

COMMENTARY

Searching a methods topic: practical challenges and implications for search design

Julian Hirt^{a,b,c}, Hannah Ewald^d, Matthias Briel^{a,e}, Stefan Schandelmaier^{a,f,*}

^aDepartment of Clinical Research, University Hospital Basel, University of Basel, Basel, Switzerland

^bInternational Graduate Academy, Institute of Health and Nursing Science, Medical Faculty, Martin Luther University Halle-Wittenberg, Halle (Saale), Germany

^cDepartment of Health, Institute of Nursing Science, Eastern Switzerland University of Applied Sciences, St.Gallen, Switzerland

^dUniversity Medical Library, University of Basel, Basel, Switzerland

^eDepartment of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Canada

^fMTA–PTE Lendület "Momentum" Evidence in Medicine Research Group, Medical School, University of Pécs, Pécs, Hungary

Accepted 24 October 2023; Published online 30 October 2023

1. Introduction

Researchers may review the methodological literature (perform a "methods review") for different reasons: to inform the choice of methods when designing a study (e.g., diagnostic study, adaptive trial); to identify solutions for methodological challenges when conducting a study (e.g., slow recruitment) or analyzing the results (e.g., handling missing data, data visualization); to find material for teaching or consulting (e.g., nontechnical explanation, worked example); or to inform methodological research (e.g., compare different methods, develop methods guidance).

Performing a methods review can be challenging and guidance remains unavailable [1–4]. With this commentary, we aim to raise awareness about common challenges that reviewers addressing methods topics may encounter during the search process. We present a case study to illustrate common challenges, suggest solutions, and discuss research needs.

2. Case study: a review of "methods guidance"

We recently conducted a review to understand better the different types of methods guidance articles for health

researchers [1]. We initially aimed to include a random sample of 100 journal articles published between 2010 and 2020. Articles were eligible if they provided guidance for understanding, applying, reporting, or assessing health research methods. Together with an information specialist, we developed a Boolean search string (i.e., combining search terms with "and" and "or"), aiming for an acceptable compromise between high sensitivity (i.e., identifying a large proportion of relevant articles) and feasibility (i.e., retrieve a number of records that a small team can screen such as 10,000). We planned to apply versions of the search string to MEDLINE (Ovid), Embase (Elsevier), and PsycINFO (Ovid) and deduplicate the results.

We started with analyzing the terminology and metadata of a seed sample, a convenience sample of 45 methods guidance articles that were all included in MEDLINE (Appendix 1). To identify index terms and expressions related to "guidance" (e.g., tutorial, guide*, recommend*, best practice), we reviewed Medical Subject Headings (MeSH) terms and author keywords using the PubMed PubReMiner [5] and Yale MeSH analyzer [6] and reviewed the titles and abstracts of the seed articles. In addition, we reviewed the search terms included in published filters for "guidelines" (all developed primarily for clinical practice guidelines) listed in the The InterTASC Information Specialists' Sub-Group Search Filter Resource [7]. Using index terms and expressions related to "guidance", we drafted a search string for MEDLINE (Ovid), executed the search, and evaluated the results in two ways: First, we screened 200 abstracts (sorted by date added to MEDLINE, "entry date", showing the most recent articles first; currently the only sorting available for large number of hits), reviewed irrelevant records, and identified irrelevant or highly unspecific terms (e.g., "standard*") that we dropped or modified in the next version of the search string. Second, we compared the search results to our seed sample,

Availability of data and materials: All data generated or analyzed during this study are included in this publication or added as supplementary data.

Ethical approval: Not necessary for this study.

Funding: We developed this methodological commentary while working on research supported through Grants from the Swiss National Science Foundation (grant No. 190566) and the University of Basel (grant No. 3MS1043).

* Corresponding author. Department of Clinical Research, University Hospital Basel, University of Basel, Totengässlein 3, 4051 Basel, Switzerland. Tel.: +41-61-265-3819.

E-mail address: s.schandelmaier@gmail.com (S. Schandelmaier).

What is new?

Key Findings

- Searching a methods topic (e.g., “risk of bias”, “subgroup analysis”) in biomedical bibliometric databases (e.g., MEDLINE, Embase) can be challenging and guidance is missing
- We illustrate key challenges such as the lack of filters for method papers, suboptimal indexing, and highly variable terminology and reporting

What this adds to what was known?

