Development and validation of study filters for identifying controlled non-randomized studies in PubMed and Ovid MEDLINE

Siw Waffenschmidt1,2 | Tamara Navarro-Ruan3 | Nick Hobson3 | Elke Hausner1 | Stefan Sauerland1 | R Brian Haynes3

1Institute for Quality and Efficiency in Health Care, Cologne, Germany
2Institute for Health Economics and Clinical Epidemiology, The University Hospital of Cologne, Cologne, Germany
3Health Information Research Unit, Department of Health Research Methods, Evidence and Impact, McMaster University, Hamilton, Canada

A retrospective analysis published by the German Institute for Quality and Efficiency in Health Care (IQWiG) in 2018 concluded that no filter for non-randomized studies (NRS) achieved sufficient sensitivity (≥92%), a precondition for comprehensive information retrieval. New NRS filters are therefore required, taking into account the challenges related to this study type. Our evaluation focused on the development of study filters for NRS with a control group (“controlled NRS”), as this study type allows the calculation of an effect size. In addition, we assumed that due to the more explicit search syntax, controlled NRS are easier to identify than non-controlled ones, potentially resulting in better performance measures of study filters for controlled NRS. Our aim was to develop study filters for identifying controlled NRS in PubMed and Ovid MEDLINE. We developed two new search filters that can assist clinicians and researchers in identifying controlled NRS in PubMed and Ovid MEDLINE. The reference set was based on 2110 publications in Medline extracted from 271 Cochrane reviews and on 4333 irrelevant references. The first filter maximizes sensitivity (92.42%; specificity 79.67%, precision 68.49%) and should be used when a comprehensive search is needed. The second filter maximizes specificity (92.06%; precision 82.98%, sensitivity 80.94%) and should be used when a more focused search is sufficient.

KEYWORDS
bibliographic databases, information storage and retrieval, MEDLINE, Non-randomized controlled trials as topic, review literature as topic
1 | BACKGROUND

When investigating the effects of medical interventions, randomized controlled trials (RCTs) show the lowest risk of bias and thus the highest certainty of results of all study types, provided their methods were correct and implemented in a way suitable to address a study's objectives. They are therefore generally the primary type of evidence included in systematic reviews. In contrast, the inclusion of non-randomized studies (NRS) leads to a markedly higher risk of bias.1 However, in cases where RCTs are lacking or cannot be conducted, it may be helpful to consider NRS.

To determine whether study filters can be used to facilitate the searches for NRS in bibliographic databases, a retrospective analysis published by the German Institute for Quality and Efficiency in Health Care (IQWiG) in 2018 identified and validated existing NRS filters in MEDLINE.2 No NRS filter achieved sufficient sensitivity (≥92%), a precondition for comprehensive information retrieval (due to insufficient sensitivity, specificity was not evaluated). The conclusion of this analysis was that it was necessary to develop new NRS filters, taking into account the challenges related to this study type.

Definitions and labelling of NRS types are inconsistent.3-5 Our focus was on the development of study filters for NRS with a control group (“controlled NRS”), as this study type allows the calculation of an effect size. Case series lack this feature and are therefore allocated to the lowest evidence level for studies.6 In addition, we assumed that due to the more explicit search syntax, controlled NRS are easier to identify than non-controlled ones, potentially resulting in better performance measures of study filters for controlled NRS.

2 | OBJECTIVES

The present analysis was a follow-up project of the IQWiG project mentioned above.2 Our aim was to develop study filters for identifying controlled NRS in PubMed and Ovid MEDLINE.

3 | METHODS

The methods used are based on guidance for the assessment of the performance of methodological search filters by Lefebvre et al7 and Bak et al8 As MEDLINE is the most frequently used bibliographic database in medicine,9 our search filter was restricted to this source. The search filter was developed and originally tested in PubMed, as the calculation of filters in the McMaster Clinical Hedges Database is based on the PubMed syntax. The final filters were then adapted for Ovid MEDLINE.

3.1 | Performance measures

We calculated different measures for search filter performance (Table 1). For the 2 x 2 table calculations, a collection of different sets of references were used: development set and validation set (relevant references, explained in Section 3.2.1) as well as irrelevant references (explained in Section 3.2.2).

