not all search strategies were as simple as the previous examples. Some of them required several search terms to retrieve all included studies. For the review “Food supplementation for improving the physical and psychosocial health of socio-economically disadvantaged children aged three months to five years” (192) from the review group Developmental, Psychosocial and Learning Problems Group (DPLPG), for example, the following search strategy was used to retrieve the 29 included studies:

<u>Ovid MEDLINE:</u>

((nutrition or nutritional or nutrient or (fortified and food)) and (diet or dietary or supplement* or meals or (energy and intake)) and (infant or child) and (malnutrition or undernourished or malnourished or height or weight or (iron and deficiency)) and (intervention or interventions or program or programme or evaluation or surveys or study)).mp

<u>PubMed:</u>

(nutrition or nutritional or nutrient or (fortified and food)) (diet or dietary or supplement* or meals or (energy and intake)) (infant or child) (malnutrition or undernourished or malnourished or height or weight or (iron deficiency)) (intervention or interventions or program or programme or evaluation or surveys or study)

This example shows that to develop simple, optimized search strategies is quite challenging for broad and complex research questions and when the language of a topic is not standardized. Such strategies, however, were not representative of the typical optimized search strategy in this study.

5.2.1.2 Number of search terms

An average of 7.2 (median=6.0) search terms was necessary to retrieve the included studies of the reviews in each simple, optimized search strategy. The average number of necessary search terms in each review group is shown in Figure 12 below.

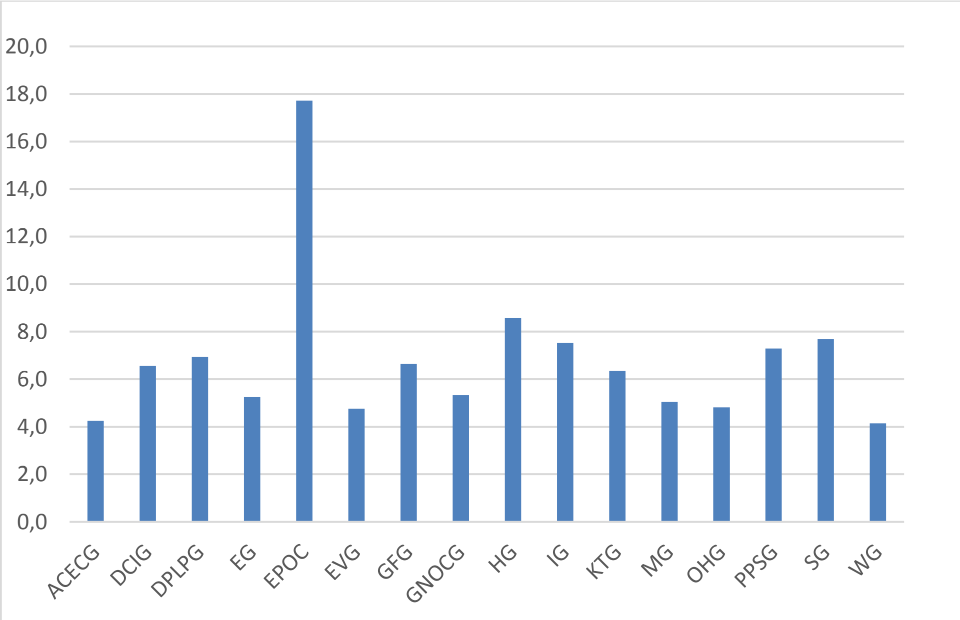


Figure 12 Average number of search terms in simple, optimized Ovid MEDLINE search strategies

As can be seen from Figure 12, Effective Practice and Organization of Care Group (EPOC) reviews needed at lot more search terms than any of the other groups. For the published Ovid MEDLINE search strategies the average number of search terms was 72 (median=53) per review. Thus, the average number of search terms necessary to find the included studies of the reviews was approximately one tenth of the number used in the published Boolean searches. The need for search terms was consistently low for almost all reviews, but the number of search terms used in the published searches varied a great deal. Average number of search terms per review per group for the simple, optimized search strategies and the published search strategies are shown in Figure 13 below.

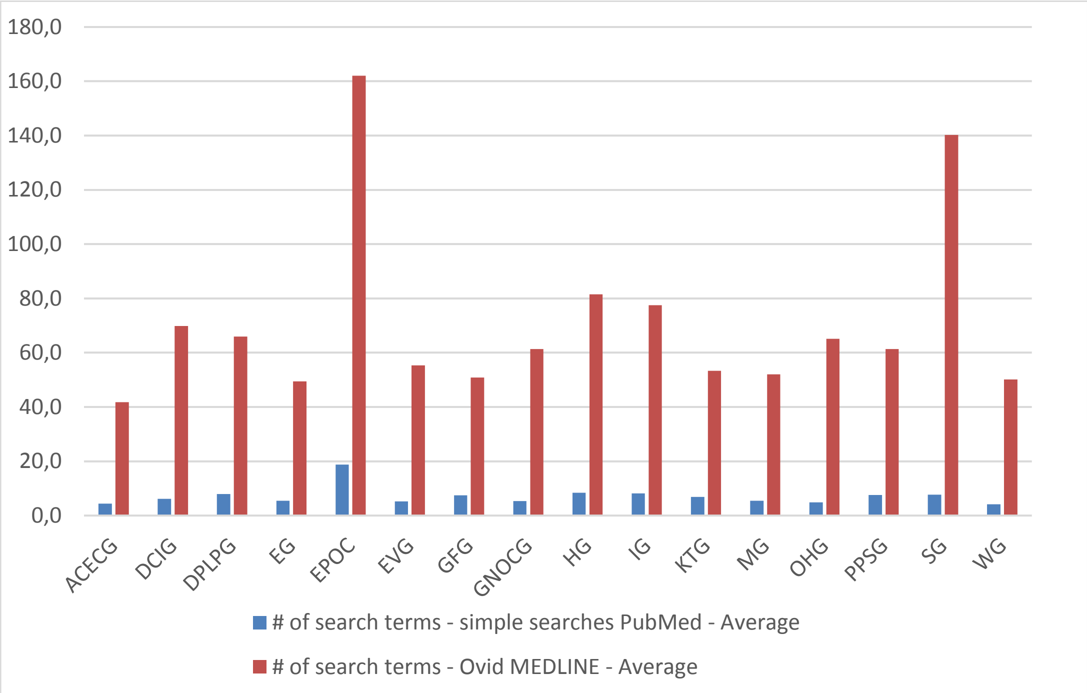


Figure 13 Average number of search terms for each group in published and simple, optimized search strategies

As can be seen from the figure above, it was possible to reduce the number of search terms considerably in all groups. There were, however, great differences between the groups.

Group differences

Considerably more search terms were used for the Effective Practice and Organization of Care Group (EPOC) and the Stroke Group (SG) reviews in the published search strategies than the other review groups. The average number of search terms per review was 163 for EPOC and 140 for SG. The average need for search terms, however, was much higher for EPOC (=17.7) than for the other review groups, but not for SG (=7.7).

5.2.1.3 Drug-related reviews

Of the 254 reviews, 113 were categorized as drug-related.

5.2.1.4 Search term sources

Most of the optimized search terms could be found in the title and protocol of the sample reviews. Across all review groups, an average of 3.0 (median=3.0) search terms per review of the total average of 7.2 occurred in the titles of the reviews with a range from 0.0 to 7.0. Another average of 2.8 (median=2.0) search terms per review per group could be found in the rest of the protocol with a range from 0 to 19. An average of 1.4 (median=0.0) search terms per review was not found in the title or rest of the review protocol with a range from 0 to 56.

The average number of search terms occurring in other sources than review title and protocol in all sample reviews was 22.9 per group. Leaving EPOC out, reduced the number to 14.3.

Number of search terms occurring in review title, rest of the protocol, and included studies is shown in Figure 14 below.

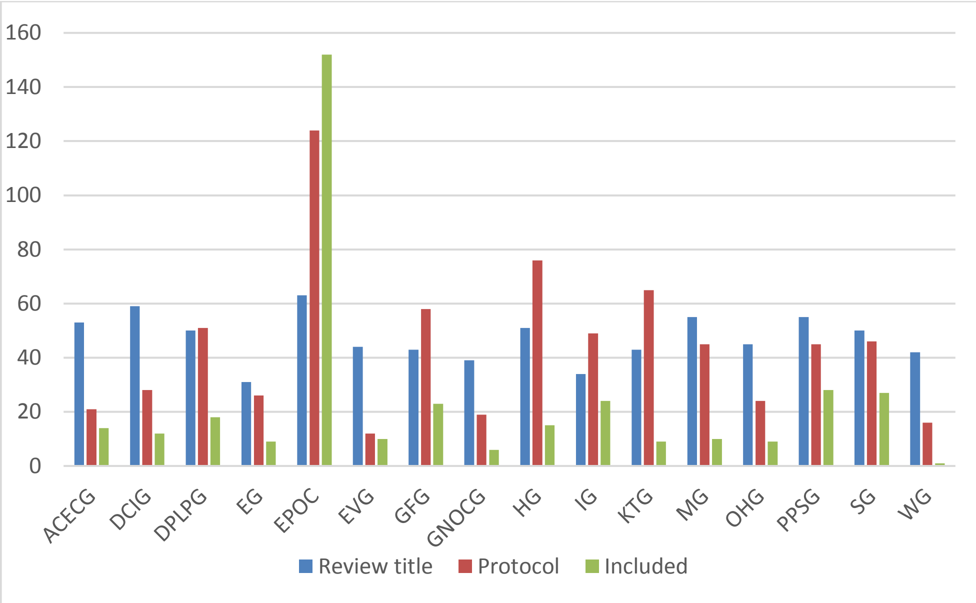


Figure 14 Average number of search terms in simple, optimized search strategies found in review title, protocol and included studies

For seven groups[3], more than half of the search terms could be found in the review title on average. In 54 reviews, all optimized search terms occurred in the review title. Of these, 36 were on drug-related questions.

In the review "Oral or parenteral iron supplementation to reduce deferral, iron deficiency and/or anaemia in blood donors" from Injuries Group (IG) (193), the following search terms were sufficient to retrieve the 26 included studies: "iron AND blood AND donors", and all three could be found in the review title. The published search strategy did also retrieve all 26 studies but contained 140 search terms. NNR was 145 for the published search strategy and 24 for the simple, optimized search strategy for this review.

[3] 1. Anaesthesia, Critical and Emergency Care Group (ACECG), 2. Dementia and Cognitive Impairment Group (DCIG), 3. Eyes and Vision Group (EVG), 4. Gynaecological, Neuro-oncology and Orphan Cancer Group (GNOCG), 5. Musculoskeletal Group (MG), 6. Oral Health Group (OHG) and 7. Wounds Group (WG)

5.2.2 Performance of simple, optimized search strategies

5.2.2.1 Ovid MEDLINE

For the simple searches in Ovid MEDLINE, there was a total of 233 495 retrieved records across the 254 reviews with an average of 916 (median=396) per review.

Average number of included MEDLINE-indexed studies found by the simple searches in Ovid was 14.9 (median=12.0) of the total of 15.8 (median=12.0) per review.

Effectiveness

Average precision across all groups was 5.6 % (median=3.1 %), and average recall was 94 % (median=100 %).

Average precision and recall for each group for the simple searches performed in Ovid MEDLINE is shown in Figure 15 below.

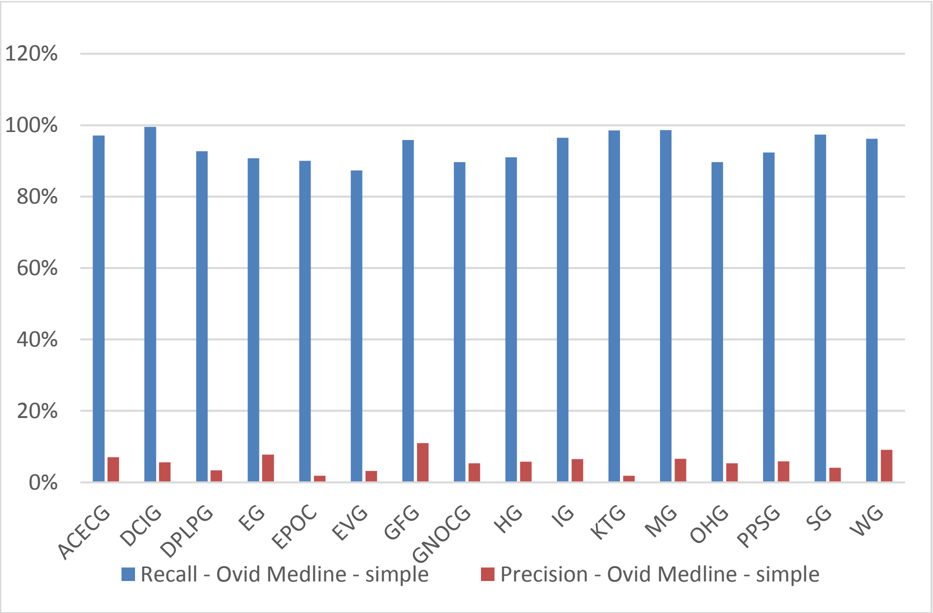


Figure 15 Precision and recall for simple, optimized search strategies in Ovid MEDLINE

Efficiency

Average NNR for each review in the simple Ovid MEDLINE searches was 66 (median=32).

5.2.2.2 PubMed

For the simple search strategies performed in PubMed, there was a total of 322 079 retrieved records across the 254 reviews with an average of 1268 (median=588) per review.

Average number of included studies found by the simple searches in PubMed was 15.3 (median=12.0) of the total number of included studies indexed in MEDLINE of 15.8 (median=12.0) per review.

Effectiveness

Average precision across all groups was 4.6 % (median=2.3 %) for the simple searches in PubMed, and average recall was 98 % (median=100 %).

Efficiency

Average NNR for each review in the simple PubMed searches was 88 (median=44).

5.2.2.3 Comparison of Ovid MEDLINE and PubMed

The simple searches in PubMed had a recall of 98 % compared to 94 % in the Ovid MEDLINE searches. Precision was slightly worse in the PubMed searches with an average of 4.6 % compared to 5.7 % in the simple Ovid MEDLINE searches. Efficiency was better for the Ovid MEDLINE searches with an NNR of 66 compared to 88 for the PubMed searches.

5.2.2.4 Filter for study design

Filter for study design was used in 179 of the 254 simple search strategies. 134 were narrow and 45 broad. The published search strategies used filter in 190 reviews.

5.2.3 Review complexity

As part of the exploratory study, the subsample of 254 was sorted into complex and non-complex reviews to investigate whether complexity influenced the results. There were 199 non-complex and 55 complex reviews. Results are presented in the sections below.

5.2.3.1 Search effectiveness

In the conventional Boolean searches, recall was 90 % and precision 3.2 % for non-complex intervention reviews. For complex intervention reviews, recall was 78 % and precision 1.2 %. Average recall and precision for all reviews was 87 % and 2.8 %, respectively, independent of complexity.

In the simple, optimized searches, recall was 95 % and precision 6.1 % for non-complex intervention reviews. For complex intervention reviews, recall was 92 % and precision 3.8 %. Average recall and precision for all reviews was 94 % and 5.6 %, respectively.

5.2.3.2 Search efficiency

NNR

In the conventional Boolean searches, NNR was 229 for non-complex intervention reviews and 321 for complex. Average NNR for all reviews was 250.

In the simple, optimized searches, NNR was 62 for non-complex intervention reviews and 82 for complex. Average NNR for all reviews was 66.

Number of search terms

In conventional Boolean searches, the average number of search terms used was 60 (median=45) for non-complex intervention reviews and 116 (median=88) for complex interventions. Average number of search terms for all reviews was 72.

In simple, optimized searches, the average number of search terms necessary to retrieve the included studies was 5.9 (median=5.0) for non-complex intervention reviews and 13.1 (median=9.0) for complex interventions. Average number of search terms for all reviews was 7.2.

5.3 Use of an automated search method – PubMed ranking

The ranking function “Best Match” in PubMed was used to rank conventional Boolean search strategies, and the results were evaluated using MAP.

The published Ovid MEDLINE search strategies translated into PubMed syntax and ranked using the “Best Match” option had an average MAP of 7.2 %. MAP for the results sorted according to “First Author” was 3.2 % and 2.2 % according to the sort option “Most Recent”. These results were based on a sample of 234 of the subsample of 254 Cochrane reviews. In PubMed, search results with more than 10 000 hits cannot be generated into a text file, which make it difficult to find the hit numbers. For 20 of the reviews, the search results exceeded this limit, and MAP could therefore not be measured.

Average MAP values for each group is shown in Figure 16 below.

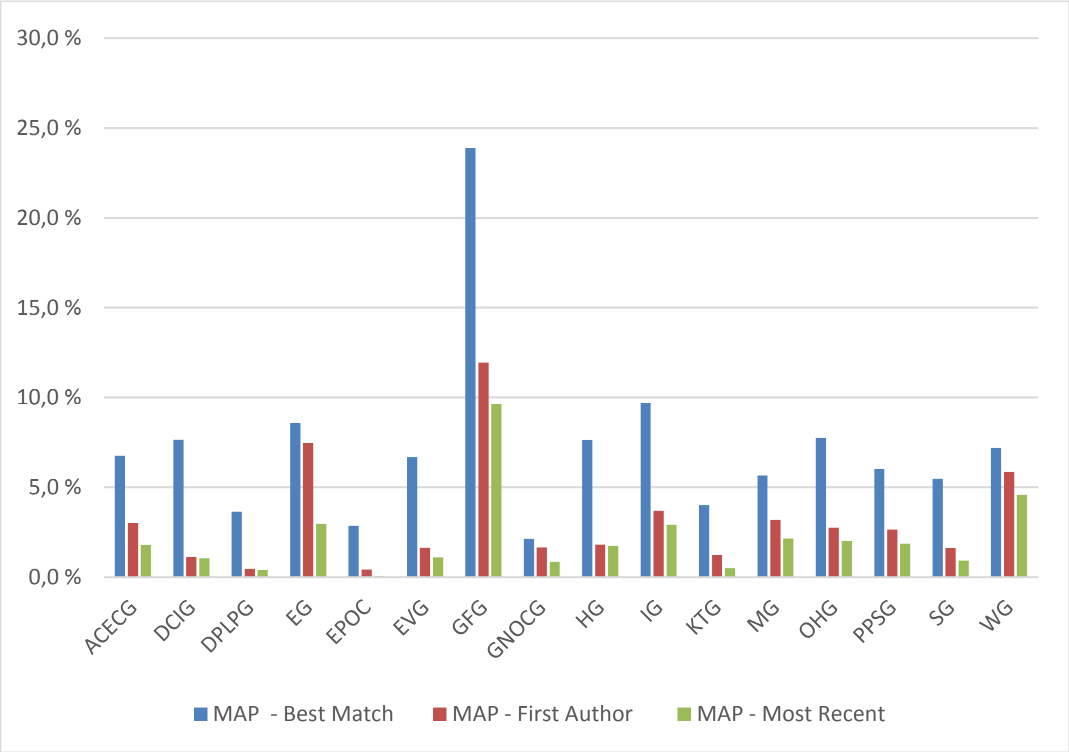


Figure 16 MAP for the published Boolean search strategies translated into PubMed syntax and ranked according to the sort option “Best Match”

Group differences

As can be seen from the figure above, the MAP values varied greatly between the different groups with “Best Match” performing better than the other sort options for all groups. To be able to decide whether this improvement in performance could be used to increase efficiency, cut-off values for the number of search results to be screened were set at 500, 1000, 1500 and 2000. Recall values can be seen at the different cut-off levels in Table 5 below.

