Complementary approaches to searching MEDLINE may be sufficient for updating systematic reviews

Sampson, Margaret; De Bruijn, Berry; Urquhart, Christine; Shojania, Kaveh

Publisher's version / Version de l'éditeur:
https://doi.org/10.1016/j.jclinepi.2016.03.004
Journal of Clinical Epidemiology, 78, October 2016, pp. 108-115, 2016-03-11

ELSEVIER

Journal of Clinical Epidemiology ■ (2016) ■

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

Complementary approaches to searching MEDLINE may be sufficient for updating systematic reviews

Margaret Sampson[a,*], Berry de Bruijn[b], Christine Urquhart[c], Kaveh Shojania[d]

[a]*Library and Media Services, Children's Hospital of Eastern Ontario, 401 Smyth Road, Ottawa, Ontario K1H 8L1, Canada*

[b]*National Research Council—Information and Communications Technologies Portfolio (NRC-ICT), 1200 Montreal Rd, bldg M-50, Ottawa, Ontario K1A 0R6, Canada*

[c]*Department of Information Studies, Sunnybrook Health Sciences Centre, Aberystwyth University, Llanbadarn Campus, Aberystwyth, Wales SY23 3AL, United Kingdom*

[d]*Centre for Quality Improvement and Patient Safety, Room H468, 2075 Bayview Avenue Toronto, Ontario M4N 3M5, Canada*

Accepted 7 March 2016; Published online xxxx

Abstract

Objectives: To maximize the proportion of relevant studies identified for inclusion in systematic reviews (recall), complex time-consuming Boolean searches across multiple databases are common. Although MEDLINE provides excellent coverage of health science evidence, it has proved challenging to achieve high levels of recall through Boolean searches alone.

Study Design and Setting: Recall of one Boolean search method, the clinical query (CQ), combined with a ranking method, support vector machine (SVM), or PubMed-related articles, was tested against a gold standard of studies added to 6 updated Cochrane reviews and 10 Agency for Healthcare Research and Quality (AHRQ) evidence reviews. For the AHRQ sample, precision and temporal stability were examined for each method.

Results: Recall of new studies was 0.69 for the CQ, 0.66 for related articles, 0.50 for SVM, 0.91 for the combination of CQ and related articles, and 0.89 for the combination of CQ and SVM. Precision was 0.11 for CQ and related articles combined, and 0.11 for CQ and SVM combined. Related articles showed least stability over time.

Conclusions: The complementary combination of a Boolean search strategy and a ranking strategy appears to provide a robust method for identifying relevant studies in MEDLINE.

Keywords: Information retrieval; Systematic reviews; Support vector machine; Clinical query; PubMed similar articles; Searches; Updating; MEDLINE

1. Introduction

Systematic review searches need to have high recall. Mechanisms to achieve this usually include expansive Boolean searches of multiple databases. This approach leads to long development times for the searches [1], necessitates accessing multiple sources that may not be accessible in some institutions [2], and entails time-consuming removal of duplicate records for articles indexed in more than one of the databases searched [3].

MEDLINE gives excellent coverage of most biomedical topics, in particular, intervention studies. However, in 1994, a landmark article by Dickersin et al. [4] established that only about half the studies included in systematic reviews were identified through MEDLINE. Recent research has demonstrated that a much higher percentage is present in MEDLINE, but sometimes their retrieval is problematic [5−8].

Successful retrieval through a Boolean search is operator dependent, with search performance being influenced by skill of the indexer and the searcher. The search of multiple databases can therefore be helpful. The target records may be indexed differently in the second, third, or subsequent source searched, increasing the probability of a match between the search terms entered and the indexing of the additional records. A text search, querying terms appearing in the title or abstract, may help improve retrieval, as may the use of indexing terms that are broader (less specific) than, or related to, the single best indexing term. These tactics on

Conflict of interest: None.

Funding: Some of this work was funded by the US Agency for Healthcare Research and Quality, Department of Health and Human Services (contract no. 290-02-0021).

* Corresponding author. Tel.: 613-737-7600.

E-mail address: msampson@cheo.on.ca (M. Sampson).

http://dx.doi.org/10.1016/j.jclinepi.2016.03.004

0895-4356/

What is new?

