



An Insider's Insight into Literature Searches

Searching the literature can take various forms, ranging from a quick scan of recent publications to a formal, systematic interrogation of all available data sources to establish the scientific consensus on a specific topic. In these days of online journal databases, the relative ease of conducting a search means that they often start informally with no thought-out search strategy or defined goal. A long list of articles can be generated almost instantaneously, but what did you miss and how long will it take to review the data? How easily can the search strategy be repeated and adapted to obtain a more complete and refined set of references?

You can burn up a great deal of time and energy searching the literature. Your search's quality and value is wholly dependent on the thought and effort you put into developing your strategy. It also determines the effort and subjectivity needed to sort through the source data it produces. As searches are often an iterative process, pre-defining your goals and strategy allows you to record your starting point and review the decision tree you finally adopt. We offer some insights from the Niche medical writing team who have been conducting literature searches for their clients since 1998.



These are systematic reviews of primary research. They investigate the effects of interventions for prevention, treatment and rehabilitation of human ailments and are published in the Cochrane Library [1, 2]. The reviews are prepared and maintained by the Cochrane Collaboration, which consists of over 31,000 specialists. Cochrane reviews are renowned for including only the highest standards of evidence-based data. There are six types of Cochrane review:

1. Intervention reviews – provide assessments of interventions in healthcare
2. Diagnostic test accuracy reviews – evaluate diagnostic tests for a particular disease
3. Methodology reviews – address how reviews and trials are conducted
4. Qualitative reviews – assess evidence other than effectiveness
5. Prognosis reviews – look at the probable course or outcome of a condition
6. Overviews of Systematic Reviews – a more recently introduced type of review that compiles data from several systematic reviews into one accessible document

Search terms (defining and refining)

When defining your search terms, carefully consider the key concepts of your topic. Think about the characteristics that define your patient population and/or their conditions. For example, if you are looking at older patients with diabetes you may want to search 'diabetes' and 'elderly'. After incorporating these key terms into the search, you can begin to address the complexities of the synonyms, alternative spellings and acronyms:

- Decide on your search's date range. It may be possible that terminology has changed over the years if you have included a broad date range
- Using filters can reduce the search hits returned and increase the degree of control you have over the search. For example, you can look specifically for reviews or meta-analyses and apply filters to include only those articles published in the last 5 years or exclude 'human' if you are looking for non-clinical studies
- The truncation feature can be used to search for different variants of a root or stem word. This option can be selected by putting an asterisk (*) at the end of a stem word. For example, searching for the term 'cardio*' will result in listings of all citations that include words that can be derived from the 'cardio' stem i.e., cardiogenic, cardiology, cardiogram, etc [3]
- Decide if you want to include common and/or scientific names, abbreviations, and any synonyms in your search strategy. Increase the likelihood of retrieving relevant results by including several alternative words for elderly, such as 'senior', 'geriatric', 'older' and/or 'aged.' Remember, the more of these terms you combine, the more results are likely to be returned
- Using booleans enables the inclusion and exclusion of multiple search terms. Booleans need to be capitalised in your search.

"AND" – narrows the search to hits containing both terms

"OR" – broadens the search to hits containing either term

"NOT" – removes terms that you don't want to include

You can combine booleans, for example:

["diabetes"] AND ["elderly" OR "senior" OR "geriatric" OR "older" OR "aged"]

- Do you need to search for both British and American English variant spellings? Some search engines only use the specific word you enter, so if you want to ensure that you search for articles by American and English authors you will need to include both language variants
- Use quotation marks when you want to search for phrases or word combinations where all words appear immediately next to each other in a specific order

Formulating a well-focused question is critical to obtaining a list of relevant articles that best address your clinical question. The Cochrane Collaboration promote the use of a framework known as PICOS, which stands for patient population, interventions or exposure, comparison and outcome or endpoint and study design, during the process of developing your question. The framework splits the search terms into five components, which can be combined to formulate a search entry relevant to the research question [4, 5].

- P** Patient population. What are the defining characteristics of the participants? What condition, age, gender or setting of care are you interested in?
- I** Intervention (exposures). This may be a treatment (so consider the dose, frequency and duration), diagnostic tool, prevention or educational intervention.
- C** Comparison. Is your intervention being compared to a placebo, standard or no treatment?
- O** Outcome. Consider if your research question needs to address the improvement of a condition, or an effective diagnosis, pain reduction or improved quality of life.
- S** Study Design. Are you focussing on randomised trials, or will you include case reports and observational studies?