Table 5 Recall values for ranked, conventional Boolean search strategies

	Review group	Cutoff at 500 hits	Cutoff at 1000 hits	Cutoff at 1500 hits	Cutoff at 2000 hits	Recall without cutoff
1	ACECG	80 %	86 %	89 %	89 %	89 %
2	DCIG	72 %	82 %	88 %	90 %	93 %
3	DPLPG	44 %	58 %	65 %	72 %	84 %
4	EG	85 %	86 %	88 %	89 %	89 %
5	EPOC	37 %	46 %	50 %	58 %	74 %
6	EVG	84 %	93 %	94 %	94 %	95 %
7	GFG	88 %	91 %	92 %	92 %	92 %
8	GNOCG	46 %	61 %	71 %	77 %	84 %
9	HG	66 %	70 %	71 %	73 %	75 %
10	IG	69 %	81 %	82 %	82 %	82 %
11	KTG	30 %	51 %	60 %	65 %	89 %
12	MG	72 %	82 %	83 %	84 %	94 %
13	OHG	76 %	80 %	82 %	83 %	84 %
14	PPSG	73 %	79 %	85 %	87 %	91 %
15	SG	57 %	73 %	80 %	87 %	95 %
16	WG	78 %	90 %	93 %	93 %	95 %
	Recall	**66 %**	**76 %**	**80 %**	**82 %**	**88 %**

5.4 Summary of results

5.4.1 Database coverage

Coverage in MEDLINE alone and MEDLINE and Embase combined for each review group based on 400 reviews is shown in Table 6 below.

Table 6 Coverage of MEDLINE alone and MEDLINE and Embase combined

	Review group	Coverage of MEDLINE alone	Coverage of MEDLINE and Embase combined
1	ACECG	89 %	95 %
2	DCIG	90 %	93 %
3	DPLPG	67 %	70 %
4	EG	82 %	84 %
5	EPOC	93 %	94 %
6	EVG	90 %	94 %
7	GNOCG	84 %	96 %
8	GFG	88 %	93 %
9	HG	87 %	92 %
10	IG	88 %	94 %
11	KTG	80 %	86 %
12	MG	82 %	90 %
13	OHG	92 %	95 %
14	PPSG	84 %	91 %
15	SG	69 %	82 %
16	WG	84 %	88 %

5.4.2 Precision

Average precision values for all searches in each review group based on 254 reviews are shown in Table 7 below.

Table 7 Precision values for all searches for all groups

	Review group	# of reviews	Ovid MEDLINE Published searches	PubMed Boolean translated from Ovid syntax	Ovid MEDLINE Simple searches	PubMed Simple searches
1	ACECG	20	2,5 %	2,6 %	7,0 %	5,4 %
2	DCIG	16	1,2 %	0,9 %	5,6 %	4,4 %
3	DPLPG	16	0,5 %	0,4 %	3,3 %	2,0 %
4	EG	12	5,0 %	4,3 %	7,7 %	5,9 %
5	EPOC	18	0,3 %	0,3 %	1,8 %	1,0 %
6	EVG	13	1,2 %	1,1 %	3,2 %	3,4 %
7	GFG	17	11,5 %	11,0 %	11,0 %	10,0 %
8	GNOCG	12	1,1 %	1,1 %	5,4 %	3,6 %
9	HG	17	1,7 %	1,6 %	5,8 %	5,1 %
10	IG	13	2,8 %	3,3 %	6,5 %	4,9 %
11	KTG	17	0,4 %	0,5 %	1,8 %	1,2 %
12	MG	20	3,0 %	3,1 %	6,6 %	5,9 %
13	OHG	16	2,6 %	2,2 %	5,3 %	4,5 %
14	PPSG	17	2,5 %	2,6 %	5,9 %	4,2 %
15	SG	16	1,0 %	1,0 %	4,1 %	3,6 %
16	WG	14	7,5 %	6,0 %	9,1 %	9,2 %

5.4.3 Recall

Average recall values for all searches in each group based on 254 review are shown in Table 8 below.

Table 8 Recall values for all searches for all groups

	Review group	# of reviews	Ovid MEDLINE Published searches	PubMed Boolean translated from Ovid syntax	Ovid MEDLINE Simple searches	PubMed Simple searches
1	ACECG	20	88 %	89 %	97 %	98 %
2	DCIG	16	91 %	94 %	100 %	100 %
3	DPLPG	16	84 %	85 %	93 %	98 %
4	EG	12	89 %	88 %	91 %	100 %
5	EPOC	18	74 %	74 %	90 %	92 %
6	EVG	13	95 %	94 %	87 %	98 %
7	GFG	17	94 %	93 %	96 %	98 %
8	GNOCG	12	82 %	82 %	90 %	96 %
9	HG	17	73 %	75 %	91 %	94 %
10	IG	13	83 %	83 %	96 %	98 %
11	KTG	17	89 %	90 %	99 %	99 %
12	MG	20	94 %	93 %	99 %	99 %
13	OHG	16	85 %	86 %	90 %	98 %
14	PPSG	17	90 %	90 %	92 %	98 %
15	SG	16	95 %	95 %	97 %	98 %
16	WG	14	93 %	93 %	96 %	97 %

5.4.4 NNR

Average NNR values for all searches in each review group based on 254 reviews are shown in Table 9 below.

Table 9 NNR for all searches for all groups

	Review group	Ovid MEDLINE Published searches	PubMed Boolean translated from Ovid syntax	Ovid MEDLINE Simple searches	PubMed Simple searches
1	ACECG	84	82	38	46
2	DCIG	307	390	89	122
3	DPLPG	828	1429	116	209
4	EG	65	58	55	72
5	EPOC	443	443	97	140
6	EVG	132	145	60	63
7	GFG	19	21	33	34
8	GNOCG	408	403	41	51
9	HG	389	530	145	203
10	IG	283	306	43	84
11	KTG	419	411	113	150
12	MG	129	129	46	41
13	OHG	77	127	36	55
14	PPSG	96	109	41	51
15	SG	247	247	62	91
16	WG	77	127	35	45

6 Discussion and conclusions

6.1 Introduction

Comprehensive searching is one of the hallmarks of systematic reviews and the advice of guidelines and handbooks on systematic review development (72, 194). The systematic search process is typically designed with the goal of finding all available primary studies that can help answer a research question, and it is generally considered that the more sources searched and the more comprehensive the search strategies, the higher the yield. This is done to minimize the possibility of missing important research studies and to avoid or at least reduce publication bias to a minimum. There are several disadvantages with comprehensive searching, though. The process of gathering relevant literature and identifying eligible studies can be time-consuming and costly, which is problematic for various reasons. One is situational, as some reviews and technology assessments are conducted to inform time-critical healthcare decisions. In such cases, one may be willing to trade off a higher likelihood of missing a small number of relevant studies to achieve more timely synthesis (56, 142). Another reason is that review resources are often scarce. Costs must be kept down, and consequently, methods should be as simple and optimal as possible. There is, however, limited empirical evidence upon which to base the extent of searching (63).

This study has therefore investigated various simplifying approaches for information retrieval in systematic reviews by both reviewing and performing empirical studies on the topics. The main aim was to improve efficiency of the search process by reducing the effort and resources used without noticeably degrading the effectiveness. To achieve this aim, both conventional and simpler, more optimized search methods were investigated using three main approaches. The first approach was to perform a cross sectional study on a sample of 400 Cochrane reviews to investigate whether searching only MEDLINE and Embase is enough for study identification in systematic reviews with respect to database coverage and search performance. The next approach used an exploratory design to find the optimal combination of search terms to retrieve the included studies in a subsample of 254 Cochrane reviews including 5 to 50 studies. The optimized search terms were used in simple search strategies and tested in Ovid MEDLINE and PubMed. Finally, the ranking function of PubMed was tested on conventional Boolean search strategies in the subsample to evaluate an automated method to support information retrieval in systematic reviews. As part of the study, three scoping reviews were performed to map the evidence on the main approaches, inform the empirical studies, and put the results of the study in context.

6.2 Comment

In the discussion chapter, results of the subsample (=254) are used in examples since some reviews had extreme values in the full sample, which skewed the results on group level.

6.3 Searching only MEDLINE and Embase for study identification in systematic reviews

The first approach to improve efficiency of the systematic search process in this study was to investigate whether searching only MEDLINE and Embase was enough for study identification in a sample of Cochrane reviews. To get an overview of previous literature on the topic, a scoping review was performed, which showed that several studies had investigated the effect of searching only MEDLINE and Embase for systematic reviews. Some of them found that it could be enough to search only MEDLINE or MEDLINE and Embase combined, while others claimed that despite their eminent reputation, a comprehensive search for information could not be limited to these two sources alone, since that would retrieve only part of the relevant literature. These conclusions, however, were based on different foundations. Some were based on database coverage and some on search performance. Therefore, to contribute to the discussion on the impact of reducing the number of databases to search in systematic reviews, this study has evaluated both the coverage in MEDLINE and Embase of primary research studies included in a sample of Cochrane reviews and the performance of the published Ovid MEDLINE search strategies of the same reviews. The results of the empirical part of the study and the scoping review will be discussed in the following sections.

6.3.1 Coverage of MEDLINE and Embase

The first approach to investigate whether searching only MEDLINE and Embase is enough for study identification in systematic reviews was to evaluate coverage of MEDLINE alone or MEDLINE and Embase combined of included studies in a sample of 400 Cochrane reviews. Coverage of Embase was only tested if studies were not indexed in MEDLINE. MEDLINE indexed 84 % of all included studies in the sample of 400 reviews and 85 % in a subsample of 254, containing 5 to 50 included studies, while MEDLINE and Embase combined held 90 % for both sample sizes. Average coverage of previous studies included in Scoping review I was 87 % for MEDLINE alone and 94 % for MEDLINE and Embase combined, which shows that the results of this study are approximately the same as that of earlier studies and confirms that most, but not all primary studies on health-related topics are indexed in MEDLINE and Embase. There are various reasons to search only MEDLINE and Embase for study identification for systematic reviews. MEDLINE and Embase are highly developed databases with sophisticated search interfaces (117), which make the searching easier and more efficient, and for many health-related systematic reviews most relevant material is indexed in MEDLINE and Embase. To reduce effort and resources used, searching either MEDLINE alone or the two combined could offer a viable solution, depending on research question.

Some topics, however, are better covered in smaller, more specialized databases. Literature from regional areas and developing countries, for example, is not well represented in MEDLINE and other well-known databases (64). One reason for this is the indexing practice of journals by the major databases. A priority system based on impact factor decides a journal’s chance of being indexed, and as western journals generally have higher impact factors, they are more likely to be indexed in MEDLINE and Embase (64). A surprising result of this study was, therefore, that only 31 of 464 primary studies in reviews from Effective Practice and Organization of Care Group (EPOC) were not indexed in MEDLINE. EPOC reviews are broad, complex and cover no clinical or biomedical topics, and several of the reviews concerned topics from middle and low-income countries. Average number of databases searched in EPOC was 10 (median=9), and other sources were used in addition, like reading reference lists of included studies. Nevertheless, most of the included studies of the EPOC reviews were indexed in MEDLINE, which most probably is because Cochrane reviews only include studies of high quality.

6.3.1.1 Should Embase be searched in addition to MEDLINE?

In this study, an average of 1.4 of the included studies of the sample reviews was indexed in Embase if not in MEDLINE. This is not too much depending on the total number of included studies. The usefulness of searching Embase in addition to MEDLINE differed highly between the groups, though. The group benefiting least from searching Embase was Effective Practice and Organization of Care Group (EPOC) with only 4 unique Embase references out of a total of 464 included studies across the 25 reviews, and the review groups Dementia and Cognitive Impairment Group (DCIG), Epilepsy Group (EG) and Eyes and Vision Group (EVG) had only 5 unique Embase references each. Other groups had more Embase unique studies, like Gynaecology and Fertility Group (GFG) with 47 and Stroke Group (SG) with 54. Consequently, the need to search Embase in addition to MEDLINE depends to a great extent on topic and probably also search quality. Lorenzetti et al. (142) wrote in their article on value of databases other than MEDLINE that when time and resources are scarce, there must be other, more specialized databases that could make a more significant contribution to study identification than Embase. Since Embase now encompasses all MEDLINE references, however, searching only Embase could be a viable solution, but might also offer some challenges. Anecdotal evidence suggests that Embase studies may be over-indexed, with many more indexing terms applied to individual studies than is typical of MEDLINE (142). This phenomenon may result in the identification of many irrelevant studies when searching Embase (142). Another, but not less important, aspect of this discussion is that Embase is only available through expensive subscriptions, while MEDLINE is free of charge through PubMed.

6.3.2 Performance of published Ovid MEDLINE searches

The next approach of this study was to evaluate search performance of the published Ovid MEDLINE search strategies of the 400 sample reviews. While it is important to know which sources contain relevant studies, this information cannot always be translated directly into recommendations, as the retrieval of studies is also dependent upon the search strategies used (131). Search techniques used to find research evidence for systematic reviews should be reliable and efficient (154). Search strategies differ widely in their performance, however, and one does not know whether a reference, although being indexed in a database, will actually be retrieved from that database using a specific search strategy.

6.3.2.1 Search effectiveness

Recall

The aim of systematic searching is to find all relevant studies on a topic. Although there is no widespread consensus on what constitutes acceptable recall in searching the literature generally, for systematic reviews there has been a strong preference for very high levels of recall, ideally 100 %, resulting in low precision (58). For the published Ovid MEDLINE search strategies of the 400 Cochrane reviews in this study, it was found that recall was quite high for most search strategies with an average of 88 %. Average recall of eight cross sectional studies included in Scoping review I was 81 %, ranging from 68 % to 93 %. None of these, however, investigated the effectiveness of search strategies in Cochrane reviews. The high recall value of this study could indicate that search strategies in Cochrane reviews have a higher quality or also that the ability of systematic search strategies to retrieve relevant studies has increased over the years. The recall values of single reviews in this study differed widely, though, and in no groups were all recall values higher than average value across all groups. In Wounds Group (WG), for example, with an average recall of 93 %, 4 out of 14 search strategies had a value lower than average for all review groups. One of them was “Honey as a topical treatment for wounds” (195) with a recall of 76 %. The following published search strategy was used in the review:

1 exp Skin Ulcer/ (18230)
2 exp Pilonidal Sinus/ (543)
3 exp Wounds, Penetrating/ (14308)
4 exp Lacerations/ (1423)
5 exp Burns/ (17087)
6 exp Wound Infection/ (14097)
7 exp Surgical Wound Dehiscence/ (2829)
8 exp "Bites and Stings"/ (7739)
9 exp Cicatrix/ (13623)
10 ((plantar or diabetic or heel$ or foot or feet or ischaemic or ischemic or venous or varicose or stasis or arterial or decubitus or pressure or skin or leg or mixed or tropical or rheumatoid or sickle cell) adj5 (wound$ or ulcer$)).ti,ab. (18800)
11 (bedsore$ or bed sore$).ti,ab. (239)
12 (pilonidal sinus$ or pilonidal cyst$).ti,ab. (437)
13 (cavity wound$ or sinus wound$).ti,ab. (37)
14 (laceration$ or gunshot or stab or stabbing or stabbed or bite$).ti,ab. (19976)
15 (burn or burns or burned or scald$).ti,ab. (19171)
16 (surg$ adj5 wound$).ti,ab. (5431)
17 (surg$ adj5 infection$).ti,ab. (8723)
18 (wound adj5 infection$).ti,ab. (10745)

19 (malignant wound$ or experimental wound$ or traumatic wound$).ti,ab. (428)

20 (infusion site$ or donor site$ or wound site$).ti,ab. (7460)

21 (skin abscess$ or skin abcess$).ti,ab. (162)

22 (hypertrophic scar$ or keloid scar$).ti,ab. (1595)

23 or/1-22 (130992)

24 exp Honey/ (1328)

25 honey.ti,ab. (3036)

26 or/24-25 (3160)

27 23 and 26 (249)

This strategy did not retrieve the following three included studies:

1. Robson V, Dodd S, Thomas S. Standardized antibacterial honey (Medihoney) with standard therapy in wound care: randomized clinical trial. J Adv Nurs. 2009;65(3):565-75.

2. Mphande AN, Killowe C, Phalira S, Jones HW, Harrison WJ. Effects of honey and sugar dressings on wound healing. J Wound Care. 2007;16(7):317-9.

3. Ingle R, Levin J, Polinder K. Wound healing with honey--a randomised controlled trial. SAMJ, S. Afr. med. j. 2006;96(9):831-5.

As can be seen from the references, all three contain both the words "wound" and "honey" in the title, and the following simple search strategy retrieved all 17 included studies of the review:

((wound OR ulcers OR burns) AND honey).mp

A title search for the two words "honey" and "wound" did not increase the screening burden to any great extent as it generated only 59 more hits but increased search effectiveness considerably as three more of the included studies were retrieved. Titles of research articles are meant to describe the content of the articles, and this study found that approximately half of the necessary number of search terms occurred in the review titles, which indicates that the title is a valuable source for feasible search terms, at least for Cochrane reviews. A search in the title field should therefore always be performed in the different databases using the problem and intervention words of the review title.

In the example described above, increasing recall was easy. In other sample reviews of this study, it would have been much more difficult, if not impossible, to increase recall, even if searching both title and abstract. For the review "Reduction in saturated fat intake for cardiovascular disease" (196), for example, the "patient word" is "cardiovascular disease". Four of the 16 included studies, however, were about cancer and two about diabetes, and neither of them mentioned any words for cardiovascular disease in the references. Another of the included studies of the same review was about "diet" and "cardiovascular disease" but did not mention "fat intake" in the bibliographic record, only in the full text article. Such studies are almost impossible to retrieve with a reasonably balanced search strategy. To search for the term "diet" would be too broad for the research question and cause too many hits. In many cases, optimal search strategies cannot be used because they retrieve an unmanageable number of results, so in practice, many systematic search strategies have to make compromises (131). Even though the aim of systematic searching is high recall, balancing a search strategy to obtain optimal recall and precision values and keep the number of hits down is important. It is difficult, however, to know when to stop searching when no relevant studies are known.