3.2 | McMaster clinical hedges database

All calculations were performed in the McMaster clinical hedges database, a well-established resource for the computation of search filters.10 It is possible to import a list of terms for analysis, as well as a set of references representing either eligible or ineligible references. The unique feature of the database is that it is possible to calculate performance measures (sensitivity, specificity, precision, and accuracy) for single as well as for multiple combinations of terms (up to 4). It is also possible to calculate the performance measures for each search query. The Clinical Hedges database was established for clinical articles published in the year 2000 and has since been revalidated.10
3.2.1 | Generation of the development and the validation set

We developed different sets of references for different purposes (Figure 1). Firstly, we generated the development and the validation set using the relative recall method. To achieve sufficient performance, a sensitivity of at least 95% for the study filter was specified in the previous study. We described there that, for a sample of 200 PMIDs per study type, if the filter's sensitivity lies within the interval of [0.92;1] it cannot be excluded that the actual sensitivity is 95%. The targeted sensitivity was therefore defined as 92% or higher. If the number of available PMIDs was lower, this was described in the results section and we estimated how the evaluation of sensitivity was affected. As stated in the previous study, a sufficient number of publications could be identified for 5 (out of 7) study types with 200 citations each (Table 2).

To generate the sets, we screened Cochrane reviews including NRS. The approach and results are described elsewhere. In short, we modified the search syntax by Ijaz et al and via PubMed searched for Cochrane reviews that largely include NRS (last search date October 2016). There was no limitation on whether the Cochrane reviews used NRS filters; however, we found that more than half of the reviews had a search block for NRS. All eligible Cochrane reviews had to fulfil the inclusion criteria (evaluation of a health care intervention, not only inclusion of RCTs or NRCTs, inclusion of NRS, and inclusion of <65 studies). We identified the citations of the NRS included in the Cochrane reviews via their bibliographies and extracted the corresponding PubMed identification numbers (PMIDs).

We used the tool by Hartling et al to classify the study types (Appendix A). We considered all controlled NRS study types (Table 2). RCTs as well as NRS without a control group were deleted from the original set of references.

For each study type, all references were randomly divided into the development set (60%) and the validation set (40%) using a random number generator.

3.2.2 | Generation of the list of terms

We used the McMaster Hedges list of terms, which was first generated in 1994, and has since been updated. The list was generated by input from clinicians and librarians and currently contains 5395 terms. All terms were tested for various study types using the Ovid Technologies searching system. We adapted the syntax for the PubMed interface (Figure 2). As the list of terms was not specifically generated or tested for controlled NRS, we conducted an additional text analysis based on the development set to add further search terms specific to the search for controlled NRS. We also excluded terms that would probably solely apply to study types that were definitely not of interest for our objective. For example, we excluded “random,” as this term would probably solely apply to RCTs.

In a next step we reduced the list of terms, as it was too large for a meaningful automated analysis. For this purpose, all single computation terms with the following performance measures were exported to Excel for further adjustment:

- sens ≥ 10%, spec ≥ 99%
- sens ≥ 15%, spec ≥ 98%
- sens ≥ 20%, spec ≥ 97%
- sens ≥ 25%, spec ≥ 95%
- sens ≥ 30%, spec ≥ 92%
- sens ≥ 35%, spec ≥ 85%
- sens ≥ 40%, spec ≥ 80%

We then summarized all individual terms and deleted duplicates.

Furthermore, we added NOTing Out terms to the list. These terms were generated by running the development set against the irrelevant references.

The irrelevant set of references that we added contained references for the McMaster PLUS database generated by hand search over several years and indexed as “not research” and “not of interest” (Table 3). We picked a random sample from the years 2012, 2015 and
2018 without further screening (as they were definitely irrelevant references) and used them as a comparator set.

In a next step, we manually scanned the terms in the McMaster Clinical HEDGES database that could be found in cell “c” (references met the criteria but were missed; for further explanation please see Table 1) but not in cell “a” (met the criteria and were included). In a final step, we also added a common NOTing Out syntax in search filters (for NOTing OUT animals).