What are MeSH headings?

Medical Subject Headings (MeSH) make up the National Library of Medicine's controlled vocabulary thesaurus and they are used for indexing Medline articles.

The hierarchical structure of MeSH enables users to search at varying degrees of specificity [6]. Over 27,000 descriptors are included in the MeSH dictionary and the list is updated daily.

The MeSH browser is a vocabulary look-up tool, which allows terms to be inserted into boxes. This is easier than using multiple brackets when incorporating several search terms and phrases in complicated searches.

If at the start of your search you are struggling to identify appropriate MeSH headings, you can obtain suggestions using MeSH on Demand. This tool will process text to identify any contained MeSH terms. You could use text from a relevant review.

The PRISMA checklist for reporting guidelines

At least 2,500 new literature reviews are added to MEDLINE annually. A recent study of these entries found that many important aspects of the methodology for literature searches are frequently not reported. For example, half of the 300 articles that the study reviewed failed to mention the terms 'systematic review' or 'meta-analysis' in the title or abstract and only two thirds reported the range of years over which the search was performed [7].

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guideline consists of a 27-point checklist, developed to help authors standardise the reporting of literature searches – thus ensuring complete and transparent reporting. Items in the checklist include a rationale for the literature search, numbers of studies screened for eligibility and limitations that were encountered. The checklist is detailed in the PRISMA statement and website and is a helpful resource for authors to follow [8].

Write a protocol...

The utility of your output is dependent on the reproducibility of your search strategy and so it is important to minimise subjectivity. Managing a multi-factorial process benefits from a system that clearly defines as many of the variables as possible. This is best addressed in the form of an objective protocol (you can get our easy-to-use template here [9]) that outlines the brief, your proposed search strategy and criteria for review. Give the precise search terms, details of any filters and search engine(s), so that the search is reproducible.

Describe how outputs of the literature search are to be recorded or stored and what information on each citation you will collect. The protocol should also describe how you will review and score each citation. Manually reviewing the titles and/or abstracts ensures that all the results adhere to the search criteria and all literature pertaining to the topic are collected. Employing stringent methods for selecting studies will limit bias, which in turn improves the reliability and accuracy of your conclusions. In your report, detail the process of 'report' selection by giving the number of reports/articles screened, reasons for exclusion and the final numbers of articles included.

With a formal protocol you can ask your colleagues to review your methodology and rationale. This will help to craft your search strategy by identifying any missing search terms or endpoints. It could save you time when sifting through the search results to eliminate indifferent or non-relevant articles. Without a protocol it will be more difficult to elicit and capture valuable feedback or to backtrack if you find that you have reached a dead end. In developing your search strategy, you may need to include a broad set of search terms to capture all the appropriate articles you need. For instance, there might be 10 different common and scientific names for a given condition, 10 more therapies you wish to include and several animal models in which relevant work may have been conducted.

Iterative protocol development, approval and finalising search strategies

As with any other scientific endeavour, you are very unlikely to achieve success on your first attempt. To have any certainty that your goal is being achieved it is necessary to build iterative search adaptations into your methodology and include checks of your outputs at each new iteration step, comparing the 'quality' of your findings. Be prepared to search and re-search the literature, using alternative search terms. Keep a detailed record of the method you followed for each step. There is no reason not to modify your protocol to see if one strategy gives a better result. For example, your change in approach is likely to affect the ratio of relevant references retrieved to non-relevant references.

“Nothing has such power to broaden the mind as the ability to investigate systematically and truly all that comes under thy observation in life”

Marcus Aurelius c. 161–180

Sleeping beauties, snowballs and rabbits...

Seekers of knowledge always hold the hope of uncovering that 'perfect' publication that confirms all aspects of your hypothesis, provides a comprehensive assessment of the background literature and yet, has previously been overlooked by most, if not all, working in the field. Termed 'sleeping beauties,' the potential value it promises drives many authors to adopt a 'boil the oceans dry' approach to literature searching [10]. For example, authors will often perform ad hoc citation mining, where they assess references that appear in any candidate manuscripts (including existing review articles) they uncover during their searches. They may also consider including material provided in the grey literature not usually found in the commercial publisher databases, e.g., government reports, conference proceedings, theses, etc. This approach has been called snowballing or citation chaining [11].

