Serving Evidence Syntheses
Improving literature retrieval in systematic reviews

Wichor M. Bramer

Serving Evidence Syntheses
Improving literature retrieval in systematic reviews

Wichor M. Bramer

ISBN/EAN: 978-94-6380-541-4

Cover design and lay-out: Astrid Sibbes

Printed by: Global Academic Press

Serving Evidence Syntheses
Improving literature retrieval in systematic reviews

Wetenschappelijk bewijs verzamelen ondersteunen
optimalisatie van zoekacties in systematic reviews

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de rector magnificus
Prof.dr. R.C.M.E. Engels
en volgens het besluit van het College voor Promoties.
De openbare verdediging zal plaatsvinden op

dinsdag 29 oktober 2019 om 13.30 uur

door

Wichor Matthijs Bramer

geboren te Vriezenveen

Erasmus University Rotterdam

promotoren
Prof.dr. O.H. Franco Duran
Prof.dr. J.M.P. Kleijnen

overige leden
Prof.dr. M.G.M. Hunink
Prof.dr. L.R. Arends
Prof.dr. R.J.P.M. Scholten

copromotor
Dr. F. Mast

Contents

Chapter 1

Introduction

1 Introduction

1.1 Systematic reviews are the foundation of evidence-based medicine

The practice of medicine ought to be based on judicious use of evidence found in the literature. Literature reviews can provide a clear view of current evidence by presenting an overview of the literature on a certain topic, thereby reducing the need for readers to read many individual studies and enhancing practitioners' decision-making processes.

The two main types of literature reviews are narrative reviews and systematic reviews. Narrative reviews are descriptive in nature and are not designed to include all relevant literature on a given topic. They are often based on the literature that is known and available to the authors, which means they are subject to selection bias and authors' individual opinions. Systematic reviews on the other hand are performed in a systematic scientific manner, minimizing selection bias and guaranteeing comprehensive evaluation of a topic. They combine the results of research from multiple papers and make use of specific methods that are clearly described and repeatable. Their goal is to provide a comprehensive summary of the literature that is relevant to a certain research question. The findings of systematic reviews are therefore considered more reliable than those of narrative reviews, and they are used to draw conclusions and make decisions (1-3).

A common misconception among researchers is that systematic reviews should always include a meta-analysis and be based on randomized controlled trials. While this is often true for systematic reviews on therapeutic topics, which may also include a meta-analysis, it is not a standard requirement. One of the librarian's tasks is therefore to ensure that researchers' understanding of reviews matches standard definitions (4, 5). Because many journal editors and peer reviewers are also unaware of these standard definitions, many reviews that are not really systematic in nature are published as systematic reviews. Given that this is misleading for readers, more effort should be made to teach authors and researchers what systematic reviews are and how they should be designed, executed and read (6). If the overall quality of systematic reviews improves, the overall quality of the treatments based on those reviews may also improve.

1.2 Systematic reviews should be based on good literature searches

One of the key features of a systematic review is a systematic search based on clearly set objectives and research questions (3). This search should be both comprehensive and exhaustive (7). Simply typing in a few basic terms in PubMed will not deliver good search results due to problems in PubMed's automatic term mapping which seems to favor journal titles over contextual searches which may result in poor sensitivity (8). This feature is still present in PubMed and generates heterogeneous results for seemingly similar search strategies. Comparisons of search strategies have shown that simple queries will miss relevant papers even though such queries at first seem appropriate (9). Good quality searches minimize the risk of missing important studies. Indeed, research has shown that two concurrent reviews on the same topic can include different references due to differences in search strategies and inclusion criteria, as well as errors in the inclusion process (10).

1.3 What is a good literature search?

In their working paper on designing search methods, Sturm and Sunyaev have identified three main criteria for a good literature search (11). The first criterion is sensitivity (or – as they call it – comprehensiveness), i.e. the percentage of the total number of relevant references retrieved by the search. Because the literature on a certain topic is unknown, sensitivity is hard to determine. The second criterion is precision, i.e. how many of the retrieved results are relevant. Depending on the goal of the search, the searcher's main priority will be either precision or sensitivity. When conducting a search for a systematic review, the search usually aims to optimize sensitivity, whereas for searches in which it is less important to find all relevant references, the focus can be more on precision (12). The third and final criterion is reproducibility, i.e. whether the search is described in such a way that others can repeat it. However, the reproducibility of many systematic reviews is poor: many systematic reviews mention the terms used in the search only briefly and fail to provide the full search strategy, making it impossible to replicate the search. Others present the search strategy within just one database, making it impossible to repeat the search in other databases (13-17).

1.4 Good literature searches can only be performed by experienced searchers

Given the misconceptions regarding literature searches and the poor quality of some reviews, it is certainly advantageous for expert searchers such as librarians to be involved in searches conducted within the field of medicine. In 2005, the Medical Library Association in the United States issued a policy statement on the role of expert searching in medical libraries (18). This statement indicates that expert searchers – apart from having basic subject knowledge and project-organizing skills – need many other skills that include knowledge of the formulation of so-called information needs, as well as knowledge of database contents and syntax. Although various end-user interfaces are now available to non-librarians, medical librarians simply have more experience in searching databases than most health professionals and medical research staff. The fact that most researchers looking for simple information can now find their own literature means that the medical librarian's role has shifted towards that of a consultant who can give advice on more complicated projects such as systematic reviews, guidelines and grants (19).

Librarians play an important and complementary role in the process of systematic review creation (20). Not only do they carry out searches for reviews and teach search methods to others, they can also help researchers to formulate good research questions and manage bibliographic data. Although a 1995 review indicated that patient outcomes and care can be improved as a result of searches carried out by end users (i.e. non-librarians) (21), and a 1988 comparison found that searches by a non-librarian were as useful as those performed by a librarian (22), not every end-user search is successful. The results of more recent studies suggest that researchers themselves are not very good at searching (23) and that health professionals perform better on searching tasks when assisted by librarians (24).

The involvement of librarians in the systematic review process clearly helps to improve the quality of systematic searches and the reporting of search strategies (25-28). It has therefore often been argued that all systematic review projects should have a librarian involved (29-32).

1.5 The quality of searches can still be improved

The quality of systematic searches by librarians can be improved. For example, guidelines for systematic reviews often recommend that search strategies undergo peer review (33). However, it can be difficult for other searchers to identify problems or assess the quality of complex searches if they have not created them themselves (34). Assessing the quality of a search can be a challenge both for the search developer/creator and for others.

The quality of systematic reviews can also be improved by increasing the number of databases searched (35). Nevertheless, a 2014 study indicated that the number of databases searched in systematic reviews of economic evaluations in health care rarely conforms to widely accepted standards (36). Whether these standard are in fact enough is still topic of debate. In addition, many studies have demonstrated numerous flaws in the quality of searches: completeness and adherence to standards are frequently suboptimal (17, 37-40); search reproducibility is not always optimal (13, 16, 41, 42); searches are poorly reported (15, 43, 44); and the overall quality of the searches can be improved (45-47). For reviews on a similar topic, differences between searches have been shown to cause discrepancies between results in terms of the references included in those reviews, which could lead to different conclusions and ultimately differences in care (48).

1.6 There is currently no complete guidance on how to perform a search

Numerous guidelines for conducting systematic reviews have been published to address such shortcomings. Nevertheless, a literature review of such guidance for literature searches in systematic reviews has shown that although these documents help in the creation and planning of a search, there are no clear step-by-step descriptions on how to create search strategies (49). Others have also found there to be few complete guidelines that assist investigators in the process of creating search strategies for systematic reviews (50, 51). Attempts to create guides

on searching have been superficial and provide search examples that are not really systematic or exhaustive (52). Although one handbook describes what a search strategy for a systematic review ultimately should look like (3), this cannot be used while creating a search strategy.
The various search styles used by searchers were described by Booth in 2008 (53). The traditional method is described as building blocks: elements (distinct concepts in a research question) are described by various synonyms, which are searched in the traditional databases. What Booth describes as citation pearl growing is a method that uses precise search terms to find key relevant publications, which are then used to find more relevant terms that can be added to the strategy. If the number of references retrieved is too small, a solution can be to drop an element or facet. In the most specific facet first method, Booth describes how sometimes only one or two elements retrieve a set of search results that is concise enough for the researchers to review. Other methods described by Booth include related articles, successive fractions, and berry picking.

The current lack of standards means that many aspects of searching remain unclear. The existing literature does not even agree which databases should be searched – most current research focuses on the coverage of databases. However, just because an article is present in a certain database does not mean it will actually be retrieved in a search strategy for a systematic review (54, 55). A researcher might search a certain database, but his search strategy might still fail to retrieve all relevant articles present in that database. Because of this, the number of databases and the specific databases that were ultimately used varies (56-58).

Retrieving a relevant reference depends on many factors. For example, if certain databases are not available to the searcher, or are not used, this can affect the completeness of the search. In addition, while a relevant reference might be present in the databases that are used, incompleteness or incorrectness in terms of keywords, abstract or title can prevent the article from being found by the search. Finally, search engine characteristics such as the absence of proximity operators in PubMed can also prevent retrieval of an article. The quality of the search and the experience of the searcher are therefore not the only factors.

1.7 The aim of this thesis

The first aim of this thesis is to provide librarians developing searches for systematic reviews with a clearly described method that has been compared with other methods. We also aim to describe a method for deduplication in EndNote, and for updates to searches. We aim to provide researchers performing systematic reviews with tools that help them carry out systematic reviews, especially in terms of the inclusion and exclusion of articles based on the relevance of titles and abstracts, and reviewing the reference lists of relevant references. Finally, we aim to give advice to librarians and researchers using our method – or other methods – on database choice and on the use of major thesaurus terms.

1.8 The research in this thesis

This thesis starts by describing a method that we have developed for creating searches for systematic reviews and for managing references using the reference software tool EndNote. Chapter 2 gives a detailed description of how to create search strategies for systematic reviews from start to finish: from analyzing research questions, determining important concepts from research questions, and identifying search terms, to combining them into a search strategy, optimizing the strategy to find all potentially relevant references, and using macros to translate the search strategy for use in other databases.

Chapter 3 describes a method for deduplicating references found after aggregating searches from different databases. Such deduplication is often complicated because similar articles can be described in different ways in different databases. Using several combinations of field names, starting from specific to more general, deduplication can be done in a way that is much faster than traditional or manual approaches and with similar sensitivity and accuracy.

Chapter 4 describes a method for screening the references retrieved by the search process in EndNote. Using groups (a standard feature in EndNote), we code references for inclusion or exclusion after which they can be compared between two or more reviewers.

Chapter 5 describes how EndNote can be used to update searches performed earlier using deduplication to remove references retrieved in earlier searches from newly retrieved references.

In Chapter 6 we describe how EndNote can be used together with Scopus or Web of Science to download reference lists – both of the references included in the review and those of relevant reviews – for subsequent screening.

In Chapter 7, the results obtained using the methods described in this thesis are compared with those obtained in other reviews that used other search methods. We compared our search method with other methods used in other reviews, focusing on three main variables: the speed of the search method, the number of references retrieved for title and abstract screening and full text evaluation, and the number of references that were ultimately included in the published reviews.

Chapter 8 was written in response to a study conducted by Gehanno et al. in 2013 that analyzed a large group of reviews and found all of the references included in those reviews to be present in Google Scholar; from which the authors concluded that the original review authors could have found all relevant references had they used Google Scholar only (59). We tested this hypothesis by selecting published systematic reviews from other institutes and determining whether we could find all of the relevant references using the original search strategies as reported for PubMed and Google Scholar.

Chapter 9 describes a follow-up experiment to that described in Chapter 8. Where in 8 we studied existing reviews that had not involved researchers from our institute – mostly without the assistance of medical librarians and therefore resulting in low-quality searches – in Chapter 9 we compared the results of searches we had performed ourselves in three databases, namely Embase, MEDLINE and Google Scholar. For searches carried out by an experienced medical librarian in these three databases, we determined database coverage and database retrieval.

As a single database cannot find all relevant references, Chapter 10 describes how we determined the most optimal combination of databases when searching for relevant references to be included in systematic reviews. Although the Cochrane Handbook's (3) general recommendation to search Embase, MEDLINE and Cochrane CENTRAL is often used as guidance for all systematic reviews, we wanted to know whether this results in the retrieval of all possible relevant references. Since a researcher deciding which databases to search does not perform multiple reviews, an average or overall conclusion cannot inform their decision. The reviewer wants to know the likelihood that using a certain combination of databases will retrieve all relevant references for their review. We calculated the probability that a review will retrieve an acceptable percentage of relevant references when using a certain database combination.

Chapter 11 describes how we determined whether it is possible to reduce the number of references for screening by limiting searches in Embase and MEDLINE to major thesaurus terms, instead of all thesaurus terms, or by limiting the search to terms in title and abstracts only. Although other studies have been done on this topic, the conclusions of those studies were mostly related to how limiting the search affects recall in a single database, rather than in the combined results of different databases. However, if a certain database search limited to major thesaurus terms misses a reference, the investigator performing the systematic review might still find that same reference in other databases. We therefore looked only at those references that were uniquely found by Embase or MEDLINE, as these are the only major databases in which searches can be limited in that way.

In Chapter 12 we discuss the results of this thesis in general and in comparison with those of other studies and other research on the topic. We identify the strengths and weaknesses in our research, mention alternative solutions for the problems described, and summarize critiques of the method. We also consider how the use of our new method would have affected the systematic reviews that we conducted in the past, and how it will affect the future of the field. Finally, we discuss the potential pitfalls of the method, and present our views on the future of systematic searches of the literature.

1.9 References

1. Oxman AD, Guyatt GH. The science of reviewing research. Ann N Y Acad Sci. 1993;703(1):125-34.
2. Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts: treatments for myocardial infarction. JAMA. 1992;268(2):240-8.
3. Higgins JPT, Green S. Cochrane handbook for systematic reviews of interventions: Wiley Online Library; 2008.
4. Foster M, Halling TD, Pepper C. Systematic reviews training for librarians: planning, developing and evaluating. Journal of EAHIL. 2018;14(1):4-8.
5. Foster MJ. An overview of the role of librarians in systematic reviews: from expert search to project manager Journal of EAHIL. 2015;11(3):3-7.
6. Haddaway NR, Land M, Macura B. “A little learning is a dangerous thing”: A call for better understanding of the term ‘ systematic review’. Environ Int. 2017;99:356-60.
7. 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 2003/01/22;7(1):76.
8. Smith AM. An examination of PubMed’s ability to disambiguate subject queries and journal title queries. J Med Libr Assoc. 2004 Jan;92(1):97-100.
9. Bramer WM. Expert Searching: Why Don’t Simple Searches [Always] Work? MLA news. 2017;57(150).
10. Rosen L, Suhami R. The art and science of study identification: a comparative analysis of two systematic reviews. BMC Med Res Methodol. 2016 02/24;16:24.
11. Sturm B, Sunyaev A, editors. If you want your research done right, do you have to do it all yourself? Developing design principles for systematic literature search systems. Designing the Digital Transformation: DESRIST 2017 Research in Progress Proceedings of the 12th International Conference on Design Science Research in Information Systems and Technology Karlsruhe, Germany 30 May-1 Jun; 2017: Karlsruher Institut für Technologie (KIT).
12. Ahmadvand A, Hadjibabaie M, Gholami K. “Precision” or “relevance”? Reporting considerations for articles focusing on evaluation of search strategies. J Clin Epidemiol. 2013;66(3):351-2.
13. Koffel JB, Rethlefsen ML. Reproducibility of Search Strategies Is Poor in Systematic Reviews Published in High-Impact Pediatrics, Cardiology and Surgery Journals: A Cross-Sectional Study. PLoS One. 2016;11(9):e0163309.
14. Faggion CM, Jr., Huivin R, Aranda L, Pandis N, Alarcon M. The search and selection for primary studies in systematic reviews published in dental journals indexed in MEDLINE was not fully reproducible. J Clin Epidemiol. 2018 Feb 21.
15. Mullins MM, DeLuca JB, Crepaz N, Lyles CM. Reporting quality of search methods in systematic reviews of HIV behavioral interventions (2000-2010): are the searches clearly explained, systematic and reproducible? Res Synth Methods. 2014 Jun;5(2):116-30.
16. Maggio LA, Tannery NH, Kanter SL. Reproducibility of literature search reporting in medical education reviews. Acad Med. 2011;86(8):1049-54.
17. Mullins MM, DeLuca JB, Crepaz N, Lyles CM. Reporting quality of search methods in systematic reviews of HIV behavioral interventions (2000–2010): are the searches clearly explained, systematic and reproducible? Research Synthesis Methods. 2014;5(2):116-30.
18. Medical Library Association. Role of expert searching in health sciences libraries. J Med Libr Assoc. 2005;93(1):42-4.
19. Cruse P, Protzko S. Librarian contributions to clinical practice guidelines. Med Ref Serv Q. 2014;33(3):327-34.
20. Spencer AJ, Eldredge JD. Roles for librarians in systematic reviews: a scoping review. J Med Libr Assoc. 2018 Jan;106(1):46-56.
21. McKibbon K, Walker-Dilks C. The quality and impact of MEDLINE searches performed by end users. Health Libr Rev. 1995;12(3):191-200.
22. Wanke LA, Hewison NS. Comparative usefulness of MEDLINE searches performed by a drug information pharmacist and by medical librarians. Am J Hosp Pharm. 1988;45(12):2507-10.
23. Rosman T, Mayer AK, Krampen G. On the pitfalls of bibliographic database searching: comparing successful and less successful users. Behav Inf Technol. 2016 2016/02/01;35(2):106-17.
24. Gardois P, Calabrese R, Colombi N, Deplano A, Lingua C, Longo F, et al. Effectiveness of bibliographic searches performed by paediatric residents and interns assisted by librarians. A randomised controlled trial. Health information and libraries journal. 2011;28(4):273-84.
25. Rethlefsen ML, Farrell AM, Osterhaus Trzasko LC, Brigham TJ. Librarian co-authors correlated with higher quality reported search strategies in general internal medicine systematic reviews. J Clin Epidemiol. 2015 Jun;68(6):617-26.
26. Rethlefsen ML, Murad MH, Livingston EH. Engaging medical librarians to improve the quality of review articles. JAMA. 2014 Sep 10;312(10):999-1000.
27. Koffel JB. Use of Recommended Search Strategies in Systematic Reviews and the Impact of Librarian Involvement: A Cross-Sectional Survey of Recent Authors. PLoS One. 2015;10(5):e0125931.
28. Meert D, Torabi N, Costella J. Impact of librarians on reporting of the literature searching component of pediatric systematic reviews. Journal of the Medical Library Association : JMLA. 2016;104(4):267-77.
29. Wade CA, Turner HM, Rothstein HR, Lavenberg JG. Information retrieval and the role of the information specialist in producing high-quality systematic reviews in the social, behavioural and education sciences. Evidence & Policy: A Journal of Research, Debate and Practice. 2006;2(1):89-108.
30. Metzendorf M-I. Why medical information specialists should routinely form part of teams producing high quality systematic reviews – a Cochrane perspective . Journal of EAHIL. 2016;12(4):6-9.
31. McGowan J, Sampson M. Systematic reviews need systematic searchers. J Med Libr Assoc. 2005 Jan;93(1):74-80.
32. McGowan J. For expert literature searching, call a librarian. Can Med Assoc J. 2001;165(10):1301-2.
33. 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.
34. 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. Evidence Based Library and Information Practice. 2011;6(4):72-87.
35. 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.

36. Wood H, Arber M, Glanville J. The quality of search methodology and search reporting in published systematic reviews of economic evaluations: Search sources. Value Health. 2014;17(7):A580.
37. Wood H, Arber M, Glanville JM. Systematic Reviews of Economic Evaluations: How Extensive Are Their Searches? Int J Technol Assess Health Care. 2017 Mar 27:1-7.
38. Layton D. A Critical Review of Search Strategies Used in Recent Systematic Reviews Published in Selected Prosthodontic and Implant-Related Journals: Are Systematic Reviews Actually Systematic? Int J Prosthodont. 2017 Jan/Feb;30(1):13-21.
39. Yaylali IE, Alacam T. Critical Assessment of Search Strategies in Systematic Reviews in Endodontics. J Endod. 2016 Jun;42(6):854-60.
40. Faggion CM, Jr., Atieh MA, Park S. Search strategies in systematic reviews in periodontology and implant dentistry. J Clin Periodontol. 2013;40(9):883-8.
41. Flores-Mir C, Major MP, Major PW. Search and selection methodology of systematic reviews in orthodontics (2000-2004). Am J Orthod Dentofacial Orthop. 2006 Aug;130(2):214-7.
42. Major MP, Major PW, Flores-Mir C. An evaluation of search and selection methods used in dental systematic reviews published in English. J Am Dent Assoc. 2006;137(9):1252-7.
43. Golder S, Loke YK, Zorzela L. Comparison of search strategies in systematic reviews of adverse effects to other systematic reviews. Health Info Libr J. 2014 Jun;31(2):92-105.
44. Roundtree AK, Kallen MA, Lopez-Olivo MA, Kimmel B, Skidmore B, Ortiz Z, et al. Poor reporting of search strategy and conflict of interest in over 250 narrative and systematic reviews of two biologic agents in arthritis: a systematic review. J Clin Epidemiol. 2009 Feb;62(2):128-37.
45. Faggion CM, Jr., Wu YC, Tu YK, Wasiak J. Quality of search strategies reported in systematic reviews published in stereotactic radiosurgery. Br J Radiol. 2016 Jun;89(1062):20150878.
46. Chen KY, Qu LN, Hu F, Zhang YH, Yang JJ, Bai X, et al. The search strategy of systematic review/meta-analysis of animal research: A survey. Chinese Journal of Evidence-Based Medicine. 2016;16(3):348-53.
47. Marler JL, Curry KY, Suda KJ. Assessing the Quality of Database Search Strategies Utilized by Meta-analyses That Evaluate Infectious Disease Drug Therapy. Therapeutic Innovation and Regulatory Science. 2014;48(3):336-40.
48. Campbell J, Bellamy N, Gee T. Differences between systematic reviews/meta-analyses of hyaluronic acid/hyaluronan/hylan in osteoarthritis of the knee. Osteoarthritis Cartilage. 2007 Dec;15(12):1424-36.
49. Cooper C, Booth A, Varley-Campbell J, Britten N, Garside R. Defining the process to literature searching in systematic reviews: a literature review of guidance and supporting studies. BMC Med Res Methodol. 2018 2018/08/14;18(1):85.
50. Bramer WM. Evaluation of instructive texts on searching medical databases. J Med Libr Assoc. 2015 Oct;103(4):208-9.
51. Droste S, Dintsios CM. Literature searching for publications used in economic evaluation in health technology assessments: An 8-steps approach to successful health economics information retrieval. Gesundheitsokonomie und Qualitatsmanagement. 2011;16(1):35-57.
52. Kable AK, Pich J, Maslin-Prothero SE. A structured approach to documenting a search strategy for publication: a 12 step guideline for authors. Nurse Educ Today. 2012;32(8):878-86.
53. Booth A. Unpacking your literature search toolbox: on search styles and tactics. Health Info Libr J. 2008 Dec;25(4):313-7.
54. Ahmadi M, Sarabi RE, Orak RJ, Bahaadinbeigy K. Information Retrieval in Telemedicine: a Comparative Study on Bibliographic Databases. Acta Inform Med. 2015;23(3):172-6.
55. Ahmadi M, Ershad-Sarabi R, Jamshidiorak R, Bahaodini K. Comparison of bibliographic databases in retrieving information on telemedicine. Journal of Kerman University of Medical Sciences. 2014;21(4):343-54.
56. Vassar M, Yerokhin V, Sinnett PM, Weiher M, Muckelrath H, Carr B, et al. Database selection in systematic reviews: an insight through clinical neurology. Health Information & Libraries Journal. 2017.
57. Vassar M, Carr B, Kash-Holley M, DeWitt E, Koller C, Day J, et al. Database choices in endocrine systematic reviews. J Med Lib Assoc. 2015;103(4):189-91.
58. Hanneke R, Young SK. Information sources for obesity prevention policy research: A review of systematic reviews. Systematic Reviews. 2017;6(1).
59. Gehanno J-F, Rollin L, Darmoni S. Is the coverage of google scholar enough to be used alone for systematic reviews. BMC Med Inform Decis Mak. 2013;13(1):7.

Chapter 2

A systematic approach to searching

Bramer WM, de Jonge GB, Rethlefsen ML, Mast F, Kleijnen J. A systematic approach to searching: an efficient and complete method to develop literature searches. J Med Libr Assoc. 2018;106(4):531-41.

2 A systematic approach to searching : An efficient and complete method to develop literature searches

Abstract

Creating search strategies for systematic reviews, finding the best balance between sensitivity and specificity, and translating search strategies between databases is challenging. Several methods describe standards for systematic search strategies, but a consistent approach for creating an exhaustive search strategy has not yet been fully described in enough detail to be fully replicable. We have established a method that describes step-by-step the process of developing a systematic search strategy as needed in the systematic review. This method describes how single line search strategies can be prepared in a text document by typing search syntax (such as field codes, parentheses, and Boolean operators) before copying and pasting search terms (keywords and free text synonyms) found in the thesaurus. To help ensure term completeness, we developed a novel optimization technique that is mainly based on comparing the results retrieved by thesaurus terms with those retrieved by the free text search words to identify potentially relevant candidate search terms. Macros in Microsoft Word have been developed to convert syntaxes between databases and interfaces almost automatically. This method helps information specialists in the development of librarian-mediated searches for systematic reviews as well as medical and health care practitioners searching for evidence to answer clinical questions. The described method can be used to create complex and comprehensive search strategies for different databases and interfaces, such as those needed when searching for relevant references for systematic reviews, and will assist both information specialists and practitioners when searching the biomedical literature.

2.1 Introduction

Librarians and information specialists are often involved in the process of preparing and completing systematic reviews (SRs) where one of their main tasks is to identify relevant references for inclusion in the review (1). Although several recommendations for the process of searching have been published (2-6), none describe the development of a systematic search strategy from start to finish.

Traditional methods of SR search strategy development and execution are highly time-consuming, reportedly requiring up to 100 hours or more (7, 8). We wanted to develop systematic and exhaustive search strategies more efficiently, while preserving the high sensitivity SR search strategies necessitate. In this article, we will describe the method developed at Erasmus MC and demonstrate its use through an example search. The efficiency of the search method and outcome of 73 searches that have resulted in published reviews are described in a separate article (9).

As we aimed to describe the creation of systematic searches in full detail, the method starts at a basic level with the analysis of the research question and the creation of search terms. Readers who are new to SR searching are advised to follow all steps described in paragraph 2. For more experienced searchers, the basic steps can be considered existing knowledge that will already be part of their normal workflow, although paragraph 2.4 probably differs from general practice. Experienced searchers will gain the most from reading about the novelties in the method as described in paragraphs 2.10-2.13 and comparing the examples given in the supplementary material to their own practice.

2.2 Creating a systematic search strategy

Our methodology for planning and creating a multi-database search strategy consists of the following steps:

1. Determining a clear and focused question
2. Describing the articles that can answer the question
3. Deciding which key concepts address the different elements of the question
4. Deciding which elements should be used for the best results
5. Choosing an appropriate database and interface to start with
6. Documenting the search process in a text document
7. Identifying appropriate index terms in the thesaurus of the first database
8. Identifying synonyms in the thesaurus
9. Adding variations in search terms
10. Using database-appropriate syntax, with parentheses, Boolean operators, and field codes
11. Optimizing the search
12. Evaluate the initial results
13. Checking for errors
14. Translating to other databases
15. Testing and reiteration

Most steps in the process are reflected by an example search described in the grey box.

Step 1: Determining a clear and focused question

A systematic search can best be applied to a well-defined and precise research or clinical question. Questions that are too broad or too vague cannot be answered easily in a systematic way and will generally result in an overwhelming number of search results. On the other hand, a question that is too specific will result into too few or even zero search results. Various papers describe this process in more detail (10-12).

The question used as an example throughout this article is:
"Does exercise therapy improve the quality of life of patients with hip osteoarthritis compared to total hip replacement?"

Step 2: Describing the articles that can answer the question

Although not all clinical or research questions can be answered in the literature, the next step is to presume that the answer can indeed be found in published studies. A good starting point for a search is hypothesizing what the research that can answer the question would look like. These hypothetical (when possible, combined with known) articles can be used as guidance for constructing the search strategy.

Our example question can be answered by clinical articles, such as, but not limited to, prospective, randomized studies that follow up on patients measuring the outcome of exercise therapy for hip osteoarthritis.

Deciding which key concepts address the different elements of the well-formulated question
In the example research question three elements can be identified: one element is exercise therapy. Because we assume we can find our outcomes of interest in clinical studies, but also want to find studies mentioning our outcomes, the key concepts of Treatment effectiveness, clinical studies and quality of life together form one element. The last element is hip osteoarthritis.

Step 3: Deciding which key concepts address the different elements of the question

Key concepts are the topics or components that the desired articles should address, such as diseases or conditions, actions, substances, settings, domains (e.g., therapy, diagnosis, etiology), or study types. Key concepts from the research question can be grouped to create elements in the search strategy.

Elements in a search strategy do not necessarily follow the Patient, Intervention, Comparison, Outcome (PICO) structure or any other related structure. Using the PICO or another similar framework as guidance can be helpful to consider, especially in the inclusion and exclusion review stage of the SR, but this not necessary for good search strategy development (13-15). Sometimes concepts from different parts of the PICO structure can be grouped together in one search element, such as when the desired outcome is frequently described in a certain study type.

Step 4: Deciding which elements should be used for the best results

Not all elements of a research question should necessarily be used in the search strategy. Some elements are less important than others or may unnecessarily complicate or restrict a search strategy. Adding an element to a search strategy will increase the chance of missing relevant references. Therefore, the number of elements in a search strategy should remain as low as possible to optimize recall.

Using the schema in Figure 2.1, elements can be ordered by their specificity and importance to determine the best search approach. Whether an element is more specific or more general can be measured objectively by the number of hits retrieved in a database when searching for a key term representing that element. Depending on the research question, certain elements are more important than others. If articles (hypothetically or known) exist that can answer the question but lack a certain element in their title, abstract, or keywords, that element is unimportant to the question. An element can also be unimportant because of expected bias or an overlap with another element.

Bias in elements

The choice of elements in a search strategy can introduce bias through use of overly specific terminology or terms often associated with positive outcomes. For the question '*does prolonged breastfeeding improve intelligence outcomes in children?*', searching specifically for the element of *duration* will introduce bias, as articles that find a positive effect of prolonged breastfeeding will be much more likely to mention time factors in their title or abstract.

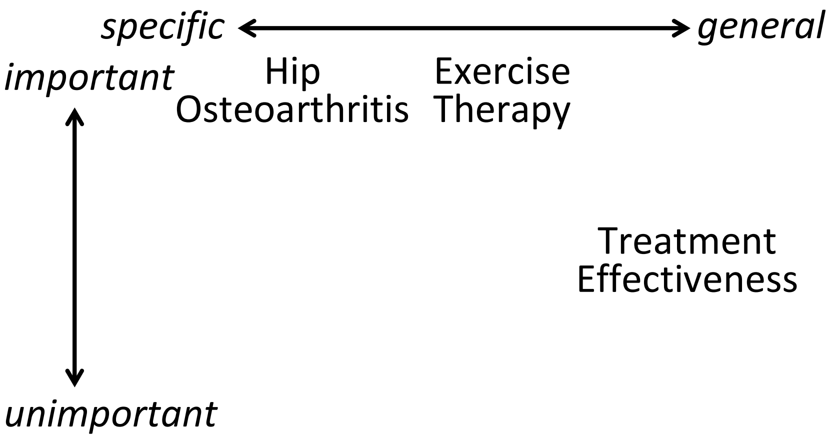
Overlapping elements

Elements in a question sometimes overlap in their meaning. Sometimes certain therapies are interventions for one specific disease. The Lichtenstein technique, for example, is a repair method for inguinal hernias. There is no need to include an element of inguinal hernias to a search for the effectiveness of the Lichtenstein therapy for inguinal hernias. Likewise, sometimes certain diseases are only found in certain populations. Adding such an overlapping element could even lead to missing relevant references.

From the plot of elements in the schema in Figure 2.1, the elements to use in a search strategy can be found by following the top row from left to right. For this method, we recommend starting with the most important and specific elements. Then, continue with more general and important elements until the number of results is acceptable for screening. Determining how many results are acceptable for screening is often a matter of negotiation with the SR team.

The plot for our research question is shown in the Figure below. The optimal search strategy will contain two elements: hip osteoarthritis and exercise therapy.

Figure 2.1: Schema for determining the optimal order of elements



Step 5: Choosing an appropriate database and interface to start with

Important factors for choosing databases to use are the coverage and the presence of a thesaurus. For medically oriented searches, the coverage and recall of Embase, which includes the MEDLINE database, are superior to those of MEDLINE.(16) Each of these two databases has its own thesaurus with its own unique definitions and structure. Because of the complexity of Emtree, the thesaurus of Embase, which contains much more specific thesaurus terms than the MEDLINE thesaurus MeSH, translation from Emtree to MeSH is easier than the other way around. Therefore, we recommend starting in Embase.

MEDLINE and Embase are available through many different vendors and interfaces. The choice of an interface and primary database is often determined by accessibility to the searcher.
For the method described here, an interface that allows for searching with proximity operators is desirable, and full functionality of the thesaurus, including explosion of narrower terms, is crucial. We recommend developing a personal workflow that always starts with one specific database and interface.

Step 6: Documenting the search process in a text document

We advise designing and creating the complete search strategies in a log document instead of directly in the database itself to register the steps taken and to make searches accountable and reproducible. From the log document, the developed search strategies can be copied and pasted into the desired databases. This way, the searcher is in control of the whole process. Any change to the search strategy should be done in the log document, assuring that the search strategy in the log is always the most recent.

Step 7: Identifying appropriate index terms in the thesaurus of the first database

Searches should start by identifying appropriate thesaurus terms for the desired elements. The thesaurus of the database is searched for matching index terms for each key concept. It is advisable to restrict the initial terms to the most important and most relevant terms. Later in the process more general terms can be added in the optimization process, in which the effect on the number of hits, and thus the desirability of adding these terms, can be evaluated more easily.

Several factors can complicate the identification of thesaurus terms. Sometimes one thesaurus term is found that exactly describes a specific element. In contrast, especially in more general elements, multiple thesaurus terms can be found to describe one element. If no relevant thesaurus terms have been found for an element, free text terms can be used, and possible thesaurus terms found in the resulting references can be added later (see Step 11).

Sometimes for a specific key concept there is no distinct thesaurus term available that describes that concept in enough detail. In Emtree, one thesaurus term often combines two or more other elements. The easiest solution for combining these terms for a sensitive search is to use such a thesaurus term in all elements where it is relevant. Examples are given in the supplementary material.

An example of a term for which no thesaurus terms can be found is Bennett's fracture, a fracture of the base of the metacarpal bone of the thumb. It can, in addition to a free text search for Bennett's fractures, be searched with a combination of the MeSH terms ("Fractures, Bone" AND "Metacarpal Bones" AND "Thumb").
An example of a term that combines two concepts is for instance the term 'cancer prevention' which is a thesaurus terms in Embase. Howeverm also 'neoplasm' and 'prevention' exist, where prevention is both an Emtree term and a subheading. The search would then become ('neoplasms'/exp OR 'cancer prevention'/exp) AND ('prevention'/exp OR 'cancer prevention'/exp). Because cancer prevention is a narrower term of prevention, the search can be shortened: ('neoplasms'/exp OR 'cancer prevention'/exp) AND ('prevention'/exp). The syntax used in these examples is that of Embase.com.

Hip Osteoarthritis
MeSH terms: Osteoarthritis, Hip
Emtree terms: hip osteoarthritis, Hip Disability and Osteoarthritis Outcome Score

Exercise therapy
MeSH terms: Exercise Therapy
Emtree terms: kinesiotherapy

Step 8: Identifying synonyms in the thesaurus

Most thesauri offer a list of synonyms on their term details page (named *Synonyms* in Emtree and *Entry Terms* in MeSH). To create a sensitive search strategy for SRs, these terms will need to be searched as free text keywords in the title and abstract fields in addition to searching their associated thesaurus term if the researcher considers the terms relevant.

The Emtree thesaurus contains more synonyms (300,000) than MeSH (220,000) (17). The difference in number of terms is even higher considering that many synonyms in MeSH are permuted terms (i.e., inversions of phrases using commas).

Thesaurus terms are ordered in a tree structure. When searching for a more general thesaurus term, the more specific (narrower) terms in the branches below that term will also be searched (this is frequently referred to as "exploding" a thesaurus term). However, to perform a sensitive search, all relevant variations of the narrower terms must be searched as free text keywords in the title or abstract in addition to relying on the exploded thesaurus term. Thus, all articles that describe a certain narrower topic in their title and abstract will already be retrieved before MeSH terms are added.

An example of a term with many synonyms is the MeSH term "observer variation", that contains among others the entry terms "observer bias" and "interobserver variability" as well as their inverted counterparts and phrases with additional spaces of hyphens, totaling to up to 41 different terms and phrases. By using phrase truncation or proximity operators wisely (see Step 9), the number of search terms needed can be limited; in the case of observer variation, four phrases are sufficient to cover all 41 entry terms.

An example of a term where narrower terms should be added is Colorectal Neoplasms. searching with the exploded MeSH term "Colorectal Neoplasms" also retrieves references indexed with the MeSH terms "Colonic Neoplasms", "Rectal Neoplasms" or "Anus Neoplasms". However, in order not to miss any relevant reference, these phrases and their variations should also be searched in title and or abstract.

In the example question, the MeSH term 'Osteoarthritis, Hip' displays the Entry Terms Hip Osteoarthritides; Hip Osteoarthritis; Osteoarthritides, Hip; Coxarthrosis; Coxarthroses; Osteoarthritis Of Hip; and Osteoarthritis Of Hips. To minimize the number of search terms needed and to capitalize on the similarity of the entry terms, we can ignore the inverted term Osteoarthritides, Hip and use truncation to search word variants. The final list of collapsed terms in a search interface that does not allow proximity search, such as PubMed, would be Hip Osteoarthrit*, Coxarthros*, and Osteoarthritis of the Hip*.
In addition, before 1989, the disease was indexed with the MeSH terms Hip, Hip Joint and Osteoarthritis. This means is that older articles can be found by combining either Hip or Hip Joint with Osteoarthritis.

Step 9: Adding variations in search terms (e.g., truncation, spelling differences, abbreviations, opposites)

Truncation allows a searcher to search for words beginning with the same word stem. A search for therap* will thus retrieve therapy, therapies, therapeutic, and all other words starting with 'therap'. Do not truncate a word stem that is too short. Also limitations of interfaces should be taken into account, especially in PubMed, where the number of search term variations that can be found by truncation is limited to 600.

Databases contain references to articles using both standard British and American English spellings. Both need to be searched as free text terms in the title and abstract. Alternatively, many interfaces offer a certain code to replace zero or one characters, allowing a search for '*pediatric*' or '*paediatric*' as '*p?ediatric*'. See Table 2.1 for detailed description of the syntax for different interfaces.

Searching for abbreviations can identify extra relevant references and retrieve more irrelevant ones. The search can be more focused by combining the abbreviation with an important word relevant to its meaning, or by using NOT to exclude frequently observed, clearly irrelevant results. It is advised that searchers do not exclude all possible irrelevant meanings, as it is very time-consuming to identify all the variations, it will result in unnecessarily complicated search strategies, and it may lead to erroneously narrowing the search and thereby reducing recall.

Searching partial abbreviations can be useful for retrieving relevant references. For example, it is very likely that an article would mention *osteoarthritis (OA)* early in the abstract, replacing all further occurrences of *osteoarthritis with OA*. Therefore, it may not contain the phrase '*hip osteoarthritis*', but only '*hip oa*'.

It is also important to search for the opposites of search terms to avoid bias. When searching for '*disease recurrence*', articles about '*disease free*' may be relevant as well. When the desired outcome is *survival*, articles about *mortality* may be relevant.

In the branch under the Emtree term kinesiotherapy (the thesaurus term most closely matching the exercise therapy element), many possibly relevant narrower terms are displayed, many of which can be used as free-text search terms, e.g.: isokinetic exercise, isometric exercise, movement therapy and muscle training.

In our example several variations exist for hip arthritis: hip artheritis, hip arthrosis etc. However, a search for hip arth* will also find hip arthroplasty, which is not a synonym for hip arthritis. Therefore that is not recommended.
To reduce the noise observed when seaching for an abbreviation such as THA (Total Hip Arthroplasty) several methods can be followed. The term can be combined with a relevant term from its meaning: (THA AND hip) or the searcher could NOT out any irrelevant meaning that occurs in the search results: (THA NOT ("Tetracosahexaenoic acid" OR "Threo-hydroxyaspartate" OR "Tetrahydroamentoflavone"))

Step 10: Using database-appropriate syntax, with parentheses, Boolean operators, and field codes

Different interfaces require different syntaxes, the special set of rules and symbols unique to each database that define how a correctly constructed search operates. Common syntax components include the use of parentheses and Boolean operators such as AND, OR, and NOT, which are available in all major interfaces. An overview of different syntaxes for four major interfaces for bibliographic medical databases (PubMed, Ovid, EBSCOhost, Embase.com, and ProQuest) is shown in Table 2.1.

Creating the appropriate syntax for each database, in combination with the selected terms as described in Steps 7-9, can be challenging. Following the method as outlined below simplifies the process:

- Create single-line queries in a text document (not combining multiple record sets); this allows for immediate checking of the relevance of retrieved references and for efficient optimization
- Type the syntax (Boolean operators, parentheses, and field codes) before adding terms; this reduces the chance that errors are made in the syntax, especially in the number of parentheses.
- Use predefined proximity structures including parentheses (such as (() ADJ3 ()) in Ovid) that can be re-used in the query when necessary
- Use thesaurus terms separately from free-text terms of each element. Start an element with all thesaurus terms (using OR) and follow with the free-text terms. This allows the unique optimization methods as described in Step 11.
- When adding terms to an existing search strategy, pay close attention to the position of the cursor. Make sure to place it appropriately either in the thesaurus terms section, in the title/abstract section, or as an addition (broadening) to an existing proximity search.

In the tables below, the method of building a query is explained in more detail step by step for different interfaces: PubMed, Ovid, EBSCOhost, Embase.com, and ProQuest. This method results in a basic search strategy designed to retrieve some relevant references upon which a more thorough search strategy can be built with optimization such as described in Step 11. Review the number of references to see if this is correct. See whether it contains relevant references, or that it is a lot of noise before you start the optimization process

Table 2.1: Field codes in five most used interfaces for biomedical literature searching

	PubMed	**Ovid**	**EBSCOhost**	**Embase.com**	**ProQuest**
Title/abstract	[tiab][1]	().ab,ti.	TI () OR AB () [2]	():ab,ti	AB,TI()
All fields	[All Fields]	.af.	[3]		ALL
Thesaurus term	[mesh:noexp]	…/	MH "…"	'…'/de	MESH(…)
Including narrower	[mesh]	exp …/	MH "…+"	'…'/exp	MESH#(…)
Combined subheading	…/sh[mesh]	exp …/sh	MH "…+/sh"	'…'/exp/dm_sh [4]	MESH(… LNK ..)
Free subheading	[sh][5]	.xs. or .fs.[5]	MW	:lnk[5]	
Publication type	[pt][6]	.pt. or exp …/[6]	PT	:it[6]	RTYPE
Proximity	[7]	ADJn	Nn	NEAR/n - NEXT/n	N/n
Exact phrase	"double quotes"	No quotes needed	"double quotes"	'single quotes'	"double quotes"
Truncated phrase	Use-hyphen*	No quote*	No quote*	'single quote*'	"Double quote*"
Truncation	End	End/ mid	End/ mid	End/ mid	End / mid / start
Infinite	*	* or $	*	*	*
0 or 1 character	-	?	#	-	$1
1 character	-	#	?	?[8]	?
Added to database since	yyyy/mm/dd:yyyy/mm/dd [edat][9] (or [mhda])	limit #N to rd=yyyymmdd-yyyymmdd [10]	EM yyyymmdd-yyyymmdd	[dd-mm-yyyy]/sd	LUPD (yyyymmdd)
Publication period (years)	yyyy:yyyy[dp]	limit #N to yr=yyyy-yyyy[10]	PY yyyy-yyyy	[yyyy-yyyy]/py	YR (yyyy-yyyy)
Record sets	#1	1[11]	S1	#1	S1

1 In PubMed, [tiab] should be placed after each search term.
2 EBSCOhost does not allow a combination of fields; all search terms for the title field need to be repeated for the abstract field.
3 EBSCOhost and Embase.com do not use an 'all fields' code; a term without a field code is searched in all fields.
4 Subheadings in Embase.com are only applied to diseases (/dm_), drugs (/dd_), or devices (/dv_).
5 [sh] and .xs. include narrower terms for subheadings; .fs. and :lnk do not.
6 [pt] and exp …/ includes narrower publication types; .pt. and :it do not.
7 In PubMed, proximity searching is not available; search the exact phrase (truncated or between double quotes) or use the Boolean AND combination.
8 The question mark does not work in combination with field codes.
9 The field [edat] refers to the entry date, when the record was added to PubMed. [mhda] refers to the MeSH date, when the record was last edited.
10 Adding a date limit can only be applied in a separate record set.
11 If a number is to be searched in the text, it should be put between double quotes (e.g., "1").

Table 2.2: Using database-appropriate syntax, with parentheses, Boolean operators and field codes. The current location of the cursor is shown by |

<table>
<tr><th></th><th>Ovid</th><th>EBSCOhost</th><th>Embase.com</th><th>PubMed sensitive</th><th>PubMed specific</th><th>ProQuest</th></tr>
<tr><td>Prepare a proximity syntax outside of the code to re-use when necessary</td><td>(() ADJ3 ())</td><td>(() N3 ())</td><td>(() NEAR/3 ())</td><td>(() AND ()) [12]</td><td></td><td>(() N/3 ())</td></tr>
<tr><td>Start by typing parentheses for the first element</td><td>(|)</td><td>(|)</td><td>(|)</td><td>(|)</td><td>(|)</td><td>(|)</td></tr>
<tr><td>Add standard syntax for thesaurus terms</td><td>(exp "|"/)</td><td>(MH "|+")</td><td>('|'/exp)</td><td>("|"[mh])</td><td>("|"[mh])</td><td>(Mesh# "|")</td></tr>
<tr><td>Paste the relevant thesaurus term(s) in the code</td><td>(exp "Osteoarthritis, Hip"/|)</td><td>(MH "Osteoarthritis, Hip+"|)</td><td>('hip osteoarthritis'/ exp|)</td><td>("Osteoarthritis, Hip"[mh] |)</td><td>("Osteoarthritis, Hip"[mh] |)</td><td>(Mesh# "Osteoarthritis, Hip"|)</td></tr>
<tr><td>Add syntax for free text terms in title abstract</td><td>(exp "Osteoarthritis, Hip"/ OR (|).ab,ti.)</td><td>(MH "Osteoarthritis, Hip+" OR AB (|)) [13]</td><td>('hip osteoarthritis'/ exp OR (|):ab,ti)</td><td>("Osteoarthritis, Hip"[mh] OR (|)) [14]</td><td>("Osteoarthritis, Hip"[mh] OR (|)) [14]</td><td>Mesh# “Osteo-arthritis, Hip” OR AB,TI(|))</td></tr>
</table>

[12] True proximity is not possible; PubMed does not allow proximity search. A very sensitive way is to search for words combined with AND, as searching for exact phrases will not retrieve all relevant hits.

[13] For our method, in EBSCOhost we advise starting with only AB as a field code. EBSCOhost cannot combine multiple fields. Search terms need to be repeated (for example TI (“hip osteoarthritis”) OR AB (“hip osteoarthritis”). This is very time consuming. After optimization copy the AB fields with and also use TI().

[14] In PubMed, this is not necessary as the field [tiab] has to be repeated after each synonym. However for optimization this will be useful.

2

<table>
<tr><th></th><th>Ovid</th><th>EBSCOhost</th><th>Embase.com</th><th>PubMed sensitive</th><th>PubMed specific</th><th>ProQuest</th></tr>
<tr><td>First add one word synonyms and exact phrases</td><td>(exp "Osteoarthritis, Hip"/ OR (Coxarthros*|).ab,ti.)</td><td>(MH "Osteoarthritis, Hip+" OR AB (Coxarthros*|))</td><td>('hip osteoarthritis'/ exp OR (Coxarthros* OR 'malum coxae sinilis'|):ab,ti)</td><td>("Osteoarthritis, Hip"[mh] OR (Coxarthros*[tiab] |))</td><td>("Osteoarthritis, Hip"[mh] OR (Coxarthros*[tiab] |))</td><td>(Mesh# "Osteoarthritis, Hip" OR AB,TI(Coxarthros*|))</td></tr>
<tr><td>Copy the proximity syntax as created above into the title abstract section</td><td>(exp "Osteoarthritis, Hip"/ OR (Coxarthros* OR ((|) ADJ3 ())).ab,ti.)</td><td>(MH "Osteoarthritis, Hip+" OR AB (Coxarthros* OR ((|) N3 ())))</td><td>('hip osteoarthritis'/ exp OR (Coxarthros* OR 'malum coxae sinilis' OR ((|) NEAR/3 ())):ab,ti)</td><td>("Osteoarthritis, Hip"[mh] OR (Coxarthros*[tiab] OR ((|) AND ())))</td><td>Not possible, instead of proximity truncated phrases are advised</td><td>(Mesh# "Osteoarthritis, Hip" OR AB,TI(Coxarthros* OR ((|) N/3 ())))</td></tr>
<tr><td>Paste or type the words in the proximity syntax</td><td>(exp "Osteoarthritis, Hip"/ OR (Coxarthros* OR ((Hip) ADJ3 (Osteoarthrit*))). ab,ti.)</td><td>(MH "Osteoarthritis, Hip+" OR AB (Coxarthros* OR ((Hip) N3 (Osteoarthrit*))))</td><td>('hip osteoarthritis'/ exp OR (Coxarthros* OR 'malum coxae sinilis' OR ((hip OR cox) NEAR/3 (arthrit* OR arthros* OR osteoarthr*))):ab,ti)</td><td>("Osteoarthritis, Hip"[mh] OR (Coxarthros*[tiab] OR ((hip*[tiab]) AND (Osteoarthrit*[tiab]))))</td><td>("Osteoarthritis, Hip"[mh] OR (Coxarthros*[tiab] OR Hip Osteoarthrit*[tiab] OR Osteoarthritis Of Hip*[tiab])) [15]</td><td>(Mesh# "Osteoarthritis, Hip" OR AB,TI(Coxarthros* OR ((hip) N/3 (osteoarthrit*))))</td></tr>
</table>

[15] In PubMed placing an asterisk (*) after a phrase combines the word with the previous word(s) into a phrase. Do not use quotes and an asterisk as when using quotes truncation will be ignored.