- We discuss pragmatic solutions such as starting the search process with selected journals; anticipating lower-than-usual sensitivity and precision of index terms; consulting with methods experts and chatbots to identify relevant search terms
- We identified research needs including a typology for methods papers; methods-specific index terms; reporting standards; and assessing the usefulness of citation-based searches and machine learning in the methods context

What is the implication and what should change now?

- The proposed solutions may be helpful for health researchers, methods reviewers, information specialists, and others who search a methods topic

identified articles that the search had missed, and reanalyzed those to identify additional expressions (e.g., “elabor*”) that we added to the next version of the search string. After repeating the process several times, the number of hits remained as high as 876,395 hits (see [Appendix 2](#) for the final version of the search string). We screened the first 500 hits (again ordered by entry date), identified 18 potentially relevant articles of which we screened the full texts, and identified only one relevant methods guidance article. With an estimated number of around 500 records needed to screen to find one relevant article, screening a random sample of 10,000 records might have yielded only 20 instead of the 100 articles needed. We considered combining our search string with a filter for “methods” or “research methodology”. Such a filter, however, is currently not available [7]. Given the low specificity of terms related to “methods”, we suspected that it would be too difficult to develop a useful filter. Therefore, given the great number of records needed to screen to find a relevant study, and the lack of a filter for “methods”, we abandoned our original plan.

In a second attempt, we changed our sampling strategy from random to purposeful, now focusing on the 100 most

recent guidance articles published in journals that frequently publish methods guidance. Also, we limited the search to the year 2020 (with the option to add more years if needed to achieve the sample size of 100). To identify relevant journals, we screened the websites of six high-impact clinical journals for journal-defined sections, article types, or series that include methods guidance. We included four journals (the British Medical Journal, Annals of Internal Medicine, Journal of the American Medical Association, and Public Library of Science Medicine) that yielded 29 methods guidance articles published in 2020 and excluded two journals (The New England Journal of Medicine, The Lancet) that did not include relevant sections, article types, or series. In addition, we applied our search strategy ([Appendix 2](#)) to eight methodology-oriented journals (Statistics in Medicine, BioMed Central Medical Research Methodology, Journal of Clinical Epidemiology, International Journal of Epidemiology, European Journal of Epidemiology, American Journal of Epidemiology, Epidemiology, and BioMed Central Trials), resulting in 741 records to screen for the year 2020 and yielding 76 additional methods guidance articles. Because the total of 105 methods guidance articles was greater than our minimum of 100, we did not widen the search and proceeded with the review. The results suggested that methods guidance articles are a highly heterogeneous type of articles with no standards regarding the development process, reporting, and disclosure of conflicts of interest. In the discussion, we acknowledge limitations regarding generalizability due to the arbitrary sample [1].

The case study illustrates challenges that may affect many searches for a methods topic. We encountered similar challenges when we reviewed the topics “subgroup analysis” [8], “feasibility checklists” [9], “strategies to obtain individual participant data” [10], or “handling missing outcome data” [11]. Other groups have mentioned similar challenges for topics including “heterogeneity of treatment effects” [12] and “competing risks” [13]. Also, the curators of the (now discontinued) Cochrane Methodology Register [14], the Core Outcome Measures in Effectiveness Trials database [15], and the Online Resource for Recruitment research in Clinical trials [16,17] acknowledged challenges when searching for a methods topic.

In [Table 1](#), we identify specific challenges and discuss, based on our experience, implications for search design and research needs.

3. Distinguishing methods papers from other papers is difficult

Both methods paper and study reports typically include methodological expressions (e.g., “randomized controlled trial”) in the title, abstract, and sometimes author keywords and index terms. Therefore, searches for methods that are applied in many studies (e.g., “subgroup analysis”) may

Table 1. Challenges when searching a methods topic, suggestions for search design, and suggestions for research and development