Traditionally this approach has been laborious requiring manual searching. The process became somewhat easier with the introduction of indexes such as Scopus and Web of Science. More recently, the process has been revolutionised by innovative citation-based online literature mapping tools like LitMaps, Connected Papers or ResearchRabbit that optimise literature mining. The applications take one or two starting papers (called seed papers) and identify similar relevant papers around your research topic (derived from the data in the seed papers). However, these are not Google-style boolean search engines and their use should be recorded in your protocol. Any studies included in your review identified through snowballing and/or ad-hoc searching must be evaluated against your inclusion and exclusion criteria at the full text screening stage and against the quality appraisal checklist in the quality appraisal stage. The number of articles located through these processes should be documented separately, along with the number of articles mined.

Where to search

It is possible to conduct online searches (on any topic pertaining to science) through academic search engines and commercial databases [12]. They provide a centralised platform and allow the researchers to acquire information on the cited literature within seconds. Although many search engines are available, some incorporate data from more trusted resources than others. They may provide information on a range of topics from engineering and technology to biology and natural science. Although these sources provide a one-stop solution to all research-related needs and literature searches, access is usually limited to academic institutional licenses.

Investigators looking for relevant source materials are generally advised to search multiple databases and use various methods to collect information around their topic of interest [13–18]. Deciding on which databases to use is of particular importance for integrative reviews. 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 [19]. Several studies have investigated the added value of using multiple databases on search efficiency [20–27]. Some concluded that using a single database is sufficient [28, 29].

Others have concluded that using one database alone is not sufficient to retrieve all references for systematic reviews [30, 31]. Most articles on this topic draw their conclusions based on database coverage of the literature [26]. A recent paper considered the benefit of adding databases to your search strategy but no true conclusion could be drawn [32]. The fact that a publication may be found in one database and not another does not necessarily translate into it being found by your search strategy.

Recent work looking to determine the optimal combination of databases needed for systematic review searches (i.e., minimising the burden for the investigators without reducing the validity of the research by missing relevant references) suggests that searches should at least involve Embase, MEDLINE, Web of Science, and Google Scholar [33].

Using multiple databases introduces its own issues. For example, it can be laborious for searchers to translate search strategies into multiple interfaces and search syntaxes, factors such as field codes and proximity operators differ between interfaces [23]. Furthermore, it is time-consuming screening additional titles and abstracts that are quite likely irrelevant or duplicates of candidates you have already identified. Finally, access to certain databases is often only available on a subscription basis.

Online Search Tools:

- BioOne
- CiteSeer
- Embase
- GetCITED
- Google Scholar
- MedWorm
- Microsoft Academic Research
- Portland Perpustakaan
- Pubmed
- OvidSP
- Science Direct
- Scopus
- Springer Links
- Web of Science

Artificial intelligence

Recently artificial intelligence (AI) models have been incorporated into tools that can augment your search strategies. In its simplest form, large language models can be used to generate alternate search terms to use in traditional lexical search strategies. More interesting are the potential benefits that may be derived beyond keyword-based searches. Termed semantic searching, large language models can be used to interpret the meaning of words and phrases used to search for information. These tools can provide maps of data sources, review the citations of papers of interest and narrow lists by the application of filters. In some cases, they may be searching more than the titles and abstracts you would use to assess a manuscripts suitability.

One example is Sysrev, a machine learning-powered platform for document review and data extraction with the ability to directly integrate with alternate data sources and databases. The tool was built to aid in the creation of system evidence reviews. It has already been used to create over 16,000 reviews. It uses FAIR principles: Findability, Accessibility, Interoperability and Reuse of digital assets [34]. Its creation is a consequence of the recognised need to reduce redundancy and the inefficient use of human time and increase the impact of evidence-based decision-making.

However, these tools have their limitations. At best, they only search the same data sets as you yourself would search (possibly searching less data if the data is locked behind paywalls). Consequently, they do not appear to result in significantly greater numbers of candidate manuscripts [35]. They can also generate results that are out of context and/or inaccurate due to a phenomenon called hallucination [36–40].

An Interview With One of Our Medical Writers

Q What are you trying to achieve with a literature search?