3.2.3 | Generation of search filters

On the basis of the final list of terms and the different sets of references (irrelevant references, development set, validation set) we developed and validated the search filters with the McMaster Clinical Hedges Database (Figure 3). We chose specificity as a performance measure, as we were mainly interested in reducing the number of irrelevant records in the search results. The aim was to develop two types of search filters:

- (1) with maximum sensitivity ($\geq 92\%$), specificity $\geq 80\%$, and PubMed hits <7 million
- (2) with maximum specificity and sensitivity $\geq 80\%$

For the filter with the best sensitivity based on the final list of terms, we first exported all “triple combinations” (combinations of 3 terms) with a sensitivity of 70% and more and a specificity of 90% and more. We then summarized all the individual terms and deleted duplicates. For similar terms, we chose the terms with

![FIGURE 1 Generation of the development and the validation set](image.png)

| Study type                      | Number of Cochrane reviews | Number of studies | Number of publications | Number of PMIDs without duplicates |
|---------------------------------|----------------------------|-------------------|-----------------------|----------------------------------|
| Quasi-randomized controlled trial | 67                        | 216               | 381                   | 331                              |
| Controlled before-after study   | 104                       | 634               | 906                   | 556                              |
| Interrupted time series (with a control group) | 31                        | 83                | 165                   | 106                              |
| Prospective cohort study        | 84                        | 384               | 493                   | 435                              |
| Retrospective cohort study      | 72                        | 436               | 480                   | 451                              |
| Non-concurrent cohort study     | 13                        | 34                | 34                    | 31                               |
| (Nested) case–control study     | 36                        | 207               | 216                   | 200                              |
| Total                           | 2110                      |                   |                       |                                  |

*aNumber of Cochrane reviews in which the study type was included at least once (multiple counting possible).
the highest sensitivity; Table 4 shows an example (“comparative stud*[all]”).

In the next step, we combined all the remaining terms with the Boolean OR and iteratively deleted individual terms until the targeted thresholds were reached. As we were not able to reach 92% sensitivity with this approach, we then manually checked the frequent terms that were in the $2 \times 2$ table (Table 1) in cell “c” (relevant references missed by the search filter) but not in cell “a” (relevant references found by the search filter). With this approach, we were able to add three further terms with sufficient performance to the syntax. Finally, we added NOTing Out terms to increase specificity.

For the filter with the best specificity, we exported the TOP best 10 “triple computations” with a sensitivity of 70% and more and a specificity of 90% and more. We then summarized all the individual terms and deleted duplicates. For similar terms, we chose the terms with the

### TABLE 3  Definition of irrelevant terms

| Category     | Reading criteria                                                                 |
|--------------|----------------------------------------------------------------------------------|
| Not research | Is this an original study, review, evidence-based guideline, or pooled analysis of original studies? |
| Not of interest | Is this original study, review, evidence-based guideline, or pooled analysis of original studies of direct interest to the health care of humans? |

### FIGURE 2  Generation of list of terms

---

### FIGURE 3  Generation of search filters
highest specificity; Table 4 shows an example (“comparative study[pt]”).

**4 | RESULTS**

Figure 4 shows the result for Section 3.2.1 (number of references used for the development set and the validation set) and Section 3.2.2 (final number of terms and NOTing Out terms), which were both used for further analysis. In total, 271 Cochrane reviews including 482 studies (4544 publications available in MEDLINE) were extracted.\(^2\) For the present study, we considered all controlled NRS study types (2110 MEDLINE references). The datasets used and/or analysed during the current study are available from the corresponding author on request. The final list of terms and NOTing Out terms are presented in Appendix B.

**4.1 | Search filter with the best sensitivity**

The combination of the individual terms in the McMaster Clinical Hedges Database resulted in a search filter with
the best sensitivity. The corresponding syntax included all NOTing Out terms and resulted in a sensitivity of 92.42%, a specificity of 79.67%, and a precision of 68.49%; the number of hits in PubMed was 7 150 309 (Table 5). Appendix C shows the performance measures of the search filter without NOTing Out terms.