Key findings

- Searches using known relevant studies and the similar articles feature of PubMed will identify and rank additional articles of potential relevance.
- For a given question, if the Boolean search has low recall, the ranking search tends to have higher recall, and vice versa. The two approaches complement each other.
- The precision of this complementary paired method appears better than the precision of exhaustive Boolean searches.

What this adds to what was known?

- The paired approach performed well regardless of which of two tested similarity searches were used. It is the use of independent retrieval methods that is important.

What is the implication and what should change now?

- Using the simple and universally available PubMed similar feature makes this paired approach practical for most systematic review teams.
- If the paired complementary approach is used, the recall may be sufficient to consider using only MEDLINE.

the part of the searcher lead to large retrievals with low precision [9]. These resulting problems may be particularly challenging for complex or newly emerging interventions with highly variant terminology, where alternatives to traditional Boolean searches have been sought [10].

Most recent systematic review information retrieval research has focused on text mining approaches [11−14]. These approaches often harvest an intentionally overinclusive set of records and then use machine learning, similarity ranking and other techniques to refine the set to identify the material most likely to be relevant, thereby reducing human screening effort. These methods show promise, but are not yet widely available to reviewers.

We examined one method, support vector machine (SVM), and compared it with a simple and readily available method based on the PubMed similar articles feature. We call this method *related articles* to distinguish the method from the similar articles feature itself. We paired both with a focused Boolean search within MEDLINE. We tested this approach in an updating context where studies included in the original review comprise the reference standard for SVM and seed articles for the PubMed-related articles search. We therefore sought to determine if a focused Boolean search paired with one of the search methods that does not depend on operator skill could provide consistently complete retrieval of relevant new studies.

Comparison of a number of searches, including two tested here, has been previously reported [15]. In this current article, two of the most successful methods in a larger sample of 72 journal-published systematic reviews, clinical query (CQ), and PubMed-related articles are tested in a cohort of six updated Cochrane reviews, as well as in a previously unreported sample of 10 Agency for Healthcare Research and Quality (AHRQ) evidence reports. The Cochrane reviews provide a true gold standard as updates were made by the review team based on evidence identified through comprehensive searches; however, the new relevant studies proved fairly easy to find. The replication in the AHRQ cohort, of more complex interventions, provided a means to validate the generalizability of the approach [16]. All records were assessed for eligibility by two reviewers, and this complete screening allowed the precision of the methods to be calculated for the first time.

Other research [17,18] suggests that searches using the PubMed similar articles feature are effective in increasing recall of relevant items for reviews or more general clinical searching when combined with a Boolean-type search of MEDLINE. We tested an additional search method, SVM, in the Cochrane and AHRQ samples to permit comparison with our PubMed-related articles search and assess whether the complementary effect generalized beyond the PubMed similar articles method.

The aim of this article was to test whether the combined approach of a focused Boolean search paired with a second search using the similar articles feature of PubMed or SVM can yield high recall with reasonable precision.

2. Methods

2.1. Formation of the study cohorts

This analysis uses a data set created for an updating study sponsored by AHRQ [16]. Methods for the selection of the cohorts, search approaches tested, and rigorous mechanisms to screen search results for relevance have been previously reported, along with the criteria to determine if a review was in need of update [16].

Briefly, Cochrane reviews were identified through a search of the ACP Journal Club database (Ovid) using the strategy:

1. review$.ti. 2. meta-analy$.mp. 3. data sources.ab. 4. (search$ or MEDLINE®).ab. 5. or/1-4 6. limit 5 to articles with commentary.

AHRQ reports were identified through the PubMed query ‘‘Evid Rep Technol Assess (Summ)’’[Journal:__jrid21544]. Screening was undertaken in two phases with two reviewers reaching consensus on eligibility. Screening continued until the predetermined sample size was reached.

For inclusion, all reviews must have included randomized controlled trials (RCTs) and provided meta-analysis for at least one outcome. The MEDLINE search strategies had to be reported in enough detail to permit replication. Fifteen AHRQ evidence reports thought likely to have important new evidence were selected and used to validate the updating methods used in the main cohort.

Cochrane reviews were selected to meet minimum quality and relevance standards, as defined by ACP Journal Club [19]. We also required that the review had been updated and that the text of both the original review and an updated version was available. The search for the original review must have included MEDLINE and at least one other electronic bibliographic database and one or more nondatabase method such as hand searching or checking reference lists. Such comprehensive searching was likely to identify most relevant literature and form a useful training set of included examples. At least 10 RCTs or quasi-RCTs must have been included in the original review to give an adequate training set for SVM.