A Although I am looking for data my main goal is efficiency. I might first do a simple search through PubMed to establish the extent of the likely output. This allows me to quickly assess what the search outputs are likely to be and tweak the search strategy to better achieve my goal. I aim to get the most precise record of the current understanding of the field of search; that equates to only the articles I need to see. Searching and re-searching the literature takes no time at all. The process of identifying relevant articles from among the many candidate citations is what takes the most effort and resource. Searching through hundreds of references can take days. Each iteration of your search should become more effective in that it results in a greater proportion of articles relevant to the objective. It can be useful to check your search terms with medical subject headings (MeSH), to cross reference whether your search terms are the same as those indexer uses [4].

Q Do you see AI removing the need for traditional literature searches?

A Clearly semantic search capabilities move us beyond traditional, keyword-based search strategies. However, there are still issues with AI and confidence in the results it can provide. As a rule of thumb, one should never solely rely on a single way of performing literature reviews. Although physical libraries may be considered outdated, one must use regular search engines and databases along with AI tools to perform a literature search. You'll have to evaluate the different options, discover the benefits and trade-offs when planning your project. A hybrid approach that combines traditional lexical search with vector-based approaches might be what you need.

Q How do you manage your search returns, do you use reference managing software?

A How one manages the many citations that are identified during your search will depend on the expected output. I tend to organise the information from each search result in a spreadsheet. There is also a function in PubMed that allows you to export into an Excel spread sheet all the available information about each article returned in a search. This is a very useful feature. For example, if you need more information you can export all the information, including the abstract.

Creating libraries within a reference management system is very helpful if you want to create bibliographies from a large number of references. There are several free and commercially available reference management systems such as Reference Manager and Endnote. These systems can take some time to set up and to incorporate your reference library of search results, but if you are planning to work in one field for some time these can be very helpful.

Reference Management Software Systems [41]

Free to use	Aigaion, Beboop, BibBase, BibDeck, BibSonomy, Bibus, CitULike, colwiz, Docear, JabRef, KBibTex, Mendeley, Pybliographer, Quiqqa*, ReadCube, rebase, RefDB, RefMe, Referencer, Wikindx, Zotero
For Purchase	Biblioscope, Bookends, Citavi, Endnote, Papers, Reference Manager, SciRef, Sente, WizFolio
Subscription	F1000Workspace, Paperpile, RefWorks

*also a premium subscription service

***“The hardest thing of all is to find a black cat in a dark room, especially if there is no cat”
Confucius***

And finally...

It is possible to develop processes by which you can determine whether your search strategy provides an acceptable capture rate, testing the sensitivity and precision of the search. Start by obtaining a recent review to see what percentage of the cited references appear in your list. You'll also need to gauge what an acceptable capture rate is. If you are looking at randomised clinical trials, sensitivity might be defined by the total number of known trials identified by the search whereas its precision would be the proportion of publications retrieved that are actually randomised clinical trials [42]. Reasons for reduced sensitivity might include limited use of MeSH terms, absent publication types, missing methodological terms and truncation of terms.

The value of any single article you find will not only be based on its intrinsic properties but also how it fits into and expands on previous work. You can identify what you feel are the most important articles by organising them to suit your requirements, such as selecting references published in journals with high impact factors or according to categories of clinical outcomes. If you rank articles manually you run the risk of introducing unacceptable subjectivity. You can minimise this by providing a formal set of criteria and applying these to your search results using independent reviewers blinded to any other information (about your candidate manuscripts) introducing an additional level of objectivity.