Issue	Suggestions for search design	Suggestions for research and development
Distinguishing methods papers from other papers can be difficult, filters remain unavailable	- May begin with and consider restricting searches only to methods-oriented journals - Consider methods-specific databases (e.g., LIGHTS, Core Outcome Measures in Effectiveness Trials, ORRCA) - Consider frequency-based best-match ordering of references (e.g., PubMed's "sort by: Best match") - Consider machine learning (ML)—supported screening tools	- Typology of methods papers to inform clear labels in titles and abstracts - Assess the performance of ML to distinguish methods papers from other papers
Methods-specific index terms may be missing, suboptimal, or applied inconsistently	- Consider searching both MEDLINE and Embase since each provide different methodological index terms - Scrutinize appropriateness of index terms together with a methods expert (questionable definition and lack of alternative terms - > risk of inappropriate and inconsistent use) - Run searches with candidate index terms to develop sense of appropriate use (lots of irrelevant articles - > risk of low precision) - Review index terms used in a sample of relevant articles to develop sense of consistency (inconsistent use - > risk of low sensitivity)	- Database publishers may consider the addition of more methodological index terms - Database publishers may review whether existing methodological index terms are being applied consistently and appropriately
Many alternative expressions for a methods topic	- Expect a potentially large number of alternative expressions - Approach different content experts (one alone may not be enough) to identify the most common alternative expressions - Analyze the terminology of a diverse set of key publications - Expect clustering of terms by academic discipline (e.g., epidemiology, statistics, psychology, oncology) - Plan, to allow adequate time to identifying relevant expressions	- Development of a thesaurus for methodological concepts to allow identification of alternative expressions - Investigate usefulness of ML-supported search tools applied to titles and abstracts in the methods context
Abstracts may be brief, unstructured, or missing	Compared to searches of clinical topics, - Expect lower sensitivity of both Boolean search strings and ML-supported searches - Expect a greater proportion of full texts needed to be screened to identify a relevant study	- Typology of methods papers to inform clear labels in titles and abstracts - Guidance for reporting abstracts of methods papers
Little is known about citation searching in the methods context	- Try citation searching, also as a source for additional search terms - Look out for common references - Be aware that methods papers may lack relevant references due to the challenges discussed in this commentary	- Investigate sensitivity of citation searching in different methods contexts

Abbreviation: ML, machine learning.

yield mostly study reports. Frequency-based best-match ordering may sometimes be useful. For instance, a search of “subgroup analysis” in PubMed yields more than 188,000 papers and mostly study reports; using PubMed’s

“sort by: Best match” moves many methods papers to the top of the list. Another good example is “*P*-value”. Formal filters for methods papers are not available, neither as a feature of the search interfaces PubMed, Ovid, or

Embase.com nor as an externally developed filter [7]. Machine learning (ML) supported approaches, including screening tools designed for systematic reviews [18], could fill that gap. The Core Outcome Measures in Effectiveness Trials working group reported a positive experience with an ML model trained to rank the results of a search for “core outcome sets” by relevance (i.e., sort the results by relevance and show papers that are likely to be relevant first). The ranking allowed the reviewers to stop after screening 25% of the references, saving 62.5 hours of screening time [16]. It should be noted, however, that ML-supported searches may be associated with new challenges such as deciding when a training dataset is of sufficient quality and agreeing on thresholds for stopping the screening process [19]. Future studies could evaluate the performance of ML specifically to distinguish methods from nonmethods papers or distinguish different types of methods papers (which would require a typology of methods papers which, to our knowledge, is not available yet). Another (untested) approach—as applied in the case study—would be to restrict the search to methods-oriented journals. Finally, one could search methods-specific databases that are available for some topics including patient recruitment [16], core outcome sets [15], reporting guidelines [20], and methods guidance [21,22].

4. Methods-specific index terms may be missing or applied inconsistently

The indexing systems of MEDLINE (MeSH) and Embase (Emtree) provide, in principle, solutions for distinguishing methods papers from other articles. One is using index terms that directly express “methods” such as the MeSH term “Epidemiologic Methods”. Another is adding the suffix “as topic” to index terms such as “Clinical Trial as topic”. Another option in MEDLINE (not Embase) is to attach subheadings to index terms such as “Clinical trial/methods” or “Clinical trial/statistics and numerical data”. Yet another solution is to use methods-specific index terms such as “sample size”. For many important topics, however, specific index terms remain unavailable or seem suboptimal (Table 2). One example is the topic “blinding” for which the MeSH system provides two terms, “Single-Blind Method” and “Double-Blind Method”, that seem too narrow for many papers addressing blinding (e.g., value of blinding, methods for blinding) and have been described as potentially misleading [23]. A general term for “blinding” is not available, nor included as an alternative term for another MeSH term (i.e., a search in PubMed for “blinding” does not automatically map to another MeSH term). We note, the Emtree system seems to provide more methods-specific index terms than the MeSH system (Table 2).