2.2. *Test searches*

2.2.1. *Clinical query*

Boolean searches were developed by a librarian experienced in systematic review searches using a protocol (Appendix A at www.jclinepi.com). Search strategies were deliberately simple, usually consisting of two or three Medical Subject Heading (MeSH) terms representing the population and intervention of the review. This search was limited by a CQ filter to "therapy (best balance of sensitivity and specificity)" that is, "Randomized controlled trial.pt. or randomized.mp. or placebo.mp." [20]. Boolean searches were run in OVID MEDLINE.

2.2.2. *Related articles*

Searches using the PubMed similar articles feature were performed using the PubMed unique identifiers (PMID) of the three newest and three largest studies included in the original review as seed articles. There was no replacement in the event of overlap between the largest and newest sets, or if one of those studies was not indexed in MEDLINE. (All studies included in the original reviews were checked for indexing status in MEDLINE.) The resulting set was limited to the publication type RCT and date limited to the period since the search date of the original review.

2.2.3. *Support vector machine*

The SVM searches used for this project have not been previously described so the methods will be reported in more detail. SVM is a well established and broadly applied classification method from machine learning [21,22]. A classification task is closely related to an information retrieval task, once the task is framed as predicting a "relevant" vs. "nonrelevant" label to instances (documents) in the collection. As SVM is a supervised algorithm, it requires training examples with known labels. It then places these training examples in a high-dimensional vector space in such a way that examples of one class are separated from examples of the other class with the greatest distance to the separating boundary. Predicting the class of a new instance happens by placing it in the same space and determining on which side of the boundary it falls. The distance to the boundary is an indication of the prediction confidence and is used to rank instances (documents) as is essential in information retrieval tasks or to provide a cutoff threshold.

In our setup, the training set consisted of search results from the original review formed with included instances being the studies included in the review (the relevant retrievals) and excluded instances being the studies found by the search used in the original review but excluded from the review (irrelevant retrievals). The new examples to be classified were articles indexed since the date of the search performed in the original review.

SVM searches were run on MEDLINE data stored locally at the National Research Council of Canada. MEDLINE was refreshed before running the searches. The MEDLINE records were represented as bags of features, where features consisted of lowercased words and word combinations (up to four words) from the title and abstract fields, full MeSH terms, and contents of all other fields in the MEDLINE record. Features were uniformly weighted. The SVM implementation used was SVM Light [23] used with a linear kernel.

The approach was piloted with one AHRQ evidence report, and various configurations were tried, observing the placement of several known new relevant studies within the retrieval. The MEDLINE set was filtered with the revised Highly Sensitive Search Strategy (HSSS) [24]. Preliminary testing indicated that this filter would retrieve 99.1% of MEDLINE-indexed relevant new evidence from the larger cohort of journal-published systematic reviews ($n = 72$) and Cochrane reviews ($n = 27$). This HSSS filter was added primarily to improve processing speed, relative to running the classifier against all of MEDLINE. With the filter, processing for one systematic review took about 10 minutes. Next, a coarse-grained filter was applied; included and excluded instances were combined in a single set labeled POS. Adjacent PubMed IDs (the next higher PMID to each member of POS, as long as that higher number was not itself a member of POS) were combined in a set to represent WORLD. The model was trained on POS vs. WORLD, using words and phrases from title, abstract, author names, and journal name. This model was run against the records passing the HSSS filter.

Four different subsets of the SVM retrieval were examined—one set included all records with a relevance score of 0.5 or more, up to a maximum of 200 retrievals (SVM200point5). Other sets consisted of all records with relevance scores of 0.95 or more (SVM95), 0.90 or more (SVM90), and 0.80 or more (SVM80). PubMed ID and relevance scores were recorded, and PubMed IDs were incorporated into the database of records for the reviewers to screen. Only the results from SVM200point5 are reported here and are described as SVM.

2.3. Determining performance of the test searches

For the updated Cochrane reviews, reference standard articles were those studies included in the updated review that were not in the original and which entered MEDLINE after the search date of the original review. For the AHRQ evidence reports, the full retrieval set was screened. Records found relevant by the consensus of reviewers were considered reference standard articles. Recall was the proportion of reference standard articles identified by the search:

$$\frac{\text{Number of reference standard articles found}}{\text{Total number of reference standard articles}}$$

BioVenn software was used to analyze the overlap between the retrieval of relevant articles from the three searches and to create Venn diagrams [25].