	Ovid	EBSCOhost	Embase.com	PubMed sensitive	PubMed specific	ProQuest
Follow the same steps for the other elements	(exp "Exercise Therapy"/ OR (((Exercise) ADJ3 (Therap*))).ab,ti.)	(MH "Exercise Therapy+" OR AB (((Exercise) N3 (Therap*))))	('kinesiotherapy'/exp OR (kinesiotherap* OR kinesitherap* OR ((exercise) NEAR/3 (technique* OR treat* OR therap*))):ab,ti)	("Exercise Therapy"[mh] OR (Exercise Therap*[tiab]))	("Exercise Therapy"[mh] OR (Exercise Therap*[tiab]))	(Mesh# "Exercise Therapy" OR AB,TI (((Exercise) N/3 (Therap*))))
Combine all elements into one search strategy	(exp "Osteoarthritis, Hip"/ OR (Coxarthros* OR ((Hip) ADJ3 (Osteoarthrit*))). ab,ti.) AND (exp "Exercise Therapy"/ OR (((Exercise) ADJ3 (Therap*))).ab,ti.)	(MH "Osteoarthritis, Hip+" OR AB (Coxarthros* OR ((Hip) N3 (Osteoarthrit*)))) AND (MH "Exercise Therapy+" OR AB (((Exercise) N3 (Therap*))))	('hip osteoarthritis'/ exp OR (Coxarthros* OR 'malum coxae sinilis' OR ((hip OR cox) NEAR/3 (arthrit* OR arthros* OR osteoarthr*))):ab,ti) AND ('kinesiotherapy'/ exp OR (kinesiotherap* OR kinesitherap* OR ((exercise) NEAR/3 (technique* OR treat* OR therap*))):ab,ti)	("Osteoarthritis, Hip"[mh] OR (Coxarthros*[tiab] OR ((hip*[tiab]) AND (Osteoarthrit*[tiab])))) AND ("Exercise Therapy"[mh] OR (Exercise Therap*[tiab]))	*("Osteoarthritis, Hip"[mh] OR (Coxarthros*[tiab] OR Hip Osteoarthrit*[tiab] OR Osteoarthritis Of Hip*[tiab])) AND ("Exercise Therapy"[mh] OR (Exercise Therap*[tiab]))*	(Mesh# "Osteoarthritis, Hip" OR AB,TI(Coxarthros* OR ((hip) N/3 (osteoarthrit*)))) AND (Mesh# "Exercise Therapy" OR AB,TI (((Exercise) N/3 (Therap*))))

Step 11: Optimizing the search

The most important question when performing a systematic search is whether all (or most) potentially relevant articles have been retrieved by the search strategy. This is also the most difficult question to answer, since it is unknown which and how many articles are relevant. It is therefore wise first to broaden the initial search strategy, making the search more sensitive, and then check if new relevant articles are found by comparing the set results (i.e., search for Strategy #2 NOT Strategy #1 to see the unique results).

A search strategy for a SR should be tested for completeness. Therefore, it is necessary to identify extra possibly relevant search terms and add them to the test search in an OR relationship within the already used elements. A good place to start, and a well-known strategy, is scanning the top retrieved articles when sorted by relevance, looking for additional relevant synonyms that could be added to the search strategy.

We have developed a unique optimization method that has not been described before in the literature. This method often adds valuable extra terms to our search strategy, and therefore extra relevant references to our search results. Extra synonyms can be found in articles that have been assigned a certain set of thesaurus terms but that lack synonyms in the title and/or abstract that are already present in the current search strategy. Searching for *major thesaurus terms NOT free-text terms* will help identify missed free-text terms in the title or abstract. Searching for *free-text terms in title NOT thesaurus terms* will help identify missed thesaurus terms. If this is done repeatedly for each element, leaving the rest of the query unchanged, this method will help add numerous relevant terms to the query. In the supplementary material, these steps are explained in detail for five different search platforms.

Table 2.3:. Optimizing the search

	Ovid	EBSCOhost	Embase.com	PubMed sensitive	PubMed specific	ProQuest
Starting query	(exp "Osteoarthritis, Hip"/ OR (Coxarthros* OR ((Hip) ADJ3 (Osteoarthrit*))). ab,ti.) AND (exp "Exercise Therapy"/ OR (((Exercise) ADJ3 (Therap*))).ab,ti.)	(MH "Osteoarthritis, Hip+" OR AB (Coxarthros* OR ((Hip) N3 (Osteoarthrit*)))) AND (MH "Exercise Therapy+" OR AB (((Exercise) N3 (Therap*))))	('hip osteoarthritis'/ exp OR (Coxarthros* OR 'malum coxae sinilis' OR ((hip OR cox) NEAR/3 (arthrit* OR arthros* OR osteoarthr*))):ab,ti) AND ('kinesiotherapy'/ exp OR (kinesiotherap* OR kinesitherap* OR ((exercise) NEAR/3 (technique* OR treat* OR therap*))):ab,ti)	("Osteoarthritis, Hip"[mh] OR (Coxarthros*[tiab] OR ((hip*[tiab]) AND (Osteoarthrit*[tiab])))) AND ("Exercise Therapy"[mh] OR (Exercise Therap*[tiab]))	("Osteoarthritis, Hip"[mh] OR (Coxarthros*[tiab] OR Hip Osteoarthrit*[tiab] OR Osteoarthritis Of Hip*[tiab])) AND ("Exercise Therapy"[mh] OR (Exercise Therap*[tiab]))	(Mesh# "Osteoarthritis, Hip" OR AB,TI(Coxarthros* OR ((hip) N/3 (osteoarthrit*)))) AND (Mesh# "Exercise Therapy" OR AB,TI (((Exercise) N/3 (Therap*))))
Optimization first element: thesaurus terms NOT title abstract terms	(exp *"Osteoarthritis, Hip"/ NOT (Coxarthros* OR ((Hip) ADJ3 (Osteoarthrit*))).ab,ti.) AND (exp "Exercise Therapy"/ OR (((Exercise) ADJ3 (Therap*))).ab,ti.)	(MJ "Osteoarthritis, Hip+" NOT AB (Coxarthros* OR ((Hip) N3 (Osteoarthrit*)))) AND (MH "Exercise Therapy+" OR AB (((Exercise) N3 (Therap*))))	('hip osteoarthritis'/ exp/mj NOT (Coxarthros* OR 'malum coxae sinilis' OR ((hip OR cox) NEAR/3 (arthrit* OR arthros* OR osteoarthr*))):ab,ti) AND ('kinesiotherapy'/ exp OR (kinesiotherap* OR kinesitherap* OR ((exercise) NEAR/3 (technique* OR treat* OR therap*))):ab,ti)	("Osteoarthritis, Hip"[mj] NOT (Coxarthros*[tiab] OR ((hip*[tiab]) AND (Osteoarthrit*[tiab])))) AND ("Exercise Therapy"[mh] OR (Exercise Therap*[tiab]))	("Osteoarthritis, Hip"[mj] NOT (Coxarthros*[tiab] OR Hip Osteoarthrit*[tiab] OR Osteoarthritis Of Hip*[tiab])) AND ("Exercise Therapy"[mh] OR (Exercise Therap*[tiab]))	(MJMesh# "Osteoarthritis, Hip" NOT AB,TI(Coxarthros* OR ((hip) N/3 (osteoarthrit*)))) AND (Mesh# "Exercise Therapy" OR AB,TI (((Exercise) N/3 (Therap*))))

	Ovid	EBSCOhost	Embase.com	PubMed sensitive	PubMed specific	ProQuest
Extra words added	(exp "Osteoarthritis, Hip"/ OR (Coxarthros* OR ((Hip OR cox) ADJ3 (Osteoarthrit* OR arthrit* OR arthros* OR oa))). ab,ti.) AND (exp "Exercise Therapy"/ OR (((Exercise) ADJ3 (Therap*))).ab,ti.)	(MH "Osteoarthritis, Hip+" OR AB (Coxarthros* OR ((Hip OR cox) N3 (Osteoarthrit* OR arthrit* OR arthros* OR oa)))) AND (MH "Exercise Therapy+" OR AB (((Exercise) N3 (Therap*))))	('hip osteoarthritis'/ exp OR (Coxarthros* OR 'malum coxae sinilis' OR ((hip OR cox) NEAR/3 (arthrit* OR arthros* OR osteoarthr* OR oa))):ab,ti) AND ('kinesiotherapy'/exp OR (kinesiotherap* OR kinesitherap* OR ((exercise) NEAR/3 (technique* OR treat* OR therap*))):ab,ti)	("Osteoarthritis, Hip"[mh] OR (Coxarthros*[tiab] OR ((hip*[tiab] OR cox[tiab]) AND (Osteoarthrit*[tiab] OR arthrit*[tiab] OR arthros*[tiab] OR oa[tiab])))) AND ("Exercise Therapy"[mh] OR (Exercise Therap*[tiab]))	("Osteoarthritis, Hip"[mh] OR (Coxarthros*[tiab] OR Hip Osteoarthrit*[tiab] OR cox Osteoarthrit*[tiab] OR Hip Osteo arthrit*[tiab] OR "Hip oa"[tiab] OR Osteoarthritis Of Hip*[tiab] OR Osteoarthritis Of the Hip*[tiab] OR "oa of the Hip"[tiab])) AND ("Exercise Therapy"[mh] OR (Exercise Therap*[tiab]))	(Mesh# "Osteoarthritis, Hip" OR AB,TI(Coxarthros* OR ((hip OR cox) N/3 (osteoarthrit* OR arthrit* OR arthros* OR oa)))) AND (Mesh# "Exercise Therapy" OR AB,TI (((Exercise) N/3 (Therap*))))
Optimization second element: title terms NOT thesaurus terms	(exp "Osteoarthritis, Hip"/ OR (Coxarthros* OR ((Hip OR cox) ADJ3 (Osteoarthrit* OR arthrit* OR arthros* OR oa))). ab,ti.) AND ((((Exercise) ADJ3 (Therap*))).ti.) NOT (exp "Exercise Therapy"/)	(MH "Osteoarthritis, Hip+" OR AB (Coxarthros* OR ((Hip OR cox) N3 (Osteoarthrit* OR arthrit* OR arthros* OR oa)))) AND (TI(((Exercise) N3 (Therap*)))) NOT (MH "Exercise Therapy+")	('hip osteoarthritis'/ exp OR (Coxarthros* OR 'malum coxae sinilis' OR ((hip OR cox) NEAR/3 (arthrit* OR arthros* OR osteoarthr* OR oa))):ab,ti) AND ((kinesiotherap* OR kinesitherap* OR ((exercise) NEAR/3 (technique* OR treat* OR therap*))):ti) NOT ('kinesiotherapy'/exp)	("Osteoarthritis, Hip"[mh] OR (Coxarthros*[tiab] OR ((hip*[tiab] OR cox[tiab]) AND (Osteoarthrit*[tiab] OR arthrit*[tiab] OR arthros*[tiab] OR oa[tiab])))) AND ((Exercise Therap*[ti])) NOT ("Exercise Therapy"[mh)	("Osteoarthritis, Hip"[mh] OR (Coxarthros*[tiab] OR Hip Osteoarthrit*[tiab] OR cox Osteoarthrit*[tiab] OR Hip Osteo arthrit*[tiab] OR "Hip oa"[tiab] OR Osteoarthritis Of Hip*[tiab] OR Osteoarthritis Of the Hip*[tiab] OR "oa of the Hip"[tiab])) AND ((Exercise Therap*[ti])) NOT ("Exercise Therapy"[mh)	(Mesh# "Osteoarthritis, Hip" OR AB,TI(Coxarthros* OR ((hip OR cox) N/3 (osteoarthrit* OR arthrit* OR arthros* OR oa)))) AND (TI (((Exercise) N/3 (Therap*)))) NOT (Mesh# "Exercise Therapy")

	Ovid	EBSCOhost	Embase.com	PubMed sensitive	PubMed specific	ProQuest
Extra words added **These steps should be performed repeatedly on all elements in the search strategy** **Note: Only after all optimization is finished AB in the EBSCOhost should be copied for all elements into both AB() OR TI()**	(exp “Osteoarthritis, Hip”/ OR (Coxarthros* OR ((Hip OR cox) ADJ3 (Osteoarthrit* OR arthrit* OR arthros* OR oa))).ab,ti.) AND (exp “Exercise Therapy”/ OR exp “Musculoskeletal Manipulations”/ OR exp “Exercise”/ OR exp “Physical Therapy Modalities”/ OR (((Exercise) ADJ3 (Therap*))).ab,ti.)	(MH “Osteoarthritis, Hip+” OR AB (Coxarthros* OR ((Hip OR cox) N3 (Osteoarthrit* OR arthrit* OR arthros* OR oa))) TI (Coxarthros* OR ((Hip OR cox) N3 (Osteoarthrit* OR arthrit* OR arthros* OR oa)))) AND (MH “Exercise Therapy+” OR MH “Musculoskeletal Manipulations+” OR MH “Exercise+” OR MH “Physical Therapy Modalities+” OR AB (((Exercise) N3 (Therap*))) OR TI (((Exercise) N3 (Therap*))))	(‘hip osteoarthritis’/exp OR (Coxarthros* OR ‘malum coxae sinilis’ OR ((hip OR cox) NEAR/3 (arthrit* OR arthros* OR osteoarthr* OR oa))):ab,ti) AND (‘kinesiotherapy’/exp OR ‘physiotherapy’/exp OR ‘manipulative medicine’/exp OR exercise/exp OR (kinesiotherap* OR kinesitherap* OR ((exercise) NEAR/3 (technique* OR treat* OR therap*))):ab,ti)	(“Osteoarthritis, Hip”[mh] OR (Coxarthros*[tiab] OR ((hip*[tiab] OR cox[tiab]) AND (Osteoarthrit*[tiab] OR arthrit*[tiab] OR arthros*[tiab] OR oa[tiab])))) AND (“Exercise Therapy”[mh] OR “Musculoskeletal Manipulations”[mh] OR “Exercise”[mh] OR “Physical Therapy Modalities”[mh] OR (Exercise Therap*[tiab]))	("Osteoarthritis, Hip"[mh] OR (Coxarthros*[tiab] OR Hip Osteoarthrit*[tiab] OR cox Osteoarthrit*[tiab] OR Hip Osteo arthrit*[tiab] OR "Hip oa"[tiab] OR Osteoarthritis Of Hip*[tiab] OR Osteoarthritis Of the Hip*[tiab] OR "oa of the Hip"[tiab])) AND ("Exercise Therapy"[mh] OR "Musculoskeletal Manipulations"[mh] OR "Exercise"[mh] OR "Physical Therapy Modalities"[mh] OR (Exercise Therap*[tiab]))	(Mesh# "Osteoarthritis, Hip" OR AB,TI(Coxarthros* OR ((hip OR cox) N/3 (osteoarthrit* OR arthrit* OR arthros* OR oa)))) AND (Mesh# "Exercise Therapy" OR Mesh# "Musculoskeletal Manipulations" OR Mesh# "Exercise" OR Mesh# "Physical Therapy Modalities" OR AB,TI (((Exercise) N/3 (Therap*))))

Step 12: Evaluate the initial results

The results should now contain relevant references. If the interface allows relevance ranking, use that in the evaluation. If you know some relevant references that should be included in the research, search for those references specifically; for example, combine a specific (first) author name with a page number and the publication year. Check whether those references are retrieved by the search. If the known relevant references are not retrieved by the search, adapt the search so that they are. If it is unclear which element should be adapted to retrieve a certain article, combine that article with each element separately.

Different outcomes are desired for different types of research questions. For instance, in the case of clinical question answering, the researcher will not be satisfied with many references that contain a lot of irrelevant references. A clinical search should be rather specific and is allowed to miss a relevant reference. In the case of an SR, the researchers do not want to miss any relevant reference and are willing to handle many irrelevant references to do so. The search for references to include in an SR should be very sensitive: no included reference should be missed. A search that is too specific or too sensitive for the intended goal can be adapted to become more sensitive or specific. Steps to increase sensitivity or specificity of a search strategy can be found in table 2.4.

Step 13: Checking for errors

Errors might not be easily detected. Sometimes clues can be found in the number of results, either when the number of results is much higher or lower than expected or when many retrieved references are not relevant. However, the number expected is often unknown, and very sensitive search strategies will always retrieve many irrelevant articles. Each query should be therefore checked for errors.

One of the most frequently occurring errors is missing the Boolean operator OR. When no OR is added between two search terms, many interfaces automatically add an AND, which unintentionally reduces the number of results and likely misses relevant references. One good strategy to identify missing ORs is to go to the web page containing the full search strategy, as translated by the database, and using Ctrl-F search for 'AND '. Check whether the occurrences of the AND operator are deliberate.

Ideally, search strategies should be checked by other information specialists (18). The Peer Review of Electronic Search Strategies (PRESS) checklist offers good guidance for this process (4). Apart from the syntax (especially Boolean operators and field codes) of the search strategy, it is wise to have the search terms checked by the clinician or researcher familiar with the topic. At Erasmus MC, researchers and clinicians are involved during the complete process of structuring and optimizing the search strategy. Each word is added after the combined decision of the searcher and the researcher, with the possibility of direct comparison of results with and without the new term.

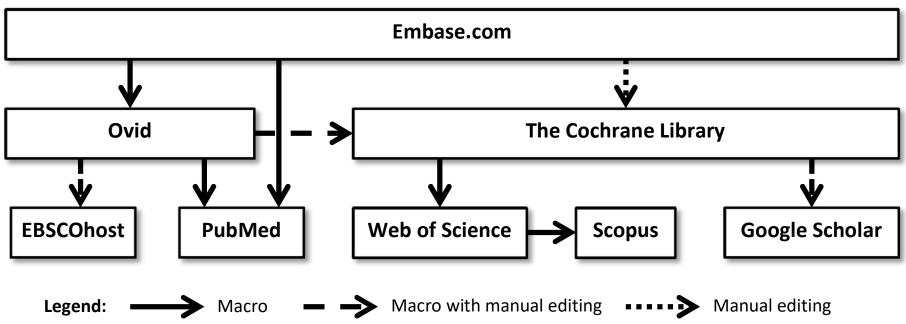
Table 2.4: Methods to increase sensitivity or specificity for a search strategy

Increase Sensitivity	Increase Specificity
Add extra terms Extra words/synonyms can be found in the retrieved references as described in this article, for instance by searching for articles that have the thesaurus terms, but lack the words in title or abstract	**Delete less relevant terms** When in doubt whether a term adds value, look at the extra retrieved references: Do they seem relevant? If not, delete these synonym(s).
Generalize specific elements Potential relevant articles may be indexed with broader thesaurus terms, because in these articles an overview of related topics has been described	**Specify more general elements** Choose a more specific thesaurus term. If the narrower terms from broad thesaurus terms cause noise, solve this by using the no explode function.
Drop less important elements Check what happens if each element is deleted from the search strategy. If the number of hits does not increase to above what can be handled, that element is not necessary and can be deleted. If the number of hits is too high, scan the extra retrieved relevant references for potentially new relevant synonyms	**Add an extra element** Extra elements will reduce the number of hits, but will increase the chance of missed relevant references. Always check the references that would be missed by introducing this element for extra relevant terms. This can be very time consuming
Use floating subheadings Searching with subheadings as a separate element (e.g. “neoplasms”[mh] AND “prevention and control”[sh]) will retrieve all references where the subheading is used in combination with other thesaurus terms.	**Combine subheadings with thesaurus terms** Floating subheadings can retrieve many irrelevant results. This noise can be reduced by combining thesaurus terms with subheadings (such as “neoplasms/prevention and control”[mh]). If using such combinations also combine the free text words of these elements in phrases or proximity.
Broaden proximity (or replace with AND) It can be determined with great precision whether to use 3, 4 or 5 words in proximity. Changes are often minimal. If starting with 3, try broadening to 6, then to 10, 15 etc. Check for each step the extra retrieved references on possible relevance. Decide which proximity gives the best results. When using AND instead of proximity often retrieves much noise, but can identify relevant references.	**Narrower proximity or phrases (instead of AND)** A combination of two or more free text words with AND is very sensitive and can lead to much noise. This noise can be reduced by combining words of two elements into phrases or proximity searches.
Add (phrase) truncation or shorten word stems Always add truncation at the end of a word and use word stems as short as possible unless this leads to irrelevant results.	**Remove truncation or lengthen word stems** Sometimes truncated words or phrases result in noise. In that case lengthen the word stem/phrase or do not truncate
Remove filters Filters should only be applied in the end of the search strategy, if the number of hits retrieved is too high or the noise is too much. The safest method is often to categorize the retrieved references by hand	**Add filters (language or date)** Restricting to language(s) may create bias in the retrieved references. So the best results are retrieved not limiting on language. If you want to restrict, only restrict to English, not to languages spoken by you or your team members. Limit on date can be done for two reasons: 1) a thorough systematic review has been carried out and needs to be updated; 2) one of the elements in the question is a new concept and did not exist before a certain date

Step 14: Translating to other databases

To retrieve as many relevant references as possible, one has to search multiple databases. Translation of complex and exhaustive queries between different databases can be very time-consuming and cumbersome. The single line search strategy approach detailed above allows for quick translations using the find and replace method in Microsoft Word (<Ctrl-H>). At Erasmus MC, macros based on the find-and-replace method in Microsoft Word have been developed for easy and fast translation between the most used databases for biomedical and health science questions. The scheme that is followed for the translation between databases is shown in Figure 2.2. Most databases simply follow the structure set by the Embase.com search strategy. The translation from Emtree terms to MeSH terms for MEDLINE in Ovid often identifies new terms that also need to be added to the Embase.com search strategy before the translation to other databases.

Figure 2.2: Schematic representation of translation between databases using available macros



Using different macros, a thoroughly optimized query in Embase.com can be relatively quickly translated into eight major databases. Basic search strategies will be created to use in many, mostly smaller, databases, as such niche databases often do not have extensive thesauri or advanced syntax options. Also there is not much need to use extensive syntax, because the number of hits, and therefore the amount of noise in these databases, is generally low. In MEDLINE (Ovid), PsycINFO (Ovid), and CINAHL (EBSCOhost), the thesaurus terms must be adapted manually, as each database has its own custom thesaurus. These macros and instructions for their installation, use, and adaptation are available at bit.ly/databasemacros.

Systematic reviews in Erasmus MC search Embase.com, Medline via Ovid, Web of Science, Cochrane CENTRAL and Google Scholar. The results of the translation between databases are shown below.

Embase.com

('hip osteoarthritis'/exp OR (Coxarthros* OR 'malum coxae sinilis' OR ((hip OR cox) NEAR/3 (arthrit* OR arthros* OR osteoarthr* OR oa))):ab,ti) AND ('kinesiotherapy'/exp OR 'physiotherapy'/exp OR 'manipulative medicine'/exp OR exercise/exp OR (kinesiotherap* OR kinesitherap* OR ((exercise) NEAR/3 (technique* OR treat* OR therap*))):ab,ti)

Medline Ovid

The macro translated the syntax of Embase into this Ovid syntax.
(hip osteoarthritis/ OR (Coxarthros* OR malum coxae sinilis OR ((hip OR cox) ADJ3 (arthrit* OR arthros* OR osteoarthr* OR oa))).ab,ti.) AND (kinesiotherapy/ OR physiotherapy/ OR manipulative medicine/ OR exercise/ OR (kinesiotherap* OR kinesitherap* OR ((exercise) ADJ3 (technique* OR treat* OR therap*))).ab,ti.)

Manually the Emtree terms are replaced by MeSH terms. If narrower terms have to be taken into account (such as in the MeSH term Physical Therapy Modalities, exp is placed before the MeSH term. Different than in Emtree where kinesiotherapy is not a narrower term of physiotherapy, the MeSH term Exercise Therapy is found below Physical Therapy Modalities, therefore does not need to be added. When searched in MeSH, the Emtree term manipulative medicine is mapped to Osteopathic Medicine, which is not the most relevant term. Musculoskeletal Manipulations is a better translation.

(Osteoarthritis, Hip/ OR (Coxarthros* OR malum coxae sinilis OR ((hip OR cox) ADJ3 (arthrit* OR arthros* OR osteoarthr* OR oa))).ab,ti.) AND (exp Physical Therapy Modalities/ OR exp Musculoskeletal Manipulations/ OR exp Exercise/ OR (kinesiotherap* OR kinesitherap* OR ((exercise) ADJ3 (technique* OR treat* OR therap*))).ab,ti.)

Cochrane CENTRAL

For a Cochrane search the syntax of Embase.com is adapted by removing the Emtree terms from the search. Cochrane only contains MeSH terms, but only copies those from Medline. Therefore, searching the Cochrane Library with MeSH terms, when Medline has already been searched, will not retrieve new articles. The syntax of the Cochrane library for title or abstract terms is similar to that of Embase.com.
((Coxarthros* OR 'malum coxae sinilis' OR ((hip OR cox) NEAR/3 (arthrit* OR arthros* OR osteoarthr* OR oa))):ab,ti) AND ((kinesiotherap* OR kinesitherap* OR ((exercise) NEAR/3 (technique* OR treat* OR therap*))):ab,ti)

Web of Science

Using the macros the syntax for Cochrane CENTRAL is translated in a syntax for Web of Science.
TS=(((Coxarthros* OR "malum coxae sinilis" OR ((hip OR cox) NEAR/2 (arthrit* OR arthros* OR osteoarthr* OR oa)))) AND ((kinesiotherap* OR kinesitherap* OR ((exercise) NEAR/2 (technique* OR treat* OR therap*)))))

Google Scholar

The macros change the syntax of Cochrane CENTRAL to that of Google Scholar. Because of limitations in the length of the search to 256 characters each OR is replaced by a vertical line.

"Coxarthros|"malum coxae sinilis"|"hip|cox arthrit|arthros|osteoarthr|oa "kinesiotherap|kinesitherap|"exercise technique|treat|therap

The Google Scholar syntax has to be edited to make it work. In stead of proximity we use the option to place an OR (|) in a quoted phrase. The macro placed some quote in the correct location, however some quotes are misplaced and have to be removed. Since Google Scholar does not allow truncation, each truncated word has to be completed.

Coxarthrosis|"malum coxae sinilis"|"hip|cox arthritis|arthrosis|osteoarthritis|oa" kinesiotherapy|kinesitherapy|"exercise technique|treatment|therapy"

Optional other databases for which Macros are available are CINAHL (EBSCOhost) and Scopus

CINAHL (EBSCOhost)
The syntax of Medline Ovid is translated into a CINAHL EBSCOhost syntax, using a macro.
(Osteoarthritis, Hip+ OR (Coxarthros* OR malum coxae sinilis OR ((hip OR cox) N2 (arthrit* OR arthros* OR osteoarthr* OR oa)))) AND (MH Physical Therapy Modalities+ OR MH Musculoskeletal Manipulations+ OR MH Exercise+ OR (kinesiotherap* OR kinesitherap* OR ((exercise) N2 (technique* OR treat* OR therap*))))

Some editing is necessary. MeSH terms have to be replaced by CINAHL headings, though there are many overlaps, as the CINAHL headings are largely based on the MeSH terms. Each thesaurus term should be preceded by the field code MH. The search terms for title abstrat have to be copied and once preceded by TI and once by AB. Sadly EBSCOhost is unable to search for the combined fields AB,TI

(MH Osteoarthritis, Hip OR TI (Coxarthros* OR malum coxae sinilis OR ((hip OR cox) N2 (arthrit* OR arthros* OR osteoarthr* OR oa))) OR AB (Coxarthros* OR malum coxae sinilis OR ((hip OR cox) N2 (arthrit* OR arthros* OR osteoarthr* OR oa)))) AND (MH Physical Therapy Modalities+ OR MH Musculoskeletal Manipulations+ OR MH Exercise+ OR TI (kinesiotherap* OR kinesitherap* OR ((exercise) N2 (technique* OR treat* OR therap*))) OR AB (kinesiotherap* OR kinesitherap* OR ((exercise) N2 (technique* OR treat* OR therap*))))

Scopus
Using the macros the syntax for Web of Science is translated in a syntax for Scopus.
TITLE-ABS-KEY(((Coxarthros* OR "malum coxae sinilis" OR ((hip OR cox) W/2 (arthrit* OR arthros* OR osteoarthr* OR oa)))) AND ((kinesiotherap* OR kinesitherap* OR ((exercise) W/2 (technique* OR treat* OR therap*)))))

Step 15: Testing and reiteration

Ideally, exhaustive search strategies should retrieve all references that are covered in a specific database. For SR search strategies, checking searches for their recall is advised. This can be done after included references have been determined by the authors of the systematic review. If additional papers have been identified through other non-database methods (i.e., checking references in included studies), results that were not identified by the database searches should be examined. If these results were available in the databases but not located by the search strategy, the search strategy should be adapted to try to retrieve these results, as they may contain terms that were omitted in the original search strategies. This may enable the identification of additional relevant results.

2.3 Discussion

A methodology for creating exhaustive search strategies has been created which describes all steps of the search process, starting with a question and resulting in thorough search strategies in multiple databases. Many of the steps described are not new, but together they form a strong method creating high quality, robust searches in a relatively short time frame. The methodology is intended to create thoroughness for literature searches. The optimization method as described in step 11 will identify missed synonyms or thesaurus terms, unlike any other method, which largely depends on predetermined keywords and synonyms. Using this method resulted in a much quicker search process, compared to our traditional methods, especially because of the easier translation between databases and interfaces (Step 14). The method is not a guarantee for speed, since speed depends on many factors, including experience. However, by following the steps and using the tools as described above, searchers can gain confidence first and increase speed through practice.

What is new?

This method is the first published, to our knowledge, that encourages searchers to start their search development process using empty syntax first, and later adding the thesaurus terms and free-text synonyms. We feel this helps the searcher to focus on the search terms, instead of on the structure of the search query. The macros to translate search strategies between interfaces are also unique in this method. The optimization method in which new terms are found in the already retrieved articles is used in some other institutes as well but has to our knowledge not been described in the literature.

What is different compared to common practice?

Traditionally, librarians and information specialists have focused on creating complex, multi-line (also called line-by-line) search strategies, consisting of multiple record sets, and this is frequently advised in the literature and handbooks (2, 19-21). The described method instead uses single-line searches, which is critical to its success. Single-line search strategies can be easily adapted by adding or dropping a term without having to recode numbers of record sets, which would be necessary in multi-line searches. They can easily be saved in a text document and repeated by copying and pasting for search updates. Single-line search strategies also allow easy translation to other syntaxes using find-and-replace technology to update field codes and other syntax elements or using macros (Step 13).

When constructing a search strategy, the searcher might experience that certain parentheses in the syntax are unnecessary, such as parentheses around all search terms in title/abstract portion if there is only one such term, or double parentheses in the proximity statement if one of the word groups exists of only one word. One might be tempted to omit those parentheses for ease of reading and managing. However, during the optimization process, the searcher is likely to find extra synonyms that might consist of one word. To add those terms to the first query (with reduced parentheses) requires adding extra parentheses (meticulously placing and counting them), whereas in the latter search it only requires proper placement of those terms.

Many search methods depend highly on the PICO framework. Research states that often PICO or PICOS is not suitable for every question (22, 23). There are other acronyms than PICO such as SPIDER (24), but each is just a variant. In our method, the most important and specific elements of a questions are being analyzed for building the best search strategy.

Though it is generally recommended that searchers search both MEDLINE and Embase, most use MEDLINE as the starting point. It is considered the gold standard for biomedical searching, partially due to historic reasons, since it was the first of its kind, and more so now that it is freely available via the PubMed interface. Our method can be used with any database as a starting point, but we use Embase instead of MEDLINE or another database for a number of reasons. First, Embase provides both unique content and the complete content of MEDLINE. Therefore, searching Embase will be, per definition, more complete than searching MEDLINE only. Secondly, the number of terms in Emtree (the Embase thesaurus) is three times as high as that of MeSH (the MEDLINE thesaurus). It is easier to find MeSH terms after all relevant Emtree terms have been identified than to starti with MeSH and translate to Emtree.

At Erasmus MC, the researchers sit next to the information specialist during most of the search strategy design process. This way, the researchers can deliver immediate feedback on the relevance of proposed search terms and retrieved references. The search team then combines knowledge about databases with knowledge about the research topic, which is an important condition to create the highest quality searches.

Limitations of the method

One disadvantage of single-line searches compared to multi-line search strategies is that errors are harder to recognize. However, with the methods for optimization as described (Step 11), errors are recognized easily because missed synonyms and spelling errors will be identified during the process. Also problematic is that more parentheses are needed, making it more difficult for the searcher and others to assess the logic of the search strategy. However, as parentheses and field codes are typed before the search terms are added (Step 10), errors in parentheses can be prevented.

Our methodology works best if used in an interface that allows for proximity searching. It is recommended that searchers with access to an interface with proximity searching capabilities select one of those as the initial database to develop and optimize the search strategy. Because the PubMed interface does not allow proximity searches, phrases or Boolean AND combinations are required. Phrase searching complicates the process and is more specific, with the risk of a higher chance of missing relevant articles, and using Boolean AND combinations increases sensitivity, but at an often high loss of specificity. Due to some searchers' lack of access to expensive databases or interfaces, the freely available PubMed interface may be necessary to use, though it should never be the sole database used for an SR (2, 16, 25). A limitation of our method is that it works best with subscription-based and licensed resources.

Another limitation is the customization of the macros to a specific institution's resources. The macros for the translation between different database interfaces only work between the interfaces as described. To mitigate this, we recommend using the find-and-replace functionality of text editors like Microsoft Word to ease the translation of syntaxes between other databases. Depending on one's institutional resources, custom macros can be developed using similar methods.

Results of the method

Whether this method results in exhaustive searches where no important article is missed is difficult to determine, because for any topic the number of relevant articles is unknown. A comparison of several parameters of 73 published reviews that were based on a search developed with this method with 258 reviews in which information specialists of other Dutch academic hospitals were acknowledged shows that the performance of the searches following the described method is comparable to those performed in other institutes but that the time needed to develop the search strategies was much shorter than the time reported for other reviews (9).

2.4 Conclusions

With the described method, searchers can gain confidence in their search strategies by finding many relevant words and creating exhaustive search strategies quickly. The approach can be used when performing SR searches or for other purposes such as answering clinical questions, with different expectations of the search's precision and recall. This method, with practice, provides a stepwise approach that facilitates the search strategy development process from question clarification to final iteration and beyond.

2.5 References

1. Harris MR. The librarian's roles in the systematic review process: A case study. J Med Libr Assoc. 2005;93(1):81-7.
2. Higgins JPT, Green S. Cochrane handbook for systematic reviews of interventions: Wiley Online Library; 2008.
3. University of York, N.H.S. Centre for Reviews and Dissemination. Systematic reviews : CRD's guidance for undertaking reviews in health care. York: CRD, University of York; 2009.
4. 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.
5. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. J Clin Epidemiol. 2009 Oct;62(10):1006-12.
6. MECIR. Methodological standards for the conduct of Cochrane Intervention Reviews. 2013 [updated December 2013; cited 2015 May, 7]; Available from: http://editorial-unit.cochrane.org/sites/editorial-unit.cochrane.org/files/uploads/MECIR_conduct_standards%202.3%2002122013.pdf.
7. Saleh AA, Ratajeski MA, Bertolet M. Grey Literature Searching for Health Sciences Systematic Reviews: A Prospective Study of Time Spent and Resources Utilized. Evidence Based Library and Information Practice. [systematic reviews; literature search; grey literature; time; health sciences]. 2014;9(3):28-50.
8. Erwin PJ. By The Clock: How Much Time Does an Expert Search Take? MLA news. 2004 October(370):1,12.
9. Bramer WM, Rethlefsen ML, Mast F, Kleijnen J. A pragmatic evaluation of a new method for librarian-mediated literature searches for systematic reviews. submitted. 2017.
10. Riva JJ, Malik KM, Burnie SJ, Endicott AR, Busse JW. What is your research question? An introduction to the PICOT format for clinicians. Journal of the Canadian Chiropractic Association. 2012;56(3):167-71.
11. Stern C, Jordan Z, McArthur A. Developing the review question and inclusion criteria. Am J Nurs. 2014 Apr;114(4):53-6.
12. Thabane L, Thomas T, Ye C, Paul J. Posing the research question: not so simple. Canadian journal of anaesthesia = Journal canadien d'anesthésie. 2009;56(1):71-9.
13. Bramer WM. Expert searching – Patient, Intervention, Comparison, Outcome (PICO) – an overrated tool – Column. MLA news 2015;55(2):6.
14. 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.
15. Methley AM, Campbell S, Chew-Graham C, McNally R, Cheraghi-Sohi S. PICO, PICOS and SPIDER: a comparison study of specificity and sensitivity in three search tools for qualitative systematic reviews. BMC Health Serv Res. 2014 Nov 21;14(1):579.
16. Bramer WM, Giustini D, Kramer BMR. Comparing the coverage, recall, and precision of searches for 120 systematic reviews in Embase, MEDLINE, and Google Scholar: a prospective study. Systematic Reviews. 2016;5(1):39.
17. Elsevier R&D Solutions. A Comparison of Emtree® and MeSH® - Whitepaper. 2015 [updated July 2015; cited 2015 November, 23]; Available from: http://supportcontent.elsevier.com/Support%20Hub/Embase/Files%20&%20Attachements/4685-Embase_White%20Paper_Comparison%20of%20Emtree%20and%20MeSH_July%202015.pdf.
18. Institute of Medicine, Committee on Standards for Systematic Reviews of Comparative Effectiveness Research, Eden J, Levit LA, Berg AO, Morton SC. Finding what works in health care standards for systematic reviews. Washington, D.C.: National Academies Press; 2011; Available from: http://www.ncbi.nlm.nih.gov/books/NBK209518/.
19. Niederstadt C, Droste S. Reporting and presenting information retrieval processes: the need for optimizing common practice in health technology assessment. Int J Technol Assess Health Care. 2010;26(04):450-7.
20. Rader T, Mann M, Stansfield C, Cooper C, Sampson M. Methods for documenting systematic review searches: a discussion of common issues. Research Synthesis Methods. 2014;5(2):98-115.
21. Hammerstrøm K, Wade A, Jørgensen A-MK. Searching for studies: A guide to information retrieval for Campbell Systematic Reviews The Campbell Collaboration 2010 [updated 19 August, 2010 cited 2016 3 April]; Available from: http://www.campbellcollaboration.org/lib/download/969/Hammerstrom_Guide_IR_Supplement.pdf.
22. Anderson LM, Oliver SR, Michie S, Rehfuess E, Noyes J, Shemilt I. Investigating complexity in systematic reviews of interventions by using a spectrum of methods. J Clin Epidemiol. 2013;66(11):1223-9.
23. Watine J, Couaillac JP, Lepargneur JP, Augereau C, Bunting PS, Oosterhuis WP. Evidence-based medicine (EBM) and clinical practice guidelines (CPG): About three misconceptions. Immuno-Analyse et Biologie Specialisee. 2012;27(5):279-82.
24. Cooke A, Smith D, Booth A. Beyond PICO: the SPIDER tool for qualitative evidence synthesis. Qual Health Res. 2012 Oct;22(10):1435-43.
25. Bramer WM, Giustini D, Kramer BM, Anderson PF. 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.

Chapter 3

De-duplication of database search results

Bramer WM, Giustini D, de Jonge GB, Holland L, Bekhuis T. De-duplication of database search results for systematic reviews in EndNote. J Med Libr Assoc. 2016;104(3):240-3.

3 De-duplication of database search results for systematic reviews in EndNote

When conducting exhaustive searches for systematic reviews, information professionals search multiple databases with overlapping content.(1-4) They typically remove duplicate records to reduce reviewer workload associated with screening titles and abstracts; sometimes the reviewers remove the duplicates. Several articles have been published recently on de-duplication methods. In our opinion, these methods are either very time consuming (5) or impractical, as they require uploading large files to an online platform.(6, 7) A recent overview article compared existing software programs, but found that none is truly satisfactory.(8)

Unique identifiers for journal articles are digital object identifiers (DOI) and PubMed IDs (PMID). However, these identifiers are not present in every database. When they are present, they often cannot be exported easily. Thus, they cannot be relied upon to identify duplicates. An alternative involves pagination because the often large page numbers in scientific journals, in combination with other fields, can serve as a type of unique identifier. However, this is complicated by variations in the way page numbers are stored. Most biomedical databases use a long format (e.g., 1008-1012), but two important databases (Medline and The Cochrane Library) use an abbreviated one (e.g., 1008-12).

The de-duplication method by Bramer et al. as presented in this article, was previously described in a brief conference paper.(9) The method consists of three stages. First, settings are changed for the displayed fields, and custom filters and export formats are installed. Second, several databases are imported into a temporary library and exported in an adapted format before being imported for de-duplication. The third stage consists of several subsequent changes in the settings for fields used to detect duplicates, followed by detailed instructions to accurately remove probable duplicates. In this article we describe this method in detail for EndNote (10), a popular reference manager.

3.1 Field settings and filters

Settings must be changed at the outset to optimize the EndNote configuration for de-duplication. These changes have to be made only the first time this method is used. Since page numbers play an important role, it is vital to show page numbers in the *library window*.

1. Go to Edit > Preferences > Display Fields.
2. Under Field, select Pages for one of the larger-numbered Columns.
3. Click OK.

We customized an EndNote style to create export files in which abbreviated page numbers are expanded, as well as an import filter to import the modified files. These filters should be installed prior to de-duplication.

1. Go to http://bit.ly/emcendnote.
2. Open the zip file.
3. Double-click _Correct Pages.ens (the file will open in EndNote).
4. In EndNote, click File > Save As.
5. Remove the text ‘copy’ from the file name and click save.
6. Close the style.
7. Repeat steps 3 through 6 for the file _Import Corrected Pages.enf.

3.2 Importing references

The exported references from MEDLINE, via PubMed (11) and Ovid,(12) and the Cochrane Library (13) are modified to adapt the page number format of references to the format used in most other databases.

1. Import all references from PubMed, Ovid MEDLINE, and the Cochrane Library into an empty EndNote library.
2. Select all references in this library (Ctrl-A).
3. Go to File > Export.
4. Select the output style: _Correct Pages and save the resulting file with the extension .txt.
5. Close the temporary library and create a final EndNote library where records from databases are imported as usual.
6. When importing MEDLINE and Cochrane reference sets, choose the newly created file and use the import filter _Import Corrected Pages.

3.3 De-duplication

1. Go to Edit > Preferences > Duplicates and select the fields to match the ones mentioned in row a of the second column of Table 3.1 and click [OK].
2. Click on All References and select one reference at random.
3. Go to References > Find Duplicates.
4. Click on [Cancel]. Duplicates will be highlighted but not yet removed.
5. Follow the steps as described in row a of the third column of Table 3.1.
6. Repeat the process again from step 1 onward for each row in Table 3.1

Table 3.1: De-duplication field settings and removal of duplicates

	Set Field Preferences	Steps to remove duplicates
a	Author | Year | Title | Secondary Title (Journal)	Press <Delete> to remove all selected duplicates without manual assessment
b	Author | Year | Title | Pages	Press <Delete> to remove all selected duplicates without manual assessment
c	Title | Volume | Pages	1. Manually assess the top references with blank title or author fields, using Ctrl-Click to deselect false duplicates. 2. Click on the column heading 'Pages' to sort all duplicate references by descending order of page numbers. 3. Review the top references without page numbers and those where the first page number is 1 for equivalent author names. If author names of subsequent references differ, deselect the marked false duplicates with <Ctrl-Click>. 4. Remove the selected duplicates with <Delete>.
d	Author |Volume | Pages	Repeat steps 1-2 as described in row c. Deselect the top references without page numbers by pressing <Ctrl-click> on the first highlighted reference, and <Ctrl-Shift-click> on the first highlighted reference with a starting page number greater than 1. Remove the remaining selected duplicates with <Delete>.
e	Year | Volume | Issue | Pages	1. Right Click on My Groups > Create Group and press <Enter>. 2. In the group Duplicate References, click on the column heading 'Pages' to sort all duplicate references by descending order of page numbers. 3. Select all references with page numbers by clicking on the top reference, holding <Shift> and then clicking on the last reference with page numbers. 4. Drag the selected references to the just created temporary 'New Group.' 5. Click on 'New Group.' Check the group for references with just one page and page numbers starting with a letter. Select from those references false duplicates and press <Delete> to remove them from the group. (They remain in All References, but are not deduplicated in this step.) 6. Select one of the references in 'New Group', click References > Find Duplicates, click Cancel and press <Delete> to remove all selected duplicates.
f	Title	Compare page numbers of consecutive references. If page numbers are present and different, examine journal titles and authors. Deselect false duplicates with <Ctrl-Click>. References with blank pages or pages starting with the number 1 are usually true duplicates, but check journal titles and author names when in doubt, especially when multiple consecutive blank pages are selected. After checking the entire list, remove the remaining selected duplicate references with <Delete>.
g	Author | Year	If a true duplicate is found, deselect all references by clicking the first true duplicate reference without holding <Ctrl>. Compare subsequent references on page numbers: if two adjacent references have the same page numbers, select the one with the largest record number with <Ctrl-Click>. After checking the complete list, remove the remaining selected references with <Delete>.

3.4 Discussion

Although the de-duplication method by Bramer et al. for EndNote resembles procedures regularly carried out by other information professionals, it is more systematic, rigorous, and reproducible. The steps may be somewhat challenging to master, but they become easy to carry out over time. The time spent de-duplicating references, as well as the error rate, are significantly reduced because just a small subset of the search results has to be assessed manually. Comparative data on the time needed and accuracy of the method relative to other methods will be presented in a forthcoming report.

To enhance efficiency and accuracy, the steps described should be followed closely in the order presented and without omission. The method's strength is based on the specificity of the first two steps which require no manual assessment. The next three steps require checking a small subset, i.e. the references that lack page numbers. The last two steps require some additional manual assessment, but screening by page numbers expedites the work, and the number of references to assess is lower than in other methods.

The limitation of this method is that it is tuned to EndNote. However, EndNote is commonly used to manage bibliographic records. The only alternative software to EndNote in which fields to be compared in the de-duplication process can be changed is Reference Manager, also provided by Thomson Reuters.(14) Reference Manager allows comparisons by start pages. However, when we tried our method in Reference Manager it failed; too many false duplicates were removed. The reason is that the comparison of start page numbers in Reference Manager appears to be flawed. Therefore, tailoring the de-duplication method by Bramer et al. for Reference Manager is not a good alternative. Additionally, Reference Manager is no longer for sale, and support for Reference Manager will likely be discontinued. We hope that EndNote will adopt some of Reference Manager's useful features, such as the option to regulate the amount of overlap in the title and other fields, and a comparison on start pages, albeit more robust than in Reference Manager.

The method described in this paper is for the most recent version of EndNote for Windows, version X7. It will also work in earlier versions (versions X3 and higher). However, in older versions, step 2 in the third column of rows c and d (Table 1) will not work as desired because duplicate references are not highlighted. Before executing the steps in row c for older versions, go to All References and sort this group by Page Numbers. Next, instead of clicking on the column heading 'Pages' as is described in step 2, go to All References and then go back to Duplicate References. Now the group Duplicate References will be sorted by Page Numbers. Then click on one of the scroll bars to reactivate the highlighting of duplicates and follow the other steps as described.

A requisite of this method is that for efficient de-duplication, page numbers should include both a start and end page. This is the reason we advise exporting the data from several databases into temporary files, which are then exported and reimported into the final EndNote library for de-duplication. Databases also differ in the format of the exported journal titles. Some databases use abbreviations, others provide full journal titles. We use customized import filters for several databases and interfaces, including Embase.com, Web of Science, CINAHL, and Scopus, to import the abbreviated journal titles into EndNote. Although this is not strictly necessary, it improves the sensitivity of the first step, and thus reduces the number of references that have to be checked manually. If databases would standardize their page numbers and journal titles, it would be possible to compare these data without the extra steps. To complicate matters, Cochrane recently switched from exporting full to abbreviated page numbers, and CINAHL recently appends the length of the article in the fields for page numbers (e.g., in the format 1008-1012 5p).

The de-duplication method by Bramer et al. is rather complicated and the learning curve is steep. However, simplification of the method (e.g., by reducing the number of different field combinations, or by not normalizing page numbers) increases the workload by increasing the number of references to manually assess. If performed frequently, librarian-mediated de-duplication services can be faster than current methods and less error prone.

3.5 References

1. Levay P, Raynor M, Tuvey D. The Contributions of MEDLINE, Other Bibliographic Databases and Various Search Techniques to NICE Public Health Guidance. Evidence Based Library and Information Practice. 2015;10(1):50-68.
2. Ahmadi M, Sarabi RE, Orak RJ, Bahaadinbeigy K. Information Retrieval in Telemedicine: a Comparative Study on Bibliographic Databases. Acta Inform Med. 2015;23(3):172-6.
3. Lorenzetti DL, Topfer L-A, Dennett L, Clement F. Value of databases other than medline for rapid health technology assessments. International journal of technology assessment in health care. 2014;30(02):173-8.
4. Beyer FR, Wright K. Can we prioritise which databases to search? A case study using a systematic review of frozen shoulder management. Health information and libraries journal. 2013;30(1):49-58.
5. Qi X, Yang M, Ren W, Jia J, Wang J, Han G, et al. Find duplicates among the PubMed, EMBASE, and Cochrane Library Databases in systematic review. PLoS One. 2013;8(8):e71838.
6. Rathbone J, Carter M, Hoffmann T, Glasziou P. Better duplicate detection for systematic reviewers: evaluation of Systematic Review Assistant-De-duplication Module. Systematic reviews. 2015;4(1):6.
7. Jiang Y, Lin C, Meng W, Yu C, Cohen AM, Smalheiser NR. Rule-based de-duplication of article records from bibliographic databases. Database : the journal of biological databases and curation. 2014;2014:bat086.
8. Kwon Y, Lemieux M, McTavish J, Wathen N. Identifying and removing duplicate records from systematic review searches. J Med Libr Assoc. 2015;103(4):184-8.
9. Bramer WM, Holland L, Mollema J, Hannon T, Bekhuis T. Removing duplicates in retrieval sets from electronic databases: comparing the efficiency and accuracy of the Bramer-method with other methods and software packages. Proc 14th European Association for Health Information and Libraries (EAHIL) Conference, 57-9; June 11-13; Rome, Italy. 2014.
10. EndNote X7. Thomson Reuters; 2015.
11. PubMed: National Library of Medicine; 2016 [cited 2016 4 Feb]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/.
12. Ovid MEDLINE: Wolters Kluwer; 2016 [cited 2016 4 Feb]. Available from: http://site.ovid.com/pdf/OvidMedline_fs.pdf.
13. The Cochrane Library: Cochrane; 2016 [cited 2016 4 Feb]. Available from: http://www.cochranelibrary.com/.
14. Reference Manager Version 12. Thomson Reuters; 2008.