When appropriate or related index terms are available, they may not be added consistently to the database record

of corresponding methods papers. Of the 105 guidance articles included in the case study, only 40 were indexed with terms expressing or related to “guidance” such as “Checklist” or with the subheading “/standards” [1]. “Methods guidance” is an example of a concept for which MeSH or Emtree currently do not provide corresponding specific index terms. The use of nonspecific terms such as “standards” (as floating subheading or as search term) may retrieve many irrelevant articles. Another potential issue is that MeSH terms developed for methods papers (e.g., “Clinical Trials as Topic”) are sometimes incorrectly applied to primary studies.

In summary, according to our experience, search strategies that are primarily based on index terms may not provide an acceptable sensitivity and precision for identifying methods papers. Hence, author-generated expressions in the title, abstract, and author keywords may be particularly relevant in the methods context. A suggestion for database publishers is to review, ideally together with methods experts, current approaches to indexing methods papers.

5. Many alternative expressions for a methods topic

Methodological terminology can be highly variable. For instance, the titles and abstracts of the 105 articles in the case study included 36 different expressions for “guidance” including nonspecific terms such as “best practice” or “how to” [1]. Another challenge is that different academic fields may prefer different expressions for the same concept. For instance, regarding the concept “effect modification”, epidemiologists may prefer “subgroup effect” or “heterogeneity of treatment effects”, statisticians “interaction”, psychologists “moderation”, and oncologists “predictive factor”. Creating a comprehensive list of alternative expressions and considering different academic fields seems a crucial step in the design of a sensitive, text-based search for a methods topic. The downside of text-based searches, as illustrated in the case study, is that the development process may require many iterations (e.g., to identify alternative expressions and test their added value) and the search results can be highly imprecise.

A comprehensive thesaurus of alternative methodological expressions is currently not available [24]. Index terms in MEDLINE and EMBASE, if available, provide one potential source of alternative expressions (“entry terms” in the MeSH system, “synonyms” in the Emtree system). Also, we are currently developing a thesaurus specifically for methods topics as part of the controlled vocabulary of the Library of Guidance for Health Scientists [22]. Emerging new techniques to identify new sources might be chatbots such as ChatGPT or Google Bard (try “What are synonyms for effect modification?”); double-checked by a methods expert within the review team, chatbots may quickly identify many relevant alternative expressions.

Table 2. Important methodological concepts with no corresponding index terms in MEDLINE or Embase. Checked in August 2023^a

Topic	MeSH vocabulary (MEDLINE)	Emtree vocabulary (Embase)
Reporting guideline	No	Candidate term (reporting)
Subgroup analysis/effect modification	No	Candidate terms (both subgroup analysis and effect modification)
Applicability/generalizability	No	Candidate terms (both applicability and generalizability)
Heterogeneity	No	Candidate term
Diagnostic accuracy	No	Yes
Meta-research	No	No
Hypothesis testing	No	Candidate term
Qualitative analysis	No	Yes
Sensitivity analysis ^b	No	Yes
Missing data	No	Candidate term
Random error	No	Yes
Individual participant data	No	Candidate term
Family-wise error rate	No	Yes
Statistical significance	No	Yes
<i>P</i> value	No	Yes
Causal inference	Only broader “Causality”	Candidate term
Directed acyclic graph	No	Yes
Validity	Alternative term for “Reproducibility of Results”	Yes
Reliability	Alternative term for “Reproducibility of Results”	Yes
Bayesian analysis	Alternative term for “Bayes Theorem”	Alternative term for “Bayes Theorem”
Hazard ratio	No	Yes
Surrogate endpoint	Alternative term for “Biomarker”	Candidate term
Blinding	Only narrower “Single-Blind Method” and Double-Blind-Method”	Only narrower “Single-Blind Procedure”, Double-Blind-Procedure”, and Triple-Blind-Procedure”
Estimand	No	Candidate term

^a MEDLINE may introduce new MeSH terms annually and Embase new Emtree terms several times per year.