In attempting to capture all relevant studies, searchers may retrieve large numbers of irrelevant references, but screening a high amount of references can be quite challenging. Sampson et al. elaborated on this in an article from 2011 (117) and wrote that to cope with large result sets, reviewers may opt to initially screen by title, rather than examining the full bibliographic record, or may opt for a single screener rather than having two reviewers independently examine each record at the first stage. Further, the authors wrote, that at high screening volumes, reviewer fatigue may contribute to missed studies, even when those studies were retrieved by the search. The time required to screen and exclude irrelevant studies may impact on the timeliness, and, ultimately, the utility of systematic reviews (142).

Total recall is the aim of systematic searching. It could, however, be time to rethink this requirement and discuss what an acceptable recall level should be. For one of the studies included in Scoping review I (58), recall for simple searches for broader clinical questions dropped to around 80 % which the authors commented may be too low for a systematic review. This depends, however, on different factors, for example, how many studies there are on the topic in question and also the quality of the retrieved studies. There are also those that claim that it is not the proportion of included studies that matters but the answer and getting to it quickly (197). To make the search process more efficient, one might have to opt for more precise and optimized search strategies.

Precision

This study found that average precision was 2.8 for both the sample of 254 and the sample of 400 reviews. The median values, however, was 1.1 and 0.8 respectively, which shows that some reviews had quite high precision values, but most had low. Previous research included in Scoping review I had an average of 3.1 % (median=2.6 %) based on four cross sectional studies. In the study by Sampson et al. performed in 2011 (117) with the objective of establishing typical values for the precision of systematic review searches, overall precision was found to be 3.0 % (median=2.9 %). The results of this and previous studies could indicate that average precision values for systematic reviews are quite stable. The precision values of single reviews of this study varied widely, though. The review “Early versus delayed post-operative bathing or showering to prevent wound complications” (198), for example, had only one included study, which was not retrieved by the published search strategy and, consequently, a precision of zero. There were also reviews with an exceptionally high precision. The review “Ozone therapy for treating foot ulcers in people with diabetes” (199) with the following search strategy generated three hits and had therefore 100 % precision since all three hits were studies included in the review:

1 exp Foot Ulcer/
2 exp Diabetic Foot/
3 (diabet* adj3 ulcer*).tw.
4 (diabet* adj3 (foot or feet)).tw.
5 (diabet* adj3 wound*).tw.
6 or/1-5
7 exp Ozone/
8 ozon*.tw.
9 or/7-8
10 6 and 9
11 randomized controlled trial.pt.
12 controlled clinical trial.pt.
13 randomi?ed.ab.
14 placebo.ab.
15 clinical trials as topic.sh.
16 randomly.ab.
17 trial.ti.

18 or/11-17

19 exp animals/ not humans.sh.

20 18 not 19

21 10 and 20

The example described above was an exception, though. Generally, the high levels of recall of the published MEDLINE searches in this study was at the cost of precision.

The average precision values varied also widely between the groups. Effective Practice and Organization of Care Group (EPOC) had the lowest average precision with 0.3 % and Gynaecology and Fertility Group (GFG) the highest with 11.5 %. The low precision of EPOC search strategies was probably because all the research questions had high complexity. The high precision of GFG, on the other hand, is not so easy to explain. Unusually high precision may indicate that the search was not sufficiently broad to identify all relevant studies. This was not the case with GFG or any of the other groups in the sample. GFG and Wounds Group (WG), the two groups with the highest precision, also had high recall. WG had a precision of 7.5 % and a recall of 93 %. The opposite, that high recall was at the cost of low precision, was true for some groups, though, for example for Stroke Group (SG) with an average recall of 95 % and precision of 1.0 %. Low precision could be due to search style or unavoidable because of the topic and its indexing (117).

An approach to increase precision and efficiency in systematic review searching that has proved successful, is to use search filters for different kinds of study designs (200-203). In this study in the sample of 400 reviews, 297 search strategies used filter for study design. In 229 of these, none of the included studies was lost, and in 39, only one study was lost, which shows that using filter for study design reduced the screening burden considerably in this sample and did not reduce search effectiveness to any great extent, except in a few reviews.

Balancing recall and precision

The aim of a systematic search is high recall but should be balanced with a reasonable precision value. It is well known, however, that there is a trade-off between recall and precision. A search with high recall will probably retrieve more relevant studies but will often result in large numbers of irrelevant references. Improving precision will reduce the number of references to screen, but the chance of losing relevant references is high. Failing to find all studies, however, does not necessarily influence effect estimates meaningfully (56, 63). This is primarily true as systematic reviews have spread from clinically focussed intervention studies to more complex and broader health and social topics such as organizational issues or social work. A comprehensive search does not guarantee the retrieval of all relevant references either. A case study by Golder and Loke (129) demonstrated the failure of a broad search strategy with numerous synonyms, text words, and indexing terms to identify all relevant references available in a database. The low precision of systematic review searches combined with constraints in research funding has become a significant barrier to updating old (58) and producing new reviews. Thus, finding a balance between having a manageable number of references to screen, while being as complete as possible, is an important consideration in systematic searching.

6.3.2.2 Search efficiency

NNR is the inverse of precision and was used to measure efficiency of the published Ovid MEDLINE search strategies. For the sample of 400 Cochrane reviews, average NNR per review was 426 (median=344) and 249 (median=190) for the subsample of 254. Both values are high above NNR in previous studies. NNR was evaluated for five of the studies included in Scoping review I with an average of 59 (median=53) and a range from 9 to 123, which is considerably lower than the average values of this study. This shows that a fair amount of the reviews in this study had very low precision. The screening burden of some of these reviews was huge and probably unnecessary since efficiency could have been increased, for example by reducing the number of search terms. NNR differed widely among the review groups, though. In the subsample of 254 reviews, Gynaecology and Fertility Group (GFG) had the lowest NNR value with an average of 19 (median=12) and Developmental, Psychosocial and Learning Problems Group (DPLPG) the highest with an average of 828 (median=345). This difference could be due to topic, search style and/or complexity in research question.

6.3.2.3 Comparison of conventional search strategies in Ovid MEDLINE and PubMed

The recall of both the published Ovid MEDLINE search strategies and the same strategies translated into PubMed syntax was 87 %. Precision was 2.8 % and 2.7 %, respectively. Thus, effectiveness of the two search interfaces was approximately the same. Efficiency, measured by NNR, was slightly worse in PubMed with 307 compared to 249 in Ovid MEDLINE. There was also a slightly higher average number of search terms in the PubMed search strategies with 83 (median=57) compared to the Ovid MEDLINE search strategies with an average of 72 (median=53). Most Cochrane reviews have their search strategies performed in Ovid

MEDLINE, but Ovid MEDLINE is a costly, commercial product. PubMed, on the other hand, is free of charge and could be used as an alternative. Altogether, the results of this study indicate that systematic searches can be performed in PubMed to save resources.

6.3.3 Challenges in systematic searching

There are many challenges and considerations to take when deciding on comprehensiveness of the systematic search process. The number of databases searched and the number of search terms used will both influence the search process. Searching many databases and using a wide selection of search terms will result in more hits and thereby increase the screening burden. Generally, it will also increase the likelihood that all or most relevant research studies are found. On the other hand, although precision tends to be inversely related to recall, screening large numbers of hits is no assurance of complete identification of relevant studies. Every search strategy may not necessarily fit the research question, and each new search will contribute to increase the number of references to screen. On the average, most included studies in the sample of this study were indexed in MEDLINE and Embase and retrieved by the published MEDLINE search strategies. Some studies, however, were not indexed in either of the two databases and not retrieved by the search strategies. The reasons for this differed and could be due to review complexity, the topics reviewed, and language variations. These issues will be discussed in the sections below.

6.3.3.1 Language, topics and review complexity

Language is a big challenge in systematic searching, and one of the most obvious is synonymy. Important articles may be overlooked if all relevant synonyms for a concept are not included in a search strategy. The same intervention or patient group may be referred to by different terms within a single body of literature, and some authors may describe their research in ways that are unfamiliar to the reader (153). Primary studies on complex research questions can be widely dispersed, and terminology and indexing practices differ from one area or field to another. Topics like public health, education and psychotherapy, for example, can cover many discipline areas (151) and be indexed in other databases than MEDLINE or Embase or in other sources, like web-sites (64). Different terminology used across disciplines can hamper identification for primary studies, especially on broad or complex research questions. As the results of this study show, literature on such questions can be more challenging to retrieve, which could be due to lack of standardization and ambiguity in language.

Intervention complexity has been discussed in the systematic review community for several years. Some claim that all interventions are more or less complex (45) and some that they fall along a spectrum from simple to more complex (204). Concerning information retrieval, it is more likely that it is easier to retrieve evidence for non-complex interventions than for complex. There is not a substantial amount of research to support this assumption, though. To investigate whether complexity had any influence on information retrieval performance in this study, the 254 reviews of the subsample were divided into complex/non-complex according to the Cochrane Handbook (184) and evaluated for effectiveness and efficiency. The results showed that search effectiveness was better for the non-complex intervention reviews. For conventional Boolean searches, non-complex reviews had a recall rate of 90 % and precision of 3.2 %. For complex reviews, recall fell to 78 % and precision to 1.2 %. Search efficiency was also better for non-complex reviews. NNR was 229 for non-complex intervention reviews and 321 for complex. Average number of search terms used was 60 (median=45) for non-complex intervention reviews and 116 (median=88) for complex interventions in conventional Boolean search strategies. Among the non-complex interventions are drugs and physical interventions, where terminology is more standardized and could be one reason why search performance was better for non-complex reviews. It seems reasonable to claim that evidence for non-complex reviews is easier to retrieve, which is not surprising and also supported by previous research (136, 154).

The more standardized the language of a research topic, the fewer synonyms are necessary to find relevant studies. Searches for pharmaceutical interventions, for example, may be associated with high levels of precision, as well as recall (136), and this could be due to the more exact and consistent description provided for pharmaceuticals. Sampson et al. (117) argued in their article “Precision of healthcare systematic review searches in a cross-sectional sample” that selection is easier, as the knowledge community has made decisions on the language to be used and the meanings ascribed, and these decisions are stable over a substantial period of time. Reviews on broader and more complex topics and in other fields, they wrote, will differ in the discovery tools available, and language variation will make the searching more challenging. The authors concluded that in subject areas where meanings change or the indexer in one country is not familiar with the implications of the context in which a word or phrase is used to describe a healthcare topic in another country, there are bound to be limitations.

If a review addresses a broad or complex question that spans multiple disciplines with a less standardized language, more synonyms have to be searched to retrieve all relevant studies, as was seen in the Effective Practice and Organization of Care Group (EPOC) reviews of this study. On the other hand, including a relevant but broad term may retrieve too many irrelevant studies (73), and while it may seem important conceptually to use a general search term, in practice it will probably add few relevant studies beyond those identified using more specific terms. All synonyms must not necessarily retrieve studies that are relevant and be included in a review and will only increase workload. This study has shown that the necessary number of search terms that will retrieve relevant studies of high quality, might be smaller than is generally assumed. A substantial reduction in number of search terms was possible to find the included studies in most of the sample reviews. It differed widely, however, how big this reduction was in and across the different groups according to research question.

6.3.3.2 Effect of reducing the number of databases to search

Scoping review I showed that conclusions differed in previous research about whether it is necessary to search more databases than MEDLINE and Embase for studies to include in systematic reviews. Already in 2003, two articles were written on the reduction of databases to search for systematic reviews, and both advised limiting searches to a core of main bibliographic databases, for example MEDLINE and Embase (59, 205). This conclusion has been supported by others (206), while some claim that a comprehensive search for information cannot be merely limited to these two sources alone as relevant studies could be missed (9). The two contradictory statements are based on different assumptions, though. Several of the studies claiming that it is necessary to search more databases, were case studies on complex research questions or research questions on broad topics like education and social care. For therapeutic intervention and diagnostic accuracy studies, searching only MEDLINE and Embase is considered enough by several previous studies (56, 63, 137, 138, 147). The results of this study also showed that search performance was better in non-complex reviews.

Reducing the number of databases to search will probably lose a few relevant studies for most reviews. None of the previous studies reviewed in Scoping review I found that MEDLINE and Embase covered all included studies of their sample reviews, and many of the published search strategies in the sample of this study did not retrieve all MEDLINE-indexed studies included in a review. What is important, though, is whether missing a few relevant studies will influence the conclusions of a review. Most studies concluded merely on coverage and/or search performance, but a few also investigated whether the conclusions of the reviews would have changed if the studies not indexed in a database or retrieved by a search strategy were omitted from the review. It appeared that there were very few cases of change in statistical significance suggesting that selective searching may not necessarily introduce bias in terms of effect estimates (56, 63, 137).

Another effect of reducing the number of databases is of a more psychological character. Some reviewers will probably not feel confident about searching only a couple of databases, because they are afraid of bias and critic, which could be reasons why the number of databases searched for systematic reviews has constantly increased over the years. In a study by Lam and McDiarmid (127) with the purpose of determining whether the number of bibliographic databases searched in systematic reviews changed over a twenty-year period, the average number of bibliographic databases searched in 1994, 2004, and 2014 were 1.62, 3.34 and 3.73, respectively. Studies that searched only one database decreased over the three milestone years (60 % in 1994, 28 % in 2004, and 10 % in 2014). In 2005, Sood et al. (146) assessed the sources of original literature contributing to Cochrane reviews on acupuncture and found that the median number of databases searched was 5, with a range of 3 to 12. In a study conducted in 2008, an average of 6 databases with a range of 3 to 19 were searched in a sample of 39 systematic reviews (143). The present study found that an average of 8 (median=7) with a range of 2 to 25 electronic databases was searched per Cochrane review in the sample of 400 reviews and confirms that the number of databases searched in systematic reviewing continues to increase. This development is not compatible with the need to simplify the search process and make better use of scarce resources and is probably not necessary since most relevant studies are found in the main bibliographic databases.

6.3.3.3 Searching databases versus other sources

While methodological guidance for systematic reviews encourages comprehensive searching, there are diminishing returns with each additional database searched. The impact of searching additional databases in terms of the final results and conclusions is not known, and searching only MEDLINE and Embase could miss important research studies. There is, however, a question of whether searching more electronic databases is the solution to secure high recall. In a study on Cochrane reviews of therapeutic interventions (177), most high quality studies could be identified by searching four standard databases: CENTRAL, MEDLINE, Embase and Science and Social Sciences Citation Indexes. Searching 26 further databases identified only an extra 2.4 % of trials. Aagaard et al. (171) evaluated search performance of 23 systematic reviews on musculoskeletal disorders. Cumulative median recall for combined searching in MEDLINE, Embase and CENTRAL was 88.9 % and increased to 90.9 % when adding 10 additional databases.

It has been suggested that searching only databases is not sufficient to identify all relevant references and that reviewers must rely upon additional sources in their literature search. Beyer et al. (131) investigated the optimal combination of databases to retrieve 31 studies on frozen shoulder management and concluded that their study showed the importance of reference checking in addition to database searching as one reference was only found in the reference list of a relevant systematic review. Rather than searching more databases, a better solution could be to increase the use of other sources that do not cause an unnecessary amount of references to screen. More serendipitous discovery methods may yield more relevant studies than more database searching (63).

6.3.4 Conclusion

Within the systematic review community, there has long been a growing interest in search efficiency, in particular whether it is possible to identify relevant research studies for systematic reviews by searching fewer databases (147). The interest in reducing the number of databases has been driven by several factors including the improved indexing, coverage and searching possibilities of databases, and the need to produce systematic reviews within time and resource constraints.

The results of both this and several of the studies included in Scoping review I indicate that searching only MEDLINE or MEDLINE and Embase combined could be enough for many health-related research questions. The conclusion of several previous studies was that searching electronic databases beyond MEDLINE and Embase did not increase yield substantially. Consequently, it has been argued that a well-structured search undertaken in only two or three databases, supported by supplementary search techniques to identify evidence, such as reference list checking and citation searching, might identify evidence more efficiently than a similar search undertaken in more databases (131, 147). The choice of databases will be influenced by a number of factors, including topic, accessibility, specificity, ease of use, and comprehensiveness (144). MEDLINE and Embase, however, have sophisticated search interfaces and cover most topics and are therefore the preferred choices when performing systematic searches. Some topic-specific research questions, though, will probably be better covered in regional or more specialized databases.

A high percentage of the included studies of the sample reviews in this study was indexed in MEDLINE and Embase, and generally, the published Ovid MEDLINE search strategies had reasonably high performance. It seems plausible to say that many searchers exaggerate the number of databases searched for systematic reviews with little or no return on the investment. It is important to construct effective searches of the main database(s), though, to find the relevant studies that are indexed in the database(s).

6.4 Selection of search terms and development of simple, optimized search strategies

Simpler and more optimized methods should be tried out to increase efficiency of the search process for systematic reviews, and this study has investigated various aspects of the search term selection process and the development and performance of simple, more optimized search strategies. More specifically, the optimal number of search terms was selected for a sample of Cochrane reviews and used in simple search strategies performed in Ovid MEDLINE and PubMed. By simple are meant strategies with only a multi-purpose field code and no adjacency operators. Further, it was investigated whether the optimized search terms occurred in the review title or rest of the protocol. Search term selection methods and the performance of simpler search strategies were reviewed in Scoping review II to get an overview of the topics and to inform the empirical studies.

6.4.1 Search term selection

The selection of search terms for systematic reviews is a crucial point of the review process in that it determines how many relevant studies that are found. The conventional way to develop a search strategy is an iterative manual process that involves continual assessment and refinement, and the usefulness of individual search terms is determined by the results. This process, however, can be both meticulous and resource demanding and limited not only to the experience and skill of the searcher but also to his or her knowledge of the topic under investigation. Therefore, more automated methods for search term selection has been tried out.

6.4.1.1 Number of optimal search terms

Little is known about the number of search terms necessary to retrieve relevant studies to include in systematic reviews. This study has investigated the optimal combination of search terms in MEDLINE search strategies in a sample of 254 Cochrane reviews using an exploratory design. The aim was to assess the actual need for search terms by investigating how many terms were necessary to find the included studies of the reviews and whether they occurred in the review title or rest of the protocol. An average of 7.2 (median=6.0) terms per review was necessary to retrieve the included studies in simple, optimized search strategies in Ovid MEDLINE and PubMed. For the conventional published Ovid MEDLINE search strategies, the average number of search terms used was 71 (median=61) per review. Thus, the average number of search terms necessary to retrieve the included studies of the sample reviews of this study was approximately one tenth of the number used in the published searches, which indicates that a reduction in number of search terms could have been recommended in most cases.