We also adapted the Pubmed filters for application in the Ovid MEDLINE interface. Appendix D explains how the different search fields were adapted. Additional search terms were added and some search lines summarized for the search syntax. The final search filter is presented in Table 6.

We also checked whether it was possible to achieve a similar value for sensitivity in Ovid MEDLINE, as sensitivity could have varied due to the different interfaces; this search resulted in a sensitivity of 92.17% (PubMed:92.42%). Since the value for the sensitivity estimate in Ovid lies within the confidence interval of the estimate in PubMed (and thus the confidence intervals overlap), no significant difference is assumed.

In addition, we checked the characteristics of the relevant studies that were not detected by the Pubmed filter (Figure 2, c = 140). We checked whether undetected studies could be detected by means of another publication in cases where more than one publication on a study existed. Forty two publications were multiple publications where at least one other publication was identified by the filter. Hence, the study could have been identified through the filter, meaning that in reality, sensitivity is likely to be higher.

### 4.2 Search filter with the best specificity

The combination of the individual terms in the McMaster Hedges Database resulted in a search filter with the best specificity (PubMed: Table 7, Ovid MEDLINE: Table 8). The syntax included all NOTing Out terms and resulted in a sensitivity of 80.94%, a specificity of 92.06%, and a precision of 82.98%; the number of hits in PubMed was 4 500 776. Appendix C shows the performance measures of the search filter without NOTing Out terms.

We also checked whether we were able to achieve a similar value for sensitivity in Ovid MEDLINE, as sensitivity could have varied due to the different interfaces;

| TABLE 6 | Search filter with best sensitivity for controlled NRS (Ovid MEDLINE, adapted from PubMed) |
| --- | --- |
| 1 | exp cohort studies/or exp epidemiologic studies/or exp clinical trial/or exp evaluation studies as topic/or exp statistics as topic/ |
| 2 | (control and [group* or study]) or (time and factors) or program or survey* or ci or cohort or comparative stud* or evaluation studies or follow-up*).mp. |
| 3 | or/1–2 |
| 4 | (animals/not humans/) or comment/or editorial/or exp review/or meta analysis/or consensus/or exp guideline/ |
| 5 | hi.fs. or case report.mp. |
| 6 | or/4–5 |
| 7 | 3 not 6 |

Abbreviation: hi.fs = floating subheading “history.”

| TABLE 7 | Search filter with best specificity for controlled NRS (PubMed) |
| --- | --- |
| **Sensitivity** | **Specificity** | **Precision** | **Accuracy** |
| (clinical trial[pt] OR comparative study[pt] OR (control[tw] AND study[tw]) OR program[tw] OR epidemiologic studies[mh]) NOT ((animals[mh:noxexp] NOT humans[mh:noxexp]) OR comment[pt] OR editorial[pt] OR review[pt] OR meta analysis[pt] OR case report [tw] OR consensus[mh] OR guideline[pt] OR history[sh])) | PubMed hits: 500.776 (September 16, 2019) |
| D 82.00% [79.85-84.14] | D 92.73% [91.73-93.73] | D 83.98% [81.91-86.05] | D 89.20% [88.22-90.18] |
| V 79.26% [76.52-82.00] | V 91.29% [89.96-92.62] | V 87.36% [78.83-93.73] | V 87.36% [86.08-88.65] |
| A 80.89% [79.20-82.58] | A 92.15% [91.35-92.95] | A 87.36% [86.08-88.65] | A 88.46% [87.68-89.25] |

Abbreviations: A, all articles; D, development set; V, validation set
this search resulted in a sensitivity of 80.01% (PubMed: 80.89%). Since the value for the sensitivity estimate in Ovid lies within the confidence interval of the estimate in PubMed (and thus the confidence intervals overlap), no significant difference is assumed.

5 | DISCUSSION

The aim of the present analysis was to develop study filters for identifying controlled NRS in PubMed and Ovid MEDLINE. We developed two search filters (one maximizing sensitivity and one maximizing specificity) for this purpose.