References

1. https://en.wikipedia.org/wiki/Systematic_review (accessed 20 Jan 2024)
2. <http://community.cochrane.org/cochrane-reviews> (accessed 20 Jan 2024)
3. http://www.nlm.nih.gov/pubs/techbull/jf06/jf06_skillkit.html (accessed 20 Jan 2024)
4. Smith D, et al. *Int Soc Pharmacoeconomic Outcome Res. News Article*. July 2009
5. Schardt C, et al. *BMC Med Inform Decis Mak* 2007; 7:16
6. http://en.wikipedia.org/wiki/Medical_Subject_Headings (accessed 20 Jan 2024)
7. Moher D, et al. *PLoS Med* 2007; 4(3):e78
8. Liberati A, et al. *Ann Intern Med* 2009; 151: W65–94
9. http://http://www.niche.org.uk/resource_centre.html (accessed 20 Jan 2024)
10. van Raan AFJ. *Scientometrics* 2004; 59, 467–472
11. Hirt J, et al. *F1000Res*. 2020 Dec 1;9:1386. doi: 10.12688/f1000research.27337.3.
12. https://en.wikipedia.org/wiki/List_of_academic_databases_and_search_engines (accessed 20 Jan 2024)
13. Levay P, et al. *Evid Based Libr Inf Pract* 2015; 10: 50–68
14. Stevinson C, Lawlor DA. *Complement Ther Med*. 2004; 12: 228–32
15. Lawrence DW. *Inj Prev*. 2008; 14: 401–4
16. Lemeshow AR, et al. *J Clin Epidemiol*. 2005; 58: 867–73
17. Zheng MH, et al. *Stroke*. 2008;39:e139. author reply e140
18. Beyer FR, Wright K. *Health Inf Libr J*. 2013; 30: 49–58
19. Higgins JPT, Green S. *Cochrane handbook for systematic reviews of interventions: The Cochrane Collaboration*, London, United Kingdom. 2011.
20. Wright K, et al. *Syst Rev*. 2015; 4: 104
21. Wilkins T, et al. *Can Fam Physician*. 2005; 51: 848–9
22. Halladay CW, et al. *J Clin Epidemiol*. 2015; 68: 1076–84
23. Ahmadi M, et al. *J Kerman Univ Med Sci*. 2014; 21: 343–54
24. Lorenzetti DL, et al. *Int J Technol Assess Health Care*. 2014; 30: 173–8
25. Beckles Z, et al. *J Clin Epidemiol*. 2013; 66: 1051–7
26. Hartling L, et al. *BMC Med Res Methodol*. 2016; 16: 1–13
27. Aagaard T, et al. *BMC Med Res Methodol*. 2016; 16: 161
28. van Enst WA, et al. *J Clin Epidemiol*. 2014; 67: 1192–9
29. Rice DB, et al. *J Psychosom Res*. 2016; 87: 7–13
30. Bramer WM, et al. *Syst Rev*. 2013; 2: 115
31. Bramer WM, et al. *Syst Rev*. 2016; 5: 39
32. Ross-White A, Godfrey C. *Health Inf Libr J*. 2017; 33: 217–24
33. Bramer WM, et al. *Syst Rev* 2017; 6:245–249
34. Bozada T, et al. *Front Artif Intell*. 2021; 4 <https://doi.org/10.3389/frai.2021.685298>
35. Schoeb D, et al. *Interact J Med Res*. 2020 Mar 30;9(1):e16606. doi: 10.2196/16606.
36. Artificial Intelligence in Medical Writing: An Insider's Insight. <https://www.niche.org.uk/asset/insider-insight/Insiders%20Insight%20GAILs.pdf> (accessed 20 Jan 2024)
37. Evolving journal guidelines for use of AI. *Editage Insights* <https://www.editage.com/insights/evolving-journal-guidelines-for-use-of-ai> (2023) (accessed 20 Jan 2024)
38. Burger B, et al. *Eur J Innov Man* 2023; 26, 233–241
39. Guittar M. *Research Guides: Using AI Tools in Your Research: Text-based AI and Machine Learning*. <https://libguides.northwestern.edu/ai-tools-research/ai-ml>.
40. Xiao, D. *Research Guides: AI-Based Literature Review Tools: Home*. <https://tamu.libguides.com/c.php?g=1289555&p=9470549>.
41. https://en.wikipedia.org/wiki/Comparison_of_reference_management_software (accessed 20 Jan 2024)
42. Dickersin K, et al. Identifying relevant studies for systematic reviews. *Systematic reviews*. In Chalmers I and Altman DG, editors. *Systematic Reviews*. London: BMJ Publishing group; 1995. p17–363

Next Steps

When done correctly, literature searching is invaluable for providing insights into research and developing evidence-based guidelines and recommendations. We created this Insider's Insight into Literature Searches to share some helpful pointers. We hope you found it useful.

We can also share with you a template for your search protocol [8], which is a useful way to start collating the results of your new literature search. If you would like advice on conducting your literature search please contact me at the email address below.

Dr Justin Cook
Head of Medical Writing
justin.cook@niche.org.uk

Get in touch



+44 (0)20 8332 2588
www.niche.org.uk