The optimal number of search terms varied widely across the 16 review groups, ranging from an average of 4.1 in Wounds Group (WG) to 17.7 in Effective Practice and Organization of Care Group (EPOC). Leaving out EPOC reduced average number of optimal search terms to 6.4 per search strategy for the other 15 groups. The maximum number of necessary search terms for an EPOC review was 74 for the review “Non-specialist health worker interventions for the care of mental, neurological and substance-abuse disorders in low- and middle-income countries” (207), which is a very complex research question, and consequently, it was difficult to keep the number of search terms low. All of the reviews in the EPOC group was of a complex nature and probably the reason why the need for search terms was high above the need for most other reviews. Both the review groups EPOC and Stroke Group (SG) used considerably more search terms in the published search strategies than the other groups. The necessary number of search terms for the simple, optimized SG search strategies, however, was much lower than for EPOC, with an average of 7.7 terms per review. This is just above the total average, indicating that this group used an unnecessary high amount of search terms for the reviews of this sample.

6.4.1.2 Search term sources

One objective of this study was to investigate whether the optimized search terms occurred in the review title or rest of the protocol. Across all review groups, an average of 3.0 (median=3.0) search terms per review of the 7.2 (median=6.0) occurred in the title of the review, and 2.8 (median=2.0) search terms in the rest of the protocol. Thus, an average of 1.4 (median=0.0) search terms per review was neither found in the title or rest of the review protocol, which is quite a small amount. There were great differences between the groups, however. For none of the groups could all search terms be found in the review title or protocol, but for the Wounds Group (WG) only one term altogether in the sample of 14 reviews did not occur in title or protocol. For 11 of the 16 Cochrane groups, an average of one search term per strategy could not be found in either the title or rest of the protocol. Only in the Effective Practice and Organization of Care Group (EPOC) most of the optimized search terms did not occur in title and protocol and consequently difficult to find.

For 54 of the 254 reviews of this study, words in the titles were enough to retrieve the included studies, and 36 of these were drug-related. Of the 254 reviews, 113 were categorized as drug-related, which shows that title was not enough as search term source to retrieve all relevant studies on drug-related questions in the sample of this study. Two previous studies included in Scoping review II (136, 154), suggested that it might be enough to use the review title as only source for developing feasible search strategies for some reviews. Day et al. (136) investigated the effect of simpler search strategies for pharmaceuticals and physical modalities in a sample of Cochrane reviews from the Back, Musculoskeletal and the Pain, Palliative Care and Supportive Care groups. They developed a simple strategy consisting of one term for the intervention and one for the condition in question. The condition and intervention terms were derived from the titles of the reviews, and it was found that simplified search strategies were an effective, efficient way to search for clinical trials when the intervention was a pharmaceutical or a well-defined physical treatment. The authors warned, however, that researchers engaged in literature searches for systematic reviews should be aware that simplified search strategies can lack sufficient sensitivity for their purposes. Waffenschmidt et al. (154) developed simple-structured search strategies for systematic reviews on drugs and wrote that “By means of allocation to specific search terms, the components of the research question in the systematic review of interest were classified as either belonging to the therapeutic indication or to the intervention. The corresponding search terms were generally the active ingredient (e.g., venlafaxine) and the disease (e.g., depression).” Probably, these terms occurred in the review title. They had, however, to be combined with the 20 first “Similar articles” offered in PubMed to prove a reliable method to assess the completeness of an evidence base of published RCTs.

In their study on Boolean versus ranked retrieval for systematic reviews, Karimi et al. (61) used information from different parts of the review protocol. They constructed three sets of bag-of-words queries, incorporating increasing amounts of information, including title only. They experienced that the more search terms they used, the better was the performance. This is quite surprising since the sample reviews were drug-related. They used a partial match search system, however, and that could be one reason why the performance was low for all searches and not suitable for systematic reviews.

For Cochrane reviews it is obvious that the review title is a good source for selecting search terms as it is required to use words for the problem and intervention of a research question in all titles. In four of the simple search strategies of this study, however, none of the optimized search terms occurred in the review title. In the review “Tubal flushing for subfertility” in the Gynaecology and Fertility Group (GFG), for example, the search terms “infertility” and “contrast medium” were most efficient to retrieve the 12 included studies. “Subfertil*” retrieved only two of the included studies and (tub* AND flush*) only three. Consequently, on some occasions review titles might not be the best source for feasible search terms. It may be that the titles do not mirror the language used by the research community, or they are not suited for the research question. The results of both this and the previous studies described above indicate that words occurring in titles are not enough to retrieve all relevant studies on a research question, not even drug-related.

If title and protocol are not enough to find relevant words, it is difficult to know where to select the most relevant search terms. One solution could be to rely on a selection of relevant references as was done in this study and has been done in previous studies (66, 93, 152). They could be generated from systematic reviews on the same topic or from a very narrow search, for example a title-only search as recommended by Hausner et al. (51). Automatic selection of search terms based on relevant publications could also be helpful, especially for reviews of complex and broad topics, like the Effective Practice and Organization of Care Group (EPOC) reviews. The use of more automated methods for search term selection will be discussed further in the next section.

6.4.1.3 Search term selection methods

Different methods are used to select search terms in systematic review searches. The conventional way is to use a manual, iterative process where feasible search terms are tested one by one until arriving at a satisfactory result. Conventional search methods in systematic reviewing have been criticized for being too subjective, especially on how search terms are selected. Several of the studies included in Scoping review II had investigated more automated methods to make the process more objective and reproducible.

Text mining or text analysis has been suggested as one way to select search terms automatically (51) and thereby increase objectivity. Using text mining software to select search terms could also be beneficial as it reduces the amount of user input. The individual iterations of traditional search term development involve significant labour input that potentially could be streamlined with more automated methods (208). Another advantage of more automatic search term selection methods is that the process is reproducible if it is strictly documented. A transparent, reproducible method for search term selection based on more automated systems is desirable to improve the overall methodical nature of the systematic review and thereby make the findings of the review more defensible (153).

The benefits of text analysis are increased speed, quality and reproducibility and has shown promising results in the development of search strategies for identifying studies, especially on diffuse or complex topics (105, 209). Relevant bibliographic references are often used as sources for the search term selection process. The benefit of using bibliographic references as search term sources will probably depend on both the topic under investigation and how many relevant studies there are on a topic. It will also depend on indexing qualities, that is, whether a reference contains an abstract, the length of the abstract, and language variation.

Incorporating text mining into the systematic review process has been recommended as an adjunct to employing experienced information professionals. O'Mara-Eves et al. (151) concluded in their article on techniques for identifying cross-disciplinary and "hard-to-detect" evidence for systematic reviews that text mining "should never be used on its own but rather in conjunction with the expertise and usual processes that are followed when developing a search strategy." Simon et al. (93) noted in their article on the use of text mining to identify primary publications on nurse staffing research that "the described development process for an empirical search strategy is a useful – though technically demanding – approach to building performance-oriented strategies." In a study on the practical application of text mining to literature on cognitive rehabilitation and enhancement through neuro-stimulation, Balan et al. (210) wrote that, “Methodologically speaking, we conclude that text-mining was helpful in getting an overall perspective on a huge corpus of literature with some level of detail, intentionally limited to handle complexity. Richer information can be extracted using more complex text-mining methods focused on narrower topics, but this requires extensive training and knowledge.” Although text mining seems promising, it also seems challenging and has not become a standard tool for supporting the development of systematic review search strategies yet.

The aim of much research on automatic search term selection has been to increase objectivity and transparency of the process. Completely objective is this method not, however, as some stages still require input from the user (153), and the selection process is based on already retrieved literature. The empirical approach described by Hausner et al. (51) was based on references either included in a previous systematic review on a similar research question or on a very precise search with the aim of finding a few relevant studies. Both these methods are based on approaches that are more or less subjective. The search terms will mirror the language used in the references that were used as a starting point. Thus, complete objectivity is difficult to obtain, and it has been suggested that an approach guiding the selection of relevant terms could help searchers develop search strategies in a more objective and systematic manner (93). An empirically guided approach to the development of a search strategy could provide a way to increase transparency and efficiency and thereby make this part of the systematic review process more scientific. Still, no guidelines are developed to help guide the search term selection process in systematic reviews. No matter which method is used for search term selection, however, when a feasible query has been formulated, results must be interpreted, the query revised and reiterated until arriving at a satisfactory solution (211). Validation of search strategies could occur as soon as all screening has been completed and the eligible studies are known, because then poorly performing search strategies can be amended, re-tested and re-run (95).

6.4.2 Development and performance of simple, optimized search strategies

Simple searches were developed and performed in Ovid MEDLINE and PubMed using the optimized search terms selected in the exploratory part of this study. In Ovid MEDLINE, the simple searches were performed using only the multi-purpose field ".mp". They had an average recall of 94 % and a precision of 5.6 %. In PubMed, searches were performed without field codes and had an average recall of 98 % and a precision of 4.6 %. The simple searches in PubMed were performed without field codes to make the searching simpler. Searching without field codes in PubMed turns the mapping function on, which expands the search to include more search terms and could be one reason why recall was higher in PubMed but with reduced precision.

The aim of the simple, optimized search strategies was 100 % recall. The results, however, show that it is difficult or almost impossible to obtain full recall when balancing the strategies to also obtain a reasonably high precision in a sample of reviews on different topics. Obviously, both recall and precision were higher for the simple, optimized search strategies in this study than for the published strategies since all relevant studies were known, and the search strategies were developed to retrieve them, but the values were not as high as expected. For several groups and single reviews, the differences in performance between the published searches and the simple, optimized searches were small, which shows that the performance of most of the published search strategies was high. For some of the groups, for example the Gynaecology and Fertility Group (GFG), search performance of the published searches was as high or almost as high as for the optimized search strategies. Why this is so, is not obvious. The research questions in this group were mostly of a non-complex nature, but so were the questions of many other groups, as well, for example the Epilepsy Group (EG), the Eyes and Vision Group (EVG), the Injuries Group (IG), the Musculoskeletal Group (MG) and the Oral Health Group (OHG). One explanation could be the search style of this group or that they have a great knowledge or overview of the research in their field.

As expected, the simple searches were also more efficient than the conventional Boolean searches. The published Ovid MEDLINE search strategies of the sample reviews were mostly quite effective, but not very efficient. They had an average NNR of 249, and the same search strategies translated into PubMed syntax had an NNR of 307, while the simple, optimized Ovid MEDLINE and PubMed strategies had an NNR of 66 and 88, respectively. The simple, optimized search strategies had approximately the same NNR values as the studies reviewed in Scoping review I, which had an average NNR of 59. None of the studies evaluating NNR in the scoping review were Cochrane reviews, however, which could indicate that in the effort for high recall, Cochrane reviews become quite inefficient.

6.4.3 Conclusion

Based on the results of this and earlier studies, there are reasons to believe that simpler, more optimized search strategies can replace conventional, comprehensive search strategies for systematic reviews at least on some occasions. Conventional Boolean search strategies seem to be unnecessary long and complicated with an exaggerated use of search terms compared to what is necessary to retrieve relevant studies. As this study has demonstrated, the actual need for search terms was approximately one tenth of what was used in the published conventional search strategies. The number of search terms used in the conventional strategies differed considerably, while the need for search terms was consistently low in all reviews. The number of search terms could be reduced in almost all reviews in the sample independent of review group and kind of research question. This shows that there is a potential to reduce the number of search terms considerably and thereby the screening burden. Previous studies have suggested that simpler search methods consisting of fewer search terms could be used for updates and rapid reviews (63, 136, 154). The results of this study indicate that simpler search methods could also be used for new reviews depending on topic and complexity.

Manual search term selection is the conventional method to select search terms for systematic reviews. To support the search term selection process, however, more automated methods have been investigated and could be a useful supplement to the conventional manual method. In this study, most relevant search terms could be found in the review title and protocol. For non-complex reviews, a manual selection method based on words in title and review protocol seems to be both effective and efficient. For more complex reviews, however, a combination of manual and automated methods, like text analysis, could provide the best result, as demonstrated by several of the included studies in Scoping review II (7, 66, 149, 151, 152). Still, however, automated methods should be used with caution as the only search term selection method.

6.5 Performance of automated search methods

For general searching, partial match retrieval with ranked search results is by far the most common method today, but for systematic reviews, conventional Boolean searching is still the preferred approach. To increase efficiency and objectivity, however, more automated search methods is a desired alternative for systematic reviews, and different automated approaches have been tried out and were presented in Scoping review III. Most of the included studies in the review were descriptive, but one of the experiments investigating ranked versus Boolean retrieval found that none of the tested ranked approaches were suitable for information retrieval in systematic reviews except a hybrid approach combining ranked and Boolean retrieval, the extended Boolean retrieval model (61, 212). To investigate extended Boolean retrieval for systematic reviews on an existing easy-to-use information retrieval system, this study has tested the ranking function “Best Match” in PubMed on conventional Boolean search strategies.

6.5.1 Extended Boolean retrieval

Extended Boolean retrieval is a transparent and reproducible method with the aim of showing the most relevant results of a hit list first. Once the ranked scores are computed, the results are presented in decreased order, and the most relevant records should tend to be towards the top of the ranking list (61). To test extended Boolean information retrieval for this study, the published search strategies in a sample of 234 Cochrane reviews including from 5 to 50 primary studies were translated into PubMed syntax and ranked, using the sort function “Best Match”. MAP was used to evaluate the results and had an average value of 7.2 %. When sorted by “First Author” and “Most Recent”, the search results had MAP values of 3.2 % and 2.2 %, respectively, which shows that the ranking function succeeded in moving the relevant hits closer to the top of the result lists.

To investigate whether the relevant studies appeared closer to the top of the result lists, cut-off values were assigned at 500, 1000, 1500 and 2000 hits. The average number of retrieved records in the 234 search strategies was 3215 (median=1082) per review. The great difference in average and median values shows that there are great variations between reviews in the number of retrieved records, which makes it problematic to set the same cut-off values for all. It appeared, however, that all review groups, except the Anaesthesia, Critical and Emergency Care Group (ACECG), the Epilepsy Group (EG), the Gynaecology and Fertility Group (GFG) and the Injuries Group (IG), had reduced recall values at all cut-off levels. A cut-off level of 1500 could have been set for the reviews of ACECG, GFG and IG in this sample without reducing recall, and for EG a cut-off value of 2000 would not have reduced recall. This is not a very great improvement, though, and therefore it seems reasonable to state that the ranking function “Best Match” in PubMed did not succeed in moving relevant studies very much closer to the top of the result lists for most searches. One reason could be that many of the included studies in Cochrane reviews are quite old, and new references are given a higher weight.

Automated information retrieval methods could have made the systematic search process more efficient, but Boolean searching is still the most important retrieval method for systematic reviews. There could be various reasons for this, and advantages and disadvantages with partial and exact match information retrieval for systematic reviews are discussed in the next section.

6.5.2 Advantages and disadvantages with partial and exact match searching

There are advantages and disadvantages with both partial and exact match systems when used for systematic reviews. First, there are some clear advantages with Boolean search systems since they have a more expressive power than ranked retrieval (212). A specific concept may be represented by a complex expression and searched in different fields such as ((intravenous* or subcutaneous*) adj3 immunoglobulin*).tw,kf. There is the possibility of using a thesaurus in the query processing, which add synonyms to a query in order to increase recall (212, 213). The use of metadata can capture the semantics of the search explicitly and in ways that free-form queries cannot, which gives the searcher more control of the search process (61). In spite of such enhancements, the problems with Boolean searching are well known: complex query logic requiring search expertise, no ranking in result sets, dichotomous retrieval, equal term weights, and the inability to control the result set size except by adding or removing query terms (212, 213). Boolean retrieval systems are characterized by the properties of the query submitted to the system, rather than the mode of indexing or algorithms employed (214).

In their article on Boolean versus ranked querying for systematic reviews, Karimi et al. discussed advantages and disadvantages with the two search systems and wrote that an important advantage of using partial match is in the process of iterative query construction (61). The items, they claimed, that are most likely to be relevant, are on top of the result list and that it should be relatively straightforward to see whether changes to queries have improved the results, for example, by examining just the first few documents in an answer set and observe whether they are relevant. There is no equivalent feature for Boolean queries. In fact, in the Boolean case, the authors argued, it is difficult and time-consuming to judge whether an alteration to a query leads to an improvement or harms the answer set. As experienced in this study, however, ranking might not always push the most relevant studies to the top of a result list as expected.

Karimi et al. also wrote that ranked queries can have some other disadvantages, for example how many search terms to use in a query (61). Since algorithms in a ranked search engine may not be known, we do not know how synonyms and word variants are handled or how the search terms are weighted. Another key drawback of partial match systems, the authors wrote, is that the size of the result set is likely to be much larger than that returned as the result of a Boolean search, and it is difficult to decide when to stop examining the result list. This is certainly a problem, the authors stated, since time saved is a desired outcome when trying out new search technology for systematic reviews, and the number of documents that a team of reviewers can examine is limited. For use in systematic reviews, ranked retrieval has still another significant disadvantage, Karimi et al. wrote, and that is lack of reproducibility. As the collection changes in a database, so does the term statistics and therefore the rankings (61). Adding even a small number of new documents to a collection could potentially impact the whole ranking score (61). Pohl (212) wrote on this problem that the typical information retrieval ranking functions face the difficulty that their ranking is dependent on properties of the whole collection and can thus be difficult to reproduce. He emphasized that reproducibility is a key requirement of systematic reviews that helps in assessing review quality, and that if ranked queries are used, reproducibility can only be assured if all aspects of the computation are reported, including term weights and within-document term frequencies. With Boolean queries all that is required is publication of the search strategy that was used, together with the database vendor and date span of the databases it was applied to (212).

One reason to simplify and automate the search process for systematic reviews is that systematic searching is quite complicated. Several studies have revealed that the quality of Boolean searches varies widely (148, 215). To develop an effective and efficient search strategy requires practice, intuition, and some trial and error and has been said to be a combination of art and science (216). Some claim that it is more of an art than a science and that there is no prescriptive approach to conducting a comprehensive search but that searching is an art that can be cultivated and practiced (73). In the article “Classical databases and knowledge organization: A case for Boolean retrieval and human decision-making during searches”, Hjørland wrote that Boolean searching requires much specialist knowledge that ordinary users seldom have time or motivation to learn and emphasized that the important role of human expertise in searching should not be ignored (217). Boolean retrieval is often considered a less efficient approach, but according to Hjørland, the Boolean retrieval model is important in order to provide users with the power to make informed searches and have full control over what is found and what is not found. In the book “Human information retrieval”, Warner differentiated between the computer science tradition, aiming at automatically transforming queries into (ranked) sets of relevant references, and the older library oriented tradition aiming at increasing the “selection power” of users (218).

As can be seen from this discussion, there are advantages and disadvantages with both partial and exact match information retrieval for systematic reviews. If Boolean searching is more like an art than a science that could be one reason why more automated search methods have not replaced the Boolean search. If searching is more like an art, however, it is also unlikely that two people will develop identical search strategies or yield identical results from a search on the same review question (73) and that is contrary to the research ideal of objectivity.