The first search filter maximizes sensitivity and should be applied when the aim is to conduct a comprehensive search. Ideally, when a comprehensive search is conducted, the search in MEDLINE is accompanied by a search in at least one other database, as well as at least one other information source, such as scanning reference lists and searches in study registries.20,21

The second search filter maximizes specificity and could be applied in more focused searches when completeness is not a primary goal. In addition, the filter could also be applied when the search for controlled NRS produces too many hits with the sensitive filter and the search needs to be restricted to make the number of hits manageable. To increase sensitivity, information retrieval could then include other search techniques, such as the “similar articles” function and co-citation searching or citation tracking. There is evidence that these techniques are efficient additional approaches.22 However, there is only limited evidence as to what extent they are able to fill the gap between a focused and a comprehensive bibliographic search.

The manual check of studies that were relevant but not found with the study filters highlights an important issue: Some studies cannot be identified, as the information in the title, abstract or MeSH terms does not indicate the study type. Glanville3 drew a similar conclusion and noted that the identification of NRS should focus on the topic investigated rather than on a specific study design. In our opinion, suitable filters are now available with the two new filters, but limitations remain. Several authors have highlighted these limitations in articles addressing indexing practices for NRS.3-5

The search for NRS using study filters will always have limitations, regardless of the resources invested in searching. Therefore, the amount of resources invested and whether a comprehensive study pool is required for a systematic review should be carefully considered. As stated, NRS are generally more prone to bias, and including them in systematic reviews provides an uncertain evidence base for conclusions. Only very large effects found in NRS can indicate a clinically meaningful effect of an intervention.23 It could therefore be questioned whether a complete study pool is necessary in principle. In our experience, the presence or absence of very large effects can be reliably determined with a few relevant NRS. In assessments where, for example, a large study pool is expected, it could be sufficient to conduct a more specific search. However, the impact of undetected studies on the conclusions of systematic reviews still needs to be further investigated before such an approach is widely adopted.

5.1 | Comparison with other filters

No other NRS filter limits the study type to controlled NRS. Therefore, the performance of other NRS filters is not comparable to our filter.

In the analysis published by IQWiG in 2018, none of the NRS filters examined achieved sufficient sensitivity (≥92%).2 One reason could be that the majority of these filters do not meet current quality standards; of the 14 NRS filters identified, only five can be classified as third-generation filters (objective approach to filter design) according to the definition by Jenkins24. The present filter would qualify as a third-generation filter, as in addition to comparing the filter with a gold standard, we employed objective word frequency analysis methods.

A recent methodological Cochrane review by Li5 assessed the performance of search strategies with methodological filters to identify observational studies. They concluded that the filters analysed had several limitations and that “Future studies should aim to use large reference standards from a range of developmental systematic reviews. They should also include external validation to help determine the generalizability of the methodological filters”.5 The number of references used in previous analyses was quite limited: Fraser4 tested their filters on 217 observational studies extracted from 1 systematic review and validated them against 69 references (included in 2 systematic reviews); the gold standard in Furlan25 comprised 93 relevant references for testing extracted from 4 Cochrane reviews; and Royle tested 424 studies (included in 20 health technology assessment reports).26 The number of references used for the development and validation of our filter exceeds the previous numbers considerably, as it was based on 2110 publications generated from 271 Cochrane Reviews by 41 different Cochrane groups and covering a wide range of topics.2 The intended minimum number of references for the development and validation set was pre-planned by means of a sample size calculation.

It is likely that the objective approach to filter design (third generation) and the sample size calculation improved the performance of the study filters as well as robustness and transferability.
5.2 Limitations

We performed an analysis at the publication level, as it would have been too time-consuming to distinguish between the publication and study level when generating the development and validation set. However, for 42 of 140 publications that were not found through the sensitive filter, at least one other publication had been identified through the filter. Thus, the sensitivity of the study filter was higher at the study level than at the publication level (95.27% vs 93.24%).

There are well-known limitations in the methods we applied. Firstly, the use of the relative recall method to generate the development and the validation set may result in limited generalizability. Secondly, more than half of the Cochrane reviews had a search block for NRS in their bibliographic searches, which might cause bias. Still, references were also detectable via the other information sources, so that this possible limitation might be negligible. Furthermore, the use of additional manual adjustments to generate the filter contradicts an objective approach and may affect performance measures.