Chapter 4

Reviewing retrieved references for inclusion

Bramer WM, Milic J, Mast F. Reviewing retrieved references for inclusion in systematic reviews using EndNote. J Med Libr Assoc. 2017;105(1):84-7

4 Reviewing retrieved references for inclusion in systematic reviews using EndNote

An important task in conducting a systematic review is reading titles and abstracts of the retrieved references, which often number in the thousands, to determine which articles meet the predefined inclusion criteria. In the past, this was performed by manually scanning through large stacks of printed titles and abstracts, followed by face-to-face meetings to discuss which references should be included. Today the workflow of the review process is more streamlined by using computer software.

Several specialized solutions for this process are available, most notably free or subscription-based online tools, such as Covidence (www.covidence.org), DistillerSR (www.distillercer.com), or Rayyan (rayyan.qcri.org). The Cochrane collaboration uses its own tool called ReviewManager. A survey in 2013 showed that more than half of all systematic reviewers use EndNote software. (1) This dominance on the market is likely to increase as sales of the second most popular tool, Reference Manager, have ceased, and its website (www.refman.com) now advises users to switch to EndNote. Many reviewers use Microsoft Excel. Some libraries have even created specialized Excel Workbooks to document the process in much detail.(2) A method for the inclusion process using EndNote is described by King et al. (3), but this process is rather complicated and time-consuming.

Below, we describe the logistics of a method to perform the title and abstract screening, verdict assignment, and result comparisons between multiple reviewers in EndNote. The process is blinded; each reviewer works in his or her own EndNote file and after the individual inclusion and exclusion process, the verdicts of the different reviewers are compared. The method described here can be performed much faster than comparable methods.

4.1 Description of the method

The method consists of several steps. First, a custom style should be installed for easy abstract scanning. Second, a field should be added to show the reviewer's name in the Library window. Third, for each systematic review, custom groups are made in the EndNote library for included and excluded references. Each reviewer drags articles to the group corresponding to their verdict. In the last steps for the final comparison of the verdicts, the included references of all reviewers are combined into one EndNote library and deduplicated. References found as duplicates were included by both reviewers, and are selected for full text reviewing, the non-duplicate references can then be discussed for inclusion or exclusion.

The steps below are written for EndNote X7 for Windows. EndNote is also available in a version for Mac users, but some of the menu items will appear on different places. We have added footnotes that guide Mac users as much as possible.

Step 1 Install the custom-made output style

A custom-made style (named _preview) facilitates easier reviewing of the titles and abstracts.

1. Visit http://bit.ly/emcendnote
2. Open the zip file
3. Double-click on the file *_preview.en*s (it will open in EndNote) [1]
4. In EndNote, click *File > Save As*
5. Remove the text '*copy*' from the file name and click [Save] [2]
6. Close the style using the cross in the top right corner [3]
7. To activate the style *_preview*, open the drop-down menu (called *Select Bibliographic Output Style*) in the top left part of the screen. [4]
8. Click on Select another style. Scroll to the top, select *_preview* and click [Choose].
9. When selecting a reference, the abstract will be displayed in the preview tab in the Tab pane

Step 2: Change the settings of the library window

When comparing the included references between reviewers, a special field will be used to document the name of the reviewer who included a certain reference. To be able to view this in the library window, this field has to be added to the preference settings.

1. Go to *Edit > Preferences > Display Fields*
2. In column 8 (or an alternative field that is not often used), under Field, select *Custom 4*
3. In the same column, under Heading, type *reviewer*

Step 3: Create EndNote library with groups for inclusion/exclusion

In the first reviewing round, two reviewers independently read titles and abstracts to decide whether a reference is potentially relevant to the review. We propose the creation of two group sets: includes and excludes, each with, at this stage, just one subgroup (Figure 4.1).

1. Right-click on My groups and select Create Group Set
2. Type excludes and hit <enter>
3. Right-click on excludes and select Create Group
4. Type title/abstract and hit <enter>
5. Repeat the process for Group Set includes with Group includes

After the groups for inclusion and exclusion have been made, two copies of these files are created (adding the name of the reviewer to the file name) and distributed to the reviewers. Each reviewer will work in his or her own copy of the EndNote file.

[1]For Mac users: The file must be download and saved locally before it can be opened.
[2]For Mac users: The text 'copy' will not appear in the file name.
[3]For Mac users: The cross appears in the top left corner.
[4]For Mac users: At the bottom right of the screen is a preview pane. The style selection menu appears at the top of the pane.

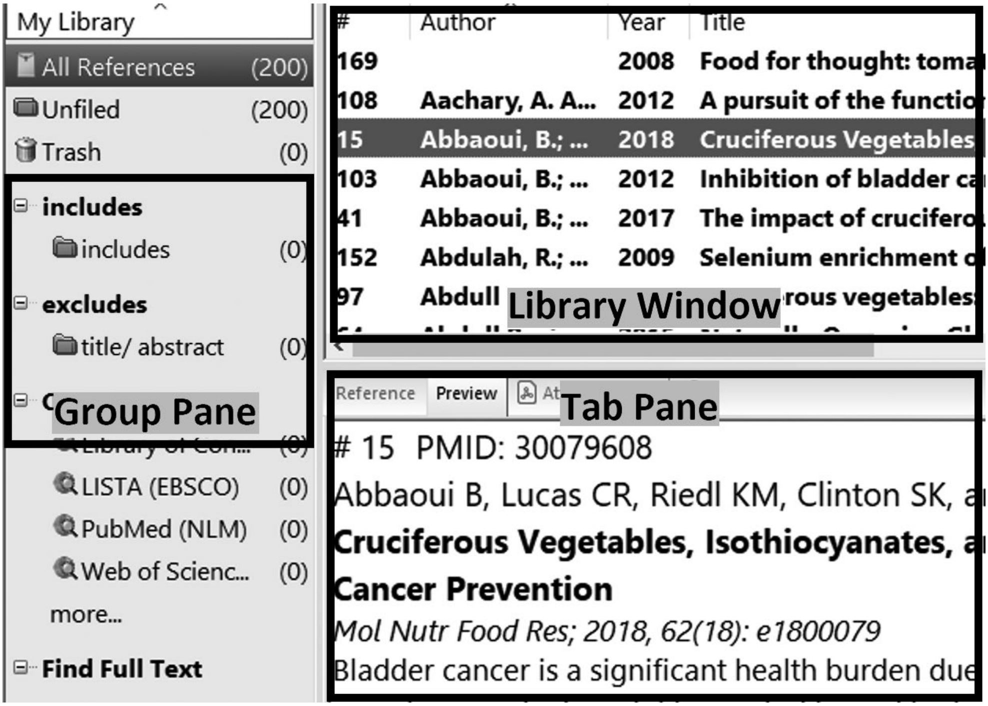


Figure 4.1: Groups for Includes and Excludes

Step 4: Title/abstract screening

The standard group named *Unfiled* contains references not yet assigned to one of the other groups. When starting the screening phase, the number of references in *Unfiled* will be equal to that in *All References*. Each reviewer reviews the relevance of references in *Unfiled* one by one based on the title and/or abstract.

1. Broaden the title field in the Library window (Figure 4.1) by dragging the column break between the columns of Title and Journal to the right until the Title column reaches an appropriate width.
2. Review the titles one by one in the Library window until a potentially relevant title is reached, without yet assigning references to the groups.
3. Click on the relevant title and read the abstract in the preview tab of the Tab pane (Figure 4.1).
4. If the abstract is irrelevant, continue reading titles in the library window.
5. If the abstract is relevant, select the article directly above that relevant article in the Library window and press <Ctrl-Shift-Home> to select all references above it. Drag all of these articles into the group excludes > *title/abstract*.
6. Next, drag the top reference in the library window (which is the reference to be included, as irrelevant references were removed from *Unfiled*) to the group includes > includes.
7. Repeat this process until all references are filed and the group Unfiled is emptied.

Step 5: Compare included references between reviewers

1. Go to the folder *includes* in the EndNote library of reviewer 1 and click on one of the references
2. Go to Tools > *Change / Move / Copy Fields*
3. In *Custom 4*, select *Insert after field's text*: and type the first name of reviewer 1 and click on [OK] in three pop-up screens.
4. Open the EndNote library from reviewer 2 without closing that from reviewer 1
5. Go to the group *includes* in the library of reviewer 2. Select all references in that group, right click on one of them and select *Copy reference to* and select the file screened by reviewer 1
6. Go to *Window* and select the file by reviewer 1.
7. Go to the group Unfiled and mark the records in that group with the name of reviewer 2 (as described above in steps 2-3)
8. Drag the references from *Unfiled* into the group *includes*
9. Check the settings for de-duplication (*Edit* > *Preferences* > *Duplicates*). At least Author, Year, Title, and Secondary Title (Journal) should be selected
10. Go to *includes*, select a random reference, and go to References > *Find Duplicates*. In the detailed comparison screen click [Cancel]. Then press <Delete> on the keyboard to remove the duplicate references
11. Right click in the group set *includes*, select *Create a group*, and name it *definite includes*
12. Select all references in the group Duplicates References and drag those to the group definite includes
13. Go to the group *definite includes* and mark the records in that group with the name of reviewer 2 (as described above in steps 2-3)
14. Right click on the group *includes* > *discussion* and select *delete group*
15. The articles currently in the group *unfiled* need to be discussed by the two reviewers
16. After consensus is reached, drag the references one by one to the appropriate group until *unfiled* is empty

Step 6: Full text reviewing

In the second round of screening, full texts of the included titles and abstracts need to be reviewed. Custom groups can be used to distinguish between various reasons for exclusion, and articles can be assigned to specific groups for certain subquestions. Each reviewer should again work in his or her own copy of this library. After reading all articles, each reference in the library should be discussed in detail, therefore no automatic comparison should be used. The steps in this round are more laborious, differ per research topic, and can hardly be generalized and optimized. Therefore, we do not describe in detail how this process can be executed.

4.2 Discussion

We advise against the use of a group 'doubt' or 'maybe' for articles for which it is not yet clear whether they should be *included*. We recommend that in cases of doubt, the article should be added to the folder includes. If the second reviewer also had doubts about the relevance or decided to include the article, the full text should be used for final judgment. If the second reviewer excluded the article, it is an item for discussion.

The process of reading titles and abstracts for inclusion and exclusion is often considered time-consuming, and the number of abstracts that can be read per hour is estimated at 120 by the Cochrane Handbook.(4) A recent study recorded the time needed for certain steps in the SR process.(5) Upon request, the authors informed us that, using specialized screening software, the median number of articles that could be reviewed per hour was 68. Another recent study estimates the time needed to screen one record on title and/or abstract at 1 minute.(6) A survey among review authors at Erasmus MC to whom we had sent an earlier draft of the article reported that the median number of titles and/or abstracts reviewed per hour with the present method was 308, with a maximum of 675.

We found that the speed of the process increases when reviewers do not document the specific reasons for excluding references during the title/abstract screening phase. It is often clear that an article is not relevant to the topic, but to determine the exact reason (or very often multiple reasons) is very time-consuming and unnecessary. According to PRISMA guidelines, reasons for exclusion, with the number of articles for each reason, should be given only in the full text screening phase.(7) We found that researchers who meticulously documented the reasons for exclusion in the first round, and those who had used other software such as Microsoft Excel, reported much slower rates (median of 60 minutes) and often later regretted their decision to do so. Also, using specialized programs or online SR management systems such as Covidence or DistillerSR unnecessarily complicate and delay the process because each abstract has to be assigned to categories individually, in contrast to our bulk assignment of non-relevant articles.

4.3 References

1. Lorenzetti D, Ghali W. Reference management software for systematic reviews and meta-analyses: an exploration of usage and usability. BMC Medical Research Methodology. 2013;13(1):141.
2. VonVille HM. Excel Workbooks for Systematic Reviews & Corresponding Handouts University of Texas; [cited 2016 April 19]; Available from: http://libguides.sph.uth.tmc.edu/excel_SR_workbook.
3. King R, Hooper B, Wood W. Using bibliographic software to appraise and code data in educational systematic review research. Medical teacher. 2011;33(9):719-23.
4. Higgins JPT, Green S. Cochrane handbook for systematic reviews of interventions: Wiley Online Library; 2008.
5. Wang Z, Asi N, Elraiyah TA, Abu Dabrh AM, Undavalli C, Glasziou P, et al. Dual computer monitors to increase efficiency of conducting systematic reviews. Journal of clinical epidemiology. 2014 12//;67(12):1353-7.
6. Shemilt I, Khan N, Park S, Thomas J. Use of cost-effectiveness analysis to compare the efficiency of study identification methods in systematic reviews. Systematic reviews. 2016 2016//;5(1):1-13.
7. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Journal of clinical epidemiology. 2009 Oct;62(10):1006-12.

Chapter 5

Updating search strategies

Bramer WM, Bain P. Updating search strategies for systematic reviews using EndNote. J Med Libr Assoc. 2017;105(3):285-9.

5 Updating search strategies for systematic reviews using EndNote

Performing, writing, and publishing a systematic review takes a long time. In a cohort of journal-published systematic reviews, Cochrane reviews, and health technology assessment reports, the median time lag between the stated last search date and publication was 61 weeks (interquartile range, 33-87 weeks) (1). In the same cohort of reviews, 7% were out of date at the time of publication (2). More recently, an examination of 182 systematic reviews performed at Erasmus MC showed that the median time between the first search and the appearance of the resulting review in PubMed was 89 weeks (interquartile range, 63-126 weeks). To maximize the currency of a review, an update of the search is recommended before submission for publication. The Methodological Expectations of Cochrane Intervention Reviews (MECIR) standards requires: "Rerun or update searches for all relevant databases within 12 months before publication" (3). Many handbooks and guidelines for performing systematic reviews state that search strategies should be updated regularly to keep track of newly added references on the topic (4-6).

Recent guidance from an international panel of authors, editors, clinicians, statisticians, information specialists, other methodologists, and guideline developers considered various aspects of updating reviews, including efficient searching. Such efficiencies include refinements based on the yield of the original search and incorporation of technological advances in searching. Garner et al. provide practical guidance on refining the original search in their Appendix 2 (7).

The Cochrane handbook (8) mentions in chapter 3.4.2.1 ("Re-executing the search") using the last date of the original search as the beginning date for the update, which is common practice, but chapter 6.4.12 ("Updating searches") does not describe a clear method. The date the record became accessible through searching, rather than publication date, is the relevant field for updating. One can choose from the thesaurus date (i.e. the date that the thesaurus terms were added), the date of the last metadata change, or the date of entry into the database. For example, the National Library of Medicine recommends using the Create date (CRDT in PubMed) field for its databases (7). However, the field MeSH date (MHDA in PubMed), which is the date MeSH terms were added to the record, also has some advantages. However, some interfaces, such as Web of Science, do not provide record dates that could guide updating. In such cases, searchers may use publication date and a safe overlapping period, resulting in extensive duplication with records retrieved in the original search.

Complicating matters further, the search may have been modified since the last search date. For instance, new words may have been added to the original search strategy based on relevant terms found in studies included in the original review. These novel terms need to be searched in all the databases that were queried in the original search from the original starting date, thus requiring even more complicated search structures and date ranges. Hence, to many authors, updating a search may seem to be a complicated and uncertain task. It need not be.

We have developed a method for updating existing reviews that uses EndNote reference management software (9). The technique uses two EndNote files: one containing the current results as they are downloaded from the complete set of databases, as if it were a first search, and one with the results of the previous or original search. By subtracting records found in the original search from the current results through EndNote's customizable de-duplication feature, only records not screened in the original search will remain in the library. Another group has previously alluded to a similar process: "…download all references from the update search and directly de-duplicate them with the references from the initial search (e.g. using Endnote)" (10). Here, we describe the process in step-by-step detail. The steps will be identical in any recent version of EndNote. The method was first developed with the Microsoft Windows version of EndNote X3 and has been fairly consistent until the current version, X8. The same steps are also applicable for the Macintosh editions by replacing the standard Windows controls with the corresponding Macintosh controls: for instance, command-A instead of control-A or command-click instead of control-click.

5.1 A new method for updating existing searches

The initial search requires no extra action. We do recommend thorough de-duplication in EndNote by the process described in an earlier article by Bramer et al. (11). This earlier article also describes how substantial differences in the way single articles are represented in various databases can be resolved. It describes how page numbers from MEDLINE and the Cochrane Library can be completed, turning abbreviated pagination (e.g., 1008-12) into full format pagination (e.g., 1008-1012) as is used in other databases. The earlier article also recommends importing full journal titles instead of abbreviations. If these steps are followed, the method described in this article will be easier to follow. The following description will still be effective if the libraries were de-duplicated using other sequences within EndNote but will require additional care to ensure that novel articles are not eliminated.

5.2 The method

The first steps in the updating process serve to create an EndNote file containing all results from the new or current search as they are retrieved on the new date. This search's date should now be considered the last date searched, and the number of records retrieved from this search will be used in the published PRISMA flow chart. In the process, a compressed library (.enlx file) can serve as an archive and be used in subsequent updates to identify and remove previously screened records.

Step 1: Re-run the search.

1. If search terms are to be changed or added, make those changes to the search strategies.
2. Run the searches in all databases from the original starting date (rather than the current date) and import all results in EndNote.
3. De-duplicate the EndNote file (preferably as described by Bramer et al. (11)).
4. Create a compressed EndNote library (.enlx file) of the complete search results and store for possible use in future updates. This set of results should be reported in the published review as the number of records reviewed by title and/or abstract.

A method to review titles and abstracts in EndNote is published in a previous article by Bramer et al. (12). Records from the original search have already been screened and need not be seen again. To remove these records from the current search results, the old records are added to the EndNote library containing the records from the new search. By means of duplicate detection, matching pairs of records (one from the original search and the other from the new search) can be identified and removed, leaving only new records that have yet to be screened.

Step 2: Copy the original search results in the new EndNote file.

1. Open the most recent version of the EndNote library containing all results found on the last search date. If this was an update of an earlier search, use the complete compressed library that was created at that time.
2. Press <Ctrl-A> to select all references. Right-click on the references and from the resulting menu, select *Copy References* to. Now select the EndNote library containing the results of the current search. This will move all the records in this library to the new library. Copied records will appear in a new group called *Copied References*. Using the *Copy References* to command instead of copying and pasting or importing records facilitates the identification of records from the previous search.

Step 3: Identify and remove records retrieved by both the previous and new searches.

1. Go to the EndNote file containing the old and new references
2. Modify the default duplicate settings by going to Edit > *Preferences* > *Duplicates* (the menu option *Preferences* can be found under the *EndNote* menu on the Mac). Select the fields: Author, Year, Title, Secondary Title (Journal).
3. Select one random reference in the group All References to ensure that this group is active.
4. Go to References > *Find Duplicates*. Click on [Cancel] in the *Find Duplicates* dialog to reveal the *Duplicate References* group.
5. Press <Ctrl-A> to select all references in the *Duplicate References* group. Press <Delete> to remove all references in the group, or drag the references to the trash.
6. Repeat step 3: 2-5 for the fields Author, Year, Title, Pages.

If the original search was executed in the recent past and had used the same methods for import into EndNote, the group *Copied References* should be nearly empty after this step. If the original search was performed long before the update took place or the initial search had been performed by another searcher, possibly using different export methods and interfaces, there may be a large number of remaining unmatched references from the original search. If so, follow the two extra steps of the method described in Bramer et al. (11): Step 3: 2-5 can be repeated with the fields Title, Volume, Pages and then with the fields Author, Volume, Pages. However, in these cases, references without page numbers should be assessed for duplicates, and true duplicates should be manually selected. For that purpose, after step 3: 4, click on the column heading Pages to sort by page number. References with page numbers can be deleted without manual assessment. If the column heading *Pages* is not visible in the reference table, go to *Edit* > *Preferences* > *Display Fields* and select in one of the columns *Pages*.

Step 4: Compare old and new records on similar titles or authors.

In the next steps, the new results are compared with the original results on title. This requires quick manual assessment to prevent removing two new, independent search results that share the exact title.

1. Go to E*dit* > *Preferences* > *Duplicates* and select only the field Title (The menu option Preferences can be found under the *EndNote* menu on the Mac).
2. Select one random reference in the group *All References* to ensure the group is active.
3. Go to *References* > *Find Duplicates* and click <*Cancel*> in the next screen.
4. Go to the group *Copied Reference*s, select one reference and press <Ctrl-A> to select all references from the previous search.
5. Go to the group *Duplicate References* to scan this set for duplicates. Click on the column heading *Title* to sort by title. True duplicates can be easily found by observing the pattern of blue and white lines. As long as white lines and blue lines are in a somewhat regular, one-by-one pattern, the pairs will consist of one reference from the old results and one reference from the new results. Occasionally groups of two blue or two white lines might appear because the sorting is irregular. If, however, a group of more than two white lines is observed, or two consecutive groups of two white references, this means that some of these references are not from the previous search results but are two current records with identical titles that represent different articles. These should not be selected for deletion. Figure 5.1 shows an example of such a pattern change. In this case, the references from authors Flexman and Afiilalo should not be deleted, as they are not duplicates.
6. Select references in sequences with regular blue-white patterns by using <Ctrl-Click> to select the beginning of such a sequence and <Shift-Click> to end a selection. This might require some practice if the user is unfamiliar with this option.
7. After all regular sequences have been selected, press <Delete> to remove all duplicate references or drag the duplicate references to the trash.

8. Repeat step 4: 1-7 for the field combination Author, Year.
9. Go to the folder Copied References and delete any references in that folder. These are the references that were found in the previous search but that were not found in the current search. They should not be reviewed again.

183	Gialanella, B....	2018	Conservative treatment of rotator cuff...	Eur J Phys Rehabil ...	900-910
7995	Gialanella, B....	2018	Conservative treatment of rotator cuff...		
7159	Baskaya, M. ...	2018	The efficacy of mirror therapy in patie...		
20	Baskaya, M. ...	2018	The efficacy of mirror therapy in patie...	J Back Musculoskel...	1177-1182
7804	Ellegaard, K.; ...	2013	Exercise therapy and ultrasound guid...	Arthritis and rheu...	S1162
5680	Ellegaard, K.; ...	2013	Exercise therapy and ultrasound guid...	Arthritis and rheu...	S1162-
1880	Gross, A.; Kay...	2015	Exercises for mechanical neck disorders	Cochrane Database...	
2606	Kay, T. M.; Gr...	2012	Exercises for mechanical neck disorders	Cochrane Database...	CD004250
8149	Handoll, H. H...	2012	Interventions for treating proximal hu...	Cochrane Database...	CD000434
2181	Handoll, H. H...	2010	Interventions for treating proximal hu...	Cochrane Database...	
8042	Goulet, C.; Ro...	2016	Sling-based Exercise for External Rota...	J Sport Rehabil	1-17
1586	Goulet, C.; Ro...	2018	Sling-based Exercise for External Rota...	J Sport Rehabil	30-36
110	Paavola, M.; ...	2018	Subacromial decompression versus di...	BMJ	k2860
9532	Paavola, M.; ...	2018	Subacromial decompression versus di...	BMJ	k2860
7844	Farfaras, S.; S...	2018	Subacromial Decompression Yields a ...	Am J Sports Med	1397-1407
164	Farfaras, S.; S...	2018	Subacromial Decompression Yields a ...	Am J Sports Med	1397-1407
1174	Lenza, M.; Bu...	2013	Surgical versus conservative intervent...	Cochrane Database...	
1094	Lenza, M.; Bu...	2019	Surgical versus conservative intervent...	Cochrane Database...	
6780	Fuller, L.; El-A...	2017	Upper limb rehabilitation after lung tr...	Transplant internat...	30-
7926	Fuller, L.; El-A...	2017	Upper limb rehabilitation after lung tr...	Transplant internat...	30
8071	Greiner, B. A.;...	2018	Work-Related Upper Limb Symptoms i...		
1337	Greiner, B. A.;...	2019	Work-Related Upper Limb Symptoms i...	Phys Ther	62-73

Figure 5.1: Recognition of pattern change in blue and white references, indicating a duplicate that should not be removed; the references in the black boxes are updates and should all be removed

5.3 Correct representation in the PRISMA flow chart

A consequence of this method is that the reference numbers and totals required for the PRISMA flow chart may not match, as the sum of the novel and old records may not equal the contents of the new EndNote library. This discrepancy can be attributed to several normal database activities: updates to individual records (such as the application of MeSH terms), global changes to controlled vocabulary, recent addition of older material to the database, or the dropping out of high relevance ranking in Google Scholar. Additionally, records with publication dates that antedate the previous search will appear among the novel records. This can be the result of both changes to controlled vocabulary and the recent addition of older material to the databases. Revision of and addition to the search strategy will also result, appropriately, in the retrieval of novel records from before the date of the previous search as well. Hence, the most reliable method of correcting the numbers is backwards correction.

Typically, the total number of unique records retrieved is reported both in the results section of a systematic review and in the PRISMA flow chart detailing the search and subsequent screening. For updated searches, the most appropriate number to report is the total number of records remaining in the updated search after removing duplicates. This number represents the minimum number of records screened for all searches. For dates of coverage, authors need only report the beginning date of coverage for each database and the date of the latest update, as the numbers of records is accurate for all searches at this point in time. The total number of full-text articles read will be the sum of the articles read in all the previous searches plus the number of articles read in current search. The numbers of records or articles excluded at each step in the work flow can be summed over the previous and updated searches.

The records described in step 4: 9 are records that had been found during the original search but that were not found in the update search. This situation may occur if a search strategy during the update is narrowed compared to the original search strategy or if certain databases that had been searched initially were not searched during the update. If the search strategy was narrowed, it is necessary to assure that the included references from the original search results are still retrieved by the current searches.

5.4 Discussion

We provide a method for the simple subtraction of previously screened results from updated systematic review searches. Our method is described for use with EndNote bibliographic management software, although conceivably it might be adapted to other bibliographic managers. However, for optimal implementation of this method, such a program requires customizable settings for duplicate detection and the option to remove both duplicate references. We are unaware of a bibliographic manager that is as flexible as EndNote in this regard. Most software uses a standard de-duplication algorithm and is set to merge duplicate references instead of removing them. By using this method, searchers spare themselves the time and effort required to reconcile update dates across platforms.

There is no consistent, widely accepted method of updating searches. Our method suggests a simple standard for both carrying out and reporting updates that requires giving only the total number of records from the inception of the original search to the date of the last update, along with any revisions to the search itself. In reporting only these details, authors give an accurate representation of the state of the database and the response to the query on the date of the last search. The only information that is lost is the number of records retrieved and screened in previous searches that did not match records in the updated query. Re-running and de-duplicating the updated search obviates the confusing task of choosing and reporting a date field for the beginning date of the new search (e.g., publication date, thesaurus date, creation date).

5.5 References

1. Sampson M, Shojania KG, McGowan J, Daniel R, Rader T, Iansavichene AE, et al. Surveillance search techniques identified the need to update systematic reviews. Journal of clinical epidemiology. 2008;61(8):755-62.
2. Shojania KG, Sampson M, Ansari MT, Ji J, Doucette S, Moher D. How quickly do systematic reviews go out of date? A survival analysis. Ann Intern Med. 2007;147(4):224-33.
3. Chandler J, Churchill R, Higgins J, Lasserson T, Tovey D. Methodological standards for the conduct of new Cochrane Intervention Reviews: The Cochrane Library; 2013 [updated 2 December 2013; cited 2017 20 January]. Available from: http://editorial-unit.cochrane.org/sites/editorial-unit.cochrane.org/files/public/uploads/MECIR_conduct_standards%202.3%2002122013_0.pdf.
4. Moher D, Tsertsvadze A, Tricco A, Eccles M, Grimshaw J, Sampson M, et al. When and how to update systematic reviews. Cochrane Database of Systematic Reviews. 2008(1).
5. Shojania KG, Sampson M, Ansari MT, Ji J, Garritty C, Rader T, et al. Updating Systematic Reviews. Rockville MD2007 Sep.
6. Stovold E, Beecher D, Foxlee R, Noel-Storr A. Study flow diagrams in Cochrane systematic review updates: an adapted PRISMA flow diagram. Systematic reviews. 2014;3:54-.
7. Garner P, Hopewell S, Chandler J, MacLehose H, Akl EA, Beyene J, et al. When and how to update systematic reviews: consensus and checklist. BMJ (Clinical research ed). 2016;354.
8. Higgins JPT, Green S. Cochrane handbook for systematic reviews of interventions: Wiley Online Library; 2008.
9. Thomson Reuters. EndNote X7. 2015.
10. European Network for Health Technology Assessment. Process of information retrieval for systematic reviews and health technology assessments on clinical effectiveness Available from: http://www.eunethta.eu/sites/5026.fedimbo.belgium.be/files/2015-07-13_Guideline_Information_Retrieval_final.pdf.
11. Bramer WM, Giustini D, de Jonge GB, Holland L, Bekhuis T. De-duplication of database search results for systematic reviews in EndNote. J Med Libr Assoc. 2016;104(3):240-3.
12. Bramer WM, Milic J, Mast F. Reviewing retrieved references for inclusion in systematic reviews using EndNote. J Med Libr Assoc. 2017;105(1):84-7.

Chapter 6

Reference Checking

Bramer WM. Reference checking for systematic reviews using Endnote. J Med Libr Assoc. 2018;106(4):542-6.

6 Reference Checking For Systematic Reviews Using Endnote

In searches for systematic reviews, it is recommended that authors review references from the reference lists of retrieved relevant reviews for possible additional, relevant references. This process can be time consuming, since there often is overlap between the reference lists and the lists contain references that were already retrieved in the initial searches. The author proposes a method in which EndNote is used in combination with the Scopus or Web of Science databases to semi-automatically download these references into an existing EndNote library.

With searches for systematic reviews, most guidelines recommend that, next to searching the traditional databases, alternative methods are used to find relevant articles. One alternative method is to review the reference lists of relevant reviews and already included citations to find articles that were not yet retrieved.(1-3) In the experience of Chapman, et al., (4) the new content can add 5-10% of the citations found. These extra references do add to the burden for the reviewer, especially since items on the reference lists often overlap and might have already been found by the database searches. Nevertheless, other studies have found that searching reference lists can add great value to database searches, and should therefore not be omitted.(5) We propose a method that allows for deduplication between these reference lists both internally and against the existing content that was downloaded from bibliographic database searches.

Chapman, et al. investigated the validity of a similar method (4), compared to manual screening of the reference lists. They found that the method saved the time needed to screen the reference lists by 62.5%. However, they do not provide enough detail on how to execute each step, and the time needed for their method seems substantial. In this article, we describe in detail the steps that can be used to efficiently perform this method. The method described by Chapman, et al., has been improved by automating the Scopus search with a specially designed output style from EndNote that allows for automatic searching articles from an existing EndNote library in Scopus, allowing for faster downloading of the reference lists than reported in their article.(4)

6.1 Automatic downloading of reference lists of included references

Several databases allow users to download cited or citing references of a list of articles. Two of such databases are Scopus by Elsevier(6) and ISI Web-of-Science by Clarivate Analytics.(7) They both need a subscription to access. If a researcher or librarian has access to one of these databases, custom made export formats from EndNote can create strategies for the selected references in those databases, leading the searcher to extra relevant articles.

For this purpose, the functionality of Scopus is superior to that of Web-of-Science, because Scopus can export the cited and citing references of a list of articles simultaneously. Web-of-

Science, on the other hand, is only able to deliver those lists per individual reference, making the process much more cumbersome and time-consuming. Using Web-of-Science, all reference lists are exported individually, and therefore the reference lists also have to be deduplicated internally before comparing to the existing list of references retrieved from the bibliographic database. Results exported from Scopus are already internally deduplicated, so that an article cited by more than one reference is present only once.
Also, our experience found that Scopus was able to retrieve more articles via this search option than Web-of-Science. Nevertheless we describe the methods in both Scopus and Web of Science.

6.2 The method

Step 1: Change the settings of EndNote as preparation

First the settings of EndNote have to be prepared for this method. Special Output Styles have to be installed. Also a column showing page numbers of references in EndNote should be made visible. This allows the selection of papers that have page numbers as they will be searched differently than those that have not.

1. Go to bit.ly/emcendnote.
2. Open the zip file.
3. Open the file _scopus search.ens
4. In EndNote select File > Save as.
5. Remove Copy from the file name and Click [Save]
6. Repeat steps 3-5 for the files _scopus pmid.ens, _scopus title.ens, _wos search.ens, _wos pmid.ens and _wos title.ens (depending on the platform you choose to use).
7. Go to Edit > Preferences > Display Fields
8. For one of the columns (preferably one of the higher numbered columns), select Pages from the drop down menu and click OK.
9. Change the deduplication settings: Go to: Edit > Preferences > Duplicates. Select Year, Title and click OK.

Step 2: Prepare an EndNote library containing the relevant references

Create a new EndNote library that contains only those articles for which you want to perform the reference check: included references and relevant reviews. In that EndNote library, create a group resolved and a group unmatched (right click on *My Groups* > *Create Group*).

Step 3: Use EndNote to create a search for the relevant references

In the following steps EndNote creates a search strategy that will search for the relevant landmark references in the database of choice. Generally the results from systematic review searches and therefore the endnote representation of included references originate from many different databases and interfaces. All of these databases store their data in different ways. It can therefore be hard to find these reference directly in the databases. This means the search has to be done in multiple steps.

1. From the drop down menu in the top left corner of EndNote (Style selection) select the style _scopus pmid when using Scopus.
   If you plan to use Web of Science, select the style _wos pmid
2. Select one random reference in All References and press Ctrl-A to select all references. And Press Ctrl-K to copy the search to your clipboard.
3. In Scopus: go to Advanced, Click on Enter query string, and press Ctrl-V to paste your search strategy and click Search

In Web of Science: Select Advanced Search from the drop down menu. Paste the search strategy in the search field. Remove the last occurrence of the Boolean operator 'OR' and click Search.

Step 4: Mark the references that have been solved

The next steps are meant to distinguish the references that we have already found in the database from the ones that we have not yet found. Import the references that have been detected in Scopus or Web of Science, and deduplicate these with the remaining articles for which references should be checked. After deduplication, the references that have been solved will be placed in a special group.

1. Download the references of the articles that were retrieved:
   In Scopus: Tick the box before 'All' in the top left corner of the results. Open the Drop down menu for Export, and select 'RIS format.' Choose 'Citation information only' from the drop down menu and click export. Open the file. The contents will be imported in EndNote.
   In Web of Science: In the drop down menu at the top of the results, select Save to EndNote Desktop. In the next screen Type in the text boxes after Records 1 to the total number of hits (if that is more than 500, limit this to 500, and come back later for 501 and further). Open the file. The contents will be imported in EndNote.
2. Go to: All References, and click on one random reference.
3. In the menu: select References > Find Duplicates
4. In the next window, the 'Find duplicates' selection window, click on [Cancel].
5. Press delete to remove all highlighted duplicates.
6. Go to the group Duplicate References and press Ctrl-A. Drag these references into the group 'resolved'.
7. Go to the Imported References, select one random reference and press Ctrl-A
8. Go to 'Unfiled' and click on the heading Title, to sort by Title.
9. Scan the list for Blue marked references, and hold Ctrl while clicking on similar references directly above or below the marked reference. If a marked reference does not have a similar reference directly above or below it, unmark it by Ctrl-Click.
10. After all duplicates have been selected, drag the marked references to resolved.

11. Go to resolved, select one random reference, Press Ctrl-A.
12. Go to Imported References, click on the scroll bar (if no scroll bar is shown, Ctrl-Click on one of the references, and Ctrl-Click again) and press Delete.
13. If references remain in the Imported References, these are unmatched references that were found in Web of Science, or Scopus, but were not among the original relevant references. If that is the case, drag the remaining references from imported to the group unmatched.

Step 5: Repeat the steps 3 and 4 for other searches

Now step 3 and 4 are repeated for other searches. The group 'Unfiled' contains references that are not yet dragged to the group 'resolved'. Follow the steps above with the following directions. To make sure the more sensitive search for page numbers and author names does not identify false references, we need to sort the library in a way so that we can select only those references with more than 2 authors and page number higher than 50.

1. Select the style:
   In Scopus: _scopus pages
   In Scopus: _wos pages
2. Go to Unfiled
3. In the menu, go to Tools > Sort Library.
4. Select Sort First by this: # of Authors, and Then by this: Pages and click [OK]
5. Scroll down and click on the first reference with 2 authors and a page number higher than 50. Now press Shift-End to select all references from there until the bottom of the list.
6. Click on the Column heading Pages. Hold Ctrl while clicking on the first reference without a page number. Next, hold Ctrl-Shift while clicking the first reference with a page number higher than 50.
7. Press Ctrl-K
8. In Scopus: go to Advanced, Click on Enter query string, and press Ctrl-V to paste your search trategy and click Search
   In Web of Science: Select Advanced Search from the drop down menu. Paste the search strategy in the search field. Remove the last occurrence of the Boolean operator 'OR' and click Search.
9. Follow Step 4 as described above
10. Select the style:
    In Scopus: _scopus title
    In Scopus: _wos title
11. Go to Unfiled, select one random reference and press Ctrl-A.
12. Press Ctrl-K
13. In Scopus: go to Advanced, Click on Enter query string, and press Ctrl-V to paste your search strategy and click Search
    In Web of Science: Select Advanced Search from the drop down menu. Paste the search strategy in the search field. Remove the last occurrence of the Boolean operator 'OR' and click Search.

14. Follow Step 4 as described above

If this error appears, The search contains an *incorrect fieldname*, this is probably because some of the titles contain *parentheses or equal* to signs. Solve that with the following steps:

   a. Go to Edit > Find and Replace
   b. Select In the field: Title, with Find type (, deselect Match Words, and click [Change].
   c. Repeat step a and b for ‘)’ and ‘=’

Not all references might have been found with these methods. If references remain in Unfiled, they have not been found in the databases, which means that the references lists of these articles have to be scanned manually.

Step 6: Remove false results

Some references might have been accidentally retrieved by the searches. They can be found in the group unmatched. If that is the case, these articles should be removed from the search results before the reference lists are exported.

1. Go to the group unmatched and click on the column heading author
2. Press Ctrl-A followed by Ctrl-K
3. In Scopus go to Advanced Search and paste and execute the search strategy

Step 7: Download the cited and citing references

Now that most relevant articles have been found in the databases, the references can be downloaded into the original EndNote file containing the search results of the original searches.

In Scopus

1. Go to advanced search and look at the search history.
2. Combine the searches from the history that contain references, usually these are the only searches in the history: #1 OR #2 OR #3
3. If unmatched references were found, subtract that search strategy from the results: #1 OR #2 OR #3 AND NOT #4
4. Click on the arrow right next to the top check box, and tick the check box before *check all*
5. Click More > *View Reference*s to view the Cited references
6. If the number of references is higher than 2,000, use filters in the column Refine Results to create two or more separate result sets that contain less than 2,000 hits. For instance select a set of publication years. First click on limit to and execute the steps below, and later select the same years and select Exclude.
7. In that overview again click on the arrow right next to the top check box, and tick the check box before *check all*
8. Click Export, Select *RIS Format* and select *Citations and abstract information*
9. Open the resulting file Scopus.ris to import the references in Endnote (import them in the file where all included and excluded are)
10. Go back to Scopus and click on *View Cited By* to see an *overview of citing references*.

11. Repeat steps 6: 5-7 to import citing references.
12. Now deduplicate the file to remove those articles that have already been reviewed from the new references

In Web-of-Science

1. If the search results are satisfying, click on the number of hits in the search history in the table at the bottom of the page
2. Now click on the first result
3. On the left side of the screen two links are visible: *X times cited* and *Y Cited References*. Clicking on the first link results in the citing references (thus newer related articles). Clicking on the second provides the searcher the complete reference list of this article Therefore finding older related citations. Both lists have to be exported separately for each article.
4. If this is done for the first record, continue with the second etc.

6.3 Discussion

Advantages of the method are that screening of the references is easier as the downloaded citations also contain the abstracts of the articles, whereas when the references are only reviewed in the reference lists of the citing articles, only bibliographic information is shown, and the reviewer has to decide on the title whether a citation might be relevant. Using the method described in another article by Bramer, et al., the screening process can be performed as is normally done for references.

Another advantage of the method is that unnecessary screening of duplicate references is avoided. References that are cited by more than one of the core articles are downloaded only once from the databases, and additionally the complete list of references is deduplicated with the references that had already been retrieved by the database searches.

Disadvantages are that not every article is indexed in Scopus, and that for some indexed articles the references lists cannot be downloaded. This means that for some core articles, still the traditional method of reference checking should be followed. However, this is not a true disadvantage of the method, as in the traditional methods, this should be done as well, but for all references. Also we noted that some of the references found in Scopus do not have all data available for download. Sometimes titles are missing. This will mostly occur with rather old and general, non-journal article references, so probably this will not result in missed relevant references.

6.4 References

1. Chandler J, Churchill R, Higgins J, Lasserson T, Tovey D. Methodological standards for the conduct of new Cochrane Intervention Reviews: The Cochrane Library; 2013 [updated 2 December 2013; cited 2017 20 January]. Available from: http://editorial-unit.cochrane.org/sites/editorial-unit.cochrane.org/files/public/uploads/MECIR_conduct_standards%20 2.3%2002122013_0.pdf.
2. Higgins JPT, Green S. Cochrane handbook for systematic reviews of interventions: Wiley Online Library; 2008.
3. University of York, N.H.S. Centre for Reviews and Dissemination. Systematic reviews : CRD's guidance for undertaking reviews in health care. York: CRD, University of York; 2009.
4. Chapman AL, Morgan LC, Gartlehner G. Semi-automating the manual literature search for systematic reviews increases efficiency. Health information and libraries journal. 2010;27(1):22-7.
5. Horsley T, Dingwall O, Sampson M. Checking reference lists to find additional studies for systematic reviews. Cochrane Database of Systematic Reviews. 2011(8):MR000026.
6. Elsevier Publishers. Scopus. Available from: www.scopus.com.
7. Clarivate Analytics. Web of Science. Available from: https://webofknowledge.com.
8. Bramer WM, Giustini D, de Jonge GB, Holland L, Bekhuis T. De-duplication of database search results for systematic reviews in EndNote. J Med Libr Assoc. 2016 Jul;104(3):240-3

Chapter 7

Evaluation of a new method

Bramer WM, Rethlefsen ML, Mast F, Kleijnen J. Evaluation of a new method for librarian-mediated literature searches for systematic reviews. Res Synth Methods. 2018;9(4): 510-20.

7 Evaluation of a new method for librarian-mediated literature searches for systematic reviews

Abstract

Objective

To evaluate and validate the time of completion and results of a new method of searching for systematic reviews, the exhaustive search method (ESM), using a pragmatic comparison.

Methods

Single-line search strategies were prepared in a text document. Term completeness was ensured with a novel optimization technique. Macros in MS Word converted the syntaxes between databases and interfaces almost automatically. We compared search characteristics, such as number of search terms and databases, and outcomes, such as number of included and retrieved references and precision, from ESM searches and other Dutch academic hospitals identified by searching PubMed for systematic reviews published between 2014 and 2016. We compared time to perform the exhaustive search method with a secondary comparator of recorded search times from published literature and contact with authors to acquire unpublished data.

Results

We identified 73 published Erasmus MC systematic reviews and 258 published by other Dutch academic hospitals meeting our criteria. We pooled search time data from 204 other systematic reviews. The ESM searches differed by using two times more databases, retrieving 44% more references, including 20% more studies in the final systematic review, but the time needed for the search was 8% of that of the control group. Similarities between methods include precision and the number of search terms.

Conclusions

The evaluated similarities and differences suggest that the exhaustive search method is a highly efficient way to locate more references meeting the specified selection criteria in systematic reviews than traditional search methods. Further prospective research is required.

7.1 Introduction

Systematic reviews are often challenged by retrieval bias, referring to failure in locating pertinent studies to include. To combat retrieval bias, comprehensive search methods are used to locate as many reports of studies as possible. Developing a comprehensive search for a systematic review can require large efforts.(1-3) Estimates suggest that database searching may take an average of 17.7 hours.(4) For each database used, it has been estimated that it may take an expert searcher two hours to translate a search adequately.(5)

There are well-established suggestions for how to plan systematic review searches, particularly the recently produced MECIR standards (Methodological standards for the conduct of new Cochrane Intervention Reviews) and the National Academy of Medicine (formerly Institute of Medicine) standards.(1, 6) However, most guidelines do not specify how to construct a comprehensive search strategy within a database. In aggregate, librarian and information specialist authored search strategies are of higher quality. However, in practice, there is still variance between librarians and information specialists in comprehensiveness and overall search quality.(7)

Biomedical information specialists of Erasmus Medical Center of the Erasmus University in Rotterdam, the Netherlands (Erasmus MC) have created a blended approach to search strategy creation that combines objective methods of search term identification with librarian or information specialist expertise, here referred to as the "exhaustive search method," or ESM.(8) This method relies on creating single-line search strategies in a text document, identifying relevant controlled vocabulary and free-terms terms using database thesauri, an optimization technique to identify possibly relevant search terms, and macros to convert searches between different database syntaxes automatically.

The best method of comparing two or more search strategies is a prospective design, whereby alternative approaches are performed on a topic, the results pooled, and recall and precision calculated after the final included studies were chosen. Indeed, this study is underway. However, to provide initial validation of the exhaustive search method as a substitute for the traditional method, we sought a more pragmatic comparison to librarian-mediated searches on a larger scale. The purpose of this study is to provide baseline evaluation and validation of the ESM search characteristics, precision, recall, retrieval, and speed.

7.2 Methods

Since early 2013, librarian-mediated searches by Erasmus MC information specialists were designed using ESM. Embase is used as the primary search design interface, via embase.com. For each important element in the research question, thesaurus terms and synonyms for title or abstract searching are collected from the Emtree thesaurus. These terms are then combined into a single-line search strategy in a Microsoft Word document using predefined syntax formats. To optimize sensitivity of the search, this preliminary search strategy is tested for completeness using an optimization method. The method examines articles indexed with identified Emtree terms, but where the titles and abstracts lack the synonyms already used in the search strategy. Relevant terms from titles and abstracts are added to the search strategy, and their added value is evaluated in concert with the requesting researcher. Further optimization is done by reversing this process: looking for new thesaurus terms in articles where the titles and/or abstracts contain one of the identified synonyms, but lack the already identified thesaurus terms.

Macros in MS Word translate the search strategies between platforms by replacing the syntax of one platform with the corresponding syntax of another platform. The standard procedure involves translation to MEDLINE via the Ovid interface, the Cochrane Library via Wiley Interscience, Web of Science Core Collection via ISI Web of Knowledge, and Google Scholar. Optional translation macros are available for Scopus and for CINAHL in the EBSCOhost interface. Using the macros we make sure there are no changes in the search terms for title and/or abstract. They remain constant between databases, only the thesaurus terms are changed between databases.

To understand whether searches performed using the ESM differ from traditionally-constructed searches, we sought to compare several attributes of complexity and time for librarian-mediated searches conducted for systematic reviews. We compared systematic reviews published by Erasmus MC that were created using the exhaustive search method, with systematic reviews published by other Dutch academic hospitals (DAH) that were assisted by medical librarians. We selected DAH as a comparison for pragmatic reasons. First, biomedical information specialists in the Netherlands, in particular those in academic settings, regularly share their expertise in biannual meetings. The general level of search expertise among Dutch academic medical information specialists is therefore considered high and rather homogeneous. Furthermore, due to the small size of the group, most individuals are personally known to the first author, making it easier to recognize their involvement, even if they were mentioned only by name and not by function. Other studies examining librarian or information specialist involvement in systematic reviews have had difficulty identifying these contributors due to ambiguous names or reporting conventions.(7, 9)

Retrieval of included and control reviews

To identify published systematic reviews for both the study and comparison groups, all systematic reviews published between 2014 and 2016 by Dutch academic hospitals, including Erasmus MC, were sought using the search strategy below in PubMed. The search was last updated July 31, 2016.

(((Medisch Centrum[ad] OR Medical Hospital[ad] OR Medical Center[ad] OR Medical Centre[ad] OR umc[ad] OR mc[ad]) AND (leiden*[ad] OR maastricht[ad] OR utrecht[ad] OR groningen[ad] OR radboud[ad] OR nijmegen[ad] OR vu[ad] OR vrije universiteit[ad] OR free university[ad] OR amsterdam[ad])) OR AMC[ad] OR VUmc[ad] OR Radboudumc[ad] OR erasmus[ad] OR umcu[ad] OR umcg[ad] OR umcl[ad] OR lumc[ad]) AND (systematic review[tiab] OR systematically review*[tiab] OR systematic literature review*[tiab] OR systematic literature search*[tiab] OR systematic search*[tiab] OR systematically search*[tiab] OR medline[tiab] OR pubmed[tiab] OR embase[tiab] OR prisma[tiab] OR google scholar[tiab]) AND 2014:2016[dp]

After completing the search, we checked if each article that was retrieved reported the results of a systematic review. Protocols for the development of systematic reviews were excluded. We considered an article a systematic review if it met one or more of the following criteria:

- The title or abstract describes the study as a "systematic (literature) review"
- The methods section describes that a systematic literature search was performed
- The article was published in a source that primarily publishes systematic reviews (e.g. Cochrane Database Syst Rev)
- The article is described in title or abstract as a review, and the process of article selection is presented in a PRISMA flow chart

If the first or corresponding author's affiliation was Erasmus MC, we verified that the search strategy used for this review was librarian-mediated by comparing Erasmus MC authors with our customer registration system. We excluded reviews where the original search was designed before January 1, 2013, as our method was developed in 2012. Articles meeting these criteria constitute the ESM group. From these identified systematic reviews, we collected the final number of references meeting their inclusion criteria and information on the topic of the review. Precision of the searches was calculated dividing the published number of final included references with the number of retrieved, deduplicated references from our search registration. Systematic reviews not reporting any relevant, retrieved references were excluded from precision calculations.