6.5.3 Conclusion

Despite many advantages, automated information retrieval is still not a viable option, compared to Boolean searching for information retrieval in systematic reviews. Extended Boolean retrieval, a hybrid between automated and Boolean retrieval, had success in an earlier study (61) and was tested in this study using the ranking option “Best Match” in PubMed. This did not improve search efficiency to any great extent in most search strategies, however, and was not a feasible approach to reduce the screening burden. It could therefore not be recommended as an alternative to conventional non-ranked Boolean searches.

Previous studies have tested different automated search methods like text mining, machine learning, and ranking algorithms for information retrieval in systematic reviews. Many of them reported on positive results of the automated methods. Based on the reviewed studies, however, it seems reasonable to say that no automated search method can replace the conventional Boolean search at the moment. More automated search methods can, however, assist in routine searching, for updating existing reviews, and for evaluating the completeness of an evidence base. According to Thomas et al. (158), specialist human skills will always be required in the creation and maintenance phase of the development of a good search strategy.

6.6 How comprehensive is comprehensive enough in systematic searching?

The previous sections of this chapter discussed the three main search approaches that were investigated in the empirical studies and reviewed in the scoping reviews. This section will discuss comprehensiveness in systematic searching more generally as this has been a concern in the review community for many years. As early as 2003, Egger et al. (54) wrote an article called "How important are comprehensive literature searches?" investigating the characteristics of clinical trials that were difficult to locate and the impact of excluding them. The authors concluded that systematic reviews that are based on a search of English language literature that are accessible in the major bibliographic databases will often produce results that are close to those obtained from reviews that are based on more comprehensive searches, which is also supported by other studies (56, 63). Egger et al. recommended that reviewers should consider the type of literature search and the degree of comprehensiveness that are appropriate for the review in question, taking into account budgetary and time constraints. They also claimed that trials that are difficult to locate are often of lower quality and that rather than preventing bias through extensive literature searches could introduce bias by including trials of low methodological quality. The authors suggested that thorough quality assessment is more important than extensive literature searching.

Andrew Booth has discussed the challenge of comprehensive searching in several articles. In 2006 (219), he wrote that "More recent developments within health technology assessment have led to recognition that it is not always possible (or indeed desirable) to expend considerable resources in the pursuit of diminishing returns from the evidence. Time and funding for systematic searching is usually finite. In many cases, 'good enough' is regarded as an acceptable substitute for the ideal." He also added, however, that "Here again, 'good enough' is both subjective and elusive." Booth wrote further that items of evidence found should not only be included in a review, but also make a difference to the overall findings of the review and concluded that "we should recognize that we are always dealing with considerable uncertainty, that we never know how large the total population of studies on a particular topic actually is. As a consequence, we never know how close we are to this total nor can we tell how many studies are missing when we decide to 'call off the search'" (219).

In the article "How much searching is enough?" from 2010, Booth (220) continued his discussion on comprehensiveness in systematic searching and asked for a much-needed evolution from comprehensive toward optimal retrieval. On the question "Why perform comprehensive searches?" he answered that there are three reasons to perform comprehensive searches and that is to maximize the chance of identifying all relevant references, to convince readers that the process underpinning the systematic review is robust, and to minimize the risk of reports being challenged for incompleteness. He wrote further that no type of review can claim to identify all relevant references, but the requirement should be to be robust. Existing methods of evidence retrieval embody strategies to minimize the risk of missing relevant studies independently from strategies for comprehensiveness, he argued. Studies with the greatest potential impact are more likely to be published, more likely to appear in high quality journals, more likely to be covered in multiple databases, and more likely to be cited. The chances that such studies will be missed are already relatively slight, and studies that are more likely to be overlooked would thus be best not found, he claimed, and concluded that the aspiration of a systematic search should not be comprehensiveness but rather the minimization of bias (220).

Others have also claimed that there must be pragmatic limits on identifying "all" research in a systematic review search (221). In an article from 2015 (222), Petticrew wrote that "more may not be better" and asked: "Do literature searches really need to be comprehensive?". He claimed that systematic reviewers conduct comprehensive literature searches in order to reduce the risk of missing key studies and to minimize publication bias and wrote that checklists (223) and guidelines (23) encourage comprehensive searching by asking "Was a comprehensive literature search performed?". He argued that there are alternatives to conducting large-scale searches in the quest for comprehensiveness and suggested that purposive searching and saturation could be an alternative to comprehensive searching also worth considering in relation to quantitative reviews, where there are usually rapidly diminishing returns from large-scale searches. In such cases, he argued that the value of the literature search often lies more in preventing future criticism of lack of assiduousness in searching, rather than in any real anticipation of finding evidence which would overturn the review's conclusions. Petticrew agreed with Booth on the importance of prioritizing database searches so that the potentially more productive databases are searched first and concluded that undoubtedly what Lefebvre (224) called "the perennial question of 'when is enough enough'" will become ever more important as reviews become more complex.

There may also be risks in abandoning comprehensiveness as a goal. In particular, bias may be introduced if the search ceases when positive conclusions are reached (222). It is well documented that studies with statistically significant results are more likely to be published and more likely to be published in journals with high impact factor than studies with non-significant results. Therefore, if a review of treatment effects only considers published studies, it is very likely to over-estimate the effectiveness of the treatment under consideration (225). Some studies are only found in “grey literature”, and others are never published at all, which makes the focus on 100 % recall problematic. This could be alarming because unpublished trials may have systematically different results from those that are published. Publication bias increases the chances that reviews based only on published reports will themselves be biased, which leads to the conclusion that the best methods may require time consuming searches and extensive enquiries to track down unpublished studies (20). The concept of saturation may also be difficult to apply when there is little research evidence to begin with (222).

This discussion shows that there are different facts and circumstances that have to be taken into consideration when deciding on the degree of comprehensiveness in systematic searching and that there is no easy solution to the question. Comprehensive searches may help mitigate publication, database, and language bias, but require substantial time, effort, and resources. Diminishing returns are expected along the continuum of increasingly comprehensive study identification efforts (56). Pragmatic searches for systematic reviews could be the future, and there are those that claim that traditional, comprehensive literature searches might even be unethical as they potentially cause significant waste (197). Maximizing the comprehensiveness of a search is a logical way to try to minimize bias in a systematic review, but in practice, it is rarely feasible to carry out a truly comprehensive search, and prioritization decisions must be made (131).

6.7 Overall summary

Although the methodological history of systematic reviews is constantly evolving, certain core principles still remain, such as the use of wide-ranging, comprehensive searches to reduce the effects of publication bias (222). There might, however, be time to reconsider the comprehensiveness issue to increase efficiency of the search process and reduce both the increasing number of databases searched and the excessive use of search terms in systematic search strategies. For many research questions, searching only MEDLINE or MEDLINE and Embase combined seems to be enough based on both this and previous research. It also seems to be safe to reduce the number of search terms for almost all systematic search strategies.

The protocol should be the main source for search terms, at least for non-complex reviews, and text analysis can help select useful search terms for broader and more complex research questions. When it comes to use of search method, Boolean searching is still the main information retrieval method for systematic reviews despite several efforts to automate the search process.

6.8 Limitations / Comments

There are several limitations and shortcomings to this study.

6.8.1 Accuracy and consistency in study results

As the author have been the only participant for the main part of this project, the results might not be as reliable as if more authors had participated in the project.

Concerning the reproduction of the published Ovid MEDLINE search strategies, there could have been changes in MeSH indexing since a search was run for the first time, which could have influenced the search results.

When finding the optimal combination of search terms in the exploratory study, searches were performed followed by scrutiny of relevant search terms. For many non-complex reviews, words in title were enough to retrieve the included studies. For more complex reviews requiring more search terms, however, the selection process was more difficult, and different search terms could probably have been selected for some of the simple, optimized search strategies.

The protocol was not accessible at the time of the investigation, but it was assumed that the background, objectives, and methods sections of the reviews had been part of the protocol and were therefore used as “substitutes”. What could be a problem is that the information in the reviews that substituted for the protocol, could have been changed as a consequence of the search results.

The results of the empirical part of this study are based on information obtained after the search process is finished and can therefore not guide search approaches for new reviews. They can, however, shed light on the actual need for number of databases and search terms.

6.8.2 Cochrane reviews

This study has only examined Cochrane reviews, and thus, the results may not be transferable to other reviews. Cochrane reviews have been found to differ in many ways from other systematic reviews on the basis of style and possibly topic coverage (168, 226). They are known to have higher methodological quality than reviews published in peer-reviewed journals and typically search multiple sources for potentially relevant studies. Major HTA programs, however, have adopted systematic review methods similar to those used by the Cochrane Collaboration (56), which could mean that the results may be relevant to their systematic review processes, as well.

7 References

1. Armstrong R, Hall BJ, Doyle J, Waters E. Cochrane Update. 'Scoping the scope' of a cochrane review. J Public Health (Oxf). 2011;33(1):147-50.

2. Thomas J, McNaught, John, Ananiadou, Sophia. Applications of text mining within systematic reviews. Res Synth Methods. 2011(2):1-14.

3. Tsafnat G, Glasziou P, Choong MK, Dunn A, Galgani F, Coiera E. Systematic review automation technologies. Syst Rev. 2014;3:74.

4. Cohen AM, Ambert K, McDonagh M. Cross-topic learning for work prioritization in systematic review creation and update. J Am Med Inform Assoc. 2009;16(5):690-704.

5. Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? PLoS Med. 2010;7(9):e1000326.

6. Marshall C, Sutton A. Systematic Review Toolbox 2015 [cited 2019 31 Oct]. Available from: http://systematicreviewtools.com/index.php.

7. Ananiadou S, Rea B, Okazaki N, Procter R, Thomas J. Supporting Systematic Reviews Using Text Mining. Soc Sci Comput Rev. 2009;27(4):509-23.

8. Lund H, Brunnhuber K, Juhl C, Robinson K, Leenaars M, Dorch BF, et al. Towards evidence based research. BMJ. 2016;355:i5440.

9. Cognetti G, Grossi L, Lucon A, Solimini R. Information retrieval for the Cochrane systematic reviews: the case of breast cancer surgery. Ann Ist Super Sanita. 2015;51(1):34-9.

10. Grimshaw JM, Eccles MP, Lavis JN, Hill SJ, Squires JE. Knowledge translation of research findings. Implement Sci. 2012;7:50.

11. Godolphin W, Towle A, McKendry R. Evaluation of the quality of patient information to support informed shared decision-making. Health Expect. 2001;4(4):235-42.

12. Bhandari M, Giannoudis PV. Evidence-based medicine: what it is and what it is not. Injury. 2006;37(4):302-6.

13. Bjørndal A, Klovning A, Flottorp S. Kunnskapshåndtering i medisin og helsefag. Oslo: Gyldendal akademisk; 2007. 195 s. : ill. p.

14. Straus SE, Sackett DL. Evidence-based medicine : how to practice and teach EBM. 3rd ed. ed. Edinburgh: Elsevier Churchill Livingstone; 2005.

15. About Cochrane Reviews - What is a systematic review? : Cochrane Library; [cited 2018 Feb 12]. Available from: http://www.cochranelibrary.com/about/about-cochrane-systematic-reviews.html.

16. Evans D, Kowanko I. Literature reviews: Evolution of a research methodology. Aust J Adv Nurs. 2000;18(2):33-8.

17. Chalmers I. Trying to Do More Good than Harm in Policy and Practice: The Role of Rigorous, Transparent, Up-to-Date Evaluations. Ann Am Acad Pol Soc Sci. 2003;589:22-40.

18. Long L. Routine piloting in systematic reviews--a modified approach? Syst Rev. 2014;3:77.

19. Creswell JW. The Purpose Statement. In: Research design: qualitative, quantitative, and mixed methods approaches. Los Angeles, Calif.: SAGE; 2014. p. 123-38.

20. Chalmers I, Dickersin K, Chalmers TC. Getting to grips with Archie Cochrane's agenda. BMJ. 1992;305(6857):786.

21. Forsetlund L. Introduction. In: Towards evidence-based public health practice. Unipubavhandlinger. Oslo: Faculty of Medicine, University of Oslo Unipub; 2004. p. 8.

22. Jamtvedt G, Norderhaug I. Informasjon om forskningsbehov fra Kunnskapssenterets systematiske oversikter 2009 [Internet]. Oslo: Nasjonalt kunnskapssenter for helsetjenesten; 2010 [cited 2019 Sept 27]. Available from: https://www.fhi.no/globalassets/dokumenterfiler/notater/2010/notat_10_forskningsbehov_2009.pdf.

23. Cochrane Handbook for Systematic Reviews of Interventions. Chichester (UK): John Wiley & Sons; 2019 [cited 2019 Sept 29]. Available from: https://training.cochrane.org/handbook.

24. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ. 2009;339:b2535.

25. Gough D, Oliver S, Thomas J. An introduction to systematic reviews. Los Angeles: SAGE; 2012. X, 288 s. : ill. p.

26. Yoshii A, Plaut DA, McGraw KA, Anderson MJ, Wellik KE. Analysis of the reporting of search strategies in Cochrane systematic reviews. J Med Libr Assoc. 2009;97(1):21-9.

27. Snilstveit B, Vojtkova M, Bhavsar A, Gaarder M. Evidence gap maps - a tool for promoting evidence-informed policy and prioritizing future research Washington, DC, USA: World Bank; 2013 [cited 2019 Sept 30]. Available from: https://openknowledge.worldbank.org/bitstream/handle/10986/16941/WPS6725.pdf?sequence=1&isAllowed=y.

28. European Platform for Investing in Children (EPIC): European Commission. Employment, Social Affairs & Inclusion; [cited 2019 Sept 29]. Available from: http://europa.eu/epic/topics/evidence-based-practices_en.htm.

29. Davies P. The relevance of systematic reviews to educational policy and practice. Oxf Rev Educ. 2000;26(3-4):365-78.

30. Bayliss HR, Beyer FR. Information retrieval for ecological syntheses. Res Synth Methods. 2015;6(2):136-48.

31. Hoorens S, Mattox T, Kilburn M, Barberi M. The use of evidence-based practices in child policy: The case of the European Platform for Investing in Children Geneve2016 [cited 2019 Sept 30]. Available from:

https://www.unige.ch/cide/files/8015/4167/0349/Publication_Adequation_Politique_enfance_jeunesse_Realite_terrain.pdf#page=50.

32. Marshall C, Brereton P, editors. Tools to support systematic literature reviews in software engineering: A mapping study. International Symposium on Empirical Software Engineering and Measurement; 2013.

33. Petticrew M. Systematic reviews from astronomy to zoology: myths and misconceptions. BMJ: British Medical Journal. 2001;322(7278):98.

34. What is Cochrane? About us: The Cochrane Collaboration; 2014 [cited 2019 Sept 30]. Available from: http://www.cochrane.org/about-us.

35. Stavrou A, Challoumas D, Dimitrakakis G. Archibald Cochrane (1909–1988): the father of evidence-based medicine. Interact Cardiovasc Thorac Surg. 2013;18(1):121-4.

36. De Brún C, Pearce-Smith N. Introduction. In: Badenoch D, Perera R, Heneghan C, editors. Searching skills toolkit: finding the evidence. Chichester, West Sussex: Wiley Blackwell; 2009. p. 1.

37. Shah HM, Chung KC. Archie Cochrane and his vision for evidence-based medicine. Plast Reconstr Surg. 2009;124(3):982.

38. Chalmers I. Archie Cochrane (1909-1988). J R Soc Med. 2008;101(1):41-4.

39. Cochrane Reviews: Cochrane Library; 2019 [cited 2019 Sept 30]. Available from: https://www.cochranelibrary.com/cdsr/reviews.

40. Methodological Expectations of Cochrane Intervention Reviews (MECIR): Cochrane Methods; [cited 2019 Sept 30]. Available from: https://methods.cochrane.org/methodological-expectations-cochrane-intervention-reviews.

41. MECIR Manual: Cochrane Community; [cited 2018 Feb 12]. Available from: http://community.cochrane.org/mecir-manual.

42. Welcome to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) website! [cited 2019 May 02]. Available from: http://www.prisma-statement.org/.

43. Munn Z, Stern C, Aromataris E, Lockwood C, Jordan Z. What kind of systematic review should I conduct? A proposed typology and guidance for systematic reviewers in the medical and health sciences. BMC Med Res Methodol. 2018;18(1):5.

44. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ. 2008;337:a1655.

45. Moore GF, Evans RE, Hawkins J, Littlecott HJ, Turley R. All interventions are complex, but some are more complex than others: using iCAT_SR to assess complexity. Cochrane Database Syst Rev. 2017;7:Ed000122.

46. 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.

47. Bartels EM. How to perform a systematic search. Baillieres Best Pract Res Clin Rheumatol. 2013;27(2):295-306.

48. Rathbone J, Carter M, Hoffmann T, Glasziou P. A comparison of the performance of seven key bibliographic databases in identifying all relevant systematic reviews of interventions for hypertension. Syst Rev. 2016;5:27.

49. PRISMA 2009 Checklist [cited 2019 May 02]. Available from: http://www.prisma-statement.org/documents/PRISMA%202009%20checklist.pdf.

50. Standards for Initiating a Systematic Review. In: Eden J LL, Berg A, et al., editor. Finding What Works in Health Care: Standards for Systematic Reviews. Washington (DC): National Academies Press (US); 2011. p. 47.

51. Hausner E, Waffenschmidt S, Kaiser T, Simon M. Routine development of objectively derived search strategies. Syst Rev. 2012;1:19.

52. McGowan J, Sampson M. Systematic reviews need systematic searchers. J Med Libr Assoc. 2005;93(1):74-80.

53. Foster M, Jewell S. The last piece: Librarianship and systematic reviews. In: Foster M, Jewell S, editors. Assembling the pieces of a systematic review : a guide for librarians. Medical Library Association books. Lanham, Maryland: Rowman & Littlefield; 2017. p. 204.

54. Egger M, Juni P, Bartlett C, Holenstein F, Sterne J. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. Health Technol Assess. 2003;7(1):1-76.

55. Chapman AL, Morgan LC, Gartlehner G. Semi-automating the manual literature search for systematic reviews increases efficiency. Health Info Libr J. 2010;27(1):22-7.

56. Halladay CW, Trikalinos TA, Schmid IT, Schmid CH, Dahabreh IJ. Using data sources beyond PubMed has a modest impact on the results of systematic reviews of therapeutic interventions. J Clin Epidemiol. 2015;68(9):1076-84.

57. O'Mara-Eves A, Thomas J, McNaught J, Miwa M, Ananiadou S. Using text mining for study identification in systematic reviews: a systematic review of current approaches. Syst Rev. 2015;4:5.

58. Rice M, Ali MU, Fitzpatrick-Lewis D, Kenny M, Raina P, Sherifali D. Testing the effectiveness of simplified search strategies for updating systematic reviews. J Clin Epidemiol. 2017;88:148-53.

59. Royle P, Milne R. Literature searching for randomized controlled trials used in Cochrane reviews: rapid versus exhaustive searches. Int J Technol Assess Health Care. 2003;19(4):591-603.