Although we used a different set of references to develop and to validate the search filters, this might not be considered a true external validation, which would normally involve the use of an independent set of references (eg, extracted from non-Cochrane systematic reviews) potentially resulting in different results for performance measures. Another limitation is that the set of irrelevant references was drawn from a different pool than the reference set and does not reflect real irrelevant references. Thus, values for specificity, precision, and accuracy are distorted and do not reflect performances in real searches. Precision in the report is provided as information for librarians, but must be interpreted as specific to the set of references created.

Finally, a potential limitation of our filters is that they summarize all NRS study types with a control group. This might not meet the needs of researchers who are only interested in specific study types (eg, cohort studies). However, our approach is in line with the Cochrane handbook, which suggests placing "emphasis on specific features of study design [...] rather than 'labels' for study design [...]". Therefore, we aimed to develop a study filter that does not specifically cover individual NRS study types, but rather identifies NRS with the important feature of a control group. Finally, cohort and case-control studies produce similar effect estimates, so jointly searching for both designs also seems reasonable.

6 CONCLUSION

We developed two new search filters that can assist clinicians and researchers in identifying controlled NRS in PubMed and Ovid MEDLINE. The first filter maximizes sensitivity and should be used when a comprehensive search is needed. The second filter maximizes specificity and should be used when a more focused search is sufficient.

ACKNOWLEDGEMENT

The authors thank Natalie McGauran for editorial support and Fabian Lotz for statistical advice.

CONFLICT OF INTEREST

The author reported no conflict of interest.

DATA AVAILABILITY STATEMENT

The datasets used and/or analysed during the current study are available from the corresponding author on request.