To be included in the comparison set, the first author had to be affiliated with a Dutch academic hospital other than Erasmus MC. Secondly, we included only articles where the assistance of a professional information specialist from the institution's medical library was mentioned in the full text. For the systematic reviews meeting these criteria, the number of databases and the names of the databases that were searched, the number of references found in MEDLINE, the total number of retrieved references after deduplication, the number of articles reviewed in full text, and the number of final included references were documented from the full text. We determined the number of search terms from the online appendices or full text. For that we copied the complete MEDLINE Ovid or PubMed search strategy into an empty Microsoft Word document. For PubMed searches that had used field codes for each synonym, we simply documented the number of instances of an opening square bracket ('['), which is a mark of the field codes in PubMed. For MEDLINE Ovid and for PubMed searches where field codes had not been used for all search terms, we used the Find function for each Boolean and proximity operator, summing the number of instances shown and adding the highest line number of the multi-line search strategy to that total. We calculated the precision of the search as above using published data. Because we did not restrict the comparison set to articles where all attributes were fully reported, we compared only those articles where data was available for each attribute individually.

Gathering the data

Starting in November 2013, for each novel ESM study, we registered the actual search time of the Erasmus MC information specialist involved, starting at the beginning of the reference interview and ending when the searches for all databases were finalized. Our registrations did not include the time needed to import results in reference management software and to deduplicate them. We also did not count time needed for handsearching, reference checking, contacting key authors or searching grey literature, as these tasks, in our institute, are generally performed by the review authors. When updates of the literature were needed, we only added extra time if changes had to be made to the search strategy, otherwise the tasks were merely rerunning the searches and importing in reference management software, tasks which are not included in the time registration. Some ESM studies partially relied on reusing search elements from other searches or published filters; using time data from these studies may have inordinately lowered the average needed time for searches. These studies were not included in the ESM time studies group, as we felt that in these cases, the time recorded would not be a good representation of the actual time needed to create such a complex search strategy.

Because data on the time needed to create the search strategies in DAH systematic reviews was unavailable, as a secondary comparison, we collated data from several published studies describing the time needed to create searches for systematic reviews.(4, 10-14) We contacted the authors for detailed information about individual systematic review projects when it was not clear from the published papers. These individual data per review were then pooled in an MS Excel file, where we calculated quartiles and median values.

We analyzed the nature of the reviews' topics using the tree structure of the MeSH databases. We searched in the MeSH database for the appropriate MeSH term for the disease described in the article and selected the corresponding high level MeSH term directly below the Diseases Category in the MeSH thesaurus. For the intervention element we chose the appropriate top level MeSH term directly below the Analytical, Diagnostic and Therapeutic Techniques and Equipment Category. If the most appropriate MeSH term for the intervention element fell within the top level MeSH term Therapeutics, we documented a MeSH term one level deeper. We also documented the domain of the review. We identified seven domains. Reviews on the effectiveness of a treatment were assigned the domain therapy; those on policies fell in the group management. Research on incidence and prevalence were documented as epidemiology, and reviews on causes of diseases as etiology. Other domains we documented were prognosis, diagnosis and prevention.

Significance of differences between numerical observations (such as the number of search terms or the number of included references) in the study data and the comparison were calculated using a two sided Mann-Whitney test. Binary data (such as research topics and database use) were compared using a Chi-square test in SPSS.(15) Differences with p-values smaller than 0.05 were considered significant.

We further validated our comparison data (DAH) by briefly comparing the outcomes of the DAH and ESM data with another dataset, which was obtained from research by Borah et al.(16) They gathered data from finished and published reviews registered in PROSPERO. We compared their data, using only the 38 reviews where we observed the acknowledgement of a medical librarian or information specialist, or when one of the co-authors' affiliation is a library, to the DAH and ESM data using range and medians.

7.3 Results

On July 31st 2016, 2422 articles were reviewed. In 206 cases, the articles were systematic reviews with an Erasmus MC employee as first author. In 141 of these reviews, we identified that the search strategy was created by an Erasmus MC medical librarian. In 92 (65%) of these reviews, the assistance of a librarian was reported in the article. In 55 articles the first author of this article (WMB) was acknowledged, 22 were co-authored by WMB, and in 20 articles, he was not mentioned by name, only by function. Sixty-eight articles of the 141 either could not be traced in our registration, or the initial search was created before January 1st, 2013. This left 73 articles in the ESM group for which all data were available. We found a total of 1271 systematic reviews where the affiliation of the first author was a Dutch academic hospital other than Erasmus MC. In 258 reviews (21%), the assistance of a medical librarian was reported. See Figure 7.1 for a flowchart of the inclusion / exclusion process. See the Supplemental data for a complete overview of all included publications and registered attributes.

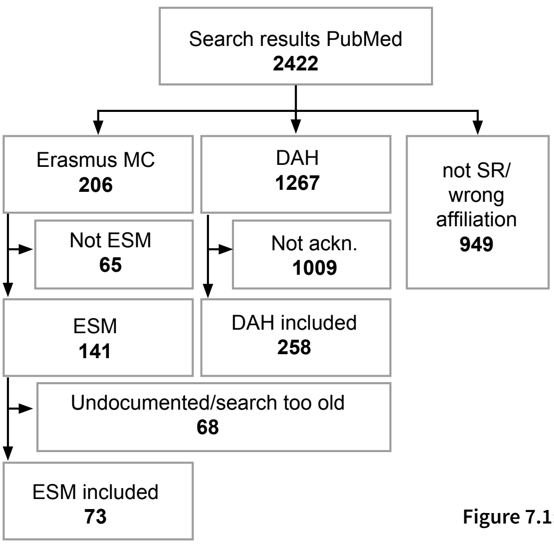


Figure 7.1: Flowchart of the inclusion exclusion process

The majority of the reviews in the DAH group were written by researchers from Academic Medical Center (AMC) Amsterdam (93). Others came from Leiden University Medical Centre (53), VU Medical Centre Amsterdam (46), and Utrecht University Medical Centre (35). The DAH librarians acknowledged or co-authoring most frequently in the DAH reviews were a librarian from Leiden (20 acknowledgements /13 co-authorships), one from Utrecht (23 acknowledgements), and an AMC information specialist (7 acknowledgements / 11 co-authorships). Of the 258 included reviews in the DAH group, 178 (68%) reported all attributes we collected. The most completely reported attribute was the number of databases searched (n=256), while the lowest was the number of search terms used (n=189). For each of the 73 reviews included in the ESM group we documented all attributes, except the topics of the research question.

Topics of the research questions

There were some significant differences regarding the research questions in the two groups (see Table 7.1). No significant differences were observed in the Patient/Population or Intervention elements of the reviews. The domain prognosis was overrepresented in the DAH group (p=.007), and epidemiology (p=.001) and etiology (p=.001) in the ESM group.

	DAH Comparison	ESM Group
Patient	**(N=215)**	**(N=63)**
Neoplasms	19%	13%
Cardiovascular diseases	13%	14%
Wounds and injuries	7%	11%
Urogenital diseases	6%	11%
Mental disorders	7%	6%
Nutritional diseases	4%	10%
Musculoskeletal diseases	9%	3%
Otorhinolaryngology	7%	3%
Signs and Symptoms	0%	6%
Skin Diseases	5%	3%
Intervention	**(N=128)**	**(N=49)**
Surgical procedures	34%	31%
Chemicals and drugs	29%	35%
Diagnostic imaging	13%	6%
Physical therapy	0%	6%
Domain	**(N=241)**	**(N=68)**
Therapy	41%	35%
Prognosis *	24%	10%
Diagnosis	15%	9%
Management	12%	15%
Epidemiology *	4%	15%
Etiology *	2%	12%

Table 7.1: Analysis of research questions between comparison (DAH) and study (ESM) group (only items where frequencies >5% are shown). * Significant difference ($p < .05$)

Databases Searched

The number of databases searched was reported in 256 DAH systematic reviews (see Figure 7.2). ESM systematic reviews searched more databases comparatively (ESM median=7; DAH median=3). There was a large variation between individual reviews (Range: ESM=3-13; DAH=1-20). The difference is highly significant in favor of ESM (Mann-Whitney U=1959.5, p<.001).

The databases that were used significantly more frequently in the ESM reviews were Cochrane CENTRAL, Web of Science, Scopus, and in particular Google Scholar, which was used in 3% of the DAH reviews, but in 76% of the ESM reviews (for each of these four databases p<.001). See Table 7.2 for an overview of frequencies of database use.

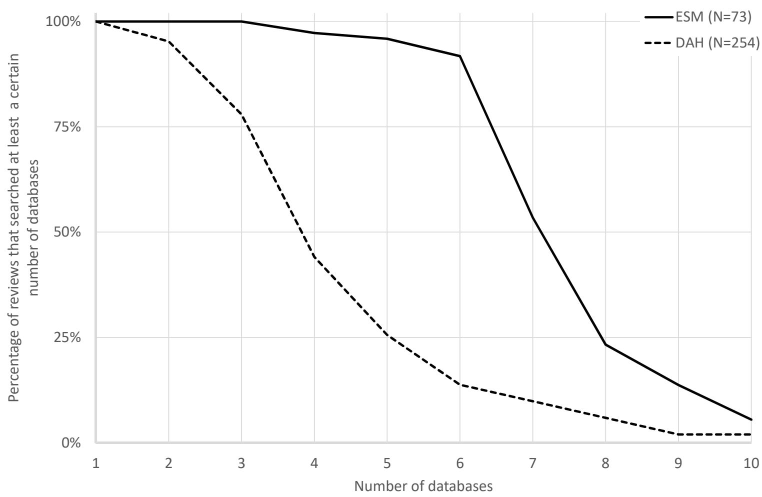


Figure 7.2: Number of databases searched

Table 7.2: Frequency of use of different databases. * Significant difference (p < .05)

	DAH (N=254)	ESM (N=73)
MEDLINE / PubMed	**100%**	**100%**
Embase	93%	100%
Cochrane CENTRAL *	56%	96%
CINAHL	29%	23%
Web of Science *	22%	89%
PsycINFO	20%	11%
Scopus *	3%	30%
Google Scholar *	3%	76%

Search Complexity

For 193 (73%) DAH reviews, we could determine the number of search terms in the MEDLINE search strategy (see Figure 7.3). The DAH search strategies are more frequently very complex; for instance, 21% of DAH searches are more than 100 terms long, versus 7% in ESM searches. The medians were nearly equal, however (ESM=50; DAH=49). There is no significant difference (Mann-Whitney U=6517.5, p=.488).

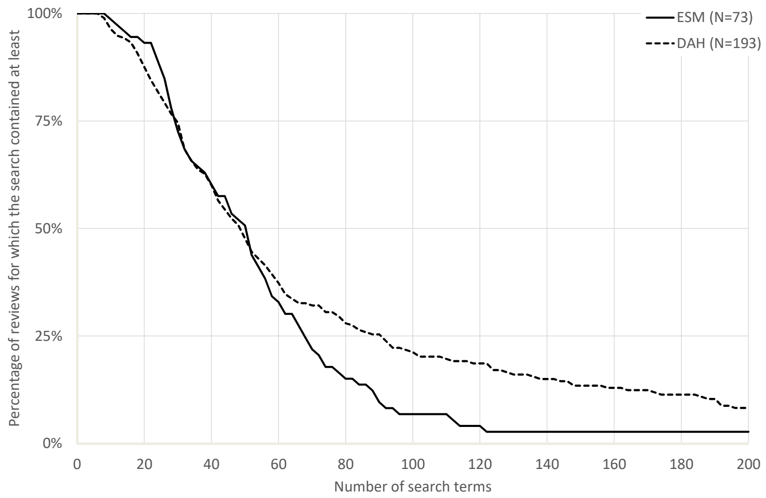


Figure 7.3: Number of search terms in the Medline search

Number of References Retrieved

The number of references after deduplication for all databases included in each review was reported in 250 (95%) DAH systematic reviews. ESM searches retrieved more references after deduplication than DAH searches (see Figure 7.4). Though the average is similar (ESM=2581; DAH=2485), the median number of hits for ESM searches differed substantially (ESM=2188; DAH=1515). There was a great deal of variation between individual searches (Range: ESM=285-9472; DAH =74-17956). The difference is significant (Mann-Whitney U=7567.5, p=.041).

For 125 reviews in the DAH group, we could find the number retrieved for Medline. The median was lower for the ESM data (ESM=875; DAH=1074), but the difference was not significant (Mann-Whitney U=4150.5, p=.363).

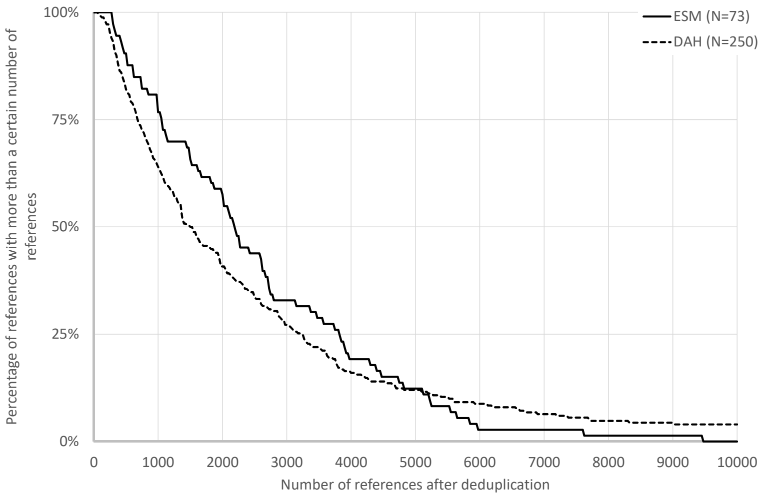


Figure 7.4: Total number of references retrieved after deduplication

Number of relevant references

In 243 reviews from the DAH group and in 62 of the ESM reviews, the number of articles screened in full text was mentioned. The median number of articles read in full text was higher in the ESM group (106) than in the DAH reviews (64). The difference is significant (Mann-Whitney U=8926.5, p=.024).

The final number of references meeting inclusion criteria was reported by 251 (98%) DAH systematic reviews (see Figure 7.5). Though there is a wide range (ESM: 4-190; DAH: 0-1342), the median was higher for ESM reviews (ESM=25; DAH=21). The difference is significant (Mann-Whitney U=10951.5, p=.004).

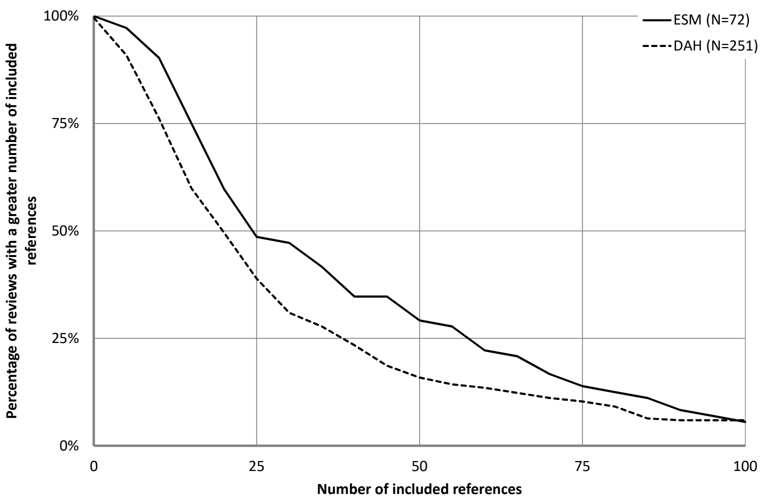


Figure 7.5: Number of included references

Precision

Precision was calculable in 241 DAH systematic reviews (93%) where both the number of included references and the number of references after deduplication were reported, not including one review in the DAH group that reported they had retrieved no relevant references. The median precision for ESM was 1.8% compared to 1.4% for DAH (see Figure 7.6). There is no significant difference (Mann-Whitney U=8354.5, p=.515).

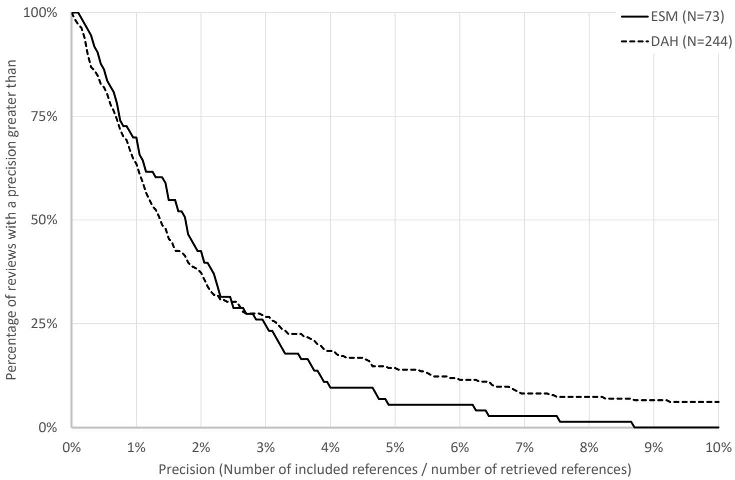


Figure 7.6: Precision (number of included references/number of retrieved references)

Time Needed to Create the Search Strategy

In 24 of the 72 included reviews in the ESM group, the initial searches were developed before November 2013, meaning before we began recording time needed to create the search strategies. In 11 cases where the search was developed after November 2013, we had reused parts of a search developed for another review, therefore we had not been able to fully document the time needed to create the search strategy. For the remaining 37 (51%) ESM systematic reviews included in this research, the time needed to create the search strategies was registered at the time of search development. For the comparison data pooled from published time studies (PTS), we identified 105 published or unpublished systematic review projects.(4, 10-14) These were combined with results from an online questionnaire in which information specialists were asked about the time they spent creating and translating searches for their last systematic review (N=99).(17)

The results are compared in Figure 7.7. The median time needed in the pooled time studies was 12.8 hours, while the median time needed for ESM searches was 60 minutes. After 2 hours the searches for 95% of all ESM reviews had been finalized compared to 2% of the PTS reviews. The difference is statistically significant in favor of ESM (Mann-Whitney U=466, p<.001).

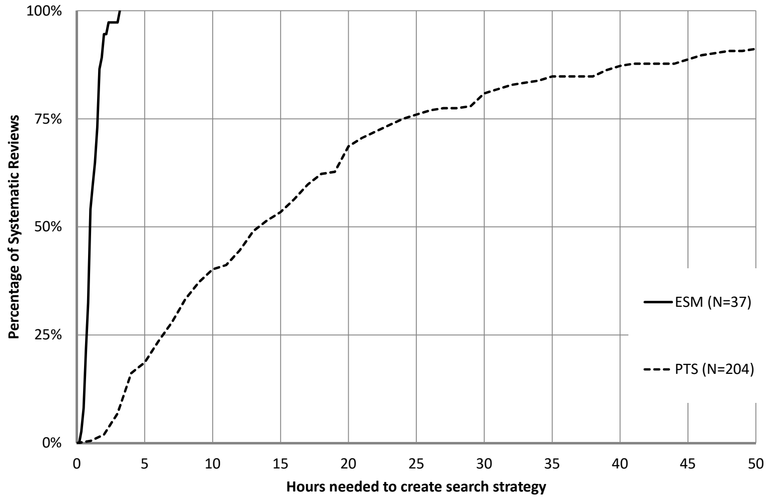


Figure 7.7: Time needed to create search strategies in hours

7.4 Discussion

Our baseline evaluation and validation of the exhaustive search method (8) against pragmatic comparators of traditional librarian-mediated searches reveals several differences and several similarities between outcomes. The ESM searches differed by using more databases, favoring the use of different databases (Web of Science, Scopus, Google Scholar), retrieving more references, including more studies in the final systematic review, and taking significantly less time to execute. Similarities between methods include similar precision and search complexity, as measured by number of search terms. These similarities and differences suggest that the exhaustive search method is a highly efficient way to locate more references meeting inclusion criteria then traditional search methods. Though ESM searches returned more results, which may require additional screening time, the searches' precision was equivalent to the DAH searches, meaning the number needed to read to find studies for inclusion remains similar.

This initial evaluation validates our hypothesis that the outcomes of the ESM search method do not negatively compare to traditional search methods. In fact, using the exhaustive search method is a faster, more efficient way of searching more databases. The precision found using both methods was statistically no different, though both groups demonstrated a lower median precision (ESM, 1.8%; DAH, 1.4%) than previously published research, which found a median precision of 2.9%.(18) Since ESM did retrieve an overall higher median number of references, though the number of references in MEDLINE was lower than in ESM reviews than in the DAH group, it appears that ESM enables searchers to cast a wider net (searching more databases, e.g.) and find more relevant references that traditional searching methods are not able to locate.

Limitations

Our validation study has several limitations. As noted above, we chose to utilize pragmatic, real-world comparisons for our initial analysis rather than begin with a large-scale prospective study. Our choice of using published systematic reviews from Dutch Academic Hospitals with librarian or information specialist-mediated searches may be considered a weakness of our study design, as it is retrospective and not a random sample or a gold standard approach. However, several factors make this a reasonable choice and comparison. First, we included over three times as many articles in our DAH group as our ESM group, representing multiple hospitals and, more importantly, multiple searchers. Having multiple searchers, each of whom may have a slightly different approach to developing a search strategy, enables us to minimize the bias that may come from comparing single searcher to single searcher. Pragmatically, using Dutch academic hospitals also enabled us to easily determine whether a systematic review included a librarian or information specialist through name recognition.

To test the validity of our comparison data, we briefly compared the outcomes of the DAH and ESM data with another dataset, which was obtained from research by Borah et al.(16) The number of included references in that set was lower than in both the DAH and ESM reviews (median 15, average 22). The number of databases searched lay between the DAH and ESM data (median: 5) as was the number of search terms (median: 52, average: 63). The number of articles retrieved was higher than both ESM and DAH data (median: 2615, average: 4981) and therefore the precision was lower (median 0.5%, average 1.5%). Using our extra validation step, we ascertained that using the DAH group is reasonable; benchmark data from Borah et al did not show dramatic differences.

The DAH group's topical coverage did differ in certain aspects from coverage in the ESM group, which is a limitation of our study design. It was, however, impractical to compare all ESM reviews with matched topic reviews. Systematic reviews are generally executed on topics for which no other review exists. Matching topics in a broader sense would not increase the validity of the control group. We investigated whether differences in topics would account for the differences in number of included references. When comparing the number of included references within certain domains (such as therapy or epidemiology), we found that the median number of included references did not differ between domains.

Another factor potentially influencing the number of included references could be whether a review limited included references to study type. We found no significant difference in the percentage of reviews that limited to randomized controlled trials (6.6% for DAH, 5.6% for ESM). In the ESM reviews, we found that reviews that included only RCTs included fewer articles overall (15 compared to 29 in reviews that did not limit to RCTs); however in the DAH reviews, the median number of included references when limited to RCTs was higher (23 compared to 20). We therefore reject the assumption that the topics of the reviews or study type limits were the main cause for the difference in number of included references.

A gold standard approach, as used to validate the Institute for Quality and Efficiency in Health Care (IQWiG) “objective approach”,(19) can offer a stronger comparison, as search topics are compared head to head, but due to the complexity of creating gold standard sets from published systematic reviews, such studies are often limited in scope. The IQWiG analysis using this approach, for example, used a sample of 13 Cochrane reviews from which gold standards were created. For this study, they also assumed comprehensiveness of the search using a proxy of “a search in at least two bibliographic databases and one additional source (e.g., a study registry) and documentation of the search strategy were required”.(19) Because a gold standard set is inherently limited by the comprehensiveness of a search, and we do not consider a search in three resources to be necessarily comprehensive, our study instead attempted to look at broader trends, similarities, and differences between the results of our method and the traditional method.

A major difference between the DAH and ESM reviews is that the searches for all ESM reviews were performed by only one person, while the DAH reviews mentioned 29 searchers by name. Nine of these information specialists were involved in 10 or more of the DAH reviews, and 15 had contributed to 5 or more reviews. To account for the difference in experience, we checked whether the number of included references differed between reviews when searchers were involved in 5 or more reviews, compared to reviews where searchers were involved in 4 or fewer reviews in the DAH dataset. We did find a small difference in number of included references, but it was not significant (22 for experienced searchers; 19.5 for less experienced searchers, Mann-Whitney U=1892, p=.263). The median number of terms used by experienced information specialists was significantly higher than that of less experienced searchers (56 vs. 41, Mann-Whitney U=807.5, p=.003).

A further limitation is our use of a second pragmatic comparator to assess time spent on creating the search strategies for systematic reviews. Because we believe that speed and efficiency are the major benefit of the exhaustive search method, it was a critical piece of our evaluation. Though we knew anecdotally that searchers spent far longer creating searches using traditional methods than we did using ESM, we were not able to ascertain the time spent creating the DAH searches retrospectively. Instead, we were required to establish a secondary comparator from published evidence and unpublished data from those studies, where necessary, to create a pool of 204 studies for which we had time data. Though this did not match our DAH data by design, the massive difference between time taken for ESM searches and the PTS searches is striking. The time difference would likely not be as striking for a novice using ESM, as indeed even ESM requires practice and expertise. Measuring time to complete a search is complicated by a myriad of factors. For instance, it is not always clear what is included as searching time in reported studies. Though we did not account for time spent searching for grey literature, deduplicating citations, handsearching, citation tracking, explaining the process of search strategy design to the researchers involved, or other tasks in our documentation for ESM searches, it is unclear whether some datasets of the comparison included these actions in their time logs. Saleh et al's study of time needed to search grey literature (4) did find that up to half of searching time could be spent searching grey literature. The data derived from a recent questionnaire among information specialists showed that on average 80% of the time for a systematic reviews was spent searching bibliographic databases (17). We altered the PTS data for studies where it was not clearly defined what was measured in the recorded search time accordingly, or where this included more tasks than we registered: for those values we used 80% of the reported time.

Though we found ESM does work faster for the information specialist, because it finds more references, the extra retrieved references mean that the researchers have to spend more time screening. However, we believe that the extra time spent screening would improve the quality of the review, as more relevant references would be included. Researchers should not restrict the number of references to reduce the time to finish the review. However, information specialists have limited time and often large numbers of clients. Many institutions only employ one or two expert searchers. If the time needed per review for the searchers is reduced, this means they are able to provide more reviews with high quality search strategies. If the expert searcher would not have had time to create a search strategy for the review project, the review authors likely would have done the searches themselves, which would lead to the loss of relevant references. Therefore, the speed of ESM has potential to increase the overall quality of systematic reviews.

When comparing the number of final included references we chose not to distinguish between references retrieved by database searches and by other methods (such as citation tracking, handsearching and contacting key authors), in large part because most reviews do not specify where each reference was found. Some of the reviews in both groups have performed these extra activities, others have not. Sometimes these differences were caused by an update of the search strategies by the authors. For example in the review by Pieterman et al.(20) 17 references were not found in our EndNote library, as they were of a later publication date than the last search date in our system. When correcting the data in the ESM results for the references not retrieved via our database searches, the number of included references showed only minor changes and the difference was still statistically significant. However, since we cannot similarly correct the DAH data, because we cannot easily distinguish where the DAH references were found, we kept all ESM final included references found outside database searches in our observations for parity.

We can only speculate why the exhaustive search method is so much faster than other methods, but still able to provide good quality searches. Using single-line search strategies allows for faster development and execution of search strategies. Searches can be easily adapted by adding extra terms without the need to regroup line numbers. Additionally, repeating a search strategy is much easier than having to meticulously type in all search statements in the correct order. (21) The single-line search strategies allow for optimization of the initial search strategy, which identifies extra relevant terms to be added to the search strategy that increase the sensitivity of the final search. The macros speed up the process of translating a search strategy across platforms and databases. Therefore, search strategies do not have to be built completely anew for each extra database.

The question remains whether ESM can be applied by other information specialists and other institutes. Our experience in teaching this method in workshops, even for experienced searchers, showed us that there might be a rather steep learning curve. One major problem that arises from single-line search strategies are inconsistent use of parentheses. Mistakes are more easily made because it can be hard to see at a glance whether the number of opening and closing parentheses match. This problem was nearly absent from our search development process because of our habit to immediately type a closing parenthesis when an opening parenthesis is typed, as well as the preparation of proximity operators, including all parentheses, before words are added. Additionally, other institutes are likely to have different subscriptions to databases and interfaces. We learnt that the process is most efficient when using the interface of embase.com compared to MEDLINE or Embase via Ovid. If other databases and interfaces are used, the macros we developed cannot be used in their current forms, and they have to be adapted to the other institute's needs.

Although the exhaustive search method allows for quick search development, the method is not a guarantee for speed, as speed varies among users of the method. Speed is dependent on several factors, most notably experience, with this method. The ESM benchmark times related in this study should not be used to set time limits on search strategy development. Instead the method should be used to improve search strategy quality, and over time can, as experience of the searcher with this method grows, result in time reduction per search.

This evaluation study is a first step towards validating the exhaustive search method. Further studies must be undertaken to provide stronger prospective evidence of the method's outcomes and speed compared to traditional searching methods. This research is currently underway. As with the IQWiG team's validation of their objective approach to searching using a similar prospective design, (19) a prospective study will be necessarily smaller in scope than this more wide-scale evaluation, though it will offer us a more robust study design.

7.5 Conclusions

The exhaustive search method currently used at Erasmus MC creates opportunities for faster development of systematic review search strategies that find more relevant studies than other methods with equivalent search precision. Using this method, librarians and information specialists can potentially help many more people with the development of exhaustive searches for systematic reviews than traditional search methods allow without loss of complexity and search precision.

Future studies should assess recall in addition to the attributes reported here. For a thorough comparison, one should perform the searches for one systematic review topic with multiple searchers, include all results for the reviewers to prepare their review, and then check whether each searcher retrieved all included references. Research is currently being undertaken on the exhaustive search method using this prospective design

7.6 References

1. Institute of Medicine, Committee on Standards for Systematic Reviews of Comparative Effectiveness Research. Finding What Works in Health Care: Standards for Systematic Reviews. Washington, DC: The National Academies Press; 2011.
2. Higgins JPT, Green S. Cochrane handbook for systematic reviews of interventions: Wiley Online Library; 2008.
3. Chandler J, Churchill R, Higgins J, Lasserson T, Tovey D. Methodological standards for the conduct of new Cochrane Intervention Reviews. 2013 [updated 2 December 2013; cited 2015 21 December]; Version 2.3:[Available from: http://editorial-unit.cochrane.org/sites/editorial-unit.cochrane.org/files/uploads/MECIR_conduct_standards%202.3%2002122013_0.pdf.
4. Saleh AA, Ratajeski MA, Bertolet M. Grey Literature Searching for Health Sciences Systematic Reviews: A Prospective Study of Time Spent and Resources Utilized. Evidence Based Library and Information Practice. [systematic reviews; literature search; grey literature; time; health sciences]. 2014;9(3):28-50.
5. Erwin PJ. By The Clock: How Much Time Does an Expert Search Take? MLA news. 2004 October(370):1,12.
6. MECIR. Methodological standards for the conduct of Cochrane Intervention Reviews. 2013 [updated December 2013; cited 2015 May, 7]; Available from: http://editorial-unit.cochrane.org/sites/editorial-unit.cochrane.org/files/uploads/MECIR_onduct_standards%202.3%2002122013.pdf.
7. Rethlefsen ML, Farrell AM, Osterhaus Trzasko LC, Brigham TJ. Librarian co-authors correlated with higher quality reported search strategies in general internal medicine systematic reviews. J Clin Epidemiol. 2015 Jun;68(6):617-26.
8. Bramer WM, de Jonge GB, Rethlefsen ML, Mast F, Kleijnen J. A systematic approach to searching: How to perform high quality literature searches more efficiently. submitted. 2017.
9. Koffel JB. Use of Recommended Search Strategies in Systematic Reviews and the Impact of Librarian Involvement: A Cross-Sectional Survey of Recent Authors. PLoS One. 2015;10(5):e0125931.
10. Allen IE, Olkin I. Estimating time to conduct a meta-analysis from number of citations retrieved. JAMA. 1999 Aug 18;282(7):634-5.
11. Gann LB, Pratt GF. Using library search service metrics to demonstrate library value and manage workload. J Med Libr Assoc. 2013 Jul;101(3):227-9.
12. Hausner E, Guddat C, Hermanns T, Lampert U, Waffenschmidt S. Prospective comparison of search strategies for systematic reviews: an objective approach yielded higher sensitivity than a conceptual one. J Clin Epidemiol. 2016 Sep;77:118-24.
13. Nauche B, Landry T. Implanter un service de soutien aux revues systématiques en milieu hospitalier - Le chemin parcouru par les bibliothécaires du CUSM. Congrès annuel de l'ABSC/CHLA; June 18; Montréal, Canada2014.
14. Lyon JA, Garcia-Milian R, Norton HF, Tennant MR. The Use of Research Electronic Data Capture (REDCap) Software to Create a Database of Librarian-Mediated Literature Searches. Med Ref Serv Q. 2014;33(3):241-52.
15. IBM. SPSS Statistics. 21 ed1989, 2012.
16. Borah R, Brown AW, Capers PL, Kaiser KA. Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry. BMJ Open. 2017;7(2).
17. Bullers K. Unpublished data (personal communication). 2017.
18. Sampson M, Tetzlaff J, Urquhart C. Precision of healthcare systematic review searches in a cross-sectional sample. Research Synthesis Methods. 2011;2(2):119-25.
19. 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.
20. Pieterman K, Plaisier A, Govaert P, Leemans A, Lequin MH, Dudink J. Data quality in diffusion tensor imaging studies of the preterm brain: a systematic review. Pediatric radiology. 2015 Aug;45(9):1372-81.
21. Bramer WM. Expert searching – Searching for Systematic Reviews: Set Numbers or One Liners? – Column. MLA news 2014;54(4):15.

Chapter 8

The comparative recall of Google Scholar versus PubMed

Bramer WM, Giustini D, Kramer BMR, Anderson PF. 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.

8 The comparative recall of Google Scholar versus PubMed in identical searches for biomedical systematic reviews: a review of searches used in systematic reviews

Abstract

Background

The usefulness of Google Scholar (GS) as a bibliographic database for biomedical systematic review (SR) searching is a subject of current interest and debate in research circles. Recent research has suggested GS might even be used alone in SR searching. This assertion is challenged here by testing whether GS can locate all studies included in 21 previously published SRs. Second, it examines the recall of GS, taking into account the maximum number of items that can be viewed, and tests whether more complete searches created by an information specialist will improve recall compared to the searches used in the 21 published SRs.

Methods

The authors identified 21 biomedical SRs that had used GS and PubMed as information sources and reported their use of identical, reproducible search strategies in both databases. These search strategies were rerun in GS and PubMed, and analyzed as to their coverage and recall. Efforts were made to improve searches that underperformed in each database.

Results

GS' overall coverage was higher than PubMed (98% versus 91%) and overall recall is higher in GS: 80% of the references included in the 21 SRs were returned by the original searches in GS versus 68% in PubMed. Only 72% of the included references could be used as they were listed among the first 1,000 hits (the maximum number shown). Practical precision (the number of included references retrieved in the first 1,000, divided by 1,000) was on average 1.9%, which is only slightly lower than in other published SRs. Improving searches with the lowest recall resulted in an increase in recall from 48% to 66% in GS, and in PubMed, from 60% to 85%.

Conclusions

Although its coverage and precision are acceptable, GS, because of its incomplete recall, should not be used as a single source in SR searching. A specialized, curated medical database such as PubMed provides experienced searchers with tools and functionality that help improve recall, and numerous options in order to optimize precision. Searches for SRs should be performed by experienced searchers creating searches that maximize recall for as many databases as deemed necessary by the search expert.

8.1 Background

For several years, information specialists have discussed which databases (and how many) should be used to perform exhaustive searches of the literature. Prior to 2004, the year of Google Scholar's release, these discussions focused primarily on traditional databases such as Embase and MEDLINE (1, 2). Further, the general consensus had developed that searching a limited number of databases was insufficient where completeness was the goal (3–6).

In addition, the type of searching that is required to support systematic reviews (SRs) is more complex and time-consuming than searching for simple clinical queries. The demands placed on a searcher for the SR are much higher than for other searches because of the specific requirements of the SR (7) which is integral to science and must therefore be performed systematically, and made repeatable, verifiable and accountable (8, 9).

Since 2004, Google Scholar (GS) has been widely-used to locate specific items and aid in cumulating the scholarly literature. In 2005, Giustini (10) stated that GS produced acceptable results for browsing routines but results of low precision meant that its use for other searching was problematic. Since then, GS has improved its scope; from 2005 to 2012, its coverage of the literature rose from 30 to 88% to 98 to 100% (11, 12). An important unanswered question about GS remains: 'Is GS advanced enough in its development to replace more sophisticated tools such as PubMed or Embase?'

In 2007, Shultz (13) provided an overview of criticism of GS that had been generated since its debut in 2004. Unfortunately, many of the original shortcomings identified between GS and traditional bibliographic databases such as MEDLINE and Embase are still in evidence: GS lacks a controlled vocabulary, search histories and sets cannot be built and manipulated and wildcards and limits (for instance study types) cannot be used precisely. Only the first 1,000 citations of any search in GS are viewable and search strings must be kept under 256 characters.

Since 2004, a number of studies have examined the value of GS in biomedical searching. Falagas, Pitsouni, et al.(14) compared four databases including GS and PubMed, and concluded that GS retrieved more obscure items than other search tools. Anders and Evans (15) focused on using advanced searching in GS and PubMed but, given major differences in the databases, their study made true comparisons difficult. In Nourbakhsh, Nugent, et al.(16), researchers found that the first 20 results in GS often produced more relevant hits than similar searches in PubMed. But since PubMed, until very recently, lists citations in chronological order (not by algorithms, as in GS) the authors' conclusions are counter-intuitive. The frequently cited study by Walters (17) covered one topic (on older person migration), which is out-of-scope in a medical database. A recent study by Shariff, Bejaimal, et al.(18) compared search strategies designed by end users and compared the first 40 hits in GS and PubMed.

In 2013, Gehanno, Rollin, et al.(19) published a paper that generated important critical discussion of the value of GS. In their paper, Gehanno et al. used GS to locate all studies originally cited in a published SR. By finding all known items, the authors argued that GS, after some improvements to increase its search precision, could be used alone in searching for SRs.

The article by Gehanno et al. drew much attention to GS and resulted in some follow-up articles. Giustini and Kamel Boulos (20) argued that a 'known-item' searching is a very different activity than locating articles by subject, as is attaining 100% recall of the (mostly unknown) relevant literature. Boeker, Vach, et al.(21) reinvestigated the results of Gehanno et al. using search strategies designed to match Medline strategies used for Cochrane systematic reviews. However, the authors designed the searches themselves and failed to account for the maximum number of results that can be retrieved in GS (1,000), although they mentioned this limitation in their manuscript. The low precision in GS as reported by Gehanno et al. and by Boeker et al., is mainly an artifact due to the large number of hits that are reported by GS. Since search results cannot be viewed beyond the first 1,000 references, actual precision in GS should be calculated as the number of relevant references found in those first 1,000 references, divided by the number of hits that can actually be viewed, which is 1,000 at most.

The research done on GS for SRs is limited in methodology, which is crucial for evaluating GS for SR searching. In Table 8.1 we critique in more detail earlier research on GS.

Table 8.1: Limitations of current published research on the usability of Google Scholar for medical purposes

Limitations of research	References
Not testing for systematic reviews	(13–16, 18, 22–26)
Limited number of searches	(14–16, 24–26)
Relevancy of results only determined by the authors	(13, 14, 16, 23, 25, 26)
Not reviewing the first 1,000 hits in Google Scholar	(15, 16, 18, 19, 21, 22)
Only using searches designed by the authors	(14–16, 21, 22)
Searches not comparable between the databases	(15, 16, 23, 25)
Published more than five years ago	(13, 14, 23, 24)
Only searching for known items	(13, 14, 19)
Only looking at coverage, not retrieval	(19)

Though it seems unlikely that an experienced information specialist would use GS as the sole database in a SR, a less experienced researcher, faced with the enormous task of performing a review without expert help, might be tempted to do so (based on the aforementioned research). At least one review is known that, after doing preliminary searches in a wide range of databases decided ultimately to use only PubMed and GS, but failed to notice their search strategy was not executable in GS, since it was over 500 characters long (27).

In this paper, the usability of GS in searching for SRs is considered, where relevancy is pre-determined by inclusion in papers that have been previously published. Both the original (identical) topical searches reported by those papers and searches improved by an information specialist are used in order to compare the recall within GS and PubMed. The aim of this paper is to discover whether the original authors would have found all included references by using GS only. When studies from the original SRs were not found, it is assessed whether a more exhaustive search strategy created by an information specialist would improve recall. Given its potential for one-stop searching, we assess whether GS can indeed replace the multiple databases required for the SR and locate all studies needed to conduct a SR.

8.2 Methods

In May 2013, PubMed and Embase were searched using the exact phrases 'systematic review' and 'google scholar' in title and/or abstract fields. Of the records identified, the full-text of relevant papers was retrieved on the open web or via subscriptions at the first author's institution. The full-text and appendices of available articles were scanned for descriptions of the strategies used to search PubMed and GS.

Articles that clearly described identical search strategies were investigated further. The queries as performed in the initial searches in the SRs were recreated. If the SRs did not discuss a medical topic, the review was excluded as it was unlikely that PubMed would have been viewed as a valuable database in those instances. When the length of a reproduced search exceeded the maximum query length allowable in GS (256 characters) the review was also excluded. All inclusion and exclusion criteria are summarized in Table 8.2.

Table 8.2: Inclusion and exclusion criteria

Inclusion criteria	Exclusion criteria
Systematic review in a medical topic	Length of search strategy greater than 256 characters
Reporting the use of both Google Scholar and PubMed	Number of hits retrieved in PubMed exceeds the reported total number of hits reviewed
Reporting in reproducible detail an identical single phrase search for these databases	

Searches that were reproducible were executed in both GS and PubMed, and the number of hits was documented accordingly. In PubMed, results were limited to before the MeSH date (field: (mhda)) of the original search date as stated in the article. The number of hits in PubMed was compared with the number originally reported (either for PubMed, or the total for all databases). If it did not exceed that number, the SR was included. In GS, search results were limited to the publication start year used by the original authors and the end publication year of the search date. Because publication dates can differ from search dates (because publication dates are generally added to the print version, while the electronic version might be available longer) we checked whether the list of includes contained articles with newer publication dates, and if so, changed the publication limits accordingly.

For each replicated search, the first 1,000 results of GS were saved in a Word document. Using the 'find' function in Word, occurrences of each included reference from the original SR were identified. Distinctive fragments of the title were searched but where no match was located, author names were searched. If a citation was not found among the first 1,000 results, GS coverage for that item was checked using author names and part of the title between double quotes. If the item was indeed present in GS, the reference was checked for retrieval in the search query (beyond the first 1,000 hits) by combining author names and distinctive title words with the full query (to check whether they had ranked low on Google's PageRank algorithm).

We did not exclude hits that were citations only (by definition) but for those references, it was checked whether the citing articles, as linked in GS, were published before the citing systematic review. It could then be assumed that the citation was present in GS when the original authors performed their searches. If the citation was only present in articles with a more recent publication year, the result was confidently discarded.

All included studies were searched in PubMed by searching for the complete reference. If PubMed did not reveal a match, a second attempt was performed using a combination of first author (1au), page number (pg) and publication year (dp). Included references were collected using the PubMed Clipboard. Once all references were retrieved, clipboard contents were checked against the results of the replicated search.

The intention of this project was not to judge the quality of the replicated searches. In a later stage, we improved some searches to investigate whether more citations could be found. An experienced information specialist (WB) created improved search strategies for GS based on the original authors' description of their research question, without taking into account the included references from that SR. A second search strategy was designed based on the frequency of words in the titles of included references of these SRs. For the searches that had missed the most included references in PubMed, an information specialist created a more comprehensive search strategy using MeSH terms and free text, without using the included references to determine search words.

8.3 Results

Of the 578 SRs retrieved, the full-text was obtained for 453 articles. A total of 84 articles described in enough detail identical searches that could be rerun in PubMed and GS. Eight articles were excluded because their search strategies exceeded the maximum search length allowed by GS (256 characters). Twenty articles were excluded because they made no mention of the number of hits retrieved. Two articles were excluded because the topic was non-medical and therefore their search strategies returned no results in PubMed. Seven articles were excluded because the authors used multiple search queries and 24 others were excluded because the numbers reported (for PubMed or total) did not match number of hits retrieved for the replicated searches.

See Figure 8.1 for a flow diagram of the in- and exclusion procedure. For one article the list of included references contained three references from beyond the search year, thus we decided to expand the publication date limits with one year.

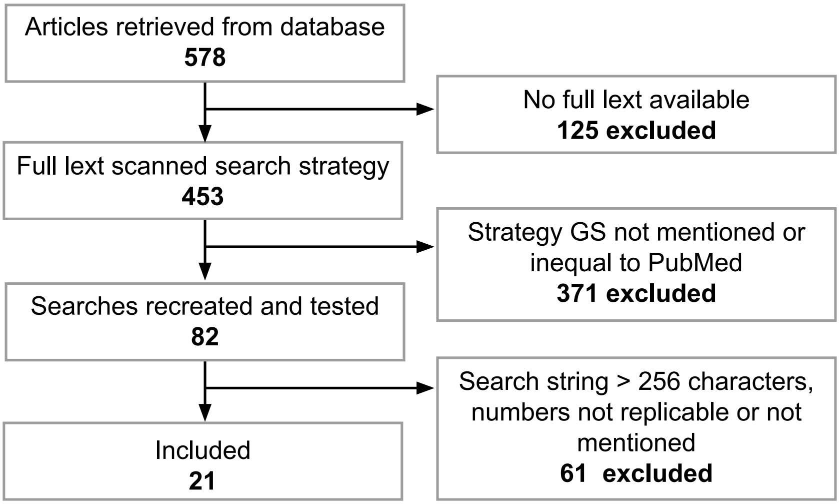


Figure 8.1: Flow diagram of reviewed articles.

In 21 articles, the cited searches for both GS and PubMed were identical, well-documented, and the number of hits in PubMed did not exceed the number of hits first reported. Additional file 1 describes the original and replicated searches along with other parameters and the detailed results. In eleven cases, the searches used in this research were exactly the same as those described in the full text or appendices of the studies. In ten instances, some minor changes had to be made. In some cases, Boolean operators (AND and OR) were not stated and nesting was not clearly laid out; proper searches cannot be performed without operators but what was intended was clear. If major changes were required, the reviews had been excluded. For five references retrieved as citations, the citing articles all had a publication date later than the original search date, so these citations were ignored. A full list of references to all SRs included in this article can be found in Additional file 2.

Coverage

The total number of studies included by the SRs was 541. In GS, ten studies were not present, thus the overall coverage of GS reached 98%. In PubMed, 48 references were not present, so PubMed had an overall coverage of 91%.

Recall

Of the total number of included studies that were reviewed (541), 389 (72%) were present in the first 1,000 hits of the original searches in GS. Forty-five articles had been retrieved by the search strategy in GS, but were not among the first 1,000 hits. If GS had allowed its users to review all search results, recall would thus have been 80%. The same searches retrieved 369 hits in PubMed (68%).

Precision

Figure 8.2 shows the distribution of observed practical precision of Google Scholar searches. The average practical precision was 1.9% and the median 1.7%. Average precision for SRs, according to Sampson, Tetzlaff, et al.(28) is around 2.9%. The practical precision in GS for the searches observed in this article is slightly below the reported average but 1.9% is nonetheless quite acceptable for SR searching, where, in order to be complete, researchers must browse through irrelevant hits to find important references.

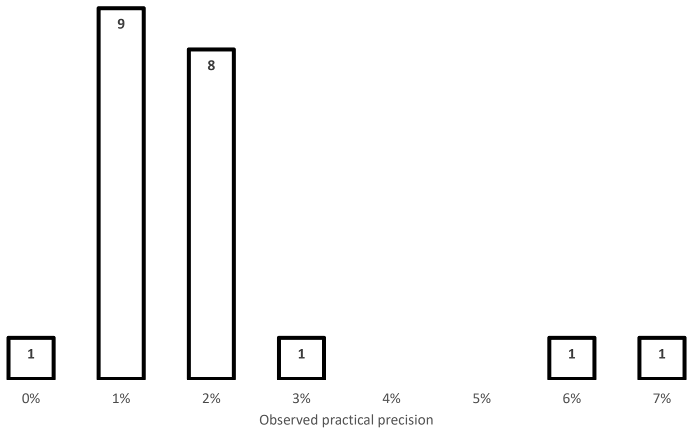


Figure 8.2: Practical precision of Google Scholar.

Improvement of search strategies in Google Scholar

163 included studies were not present in the first 1,000 hits of the original searches in GS. For five SRs where GS had missed more than ten included references (in total 110), we tried to improve the search strategies. Using the first improved search strategies, created without taking into account the included studies, retrieval in GS for these five SRs increased from 53% to 60%. The search strategies designed to capture as much included studies as possible (when the search strategies were based on words in the title of the included studies) resulted in 66% retrieval for these five SRs (Table 8.3). The improved search strategies for GS are available in Additional file 3.

Table 8.3: Systematic reviews of which more than ten included references were not retrieved in Google Scholar; performance of improved searches

	Number of included references	Number of includes retrieved by		
		Authors' search	Improved search #1	Improved search #2
Hasani	78	26	26	37
Novak	30	10	12	12
Verhoeven	89	72	75	72
Navarese	17	6	13	16
Belsey	20	10	15	16
	234	124 (53%)	141 (60%)	154 (66%)

Improvement of search strategies in PubMed

Of the references found in PubMed, 124 were not retrieved by replicating the original searches. Of these, 111 were included by the seven SRs in Table 8.4. The other SRs each had less than three included references that were not found in PubMed. We tried to create better searches for the research questions of these seven SRs to see if this would increase retrieval in PubMed. Using the improved search strategies increased retrieval for the seven SRs from 61% to 85%. The improved strategies for PubMed are available in Additional file 4.

Table 8.4: Systematic reviews that contributed most to the 'not retrieved articles' in PubMed; performance of improved searches

	Number of included references	Number of includes retrieved by	
		Authors' search	Improved search
Javan	68	29	61
Hasani	78	51	67
Verhoeven	89	67	81
Navarese	17	5	5
Novak	30	19	26
Gupta	16	8	12
Hardefeldt	26	18	23
	324	197 (61%)	275 (85%)

8.4 Discussion

Literature searching in multiple databases can often be cumbersome and is always time-consuming if it is done well. GS offers an easy-to-use, familiar interface and relevance ranking, making simple searching for a few good articles much easier. However, the use of GS as a robust search tool is not without its challenges.