60. Drucker AM, Fleming P, Chan A-W. Research techniques made simple: assessing risk of bias in systematic reviews. J Invest Dermatol. 2016;136(11):e109-e14.

61. Karimi S, Pohl S, Scholer F, Cavedon L, Zobel J. Boolean versus ranked querying for biomedical systematic reviews. BMC Med Inf Decis Mak. 2010;10:58.

62. van Enst WA, Scholten RJ, Whiting P, Zwinderman AH, Hooft L. Meta-epidemiologic analysis indicates that MEDLINE searches are sufficient for diagnostic test accuracy systematic reviews. J Clin Epidemiol. 2014;67(11):1192-9.

63. Hartling L, Featherstone R, Nuspl M, Shave K, Dryden DM, Vandermeer B. The contribution of databases to the results of systematic reviews: a cross-sectional study. BMC Med Res Methodol. 2016;16(1):127.

64. Betran AP, Say L, Gulmezoglu AM, Allen T, Hampson L. Effectiveness of different databases in identifying studies for systematic reviews: experience from the WHO systematic review of maternal morbidity and mortality. BMC Med Res Methodol. 2005;5(1):6.

65. Whiting P, Westwood M, Burke M, Sterne J, Glanville J. Systematic reviews of test accuracy should search a range of databases to identify primary studies. J Clin Epidemiol. 2008;61(4):357-64.

66. 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. J Clin Epidemiol. 2015;68(2):191-9.

67. Bui DD, Jonnalagadda S, Del Fiol G. Automatically finding relevant citations for clinical guideline development. J Biomed Inform. 2015;57:436-45.

68. Smalheiser NR, Lin C, Jia L, Jiang Y, Cohen AM, Yu C, et al. Design and implementation of Metta, a metasearch engine for biomedical literature retrieval intended for systematic reviewers. Health Inf Sci Syst. 2014;2:1.

69. Fernández-Sáez AM, Bocco MG, Romero FP. SLR-Tool a tool for performing systematic literature reviews Athens, Greece.: ICSOFT 2010 - The 5th International Conference on Software and Data Technologies; 2010 [cited 2019 Sept 30]. Available from: https://pdfs.semanticscholar.org/d3c1/d9d5f31e0a3beaeff5cfaa9079de87f301e5.pdf?_ga=2.190093163.1136061249.1569845349-2092808879.1562391061.

70. Sampson M, Shojania KG, Garritty C, Horsley T, Ocampo M, Moher D. Systematic reviews can be produced and published faster. J Clin Epidemiol. 2008;61(6):531-6.

71. Israel M, Hay I. Relationships, integrity and ethics of care. In: Research ethics for social scientists: between ethical conduct and regulatory compliance. London: Sage; 2006. p. 112-28.

72. Lefebvre C, Manheimer E, Glanville J. Searching for studies. 2011. In: Cochrane Handbook [Internet]. Cochrane Collaboration, [cited 2018 May 12]. Available from: https://handbook-5-1.cochrane.org/chapter_6/6_searching_for_studies.htm.

73. Aromataris E, Riitano D. Systematic reviews: constructing a search strategy and searching for evidence. Am J Nurs. 2014;114(5):49-56.

74. Bullers K, Howard AM, Hanson A, Kearns WD, Orriola JJ, Polo RL, et al. It takes longer than you think: librarian time spent on systematic review tasks. J Med Libr Assoc. 2018;106(2):198.

75. Snilstveit B, Stevenson J, Shemilt I, Clarke M, Jimenez E, Thomas J. Timely, Efficient, and Living Systematic Reviews: Opportunities in International Development2018 [cited 2019 Oct 20]. Available from: https://cedilprogramme.org/wp-content/uploads/2018/11/Inception-Paper-7-Birte-Snilsveit-Timely-Efficient-and-living-systematic-reviews.pdf.

76. Golder S, Loke YK. Sources of information on adverse effects: a systematic review. Health Info Libr J. 2010;27(3):176-90.

77. About MEDLINE® and PubMed®: The Resources Guide: National Institutes of Health, US National Library of Medicine; [cited 2019 Oct 8]. Available from: https://www.nlm.nih.gov/bsd/pmresources.html.

78. MEDLINE®: Description of the Database: National Institutes of Health. US National Library of Medicine; [cited 2019 Sept 30]. Available from: https://www.nlm.nih.gov/bsd/medline.html.

79. PubMed Help: National Center for Biotechnology Information, U.S. National Library of Medicine; 2005 [cited 2017 May 04]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK3827/.

80. Welcome to Medical Subject Headings! : NIH U.S. National Library of Medicine; [cited 2019 Jan 03]. Available from: https://www.nlm.nih.gov/mesh/meshhome.html.

81. Chapman D. Advanced search features of PubMed. J Can Acad Child Adolesc Psychiatry. 2009;18(1):58-9.

82. Embase: Elsevier; [cited 2017 May 04]. Available from: https://www.elsevier.com/solutions/embase-biomedical-research.

83. How CENTRAL is created: Cochrane Library; [cited 2019 Mar 27]. Available from: https://www.cochranelibrary.com/central/central-creation.

84. Bramer WM, Giustini D, Kramer BM. Comparing the coverage, recall, and precision of searches for 120 systematic reviews in Embase, MEDLINE, and Google Scholar: a prospective study. Systems Review. 2016;5:39.

85. Kelley GA, Kelley KS. The Association between Databases for Indexing Studies Intended for an Exercise Meta-Analysis of Arthritis Randomized Controlled Trials. Arthritis [Internet]. 2012 [cited 2019 Sept 30]. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3424631/.

86. Rethlefsen ML, Murad MH, Livingston EH. Engaging medical librarians to improve the quality of review articles. JAMA. 2014;312(10):999-1000.

87. Sampson M, McGowan J, Cogo E, Grimshaw J, Moher D, Lefebvre C. An evidence-based practice guideline for the peer review of electronic search strategies. J Clin Epidemiol. 2009;62(9):944-52.

88. Booth A. Unpacking your literature search toolbox: on search styles and tactics. Health Info Libr J. 2008;25(4):313-7.

89. Schardt C, Adams MB, Owens T, Keitz S, Fontelo P. Utilization of the PICO framework to improve searching PubMed for clinical questions. BMC Med Inform Decis Mak. 2007;7(1):16.

90. Bayliss SE, Dretzke J. Health technology assessment in social care: a case study of randomized controlled trial retrieval. Int J Technol Assess Health Care. 2006;22(1):39-46.

91. Ramer SL. Site-ation pearl growing: methods and librarianship history and theory. J Med Libr Assoc. 2005;93(3):397-400.

92. Schlosser RW, Wendt O, Bhavnani S, Nail-Chiwetalu B. Use of information-seeking strategies for developing systematic reviews and engaging in evidence-based practice: The application of traditional and comprehensive Pearl Growing. A review. Int J Lang Commun Disord. 2006;41(5):567-82.

93. Simon M, Hausner E, Klaus SF, Dunton NE. Identifying nurse staffing research in Medline: development and testing of empirically derived search strategies with the PubMed interface. BMC Med Res Methodol. 2010;10:76.

94. Gargon E, Williamson PR, Clarke M. Collating the knowledge base for core outcome set development: developing and appraising the search strategy for a systematic review. BMC Med Res Methodol. 2015;15:26.

95. Sampson M, McGowan J. Inquisitio validus Index Medicus: A simple method of validating MEDLINE systematic review searches. Res Synth Methods. 2011;2(2):103-9.

96. White VJ, Glanville JM, Lefebvre C, Sheldon TA. A statistical approach to designing search filters to find systematic reviews: objectivity enhances accuracy. J Inf Sci. 2001;27(6):357-70.

97. Booth A. Searching for qualitative research for inclusion in systematic reviews: a structured methodological review. Syst Rev. 2016;5:74.

98. Salton G, Fox EA, Wu H. Extended Boolean information retrieval. Commun ACM. 1983;26(11):1022-36.

99. Thomas J, McNaught J, Ananiadou S. Applications of text mining within systematic reviews. Res Synth Methods. 2011;2(1):1-14.

100. Ananiadou S, Thompson P, Thomas J, Mu T, Oliver S, Rickinson M, et al. Supporting the education evidence portal via text mining. Philosophical transactions Series A, Mathematical, physical, and engineering sciences. 2010;368(1925):3829-44.

101. Zhou W, Smalheiser NR, Yu C. A tutorial on information retrieval: basic terms and concepts. J Biomed Discov Collab. 2006;1(1):2.

102. Canese K. PubMed relevance sort. NLM Tech Bull. 2013;394:e2.

103. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med. 2009;151(4):264-9.

104. Documenting the search process and reporting the search results. 2011. In: Cochrane Handbook for Systematic Reviews of Interventions [Internet]. Cochrane Training. Version

5.1. [cited 2019 Sept 30]. Available from: https://training.cochrane.org/trials-search-co-ordinator-handhook/6-author-support/65-documenting-search-process-and-reporting.

105. Paynter R, Bañez LL, Berliner E, Erinoff E, Lege-Matsuura J, Potter S, et al. EPC methods: an exploration of the use of text-mining software in systematic reviews: Agency for Healthcare Research and Quality (US); 2016 [cited 2019 Sept 30]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK362044/.

106. Cooper C, Dawson S, Peters J, Varley-Campbell J, Cockcroft E, Hendon J, et al. Revisiting the need for a literature search narrative: A brief methodological note. Res Synth Methods. 2018;9(3):361-5.

107. Craven J, Levay P. Recording database searches for systematic reviews-what is the value of adding a narrative to peer-review checklists? A case study of NICE Interventional Procedures Guidance. Evid Based Libr Inf Pract. 2011;6(4):72-87.

108. Clough P, Sanderson M. Evaluating the performance of information retrieval systems using test collections. Inf Res. 2013;18(2).

109. Van Rijsbergen C. Retrieval effectiveness. In: Sparck Jones K, editor. Information retrieval experiment. London: Butterworths; 1981. p. 32.

110. Borlund P. The concept of relevance in IR. J Am Soc Inf Sci Technol. 2003;54(10):913-25.

111. Ingwersen P, Järvelin K, Croft WB. Introduction. In: Croft WB, editor. The Turn: Integration of Information Seeking and Retrieval in Context. The Information Retrieval Series. 18. Dordrecht: Dordrecht: Springer Netherlands; 2005. p. 1.

112. Sampson M, Zhang L, Morrison A, Barrowman NJ, Clifford TJ, Platt RW, et al. An alternative to the hand searching gold standard: validating methodological search filters using relative recall. BMC Med Res Methodol. 2006;6:33.

113. Frické M. Measuring recall. J Inf Sci. 1998;24(6):409-17.

114. Moseley AM, Elkins MR, Herbert RD, Maher CG, Sherrington C. Cochrane reviews used more rigorous methods than non-Cochrane reviews: survey of systematic reviews in physiotherapy. J Clin Epidemiol. 2009;62(10):1021-30.

115. Baeza-Yates R, Ribeiro-Neto B. Retrieval evaluation. In: Modern information retrieval: the concepts and technology behind search. Harlow: Addison Wesley; 2011. p. 131.

116. Kekäläinen J, Järvelin K, editors. Evaluating information retrieval systems under the challenges of interaction and multidimensional dynamic relevance. Proceedings of the 4th CoLIS conference; 2002.

117. Sampson M, Tetzlaff J, Urquhart C. Precision of healthcare systematic review searches in a cross-sectional sample. Res Synth Methods. 2011;2(2):119-25.

118. Baeza-Yates R, Ribeiro-Neto B. Retrieval evaluation. In: Modern information retrieval: the concepts and technology behind search. Harlow: Addison Wesley; 2011. p. 140.

119. Hersh W. Evaluation of biomedical text-mining systems: lessons learned from information retrieval. Brief Bioinform. 2005;6(4):344-56.

120. Hua X-S, Worring M, Chua T-S. Experimental Setup. 2012. In: Internet Multimedia Search and Mining [Internet]. [cited 2019 Sept 30]; [8-15]. Available from: http://doras.dcu.ie/19610/1/imsmbook.pdf.

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

122. Colquhoun HL, Levac D, O'Brien KK, Straus S, Tricco AC, Perrier L, et al. Scoping reviews: time for clarity in definition, methods, and reporting. J Clin Epidemiol. 2014;67(12):1291-4.

123. Daudt HM, van Mossel C, Scott SJ. Enhancing the scoping study methodology: a large, inter-professional team's experience with Arksey and O'Malley's framework. BMC Med Res Methodol. 2013;13:48.

124. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol. 2005;8(1):19-32.

125. Royle P, Waugh N. A simplified search strategy for identifying randomised controlled trials for systematic reviews of health care interventions: a comparison with more exhaustive strategies. BMC Med Res Methodol. 2005;5:23.

126. Lefebvre C, Manheimer E, Glanville J. Searching CENTRAL, MEDLINE and EMBASE: specific issues. In: The Cochrane Handbook - Version 510 [Internet]. The Cochrane Collaboration, [cited 01.10.2019]. Available from: https://handbook-5-1.cochrane.org/chapter_6/6_3_3_searching_central_medline_and_embase_specific_issues.htm.

127. Lam MT, McDiarmid M. Increasing number of databases searched in systematic reviews and meta-analyses between 1994 and 2014. J Med Libr Assoc. 2016;104(4):284-9.

128. Covidence [cited 2019 Oct 01]. Available from: https://www.covidence.org/.

129. Golder S, Loke YK. The contribution of different information sources for adverse effects data. Int J Technol Assess Health Care. 2012;28(2):133-7.

130. Lemeshow AR, Blum RE, Berlin JA, Stoto MA, Colditz GA. Searching one or two databases was insufficient for meta-analysis of observational studies. J Clin Epidemiol. 2005;58(9):867-73.

131. Beyer FR, Wright K. Can we prioritise which databases to search? A case study using a systematic review of frozen shoulder management. Health Info Libr J. 2013;30(1):49-58.

132. Kwon Y, Powelson SE, Wong H, Ghali WA, Conly JM. An assessment of the efficacy of searching in biomedical databases beyond MEDLINE in identifying studies for a systematic review on ward closures as an infection control intervention to control outbreaks. Syst Rev. 2014;3:135.

133. Bramer WM, Rethlefsen ML, Kleijnen J, Franco OH. Optimal database combinations for literature searches in systematic reviews: a prospective exploratory study. Syst Rev. 2017;6(1):245.

134. Goossen K, Tenckhoff S, Probst P, Grummich K, Mihaljevic AL, Buchler MW, et al. Optimal literature search for systematic reviews in surgery. Langenbecks Arch Surg. 2018;403(1):119-29.

135. Bramer WM, Giustini D, Kramer BM, Anderson P. The comparative recall of Google Scholar versus PubMed in identical searches for biomedical systematic reviews: a review of searches used in systematic reviews. Syst Rev. 2013;2:115.

136. Day D, Furlan A, Irvin E, Bombardier C. Simplified search strategies were effective in identifying clinical trials of pharmaceuticals and physical modalities. J Clin Epidemiol. 2005;58(9):874-81.

137. Rice DB, Kloda LA, Levis B, Qi B, Kingsland E, Thombs BD. Are MEDLINE searches sufficient for systematic reviews and meta-analyses of the diagnostic accuracy of depression screening tools? A review of meta-analyses. J Psychosom Res. 2016;87:7-13.

138. Rollin L, Darmoni S, Caillard JF, Gehanno JF. Searching for high-quality articles about intervention studies in occupational health--what is really missed when using only the Medline database? Scand J Work Environ Health. 2010;36(6):484-7.

139. Shariff SZ, Sontrop JM, Iansavichus AV, Haynes RB, Weir MA, Gandhi S, et al. Availability of renal literature in six bibliographic databases. Nephrology Dialysis Transplantation Plus. 2012;5(6):610-7.

140. Slobogean GP, Verma A, Giustini D, Slobogean BL, Mulpuri K. MEDLINE, EMBASE, and Cochrane index most primary studies but not abstracts included in orthopedic meta-analyses. J Clin Epidemiol. 2009;62(12):1261-7.

141. Hanneke R, Young SK. Information sources for obesity prevention policy research: a review of systematic reviews. Syst Rev. 2017;6(1):156.

142. Lorenzetti DL, Topfer LA, Dennett L, Clement F. Value of databases other than medline for rapid health technology assessments. Int J Technol Assess Health Care. 2014;30(2):173-8.

143. Michaleff ZA, Costa LO, Moseley AM, Maher CG, Elkins MR, Herbert RD, et al. CENTRAL, PEDro, PubMed, and EMBASE are the most comprehensive databases indexing randomized controlled trials of physical therapy interventions. Phys Ther. 2011;91(2):190-7.

144. Moseley AM, Sherrington C, Elkins MR, Herbert RD, Maher CG. Indexing of randomised controlled trials of physiotherapy interventions: a comparison of AMED, CENTRAL, CINAHL, EMBASE, hooked on evidence, PEDro, PsycINFO and PubMed. Physiotherapy. 2009;95(3):151-6.

145. Pilkington K. Searching for CAM evidence: an evaluation of therapy-specific search strategies. J Altern Complement Med. 2007;13(4):451-9.

146. Sood A, Sood R, Bauer BA, Ebbert JO. Cochrane systematic reviews in acupuncture: methodological diversity in database searching. J Altern Complement Med. 2005;11(4):719-22.

147. Preston L, Carroll C, Gardois P, Paisley S, Kaltenthaler E. Improving search efficiency for systematic reviews of diagnostic test accuracy: an exploratory study to assess the viability of limiting to MEDLINE, EMBASE and reference checking. Syst Rev. 2015;4(1):82.

148. Sampson M, McGowan J. Errors in search strategies were identified by type and frequency. J Clin Epidemiol. 2006;59(10):1057-63.

149. Bradley S, Giustini D, editors. GoPubMed versus PubReMiner for analyzing PubMed search results: a head to head comparison of two free web ‘data mining’ tools. CHLA/ABSC Conference; 2011: Canadian Health Libraries Association.

150. Bramer WM, Giustini D, Kleijnen J, Franco OH. Searching Embase and MEDLINE by using only major descriptors or title and abstract fields: a prospective exploratory study. Syst Rev. 2018;7(1):200.

151. O'Mara-Eves A, Brunton G, McDaid D, Kavanagh J, Oliver S, Thomas J. Techniques for identifying cross-disciplinary and ‘hard-to-detect’evidence for systematic review. Res Synth Methods. 2014;5(1):50-9.

152. Stansfield C, O'Mara-Eves A, Thomas J. Text mining for search term development in systematic reviewing: A discussion of some methods and challenges. Res Synth Methods. 2017;8(3):355-65.

153. Thompson J, Davis J, Mazerolle L. A systematic method for search term selection in systematic reviews. Res Synth Methods. 2013;5(2):87-97.

154. Waffenschmidt S, Janzen T, Hausner E, Kaiser T. Simple search techniques in PubMed are potentially suitable for evaluating the completeness of systematic reviews. J Clin Epidemiol. 2013;66(6):660-5.