All included reviews were classified into clinical area based on factors such as ISI journal classification, the Cochrane Collaboration Review Group where the topic might be placed and the high level MeSH term under which the population (condition) would be indexed. Performance of the searches in different clinical areas was displayed graphically for the searches both alone and in combinations, to allow examination of differences in parallelism, level, and flatness.

Precision is the proportion of all retrieved records that are relevant:

$$\frac{\text{Number of reference standard articles found}}{\text{Total number of records retrieved}}$$

The inverse of precision is the number needed to read to find one eligible study, thus precision influences the work of the review. Precision was calculated only for the AHRQ evidence reviews. Because not all candidates retrieved by the Cochrane searches were assessed, precision could not be established in that sample.

2.4. Determining stability of results over time

The CQ and related article searches, originally run in March 2008, were repeated in February 2015 in the AHRQ cohort. Recall of relevant items was compared with the original retrievals to determine if any changes introduced by National Library of Medicine might invalidate findings. As SVM configuration is under the control of the investigators, it was not retested.

3. Results

Six Cochrane reviews and 10 AHRQ evidence reports met all inclusion criteria (Appendix B at www.jclinepi.com). Characteristics of the included reviews are shown in Table 1.

Table 1. Characteristics of the included reviews

Characteristic	Cochrane reviews; *N*	AHRQ evidence reports; *N*
Therapy evaluated		
Medications	5	9[a]
Medical devices	1	1
Procedures	—	3
Clinical topic area		
Cardiac and cardiovascular systems	—	3
Critical care	1	—
Endocrinology and metabolism	—	1
Infectious disease	3	—
Clinical neurology	—	1
Obstetrics and gynecology	—	2
Oncology	—	1
Peripheral vascular diseases	—	1
Psychiatry	—	1
Respiratory systems	1	—
Urology and nephrology	1	—
Publication period		
March 1997—April 1999	3	—
May 1999—June 2001	1	3
July 2001—August 2003	2	3
September 2003—December 2005	—	4
Median included trials	17 (IQR, 14—20)	96 (IQR, 31.75—121.5)
Median included participants	8,679 (IQR, 4,085—50,109)	22,830 (IQR, 14, 172—49,687)
MEDLINE coverage of included articles	99/107 (92.5%)	969/980 (98.8%)

Abbreviations: AHRQ, Agency for Healthcare Research and Quality; IQR, interquartile range.

[a] Some AHRQ reviews included more than one class of therapy.

Recall of new relevant studies is shown in Table 2. The updated Cochrane reviews had 20 new studies identified by the review authors and added in the updates. Recall of these 20 ranged from 1.00 for the related articles method to a low of 0.80 CQs (Table 2). Our team identified 277 new studies as relevant for inclusion in the 10 evidence reports that were updated. All test searches showed lower recall for this cohort than for the Cochrane reviews and here the CQ outperformed both ranking searches (Table 2).

3.1. Recall of CQ combined with a ranking method

Of the 297 new relevant studies identified, the combination of CQ and related articles searches identified 270 (overall recall of 0.91). Recall was 1.00 (20/20) in the Cochrane cohort and 0.90 (250/277) in the AHRQ cohort.

The combination of CQ and SVM identified 263 studies (overall recall of 0.89). Recall was 1.00 (20/20) in the Cochrane cohort and 0.88 (243/277) in the AHRQ cohort. When all three methods were used together, recall was 0.997 with 296 of the 297 studies identified.

Table 2. Recall of eligible studies by the search methods

	Cochrane reviews		AHRQ evidence reports	
Retrieval method	***N***	**Recall**	***N***	**Recall**
Clinical query	16	0.80	188	0.68
Related articles	20	1.00	176	0.64
SVM	19	0.95	129	0.47
Total	20		277	

Abbreviations: AHRQ, Agency for Healthcare Research and Quality; SVM, support vector machine.

Across the two cohorts, the overlap and unique component from the retrieval of relevant records by the test searches was examined (Fig. 1).