To focus on the differences between the databases, the research was restricted to SRs that used both GS and PubMed with identical search strategies. Although this research represents a small sample of all published SRs, and is not representative, we nonetheless believe our findings to be of indicative of a trend. Had SRs been selected that did not describe their GS search strategies, it would have been necessary to create them which was not the purpose of this research. If SRs reported non-identical searches for GS and PubMed, the findings would be reviewing the ability of the original reviewers to translate their searches, which again was not the intention.

One of the most challenging aspects is the frequency with which Google changes its functionality without giving any prior notice to its users. In 2012, GS changed its advanced searching features and removed the ability to limit results to specific domains and disciplines (for instance medicine). In March 2013, GS reduced the maximum number of articles it can show per page from 100 to 20, and in June the tilde operator, that was very usable to search for synonyms (frequently used as a replacement for truncation), was removed from the regular search engine (google.com), at the moment it is still available in GS. These changes have a major impact on searching, and when users asked whether some of these features would be reinstated, Google said little (29). Even more threatening is the fact that since the end of 2011, GS has disappeared from the menu of regular Google, thus making it harder to find for those users who do not already know of its existence, although results from GS and a link to a search in GS ('Scholarly articles for…') often appear in the search results of regular Google. Because of Google's tendency to shut down applications it considers less frequently used (like recently Google Reader in June 2013), this might be a threat for the continuity of GS. And if GS were to be shut down, this would be a major threat to the replicability of the methodology of the SRs that were performed with a GS search.

Many published SRs report on the total number of hits in the databases they have used. However, the ratio between the number of hits we retrieved in PubMed and the total number of hits that were reported by the original authors varied a lot. In one case the reported total was 108 times higher than the number we found in PubMed, while for many other SRs the ratio between numbers reported and what was retrieved was equal to one. Some authors opted to report the total number of citations found in GS, while others took into account the first 1,000, or just reported the number they felt was necessary to view. Still, others ignored number of hits in GS in their reporting and counted only relevant, unique hits. Consequently, the resulting numbers in our searches were at odds with what the original authors reported in their published reviews. It is recommended that authors of SRs that use GS as a primary source report only those hits that were actually reviewed from GS.

Search reproducibility was low due to inaccurate or incomplete reporting of search strategies. Many papers that were examined referred to their search strategies by listing the keywords and Boolean operators used in an illogical order (30). Even in cases where searches were explicitly stated, the number of hits did not match the number of hits retrieved using the exact limiters and search parameters. To ensure transparency and reproducibility, authors of SRs should take care to follow the guidelines in the PRISMA Statement (7) for reporting search strategies. This states that the number of studies screened should be stated and not the number a database claims to have found.

Importing all references into reference management software is now a standard feature in bibliographic databases such as PubMed and Embase. GS does not offer such a feature. With Zotero, all results from one page can be imported from GS. However, recently the maximum number of hits shown per page changed from 100 to 20, making downloading the full set of hits more time consuming. When the authors of this article used Zotero to import the contents of a single search into Endnote, after downloading 200 references a 'Captcha' was shown, as Google had detected that our 'computer or network may be sending automated queries'. GS seemed disinclined to provide the flexibility required to properly search the literature for the SRs.

We observed that many items were found because GS indexes content beyond the abstract into the full-text of articles, including their references. Excerpts often show a part of the article containing the reference list with the searched words found in titles of referred articles. When searching for included references in the first 1,000, numerous false hits were encountered showing the included articles in the reference list of other retrieved articles. GS seems to perform citation tracking for articles by using search words in the title. This accounts for much of the extra hits (or 'noise') in GS' results.

Limitations of research

One limitation in this research (as in all retrospective research involving GS) is one can never be certain of GS' coverage at any specific point in time, which is a serious problem for searchers. We performed searches several years after the original reviews were performed (average search date was September 2010 but ranged from Jan 2007 to October 2012, while we searched GS in May/October 2013). Although we limited GS searches to the publication year of the original search date, the results will probably differ from those retrieved by the original authors. Not surprisingly, the number of hits retrieved during our searching was often at odds with what was originally reported. Since replicable searching is essential in performing research, the inability to reproduce searches in GS severely limits its value to researchers. Bibliographic databases such as PubMed offer additional database management dates next to publication dates. These databases keep track of content, and note when changes are made. In PubMed using the field mesh date ((mhda)), one can be rather certain what a search result would have been on a given date. GS lacks these management dates, and only offers publication dates. Because of the absence of a clear date restriction feature, replicability, the search process (which is crucial to SR searching) is very is rendered far more problematic than would be the case in curated databases such as Medline and Embase.

We limited our searches to the publication dates of the last search date. However, this reduced the reported hits substantially. Even when a theoretical limit to publication years 1800 to 2099 was applied, the number of hits dropped. In simple queries, this seemed to affect only the number of hits, but not the resulting references. The first 1,000 articles remain largely the same, making this a good replacement for date limits in bibliographic databases. However, on the improved search strategies that were more complex, the number of hits reported often dropped a factor ten, even with the theoretical limit to all publications dates. Resulting references differed immensely, and the first 1,000 hits hardly contained any of the included references. Therefore, for the improved searches, publication date limits were not used. This is a newly identified problem when using GS for SR searching.

Though it was not our intention to judge the quality of the searches of the SRs, we believe that the quality of the searches was poor as they often only combined a few words with hardly any synonyms. This is of course also due to the selection process: if an identical search is used in GS, none of the more sophisticated tools of PubMed (for instance MeSH terms, field codes and truncation) could have been used.

Limitations of Google Scholar

Improving search strategies is a major challenge in GS. As GS cannot store search histories, it is mostly impossible to build multi-set queries or evaluate changes made to search queries. Search strings are momentarily limited to 256 characters, and searchers have to select their keywords accordingly. This is further complicated by the fact that GS does not allow truncation as required by searchers. A feature that could replace truncation in GS is the tilde (~), which automatically searches for word variants. Although useful, the feature does not work in combination with Boolean operators like OR, and therefore cannot be used in exhaustive search queries. In addition, the feature was recently deprecated from Google's regular search engine, making its continued availability in GS uncertain. Using | instead of OR and by simply leaving out ANDs it is possible to use as many of those characters for search words as possible. Finally, GS has no feature for proximity or adjacency searching; parentheses can be used to search word variants in a double-quoted phrase ("(myocardial|heart) (infarct|attack)"), as well as asterisks (*), but the number of asterisks used marks the exact number of words allowed, where proximity searching in other databases generally describes a range of words. These missing features make a translation of a proper search query as designed for other databases difficult if not impossible. The limitations experienced in this research are presented in Table 8.5.

Table 8.5: Comparison of Google Scholar and PubMed in systematic review searching

Google Scholar	PubMed
Shows up to 1,000 results	Shows all results
Searches the full text of the article, and words on the webpage	Searches only bibliographic data and controlled vocabulary (MeSH terms)
No controlled vocabulary available	Controlled vocabulary (MeSH terms) added by skilled indexers and searchable (including 'explode')
Searches the broad aspect of science (filters limiting results to medical articles were removed)	Only contains articles on medical topics
No search history available (unable to compare or combine record sets)	Detailed search history available, flexibility in combining record sets to create complicated search strategies
Search queries limited to 256 characters	No limits on the length of search queries
No truncation allowed. Tilde can be used to search for variants, but cannot be used in OR relationship with other words. GS is said to search for word variants, but this is very rare and the mechanism is unclear.	Truncation allowed
Possibly automatic searching for synonyms (details unclear)	Automatic Term Mapping (details available)
Field names only for title (complete query) and author names	Field names for many fields can be assigned per synonym
No advanced limits (for publication type, human studies and so on)	Multiple advanced limits in the database itself, or available from third parties
Cannot accurately limit to search dates (no controlled updates)	Different date fields available to limit searches to results before a certain date
Cannot download results in bulk to reference management software	Multiple options to download the complete results set to reference management software
Proximity searching only with exact order and exact number of connecting words	No proximity search possible

Though we wanted to improve the searches that returned the least included references, we could not draw conclusions about the effects of our improvements. We do not know, for example, if the authors missed important references in their original searches. We assume that within the extra hits retrieved, extra relevant studies might have been found. Because GS was unable to retrieve all articles found by the authors, its results cannot be considered complete.

8.5 Conclusion

As we've shown, the coverage and precision of GS are acceptable. Although coverage is not 100%, for many investigators 98% might suffice for simple literature or narrative reviews of a topic. The overall precision of GS (using the total reported number of hits as denominator) is rather low, but the practical precision, calculated by the number of relevant hits in the first 1,000, with 1,000 as denominator is theoretically acceptable for SR searching, and highly dependent on the number of included references.

Our a priori question was 'Is GS's recall sufficient to be used on its own in systematic review (SR) searching?' Overall retrieval in GS is 72%, which is too low for it to be used as a single database for the SR. PubMed fared similarly at 68%. The creation of better searches in GS proved to be a constant challenge. A high of 66% recall was achieved for the five searches that initially missed the most references from the SRs. In PubMed, our improved searches reached 85% recall. Neither database was sufficient on its own to find all articles from previously-published SRs.

Researchers from other disciplines might find results in GS that are 'good enough'. Similarly, medical professionals might find using PubMed (or another bibliographic database) on its own is good enough for day-to-day searching. Some may even prefer GS for initial searching to find 'a few good articles' due to its excellent relevance ranking.

However, SRs require a complete view of all existing literature in a given area. This can only be achieved by performing exhaustive searches of relevant databases and websites in consultation with a trained information specialist. These important searches that support the SR methodology must be repeatable, verifiable and accountable, which poses a problem with GS.

There is therefore no reason why GS should be viewed as more suitable for performing SR searches than PubMed (or any other specialized database). We hope that our research will inform future authors and guide their use of GS. Authors of future SRs should continue to use GS but in concert with multiple other databases, not as a replacement of other databases.

8.6 References

1. Sampson M, Barrowman NJ, Moher D, Klassen TP, Pham B, Platt R, St John PD, Viola R, Raina P: Should meta-analysts search Embase in addition to Medline? J Clin Epidemiol 2003, 56:943–955.
2. Watson RJ, Richardson PH: Identifying randomized controlled trials of cognitive therapy for depression: comparing the efficiency of Embase, Medline and PsycINFO bibliographic databases. Br J Med Psychol 1999,72(Pt 4):535–542.
3. 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, 58:867–873.
4. Topfer LA, Parada A, Menon D, Noorani H, Perras C, Serra-Prat M: Comparison of literature searches on quality and costs for health technology assessment using the MEDLINE and EMBASE databases. Int J Technol Assess Health Care 1999, 15:297–303.
5. Crumley ET, Wiebe N, Cramer K, Klassen TP, Hartling L: Which resources should be used to identify RCT/CCTs for systematic reviews: a systematic review. BMC Med Res Methodol 2005, 5:24.
6. 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, 5:6.
7. Moher D, Liberati A, Tetzlaff J, Altman DG: The Prisma Group: Preferred Reporting Items for Systematic Reviews and Meta-Analyses: the PRISMA Statement. PLoS Med 2009, 6:e1000097.
8. McGowan J, Sampson M: Systematic reviews need systematic searchers. J Med Libr Assoc 2005, 93:74–80.
9. Higgins JPT, Green S: Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration 2011. http://www.cochrane-handbook.org
10. Giustini D: How Google is changing medicine. BMJ 2005, 331:1487–1488.
11. Chen X: Google Scholar's dramatic coverage improvement five years after debut. Ser Rev 2010, 36:221–226.
12. Harzing AW: A longitudinal study of Google Scholar coverage between 2012 and. Scientometrics 2013, 2013:1–11.
13. Shultz M: Comparing test searches in PubMed and Google Scholar. J Med Libr Assoc 2007, 95:442–445.
14. Falagas ME, Pitsouni EI, Malietzis GA, Pappas G: Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. FASEB J 2008, 22:338–342.
15. Anders ME, Evans DP: Comparison of PubMed and Google Scholar literature searches. Respir Care 2010, 55:578–583.
16. Nourbakhsh E, Nugent R, Wang H, Cevik C, Nugent K: Medical literature searches: a comparison of PubMed and Google Scholar. Health Info Libr J 2012, 29:214–222.
17. Walters WH: Google scholar search performance: comparative recall and precision. Portal 2009, 9:5–24.
18. Shariff SZ, Bejaimal SA, Sontrop JM, Iansavichus AV, Haynes RB, Weir MA, Garg AX: Retrieving clinical evidence: a comparison of PubMed and Google Scholar for quick clinical searches. J Med Internet Res 2013, 15:e164.

19. Gehanno J-F, Rollin L, Darmoni S: Is the coverage of Google Scholar enough to be used alone for systematic reviews. BMC Med Inform Decis Mak 2013, 13:7.
20. Giustini D, Kamel Boulos MN: Google Scholar is not enough to be used alone for systematic reviews. Online J Public Health Inform 2013, 5:214.
21. Boeker M, Vach W, Motschall E: Google Scholar as replacement for systematic literature searches: good relative recall and precision are not enough. BMC Med Res Meth 2013, 13:131.
22. Freeman MK, Lauderdale SA, Kendrach MG, Woolley TW: Google Scholar versus PubMed in locating primary literature to answer drug-related questions. Ann Pharmacother 2009, 43:478–484.
23. Haase A, Follmann M, Skipka G, Kirchner H: Developing search strategies for clinical practice guidelines in SUMSearch and Google Scholar and assessing their retrieval performance. BMC Med Res Methodol 2007, 7:28.
24. Henderson J: Google Scholar: a source for clinicians? Can Med Assoc J 2005, 172:1549–1550.
25. Mastrangelo G, Fadda E, Rossi CR, Zamprogno E, Buja A, Cegolon L: Literature search on risk factors for sarcoma: PubMed and Google Scholar may be complementary sources. BMC Res Notes 2010, 3:131.
26. Tober M: PubMed, ScienceDirect, Scopus or Google Scholar - Which is the best search engine for an effective literature research in laser medicine? Med Laser Appl 2011, 26:139–144.
27. Kennedy CM, Powell J, Payne TH, Ainsworth J, Boyd A, Buchan I: Active assistance technology for health-related behavior change: an interdisciplinary review. J Med Internet Res 2012, 14:e80.
28. Sampson M, Tetzlaff J, Urquhart C: Precision of healthcare systematic review searches in a cross-sectional sample. Res Synth Meth 2011, 2:119–125.
29. Google Scholar Removes Search By Discipline. [http://www.seroundtable.com/google-scholar-discipline-16631.html]
30. Golder S, Loke Y, McIntosh HM: Poor reporting and inadequate searches were apparent in systematic reviews of adverse effects. J Clin Epidemiol 2008, 61:440–448.

Chapter 9

Comparing Embase, MEDLINE and Google Scholar

Bramer WM, Giustini D, Kramer BMR. Comparing the coverage, recall, and precision of searches for 120 systematic reviews in Embase, MEDLINE, and Google Scholar: a prospective study. Syst Rev. 2016;5(1):39.

9 Comparing the coverage, recall and precision of searches for 120 systematic reviews in Embase, MEDLINE and Google Scholar: a prospective study

Abstract

Introduction

Previously, we reported on the low recall of Google Scholar (GS) for systematic review (SR) searching. Here, we test our conclusions further in a prospective study by comparing the coverage, recall and precision of SR search strategies previously performed in Embase, MEDLINE and GS.

Methods

The original search results from Embase and MEDLINE and the first 1,000 results of GS for librarian-mediated SR searches were recorded. Once the inclusion-exclusion process for the resulting SR was complete, search results from all three databases were screened for the SR's included references. All three databases were then searched post-hoc for included references not found in the original search results.

Results

We checked 4,795 included references from 120 SRs against the original search results. Coverage of GS was high (97.2%) but marginally lower than Embase and MEDLINE combined (97.5%). MEDLINE on its own achieved 92.3% coverage. Total recall of Embase/MEDLINE combined was 81.6% for all included references, compared to GS at 72.8% and MEDLINE alone at 72.6%. However, only 46.4% of the included references were among the downloadable first 1,000 references in GS. When examining data for each SR, the traditional databases' recall was better than GS, even when taking into account included references listed beyond the first 1,000 search results. Finally, precision of the first 1,000 references of GS is comparable to searches in Embase and MEDLINE combined.

Conclusions

Although overall coverage and recall of GS are high for many searches, the database does not achieve full coverage as some researchers found in previous research. Further, being able to view only the first 1,000 records in GS severely reduces its recall percentages. If GS would enable the browsing of records beyond the first 1000 its recall would increase but not sufficiently to be used alone in SR searching. Time needed to screen results would also increase considerably. These results support our assertion that neither GS, nor one of the other databases investigated, is on its own, an acceptable database to support systematic review searching.

9.1 Background

In 2013, an article by Gehanno et al (1) prompted a discussion around the utility of Google Scholar (GS) to support systematic review (SR) searching. In response, we examined the recall of GS and PubMed search strategies for included references of published biomedical SRs.(2) There, we determined that the recall of all included references found among the first 1,000 search results in GS was insufficient for it to be used on its own to support SR searching.

In our 2013 study we intentionally selected search strategies that were identical in PubMed and GS, in an effort to study the effect of the database, instead of the quality of the query translation. Therefore, the search strategies used for PubMed in our previous study did not fully use the possibilities of a traditional database search strategy. Librarian-mediated searches (combining MeSH terms and free text terms in traditional databases) achieve better results than non-librarian searches.(3, 4) References that would have been included, had they been retrieved, could have been missed in PubMed due to a lack of MeSH terms in the search strategies. In our previous paper, we showed that optimization of search strategies in PubMed (adding MeSH terms and more synonyms) led to more improvement than a similar process in GS (adding extra terms found in the included references). Here we investigate whether an experienced information specialist, an expert at performing systematic review searches, can find all included references using only one database.

Our prior research replicated search strategies used in previously published systematic reviews. One of GS' shortcomings is that searches are never wholly replicable later, as the search algorithm is constantly changing day to day. GS can only limit search results to publication date ranges. In traditional databases such as PubMed, search results can be limited to specific dates, such as MeSH date (date when MeSH terms were altered), or entry date (date a record was added to database). Search results can be reproduced in PubMed as they were performed on a specific day, month and year.

GS does not only index papers which it found as full text, but it also finds references merely because they were cited by papers (which are then marked as (citation)). In this article, we refer to references marked by GS as (citation) as 'citation only'. As we used only published SRs in our previous paper, for which most of the full text had already been indexed by GS, we hypothesize that GS probably covered all included references at least as citation only. Due to GS's ever-changing database, search engine and relevance ranking algorithm, searchers are never confident these citation only results were present at the time of the original search.

As a follow-up, we aim to evaluate the search results of systematic review search strategies created by an experienced information specialist at the time they were conducted in MEDLINE via the Ovid interface, combined searching in Embase and MEDLINE via Embase.com and GS. Our goal is to compare the coverage of these databases and their performance in terms of precision and recall for included references in SRs.

9.2 Methods

The first author regularly performs librarian-mediated searches to support SRs in the academic hospital setting in which he works. The reviews generally cover a wide range of medical topics, from therapeutic effectiveness and diagnostic accuracy to ethics and public health. The methods used at Erasmus MC to create systematic review search strategies will be described in detail in a separate paper.

In short, the first author performs single-line search strategies in Embase.com, which are developed using a unique optimization method. The Embase.com search strategies are translated into other databases and interfaces using macros. These macros are developed in MS Word to search for syntax from one interface and replace it with an appropriate syntax for another interface. After automatic translation of syntax from Embase.com to MEDLINE in Ovid, Emtree terms for Embase are manually replaced with appropriate MeSH terms.

Search strategies for GS are derived from an array of words searched in titles and abstracts in Embase.com. All relevant search terms are copied and truncated terms are expanded to the most common term(s). To adhere to the limitation of 256 characters, the length of each search strategy is reduced by replacing all Boolean operators OR, including its surrounding spaces with |, effectively reducing the number of characters per synonym by three. Proximity operators in Embase.com are replaced in GS by combining optional search terms in quoted phrases. Thus if an Embase.com search strategy for liver cancer contains ((liver OR hepatic) NEAR/3 (cancer* OR tumor* OR neoplasm*)) this is translated to "liver|hepatic cancer|tumor|neoplasms" in GS. If the total number of characters in the GS search exceeds 256, the information specialist (often together with the reviewer) decides which search terms are likely to be least relevant and deletes them one at a time, until the threshold is reached. In the supplementary material some examples of search strategy translations between the three databases mentioned are provided.

SR searches were documented at the institution of the first author before researchers began to screen articles for inclusion. The total search results from two major biomedical databases: MEDLINE in the Ovid interface, Embase at Embase.com (searching both Embase and MEDLINE records) and Google Scholar (where Publish or Perish software (5) allowed downloading of the first 1,000 search results) were imported into EndNote after searching was concluded (at the start of the systematic review project).

Reviewers obtained full search results from all databases which involved at least the Cochrane Registry of Trials, Web of Science and a subset of PubMed to find recent articles. Occasionally additional databases were used such as Scopus, CINAHL (via EBSCOhost) or PsycINFO (via Ovid). Reviewers were advised to seek other sources of included references by using cited and citing references tracking, contacting key authors in the field, and hand-searching journals, but the decision to do so was up to the researchers. In the first author's institution, as in many other institutes, these tasks are generally performed by the researchers, not as a library service.

After the process of collecting included references was completed, reviewers provided us with a list of included references. Alternatively, these were retrieved using the reference lists of resulting publications. We searched for all included references one-by-one in the original files in EndNote, using author names, year and if necessary parts of the title. Record numbers of positive matches in EndNote were used to determine the database(s) from which each included reference was retrieved.

For included references not found in the first 1,000 results from GS, post-hoc GS searches were conducted. Original search strategies used for the SR were combined with author names preceded by 'author:' and distinct words or phrases from titles, preceded by 'intitle:'. If included references were retrieved by this search, they were identified as part of overall recall of the total number of hits reported. Where combinations of these data elements for the included reference, together with the original search strategies exceeded 256 characters the original search strategies were divided into separate searches. Positive hits were confirmed when both separate searches, combined with the article's metadata, retrieved the item.

When included references were present in GS as citations only, this was documented regardless of whether they had been found in the first 1,000 search results, in the total search results or as a positive coverage. When included references were found as citations only, all citing articles were checked. When the single article citing this included reference was the published review for which the search strategy was first designed, we concluded that the result must have been indexed after the search strategy was originally performed. This included reference was thus not taken into account in the overall coverage of GS. For all three databases, coverage of non-retrieved included references from the inclusion sets was checked thoroughly by searching the databases for author names, distinct words from titles and publication year, using multiple combinations if necessary to ensure no included references were missed.

From these results, overall coverage (number of included references available in the database divided by the total number of included references), recall (number of included references found in the search results for the original search strategies for a database divided by the total number of included references retrieved by all databases together) and precision (number of included references retrieved by a certain database divided by the total number of search results retrieved by that database) of the three databases were calculated. We additionally calculated recall and precision for the first 1,000 hits of GS. All data were calculated for the total set of included references (overall values), as well as per review. After we determined which search strategies scored exceptionally well or low on recall in the first 1,000 search results in GS, we examined the characteristics of the search strategies (topics and number of search terms) and search results (number of hits and number of included references).

We visualized most data in boxplot figures. A general legend can be found in figure 9.1.

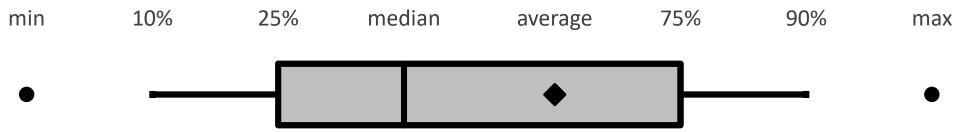


Figure 9.1: Legend of boxplot figures

9.3 Results

Between May 2013 and August 2015, 520 exhaustive searches designed for SRs by the first author were saved and documented. In August 2015 the reviewers of 120 SRs had screened all search results against their review's unique inclusion and exclusion criteria. In aggregate, these reviews had included a total of 4,795 references. The results for overall recall and coverage of the original 120 search strategies for all three databases for these 4,795 included references are summarized in figure 9.2.

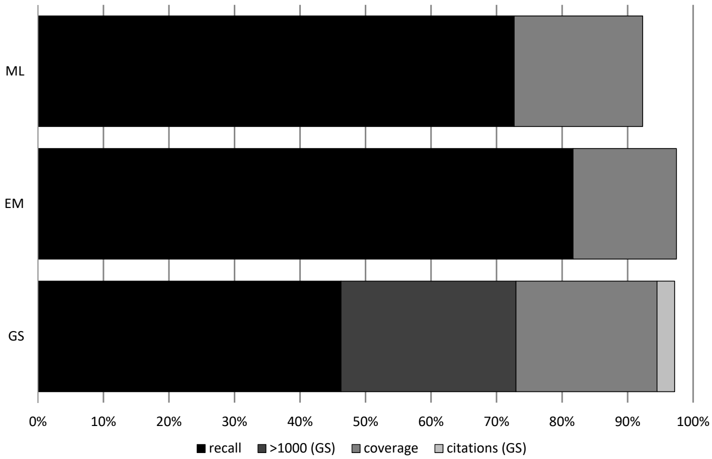


Figure 9.2: Total coverage and recall of MEDLINE (ML), Embase/MEDLINE (EM), and Google Scholar (GS)

Overall coverage

Overall, GS contained 4,708 of the total number of included references (N=4,795). However, 179 of these were present as citations only. In 49 of these citation only results, the only citing paper in GS was the review based on our search strategy. These 49 search results could not have been covered in GS at the time of the original search. Therefore, overall coverage of the included references was 4,659 (97.2%). In Embase the percentage of included references found in the database was slightly above GS at 97.5% while MEDLINE produced 92.3% of all included references (See Figure 9.2).

9

Coverage per SR

For individual SRs the percentage of included references present in the three databases varied. For 68% of all SRs the coverage of GS was 100%, Embase contained 100% of all included references for 63% of all reviews, compared to 34% for MEDLINE. For individual SRs the coverage of GS can be as low as 72%, 77% was the lowest observation in Embase, and 61% in MEDLINE. See figure 9.3 for a visualization of the coverage per SR.

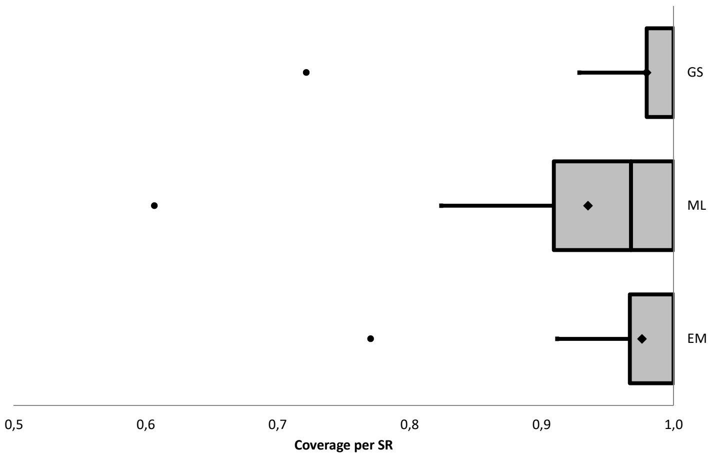


Figure 9.3: Coverage per SR for GS, MEDLINE, and Embase/MEDLINE

Overall recall

In terms of overall recall, Embase/MEDLINE was the most complete, retrieving 3,914 of all included references (81.6%), while MEDLINE alone retrieved 3,481 included references (72.6%). Counting all search results found by the search strategies, GS retrieved 3,493 included references (72.8%). However, only 2,224 of those were downloaded with the combined first 1,000 search results for the 120 SRs, so the actual recall of GS is much lower at 46.4% (See Figure 9.2).

Recall per SR

For individual SRs, the percentage of included references present in the first 1,000 search results in GS varied by a wide margin. In fifteen SRs, fewer than 25% of all included references that had been identified through database searches were found in the first 1,000 search results of GS, but nine SRs achieved the maximum 100%.

Recall fared much better in GS when all search results were taken into account (see Figure 9.4). A rate of at least 100% was reached in 24 SRs (20.0%). For four SRs the recall of all search results in GS was even higher than 100% because GS was able to find included references that had not been found in the traditional databases, but were identified via other sources (e.g. reference checking or hand searching). The recall of traditional databases such as Embase and MEDLINE was more consistent of which Embase / MEDLINE performed the best, although its minimum recall was only 43%.

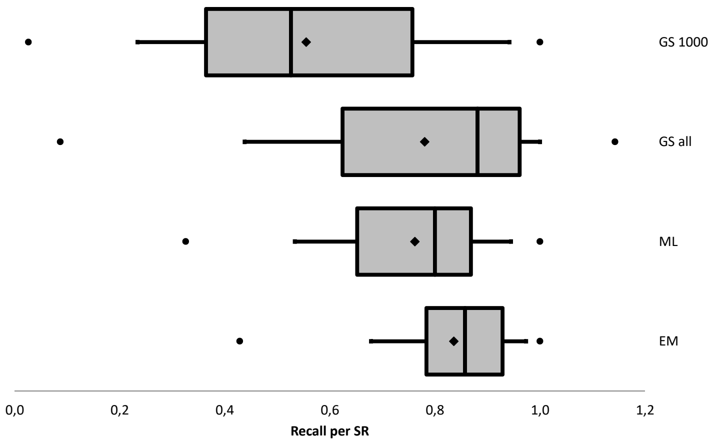


Figure 9.4: Recall per SR for the first 1000 references in GS, total GS, MEDLINE, and Embase/MEDLINE

Overall precision

The total number of search results that were downloaded from GS was 118,509 (in 4 of 120 reviews the number of hits in GS was lower than 1,000). These search results together contained 2,224 of the included references in the SRs, thus the overall precision of the first 1,000 search results of GS is 1.9%. The total reported number of search results in GS was 10,092,939 of which 3,493 were included references, thus the overall precision of the complete search results of GS was 0.03%. The precision of Embase was 3,940 / 192,935 = 2.0%, and for MEDLINE 3,506 / 126,657 = 2.8%. These data are visualized within figure 9.5.

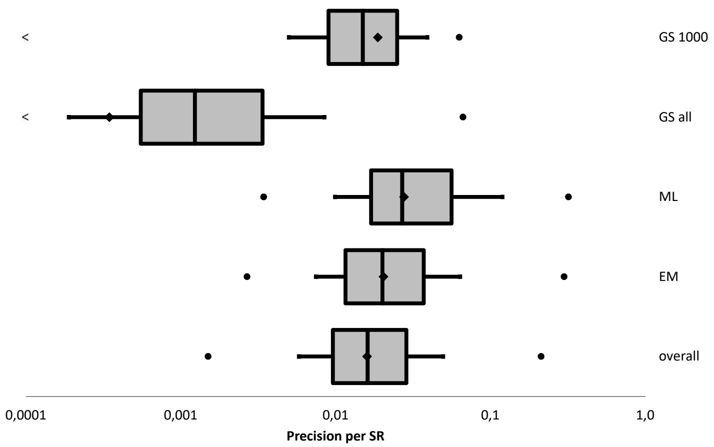


Figure 9.5: Precision per SR for the first 1000 references in GS, GS, MEDLINE, Embase/MEDLINE, and overall

Precision per SR

The precision of GS' first 1,000 search results (1.9%) did not differ much from precision as observed in all databases (1.6%) that were searched in the review process (see Figure 9.5). However the precision of the total set of search results in GS was much lower than that of the other databases (0.03%).

Why GS scored low is a valid question. Some reasons are discussed below. In some cases (Ahmadi et al. (6), Ambagtsheer (not yet published), Leermakers, Moreira (7)) the first author together with the reviewer had not been able to translate a complicated embase.com search strategy into a GS search, due to the lack of proximity operators. Another reason was that the search strategy of Embase was too long for all important search terms to be used in the GS search strategy. In some cases recall in the traditional databases was possibly higher because of the use of thesaurus terms for a broad topic (such as sexual risk behavior, Legemate, not yet published) or because the fact that topic was very broad which could have resulted in many non-medical references in GS (music in premature infants, Oliai Araghi, not yet published). In other cases it is unclear why there is such a vast difference between recall in Embase and GS (Bramer (7)). SRs where GS scored exceptionally well often try to answer well-defined topics, such as cashew nut allergy (van der Valk et al. (8)), or platelet-rich plasma injections for tennis elbow (de Vos et al. (9)).
Search strategies for Embase.com, Medline via Ovid and GS for already published reviews where GS scored exceptionally high or low (as cited in the paragraph above) are shown in the supplementary materials.

9.4 Discussion

GS covers a vast amount of literature but, when excluding citation only results first indexed after publication of the reviews used in this research, overall coverage of Embase is slightly higher. Overall recall of GS is not higher than when searching MEDLINE only and much lower than when searching both MEDLINE and Embase. Since only the first 1,000 search results of GS can be used, practical recall of GS is exceptionally low, which makes GS unacceptable as a single database to support the SR. If all search results in GS were made available to users, recall would still be too low for SRs, but reviewer burden would increase due to loss of precision. In fact, none of the observed databases can be used as single databases for SR searching, as the best performing database (Embase/MEDLINE) for individual SRs can result in a recall of less than 50%. Our observations are similar to recent observations made by Haddaway et al. (10) who compared the recall of GS to that of Web of Science for SRs in the field of environmental science.

Low precision has always been considered a problem in GS (11), but when accounting for actually usable search results (i.e. the first listed 1,000), we observed precision to be only slightly lower than the 2.9% as reported by Sampson et al in 2011, and comparable to that observed in the other databases.(12) Further, the precision observed in all databases in our study was nearly equal to the practical precision of the first 1,000 hits of GS.

The results of this prospective research are for the most part comparable to our previous, retrospective, study[2]. The coverage of GS and recall in MEDLINE are similar to those observed in 2013. However, recall in Google Scholar is much lower for our original searches than for the reconstructed searches in our previous study (45.1% vs. 72%). That is probably because our search strategies, as they were designed by an experienced information specialist, were optimized to find as many included references as possible in the traditional databases. We translated these search strategies with our best of knowledge into a GS search strategy, but were unable to reach a high recall in GS as we had succeeded in the traditional databases.

In SRs, ideally, extended search methods that go beyond traditional databases are used to find included references. Total number of included references is therefore sometimes higher than the number of included references retrieved in the downloaded search results from traditional databases. In evaluating results of GS search strategies for known items from the included references, some articles did meet the search strategy's criteria, but were not considered relevant enough by GS to be among the first 1000 viewable search results. For some SRs therefore the recall of the complete GS search results was higher than 100%.

The current research could be improved by using search strategies created by multiple independent information specialists at baseline but we question whether such a change would alter our conclusions. Research on GS for SRs should not focus on whether to use the search tool as a single source, but whether it adds value to the search results from other databases. The authors are currently collecting data for a follow up study that can answer the question whether GS is able to locate included references unidentified by the traditional databases.

9.5 Conclusions

Despite its vast coverage of the scholarly literature, Google Scholar is not sufficient to be used on its own as a single database to support SR searching. The reason for this is not low precision in GS searching, which is comparable to traditional databases. More problematic is GS' low recall capabilities which are related to the viewable 1,000 search results only policy of the search engine. Even if Google Scholar were to allow users to browse beyond the first 1,000 search results, its overall recall would still be too low to locate all included references to support the systematic review. We conclude similarly that neither Embase nor MEDLINE on its own is sufficient in retrieving all included references for SRs.

9.6 References

1. Gehanno J-F, Rollin L, Darmoni S. Is the coverage of google scholar enough to be used alone for systematic reviews. BMC Medical Informatics and Decision Making. 2013;13(1):7.
2. Bramer WM, Giustini D, Kramer BM, Anderson PF. The comparative recall of Google Scholar versus PubMed in identical searches for biomedical systematic reviews: a review of searches used in systematic reviews. Systematic reviews. 2013;2:115. doi:10.1186/2046-4053-2-115.
3. Gardois P, Calabrese R, Colombi N, Deplano A, Lingua C, Longo F et al. Effectiveness of bibliographic searches performed by paediatric residents and interns assisted by librarians. A randomised controlled trial. Health information and libraries journal. 2011;28(4):273-84.
4. Brettle A, Hulme C, Ormandy P. Effectiveness of information skills training and mediated searching: qualitative results from the EMPIRIC project. Health information and libraries journal. 2007;24(1):24-33. doi:10.1111/j.1471-1842.2007.00702.x.
5. Harzing AW. Publish or Perish. 2007. http://www.harzing.com/pop.htm Accessed 27 October 2015.
6. Ahmadi AR, Lafranca JA, Claessens LA, Imamdi RM, JN IJ, Betjes MG et al. Shifting paradigms in eligibility criteria for live kidney donation: a systematic review. Kidney Int. 2015;87(1):31-45. doi:10.1038/ki.2014.118.
7. Bramer WM. Evaluation of instructive texts on searching medical databases. J Med Libr Assoc. 2015;103(4):208-9. doi:10.3163/1536-5050.103.4.010.
8. van der Valk JP, Dubois AE, Gerth van Wijk R, Wichers HJ, de Jong NW. Systematic review on cashew nut allergy. Allergy. 2014;69(6):692-8. doi:10.1111/all.12401.
9. de Vos RJ, Windt J, Weir A. Strong evidence against platelet-rich plasma injections for chronic lateral epicondylar tendinopathy: a systematic review. Br J Sports Med. 2014;48(12):952-6. doi:10.1136/bjsports-2013-093281.
10. Haddaway NR, Collins AM, Coughlin D, Kirk S. The Role of Google Scholar in Evidence Reviews and Its Applicability to Grey Literature Searching. PLoS ONE. 2015;10(9):e0138237. doi:10.1371/journal.pone.0138237.
11. Boeker M, Vach W, Motschall E. Google Scholar as replacement for systematic literature searches: good relative recall and precision are not enough. BMC Medical Research Methodology. 2013;13(1):131.
12. Sampson M, Tetzlaff J, Urquhart C. Precision of healthcare systematic review searches in a cross-sectional sample. Research Synthesis Methods. 2011;2(2):119-25. doi:10.1002/jrsm.42.

Chapter 10

Optimal database combinations

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.

10 Optimal database combinations for literature searches in systematic reviews: a prospective exploratory study

Abstract

Background

Within systematic reviews, when searching for relevant references, it is advisable to use multiple databases. However, searching databases is laborious and time consuming, as syntax of search strategies are database specific. We aimed to determine the optimal combination of databases needed to conduct efficient searches in systematic reviews, and whether the current practice in published reviews is appropriate. While previous studies determined coverage of databases, we analyzed the actual retrieval from the original searches for systematic reviews.

Methods

Since May 2013, the first author prospectively recorded results from systematic review searches that he performed at his institution. PubMed was used to identify systematic reviews published using our search strategy results. For each published systematic review, we extracted the references of the included studies. Using the prospectively recorded results and the studies included in the publications, we calculated recall, precision, and number needed to read for single databases and databases in combination. We assessed the frequency at which databases and combinations would achieve varying levels of recall (i.e., 95%). For a sample of 200 recently published systematic reviews we calculated how many had used enough databases to ensure 95% recall.

Results

A total of 58 published systematic reviews were included, totaling 1746 relevant references identified by our database searches, while 84 included references had been retrieved by other search methods. Sixteen percent of the included references (291 articles) were only found in a single database; Embase produced the most unique references (n=132). The combination of Embase, MEDLINE, Web of Science Core Collection, and Google Scholar performed best, achieving an overall recall of 98.3% and 100% recall in 72% of systematic reviews. We estimate that 60% of published systematic reviews do not retrieve 95% of all available relevant references as many fail to search important databases. Other, specialized databases, such as CINAHL or PsycINFO add unique references to some reviews where the topic of the review is related to the focus of the database.

Conclusions

Optimal searches in systematic reviews should search at least Embase, MEDLINE, Web of Science and Google Scholar as a minimum requirement to guarantee adequate and efficient coverage.

10.1 Background

Investigators and information specialists searching for relevant references for a systematic review (SR) are generally advised to search multiple databases and to use additional methods to be able to adequately identify all literature related to the topic of interest (1-6). The Cochrane Handbook, for example, recommends the use of at least MEDLINE and Cochrane Central and, when available, Embase for identifying reports of randomized controlled trials (7). There are disadvantages to using multiple databases. It is laborious for searchers to translate a search strategy into multiple interfaces and search syntaxes, as field codes and proximity operators differ between interfaces. Differences in thesaurus terms between databases adds another significant burden for translation. Furthermore, it is time-consuming for reviewers who have to screen more, and likely irrelevant, titles and abstracts. Lastly, access to databases is often limited and only available on subscription basis.

Previous studies have investigated the added value of different databases on different topics (8-15). Some concluded that searching only one database can be sufficient as searching other databases has no effect on the outcome (16, 17). Nevertheless others have concluded that a single database is not sufficient to retrieve all references for systematic reviews (18, 19). Most articles on this topic draw their conclusions based on the coverage of databases (14). A recent paper tried to find an acceptable number needed to read for adding an additional database, sadly however, no true conclusion could be drawn (20). However, whether an article is present in a database may not translate to being found by a search in that database. Because of this major limitation, the question of which databases are necessary to retrieve all relevant references for a systematic review remains unanswered. Therefore, we research the probability that single or various combinations of databases retrieve most relevant references in a systematic review by studying actual retrieval in various databases.

The aim of our research is to determine the combination of databases needed for systematic review searches to provide efficient results (i.e. to minimize the burden for the investigators without reducing the validity of the research by missing relevant references). A secondary aim is to investigate the current practice of databases searched for published reviews. Are included references being missed because the review authors failed to search a certain database?

10.2 Methods

Development of search strategies

At Erasmus MC, search strategies for systematic reviews are often designed via a librarian-mediated search service. The information specialists of Erasmus MC developed an efficient method that helps them perform searches in many databases in a much shorter time than other methods. This method of literature searching and a pragmatic evaluation thereof are published in separate journal articles (21, 22). In short, the method consists of an efficient way to combine thesaurus terms and title/abstract terms into a single line search strategy. This search is then optimized. Articles that are indexed with a set of identified thesaurus terms, but do not contain the current search terms in title or abstract, are screened to discover potential new terms. New candidate terms are added to the basic search and evaluated. Once optimal recall is achieved, macros are used to translate the search syntaxes between databases, though manual adaptation of the thesaurus terms is still necessary.

Review projects at Erasmus MC cover a wide range of medical topics, from therapeutic effectiveness and diagnostic accuracy to ethics and public health. In general, searches are developed in MEDLINE in Ovid (Ovid MEDLINE® In-Process & Other Non-Indexed Citations, Ovid MEDLINE® Daily and Ovid MEDLINE®, from 1946); Embase.com, (searching both Embase and MEDLINE records, with full coverage including Embase Classic); the Cochrane Central Register of Controlled Trials (CENTRAL) via the Wiley Interface; Web of Science Core Collection (hereafter called Web of Science); PubMed restricting to records in the subset "as supplied by publisher" to find references that not yet indexed in MEDLINE (using the syntax publisher(sb)); and Google Scholar. In general, we use the first 200 references as sorted in the relevance ranking of Google Scholar. When the number of references from other databases was low, we expected the total number of potential relevant references to be low. In this case, the number of hits from Google Scholar was limited to 100. When the overall number of hits was low, we additionally searched Scopus, and when appropriate for the topic, we included CINAHL (EBSCOhost), PsycINFO (Ovid) and SportDiscus (EBSCOhost) in our search.

Beginning in May 2013, the number of records retrieved from each search for each database was recorded at the moment of searching. The complete results from all databases used for each of the systematic reviews were imported into a unique EndNote library upon search completion and saved without deduplication for this research. The researchers that requested the search received a de-duplicated EndNote file from which they selected the references relevant for inclusion in their systematic review. All searches in this study were developed and executed by W.M.B.

Determining relevant references of published reviews

We searched PubMed in July 2016 for all reviews published since 2014 where first authors were affiliated to Erasmus MC, Rotterdam, the Netherlands and matched those with search registrations performed by the medical library of Erasmus MC. This search was used in earlier research (21). Published reviews were included if the search strategies and results had been documented at the time of the last update, and if, at minimum, the databases Embase, MEDLINE, Cochrane CENTRAL, Web of Science, and Google Scholar had been used in the review. From the published journal article, we extracted the list of final included references. We documented the department of the first author. To categorize the types of patient/population and intervention, we identified broad MeSH terms relating to the most important disease and intervention discussed in the article. We copied from the MeSH tree the top MeSH term directly below the disease category, or, in to case of the intervention, directly below the therapeutics MeSH term. We selected the domain from a pre-defined set of broad domains, including therapy, etiology, epidemiology, diagnosis, management, and prognosis. Lastly, we checked whether the reviews described limiting their included references to a particular study design.

To identify whether our searches had found the included references, and if so, from which database(s) that citation was retrieved, each included reference was located in the original corresponding EndNote library using the first author name combined with the publication year as a search term for each specific relevant publication. If this resulted in extraneous results, the search was subsequently limited using a distinct part of the title or a second author name. Based on the record numbers of the search results in EndNote, we determined from which database these references came. If an included reference was not found in the EndNote file, we presumed the authors used an alternative method of identifying the reference (e.g., examining cited references, contacting prominent authors, or searching grey literature), and we did not include it in our analysis.

Data analysis

We determined the databases that contributed most to the reviews by the number of unique references retrieved by each database used in the reviews. Unique references were included articles that had been found by only one database search. Those databases that contributed most unique included references were then considered candidate databases to determine the most optimal combination of databases in the further analyses.

In Excel, we calculated the performance of each individual database and various combinations. Performance was measured using recall, precision, and number needed to read. See Table 10.1 for definitions of these measures. These values were calculated both for all reviews combined and per individual review.

Table 10.1: Definitions of general measures of performance in searches

Recall	# included references retrieved by a database/combination# included references retrieved by all databases
Precision	# included references retrieved by a database/combination# total references retrieved by those database(s)
Number Needed to Read	# total references retrieved by a database/combination# included references retrieved by those database(s)

Performance of a search can be expressed in different ways. Depending on the goal of the search, different measures may be optimized. In the case of a clinical question, precision is most important, as a practicing clinician does not have a lot of time to read through many articles in a clinical setting. When searching for a systematic review, recall is the most important aspect, as the researcher does not want to miss any relevant references. As our research is performed on systematic reviews, the main performance measure is recall.

We identified all included references that were uniquely identified by a single database. For the databases that retrieved the most unique included references, we calculated the number of references retrieved (after deduplication) and the number of included references that had been retrieved by all possible combinations of these databases, in total and per review. For all individual reviews, we determined the median recall, the minimum recall, and the percentage of reviews for which each single database or combination retrieved 100% recall.

For each review that we investigated, we determined what the recall was for all possible different database combinations of the most important databases. Based on these we determined the percentage of reviews where that database combination had achieved 100% recall, more than 95%, more than 90%, and more than 80%. Based on the number of results per database both before and after deduplication as recorded at the time of searching, we calculated the ratio between the total number of results and the number of results for each database and combination.

Improvement of precision was calculated as the ratio between the original precision from the searches in all databases and the precision for each database and combination.

To compare our practice of database usage in systematic reviews against current practice as evidenced in the literature, we analyzed a set of 200 recent systematic reviews from PubMed. On 5 January 2017, we searched PubMed for articles with the phrase “systematic review” in the title. Starting with the most recent articles, we determined the databases searched either from the abstract or from the full text until we had data for 200 reviews. For the individual databases and combinations that were used in those reviews, we multiplied the frequency of occurrence in that set of 200 with the probability that the database or combination would lead to an acceptable recall (which we defined at 95%) that we had measured in our own data.

10.3 Results

Our earlier research had resulted in 206 systematic reviews published between 2014 and July 2016, in which the first author was affiliated with Erasmus MC(21). In 73 of these, the searches and results had been documented by the first author of this article at the time of the last search. Of those, 15 could not be included in this research, since they had not searched all databases we investigated here. Therefore, for this research, a total of 58 systematic reviews were analyzed. The references to these reviews can be found in the supplementary material. An overview of the broad topical categories covered in these reviews is given in Table 10.2. Many of the reviews were initiated by members of the departments of surgery and epidemiology. The reviews covered a wide variety of disease, none of which was present in more than 12% of the reviews. The interventions were mostly from the chemicals and drugs category, or surgical procedures. Over a third of the reviews were therapeutic, while slightly under a quarter answered an etiological question. Most reviews did not limit to certain study designs, 9% limited to RCTs only, and another 9% limited to other study types.

Together, these reviews included a total of 1830 references. Of these, 84 references (4.6%) had not been retrieved by our database searches and were not included in our analysis, leaving in total 1746 references. In our analyses, we combined the results from MEDLINE in Ovid and PubMed (the subset as supplied by publisher) into one database labelled MEDLINE.

Table 10.2: Description of topics of included references (only values above 5% are shown)

Department (N=55)	
Surgery	13 (24%)
Epidemiology	10 (18%)
Internal Medicine	3 (5%)
Orthopedics	3 (5%)
Patient (N=52)	
Neoplasms	6 (12%)
Wounds and Injuries	6 (12%)
Musculoskeletal Diseases	5 (10%)
Cardiovascular Diseases	5 (10%)
Nutritional and Metabolic Diseases	5 (10%)
Intervention (N=31)	
Chemicals and Drugs Category	12 (39%)
Surgical Procedures, Operative	8 (26%)
Food and Beverages	2 (6%)
Biological Factors	2 (6%)
Domain (N=54)	
Therapy	19 (35%)
Etiology	13 (24%)
Epidemiology	6 (11%)
Diagnosis	6 (11%)
Management	5 (9%)
Prognosis	5 (9%)
Study Types (N=58)	
No limits mentioned	48 (83%)
RCTs	5 (9%)
RCTs and cohort studies / case control studies	5 (9%)

Unique references per database

A total of 292 (17%) references were found by only one database. Table 10.3 displays the number of unique results retrieved for each single database. Embase retrieved the most unique included references, followed by MEDLINE, Web of Science, and Google Scholar. Cochrane CENTRAL is absent from the table, as for the five reviews limited to randomized trials, it did not add any unique included references. Subject-specific databases such as CINAHL, PsycINFO and SportDiscus only retrieved additional included references when the topic of the review was directly related to their special content, respectively nursing, psychiatry and sports medicine.