155. Sampson MJ. Updating searches for systematic reviews [PhD]. Aberystwyth: Aberystwyth University; 2009.

156. Hartley J. To attract or to inform: what are titles for? Journal of Technical Writing and Communication. 2005;35(2):203-13.

157. Bavdekar SB. Formulating the right title for a research. J Assoc Physicians India. 2016;64:53.

158. Thomas J, Noel-Storr A, Marshall I, Wallace B, McDonald S, Mavergames C, et al. Living systematic reviews: 2. Combining human and machine effort. J Clin Epidemiol. 2017.

159. Beller E, Clark J, Tsafnat G, Adams C, Diehl H, Lund H, et al. Making progress with the automation of systematic reviews: principles of the International Collaboration for the Automation of Systematic Reviews (ICASR). Syst Rev. 2018;7(1):77.

160. O'Connor AM, Tsafnat G, Gilbert SB, Thayer KA, Wolfe MS. Moving toward the automation of the systematic review process: a summary of discussions at the second meeting of International Collaboration for the Automation of Systematic Reviews (ICASR). Syst Rev. 2018;7(1):3.

161. Young S, Duffull SB. A learning-based approach for performing an in-depth literature search using MEDLINE. J Clin Pharm Ther. 2011;36(4):504-12.

162. Sampson M, Shojania KG, McGowan J, Daniel R, Rader T, Iansavichene AE, et al. Surveillance search techniques identified the need to update systematic reviews. J Clin Epidemiol. 2008;61(8):755-62.

163. Cohen AM, Adams CE, Davis JM, Yu C, Yu PS, Meng W, et al. Evidence-based medicine, the essential role of systematic reviews, and the need for automated text mining tools Arlington, VA, USA2010 [cited 2019 Oct 30]. Available from: https://www.researchgate.net/profile/Neil_Smalheiser/publication/221629869_Evidence-based_medicine_the_essential_role_of_systematic_reviews_and_the_need_for_automated_text_mining_tools/links/00b7d53550ec2dfade000000/Evidence-based-medicine-the-essential-role-of-systematic-reviews-and-the-need-for-automated-text-mining-tools.pdf?origin=publication_detail.

164. Ramampiaro H, Cruzes D, Conradi R, Mendona M. Supporting evidence-based Software Engineering with collaborative information retrieval. In: Proceedings of the 6th International Conference on Collaborative Computing: Networking, Applications and Worksharing, CollaborateCom 2010; Chicago, IL, USA2010.

165. Sturm B, Schneider S, Sunyaev A, editors. Leave no stone unturned: introducing a revolutionary meta-search tool for rigorous and efficient systematic literature searches. ECIS - European Conference on Information Systems 2015; Münster, Germany: Association for Information Systems.

166. Bowes D, Hall T, Beecham S, editors. SLuRp: a tool to help large complex systematic literature reviews deliver valid and rigorous results. Proceedings of the 2nd international workshop on Evidential assessment of software technologies; 2012: ACM.

167. HTTP Status 404 - /SLRTool/ [cited 2019 Oct 01]. Available from: http://alarcosj.esi.uclm.es/SLRTool/.

168. Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG. Epidemiology and reporting characteristics of systematic reviews. PLoS Med. 2007;4(3):e78.

169. More about us: Cochrane Methods. Information Retrieval; [cited 2019 Aug 08]. Available from: https://methods.cochrane.org/irmg/more-about-us.

170. Information Retrieval Methods: Cochrane Methods. Information Retrieval; [cited 2019 Aug 08]. Available from: https://methods.cochrane.org/irmg/.

171. Aagaard T, Lund H, Juhl C. Optimizing literature search in systematic reviews - are MEDLINE, EMBASE and CENTRAL enough for identifying effect studies within the area of musculoskeletal disorders? BMC Med Res Methodol. 2016;16(1):161.

172. Adams SP, Tsang M, Wright JM. Lipid lowering efficacy of atorvastatin. Cochrane Database Syst Rev. 2012(12).

173. Leeflang MM, Scholten RJ, Rutjes AW, Reitsma JB, Bossuyt PM. Use of methodological search filters to identify diagnostic accuracy studies can lead to the omission of relevant studies. J Clin Epidemiol. 2006;59(3):234-40.

174. Beynon R, Leeflang MM, McDonald S, Eisinga A, Mitchell RL, Whiting P, et al. Search strategies to identify diagnostic accuracy studies in MEDLINE and EMBASE. Cochrane Database Syst Rev. 2013(9):Mr000022.

175. Forrester LT, Maayan N, Orrell M, Spector AE, Buchan LD, Soares-Weiser K. Aromatherapy for dementia. Cochrane Database Syst Rev. 2014(2):Cd003150.

176. PubMed Tutorial: Truncation: NIH. U.S. National Library of Medicine; [cited 2019 May 02]. Available from: https://www.nlm.nih.gov/bsd/disted/pubmedtutorial/020_460.html.

177. Greenhalgh T, Peacock R. Effectiveness and efficiency of search methods in systematic reviews of complex evidence: audit of primary sources. BMJ. 2005;331(7524):1064-5.

178. Controlled vocabulary and text words. In: The Cochrane Handbook: Version 510 [Internet]. The Cochrane Collaboration, [cited 2018 Oct 02]. Available from: https://handbook-5-1.cochrane.org/chapter_6/6_4_5_controlled_vocabulary_and_text_words.htm.

179. Kelly D. Methods for evaluating interactive information retrieval systems with users. Found Trends Inf Ret. 2009;3(1—2):1-224.

180. Cochrane Work. Registering a new title: Cochrane Collaboration; [cited 2017 Nov 23]. Available from: http://work.cochrane.org/registering-new-title.

181. Advanced Search [Ovid Tutorial]: Wolters Kluwer; 2019 [cited 2019 Jan 14]. Available from: http://resourcecenter.ovid.com/site/help/documentation/ospa/en/Content/advanced.htm.

182. PubMed Help: All Fields [ALL]: NCBI; [updated December 10, 2018; cited 2019 Feb 07]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK3827/#pubmedhelp.All_Fields_ALL.

183. Cochrane Highly Sensitive Search Strategy for identifying randomized trials in MEDLINE: sensitivity-maximizing version (2008 revision); PubMed format [cited 2017 Nov 23]. Available from: http://handbook-5-1.cochrane.org/chapter_6/box_6_4_a_cochrane_hsss_2008_sensmax_pubmed.htm.

184. Interventions. 2011. In: Cochrane Handbook for Systematic Reviews of Interventions Version 510 (updated March 2011) [Internet]. The Cochrane Collaboration, [cited 2019 Jul 31]. Available from: https://handbook-5-1.cochrane.org/chapter_7/7_3_4_interventions.htm.

185. Tsafnat G, Dunn A, Glasziou P, Coiera E. The automation of systematic reviews. BMJ. 2013;346:f139.

186. Winokur M, Holtan A, Batchelder KE. Kinship care for the safety, permanency, and well-being of children removed from the home for maltreatment. Cochrane Database Syst Rev. 2014(1):Cd006546.

187. Structure of a search strategy. [updated March 2011]. In: Cochrane Handbook for Systematic Reviews of Interventions [Internet]. The Cochrane Collaboration. Version 5.1.0 [cited 2019 Jan 14]. Available from: https://handbook-5-1.cochrane.org/chapter_6/6_4_2_structure_of_a_search_strategy.htm.

188. Seth R, Kydd AS, Buchbinder R, Bombardier C, Edwards CJ. Allopurinol for chronic gout. Cochrane Database Syst Rev. 2014(10):Cd006077.

189. Do DV, Wang X, Vedula SS, Marrone M, Sleilati G, Hawkins BS, et al. Blood pressure control for diabetic retinopathy. Cochrane Database Syst Rev. 2015;1:Cd006127.

190. Allingstrup M, Afshari A. Selenium supplementation for critically ill adults. Cochrane Database Syst Rev. 2015(7):Cd003703.

191. Ona XB, Osorio D, Cosp XB. Drug therapy for treating post-dural puncture headache. Cochrane Database Syst Rev. 2015(7).

192. Kristjansson E, Francis DK, Liberato S, Benkhalti Jandu M, Welch V, Batal M, et al. Food supplementation for improving the physical and psychosocial health of socio-economically disadvantaged children aged three months to five years. Cochrane Database Syst Rev. 2015(3):Cd009924.

193. Smith GA, Fisher SA, Doree C, Di Angelantonio E, Roberts DJ. Oral or parenteral iron supplementation to reduce deferral, iron deficiency and/or anaemia in blood donors. Cochrane Database Syst Rev. 2014(7):Cd009532.

194. Kugley S, Wade A, Thomas J, Mahood Q, Jørgensen A-MK, Hammerstrøm K, et al. Searching for Studies: A Guide to Information Retrieval for Campbell Systematic Reviews: The Campbell Collaboration; 2016 [cited 2019 Oct 01]. Available from: http://www.campbellcollaboration.org/images/Campbell_Methods_Guides_Information_Retrieval.pdf.

195. Jull AB, Cullum N, Dumville JC, Westby MJ, Deshpande S, Walker N. Honey as a topical treatment for wounds. Cochrane Database Syst Rev. 2015(3):Cd005083.

196. Hooper L, Martin N, Abdelhamid A, Davey Smith G. Reduction in saturated fat intake for cardiovascular disease. Cochrane Database Syst Rev. 2015(6):Cd011737.

197. Brassey J. New post: Testing the effectiveness of simplified search strategies for updating systematic reviews 2017 [cited 2019 Oct 01]. Available from: https://rapid-reviews.info/2017/11/03/new-post-testing-the-effectiveness-of-simplified-search-strategies-for-updating-systematic-reviews/.

198. Toon CD, Sinha S, Davidson BR, Gurusamy KS. Early versus delayed post-operative bathing or showering to prevent wound complications. Cochrane Database Syst Rev. 2015(7):Cd010075.

199. Liu J, Zhang P, Tian J, Li L, Li J, Tian JH, et al. Ozone therapy for treating foot ulcers in people with diabetes. Cochrane Database Syst Rev. 2015(10):Cd008474.

200. Prady SL, Uphoff EP, Power M, Golder S. Development and validation of a search filter to identify equity-focused studies: reducing the number needed to screen. BMC Med Res Methodol. 2018;18(1):106.

201. Budhram D, Navarro-Ruan T, Haynes RB. The efficiency of database searches for creating systematic reviews was improved by search filters. J Clin Epidemiol. 2018;95:1-6.

202. Golder S, Wright K, Loke YK. The development of search filters for adverse effects of surgical interventions in medline and Embase. Health Info Libr J. 2018;35(2):121-9.

203. 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 Technol Assess. 2017;21(69):1-148.

204. Lewin S, Hendry M, Chandler J, Oxman AD, Michie S, Shepperd S, et al. Assessing the complexity of interventions within systematic reviews: development, content and use of a new tool (iCAT_SR). BMC Med Res Methodol. 2017;17(1):76.

205. Moher D, Pham B, Lawson ML, Klassen TP. The inclusion of reports of randomised trials published in languages other than English in systematic reviews. Health Technol Assess. 2003;7(41):1-90.

206. Booth A. Over 85% of included studies in systematic reviews are on MEDLINE. J Clin Epidemiol. 2016;79:165-6.

207. van Ginneken N, Tharyan P, Lewin S, Rao GN, Meera SM, Pian J, et al. Non-specialist health worker interventions for the care of mental, neurological and substance-abuse disorders in low- and middle-income countries. Cochrane Database Syst Rev. 2013(11):Cd009149.

208. Porter AL, Huang Y, Schuehle J, Youtie J. Meta data: Big Data research evolving across disciplines, players, and topics: IEEE International Congress on Big Data; 2015 [cited 2019 Oct 01]. Available from: https://www.researchgate.net/publication/280529689_MetaData_BigData_Research_Evolving_Across_Disciplines_Players_and_Topics.

209. Petrova M, Sutcliffe P, Fulford KW, Dale J. Search terms and a validated brief search filter to retrieve publications on health-related values in Medline: a word frequency analysis study. J Am Med Inform Assoc. 2012;19(3):479-88.

210. Balan PF, Gerits A, Vanduffel W. A practical application of text mining to literature on cognitive rehabilitation and enhancement through neurostimulation. Front Syst Neurosci. 2014;8:182.

211. Toms EG. Task-based information searching and retrieval. In: Ruthven I, Kelly, D. , editor. Interactive Information Seeking, Behaviour and Retrieval. Facet Publishing; 2011. p. 43-59.

212. Pohl S, Zobel J, Moffat A. Extended Boolean retrieval for systematic biomedical reviews 2010 [cited 2019 Oct 01]. Available from: https://people.eng.unimelb.edu.au/jzobel/fulltext/acsc10.pdf.

213. Martinez D, Karimi S, Cavedon L, Baldwin T. Facilitating biomedical systematic reviews using ranked text retrieval and classification 2008 [cited 2019 Oct 01]. Available from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.142.2305&rep=rep1&type=pdf.

214. Jackson P, Moulinier I. Document retrieval. In: Natural language processing for online applications: text retrieval, extraction and categorization. Amsterdam: John Benjamins; 2007. p. 29.

215. Golder S, Loke Y, McIntosh HM. Poor reporting and inadequate searches were apparent in systematic reviews of adverse effects. J Clin Epidemiol. 2008;61(5):440-8.

216. Ecker ED, Skelly AC. Conducting a winning literature search. Evid Based Spine Care J. 2010;1(01):9-14.

217. Hjørland B. Classical databases and knowledge organization: A case for boolean retrieval and human decision-making during searches. J Assoc Inf Sci Technol. 2014.

218. Warner J. Introduction. In: Human information retrieval. Cambridge, Mass.: MIT press; 2010. p. 1-15.

219. Booth A. The number needed to retrieve: a practically useful measure of information retrieval? Health Info Libr J. 2006;23(3):229-32.

220. Booth A. How much searching is enough? Comprehensive versus optimal retrieval for technology assessments. Int J Technol Assess Health Care. 2010;26(4):431-5.

221. Elliott JH, Synnot A, Turner T, Simmonds M, Akl EA, McDonald S, et al. Living systematic review: 1. Introduction—the why, what, when, and how. J Clin Epidemiol. 2017;91:23-30.

222. Petticrew M. Time to rethink the systematic review catechism? Moving from 'what works' to 'what happens'. Syst Rev. 2015;4(1):36.

223. AMSTAR Checklist: A MeaSurement Tool to Assess Systematic Reviews; [cited 2019 Aug 23]. Available from: https://amstar.ca/Amstar_Checklist.php.

224. Lefebvre C, Glanville J, Wieland LS, Coles B, Weightman AL. Methodological developments in searching for studies for systematic reviews: past, present and future? Syst Rev. 2013;2:78.

225. Wiysonge CS, Kamadjeu R, Tsague L. Systematic reviews in context: highlighting systematic reviews relevant to Africa in the Pan African Medical Journal. Pan Afr Med J. 2016;24(1).

226. Jadad AR, Cook DJ, Jones A, Klassen TP, Tugwell P, Moher M, et al. Methodology and reports of systematic reviews and meta-analyses: a comparison of Cochrane reviews with articles published in paper-based journals. JAMA. 1998;280(3):278-80.

8 Appendices

8.1 Appendix I: Reviews not eligible for the cross sectional study

		Review groups	Zero studies included or indexed in MEDLINE	Not performed in Ovid	Not reproducible
1	ACECG	Anaesthesia Group	1	2	0
2	DCIG	Dementia and Cognitive Improvement Group	1	0	2
3	DPLPG	Developmental, Psychosocial and Learning Problems Group	5	0	0
4	EG	Epilepsy group	4	2	0
5	EPOC	Effective Practice and Organization of Care Group	3	5	5
6	EVG	Eyes and Vision Group	7	1	0
7	GFG	Gynaecology and Fertility Group	0	1	2
8	GNOCG	Gynaecological, Neuro-oncology and Orphan Cancer Group	6	5	1
9	HG	Heart Group	1	3	0
10	IG	Injuries Group	13	1	0
11	KTG	Kidney and Transplant Group	3	0	1
12	MG	Musculoskeletal Group	2	3	1
13	OHG	Oral Health Group	1	0	0
14	PPSG	Pain, Palliative and Supportive Care Group	2	1	0
15	SG	Stroke Group	0	1	4
16	WG	Wounds Group	6	0	1
		Sum	55	25	17

8.2 Appendix II: Reviews eligible for the exploratory study and the experiment

Number of studies included in the experiment:

	Cochrane Group	>4 incl. in MEDLINE	<5 incl. in MEDLINE	>50 incl. in MEDLINE	Over-views	With-drawn	Total # of incl. for exp.
1	ACECG	20	5				20
2	DCIG	16	9				16
3	DPLPG	17	7	2			16
4	EG	12	13				12
5	EPOC	19	6	1			18
6	EVG	13	12				13
7	GFG	19	6	2			17
8	GNOCG	12	13				12
9	HG	19	6	1		1	17
10	IG	13	11	1			13
11	KTG	17	8				17
12	MG	20	5				20
13	OHG	18	7	2			16
14	PPSG	17	8				17
15	SG	16	9				16
16	WG	15	10		1		14
	Sum	263	137	8	1	1	254

Number of reviews with 5-50 included studies indexed in MEDLINE: 254.

Number of reviews with less than 5 included studies indexed in MEDLINE: 137.

Number of reviews with more than 50 included studies indexed in MEDLINE: 8.

Number of overview of reviews: 1

Number of withdrawn reviews: 1

254 reviews eligible for the experiment.