3.2. Consistency across clinical areas

Thirteen clinical areas were represented in the larger study that included 72 journal-published reviews, but only eight of those clinical areas had two or more new relevant studies in the smaller Cochrane and AHRQ cohorts. Fig. 2 shows recall of new studies by the three types of searches, for these eight clinical areas.

Combined recall of the CQ Boolean search paired with a ranking search is shown in Fig. 3. The combination showed complete recall of relevant new studies in three of eight areas when CQ was paired with related articles and recall of 0.67 or higher in all areas. Four of eight clinical areas had complete recall for the combination of CQ and SVM, and recall was 0.80 or higher in all areas. Even for peripheral vascular disease, where all three searches performed fairly poorly individually, recall was 0.76 for the CQ/related articles and 0.90 in the CQ/SVM pairing.

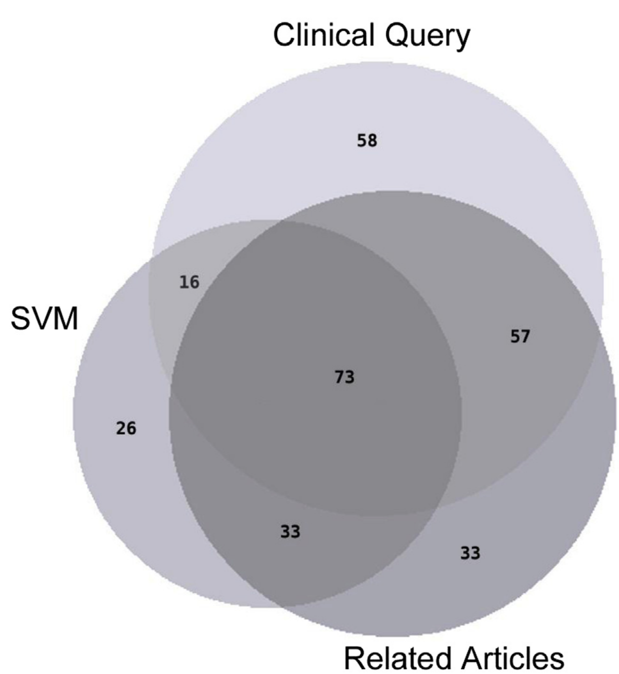

Fig. 1. Overlapping and unique retrieval of relevant studies by each search method. The size of the circle is proportional to the size of the retrieval, but the overlap and unique portions are approximations. Exact figures for the components are labeled. *Abbreviations:* SVM, support vector machine.

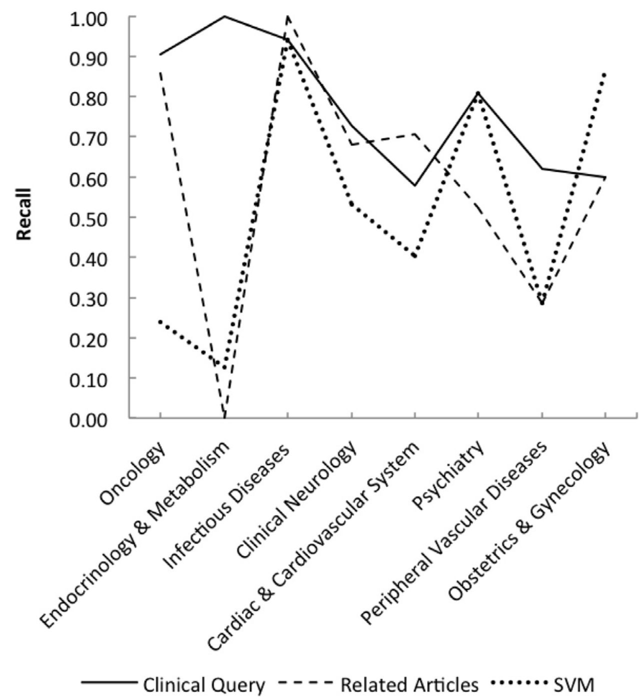

Fig. 2. Recall of new studies by clinical area for each search method. *Abbreviations:* SVM, support vector machine.

3.3. Search precision

Precision, across the 10 AHRQ reviews, was 0.11 for the CQ, 0.22 for related articles, and 0.19 for SVM (Table 3).