Table 10.3: Number of unique included references by each specific database

Database	Number of reviews that used the database	Number of reviews with unique references	Number of unique references
Embase	58	29 (50%)	132 (45%)
MEDLINE	58	27 (47%)	69 (24%)
Web of Science	58	19 (33%)	37 (13%)
Google Scholar	58	24 (41%)	37 (13%)
CINAHL	18	1 (6%)	6 (2%)
Scopus	24	3 (13%)	5 (2%)
PsycINFO	11	1 (9%)	2 (1%)
SportDiscus	2	2 (100%)	3 (1%)

Overall performance

The four databases that had retrieved the most unique references (Embase, MEDLINE, Web of Science, and Google Scholar) were investigated individually and in all possible combinations (see Table 10.4). Of the individual databases, Embase had the highest overall recall (85.9%). Of the combinations of two databases, Embase and MEDLINE had the best results (92.8%). Embase and MEDLINE combined with either Google Scholar or Web of Science scored similarly well on overall recall (95.9%). However, the combination with Google Scholar had a higher precision and higher median recall, a higher minimum recall, and a higher proportion of reviews that retrieved all included references. Using both Web of Science and Google Scholar in addition to MEDLINE and Embase increased the overall recall to 98.3%. The higher recall from adding extra databases came at a cost in number needed to read (NNR). Searching only Embase produced an NNR of 57 on average, whereas, for the optimal combination of four databases, the NNR was 73.

Table 10.4: Performance of several databases and database combinations in terms of sensitivity and precision20

	# results	# includes (N=1746)	overall recall[1]	median recall[2]	minimum recall[3]	percentage 100% recall [4]	precision [5]	Number Needed to Read [6]
Embase (EM)	85521	1500	85.9%	87.3%	45.8%	13.8%	1.8%	57
MEDLINE (ML)	56340	1375	78.8%	82.9%	50.0%	8.6%	2.4%	41
Web of Science (WoS)	48561	1189	68.1%	72.5%	13.2%	6.9%	2.4%	41
Google Scholar (GS)	10342	601	34.4%	38.0%	8.3%	5.2%	5.8%	17
EM-ML	100444	1621	92.8%	94.6%	66.7%	24.1%	1.6%	62
EM-WoS	104444	1585	90.8%	93.8%	57.9%	27.6%	1.5%	66
EM-GS	91411	1570	89.9%	93.3%	65.8%	25.9%	1.7%	58
ML-WoS	75263	1481	84.8%	87.1%	60.0%	15.5%	2.0%	51
ML-GS	62230	1459	83.6%	89.8%	63.2%	15.5%	2.3%	43
WoS-GS	54451	1320	75.6%	85.7%	23.7%	13.8%	2.4%	41
EM-ML-GS	106334	1674	95.9%	97.4%	78.9%	41.4%	1.6%	64
EM-ML-WoS	119367	1674	95.9%	97.1%	71.1%	37.9%	1.4%	70
EM-WoS-GS	110334	1638	93.8%	98.1%	65.8%	44.8%	1.5%	67
ML-WoS-GS	81153	1528	87.5%	92.6%	70.0%	29.3%	1.9%	53
EM-ML-GS-WoS	125257	1716	98.3%	100.0%	78.9%	74.1%	1.4%	73

[1] Overall recall: The total number of included references retrieved by the databases divided by the total number of included references retrieved by all databases.

[2] Median recall: The median value of recall per review

[3] Minimum recall: The lowest value of recall per review

[4] Percentage 100% recall: The percentage of reviews for which the database combination retrieved all included references

[5] Precision: The number of included references divided by the total number of references retrieved

[6] Number Needed to Read: The total number of references retrieved divided by the number of included references

Probability of appropriate recall

We calculated the recall for individual databases and databases in all possible combination for all reviews included in the research. Figure 10.1 shows the percentages of reviews where a certain database combination led to a certain recall. For example, in 48% of all systematic reviews, the combination of Embase and MEDLINE (with or without Cochrane CENTRAL; Cochrane CENTRAL did not add unique relevant references) reaches a recall of at least 95%. In 72% of studied systematic reviews, the combination of Embase, MEDLINE, Web of Science, and Google Scholar retrieved all included references. In the top bar, we present the results of the complete database searches relative to the total number of included references. This shows that many database searches missed relevant references.

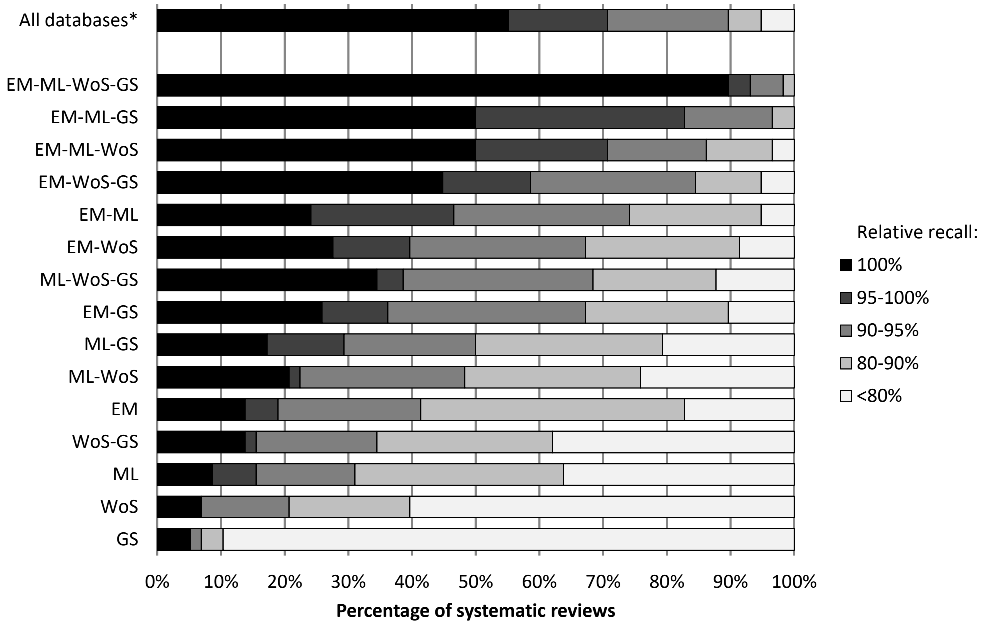


Figure 10.1 : Percentage of systematic reviews for which a certain database combination reached a certain recall. The X-axis represents the percentage of reviews for which a specific combination of databases, as shown on the y-axis, reached a certain recall (represented with bar colors). Abbreviations: EM Embase, ML MEDLINE, WoS Web of Science, GS Google Scholar. Asterisk indicates that the recall of all databases has been calculated over all included references. The recall of the database combinations was calculated over all included references retrieved by any database

Differences between domains of reviews

We analyzed whether the added value of Web of Science and Google Scholar was dependent of the domain of the review. For 55 reviews, we determined the domain. See Figure 10.2 for the comparison of the recall of Embase, MEDLINE, and Cochrane CENTRAL per review for all identified domains. For all but one domain, the traditional combination of Embase, MEDLINE, and Cochrane CENTRAL did not retrieve enough included references. For 4 out of five systematic reviews that limited to randomized controlled trials (RCTs) only, the traditional combination retrieved 100% of all included references. However, for one review of this domain, the recall was 82%. Of the 11 references included in this review, one was found only in Google Scholar, one only in Web of Science.

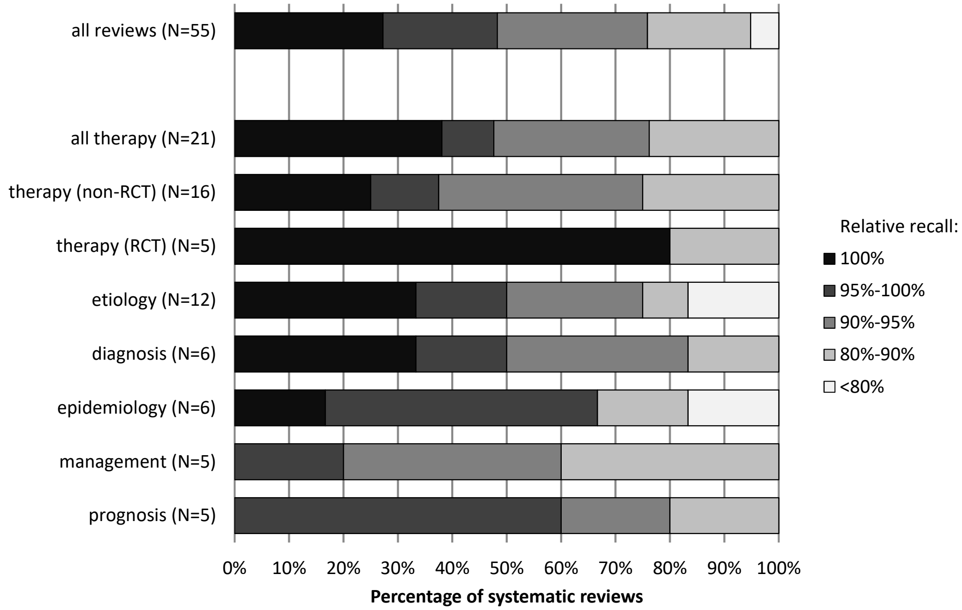


Figure 10.2: Percentage of systematic reviews of a certain domain for which the combination Embase, MEDLINE and Cochrane CENTRAL reached a certain recall

Reduction in number of results

We calculated the ratio between the number of results found when searching all databases, including databases not included in our analyses, such as Scopus, PsycINFO, and CINAHL, and the number of results found searching a selection of databases. See Figure 10.3 for the legend of the plots in Figure 10.4 and 10.5. Figure 10.4 shows the calaculated ratio between the number of results for individual reviews. The database combinations with the highest recall did not reduce the total number of results by large margins. Moreover, in combinations where the number of results was greatly reduced, the recall of included references was lower.

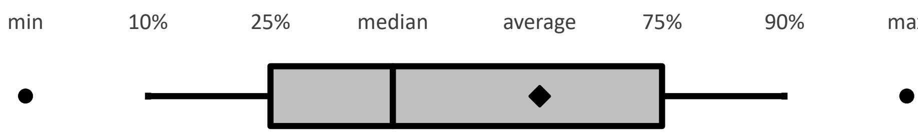


Figure 10.3: Legend of Figs.10. 4 and 10. 5

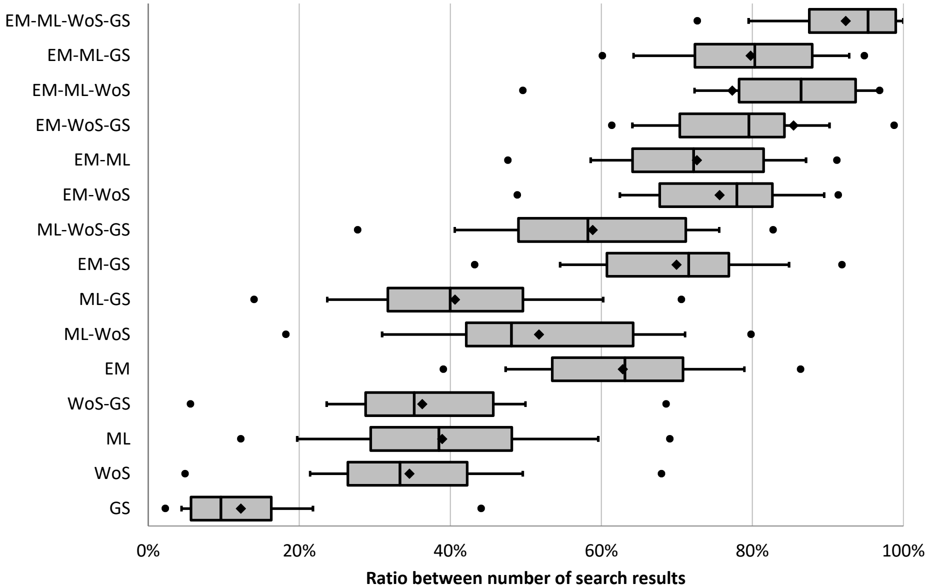


Figure 10.4: The ratio between number of results per database combination and the total number of results for all databases

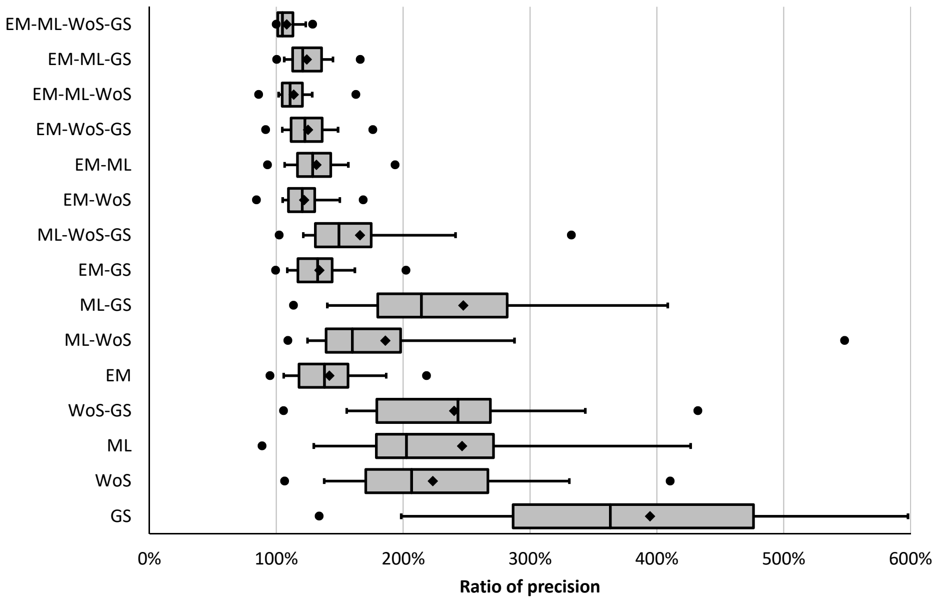


Figure 10.5: The ratio between precision per database combination and the total precision for all databases

Improvement of precision

To determine how searching multiple databases affected precision, we calculated for each combination the ratio between the original precision, observed when all databases were searched, and the precision calculated for different database combinations. Figure 10.5 shows the improvement of precision for 15 databases and database combinations. Because precision is defined as the number of relevant references divided by the number of total results, we see a strong correlation with the total number of results.

Status of current practice of database selection

From a set of 200 recent SRs identified via PubMed, we analyzed the databases that had been searched. Almost all reviews (97%) reported a search in MEDLINE. Other databases that we identified as essential for good recall were searched much less frequently; Embase was searched in 61%, Web of Science in 35%, and Google Scholar was only used in 10% of all reviews. For all individual databases or combinations of the four important databases from our research (MEDLINE, Embase, Web of Science and Google Scholar), we multiplied the frequency of occurrence of that combination in the random set, with the probability we found in our research that this combination would lead to an acceptable recall of 95%. The calculation is shown in Table 10.5. For example, around a third of the reviews (37%) relied on the combination of MEDLINE and Embase. Based on our findings, this combination achieves acceptable recall about half the time (47%). This implies that 17% of the reviews in the PubMed sample would have achieved an acceptable recall of 95%. The sum of all these values is the total probability of acceptable recall in the random sample. Based on these calculations, we estimate that the probability that this random set of reviews retrieved more than 95% of all possible included references was 40%. Using similar calculations, also shown in Table 10.5, we estimated the probability that 100% of relevant references were retrieved is 23%.

Table 10.5. Calculation of probability of acceptable recall of a PubMed sample of systematic reviews

	Frequency	Frequency percentage (a)	Probability recall > 95% (b)	a x b	Probability recall 100% (c)	c x b
EM-ML	73	37%	47%	17%	24%	9%
ML	41	21%	16%	3%	9%	2%
EM-ML-WoS	40	20%	64%	13%	36%	7%
ML-WoS	21	11%	21%	2%	16%	2%
ML-GS	7	4%	26%	1%	16%	1%
ML-WoS-GS	7	4%	37%	1%	29%	1%
EM-ML-GS	5	3%	76%	2%	41%	1%
EM	2	1%	19%	0%	14%	0%
EM-WoS	1	1%	40%	0%	28%	0%
WoS	1	1%	7%	0%	7%	0%
Total	**198***	**100%**		**40%**		**23%**

* Two reviews did not use any of the databases used in this evaluation.

10.4 Discussion

Our study shows that, to reach maximum recall, searches in systematic reviews ought to include a combination of databases. To ensure adequate performance in searches (i.e. recall, precision, and number needed to read), we find that literature searches for a systematic review should, at minimum, be performed in the combination of the following 4 databases: Embase, MEDLINE (including Epub ahead of print), Web of Science Core Collection, and Google Scholar. Using that combination, 93% of the systematic reviews in our study obtained levels of recall that could be considered acceptable (>95%). Unique results from specialized databases that closely match systematic review topics, such as PsycINFO for reviews in the fields of behavioral sciences and mental health or CINAHL for reviews on the topics of nursing or allied health, indicate that specialized databases should be used additionally when appropriate.

We find that Embase is critical for acceptable recall in a review and should always be searched for medically-oriented systematic reviews. However, Embase is only accessible via a paid subscription, which generally makes it challenging for review teams not affiliated with academic medical centers to access. The highest scoring database combination without Embase is a combination of MEDLINE, Web of Science, and Google Scholar, but that reaches satisfactory recall for only 39% of all investigated systematic reviews, while still requiring a paid subscription to Web of Science. Of the five reviews that included only RCTs, four reached 100% recall if MEDLINE, Web of Science, and Google Scholar combined were complemented with Cochrane CENTRAL.

The Cochrane Handbook recommends searching MEDLINE, Cochrane CENTRAL, and Embase for systematic reviews of RCTs. For reviews in our study that included RCTs only, indeed, this recommendation was sufficient for four (80%) of the reviews. The one review where it was insufficient was about alternative medicine, specifically meditation and relaxation therapy, where one of the missed studies was published in the *Indian Journal of Positive Psychology*. The other study from the *Journal of Advanced Nursing* is indexed in MEDLINE and Embase, but was only retrieved because of the addition of KeyWords Plus in Web of Science. We estimate more than 50% of reviews that include more study types than RCTs would miss more than 5% of included references if only traditional combination of MEDLINE, Embase, and Cochrane CENTAL is searched.

10

We are aware that the Cochrane Handbook (7) recommends more than only these databases, but further recommendations focus on regional and specialized databases. Though we occasionally used the regional databases LILACS and SciELO in our reviews, they did not provide unique references in our study. Subject-specific databases like PsycINFO only added unique references to a small percentage of systematic reviews when they had been used for the search. The third key database we identified in this research, Web of Science, is only mentioned as a citation index in the Cochrane Handbook, not as a bibliographic database. To our surprise, Cochrane CENTRAL did not identify any unique included studies that had not been retrieved by the other databases, not even for the five reviews focusing entirely on RCTs. If Erasmus MC authors had conducted more reviews that included only RCTs, Cochrane CENTRAL might have added more unique references.

MEDLINE did find unique references that had not been found in Embase, although our searches in Embase included all MEDLINE records. It is likely caused by difference in thesaurus terms that were added, but further analysis would be required to determine reasons for not finding the MEDLINE records in Embase. Although Embase covers MEDLINE, it apparently does not index every article from MEDLINE. Thirty-seven references were found in MEDLINE (Ovid), but were not available in Embase.com. These are mostly unique PubMed references, which are not assigned MeSH terms, and are often freely available via PubMed Central.

Google Scholar adds relevant articles not found in the other databases, possibly because it indexes the full text of all articles. It therefore finds articles in which the topic of research is not mentioned in title, abstract, or thesaurus terms, but where the concepts are only discussed in the full text. Searching Google Scholar is challenging as it lacks basic functionality of traditional bibliographic databases, such as truncation (word stemming), proximity operators, the use of parentheses, and a search history. Additionally, search strategies are limited to a maximum of 256 characters, which means that creating a thorough search strategy can be laborious.

Whether Embase and Web of Science can be replaced by Scopus remains uncertain. We have not yet gathered enough data to be able to make a full comparison between Embase and Scopus. In 23 reviews included in this research, Scopus was searched. In 12 reviews (52%), Scopus retrieved 100% of all included references retrieved by Embase or Web of Science. In the other 48%, the recall by Scopus was suboptimal, in one occasion as low as 38%.

Of all reviews in which we searched CINAHL and PsycINFO, respectively for 6% and 9% of the reviews, unique references were found. For CINAHL and PsycINFO, in one case each, unique relevant references were found. In both these reviews the topic was highly related to the topic of the database. Although we did not use these special topic databases in all of our reviews, given the low number of reviews where these databases added relevant references, and observing the special topics of those reviews, we suggest that these subject databases will only add value if the topic is related to the topic of the database.

Many articles written on this topic have calculated overall recall of several reviews, instead of the effects on all individual reviews. Researchers planning a systematic review generally perform one review, and they need to estimate the probability that they may miss relevant articles in their search. When looking at the overall recall, the combination of Embase and MEDLINE and either Google Scholar or Web of Science could be regarded sufficient with 96% recall. This number however is not an answer to the question of a researcher performing a systematic review, regarding which databases should be searched. A researcher wants to be able to estimate the chances that his or her current project will miss a relevant reference. However, when looking at individual reviews, the probability of missing more than 5% of included references found through database searching is 33% when Google Scholar is used together with Embase and MEDLINE and 30% for the Web of Science, Embase MEDLINE combination. What is considered acceptable recall for systematic review searches is open for debate and can differ between individuals and groups. Some reviewers might accept a potential loss of 5% of relevant references; others would want to pursue 100% recall, no matter what cost. Using the results in this research, review teams can decide, based on their idea of acceptable recall and the desired probability which databases to include in their searches.

Strengths and limitations

We did not investigate whether the loss of certain references had resulted in changes to the conclusion of the reviews. Of course, the loss of a minor non-randomized included study that follows the systematic review's conclusions would not be as problematic as losing a major included randomized controlled trial with contradictory results. However, the wide range of scope, topic, and criteria between systematic reviews and their related review types make it very hard to answer this question.

We found that two databases previously not recommended as essential for systematic review searching, Web of Science and Google Scholar, were key to improving recall in the reviews we investigated. Because this is a novel finding, we cannot conclude whether it is due to our dataset or to a generalizable principle. It is likely that topical differences in systematic reviews may impact whether databases such as Web of Science and Google Scholar add value to the review. One explanation for our finding may be that if the research question is very specific, the topic of research might not always be mentioned in title and or abstract. In that case, Google Scholar might add value by searching the full text of articles. If the research question is more interdisciplinary, a broader science database such as Web of Science is likely to add value. The topics of the reviews studied here may simply have fallen into those categories, though the diversity of the included reviews may point to a more universal applicability.

Although we searched PubMed as supplied by publisher separately from MEDLINE in Ovid, we combined the included references of these databases into one measurement in our analysis. Until 2016, the most complete MEDLINE selection in Ovid still lacked the electronic publications that were already available in PubMed. These could be retrieved by searching PubMed with the subset as supplied by publisher. Since the introduction of the more complete MEDLINE collection *Epub Ahead of Print, In-Process & Other Non-Indexed Citations and Ovid MEDLINE®*, the need to separately search PubMed as supplied by publisher has disappeared. According to our data, PubMed's "as supplied by publisher" subset retrieved 12 unique included references, and it was the most important addition in terms of relevant references to the four major databases.

It is therefore important to search MEDLINE including the “Epub Ahead of Print, In-Process, and Other Non-Indexed Citations” references.
These results may not be generalizable to other studies for other reasons. The skills and experience of the searcher is one of the most important aspects in the effectiveness of systematic review search strategies.(23-25) The searcher in the case of all 58 systematic reviews, is an experienced biomedical information specialist. Though we suspect that searchers who are not information specialists or librarians would have a higher possibility of less well-constructed searches and searches with lower recall, even highly trained searchers differ in their approaches to searching. For this study, we searched to achieve as high a recall as possible, though our search strategies, like any other search strategy, still missed some relevant references because relevant terms had not been used in the search. We are not implying that a combined search of the four recommended databases will never result in relevant references being missed, rather that failure to search any one of these four databases will likely lead to relevant references being missed. Our experience in this study shows that additional efforts, such as hand searching, reference checking and contacting key players, should be made to retrieve extra possible includes.

Based on our calculations made by looking at random systematic reviews in PubMed, we estimate that 60% of these reviews are likely to have missed more than 5% of relevant references only because of the combinations of databases that were used. That is with the generous assumption that the searches in those databases had been designed sensitively enough. Even when taking into account that many searchers consider the use of Scopus as a replacement of Embase, plus taking into account the large overlap of Scopus and Web of Science, this estimate remains similar. Also, while the Scopus and Web of Science assumptions we made might be true for coverage, they are likely very different when looking at recall, as Scopus does not allow the use of the full features of a thesaurus. We see that reviewers rarely use Web of Science and especially Google Scholar in their searches, though they retrieve a great deal of unique references in our reviews. Systematic review searchers should consider using these databases if they are available to them, and if their institution lacks availability, they should ask other institutes to cooperate on their systematic review searches.

The major strength of our paper is that it is the first large-scale study we know of to assess database performance for systematic reviews using prospectively collected data. Prior research on database importance for systematic reviews has looked primarily at whether included references could have theoretically been found in a certain database, but most have been unable to ascertain whether the researchers actually found the articles in those databases.(10, 12, 16, 17, 26) Whether a reference is available in a database is important, but whether the article can be found in a precise search with reasonable recall is not only impacted by the database’s coverage. Our experience has shown us that it is also impacted by the ability of the searcher, the accuracy of indexing of the database, and the complexity of terminology in a particular field. Because these studies based on retrospective analysis of database coverage do not account for the searchers’ abilities, the actual findings from the searches performed, and the indexing for particular articles, their conclusions lack immediate translatability into practice. This research goes beyond retrospectively assessed coverage to investigate real search performance in databases. Many of the articles reporting on previous research concluded that one database was able to retrieve most included references. Halladay et al. (10) and van Enst et al. (16) concluded that databases other than MEDLINE/PubMed did not change the outcomes of the review, while Rice et al. (17) found the added value of other databases only for newer, non-indexed references. In addition, Michaleff et al. (26) found that Cochrane CENTRAL included 95% of all RCTs included in the reviews investigated. Our conclusion that Web of Science and Google Scholar are needed for completeness has not been shared by previous research. Most of the previous studies did not include these two databases in their research.

10.5 Conclusions

We recommend that, regardless of their topic, searches for biomedical systematic reviews should combine Embase, MEDLINE (including electronic publications ahead of print), Web of Science (Core Collection) and Google Scholar (the 200 first relevant references) at minimum. Special topics databases such as CINAHL and PsycINFO should be added if the topic of the review directly touches the primary focus of a specialized subject database, like CINAHL for focus on nursing and allied health or PsycINFO for behavioral sciences and mental health. For reviews where RCTs are the desired study design, Cochrane CENTRAL may be similarly useful. Ignoring one or more of the databases that we identified as the four key databases will result in more precise searches with a lower number of results, but the researchers should decide whether that is worth the increased probability of losing relevant references. This study also highlights once more that searching databases alone is, nevertheless, not enough to retrieve all relevant references.

Future research should continue to investigate recall of actual searches beyond coverage of databases and should consider focusing on the most optimal database combinations, not on single databases.

10.6 References

1. Levay P, Raynor M, Tuvey D: The Contributions of MEDLINE, Other Bibliographic Databases and Various Search Techniques to NICE Public Health Guidance. Evidence Based Library and Information Practice 2015, 10:50-68.
2. Stevinson C, Lawlor DA: Searching multiple databases for systematic reviews: Added value or diminishing returns? Complement Ther Med 2004, 12:228-232.
3. Lawrence DW: What is lost when searching only one literature database for articles relevant to injury prevention and safety promotion? Inj Prev 2008, 14:401-404.
4. 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, 58:867-873.
5. Zheng MH, Zhang X, Ye Q, Chen YP: Searching additional databases except PubMed are necessary for a systematic review. Stroke 2008, 39:e139; author reply e140.
6. 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, 30:49-58.
7. Higgins JPT, Green S: Cochrane handbook for systematic reviews of interventions. Wiley Online Library; 2008.
8. Wright K, Golder S, Lewis-Light K: What value is the CINAHL database when searching for systematic reviews of qualitative studies? Syst Rev 2015, 4:104.
9. Wilkins T, Gillies RA, Davies K: EMBASE versus MEDLINE for family medicine searches: Can MEDLINE searches find the forest or a tree? Can Fam Physician 2005, 51:848-849.

10. 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, 68:1076-1084.
11. Ahmadi M, Ershad-Sarabi R, Jamshidiorak R, Bahaodini K: Comparison of bibliographic databases in retrieving information on telemedicine. Journal of Kerman University of Medical Sciences 2014, 21:343-354.
12. Lorenzetti DL, Topfer L-A, Dennett L, Clement F: Value of databases other than medline for rapid health technology assessments. Int J Technol Assess Health Care 2014, 30:173-178.
13. Beckles Z, Glover S, Ashe J, Stockton S, Boynton J, Lai R, Alderson P: Searching CINAHL did not add value to clinical questions posed in NICE guidelines. J Clin Epidemiol 2013, 66:1051-1057.
14. 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, 16:1-13.
15. 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, 16:161.
16. 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, 67:1192-1199.
17. 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.
18. Bramer WM, Giustini D, Kramer BM, Anderson PF: 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.
19. Bramer WM, Giustini D, Kramer BMR: Comparing the coverage, recall, and precision of searches for 120 systematic reviews in Embase, MEDLINE, and Google Scholar: a prospective study. Syst Rev 2016, 5:39.
20. Ross-White A, Godfrey C: Is there an optimum number needed to retrieve to justify inclusion of a database in a systematic review search? Health Info Libr J 2017.
21. Bramer WM, Rethlefsen ML, Mast F, Kleijnen J: A pragmatic evaluation of a new method for librarian-mediated literature searches for systematic reviews. submitted 2017.
22. Bramer WM, de Jonge GB, Rethlefsen ML, Mast F, Kleijnen J: A systematic approach to searching: How to perform high quality literature searches more efficiently. submitted 2017.
23. Rethlefsen ML, Farrell AM, Osterhaus Trzasko LC, Brigham TJ: Librarian co-authors correlated with higher quality reported search strategies in general internal medicine systematic reviews. J Clin Epidemiol 2015, 68:617-626.
24. McGowan J, Sampson M: Systematic reviews need systematic searchers. J Med Libr Assoc 2005, 93:74-80.
25. McKibbon KA, Haynes RB, Dilks CJW, Ramsden MF, Ryan NC, Baker L, Flemming T, Fitzgerald D: How good are clinical MEDLINE searches? A comparative study of clinical end-user and librarian searches. Comput Biomed Res 1990, 23:583-593.
26. Michaleff ZA, Costa LO, Moseley AM, Maher CG, Elkins MR, Herbert RD, Sherrington C: CENTRAL, PEDro, PubMed, and EMBASE are the most comprehensive databases indexing randomized controlled trials of physical therapy interventions. Phys Ther 2011, 91:190-197.

Chapter 11

Searching using only major descriptors or title and abstract

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.

11 Searching Embase and MEDLINE by using only major descriptors or title and abstract fields: a prospective exploratory study

Abstract

Background

Researchers performing systematic reviews (SRs) must carefully consider the relevance of thousands of citations retrieved from bibliographic database searches, the majority of which will be excluded later on close inspection. Well-developed bibliographic searches are generally created with thesaurus or index terms in combination with keywords found in the title and/or abstract fields of citation records. Records in the bibliographic database Embase contain many more thesaurus terms than MEDLINE. Here we aim to examine how limiting searches to major thesaurus terms (in MEDLINE called focus 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 SR searches.

Methods

To examine the impact of using search techniques aimed at higher precision, we analyzed previously-completed SRs and focused our original searches to major thesaurus terms or terms in title and/or abstract only in Embase.com or in Embase.com and MEDLINE (Ovid) combined. We examined the total number of search results in both Embase and MEDLINE, and checked whether included references were retrieved by these more focused approaches.

Results

For 73 SRs, we limited Embase searches to major terms only while keeping the search in Medline and other databases such as Web of Science as they were. The overall search yield (or total number of search results) was reduced by 8%. Six reviews (9%) lost more than 5% of the relevant references. Limiting Embase and MEDLINE to major the number of references was 13% lower. For 15% of the reviews the loss of relevant references was more than 5%. Searching Embase for title and abstract caused a loss of more than 5% in 16 reviews (22%), while limiting Embase and MEDLINE that way this happened in 24 reviews (33%).

Conclusions

Of the four search options, two options substantially reduced the overall search yield. However this also resulted in a greater chance of losing relevant references, even though many references were still found in other databases such as Web of Science.

11.1 Background

Performing high-quality systematic reviews (SRs) is an exacting and time-consuming process because biomedical researchers are required to review thousands of titles and abstracts from their searches in traditional bibliographic databases. Ultimately, the percentage of references selected for inclusion in a SR is around 2% (1). The time and resources needed for selection and screening of papers can be considerably reduced by reducing the number of papers found by a more targeted search. Still, targeted searching and its benefits should be evaluated against the likelihood of missed relevant studies.

Many peer-reviewed articles are retrieved by searching bibliographic databases such as MEDLINE and Embase (2-4). Including terms from a controlled vocabulary, or thesaurus such as Medical Subject Headings (MeSH) in MEDLINE and Excerpta Medica Thesaurus (Emtree terms) in Embase is critical in creating robust, sensitive searches. “A controlled vocabulary is an organized arrangement of words and phrases used to index content and/or to retrieve content through browsing or searching. It typically includes preferred and variant terms and has a defined scope or describes a specific domain.”(5) In the databases Embase and MEDLINE either indexers read articles and select predefined terms that closely describe their content, or this is done by an automated algorithm. A few of the most important assigned thesaurus terms (usually around 25%) are marked as major descriptors or focus concepts.(6) There are many ways to describe the phenomenon (such as major topic, major terms, focus terms, major headers) or clarity we will use the term major descriptor throughout the article for both databases.

When conducting SRs, searches are designed to be more sensitive in order to retrieve as many relevant papers as possible. Combining thesaurus terms searched in subject fields with words or key terms in the title and abstract fields are recommended for good recall.(7)

Embase records generally have more thesaurus terms attached to them than records in MEDLINE. Thus, searching with Embase (Emtree) terms can increase the likelihood of finding more references that might not necessarily be relevant.(8-10) The number of search results retrieved by a search strategy can be reduced by searching for index terms as major descriptors or by searching title and abstract fields only (11). It is unclear how focusing search strategies affects the total number of search results for SRs, and whether potentially relevant references would be missed.

Previous research on focusing search strategies is not clear about the impact on recall when using or targeted field search strategies (12, 13). In the existing literature, overall performance of searches is determined by examining the final set of relevant references. Data for individual SRs suggests that there are many differences in the performance of searches aimed at a higher precision. However, only Embase is usually taken into account (13). Still, in biomedical SRs, any articles missed in Embase may be found by other databases such as MEDLINE or Web of Science.

We aimed to determine whether focused Embase and MEDLINE searches negatively impacted the retrieval of relevant studies across a broad range of SRs.

11.2 Methods

This cross-sectional study is reported along the recommendation of the STROBE checklist (see supplementary materials). To examine more focused searches and their impact we looked at individual reviews, rather than on overall performance combining multiple reviews. We took into account the search results from additional databases, which might find articles not retrieved in Embase and MEDLINE.

Registration of systematic review searches

From May 2013 to present, we registered SR search strategies and documented results for researchers at Erasmus MC (Erasmus University Medical Centre, Rotterdam, the Netherlands). These data were tested in other research projects regarding the coverage and retrieval of certain bibliographic databases in supporting SRs (14). Once these searches were performed, the results were imported into EndNote (15) and saved before and after deduplication. Deduplicated EndNote files for each SR were assessed by the researchers who requested the search in order to determine the references most relevant to their specific research question, for inclusion in their review. EndNote files that had not been deduplicated were saved for later analysis.

Collecting Erasmus MC published reviews

We searched PubMed retrospectively for SRs published by researchers from Erasmus MC. We included published reviews where the Erasmus MC affiliation was listed for the first or corresponding author and for which we had documented the searches. In all these reviews we searched at least in Embase.com, Ovid MEDLINE and Web of Science and often additional databases (in particular the Cochrane Library and Google Scholar). All search strategies were designed by the same Erasmus MC librarian (Wichor Bramer) and all searches combined thesaurus terms and searches for words or phrases in titles and abstract fields in Embase and MEDLINE.

Of the SRs analyzed here, all references that were selected for inclusion were aggregated. All included references were searched one-by-one in the EndNote file containing the original downloaded search results from all databases for the related SR. By using record numbers of search results in EndNote, we determined which databases had retrieved the specific reference in question.

Adaptation of original search strategies

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. For searches where major thesaurus terms were used to limit retrieval, we searched non-major thesaurus terms for concepts related to study type and age groups such as children. These are so-called check tags, which are by default never marked as major terms. In the title and/or abstract only searches we removed all thesaurus terms, including the check tags for age groups and searched those elements of the search also as title and/or abstract only. In table 1 of the supplementary materials an example is given of adapted search strategies.

Characteristics of the included reviews

We analyzed review topics by using the MeSH tree structures in in the MeSH Database (on entrez)(16). For the most important disease aspect, we selected the broadest or highest level MeSH term below the *Diseases Category* in the tree. For the most important intervention we chose the MeSH term directly below the *Analytical, Diagnostic and Therapeutic Techniques and Equipment Category* tree. However, for terms designated under *Therapeutics*, we used those MeSH term at one level deeper. We also documented the domain of the review out of seven predefined domains. Reviews on the effectiveness of a treatment were designated *therapy*; policies were designated *management*. Research on incidence and prevalence were designated as *epidemiology*, and reviews on causes of diseases as *etiology*. Other domains we documented were *prognosis, diagnosis* and *prevention*.

Analysis of included references

Two sets of included references were created for each review: 1) a set with included references retrieved only by searching in Embase; and 2) a set of included references that were uniquely retrieved by Embase and/or MEDLINE. Included references that had been identified by other searches (such as Web of Science, Cochrane CENTRAL or handsearching) were unaffected as we did not change those searches. Specific references were searched in both databases by combining authors' names with distinct title words. In Embase, we combined all references uniquely retrieved by Embase into a single set, while in MEDLINE we did the same for all references uniquely retrieved through a combination of Embase and MEDLINE. We combined each modified focused search in Embase and MEDLINE (two per database) with the set of references found in those databases for each review.

We calculated the sensitivity of these more focused searches by dividing the number of references found (sum of the total unique included references found by the focused search and total number of included references found by other databases) by the total number of references included (Figure 11.1).

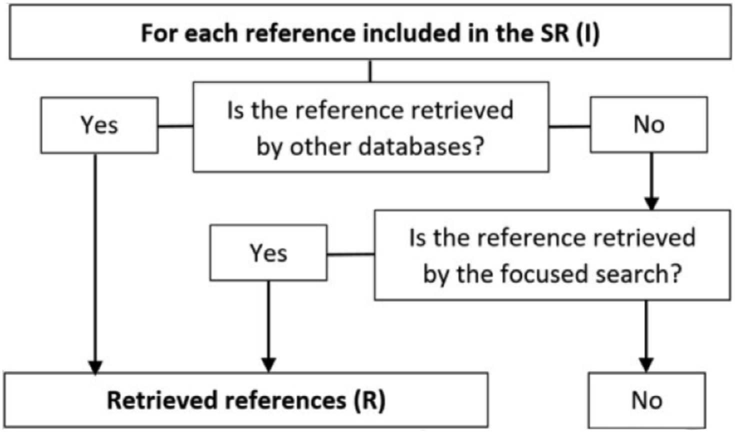


Figure. 11.1: Flow chart of references included and found by focused searches

We could not determine the exact number of references retrieved with focused search strategies because the focused searches were created retrospectively for reviewsthat had been published for which the searches had been executed months or years before this research project. We therefore estimated the relative reduction in overall number of search results based on the number of unique results from both databases, combined multiplied by the ratio in reduction of number of results by examining the focused searches.

Table 11.1: Abbreviations used in estimation of number of search results of a focused search for a systematic review

	Before deduplication	After deduplication
Embase.com	Eb	Ea
Medline Ovid	Mb	Ma
Web of Science	Wb	Wa
Other databases	Ob	Oa
Total	Tb	Ta

Deduplication for the reviews always done as follows, Embase records were imported first, followed by Medline and then other databases. The number of records from other databases never changes. The number of unique references in Embase after deduplication (Eu) is Ea – (Mb – Ma). The number of unique references in Medline after deduplication (Mu) is Ma – (Wb – Wa). We calculated the ratio that a focused search in Embase (rE) and Medline (rM) as the number retrieved with the focused search devided by the number of the original search. The total number of references after deduplication with a focused search in embase = Oa + Wa + Ma + rE*Ea, and for the focused search in Medline and Embase combined: Oa + Wa + rM*Ma + rE*Ea.

11.3 Results

We collated all included references from 73 published systematic reviews for a total of 2551 references. Of these included references, 163 (6.4%) were retrieved by Embase alone, and 369 references (14.5%) were not retrieved by any other database than Embase or MEDLINE. The topics of the included reviews are described in Table 11.2. By limiting Embase to major thesaurus terms, the overall number of search results for individual reviews was reduced by a median of 8%. Those searches retrieved a total of 2515 references (98.6%). Of the 163 references that had been retrieved uniquely by Embase, 127 references (77.9%) were found if the searches were limited to major thesaurus terms. In 57 of 73 reviews (78%), all included references would still have been found in Embase if searches had been limited to major terms only. In the supplementary materials an overview is given of the included reviews and the results.

Table 11.2: Description of reviews included in the research

Patient (N=62)	
Wounds and Injuries	7 (11%)
Cardiovascular Diseases	7 (11%)
Musculoskeletal Diseases	6 (10%)
Nutritional and Metabolic Diseases	5 (8%)
Neoplasms	5 (8%)
Pathological Conditions, Signs and Symptoms	4 (6%)
Female Urogenital Diseases and Pregnancy Complications	4 (6%)
Intervention (N=41)	
Chemicals and Drugs	12 (29%)
Operative Surgical Procedures	12 (29%)
Domain (N=68)	
etiology	17 (25%)
therapy (non-RCT)	13 (19%)
therapy (RCT)	10 (15%)
management	9 (13%)
epidemiology	7 (10%)
diagnosis	7 (10%)
prognosis	5 (7%)

By limiting Embase searching to title and abstract fields, the median reduction in number of search results per review was 11%, retrieving a total of 2472 references (96.9%). Of the 163 references that were uniquely retrieved by Embase, 84 (51.1%) were retrieved by the search in title and / or abstract only. In those searches, 50 reviews (68%) were unaffected in terms of lost included articles. When we limited both Embase and MEDLINE to major thesaurus terms median reduction in number of search results was 13% retrieving 2487 references (97.5%). Of the 369 included references that had not been retrieved by other databases than MEDLINE or

Embase, 305 (82.7%) were retrieved when both Embase and MEDLINE had been limited to major thesaurus terms only. In 46 reviews (63%) no included references were missed. Limiting both Embase and MEDLINE to title and abstract field searches reduced the number of search results by 20% and retrieved 2381 included references (93.3%). Of the 369 included references that had not been retrieved by databases beyond MEDLINE or Embase, 199 (53.9%) were retrieved when both the MEDLINE and Embase searches were limited to words in title and/or abstract only. In other words, 38 reviews (52%) would still have retrieved all included references using these more focused strategies.

In Figure 11.2 we show the percentage of reviews for which the focused searches reached a certain recall threshold. For instance, 91% of all reviews achieved 95% recall or greater when Embase searches were limited to major Emtree index terms. Similarly, 80% of all reviews had at least 90% recall when both Embase and MEDLINE were used to search titles and/or abstracts only.

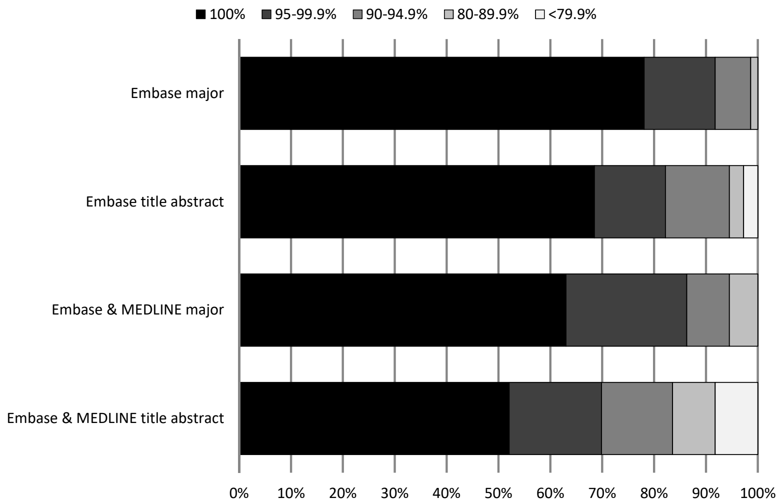


Figure 11.2: Percentage of reviews that reached a certain recall threshold using focused searches

Figure 11.3 shows the change in the total number of search results retrieved by the more focused searches. Limiting Embase to major thesaurus terms retrieved the highest number of search results and limiting both databases to titles and abstracts retrieved the lowest. If both Embase and MEDLINE were restricted to major thesaurus terms fewer search results were retrieved than by limiting only Embase to titles and abstracts.

11

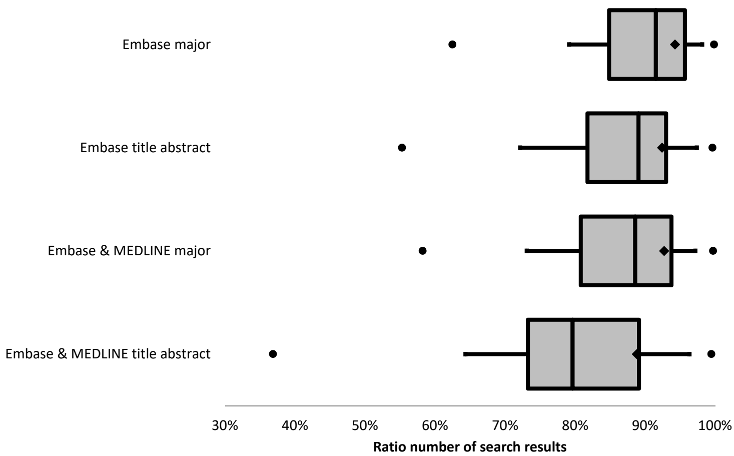


Figure 11.3: Ratio between total number of search results for systematic reviews for focused searches and the original unchanged searches. The figure shows from left to right: minimum, 10th percentile, 25th percentile, median average, 75th percentile, 90th percentile, and maximum

Since limiting Embase to major descriptors resulted in the most acceptable results we decided to investigate that search type further for different review types. For the review types of which we identified 5 or more reviews we analyzed these results separately. Figure 11.4 shows the results of limiting Embase to major thesaurus terms only for the reviews that performed a meta-analysis, all 7 domains we identified, 5 of the most frequently observed types of diseases and 2 types of interventions. For two subsets of reviews the searches in Embase limited to major thesaurus terms retrieved all included references for all reviews. The 10 reviews in the therapeutic domain that included only randomized controlled trials, and the 5 reviews about neoplasms were not affected by limiting Emtree terms to major. Here more focused searches might still generate acceptable results.

For these two reviews types we analyzed in more detail the effects of the focused searches by searching in the title and/or abstract fields only in Embase, and by focusing MEDLINE searches further. Figure 11.5 and Figure 11.6 show the results of these analyses. Both types show that focusing Embase did not negatively affect the included references. When MEDLINE was limited further some reviews missed included references. This was because Embase did not find any unique references for the therapeutic reviews that included only Randomized Controlled Trials (RCTs); and Embase found only one unique included reference for reviews in neoplasms.

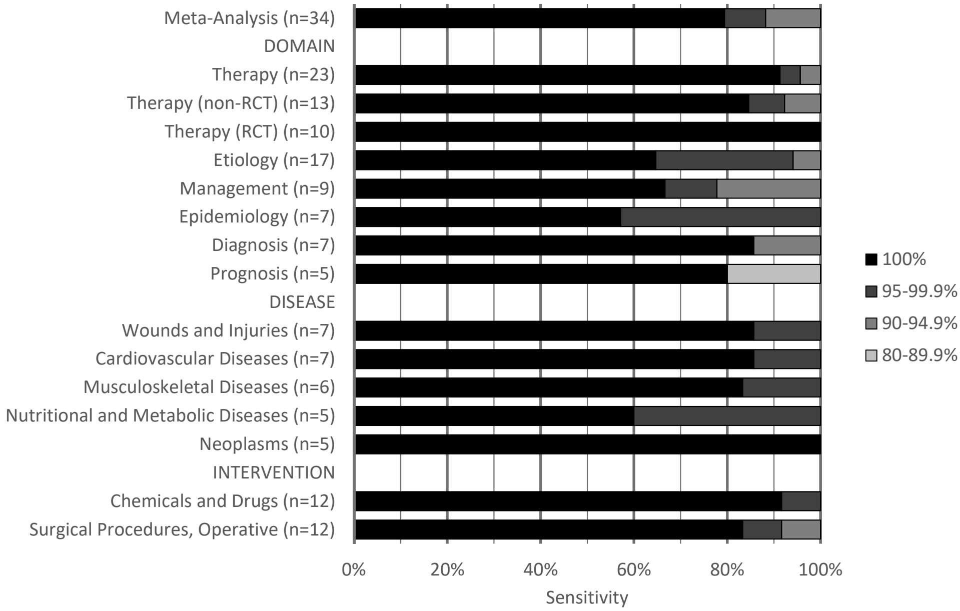


Figure 11.4: The sensitivity of searches for all databases where Embase was focused to major thesaurus terms for different review types and topics

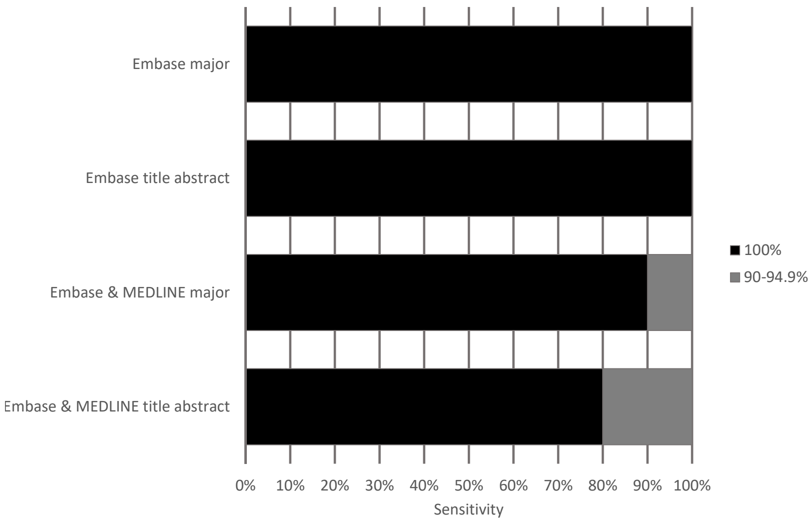


Figure 11.5: The effect of the four focused search methods on therapeutic reviews that included only RCTs (N = 10)

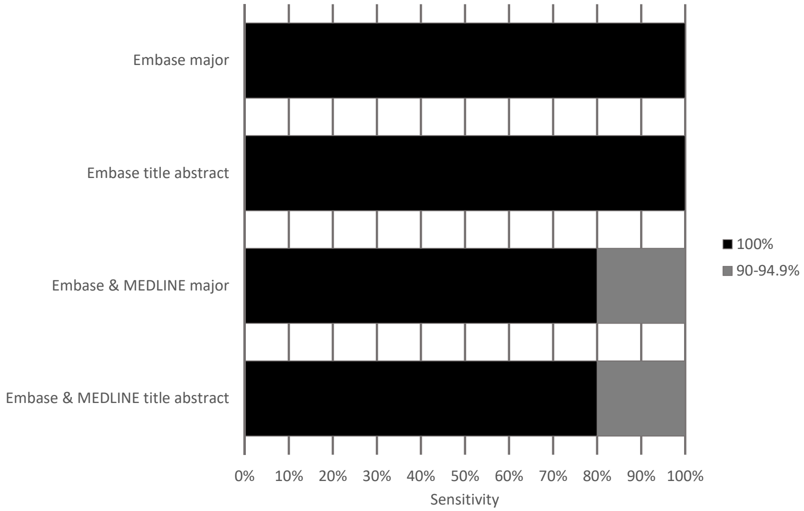


Figure 11.6: The effect of the four focused search methods on reviews about oncology (N = 5)

11.4 Discussion

In this paper, we have compared four different search methods to reduce the total yield of articles for screening. For each method, the aim was to determine the likelihood that a more focused search might miss relevant articles. The method with the fewest negative consequences (option 1: limiting searches in Embase to major thesaurus terms) only reduced the overall number of search results by 8%. Two methods that resulted in a greater reduction of search results but with acceptable recall for most SRs were: 2) searching Embase for terms in the titles and abstract fields only and 3) limiting searches in both Embase and MEDLINE to major descriptors. The method that found the fewest search results is method 4: limiting Embase and MEDLINE searches to terms in titles or abstracts alone which resulted in a substantial loss of relevant papers where more than 95% of the included references in fewer than 67% of the SRs would have been retrieved.