8.3 Appendix III: Results of the cross sectional study

Cross sectional study – average results for the sample of 400 reviews

	Name of group	# of hits MEDLINE Ovid	# of hits without filter for study design	# of included studies	# of included studies in MEDLINE	# of studies in Embase if not in MEDLINE	# of studies not in MEDLINE or Embase	# of studies not found by MEDLINE search	# of studies not found without filter	# of databases searched
1	ACECG	871	6088	15	13.4	1.4	0.7	1.8	1.5	6.5
2	DCIG	2400	4424	8.0	7.2	0.5	0.6	0.5	0.9	12.2
3	DPLPG	5227	24808	21.3	14.2	1.1	6.3	1.8	0.8	14.1
4	EPOC	4090	19662	18.6	17.3	0.3	1.1	4.2	3.0	10.5
5	EG	468	2099	7.2	5.9	0.5	1.1	0.6	0.2	5.2
6	EVG	988	3791	7.2	6.4	0.6	0.4	0.3	0.3	7.5
7	GNOCG	3155	6296	10.5	9.2	1.8	0.4	2.2	0.6	7.2
8	GFG	467	4108	22.1	18.7	2.4	1.5	0.8	0.4	9.6

	Name of group	# of hits MEDLINE Ovid	# of hits without filter for study design	# of included studies	# of included studies in MEDLINE	# of studies in Embase if not in MEDLINE	# of studies not in MEDLINE or Embase	# of studies not found by MEDLINE search	# of studies not found without filter	# of databases searched
9	HG	2934	10522	25.0	21.6	1.9	1.9	3.0	1.8	8.5
10	IG	4147	12772	14.6	12.9	1.1	0.9	2.4	2.3	8.9
11	KTG	6304	-	18.2	14.6	1.4	2.6	1.2	-	5.9
12	MG	1274	3966	18.2	14.9	1.8	1.9	1.4	0.6	6.8
13	OHG	1220	6275	21.2	19.6	0.9	1.2	4.2	1.4	5.9
14	PPSG	1313	4705	15.6	13.1	1.2	1.4	1.8	1.8	4.7
15	SG	1904	9264	16.1	11.1	2.8	2.8	0.6	0.6	11.3
16	WG	572	1421	10.3	8.6	0.7	1.2	1.0	0.7	7.2
	Total - average	2326	8012	15.6	13.1	1.4	1.6	1.7	1.1	8.3
	Total - median	798	3485	9.5	8.0	1.0	0.0	0.0	0.0	7.0

8.4 Appendix IV: Scoping review I – search strategy

Database(s): **Ovid MEDLINE(R) and Epub Ahead of Print, In-Process & Other Non-Indexed Citations and Daily** 1946 to October 19, 2018

Search Strategy:

#	Searches	Results
1	"Review Literature as Topic"/	7534
2	exp Evidence-Based Practice/mt [Methods]	6303
3	Technology Assessment, Biomedical/	9410
4	Meta-Analysis as Topic/	16473
5	"review literature as topic".kf.	110
6	((systemic* or systematic*) adj2 review*).ab,kf.	109334
7	((systemic* or systematic*) adj2 reviews).ti.	3112
8	((summari* or synthes*) adj2 (research or evidence or knowledge or literature)).mp.	28453
9	((evidence or scoping) adj review*).tw,kf.	5450
10	(knowledge adj (base or bases)).tw,kf.	7669
11	((technolog* adj2 assessment*) or hta).tw,kf.	8005
12	(meta-analys?s or metaanalys?s).tw,kf.	134689
13	evidence-based.tw,kf.	97226
14	or/1-13 [Research syntheses]	346498
15	"Information Storage and Retrieval"/st	973
16	((database* or source* or resource*) adj3 (additional* or contribut* or coverage or efficiency or impact or indexed or number or performance)).tw,kf.	23973
17	restrict* search*.tw,kf.	198
18	(searching or information retrieval).ti.	5927
19	(search* and (coverage or recall or Precision)).ti.	101

20	(included adj2 studies).ti,kf.	29
21	or/15-20	31056
22	Databases, Bibliographic/	5597
23	(Embase or Medline or PubMed).tw,kf.	162279
24	23 or 23	162279
25	14 and 21 and 24	877
26	limit 25 to yr="2005 -Current"	799

8.5 Appendix V: Scoping review II – search strategy

Database(s): **Ovid MEDLINE(R) and Epub Ahead of Print, In-Process & Other Non-Indexed Citations and Daily** 1946 to December 06, 2018

Search Strategy:

#	Searches	Results
1	"Review Literature as Topic"/	7585
2	exp Evidence-Based Practice/mt [Methods]	6357
3	Technology Assessment, Biomedical/	9444
4	Meta-Analysis as Topic/	16561
5	"review literature as topic".kf.	111
6	((systemic* or systematic*) adj2 review*).ab.	108167
7	((systemic* or systematic*) adj2 reviews).ti,kf.	3728
8	((summari* or synthes*) adj2 (research or evidence or knowledge or literature)).mp.	28930
9	((evidence or scoping) adj review*).tw.	5598
10	(knowledge adj (base or bases)).tw.	7699
11	((technolog* adj2 assessment*) or hta).tw.	7633
12	(meta-analys?s or metaanalys?s).tw.	135936

13	evidence-based.tw.	95783
14	or/1-13 [Research syntheses]	347370
15	(search* adj2 term? adj2 select*).tw.	114
16	(simpl* adj3 search*).tw.	940
17	(search adj2 strateg*).ti.	476
18	(searches or searching).kf.	276
19	Boolean search*.tw,kf.	113
20	or/15-19	1894
21	14 and 20	309
22	limit 21 to yr="2005 -Current"	277
23	limit 22 to (danish or english or norwegian or swedish)	273

8.6 Appendix VI: Scoping review III – search strategy

Database(s): **Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R)** 1946 to Present

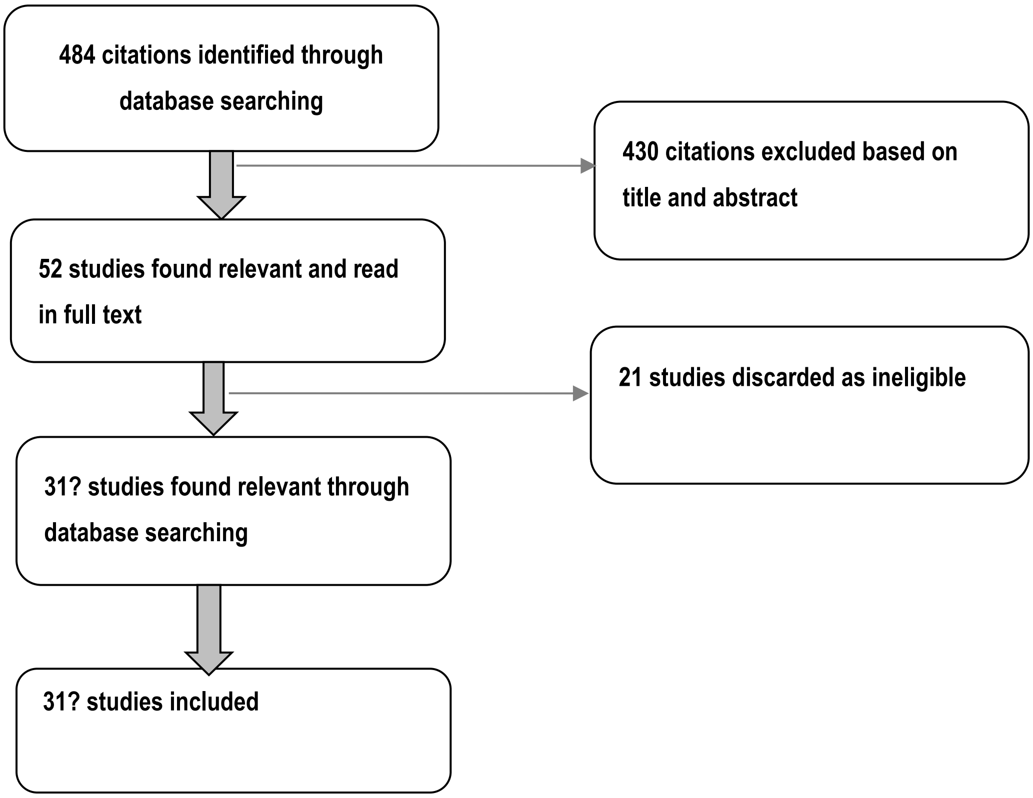
#	Searches	Results
1	"Review Literature as Topic"/	6110
2	exp Evidence-Based Practice/mt [Methods]	4982
3	Technology Assessment, Biomedical/	8504
4	Meta-Analysis as Topic/	14637
5	"review literature as topic".kf.	53
6	((systemic* or systematic*) adj2 review*).ab.	65446
7	((systemic* or systematic*) adj2 reviews).ti.	2176
8	((summari* or synthes*) adj2 (research or evidence or knowledge or literature)).mp.	19590
9	((evidence or scoping) adj review*).tw.	2725

10	(knowledge adj (base or bases)).tw.	6332
11	((technolog* adj2 assessment*) or hta).tw.	5767
12	(meta-analys?s or metaanalys?s).tw.	86518
13	evidence-based.tw.	70341
14	or/1-13 [Research syntheses]	238946
15	"information storage and retrieval"/	17369
16	((article* or document* or information) adj3 (retriev* or seek* or search*)).mp.	41270
17	((systemic* or systematic*) adj2 search*).tw.	15833
18	(alternativ* adj2 search*).tw.	1422
19	((100% or high or total) adj recall).tw.	337
20	((captur* or find or finding or identif* or locat* or retriev* or search*) adj4 ((all or as many or as much or developed or primary or relevant or total) adj4 (articles or documents or evidence or information or knowledge or literature or research or studies))).tw.	20814
21	((approach* or complex or comprehensiv* or extensive or literature or method* or strateg* or technique*) adj2 search*).mp.	66499
22	(identif* adj2 data).tw.	15250
23	online search*.tw.	686
24	search*.ti. /freq=2	347
25	search*.ab. /freq=6	2952
26	or/15-25 [Information retrieval (in general)]	141304
27	Data mining/	4286
28	Natural Language Processing/	2963
29	Word sense disambiguation.kf.	10
30	Text-mining.kf.	171
31	(rank* adj5 (quer* or result* or retriev* or search*)).mp.	4537
32	(textmining or ((text or literature) adj2 mining)).mp.	1702
33	(natural* adj3 language* adj3 process*).tw.	1388

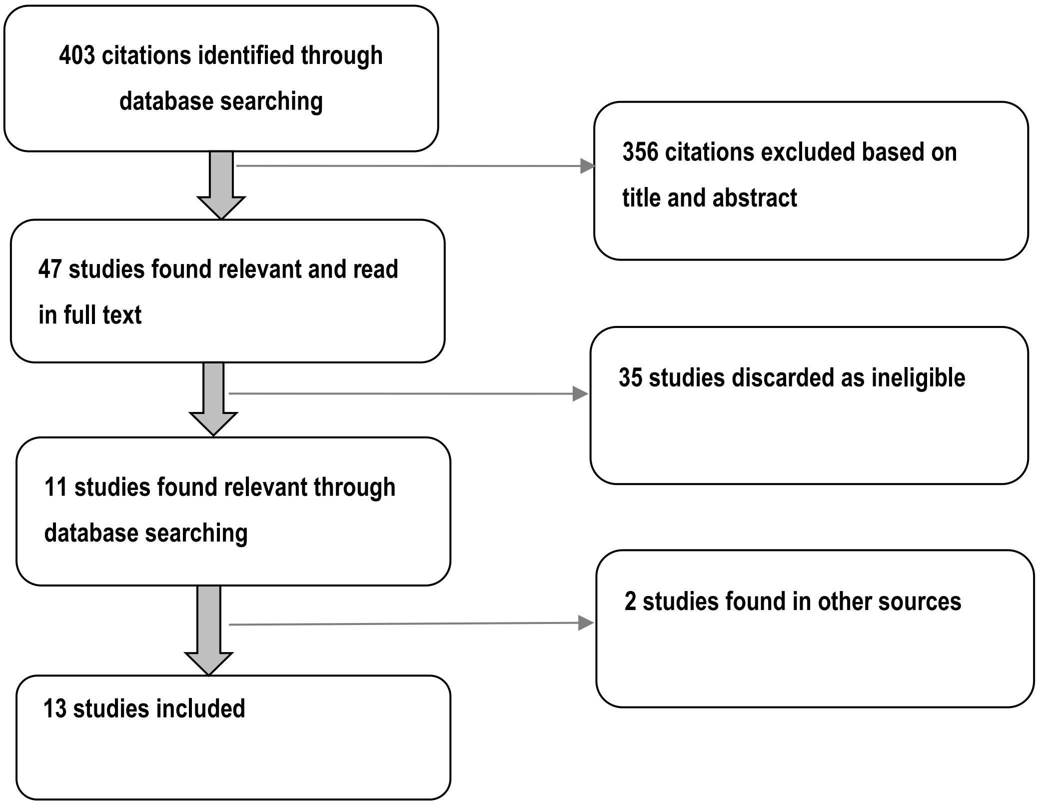
34	text analys*.tw.	547
35	search tool*.tw.	1253
36	(automatic* adj3 term? adj3 recogni*).tw.	11
37	machine-learning.tw.	7207
38	(Google Scholar or MEDLINE or PubMed).ti.	1384
39	named entity recognition.tw.	160
40	machine-assist*.tw.	72
41	term association*.tw.	522
42	(automat* adj3 information adj3 extract*).tw.	140
43	capture-mark-recapture.tw.	201
44	or/27-43 [Alternative IR techniques]	23143
45	14 and 26 and 44	402
46	*Review Literature as Topic/	2752
47	*"Information Storage and Retrieval"/	9755
48	search*.ti.	26769
49	46 and (47 or 48)	213
50	Comprehensive literature search for SRs.kf.	1
51	(automat* adj5 ((systemic* or systematic*) adj2 review*)).tw.	24
52	*Data Mining/mt	1548
53	(text mining or textmining).ti.	322
54	52 or 53	1763
55	14 and 54	105
56	26 and 54	322
57	55 or 56	395
58	methods.fs.	2948161
59	15 and 27 and 58	137

60	or/45,49-51,57,59	1013
61	remove duplicates from 60	998
62	limit 61 to yr="2005 -Current"	870
63	limit 62 to (danish or english or norwegian or swedish)	857

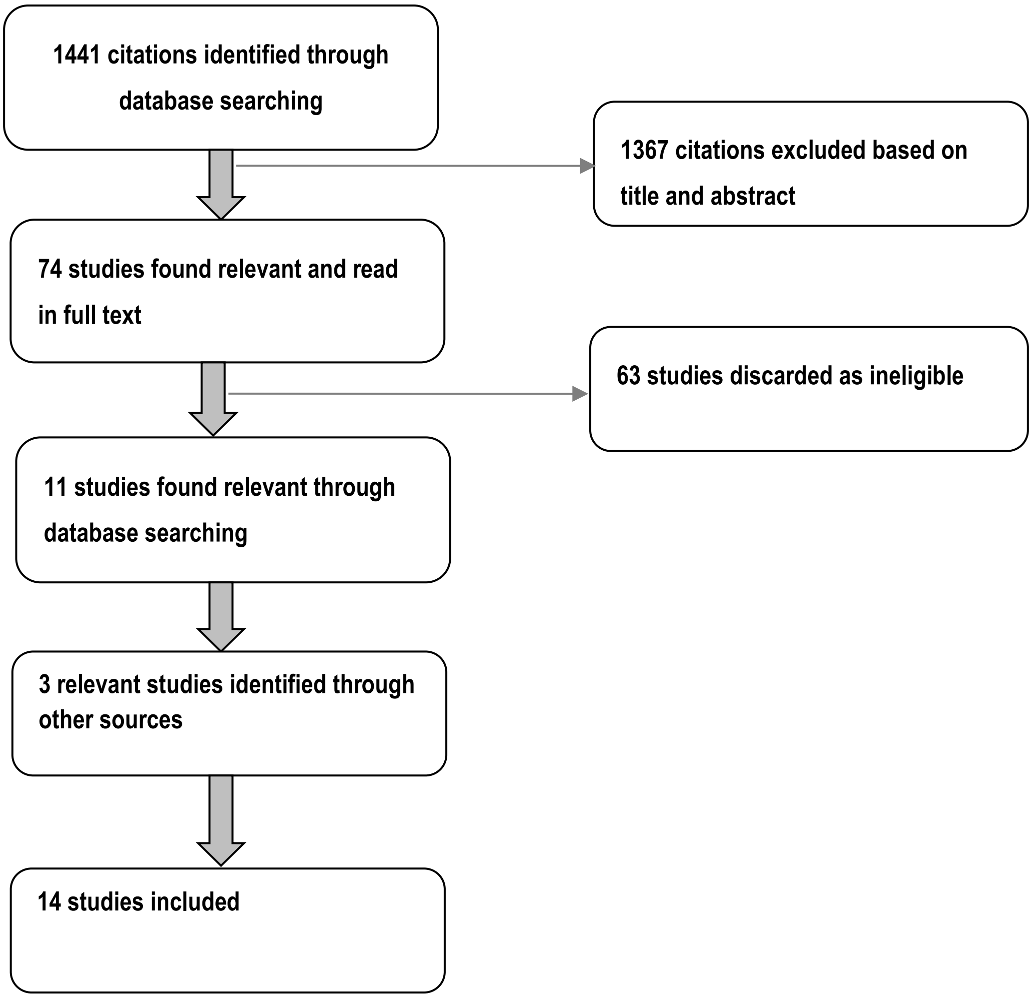
8.7 Appendix VII: PRISMA flow diagram on study identification: Scoping review I



8.8 Appendix VIII: PRISMA flow diagram on study identification: Scoping review II



8.9 Appendix IX: PRISMA flow diagram on study identification - Scoping review III



Glossary

Bias: flaws in a study that can lead to invalid conclusions.

Case study: a detailed investigation of one unit, here: one systematic review.

Cochrane Collaboration: an international collaboration that develops and maintains systematic reviews in healthcare.

Cross sectional study: a type of observational study that analyses data from a population, or a representative subset, at a specific point in time.

Effectiveness: the degree to which something is successful in producing a desired result.

Efficiency: the ability to avoid wasting materials, energy, efforts, money, and time in doing something or in producing a desired result.

Empirical research: is based on observed and measured phenomena and derives knowledge from actual experience rather than from theory or belief.

Experiment: a scientific procedure undertaken to make a discovery, test a hypothesis, or demonstrate a known fact

Exploratory study: studies conducted when little is known about a particular phenomenon. Employs a variety of research methods with the goal of learning more about a phenomenon, rather than making specific predictions.

Extended Boolean (information) retrieval: an intermediate between the Boolean system of query processing and the vector processing model.

Evidence Based Practice (EBP): practice that is based on the best available evidence, moderated by patient preferences.

Evidence selection bias: occurs when a systematic review does not identify all available data on a topic.

Leximancer: a semantic concept recognition software.

Precision: the fraction of retrieved documents that is relevant and is calculated as the number of relevant retrieved records divided by the total amount of retrieved records.

Protocol: a plan to be followed in a study and should describe the rationale for the review, the objectives and the methods that will be used to locate, select and critically appraise studies, and to collect and analyse data from the included studies.

Proximity (adjacency) operator: a Boolean operator that searches for terms near each other with a specified number of words between the search terms.

Publication bias: the likelihood of a study being published based on the findings of the study.

Randomized controlled trial (RCT): a trial that has randomly assigned groups in order to determine the effectiveness of an intervention.

Recall: the fraction of relevant documents that has been retrieved and is used to evaluate the ability of a search to identify all relevant studies.

Reliability: degree to which a research method produces stable and consistent results.

Research question: an answerable inquiry into a specific concern or issue and the initial step in a research project.

Search strategy: a predefined plan for searching for information or research on a topic.

Search term: a single word or phrase that is part of a search strategy.

Subject heading: controlled lists of terms or phrases used to describe the subjects of records in a database.

Systematic review: a review of the literature that is undertaken according to a defined and systematic approach.

Termine: a term extraction program that identifies the key terms and compound terms in a body of text and produces a list of terms in order of their frequency and significance of term occurrence.

Text word: words in title and abstract in a bibliographic record.

Validity: the extent to which a study or an intervention measures what it is intended to measure.