ORCID

Siw Waffenschmidt https://orcid.org/0000-0001-6860-6699

REFERENCES

1. Deeks JJ, Dinnes J, D’Amico R, et al. Evaluating non-randomised intervention studies. Health Technol Assess. 2003;7(27):1-173.
2. Hausner E, Metzendorf MJ, Richter B, Lotz F, Waffenschmidt S. Study filters for non-randomized studies of interventions consistently lacked sensitivity upon external validation. BMC Med Res Methodol. 2018;18(1):171.
3. Glanville J, Eysers J, Jones AM, et al. Quasi-experimental study designs series; paper 8: identifying quasi-experimental studies to inform systematic reviews. J Clin Epidemiol. 2017;89:67-76.
4. Fraser C, Murray A, Burr J. Identifying observational studies of surgical interventions in MEDLINE and EMBASE: a validation of a search filter. BMC Med Res Methodol. 2006;6:41.
5. Li L, Smith HE, Atun R, Tudor Car L. Search strategies to identify observational studies in MEDLINE and Embase. Cochrane Database Syst Rev. 2019;(3):MR000041.
6. OCEBM Levels of Evidence Working Group. The Oxford 2011 levels of evidence. http://www.cebm.net/index.aspx?o=5653. Accessed October 21, 2019.
7. Lefebvre C, Glanville J, Beale S, et al. Assessing the performance of methodological search filters to improve the efficiency of evidence information retrieval: five literature reviews and a qualitative study. Health Technol Assess. 2017;21(69):1-148.
8. Bak G, Mierzwiniski-Urban M, Fitzsimmons H, Morrison A, Maden-Jenkins M. A pragmatic critical appraisal instrument for search filters: introducing the CADTH CAI. Health Info Libr J. 2009;26(3):211-219.
9. 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):284-289.
10. Wilczynski NL, McKibbon KA, Walter SD, Garg AX, Haynes RB. MEDLINE clinical queries are robust when searching in recent publishing years. J Am Med Inform Assoc. 2013;20(2):363-368.
11. Sampson M, Zhang L, Morrison A, et al. An alternative to the hand searching gold standard: validating methodological search filters using relative recall. *BMC Med Res Methodol*. 2006; 6:33.
12. Ijaz S, Verbeek JH, Mischke C, Ruotsalainen J. Inclusion of nonrandomized studies in Cochrane systematic reviews was found to be in need of improvement. *J Clin Epidemiol*. 2014;67 (6):645-653.
13. Hartling L, Bond K, Santaguida PL, Viswanathan M, Dryden DM. Testing a tool for the classification of study designs in systematic reviews of interventions and exposures showed moderate reliability and low accuracy. *J Clin Epidemiol*. 2011;64(8):861-871.
14. Haynes RB, Wilczynski N, McKibbon KA, Walker CJ, Sinclair JC. Developing optimal search strategies for detecting clinically sound studies in MEDLINE. *J Am Med Inform Assoc*. 1994;1:447-458.
15. Wilczynski NL, Haynes RB, Lavis JN, Ramkissoonsingh R, Arnold-Oatley AE. Optimal search strategies for detecting health services research studies in MEDLINE. *CMAJ*. 2004;171 (10):1179-1185.
16. Hausner E, Waffenschmidt S, Kaiser T, Simon M. Routine development of objectively derived search strategies. *Syst Rev*. 2012;1:19.
17. McMaster University. McMaster+. https://hiru.mcmaster.ca/hiru/HIRU_McMaster_PLUS_projects.aspx. Accessed August 26, 2019.
18. McMaster University. Inclusion Criteria. https://hiru.mcmaster.ca/hiru/InclusionCriteria.html. Accessed October 30, 2019.
19. Health Information Research Unit, McMaster University. Reading criteria: full [unpublished]. 2017.
20. Shea BJ, Grimshaw JM, Wells GA, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol*. 2007; 7:10.
21. European Network for Health Technology Assessment. Process of information retrieval for systematic reviews and health technology assessments on clinical effectiveness. https://www.eunethta.eu/wp-content/uploads/2018/01/Guideline_Informati on_Retrieval_V1-2_2017.pdf. Updated December 2017. Accessed June 6, 2019.
22. Hausner E, Waffenschmidt S. Value of using different search approaches. http://vortal.htai.org/?q=node/993. Updated May 28, 2019. Accessed September 19, 2019.
23. Guyatt GH, Oxman AD, Sultan S, et al. GRADE guidelines: 9. Rating up the quality of evidence. *J Clin Epidemiol*. 2011;64 (12):1311-1316.
24. Jenkins M. Evaluation of methodological search filters: a review. *Health Info Libr J*. 2004;21(3):148-163.
25. Furlan AD, Irvin E, Bombardier C. Limited search strategies were effective in finding relevant nonrandomized studies. *J Clin Epidemiol*. 2006;59(12):1303-1311.
26. Royle P, Waugh N. Literature searching for clinical and cost-effectiveness studies used in health technology assessment reports carried out for the National Institute for clinical excellence appraisal system. *Health Technol Assess*. 2003;7(34):1-51.
27. Boynton J, Glenville J, McDaid D, Lefebvre C. Identifying systematic reviews in MEDLINE: developing an objective approach to search strategy design. *J Inform Sci*. 1998;24:137-154.
28. Reeves BC, Deeks JJ, Higgins JPT, Shea B, Tugwell P, Wells GA. Chapter 24: including non-randomized studies on intervention effects. In: Higgins JPT, Thomas J, Chandler J, et al., eds. Cochrane handbook for systematic reviews of interventions version 6.0; 2019: http://handbook.cochrane.org. (updated July 2019)
29. Lanza A, Ravaud P, Riveros C, Dechartres A. Comparison of estimates between cohort and case-control studies in meta-analyses of therapeutic interventions: a meta-epidemiological study. *PLoS One*. 2016;11(5):e0154877.

**SUPPORTING INFORMATION**

Additional supporting information may be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Waffenschmidt S, Navarro-Ruan T, Hobson N, Hausner E, Sauerland S, Haynes RB. Development and validation of study filters for identifying controlled non-randomized studies in PubMed and Ovid MEDLINE. *Res Syn Meth*. 2020;11:617–626. [https://doi.org/10.1002/jrsm.1425](https://doi.org/10.1002/jrsm.1425)