To our knowledge, this is the first study to investigate the possible effects of using more targeted search strategies in Embase and MEDLINE. In 2015, investigators from the Canadian Agency for Drugs and Technologies in Health (CADTH) published an extensive report on the topic (12). Still, because sensitivity scores in their original searches were low to begin with, sometimes even less than 10%, focusing them further would have been problematic. We broadened the scope of the previous research by looking at both Embase and MEDLINE, and bij searching for title/abstracts only terms.

We observed that for therapeutic reviews that included RCTs only, and for reviews on neoplastic diseases, focusing the Embase searches did not have a negative effect on the set of included references. However, the number of reviews in both of these groups is small. The generalizability of this observation might not be high.

In this research, it was not necessary to confirm that all included references from all reviews were found. Here, data collected during previous research was reused and examined in different detail.(14) This previous work aimed to identify the databases from which all included references had been retrieved. For the references that had been retrieved from databases other than Embase and MEDLINE, most notably Web of Science, we did not change our search strategies. The goal was to determine the recall of the more targeted searches by examining only those references that were uniquely retrieved by Embase and / or MEDLINE. For reviews in which all included references had also been retrieved by databases other than Embase and MEDLINE, recall of all targeted searches was determined to be 100%.

We decided against the testing of these more focused search approaches in our MEDLINE strategies (ie., on their own) while keeping the high recall Embase search strategies. Embase is known for more exhaustive indexing than MEDLINE, therefore focusing the search in MEDLINE only would be unlikely to reduce the number of references. MEDLINE is considered the gold standard for biomedical database searching and therefore did not make sense for us to limit our searches in MEDLINE while keeping the Embase search as sensitive as possible.

Further, we did not determine the effect of limiting specific PICO (Patient, Intervention, Comparison, Outcome) elements to major thesaurus terms done by previous researchers.(12) At CADTH, they concluded that it was safe to limit intervention concepts to major terms, but that limiting population elements to major terms should be avoided. However, the reviews included in our research covered a wider range of topics, many of which cannot easily be translated into the PICO framework. In this research, the consequences of focusing all elements in a search strategy are smaller than those by CADTH. Therefore, it was unnecessary to limit specific elements only. We left certain elements unchanged such as study types and population characteristics (such as age, gender or species) because these are check tags which are never assigned as major terms.

.

Another limitation of this research is that all search strategies were designed by one searcher working at one institute. Methods used at Erasmus MC are designed to create highly-sensitive sets of search terms in both thesaurus terms and in titles and abstracts.(17) It is unclear whether our conclusions are generalizable to other settings and institutions since other information specialists and review authors may not have have similar emphasis on titles and abstract terms. It must be stressed that decisions to limit search strategies for SRs should be made with close consideration to local context and resources. If review teams aim to find every article relevant to their research questions, imposing search limits should be carefully tested. Focusing search strategies should only be considered in those instances when the number of references retrieved is well beyond project resources. If the available resources permit screening all retrieved references, and their aim is to retrieve all relevant references, then focusing search results is not recommended, due to the, albeit small, chance of missing relevant references.

Further, reviewing high numbers of references may not be cost prohibitive for some SR teams. When reviewers use methods to review titles and abstracts within EndNote, the median number of articles reviewed per hour can be as high as 300, which is three times that observed for other methods.(18) Also other dedicated tools such as DistillerSR, Covidence or Rayyan show promisin results in speeding up the review process. When reviewers aim to reduce the time needed to screen titles and abstracts, changing methods used to screen titles and abstract, rather than limiting the recall of searches, may be a suitable alternative. Changing screening methods can reduce the total amount of time needed than focusing the searches, and is less likely to result in relevant papers being missed.

Before focusing searches to obtain greater precision, SR searchers should remember that the number of relevant references retrieved is highly dependent on the selection of terms used to search titles and/or abstract fields. When limiting database searches to major terms, we recommend using highly sensitive search strategies that contain sufficient words and phrases in titles and/or abstracts. The methods used to develop search strategies that were used for these SRs are described in detail in a separate report (17). This report describes a method developed by Erasmus MC librarians to help optimize the terms in the title and/or abstract fields. Because of the unique methods used to construct our searches, our conclusions should be viewed with caution, and we invite other information specialists to review their own data in a manner similar to our own.

11.5 Conclusion

Systematic review searching aims to be optimally sensitive but the resulting size of the biomedical literature retrieved often challenges the review teams to perform thorough screening. The question we aimed to answer is: can researchers use focused searches to reduce the time burden of the screening process? 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 both Embase and MEDLINE can be focused by searching for thesaurus terms as major descriptors. This approach may not ultimately have negative consequences in SRs, as long as a thorough search in other databases such as Web of Science next to the search in Embase and MEDLINE is performed. However, the reduction in the number of search results will in all likelihood be limited. Searching Embase and MEDLINE using terms in titles and abstracts alone results in too many relevant articles being missed, and is therefore not recommended.

11.6 References

1. Sampson M, Tetzlaff J, Urquhart C. Precision of healthcare systematic review searches in a cross-sectional sample. Research Synthesis Methods. 2011;2(2):119-25. doi:10.1002/jrsm.42.
2. 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;4:82. doi:10.1186/s13643-015-0074-7.
3. Sampson M, Barrowman NJ, Moher D, Klassen TP, Pham B, Platt R et al. Should meta-analysts search Embase in addition to Medline? J Clin Epidemiol. 2003;56(10):943-55.
4. Topfer LA, Parada A, Menon D, Noorani H, Perras C, Serra-Prat M. Comparison of literature searches on quality and costs for

health technology assessment using the MEDLINE and EMBASE databases. Int J Technol Assess Health Care. 1999;15(2):297-303.
5. Harping P. Introduction to controlled vocabularies terminology for art, architecture, and other cultural works; terminology. Los Angeles: Getty Trust Publ.; 2013.
6. Embase Indexing Guide 2015 : A comprehensive guide to Embase indexing policy 2015. http://supportcontent.elsevier.com/Support%20Hub/Embase/Files%20&%20Attachements/4683-Embase%20indexing%20guide%202015.pdf.Accessed 12 September 2018.
7. Jenuwine ES, Floyd JA. Comparison of Medical Subject Headings and text-word searches in MEDLINE to retrieve studies on sleep in healthy individuals. Journal of the Medical Library Association. 2004;92(3):349-53.
8. Elsevier R&D Solutions. A Comparison of Emtree® and MeSH® - Whitepaper. 2015. http://supportcontent.elsevier.com/Support%20Hub/Embase/Files%20&%20Attachements/4685-Embase_White%20Paper_Comparison%20of%20Emtree%20and%20MeSH_July%202015.pdf. Accessed November, 23 2015.
9. Wilkins T, Gillies RA, Davies K. EMBASE versus MEDLINE for family medicine searches: Can MEDLINE searches find the forest or a tree? Can Fam Physician. 2005;51(JUNE):848-9.
10. Elsevier. Embase Indexing Guide 2015: A comprehensive guide to Embase indexing policy 2015. https://www.elsevier.com/__data/assets/pdf_file/0016/92104/Embase-indexing-guide-2015.pdf. Accessed 30 Sep 2016.
11. Hausner E, Simon M, Kaiser T. Focus EMTREE terms vs. Non-focused EMTREE terms to search for clinical drug trials in EMBASE. 2008. http://abstracts.cochrane.org/2008-freiburg/focus-emtree-terms-vs-non-focused-emtree-terms-search-clinical-drug-trials-embase.
12. Canadian Agency for Drugs and Technologies in Health (CADTH). Pruning Emtree: Does Focusing Embase Subject Headings Impact Search Strategy Precision and Sensitivity? 2015. https://www.cadth.ca/pruning-emtree-embase.
13. Duffy S, Ross J, Misso K, de Kock S, Noake C, Stirk L. Is it possible to focus EMTREE without loss of sensitivity when searching Embase for systematic reviews? Analysis of Cochrane Reviews and HTA reports. EAHIL Conference; 6-11 Jun 2016; Seville, Spain2016.
14. Bramer WM, Giustini D, Kramer BMR. Comparing the coverage, recall, and precision of searches for 120 systematic reviews in Embase, MEDLINE, and Google Scholar: a prospective study. Systematic Reviews. 2016;5(1):39. doi:10.1186/s13643-016-0215-7.
15. Thomson Reuters. EndNote X7. 2015.
16. National Library of Medicine. MeSH database. https://www.ncbi.nlm.nih.gov/mesh. Accessed 15 september 2018.
17. Bramer WM, de Jonge GB, Rethlefsen ML, Mast F, Kleijnen J. A systematic approach to searching: How to perform high quality literature searches more efficiently. submitted. 2017.
18. Bramer WM, Milic J, Mast F. Reviewing retrieved references for inclusion in systematic reviews using EndNote. J Med Lib Assoc. 2017;105(1).

Chapter 12

General discussion

12 General discussion

Information specialists who assist researchers with librarian-mediated searches of the literature for systematic reviews are facing increasingly heavy workloads and more complex tasks. The objective of this thesis was to provide better guidance for information specialists about how to perform these searches more efficiently. Accordingly, this thesis presents and evaluates several tools and methods that searchers – medical librarians, information specialists and researchers – can apply when creating comprehensive searches and when deduplicating and updating search results. A secondary goal was to help researchers during the reference inclusion and exclusion process and to reduce, where possible, the burden of reviewing references. For searchers not planning on using our advanced searching methods, we also provide guidance on for the selection of databases to search, and for limiting searches to major thesaurus terms, or title and abstract fields.

12.1 Principal Findings

A new method for conducting searches

In Chapter 2 we presented a new way of searching for references to be included in systematic reviews. The process starts with a research or clinical question and ends with optimal searches in relevant databases. A step-by-step protocol helps librarians to create single-line search strategies in a text document that combines thesaurus terms and terms searched in titles or abstracts. To accomplish high sensitivity, we developed an optimization method to find additional relevant search terms. Temporary changes in a search strategy highlight potentially relevant terms in title and abstract fields, and in the controlled vocabulary fields of articles that have already been retrieved. Adding these terms to an existing search strategy allows the researcher to evaluate the extra articles thus retrieved. Macros ease the translation between different databases and interfaces.

While such a method overlaps with current practices articulated by other information specialists, to the best of our knowledge the steps described in our protocol have not been described fully elsewhere. Several other features of our approach differ from those of more generally accepted methods. For example, our method does not use the PICO model, the preferred starting database is Embase and it creates single-line search strategies. Three features of the method are unique: the fact that searches are constructed using predefined syntaxes, the additional optimization steps, and the use of macros to translate search strategies.

Using EndNote when performing systematic reviews

We recommend that researchers and librarians use EndNote for the bibliographic management of a systematic review, and we have developed various methods that speed up several stages of the process. In Chapter 3 we described a method in EndNote for faster deduplication of references: fields used to compare duplicates in EndNote are adapted in a specific sequence, after which a selection of references can be manually checked for duplicates. This method works faster than other methods whereby screening is usually done manually to ensure that falsely positive duplicates are not discarded. In Chapter 4 we described how researchers conducting a systematic review can screen the titles and abstracts of references they have retrieved. Records in EndNote are grouped according to the reviewer's verdict on whether to include or exclude each one. Deduplication then allows the references to be compared, and discrepancies between first and second reviewers identified. When interviewed, systematic reviewers who used the method indicated that they could review a median of 300 references per hour, a figure much higher than generally considered possible.

A systematic review should be up to date when it is published. However, performing a systematic review takes substantial time, and the search may miss current references. Most searches must therefore be updated to make sure the most recent articles are included in the resulting publication. Traditionally the process of updating searches relies on date selection in the databases. We described in Chapter 5 a method that uses EndNote to update search strategies. When an update is needed, results from all databases are imported and deduplicated in EndNote. The contents of the EndNote file that was created earlier are then copied into the new file. After deduplication, both the old records and the new records are removed, leaving only the new references. EndNote is also applied in the methods described in Chapter 6, which researchers conducting a systematic review can use to download reference lists of previously identified relevant articles or articles that refer to those articles. Output styles were created in EndNote which can then be used to generate search strategies in Web of Science or Scopus aimed to find those previously identified relevant articles. From those databases the reference lists can then be downloaded into EndNote and deduplicated with the results of the earlier searches.

The search method performs well in comparison with searches by other information specialists

In Chapter 7 we compared the performance of the new search method with that of searches created by other information specialists, in terms of the time needed to create the searches, and the outcomes of searches. The searches performed with our method were generated more quickly than those performed by other information specialists. However, the effectiveness of our searches was just as good. While the number of terms used in our searches was equal to that used in the control methods, we searched within a higher number of databases. As might be expected, the searches created with our method retrieved more references, but on average they also identified a significantly higher number of relevant references. The specificity (percentage of retrieved references that are eventually included) showed little difference between our method and other methods.

General advice for improving searches for reviews

In 2012, Gehanno et al. found that all references included in Cochrane reviews were indexed in Google Scholar.(1) The conclusion of this article – that therefore Google Scholar alone was sufficient for retrieving all relevant references in systematic reviews – was what led to us conceiving our first research projects. In Chapter 8, we investigated whether searches for published reviews would have found all relevant references if the search had been conducted in Google Scholar only. Contrary to the findings of other Gehanno et al, our results suggested that a single search in Google Scholar is not sufficient for systematic reviews. We then wondered whether the outcomes of our research were generalizable, as the quality of the searches conducted in the systematic reviews included in our first study, where we looked at reviews by others, was rather low. As a follow-up, we investigated systematic reviews for which we had developed the search strategies ourselves. We compared the coverage and retrieval of Embase, MEDLINE and Google Scholar, as described in Chapter 9. While coverage was high for each of the databases included in the research retrieval within a single database was far from complete. We then tested optimal combinations of databases, as described in Chapter 10. To find an acceptable percentage of relevant references, we found that as a minimum in Embase, MEDLINE, Web of Science and Google Scholar should be searched. Our findings indicate that searching a higher number of databases results in a higher number of references being retrieved. In Chapter 11 we investigated whether we could reduce this the number of hits, and therefore the time needed by researchers to screen articles for a systematic review, by searching for major thesaurus terms, or by searching for keywords in the titles and abstracts only. We found that any restriction that substantially reduces the number of hits would result in the loss of too many relevant references.

12.2 Methodological Considerations

Testing the methodology described

The key part of the research presented in this thesis is our method for developing systematic search strategies. We went back and looked at seventy-three published systematic reviews in which we had used this method to design the searches. We used these reviews to test our new method in terms of two different comparisons. Firstly, we compared the time spent on these searches with the time spent by other information specialists in 204 other systematic review projects from registries by other medical university libraries. Secondly, for 258 reviews that mention receiving assistance from information specialists at other Dutch universities, we determined the search effectiveness, such as the number of references retrieved and the number of articles included in these reviews. Unfortunately, other universities' registries for time spent on searches lacked links to the resulting publication, and the information specialists mentioned in the reviews for which we determined effectiveness had not registered the time they spent on

their searches. However, we found that a key advantage of our method was that it could retrieve a higher number of relevant references in a much shorter time.
We have made plans to test the method in a more comprehensive way. In our opinion, the best way to compare search methods would be to have multiple information specialists using different methods simultaneously performing a search on a given topic. However, comparing search methods is challenging. There is currently no standard method to create search strategies, and there is no standard education for it. Information specialists generally learn to search on the job from senior colleagues. Therefore, research comparing search methods will mainly compare search skills. A further difficultly in the comparison is that the search creation relies heavily on optimal communication with the researcher about the topic of the search. Standardizing such information for each searcher would be rather complicated. Apart from that, it is difficult to find volunteers to invest several hours of their time in a search that is not for one of their own clients.

In Chapters 3 to 6 we described several methods which use EndNote in the process of performing systematic reviews. The method for deduplication of the search results from various databases has been evaluated against several gold standards derived from previous research on deduplication methods (2-4). The results of this have been presented by us at a conference (5). We found that the three previous publications all describe methods that would take significantly more time than our method would. Deduplication of a set of 10,000 references might take up to a week. When we interviewed information specialists from other institutes they confirmed that deduplicating such a large set of references would take them several days. In contrast, using our method, deduplication of an initial set of 10,000 references takes approximately 15 minutes. When we compared the results of our deduplication method with those from other methods, we saw that the margin of error of our methods is about 1 false duplicate in 3000 records. Results from several systematic reviews conducted at Erasmus MC show that the sensitivity is 99.5%. We can thus draw the preliminary conclusion that the method is not only fast but also accurate. We have planned to perform a more detailed comparison of various deduplication methods, in term of accuracy, speed and ease of use.

Although we have not published evaluations of the other EndNote methods described in this thesis, we can share our experiences as we have used them in several review projects at Erasmus MC. There are several dedicated software programs that can be used to assist reviewers in the process of including and excluding references. Our experience is that assigning a verdict in these programs requires one or more actions per reference. For instance, the reviewer has to select a reason for exclusion from a dropdown list and click a button before being able to proceed to the next title or abstract. In our method, verdicts can be assigned to larger groups of references simultaneously. Since normally around 95% of the references retrieved are excluded based on title and abstract, this reduces not only the time it takes to review references, but also the burden on reviewers. A downside we experienced to our method is that if EndNote unexpectedly stops working, the selection of references to be included is often not saved. In addition, an error frequently made by EndNote users is that when opening the EndNote file, instead of opening the EndNote library (extension .enl), they reopen the compressed library (.enlx). In this instance, the selection of references to be included and excluded can also be lost. We have asked the manufacturer of EndNote to implement software enhancements that will prevent these problems.

Guidance for search methods

In Chapters 8 to 11 of this thesis, we provided guidance for searchers on the choices they need to make when performing a search, especially when selecting databases and deciding whether or not to limit the search to major thesaurus terms.
The first published study in this thesis was performed as a response to an earlier study that had the following two conclusions: "The coverage of GS [Google Scholar] for the studies included in the systematic reviews is 100%. If the authors of the 29 systematic reviews had used only GS, no reference would have been missed", and "With some improvement in the research options, to increase its precision, GS could become the leading bibliographic database in medicine and could be used alone for systematic reviews" (1). By focusing our research on the retrieval and not the coverage of the databases, we were able to refute these conclusions, and showed that – for both our own reviews and those done by others – no single database can retrieve all relevant references for systematic reviews.

The number of databases that can be used in a search, and which ones, depends on the databases that an institution has access to. While many hospitals and academic institutions cannot afford access to major databases such as Embase, for example, they can use freely available databases and search engines such as PubMed or Google Scholar (6). In addition, search experts are more likely to be employed by academic hospitals than by general hospitals, where less emphasis is put on research (7). It is unsurprising, therefore, that the number of databases generally used in systematic reviews is lower than the number used when information specialists are involved. Our own data on this topic is consistent with that of previous publications (8, 9).

The work in this thesis includes systematic reviews on medical topics only, which limits the generalizability of our results and the application of the methods we have proposed. For medically oriented systematic reviews, we consider the retrieval provided by four databases (Embase, MEDLINE, Web of Science and Google Scholar) to be good enough. However, because the searches with which we tested retrieval from these databases were developed with our own search method, retrieval from these databases might not be optimal for other information specialists using other search methods. We also note that databases such as Web of Science and Google Scholar do not contain thesaurus terms. Therefore, retrieval from these databases will rely more heavily on the completeness of the terms searched in title and abstract. In our method, we optimized our searches such that, where possible, they contained all relevant thesaurus terms. But we also spent much time optimizing the terms to be searched in the title and abstract. Information specialists who use other methods might fail to identify all relevant terms in the title and abstract, thereby reducing the added value of using Web of Science and Google Scholar. Of all four databases, we found that Embase identified the highest number of unique relevant references. It should be noted that in our method the search strategy is optimized for Embase, after which the search is translated semi-automatically for use in other databases. Our findings that Embase was the database that provided the highest number of unique hits is therefore not unexpected. Other information specialists often start their searches in MEDLINE and might find fewer added value of Embase that we did.

12.3 Findings In Perspective

Our method for the creation of search strategies can retrieve many relevant references in a short time period. We must bear in mind, however, that it may be difficult, both for researchers conducting a systematic review and for other information specialists to learn the method correctly. While the approach has been taught at international workshops, it has not undergone large-scale evaluation. And while the new aspects of the method (the macros and optimization method, in particular) can probably be incorporated in practice, searchers may be reluctant to accept those parts of the approach that differ from their daily practice. For example, many searchers might be reluctant to create single-line search strategies and to start in Embase, yet these are crucial aspects of our approach. Such obstacles may prevent searchers from adopting the method, or for those that are willing to give some aspects of our method a chance, might result in lower appreciation of the method.

Limitations of the research

We acknowledge that our research has a number of limitations. Firstly, our method for search strategy development relates to bibliographic databases only, as is traditionally the case when conducting a systematic review. However, obtaining a truly exhaustive overview of all literature on a certain topic requires additional sources: the review synthesis should also include so-called gray literature (10-15), hand searching (16), and data from unpublished studies as found in trial registries (17-21). Additional relevant references can be identified by reviewing the reference lists of relevant articles found by the search (22-25). In this Chapter 6, we have described methods to ease the scanning of reference lists in such articles. However, in our institute, citation searches and other retrieval methods mentioned above are normally the task of researchers themselves and not of medical librarians. For this reason, the effectiveness of using these improved reference list screening methods to identify additional references was not evaluated in this thesis. Alternative methods for reference collection use text mining of known relevant references to identify search terms. Although a study has shown text mining to be faster than traditional methods (26), the search times reported in this study were still much longer than ours. As part of a future project, we hope to be able to compare our own method to the text mining method.

The usefulness and replicability of the methods described in this thesis are highly dependent on the databases and interfaces available to searchers. MEDLINE is available for free via the interface

PubMed or as a paid subscription via other interfaces. Because of the lack of proximity operators, PubMed is not the interface that is most suitable for this method. Embase is only available through a subscription: Embase.com is available directly via Elsevier, but it also available via the Ovid interface from Wolters Kluwer, which can hosts many other databases, such as MEDLINE. At Erasmus MC, the primary database is Embase using the embase.com interface, and most of the time and effort spent in the crea tion of search strategies takes place within that interface. On the face of it, there is no notable difference between Embase content at embase.com and Embase via the Ovid interface (27). However, from personal experience during a period when the embase.com interface was temporarily unavailable, it is clear that using our search methods in the Ovid interface is more challenging, because of the lack or relevance ranking.

For our method, we also developed macros to translate syntaxes for use in different interfaces. Other researchers have developed an online tool called Systematic Review Accelerator, that can also translate queries from PubMed or Ovid MEDLINE for use in many other databases (28). This tool requires no manual editing by the user, as some of our macros do. A disadvantage of using the online tool in combination with our method is that it uses MEDLINE as the starting point, whereas we recommend to use Embase as the first database. Regardless of the method used to translate the syntax, a manual translation of MeSH terms into Emtree terms or vice versa is therefore required whenever starting in a database other than that for which the means of syntax translation was designed. Although the macros developed to translate search strategies for use in different databases are custom-made for the databases and interfaces used at Erasmus MC, institutions that subscribe to the same databases can still use the macros, which are available online. While the databases Erasmus MC subscribes to are major databases, many institutions have other subscriptions, such as Ovid Embase. Information specialists at those institutions will therefore have to adapt the macros to their own needs accordingly.

The online information accompanying the macros describes how they can be adapted for different database combinations. We have recently added extra macros that are not used by us, but that might be relevant to other institutes.

A final limitation relates to our use of EndNote for the bibliographic management of a systematic review. While others have published an alternative method for the inclusion and exclusion of articles that also makes use of EndNote (29), this method appears to be much more time consuming and complicated than our method. Alternatives for our method can also be found in dedicated software such as Rayyan (30). However, the use of dedicated software is often more time consuming, as the reviewer has to perform an action, such as select an item from a drop down menu or tick a box for each article. Our method is faster since the majority of articles retrieved are excluded in the title and or abstract screening phase, which allows the reviewer to mark irrelevant records in bulk. An advantage of Rayyan is that it uses machine learning to sort the remaining references in such a way that those most likely to be relevant are shown first. After a while, the program indicates that the remaining set contains no other likely relevant references. If a future version of EndNote allows the reviewer to highlight relevant and irrelevant words in abstracts, this will certainly increase its performance.

Generalizability of the methods

Although the research presented in this thesis was performed in the context of the medical literature, we hope that our methods will be used by information specialists from different fields and in different settings, and for different purposes. To this end, we have made a start at studying how generalizable our methods might be. The methods presented here have been taught to information specialists in other fields at general conferences for librarians. However, the search methods have proved to be less practical in other fields. The method's success in the medical databases is caused by one key factor, namely its reliance on terms in a thesaurus (including synonyms and narrower terms). These thesaurus terms are assigned by specialists to each article when it is added to a database. In the field of medicine, there are two major thesauri available, both of which have good coverage of a wide range of topics: MeSH for MEDLINE and Emtree for Embase. The thesaurus terms and synonyms provided by these thesauri form the basis of the initial searches our method, and this is also the case for most other medical search strategies. We also note that when using the additional optimization step described above, it is vital that interface and database enable separate searches for thesaurus terms and for terms in title and/or abstract. General scientific databases such as Web of Science and Scopus do not assign thesaurus terms to their articles, and they do not allow for such separate searching of the keywords.

Although the methods in this thesis have been developed for searches conducted for systematic reviews, they can also be used for other literature syntheses. A working group from Canada has proposed a code of practice for conducting librarian-mediated searches, ranging from level one – simple searches to bridge an information gap – to level five – searches to support a systematic review (31). The code consists of 87 unique steps divided into seven search stages. Although the steps are not described in great detail, most of the steps in both the more basic searches and the systematic review searches are also described in our method.

12.4 Directions for Future Research

The research presented in this thesis suggests that the effectiveness of searches created with our method is superior to that of searches conducted for other reviews. When we compared the average effectiveness of numerous searches, we found our method to be faster, and to retrieve a higher number of relevant references. We note, however, that we were unable to make direct comparisons on the same topics, and so had to compare searches carried out on a wide range of different topics. The questions remains, therefore, whether or not the method we have developed retrieves relevant references missed by other search methods, and also whether our method might in fact still be missing relevant references identified using other methods. We envisioned addressing these questions in a research project whereby multiple information specialists would conduct simultaneous searches on the same research topic. It turns out, however, that it is hard to find volunteers to do the searches given the considerable time investment. Apart from that, it is almost impossible to perform a high quality search without the researcher present to provide detailed information on the topic. In theory a researcher could meet with 3 different information specialists, however, the information presented to the second and third information specialist might be influenced by the searches created by the first meeting. Indeed, part of our method requires the researcher to be involved in the search process: the researcher is present during the search development for instant feedback on the relevance of potential new terms and newly retrieved references. A complicating factor for the envisioned research project is that systematic review projects take a long time to be completed. The searches' effectiveness can only be evaluated once the final list of references is known. In addition, when designing such a study we must bear in mind that not all search requests for systematic reviews result in that review getting published: although a pilot study was carried out by three information specialists, the search did not result in a finished systematic review as the researchers found little evidence for their hypotheses. Because the researchers did not decide on a final list of relevant references this pilot project was unusable. If we can overcome the barriers mentioned above we plan to continue this research.

Assisting each review project with in-person meetings to create a search is time consuming for information specialists (32). However, unless they have received proper training, most researchers do not have enough knowledge to search the databases themselves in a way that is sufficient for a systematic review (33). Our research has shown that the method we have developed can lead to the creation of high-quality searches that can be completed in a short time period. However, since the method was tested only as executed by its developer, it is unclear how the same method would perform when used by other people. The question still remains whether our method can be effectively taught to others, and this will be addressed in research planned as a follow-up to this thesis. One of the planned follow-up projects entails a before-and-after study in which information specialists are asked to create a search strategy on a given research topic. Participants then attend a workshop in which they are taught how to use the method. Immediately after the workshop, another research question is given to them. And as a final testing point, six months after the workshop, they receive a third research question for which they are asked to create a search strategy. This study design will allow us to determine whether the workshop whether the workshop achieves lasting improvements in terms of the completeness of the search strategy (search terms used) and its effectiveness (percentage of relevant references retrieved).
Another research project that we have started focuses on the optimization steps in our method. This is related to the fact that in general people searching references for a systematic review want to know whether their current strategy has already identified all relevant references, or whether additional efforts will still find additional references. To address this need, for several search strategies, we have registered the steps leading to the final search strategy, storing the

complete search strategy roughly every five minutes, or after making major changes. Once the systematic reviews for which these searches were developed have been completed, we can determine where in the process each of the references included in that review was retrieved. If all relevant references are already being retrieved during the initial search, the optimization process the number of optimization steps can be reduced. However, if relevant references are being added right up until the last optimization steps, an even longer optimization period might be needed to ensure the inclusion of such additional relevant references. This might prove difficult to generalize though, as each review topic and therefore each search is different.

The future of searching biomedical databases

Numerous reports in the literature have speculated on how searches for systematic reviews will be conducted in the future. Some researchers envision a future where meta-search engines take over the tasks of translating a search between databases (34). One overview article on the future of literature research (35) has suggested that text mining (26, 36-38) or relevance ranking (39, 40) may well take the place of Boolean searches. However, since completeness and repeatability of the searches is crucial for a systematic review (41, 42), it is unlikely that web search engines will replace the traditional database searches in the near future (43). Traditional Boolean searches – as also created using our method – allow the searcher to have full control over the results of the search, and such searches are repeatable (44).

In our view, in the near future the practice of librarian-mediated searches will not differ much from current search practices. In our opinion, medical information specialists will not be replaced by tools. Translating an often complicated research question into a structured search remains a human task, although automated online tools can help the information specialist to create better searches, whereby text mining can provide lists of potentially relevant search terms (45). However, we envision that machine learning will be able to assist or replace in the screening of references, severely reducing the burden for the reviewers in terms of the tedious task of reading thousands of titles and abstracts. Indeed, tools such as Rayyan already use artificial intelligence to rank remaining references according to relevance to the topic. A system based on machine learning may be able to decide which of the retrieved titles and abstracts are most likely to be included, and which ones will probably be excluded (30, 46-48). After a certain cut-off point, there are no more potentially relevant references to screen. Because not all retrieved references have to be screened researchers can accept larger numbers of references retrieved for their systematic review. When developing searches, the trade-off between specificity and sensitivity is challenging: in nearly all cases, researchers want not only to find all relevant articles, but also not to find too many irrelevant references. If the researcher can handle larger numbers of references the information specialists does not have to optimize the specificity but focus mostly on the sensitivity of the search.

12.5 Conclusions and Recommendations

We have developed a method that explains from start to finish how to create a systematic review search strategy. We have also developed methods that help both information specialists and authors of systematic reviews in the process of reviewing search results. The quality and effectiveness of searches created with the search method is at least as good as those of other methods and librarians. Although the coverage of some databases approaches 100%, retrieval in real systematic review searches of the major databases investigated here is not sufficient for them to be used on their own. We therefore recommend the use at least the databases Embase, MEDLINE, and Web of Science, complemented with results from the search engine Google Scholar. Finally, focusing controlled vocabulary on major terms or searching title and abstract only – to reduce the number of references found in these databases – results in the loss of too many relevant references.

12.6 References

1. Gehanno J-F, Rollin L, Darmoni S. Is the coverage of google scholar enough to be used alone for systematic reviews. BMC Med Inform Decis Mak. 2013;13(1):7.
2. Jiang Y, Lin C, Meng W, Yu C, Cohen AM, Smalheiser NR. Rule-based deduplication of article records from bibliographic databases. Database (Oxford). 2014;2014:bat086.
3. Rathbone J, Carter M, Hoffmann T, Glasziou P. Better duplicate detection for systematic reviewers: evaluation of Systematic Review Assistant-Deduplication Module. Systematic Reviews. 2015;4(1):6.
4. Qi X, Yang M, Ren W, Jia J, Wang J, Han G, et al. Find duplicates among the PubMed, EMBASE, and Cochrane Library Databases in systematic review. PLoS One. 2013;8(8):e71838.
5. Bramer WM, Holland L, Mollema J, Hannon T, Bekhuis T. Removing duplicates in retrieval sets from electronic databases: comparing the efficiency and accuracy of the Bramer-method with other methods and software packages. Proc 14th European Association for Health Information and Libraries (EAHIL) Conference, 57-9; June 11-13; Rome, Italy. 2014.
6. Lam MT, De Longhi C, Turnbull J, Lam HR, Besa R. Has Embase replaced MEDLINE since coverage expansion? J Med Libr Assoc. 2018 Apr;106(2):227-34.
7. Thibodeau PL, Funk CJ. Trends in hospital librarianship and hospital library services: 1989 to 2006. J Med Libr Assoc. 2009 Oct;97(4):273-9.
8. Vassar M, Yerokhin V, Sinnett PM, Weiher M, Muckelrath H, Carr B, et al. Database selection in systematic reviews: an insight through clinical neurology. Health Information & Libraries Journal. 2017.
9. Wood H, Arber M, Glanville JM. Systematic Reviews of Economic Evaluations: How Extensive Are Their Searches? Int J Technol Assess Health Care. 2017 Mar 27:1-7.
10. Al Bedah A, Elolemy A, Al Mudaiheem A, Alrasheid M, Khalil M. Searching grey literature for evidence: Ozone therapy as a model. J Altern Complement Med. 2014;20(5):A142-A3.
11. Wright JM, Cottrell DJ, Mir G. Searching for religion and mental health studies required health, social science, and grey literature databases. J Clin Epidemiol. 2014;67(7):800-10.
12. Godin K, Stapleton J, Kirkpatrick S, Hanning R, Leatherdale S. Applying systematic review search methods to the grey literature: a case study examining guidelines for school-based breakfast programs in Canada. Systematic Reviews. 2015;4(1):138.
13. Mahood Q, Eerd DV, Irvin E. Searching for grey literature for systematic reviews: challenges and benefits. Research Synthesis Methods. 2014;5(3):221-34.
14. Enticott J, Buck K, Shawyer F. Finding "hard to find" literature on hard to find groups: A novel technique to search grey literature on refugees and asylum seekers. Int J Methods Psychiatr Res. 2017 Sep 04.
15. Adams J, Hillier-Brown FC, Moore HJ, Lake AA, Araujo-Soares V, White M, et al. Searching and synthesising 'grey literature' and 'grey information' in public health: critical reflections on three case studies. Systematic Reviews. [journal article]. 2016;5(1):164.
16. Richards D. Handsearching still a valuable element of the systematic review. Evidence-based Dentistry. 2008;9(3):85.
17. Wolfe N, Gotzsche PC, Bero L. Strategies for obtaining unpublished drug trial data: a qualitative interview study. Syst Rev. 2013;2:31.
18. Knottnerus JA, Tugwell P. The potential impact of unpublished results. J Clin Epidemiol. 2013 Oct;66(10):1061-3.
19. Ziai H, Zhang R, Chan AW, Persaud N. Search for unpublished data by systematic reviewers: An audit. BMJ Open. 2017;7(10).
20. Baudard M, Yavchitz A, Ravaud P, Perrodeau E, Boutron I. Impact of searching clinical trial registries in systematic reviews of pharmaceutical treatments: methodological systematic review and reanalysis of meta-analyses. BMJ. 2017 Feb 17;356:j448.
21. Le Cleach L, Doney E, Katz KA, Williams HC, Trinquart L. Research Techniques Made Simple: Workflow for Searching Databases to Reduce Evidence Selection Bias in Systematic Reviews. J Invest Dermatol. 2016 Dec;136(12):e125-e9.
22. Horsley T, Dingwall O, Sampson M. Checking reference lists to find additional studies for systematic reviews. Cochrane Database Syst Rev. 2011(8):MR000026.
23. Papaioannou D, Sutton A, Carroll C, Booth A, Wong R. Literature searching for social science systematic reviews: consideration of a range of search techniques. Health Info Libr J. 2010 Jun;27(2):114-22.
24. Wright K, Golder S, Rodriguez-Lopez R. Citation searching: a systematic review case study of multiple risk behaviour interventions. BMC

Med Res Methodol. 2014;14(1):73.
25. Belter CW. Citation analysis as a literature search method for systematic reviews. Journal of the Association for Information Science and Technology. 2016.
26. 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.
27. Fortier KJ, Kiss N, Tongbram V. What is the optimal search engine for results from Embase and medline: Ovid or embase.com? Value Health. 2013;16(3):A25.
28. Systematic Review Accellerator. Available from: http://crebp-sra.com/.
29. Peters MDJ. Managing and Coding References for Systematic Reviews and Scoping Reviews in EndNote. Med Ref Serv Q. 2017 2017/01/02;36(1):19-31.
30. Olofsson H, Brolund A, Hellberg C, Silverstein R, Stenstrom K, Osterberg M, et al. Can abstract screening workload be reduced using text mining? User experiences of the tool Rayyan. Res Synth Methods. 2017 Sep;8(3):275-80.
31. Baer S, Farrell A, Lee P, MacDonald J, Rabb D, Ballantyne Scott B, et al. Mediated Searching: A Code of Practice May 2018. 2018 [updated May 16, 2018; cited 2018 May 31]; Available from: https://standardsworkinggroup.blogspot.com.
32. 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 Apr;106(2):198-207.
33. Mikalef P, Kourouthanassis PE, Pateli AG. Online information search behaviour of physicians. Health Information & Libraries Journal.2017.
34. 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.
35. Lefebvre C, Glanville J, Wieland L, Coles B, Weightman A. Methodological developments in searching for studies for systematic reviews: past, present and future? Systematic Reviews. 2013;2(1):78.
36. 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 Jun 29.
37. 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 Jan 14;4:5.
38. Thomas J, McNaught J, Ananiadou S. Applications of text mining within systematic reviews. Research Synthesis Methods. 2011;2(1):1-14.
39. Sampson M, Barrowman NJ, Moher D, Clifford TJ, Platt RW, Morrison A, et al. Can electronic search engines optimize screening of search results in systematic reviews: an empirical study. BMC Med Res Methodol. 2006;6:7.
40. Karimi S, Pohl S, Scholer F, Cavedon L, Zobel J. Boolean versus Ranked Querying for Biomedical Systematic Reviews. BMC Med Inform Decis Mak. 2010;10:58.
41. Briscoe S. A review of the reporting of web searching to identify studies for Cochrane systematic reviews. Res Synth Methods. 2017 Oct 24.
42. Stansfield C, Dickson K, Bangpan M. Exploring issues in the conduct of website searching and other online sources for systematic reviews: how can we be systematic? Systematic Reviews. [journal article]. 2016;5(1):191.
43. Bates J, Best P, McQuilkin J, Taylor B. Will Web Search Engines Replace Bibliographic Databases in the Systematic Identification of Research? The Journal of Academic Librarianship. 2017 1//;43(1):8-17.
44. Hjørland B. Classical databases and knowledge organization: A case for boolean retrieval and human decision-making during searches. Journal of the Association for Information Science and Technology. 2015;66(8):1559-75.
45. Thompson J, Davis J, Mazerolle L. A systematic method for search term selection in systematic reviews. Research Synthesis Methods. 2014;5(2):87-97.
46. Shemilt I, Simon A, Hollands Gareth J, Marteau Theresa M, Ogilvie D, O'Mara-Eves A, et al. Pinpointing needles in giant haystacks: use of text mining to reduce impractical screening workload in extremely large scoping reviews. Research Synthesis Methods. 2013 2014/03/01;5(1):31-49.
47. Przybyła P, Brockmeier AJ, Kontonatsios G, Le Pogam MA, McNaught J, von Elm E, et al. Prioritising references for systematic reviews with RobotAnalyst: a user study. Research synthesis methods. 2018.
48. Gates A, Johnson C, Hartling L. Technology-assisted title and abstract screening for systematic reviews: a retrospective evaluation of the Abstrackr machine learning tool. Systematic Reviews. [journal article]. 2018 March 12;7(1):45.

Appendices

Appendices

English Summary

Systematic reviews are an important part of evidence based medicine. The quality of a systematic review depends on the quality of the search strategy. For high quality searches systematic reviewers are advised to seek the help of medical librarians. Librarians are involved in other parts of the systematic review process, such as deduplication of search results, updating searches and retrieving reference lists of included papers. However no clear guidance exists on how to create a search strategy for a systematic review, nor is there evidence of the best method to create a search strategy. In this thesis we aim to provide new and experienced searchers with a method to create searches for systematic reviews as well as methods in EndNote for other tasks librarians perform in the review process. We aimed to evaluate the search process and to provide guidance to searchers to improve the efficiency of their searches.

One main article of this thesis is chapter 2 which provides guidance on how to create a search strategy for systematic reviews. Especially important in the method is the creation of single line search strategies in a text editor which are then optimized, comparing the articles retrieved by the thesaurus terms to the ones retrieved by the title and abstract terms. This identifies potential new relevant terms to be added to the search strategy. Ultimately the researcher sitting next to the librarian makes the decision whether to add the term or not. Once the search strategy in the first database is optimized it is translated with the use of macros into other databases and interfaces.

The chapters 3-6 deal with the use of EndNote for assistance of the systematic review process. EndNote is traditionally already used in many libraries as a tool for researchers to format the bibliographies of their journal articles, yet it can be used in a wide variety of ways. Chapter 3 describes our method of deduplication. This task isn't always done by the medical library, especially because it is rather time consuming. The time needed to perform this task is greatly reduced with our method, which would allow librarians to perform this service for their clients. Chapter 4 guides researchers in the reviewing process within EndNote. It is widely accepted that reviewing titles and abstracts is done in an average rate of 120 references per hour. However, using our method researchers are able to review a median of 300 titles and abstracts per hour. An important task for librarians is to update searches before publication of the review. Chapter 5 describes how we use EndNote for that. By deduplicating the new references with the earlier retrieved references, and removing both duplicates, we avoid the often difficult date limitation in the databases. Finally, in chapter 6 we explain how we use custom made output styles in EndNote to find the included references and relevant reviews in Scopus and Web of Science, which allows us to download the reference list of these papers into EndNote for additional title and abstract screening, as required in systematic reviews.

Chapter 7 contains the evaluation of our search methods. At our institute we keep a documentation of review projects that we assisted. From that we could determine the time needed for the search and the complexity of the search strategies, and from the resulting publications we extracted the numbers of relevant references. Several metadata sets of review projects were shared by colleagues from other institutes. From these data we learned that the median time needed for information specialists to create the search strategies in these projects was 12 hours, which is significantly longer than our median of 70 minutes. We collected reviews published by Dutch academic hospitals that mentioned the involvement of a medical librarian either as a co-author or in an acknowledgement. Despite our faster approach our searches proved to be more effective as we retrieved significantly more relevant references than the reviews from the control group. Sadly direct comparison is difficult, as we could not obtain a set of reviews for which we know both the time used for the search strategies as well as the outcome of the final review. Also the topics between the control group and our reviews varied, which makes comparison difficult.

Chapters 8-11 of the thesis give advice to researchers and librarians performing systematic review searches, even when they don't use the full methods as described in chapter 2. Chapter 8 was the first article of the thesis that was published. It originated as a response to earlier research that concluded that since all included references of a large set of Cochrane systematic reviews are indexed in Google Scholar, Google Scholar could have been used to find all references. We investigated reviews from other institutes that used at least PubMed and Google Scholar and presented search strategies in a replicable way. Our data showed that a search in Google Scholar alone was not able to retrieve all relevant references. This was however partially due to the low quality of the search strategies used for these reviews. As a follow up for this research we investigated the effectiveness of our own searches in Embase, Medline and Google Scholar for our own systematic reviews. In chapter 9 we concluded that the retrieval of none of these databases is enough for it to serve as the single database in a systematic review. We therefore investigated in chapter 10 what would be the optimal combination of databases. We determined the probability of missed included references, if published systematic reviews for which we had developed the searches would have been limited to a certain combination of databases. We found that the number one important database was Embase, followed by Medline, Web of Science and Google Scholar. Given the overindexing often seen in Embase, where many more thesaurus terms are added to records than in Medline, we investigated in chapter 11 whether we could reduce the number of references without losing relevant references by focusing on major terms, or terms in title and or abstract only. We found that none of these methods would reduce the number of references to screen a lot, but all of them would cause systematic reviews to miss relevant references.

Finally chapter 12 is the general discussion. Whether the method can be effectively taught to others, still has to be investigated. We describe several alternative methods that we did not test in the thesis. Ideally we would evaluate the method by ways of side-by-side search creation with multiple information specialists working with different models. This research is planned for the future. In the future we still see an important task for information specialists in the creation of systematic review searches, however, the development of tools that facilitate screening of large sets of references can reduce the need to limit the search so the information specialist can focus only on retrieving all relevant references. We hope this thesis has provided an effective method for experienced and novice medical librarians.

Nederlandstalige Samenvatting

Systematic reviews zijn belangrijk binnen evidence-based medicine. De kwaliteit van systematic reviews wordt voor een belangrijk deel bepaald door de kwaliteit van de zoekacties. Voor zoekacties van goede kwaliteit wordt onderzoekers geadviseerd de hulp in te schakelen van medisch informatiespecialisten. Deze zijn niet alleen betrokken bij de zoekacties, maar meer en meer ook bij andere onderdelen van het review proces, zoals het ontdubbelen van resultaten in EndNote, het updaten van zoekacties en het downloaden van referentielijsten van geïncludeerde artikelen. Er zijn echter geen duidelijke richtlijnen hoe een goede zoekactie kan worden opgezet. In dit proefschrift willen wij nieuwe en ervaren zoekers een methode bieden om zoekacties voor systematic reviews maar ook voor andere doeleinden te maken. Daarnaast presenteren we methoden in EndNote voor de andere taken die door bibliothecarissen gedaan worden. We evalueren de nieuwe methode en bieden zoekers die niet van onze nieuwe methode gebruiken houvast om de efficiëntie van hun searches te verbeteren.

Het belangrijkste hoofdstuk is hoofdstuk 2, waarin we een gedetailleerde methode beschrijven voor het opstellen van een search, zoals voor een systematic review. Vooral belangrijk in de methode is het opstellen van eenregelige zoekacties in een tekstverwerker. De zoekacties kunnen dan worden geoptimaliseerd, waarbij artikelen die worden gevonden met de thesaurus termen worden vergeleken met artikelen die gevonden worden op basis van de titel en samenvatting. Hiermee kan de informatie specialist nieuwe, potentieel relevante termen identificeren die dan aan de onderzoeker worden voorgelegd om ze te beoordelen op relevantie. Als de zoekactie in de eerste database optimaal is in de ogen van zowel de informatiespecialist als de onderzoekers kan deze worden vertaald met behulp van macro's in diverse andere databases en interfaces.

Hoofdstuk 3-6 beschrijven het gebruik van EndNote tijdens andere onderdelen van het systematic review proces. Endnote wordt traditioneel gebruikt door onderzoekers als een middel om bibliografieën aan hun artikelen toe te voegen in de stijl die een tijdschrift wil, maar de mogelijkheden zijn veel uitgebreider. Hoofdstuk 3 beschrijft een nieuwe methode van ontdubbelen. Deze taak wordt niet in elk instituut door de bibliotheek gedaan, met name omdat het als vrij tijdrovend wordt ervaren. De tijd die dit kost is met onze methode sterk gereduceerd, waardoor het makkelijker wordt deze dienst aan te bieden voor onderzoekers. Hoofdstuk 4 legt uit hoe onderzoekers het proces van includeren en excluderen van artikelen in EndNote kunnen uitvoeren. In het algemeen wordt gesteld dat het beoordelen van 120 titels en abstracts van wetenschappelijke ongeveer een uur kost. Met deze methode was de gemiddelde snelheid van onderzoekers 300 artikelen per uur. Met de vaak hoge aantallen resultaten die uit systematic review zoekacties komen is dit een aanzienlijke tijdswinst. Voor publicatie van een review is het belangrijk om de zoekacties te updaten. Hoofdstuk 5 beschrijft hoe EndNote daarvoor gebruikt kan worden. Door nieuwe referenties te ontdubbelen met de eerder gevonden artikelen en daarbij beide referenties te verwijderen blijven alleen de nieuwe referenties over. Op die manier vermijd je het lastige gebruik van data beperkingen in de databases. Tot slot wordt in hoofdstuk

6 uitgelegd hoe met behulp van EndNote uit Scopus of Web of Science lijsten met geciteerde en citerende artikelen van geïncludeerde artikelen kunnen worden geïmporteerd voor extra screening van titels en samenvattingen.