^b Not to be confused with the MeSH terms “sensitivity and specificity” that refers to measurement but not to the different concept “sensitivity analysis”.

6. Abstracts may be brief, unstructured, or missing

In contrast to study reports that usually include an abstract with the sections Introduction, Methods, Results, and, Discussion and provide a clear objective, abstracts of methods papers may not follow a specific structure. There are no published formal reporting standards for methods papers [20]. Of the 105 methods guidance articles included in the case study, 44 had a structured abstract, 45 an unstructured abstract, and 16 no abstract. Structured abstracts are preferable from a searcher’s point of view because, similar to a checklist, they are likely to increase the amount of information and consistency across journals. Missing or unstructured abstracts increase the risk that reviewers, but also ML-powered approaches, miss relevant articles.

Another challenge in the methods context can be titles designed to stimulate interest rather than clearly expressing the focus of a paper such as “Clinicians are right not to like Cohen’s κ ” [25] or “Why carve up your continuous data?” [26]. Such titles are difficult to retrieve. A reporting guideline could support the retrieval of methods papers through harmonizing the structure of abstracts and make the titles of methods papers more consistent.

7. Citation searching may yield additional articles and search terms

Preliminary evidence suggests that citation searching adds value to most systematic reviews [27]. Therefore,

and given the challenges associated with Boolean search strings, citation searching may be useful in the methods context. In our case study, for reasons related to our sampling strategy, we did not screen references (backward citation searching) or citing articles (forward citation searching). We noted, however, that eight of the 13 included reporting guidelines shared a common reference. [28] In a systematic search for reporting guidelines, screening the articles that cite that common reference could be an efficient approach. A strength of citation searching is that it mitigates the aforementioned challenge of heterogeneous terminology and also provides a potential source of additional search terms. A potential caveat is that the authors of methods papers, due to the challenges outlined in this commentary, may miss relevant references. Consequently, the sensitivity of citation searching in a methods review may be lower than in a systematic review of a clinical topic. More research is needed to better understand citation searching in the methods context.

In summary, we have highlighted that searching a methods topic in general biomedical databases can be associated with a number of challenges. Although the individual challenges are not specific to the methods context, if they occur together, as in our case, they can result in searches with poor sensitivity and precision. Reviewers addressing methods topics and information specialists who design a search strategy should be aware of the challenges; may adjust their expectations regarding sensitivity and precision; anticipate extra time for identifying alternative expressions, consulting with methods experts, and screening full texts; and be open to experimenting with citation search and ML. More research is required to improve the retrieval of methods papers.

CRedit authorship contribution statement

All authors have made substantial contributions to the conception of the commentary, drafting the commentary or revising it critically for important intellectual content, and final approval of the version to be submitted.

Data availability

No additional data or code.

Declaration of competing interest

None declared.

Acknowledgments

We sincerely thank Julie Glanville and Susan Wieland for their detailed feedback and many constructive suggestions during the peer review process.

Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jclinepi.2023.10.017>.