Hoofdstuk 7 bevat de evaluatie van onze zoekmethode. Reviews waarvoor de zoekactie is ontwikkeld met de hier beschreven methode worden vergeleken met controle sets bestaande uit andere reviews. In ons instituut houden wij documentatie bij van review projecten waarvoor wij worden gevraagd om een zoekactie te maken. Daaruit kunnen we bepalen wat de tijd is die nodig was voor de zoekactie, en de complexiteit van de zoekacties. Verschillende metadata sets van review projecten zijn met ons gedeeld door collega's van andere instituten. Uit die data konden wij bepalen dat de mediane tijd die anderen nodig hebben voor het creëren van een search voor een systematic review 12 uur bedraagt. Dat is significant langer dan onze mediaan van 70 minuten. We verzamelden reviews van andere Nederlandse academische ziekenhuizen die de betrokkenheid van een medisch bibliothecaris vermeldden, hetzij als co-auteur, hetzij in een acknowledgement. Ondanks onze snellere aanpak bleken onze zoekacties minstens zo efficiënt, aangezien we significant meer relevante referenties vonden dan de reviews uit de controle groep. Helaas is een directe vergelijking erg lastig omdat we niet voldoende reviews konden vinden waarvan we zowel de tijd benodigd voor de search, als de uiteindelijke geincludeerde artikelen van de publicatie. Ook omdat de onderwerpen soms verschillen is het lastig om een directe vergelijking te maken.

Hoofdstuk 8-11 geeft advies aan onderzoekers en informatie specialisten die een zoekactie voor een systematic review maken, ook wanneer ze niet de volledige methode gebruiken zoals beschreven in hoofdstuk 2 Hoofdstuk 8 was het eerste artikel van het proefschrift dat is gepubliceerd. De oorsprong ligt in een reactie op eerder onderzoek dat concludeerde dat, omdat alle geïncludeerde artikelen in Cochrane systematic reviews waren geïndexeerd in Google Scholar, een zoekactie in Google Scholar voldoende zou zijn geweest om alle relevante artikelen te vinden. Wij onderzochten reviews van andere instituten die in ieder geval PubMed en Google Scholar hadden gebruikt en die zodanig werden gepubliceerd dat ze konden worden herhaald. We concludeerden dat de zoekacties in Google Scholar bij lange na niet effectief genoeg waren om alle geïncludeerde artikelen te vinden. Dit was deels terug te voeren op de lage kwaliteit van de zoekacties in deze reviews. Als een vervolg op dit onderzoek hebben we de effectiviteit van onze eigen zoekacties in Embase, Medline en Google Scholar onderzocht. In hoofdstuk 9 concluderen we daarmee dat de retrieval van geen enkele van deze databases voldoende is om te worden gebruikt als de enige database in een systematic review. Daarom hebben we in hoofdstuk 10 onderzoek gedaan naar de optimale combinatie of databases. We bepaalden daarvoor de waarschijnlijkheid dat geïncludeerde artikelen zouden worden gemist voor gepubliceerde systematic reviews waarvan wij de zoekacties hadden ontwikkeld, als die zoekacties zouden zijn beperkt gebleven tot bepaalde database combinaties. We vonden dat de belangrijkste database Embase was, op de voet gevolgd door Medline, Web of Science en Google Scholar. Omdat Embase de neiging heeft te overindexeren (er worden veel meer thesaurus termen toegekend aan artikelen dan bijvoorbeeld in Medline) onderzochten we in hoofdstuk 11 of we het aantal resultaten konden verminderen zonder daarbij relevante referenties te verliezen door de zoekacties te beperken tot major thesaurus termen in plaats van alle termen, of tot alleen woorden in titel en samenvatting. We vonden dat geen van de methoden het totaal aantal artikelen sterk verminderde, maar ze zouden wel alle zorgen voor gemiste relevante referenties.

Tot slot in hoofdstuk 12 bediscussiëren we het gehele proefschrift. We beschrijven diverse alternatieve oplossingen voor de problemen die we met dit proefschrift wilden oplossen die niet door ons zijn getest. Of de methode effectief kan worden overgedragen aan anderen moet nog worden onderzocht. Idealiter zouden we de methode grondig testen door twee informatiespecialisten zij-aan-zij een search te laten doen op hetzelfde onderwerp. Dit onderzoek staat gepland voor de toekomst. In de toekomst zien we nog steeds een belangrijke rol weggelegd voor de informatie specialist in het maken van zoekacties voor systematic reviews. Anderzijds kan de ontwikkeling van hulpmiddelen die het mogelijk maken om gemakkelijker grotere aantallen artikelen te beoordelen ervoor zorgen dat de onderzoekers niet langer druk leggen op informatiespecialisten om het aantal resultaten terug te brengen, waardoor de informatiespecialist kan focussen op het vinden van alle relevante artikelen. Wij hopen dat dit proefschrift een handreiking biedt aan zowel beginnende als ervaren informatiespecialisten in het zoeken voor systematic reviews.

Acknowledgements / Dankwoord

Een proefschrift schrijf je niet alleen. Zonder een goede achterban kun je het niet volbrengen. Daarom is het niet meer dan logisch om hier de belangrijkste personen te bedanken.

Oscar, I vividly remember our first encounter back in 2012. When you came to Rotterdam, the first thing you wanted was an appointment with the library. When you moved to Bern I told the librarians there to expect an early meeting with you, only to hear that you already had a meeting with them before you started your position. I never envisioned back in those days that we would be standing here today. The method presented in this thesis was still in development when we first met, and I didn't have any experience in scientific research yet. Today we have co-authored almost 30 articles. Despite your monthly reliefs that I was the easiest PhD student you ever supervised, you really lifted my research with your spot-on comments. Thank you for believing in me and my methods and for all opportunities and inspiration you gave me.

Jos, we ontmoetten elkaar voor het eerst in Utrecht in de kelder van het Leeuwenbergh Gasthuis in 2015. Kort nadat je mijn promotor werd waren er problemen met een editor en peer reviewer. Toen ik dat met je besprak stelde je voor dat als ze moeilijk bleven doen, je hen zou bellen. Het waren beide oud werknemers van je. Toen wist ik dat ik de juiste persoon had gevonden. Dank voor al je inzet en adviezen, al is het niet al eerste promotor in Maastricht zoals jij het graag had gezien.

Frans, toen je me in 2009 aannam hoopte je de nieuwe Guus van den Brekel te hebben gevonden die de innovatie van de medische bibliotheek zou kunnen initiëren. Gelukkig gaf je me tegelijk carte blanche om mijn eigen weg binnen de organisatie te ontwikkelen. We ontdekten samen de grote behoefte aan goed literatuuronderzoek voor systematic reviews en dat werd mijn speerpunt. In de innovatie ben ik altijd meer een volger gebleven, maar onder jou ben ik in het literatuuronderzoek tot een voortrekker uitgegroeid. Dank voor je adviezen en steun.

Ik wil de leescommissie **prof. Hunink**, **prof. Arends** en **prof. Scholten** hartelijk danken voor het lezen en goedkeuren van mijn proefschrift. Beste **Myriam**, bedankt dat je zo'n trouwe klant bent die al haar PhD studenten steevast naar mij doorverwijst voor hun zoekacties. Beste **Lydia**, jij was al supervisor van het eerste systematic review project waar ik als co-auteur aan verbonden was, tevens mijn eerste PubMed vermelding in 2012. Ik refereer in mijn cursussen nog regelmatig aan de kleine fout die ik toen maakte, waarbij door het vergeten van een woordje OR de hele search overnieuw gedaan moest worden. Ik leer de cursisten tegenwoordig hoe ze dat kunnen voorkomen. Beste **prof. Koenderink**, lieve **Gijsje**, bijzonder om je aan het werk te zien als hoogleraar en een eer om je in mijn oppositie te hebben.

Wetenschappelijke tijdschriftartikelen schrijf je niet alleen. Daarvoor heb je co-auteurs nodig. Als een buitenpromovendus zonder onderzoeksgroep om je heen zijn die mensen lastiger te vinden. Het is daardoor ook een heel heterogene groep geworden.

I am lucky to know some of the best co-authors of scientific journal articles.

Dean, your responses are always so timely, and full of marvelous comments that lifted many of the publications in this thesis to a much higher level. Though we have never met in person I feel a deep friendship for you. Thank you for all your hard labour, not only in this thesis, but in general to the medical library community.

Melissa, you are a great inspiration for any librarian that wants to start writing scientific journal articles. Your comments as co-author or peer reviewer are always spot on and though sometimes at first glance your comments may appear a bit harsh you have turned around many of my original texts and really added scientific value to my research.

Patricia, you were a witness of my coming out as a researcher. During one of the then weekly #medlibs twitter chat nights in early 2012 (which sadly for me ran at 3 AM for my time zone) we discussed the Gehanno article and developed the plan to perform a follow up research and advised me to contact Dean for that. Though you ultimately only co-authored one of the papers in the thesis your infinite energy is evident when I see you presenting 4 posters and 3 presentations at MLA. Thank you for your inspiration and versatility.

Een aantal naaste collega's hebben hun sporen nagelaten in dit proefschrift.

Louis, veel van wat ik weet over PubMed en EndNote heb ik van jou geleerd. Jouw pensionering vroeg van ons een efficiency slag die de aanzet was tot de verbeterde zoekmethode, en de aanpassingen aan EndNote die jij hebt ingezet, vormden de aanzet voor een aantal hoofdstukken in dit boekje. Dank je wel voor de inspiratie.

Gerdien, je hebt aan de wieg gestaan van de methoden die in dit proefschrift zijn beschreven. Dit jaar ben ook jij Louis achterna gegaan. Ik heb het je niet altijd even gemakkelijk gemaakt als we samen een cursus presenteerden en onze aanpak nogal verschilde. Sorry daarvoor. Dank voor het samen sparren op lastige onderzoeksvragen en het verder aanscherpen van de methoden.

Many others have cooperated with me on various projects that ultimately did or did not end up in this thesis. Thank you: **Agnes, Arthur, Bianca, Carolien, David, Eric, Floris, Frederick, Hans, Jacqueline, Jelena, Judith, Leslie, Patrick, Paul, Tanja** and **Wendy**.

Ferry, wat had ik het je gegund om mijn proefschrift in te voeren in je jarenlang zorgvuldig bijgehouden collectie. Helaas heb je deze dag niet meer mee mogen maken, maar daarnaast is je verzameling proefschriften na de reorganisatie van 2017 ook niet meer wat het eerder was. Rust zacht!

Oud-collega's **Hans, Nico, Petra, Wim V. en Youssou** bedankt voor de heerlijke gezelligheid op de oude gang. Gelukkig hebben jullie allemaal inmiddels jullie draai gevonden. **Wim M**. en **Christa**, gelukkig zijn er nog mensen van de oude garde als kamergenoot overgebleven, bedankt voor de dagelijkse gezelligheid. **Sabrina, Elise** en **Maarten** bedankt dat jullie proefkonijn wilden zijn voor mijn test hoe goed de methode is over te dragen op "leken". Dat blijkt wonderwel gelukt. Binnen het jaar begonnen de acknowledgements van systematic reviews binnen te rollen.

Systematic searches are generally being performed not for the librarian themselves but for their clientele. If it were not for the demanding faculty and students at Erasmus MC with their large variety of research questions the method described here had not been as necessary as it is for us. It is impossible to mention each client by name that I have helped in the last 10 years, therefore I will limit to the authors of articles in which I have co-authored (or started a not finished project with) and the clients that regularly returned: **Ali, Alina, Alison, Alwine, Anne, Annelies, Arianne, Audrey, Ayesha, Barry, Beth, Caroline, Christina, Corinne, Debora, Deepak, Edeltraud, Eline, Elisabeth, Els, Emeline, Epke, Erwin, Esmee, Faiz, Ferdows, Franka, Frederieke, Herjan, Hester, Ivo, Jan-Willem, Jasper, Jeff, Jiaye, Juan, Judith, Julia, Ke-Xin, Kim, Kimberly, Kirtie, Klaas, Kosei, Kris, Layal, Lennard, Loes, Lotte, Lottie, Luis, Magda, Maria, Marianne, Marie-Christine, Marija, Mark, Max, Mohamad, Monique, Myrthe, Nico, Niels, Nienke, Onno, Osemeke, Paola, Petek, Qin, Rintje, Ruud, Sanne, Taulant, Tim, Trudy, Valentina, Victor, Vincent, Winnifred, Zeynep** and **Zsuzsanna.**

Een tweetal personen hebben mij (tegen een zeer schappelijke vergoeding) geholpen bij het opstellen en afmaken van het boekje. Ik bedank hier met name **Sally** voor je zeer uitgebreide revisie van de introductie en discussie (ik moet toch nog veel leren) en **Astrid** voor de professionele opmaak.

Verder iedereen die ik vergeten ben. Zoals ik zo vaak na afloop van een zoekactie zeg, vergeef het me als ik je naam en gezicht niet meer weet. Het spiegeltje naast mijn computer is er nog steeds niet gekomen.

Lieve **Marnix** en **Isabell**, bedankt voor het niet doorvragen naar een datum voor mijn promotie (een beginners fout die ik zelf jaren geleden maakte). Jullie zijn mij al voorgegaan, Isabell zelfs al met op deze plek de uitdaging dat de beurt nu aan mij was. Ben je zo tevreden? Heerlijk om te zien hoe jullie genieten als gezin met **Marit** en **Merle** in jullie hart.

Lieve **Hans** en **Agnès**, bedankt voor jullie niet aflatende steun en interesse. Bedankt voor alle oppasmomenten, meestal afgerond met die spaghetti van Menetti en van oma.

Lieve **BraBorBrus**: **Elsbeth** en **Arjen, Marjon** en **Hans,** en **Jan** en **Lianne**, bedankt voor jullie interesse en meeleven (ook Marjon weet wat het is om te promoveren). Door de afstand zien we elkaar helaas niet heel vaak, maar dat neemt niet weg dat we, door alles wat we hebben meegemaakt een hechte band hebben. Het is altijd heerlijk om jullie te spreken en onze kinderen met **Jim, Thije, Stijn, Yve** en **Veerle** te zien spelen.

Lieve **pap** en **mam**. Opvoeden gaat met ups en downs, dat heb ik zelf ook geleerd. In het verleden hebben we genoeg downs gehad. Toen hadden we niet verwacht hier vandaag te staan. Laten we vanaf nu voor alleen maar ups gaan. Bedankt voor jullie onvoorwaardelijke hulp en liefde.

De belangrijkste persoon die promoveren mogelijk maakt is je levenspartner. Lieve **Saskia**, maatje, liefste, alles. Ik zei het jaren geleden al, als ik weer eens tot 2 uur 's nachts had gewerkt: "het lijkt wel promoveren". Nu is het dan afgerond, maar zonder jou had ik het niet gekund. Ik beloof je dat ik niet meer (zo vaak) tot 2 uur 's nachts zal werken. Ik beloof je dat ik weer met meer dan een half oor zal luisteren als je over vioolles praat. Ik beloof je dat ik weer meer aan mijn gezondheid zal denken (minder en vegetarisch eten). En ik beloof je jou altijd zo te steunen als je mij de afgelopen 13 jaar hebt gesteund.

Tot slot de belangrijkste personen in mijn leven. Lieve **Joris** en **Thirza**, nu is papa's boekje eindelijk klaar. Nu is er weer tijd voor leuke dingen. Nu kan ik altijd gewoon weer mee naar paardrijden, schermen, zwemmen, zangles en accordeonles. Kunnen we eindelijk dat draaiorgel afbouwen waar we jaren geleden aan begonnen, of die boot, of dat vliegtuig, of wat jullie ook maar in jullie geweldige hoofd halen. Dank jullie wel voor jullie heerlijke energie en ontzettend dikke knuffels.

PhD portfolio

Courses and Workshops		
2016	Biomedical English Writing and Communication	1.0
	Workshop: Dealing with groups	0.1
2015	Writing for Publication: Getting Started, Getting Help and Getting Published	0.2
Attended conferences		
2019	European Association of Health Information and Libraries, Basel, CH	1.0
	Medical Library Association, Chicago, US	1.0
2018	Medical Library Association, Atlanta, US	1.0
2017	European Association of Health Information and Libraries, Dublin, IR	1.0
	Medical Library Association, Seattle, US	1.0
2016	European Association of Health Information and Libraries, Sevilla, ES	1.0
	Medical Library Association, Toronto, CA	1.0
2015	European Association of Health Information and Libraries, Edinburg, UK	1.0
	Medical Library Association, Austin, US	1.0
	Arbeitsgemeinschaft für Medizinisches Bibliothekswesen, Basel, CH	0.5
Teaching		
2015-2019	Systematic Literature Retrieval in PubMed, Erasmus MC	10.0
	Systematic Literature Retrieval in Embase, Erasmus MC	5.0
	EndNote, Erasmus MC	5.0
2018-2019	Systematic reviews, Erasmus MC	2.0
2019	Systematic literature review: Basic & Beyond, EACTS guidelines workshop, London, UK	0.2
	Systematic literature research in multiple databases, Commissie Bij- en Nascholing, Rotterdam	0.2
2018	Stop muddling through the databases - find relevant research literature faster, Bern, CH	0.5
	Systematic literature research in multiple databases, Commissie Bij- en Nascholing, Rotterdam	0.2
2017	Systematic Literature retrieval in PubMed, Bogota, CO	0.4
	Searching bibliographic databases efficiently and systematically (VOGIN lezing), Amsterdam	0.2
2016	Efficiently searching in biomedical databases : improving speed and accuracy, Stockholm, SE	0.2
	Searching bibliographic databases efficiently and systematically, Nieuwegein	0.2
	Embase.com for biomedical information specialists, Rotterdam	0.2
	Systematic searching and the use of EndNote for systematic reviews, New Delhi, IN	1.0
	A unique method for fast, high-quality systematic searching, Webinar Embase	0.2
	Adapting EndNote for use in a large Academic Hospital, Amsterdam	0.2
Other		
2015-2019	Peer review of articles for scientific journals	2.0
	Member of the editorial board of Journal of the Medical Library Association	0.5

List of publications

Original research

Rietjens JA, **Bramer WM**, Geijteman EC, van der Heide A, Oldenmenger WH. Development and validation of search filters to find articles on palliative care in bibliographic databases. Palliat Med; 2019, 33(4): 470-474

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 Nov 20;7(1):200

Pols D, **Bramer W**, Bindels P, van de Laar F, Bohnen A. Nieuwe zoekfilters voor het vinden van relevante publicaties. Huisarts en Wetenschap. 2018 Aug.

Bramer WM, de Jonge GB, Rethlefsen ML, Mast F, Kleijnen J. A systematic approach to searching: How to perform high quality searches for systematic reviews more efficiently. J Med Libr Assoc. 2017;105(4):531-541.

Bramer WM. Reference Checking For Systematic Reviews Using Endnote. J Med Libr Assoc. 2017;105(4):542-546.

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.

Bramer WM, Rethlefsen ML, Mast F, Kleijnen J. Evaluation of a new method for librarian-mediated literature searches for systematic reviews. Res Synth Methods. 2017 Oct 26.

Bramer W, Bain P. Updating search strategies for systematic reviews using EndNote. Journal of the Medical Library Association : J Med Libr Assoc. 2017;105(3):285-9.

Bramer WM, Milic J, Mast F. Reviewing retrieved references for inclusion in systematic reviews using EndNote. J Med Libr Assoc. 2017 Jan;105(1):84-7.

Bramer WM, Giustini D, Kramer BMR. Comparing the coverage, recall, and precision of searches for 120 systematic reviews in Embase, MEDLINE, and Google Scholar: a prospective study. Syst Rev. 2016;5(1):39.

Bramer WM, Giustini D, de Jonge GB, Holland L, Bekhuis T. De-duplication of database search results for systematic reviews in EndNote. J Med Libr Assoc. 2016 Jul;104(3):240-3.

Bramer WM. Variation in number of hits for complex searches in Google Scholar. J Med Libr Assoc. 2016 Apr;104(2):143-5.

Thielen FW, Van Mastrigt G, Burgers LT, **Bramer WM,** Majoie H, Evers S, Kleijnen J. How to prepare a systematic review of economic evaluations for clinical practice guidelines: database selection and search strategy development (part 2/3). Expert Rev Pharmacoecon Outcomes Res. 2016 Nov 2:1-17.

Bramer WM. Evaluation of instructive texts on searching medical databases. J Med Libr Assoc. 2015 Oct;103(4):208-9.

Pols DH, **Bramer WM**, Bindels PJ, van de Laar FA, Bohnen AM. Development and Validation of Search Filters to Identify Articles on Family Medicine in Online Medical Databases. Ann Fam Med. 2015 Jul;13(4):364-6.

Bramer WM, Giustini D, Kramer BMR, Anderson PF. 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

Letters and Columns

Bramer WM. Expert Searching: Why Don't Simple Searches [Always] Work? – Column. MLA news. 2017;57(150).

Bramer WM. Expert searching – Patient, Intervention, Comparison, Outcome (PICO) – an overrated tool – Column. MLA news 2015;55(2):6.

Bramer WM. Expert searching – Searching for Systematic Reviews: Set Numbers or One Liners? – Column. MLA news 2014;54(4):15.

Bramer WM, Ket JC, Kramer BM. Contradictory results of a more extensive search on cruciferous vegetables and breast cancer – Letter to the editor. Breast. 2012;21(6):783.

Articles about the research in this thesis

van Houts I. Cursus systematisch literatuuronderzoek vooral heel praktisch (Workshop systematic literature retrieval above all very practical). IP - vakblad voor informatieprofessionals 2018 (5):33.

de Nijs R. Biomedisch informatie specialisten in de spotlights. IP - vakblad voor informatieprofessionals 2018 (6):24-27.

Sieverts E. Do it yourself: Systematische zoeker (Systematic searcher). IP - vakblad voor informatieprofessionals 2016 (8):5.

Conference presentations, posters and workshops

Bramer WM, Rethlefsen ML Effectiveness and Efficiency in Exhaustive Searches. Workshop given at Medical Library Association - Chicago, IL (US), 2019

Bramer WM, de Jonge G Effectiveness and Efficiency in Exhaustive Searches. Workshop given at European Association for Health Information and Libraries (EAHIL) - Basel (CH), 2019

Bramer WM, de Jonge G. EndNote – More than a reference tool. Workshop given at European Association for Health Information and Libraries (EAHIL) - Basel (CH), 2019

Bramer WM. Systematic literature review: Basic & Beyond - Searching for the evidence. Presentation at EACTS Guidelines Workshop –Windsor (UK), 2019

Bramer WM. EndNote, more than a reference tool. Presentation at Medical Library Association - Chicago, IL (US), 2019

Bramer WM, de Jonge GB Systematic literature research in multiple databases. Workshop for Commissie Bij- en Nascholing – Rotterdam, 2019

Bramer WM, Rethlefsen ML. Effectiveness and Efficiency in Exhaustive Searches. Workshop given at Medical Library Association - Atlanta, GA (US), 2018

Bramer WM. Stop muddling through the databases – find relevant research literature faster! Workshop given at Institute of Social and Preventive Medicine - Universität Bern - Bern (CH), 2018

Bramer WM, de Jonge GB. Systematic literature research in multiple databases. Workshop for Commissie Bij- en Nascholing – Rotterdam, 2018

Bramer WM. Systematic Literature retrieval in PubMed. Workshop given for Fundación Universitaria de Ciencias de la Salud: FUCS – Bogota (CO), 2017

Bramer WM,de Jonge G. Improving efficiency and confidence in systematic searching through an innovative way of searching bibliographic databases. Workshop given at International Congress of Medical Librarianship (ICML) & European Association for Health Information and Libraries (EAHIL) - Dublin (IE), 2017

Kuhn I, **Bramer WM**, Foster M. Search Strategy Development - compare, contrast and advance. Workshop given at International Congress of Medical Librarianship (ICML) & European Association for Health Information and Libraries (EAHIL) - Dublin (IE), 2017

Bramer WM. Improving efficiency and confidence in systematic literature searching. Workshop given at Medical Library Association - Seattle, WA (US), 2017

Foster M, **Bramer WM**, Jewel S. Systematic Searching Live. Workshop given at Medical Library Association - Seattle, WA (US), 2017

Bramer WM, Rethlefsen ML, Giustini D, Kleijnen J, Franco OH Comparing the effectiveness of conceptual search methods: is a fast approach sufficient. for the production of (sound) systematic reviews: a prospective, double-blinded, controlled study. Presentation at International Congress of Medical Librarianship (ICML) & European Association for Health Information and Libraries (EAHIL) - Dublin, (IE) 2017

Bramer WM, Rethlefsen ML, Giustini D, Kleijnen J, Franco OH. Comparing the effectiveness of conceptual search methods: is a fast approach sufficient for the production of (sound) systematic reviews: a prospective, double-blinded, controlled study. Presentation at Medical Library Association - Seattle, WA (US), 2017

Bramer WM, Giustini D, Kleijnen J, Franco OH. Searching major descriptors or words in titles and abstracts only: the likelihood of losing relevant references in systematic reviews. Poster presented at International Congress of Medical Librarianship (ICML) & European Association for Health Information and Libraries (EAHIL) - Dublin, (IE), 2017

Bramer WM, Giustini D, Kleijnen J, Franco OH. Searching major descriptors or words in titles and abstracts only: the likelihood of losing relevant references in systematic reviews. Poster presented at Medical Library Association - Seattle , WA (US), 2017

Bramer WM. Searching bibliographic databases efficiently and systematically. Workshop at VOGIN-IP Lezing – Amsterdam, 2017

Bramer WM, de Jonge G. Improving efficiency and confidence in systematic searching. Workshop given at European Association for Health Information and Libraries (EAHIL) – Sevilla (ES), 2016

Bramer WM. Improving efficiency and confidence in systematic literature searching. Workshop given at Medical Library Association / Canadian Health Library Association – Toronto (CA), 2016

Bramer WM. The added value of searching multiple databases in searching for systematic reviews: a prospective study. Presentation at European Association for Health Information and Libraries Conference – Sevilla (ES), 2016

Bramer WM, Mast F, Kleijnen J. Librarian-mediated literature searches for systematic reviews at Erasmus MC : Much faster but with equal quality. Presentation at European Association for Health Information and Libraries Conference – Sevilla (ES), 2016

Bramer WM. The Added Value of Multiple Databases in Searching for Exhaustiveness : A Prospective Study. Presentation at Medical Library Association, the Canadian Health Libraries Association (CHLA) International Clinical Librarian Conference (ICLC) Mosaic '16 - Toronto (CA), 2016

Bramer WM. Efficiently searching in biomedical databases : improving speed and accuracy. Presentation for Swedish Biomedical Information Specialists – Karolinska Institutet, Stockholm (SE), 2016

Bramer WM. Optimization of Systematic Review Searches : When Can You Stop? Poster presented at Medical Library Association (MLA), the Canadian Health Libraries Association (CHLA) International Clinical Librarian Conference (ICLC) Mosaic '16, Toronto (CA), 2016

Bramer WM, Rethlefsen ML, Giustini D. Comparing the Effectiveness of Conceptual Search Methods: Is a Fast Approach Sufficient for the Production of (sound) Systematic Reviews: A Prospective, Double-blinded, Controlled Study. Poster presented at Medical Library Association (MLA), the Canadian Health Libraries Association (CHLA) International Clinical Librarian Conference (ICLC) Mosaic '16, Toronto (CA), 2016

Bramer WM. Searching bibliographic databases efficiently and systematically. Workshop at Koninklijke Nederlandse Vereniging van Informatiespecialisten – Nieuwegein, 2016

Bramer WM, de Jonge GB. Embase.com for biomedical information specialists. Workshop for Commissie bij- en nascholing – KNVI / BMI – Rotterdam, 2016

Bramer WM. Systematic searching and the use of EndNote for systematic reviews. Web-training and mentorship for CHEORS (Complete Health Economics Outcomes Research Solutions) – New Delhi (IN), 2015

Bramer WM. Efficient searching methods demonstrated - Live searches on participants' research questions. Workshop given at Arbeitsgemeinschaft für Medizinisches Bibliothekswesen – Basel (CH), 2015

Bramer WM, de Jonge G. Improving efficiency and confidence in systematic literature searching. Workshop given at European Association for Health Information and Libraries 2015 – Edinburgh (UK), 2015

Bramer WM. Sunrise seminar: A unique method for fast, high-quality systematic searching. Workshop given at Medical Library Association – Austin TX (US), 2015

Bramer WM. A unique method for fast, high-quality systematic searching. Webinar presented for Embase.com, 2015

Bramer WM, de Jonge GB, Mast F. Efficiently Searching for Systematic Reviews : How to Perform High Quality Searches More Efficiently. Presentation at 23rd Cochrane Colloquium – Vienna (AU), 2015

Bramer WM. Efficiently Searching for Systematic Reviews. Presentation at Arbeitsgemeinschaft für Medizinisches Bibliothekswesen – Basel (CH), 2015

Bramer WM, de Jonge GB, Mast F. Efficiently Searching for Systematic Reviews : How to Perform High Quality Searches More Efficiently. Presentation at Medical Library Association – Austin TX (US), 2015

Bramer WM. Adapting EndNote for use in a large Academic Hospital. Workshop for Dutch EndNote users – Amsterdam, 2015

Bramer WM. Systematic Searching for Medical Information Specialists. Invited workshop for Swiss Biomedical Information Specialists – Centre Hospitalier Universitaire Vaudois, Lausanne (CH), 2014

Bramer WM, Holland L, Mollema J, Hannon T, Bekhuis T. Removing duplicates in retrieval sets from electronic databases : comparing the efficiency and accuracy of the Bramer-method with other methods and software packages. Presentation at European Association for Health Information and Libraries Conference – Rome (IT), 2014

Bramer WM, de Jonge GB, Mast F. Efficiently searching for systematic reviews : Quick search optimization and semi-automatic translation between multiple databases. Poster presented at European Association for Health Information and Libraries Conference – Rome (IT), 2014

Bramer WM. Tips and tricks on systematic searching in embase.com. Workshop for Working group Biomedical Information (BMI) of the Royal Dutch Society of Information professionals – Leiden, 2014

Bramer WM. Systematic Searching for Medical Information Specialists. Workshop for Medical information specialists VU and AMC Amsterdam, 2014

Bramer WM. Education in Literature searches: Do it yourself or teach others how to do it? Presentation at Working group Electronic Sources and Searching – Rotterdam, 2012

Bramer WM. Scientific output on Erasmus MC website and the use of XML. Presentation at Emerging Technologies in Academic Libraries (EMTACL) - Trondheim (NO), 2010

Publications co-authored as search coordinator

Klompenhouwer AJ, Ten Cate DWG, Willemssen F, **Bramer WM**, Doukas M, de Man RA, Ijzermans JNM. The impact of imaging on the surgical management of biliary cystadenomas and cystadenocarcinomas; a systematic review. HPB (Oxford). 2019 May 10.

Liu J, Ma B, Cao W, Li M, **Bramer WM**, Peppelenbosch MP, Pan Q. Direct-acting antiviral agents for liver transplant recipients with recurrent genotype 1 hepatitis C virus infection: systematic review and meta-analysis. Transpl Infect Dis. 2019 Jan 7:e13047.

Yang Q, Liu J, Ma W, Wang J, Li F, **Bramer WM**, Peppelenbosch MP, Pan Q. Efficacy of Different Endoscopic Stents in the Management of Postoperative Biliary Strictures: A Systematic Review and Meta-analysis. J Clin Gastroenterol. 2019 Feb 22.

Dall'Aglio L, Muka T, Cecil CAM, **Bramer WM**, Verbiest M, Nano J, Hidalgo AC, Franco OH, Tiemeier H. The Role of Epigenetic Modifications in Neurodevelopmental Disorders: A Systematic Review. Neurosci Biobehav Rev. 2018 Jul 29.

Glisic M, Kastrati N, Gonzalez-Jaramillo V, **Bramer WM**, Ahmadizar F, Chowdhury R, Danser AJ, Roks AJ, Voortman T, Franco OH, Muka T. Associations between Phytoestrogens, Glucose Homeostasis, and Risk of Diabetes in Women: A Systematic Review and Meta-Analysis. Adv Nutr. 2018 Nov 1;9(6):726-40.

Glisic M, Kastrati N, Musa J, Milic J, Asllanaj E, Portilla Fernandez E, Nano J, Ochoa Rosales C, Amiri M, Kraja B, Bano A, **Bramer WM**, Roks AJM, Danser AHJ, Franco OH, Muka T. Phytoestrogen supplementation and body composition in postmenopausal women: A systematic review and meta-analysis of randomized controlled trials. Maturitas. 2018 Sep;115:74-83.

Pastoor H, Timman R, de Klerk C, **Bramer WM**, Laan ET, Laven JS. Sexual function in women with polycystic ovary syndrome: a systematic review and meta-analysis. Reprod Biomed Online. 2018 Oct 19.

Rojas LZ, Glisic M, Pletsch-Borba L, Echeverria LE, **Bramer WM**, Bano A, Stringa N, Zaciragic A, Kraja B, Asllanaj E, Chowdhury R, Morillo CA, Rueda-Ochoa OL, Franco OH, Muka T. Electrocardiographic abnormalities in Chagas disease in the general population: A systematic review and meta-analysis. PLoS Negl Trop Dis. 2018 Jun;12(6):e0006567.

Tuenter A, Bautista Nino PK, Vitezova A, Pantavos A, **Bramer WM**, Franco OH, Felix JF. Folate, vitamin B12, and homocysteine in smoking-exposed pregnant women: A systematic review. Matern Child Nutr. 2018 Sep 4:e12675.

Ultee KHJ, Buttner S, Huurman R, Bastos Goncalves F, Hoeks SE, **Bramer W**M, Schermerhorn ML, Verhagen HJM. Systematic Review and Meta-Analysis of the Outcome of Treatment for Type II Endoleak Following Endovascular Aneurysm Repair. Eur J Vasc Endovasc Surg. 2018 Aug 10.

Wieten E, Schreuders EH, Grobbee EJ, Nieboer D, **Bramer WM**, Lansdorp-Vogelaar I, Bruno MJ, Kuipers EJ, Spaander MCW. Incidence of faecal occult blood test interval cancers in population-based colorectal cancer screening: a systematic review and meta-analysis. Gut. 2018 Jun 22.

Balak DM, Hajdarbegovic E, **Bramer WM**, Neumann MH, Thio HB. Progressive multifocal leukoencephalopathy associated with fumaric acid esters treatment in psoriasis patients. J Eur Acad Dermatol Venereol. 2017 Mar 21.

Cepeda M, Schoufour J, Freak-Poli R, Koolhaas CM, Dhana K, **Bramer WM**, Franco OH. Levels of ambient air pollution according to mode of transport: a systematic review. Lancet Public Health. 2017 Jan;2(1):e23-e34.

Groen JW, Grosfeld S, **Bramer WM**, Ernst MF, Mullender MM. Cyclic and non-cyclic breast-pain: A systematic review on pain reduction, side effects, and quality of life for various treatments. Eur J Obstet Gynecol Reprod Biol. 2017 Dec;219:74-93.

Hakim MS, Wang W, **Bramer WM**, Geng J, Huang F, de Man RA, Peppelenbosch MP, Pan Q. The global burden of hepatitis E outbreaks: a systematic review. Liver international : official journal of the International Association for the Study of the Liver. 2017 Jan;37(1):19-31.

Janssen ML, van Broekhoven DL, Cates JM, **Bramer WM**, Nuyttens JJ, Gronchi A, Salas S, Bonvalot S, Grunhagen DJ, Verhoef C. Meta-analysis of the influence of surgical margin and adjuvant radiotherapy on local recurrence after resection of sporadic desmoid-type fibromatosis. Br J Surg. 2017 Mar;104(4):347-57.

Kis Z, Muka T, Franco OH, **Bramer WM**, De Vries LJ, Kardos A, Szili-Torok T. The Short and Long-Term Efficacy of Pulmonary Vein Isolation as a Sole Treatment Strategy for Paroxysmal Atrial Fibrillation: A Systematic Review and Meta-Analysis. Curr Cardiol Rev. 2017 Jan 17;13(3):199-208.

Klompenhouwer AJ, Verver D, Janki S, **Bramer WM**, Doukas M, Dwarkasing RS, de Man RA, IJzermans J. Management of hepatic angiomyolipoma: A systematic review. Liver Int. 2017 Sep;37(9):1272-80.

Moraes AAI, Baena CP, Muka T, Bano A, Buitrago-Lopez A, Zazula A, Erbano BO, Schio NA, Guedes MH, **Bramer WM**, Franco OH, Faria-Neto JR. Achieved systolic blood pressure in older people: a systematic review and meta-analysis. BMC Geriatr. 2017 Dec 5;17(1):279.

Muka T, Nano J, Jaspers L, Meun C, **Bramer WM**, Hofman A, Dehghan A, Kavousi M, Laven JS, Franco OH. Associations of Steroid Sex Hormones and Sex Hormone-Binding Globulin With the Risk of Type 2 Diabetes in Women: A Population-Based Cohort Study and Meta-analysis. Diabetes. 2017 Mar;66(3):577-86.

Te Veldhuis EC, Te Veldhuis AH, **Bramer WM**, Wolvius EB, Koudstaal MJ. The effect of orthognathic surgery on the temporomandibular joint and oral function: a systematic review. Int J Oral Maxillofac Surg. 2017 May;46(5):554-63.

Ambagtsheer F, de Jong J, **Bramer WM**, Weimar W. On patients who purchase organ transplants abroad. American journal of transplantation : official journal of the American Society of Transplantation and the American Society of Transplant Surgeons. 2016 Mar 1.

Balak DM, Fallah Arani S, Hajdarbegovic E, Hagemans CA, **Bramer WM**, Thio HB, Neumann HA. Efficacy, effectiveness and safety of fumaric acid esters in the treatment of psoriasis: a systematic review of randomized and observational studies. Br J Dermatol. 2016 Aug;175(2):250-62.

Braun KV, Voortman T, Dhana K, Troup J, **Bramer WM**, Troup J, Chowdhury R, Dehghan A, Muka T, Franco OH. The role of DNA methylation in dyslipidaemia: A systematic review. Prog Lipid Res. 2016 Oct 13;64:178-91.

de Jong MH, Kamperman AM, Oorschot M, Priebe S, **Bramer W**, van de Sande R, Van Gool AR, Mulder CL. Interventions to Reduce Compulsory Psychiatric Admissions: A Systematic Review and Meta-analysis. JAMA Psychiatry. 2016 Jul 1;73(7):657-64.

Garcia AH, Voortman T, Baena CP, Chowdhurry R, Muka T, Jaspers L, Warnakula S, Tielemans MJ, Troup J, **Bramer WM**, Franco OH, van den Hooven EH. Maternal weight status, diet, and supplement use as determinants of breastfeeding and complementary feeding: a systematic review and meta-analysis. Nutr Rev. 2016 Aug;74(8):490-516.

Jaspers L, Feys F, **Bramer WM**, Franco OH, Leusink P, Laan ET. Efficacy and Safety of Flibanserin for the Treatment of Hypoactive Sexual Desire Disorder in Women: A Systematic Review and Meta-analysis. JAMA Intern Med. 2016 Apr 1;176(4):453-62.

Karim F, Loeffen J, **Bramer W**, Westenberg L, Verdijk R, van Hagen M, van Laar J. IgG4-related disease: a systematic review of this unrecognized disease in pediatrics. Pediatr Rheumatol Online J. 2016;14(1):18.
Leermakers ET, Darweesh SK, Baena CP, Moreira EM, Melo van Lent D, Tielemans MJ, Muka T, Vitezova A,

Chowdhury R, **Bramer WM**, Kiefte-de Jong JC, Felix JF, Franco OH. The effects of lutein on cardiometabolic health across the life course: a systematic review and meta-analysis. The American journal of clinical nutrition. 2016 Feb;103(2):481-94.

Maas NM, van der Grinten M, **Bramer WM**, Kleinrensink GJ. Metatarsophalangeal joint stability: a systematic review on the plantar plate of the lesser toes. J Foot Ankle Res. 2016;9:32.

Melo van Lent D, Leermakers ETM, Darweesh SKL, Moreira EM, Tielemans MJ, Muka T, Vitezova A, Chowdhury R, **Bramer WM**, Brusselle GG, Felix JF, Kiefte-de Jong JC, Franco OH. The effects of lutein on respiratory health across the life course: A systematic review. Clinical Nutrition ESPEN. 2016;13:e1-e7.

Muka T, Koromani F, Portilla E, O'Connor A, **Bramer WM**, Troup J, Chowdhury R, Dehghan A, Franco OH. The role of epigenetic modifications in cardiovascular disease: A systematic review. Int J Cardiol. 2016 Jun 1;212:174-83.

Muka T, Nano J, Voortman T, Braun KV, Ligthart S, Stranges S, **Bramer WM**, Troup J, Chowdhury R, Dehghan A, Franco OH. The role of global and regional DNA methylation and histone modifications in glycemic traits and type 2 diabetes: A systematic review. Nutr Metab Cardiovasc Dis. 2016 Jul;26(7):553-66.

Muka T, Vargas KG, Jaspers L, Wen KX, Dhana K, Vitezova A, Nano J, Brahimaj A, Colpani V, Bano A, Kraja B, Zaciragic A, **Bramer WM**, van Dijk GM, Kavousi M, Franco OH. Estrogen receptor beta actions in the female cardiovascular system: A systematic review of animal and human studies. Maturitas. 2016 Apr;86:28-43.

Osokogu OU, Dukanovic J, Ferrajolo C, Dodd C, Pacurariu AC, **Bramer WM**, tJong G, Weibel D, Sturkenboom MC, Kaguelidou F. Pharmacoepidemiological safety studies in children: a systematic review. Pharmacoepidemiology and drug safety. 2016 Aug;25(8):861-70.

Taneri PE, Kiefte-de Jong JC, **Bramer WM**, Daan NM, Franco OH, Muka T. Association of alcohol consumption with the onset of natural menopause: a systematic review and meta-analysis. Human reproduction update. 2016 Jun;22(4):516-28.

Tielemans MJ, Garcia AH, Peralta Santos A, **Bramer WM**, Luksa N, Luvizotto MJ, Moreira E, Topi G, de Jonge EA, Visser TL, Voortman T, Felix JF, Steegers EA, Kiefte-de Jong JC, Franco OH. Macronutrient composition and gestational weight gain: a systematic review. The American journal of clinical nutrition. 2016 Jan;103(1):83-99.

Vargas KG, Milic J, Zaciragic A, Wen KX, Jaspers L, Nano J, Dhana K, **Bramer WM**, Kraja B, van Beeck E, Ikram MA, Muka T, Franco OH. The functions of estrogen receptor beta in the female brain: A systematic review. Maturitas. 2016 Jun 4.

Wen KX, Milic J, El-Khodor B, Dhana K, Nano J, Pulido T, Kraja B, Zaciragic A, **Bramer WM**, Troup J, Chowdhury R, Ikram MA, Dehghan A, Muka T, Franco OH. The Role of DNA Methylation and Histone Modifications in Neurodegenerative Diseases: A Systematic Review. PLoS One. 2016;11(12):e0167201.

Chaker L, Falla A, van der Lee SJ, Muka T, Imo D, Jaspers L, Colpani V, Mendis S, Chowdhury R, **Bramer WM**, Pazoki R, Franco OH. The global impact of non-communicable diseases on macro-economic productivity: a systematic review. Eur J Epidemiol. 2015 May;30(5):357-95.

De Lima A, Galjart B, Wisse PH, **Bramer WM**, van der Woude CJ. Does lower gastrointestinal endoscopy during pregnancy pose a risk for mother and child? - a systematic review. BMC Gastroenterol. 2015;15(1):15.

Harrison SL, Sajjad A, **Bramer WM**, Ikram MA, Tiemeier H, Stephan BC. Exploring strategies to operationalize cognitive reserve: A systematic review of reviews. Journal of clinical and experimental neuropsychology. 2015 Mar 9;37(3):253-64.

Jaspers L, Colpani V, Chaker L, van der Lee SJ, Muka T, Imo D, Mendis S, Chowdhury R, **Bramer WM**, Falla A, Pazoki R, Franco OH. The global impact of non-communicable diseases on households and impoverishment: a systematic review. Eur J Epidemiol. 2015 Mar;30(3):163-88.

Leermakers ET, Moreira EM, Kiefte-de Jong JC, Darweesh SK, Visser T, Voortman T, Bautista PK, Chowdhury R, Gorman D, **Bramer WM**, Felix JF, Franco OH. Effects of choline on health across the life course: a systematic review. Nutr Rev. 2015 Aug;73(8):500-22.

Li J, Hernanda PY, **Bramer WM**, Peppelenbosch MP, van Luijk J, Pan Q. Anti-Tumor Effects of Metformin in Animal Models of Hepatocellular Carcinoma: A Systematic Review and Meta-Analysis. PLoS One. 2015;10(6):e0127967.

Luk F, de Witte SF, **Bramer WM**, Baan CC, Hoogduijn MJ. Efficacy of immunotherapy with mesenchymal stem cells in man: a systematic review. Expert review of clinical immunology. 2015 Mar 27;11(5):617-36.

Malfliet A, Kregel J, Cagnie B, Kuipers M, Dolphens M, Roussel N, Meeus M, Danneels L, **Bramer WM**, Nijs J. Lack of Evidence for Central Sensitization in Idiopathic, Non-Traumatic Neck Pain: A Systematic Review. Pain physician. 2015 May-Jun;18(3):223-36.

Muka T, Imo D, Jaspers L, Colpani V, Chaker L, van der Lee SJ, Mendis S, Chowdhury R, **Bramer WM**, Falla A, Pazoki R, Franco OH. The global impact of non-communicable diseases on healthcare spending and national income: a systematic review. Eur J Epidemiol. 2015 Apr;30(4):251-77.

Osokogu OU, Fregonese F, Ferrajolo C, Verhamme K, de Bie S, 't Jong G, Catapano M, Weibel D, Kaguelidou F, **Bramer WM**, Hsia Y, Wong ICK, Gazarian M, Bonhoeffer J, Sturkenboom M. Pediatric Drug Safety Signal Detection: A New Drug-Event Reference Set for Performance Testing of Data-Mining Methods and Systems. Drug Safety. 2015 Feb;38(2):207-17.

Voortman T, van den Hooven EH, Braun KV, van den Broek M, **Bramer WM**, Chowdhurry R, Franco OH. Effects of polyunsaturated fatty acid intake and status during pregnancy, lactation, and early childhood on cardiometabolic health: A systematic review. Prog Lipid Res. 2015 May 27;59:67-87.

Voortman T, Vitezova A, **Bramer WM**, Ars CL, Bautista PK, Buitrago-Lopez A, Felix JF, Leermakers ETM, Sajjad A, Sedaghat S, Tharner A, Franco OH, van den Hooven EH. Effects of protein intake on blood pressure, insulin sensitivity and blood lipids in children: a systematic review. Brit J Nutr. 2015 Feb 14;113(3):383-402.

Weiland A, Van de Kraats RE, Blankenstein AH, Van Saase JL, Van der Molen HT, **Bramer WM**, Van Dulmen AM, Arends LR. Encounters between medical specialists and patients with medically unexplained physical symptoms; influences of communication on patient outcomes and use of health care: a literature overview. Perspectives on medical education. 2012 Nov;1(4):192-206.

About the author

Wichor Bramer was born in Vriezenveen in 1974, but grew up in Bergentheim. He attended the secondary school in Hardenberg at Christelijke Scholengemeenschap Jan van Arkel. He studied biology at rijksuniversiteit Groningen. After several years he made the switch to study Information Services and Management at Saxion Hogeschool in Deventer, from which he graduated in 2002.

Since then he has worked as an information specialist in several health care related organizations in the Netherlands. In 2009 he started as biomedical information specialist at the Medical Library of Erasmus MC. His main tasks are to create exhaustive searches for systematic reviews, and to develop and conduct workshops for PhD and master students on systematic searching in various databases.

At Erasmus MC, with the help of his colleagues, he has developed a method that helps him create exhaustive and high quality searches in a much shorter time than usual, and a method for fast but accurate deduplication of obtained references. This allows him to design, perform, execute, and import into reference software and deduplicate the results of the searches for over 250 systematic reviews on a yearly basis. After the publication of his first scientific article, besides his daily job and the plans for several more, he sought the help of promotor Jos Kleijnen in November 2015.

In his spare time Wichor manages the website www.stationsweb.nl, dedicated to photographs of Dutch (former) railway stations. He plays the organ and piano, is a singer and occasional conductor of choirs and likes to swim, run and ride a bike.

Over de schrijver

Wichor Bramer is geboren in Vriezenveen in 1974, en groeide op in Bergentheim. Hij volgde de middelbare school aan Christelijke Scholengemeenschap Jan van Arkel in Hardenberg. Hij studeerde biologie aan de rijksuniversiteit Groningen. Na verschillende jaren stapte hij over naar Informatiedienstverlening en –management aan de Saxion Hogeschool in Deventer, waaraan hij afstudeerde in 2002.

Sindsdien werkte hij als in informatiespecialist in verschillende gezondheidszorg gerelateerde organisaties in Nederland. In 2009 begon hij als biomedisch informatiespecialist bij de Medische Bibliotheek van het Erasmus MC. Hij houdt zich daar bezig met het creëren van zoekacties voor systematic reviews, en met het ontwikkelen en onderwijzen van cursussen voor promovendi en studenten over systematische zoekacties in verschillende databases.

Bij het Erasmus MC ontwikkelde hij, met hulp van zijn collega's, een methode die hen helpt uitputtende zoekacties van hoge kwaliteit veel sneller dan gebruikelijk op te stellen, en hij ontwikkelde een methode om snelle maar nauwkeurige de gevonden artikelen te ontdubbelen. Hierdoor is hij in staat zoekacties voor meer dan 250 reviews per jaar te ontwerpen en uit te voeren en in referentie software te importeren en te ontdubbelen. Naast zijn dagelijks werk begon bij onderzoek te doen en artikelen te schrijven, en in november 2015 zocht bij contact met promotor Jos Kleijnen.

In zijn vrije tijd onderhoudt Wichor de website www.stationsweb.nl, gewijd aan foto's van (voormalige) Nederlandse stationsgebouwen. Hij speelt orgel en piano, zingt en is bij gelegenheid dirigent van koren en houdt van zwemmen, hardlopen en fietsen.