References

- [1] Hirt J, Ewald H, Lawson DO, Hemkens LG, Briel M, Schandelmaier S. A systematic survey of methods guidance suggests areas for improvement regarding access, development, and transparency. *J Clin Epidemiol* 2022;149:217–26.
- [2] Gentles SJ, Charles C, Nicholas DB, Ploeg J, McKibbin KA. Reviewing the research methods literature: principles and strategies illustrated by a systematic overview of sampling in qualitative research. *Syst Rev* 2016;5(1):172.
- [3] Furlong DE, Lester JN. Toward a practice of qualitative methodological literature reviewing. *Qual Inq* 2023;29(6):669–77.
- [4] Heinze G, Boulesteix AL, Kammer M, Morris TP, White IR, Simulation Panel of the STRATOS initiative. Phases of methodological research in biostatistics-Building the evidence base for new methods. *Biom J* 2023:e2200222.
- [5] PubMed PubReMiner: a tool for PubMed query building and literature mining. Available at <https://hgserver2.amc.nl/cgi-bin/miner/miner2.cgi>. Accessed July 25, 2023.
- [6] Yale MeSH Analyzer. Available at <https://mesh.med.yale.edu/>. Accessed July 25, 2023.
- [7] ISSG search filter resource. Available at <https://sites.google.com/a/york.ac.uk/issg-search-filters-resource/>. Accessed July 20, 2020.
- [8] Schandelmaier S, Chang Y, Devasenapathy N, Devji T, Kwong JS, Colunga Lozano LE, et al. A systematic survey identified 36 criteria for assessing effect modification claims in randomized trials or meta-analyses. *J Clin Epidemiol* 2019;113:159–67.
- [9] Gloy V, Speich B, Griessbach A, Taji Heravi A, Schulz A, Fabbro T, et al. Scoping review and characteristics of publicly available checklists for assessing clinical trial feasibility. *BMC Med Res Methodol* 2022;22:142.
- [10] Ventresca M, Schünemann HJ, Macbeth F, Clarke M, Thabane L, Griffiths G, et al. Obtaining and managing data sets for individual participant data meta-analysis: scoping review and practical guide. *BMC Med Res Methodol* 2020;20:113.
- [11] Kahale LA, Guyatt GH, Agoritsas T, Briel M, Busse JW, Carrasco-Labra A, et al. A guidance was developed to identify participants with missing outcome data in randomized controlled trials. *J Clin Epidemiol* 2019;115:55–63.
- [12] Varadhan R, Stuart EA, Louis TA, Segal JB, Weiss CO. Review of guidance documents for selected methods in patient centered outcomes research: standards in addressing heterogeneity of treatment effectiveness in observational and experimental patient centered outcomes research 2012. Available at pcori.org. Accessed December 11, 2023.
- [13] Manja V, AlBashir S, Guyatt G. Criteria for use of composite end points for competing risks-a systematic survey of the literature with recommendations. *J Clin Epidemiol* 2017;82:4–11.
- [14] Cochrane methodology register. Available at <https://cmr.cochrane.org/>. Accessed January 22, 2020.

- [15] Home: Core outcome Measures in effectiveness trials initiative (COMET). Available at <http://www.comet-initiative.org/>. Accessed July 3, 2019.
- [16] ORRCA: Home. Available at <http://www.orrca.org.uk/Home/Index/>. Accessed January 22, 2020.
- [17] Muhammad I, Bollegala D, Coenen F, Gamble C, Kearney A, Williamson P. Maintaining curated document databases using a learning to rank model: the ORRCA experience. In: Bramer M, Ellis R, editors. Artificial intelligence XXXVII12498. Cham: Springer International Publishing; 2020:345–57. https://doi.org/10.1007/978-3-030-63799-6_26.
- [18] Chappell M, Edwards M, Watkins D, Marshall C, Graziadio S. Machine learning for accelerating screening in evidence reviews. *Cochrane Evidence Synthesis and Methods* 2023;1(5): e12021.
- [19] Stansfield C, Stokes G, Thomas J. Applying machine classifiers to update searches: analysis from two case studies. *Res Synth Methods* 2022;13(1):121–33.
- [20] The EQUATOR Network. Available at <https://www.equator-network.org/>. Accessed September 13, 2021.
- [21] LIGHTS - search. Available at <https://lights.science/app/search/randomized>. Accessed March 30, 2023.
- [22] Hirt J, Schönenberger CM, Ewald H, Lawson DO, Papola D, Rohner R, et al. Introducing the library of guidance for health Scientists (LIGHTS): a living database for methods guidance. *JAMA Netw Open* 2023;6(2):e2253198.
- [23] Lang TA, Stroup DF. Who knew? The misleading specificity of “double-blind” and what to do about it. *Trials* 2020;21(1):697.
- [24] BARTOC.Org. Available at <http://bartoc.org>. Accessed July 31, 2023.
- [25] de Vet HCW, Mokkink LB, Terwee CB, Hoekstra OS, Knol DL. Clinicians are right not to like Cohen’s κ . *BMJ* 2013;346:f2125.
- [26] Owen SV, Froman RD. Why carve up your continuous data? *Res Nurs Health* 2005;28:496–503.
- [27] Hirt J, Nordhausen T, Appenzeller-Herzog C, Ewald H. Citation tracking for systematic literature searching: a scoping review. *Res Synth Methods* 2023;14(3):563–79.
- [28] Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. *PLoS Med* 2010;7(2): e1000217.