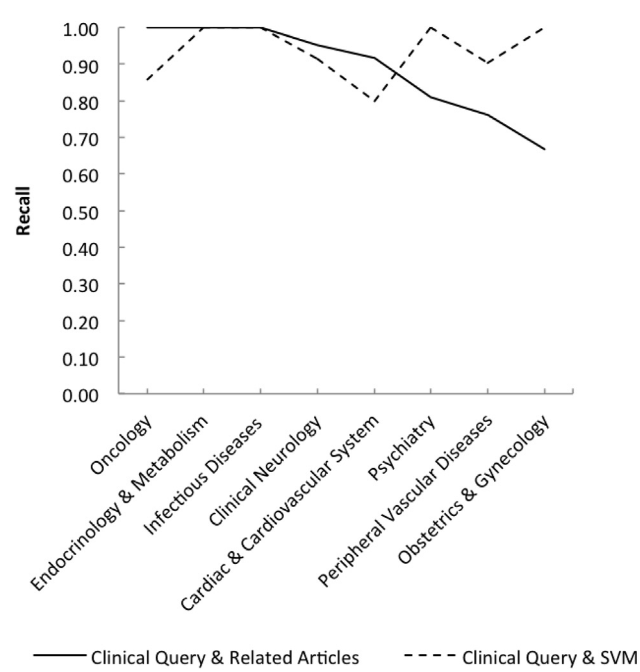

Fig. 3. Recall of new studies by clinical area for search methods in combination. *Abbreviations:* CQ, clinical query; SVM, support vector machine.

Table 3. Overall precision in the AHRQ Sample

Retrieval method	Eligible studies retrieved	No. of candidates retrieved	Precision
Clinical query	187	1,637	0.11
Related articles	176	814	0.22
SVM	128	659	0.19
Clinical query + related articles	250	2,318[a]	0.11
Clinical query + SVM	243	2,247[a]	0.11
Clinical query + related articles + SVM	276	3,264[a]	0.08

Abbreviations: AHRQ, Agency for Healthcare Research and Quality; SVM, support vector machine.

[a] Number of candidates after removal of duplicate records.

Precision for the individual reviews ranged from 0.0 to 0.38 for CQs, 0.004 to 0.80 for related articles, and 0.01 to 0.43 for SVM. Overall precision was 0.11 for CQ and either related articles or SVM. Overall precision was 0.08 when CQ and both related articles and SVM were used. For ranked searches where a fixed number of records will be screened, that number is always the denominator, so precision will tend to increase as the number of relevant records increases—the maximum retrieval size for SVM was capped at 200 records.

3.4. Retrieval consistency over time

Searches were retested for the AHRQ cohort using PubMed results obtained February 24, 2015. Recall of the related articles searches across the 10 AHRQ reports was 0.64 originally and 0.50 when repeated (Fig. 4). Of the three extreme cases, one rose from 0.00 to 0.80, one remained at 0.00, and one dropped from 1.00 to 0.60. CQ had overall recall of 0.68 originally and 0.70 when repeated. Changes in extreme cases were minimal. Overall combined recall of CQ and related article was 0.90 originally, falling to 0.85 when repeated. Considering the extreme values, all review with combined recall of 1.00 originally remained at 1.00, whereas the one review with 0.00 recall in combination showed recall of 0.80 when repeated.

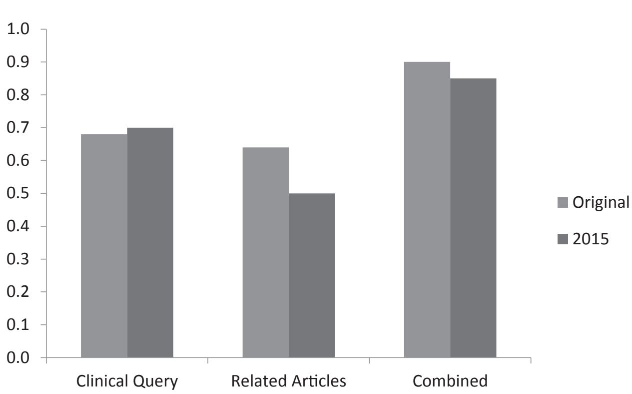

Fig. 4. Recall of new studies for search methods, alone and in combination originally (left) and when retested in 2015 (right).

4. Discussion

We expected that the sophisticated SVM approach would perform well when paired with more stripped-down, focused Boolean searches. Indeed, the two together were able to replace the multidatabase, multimodel searches used by the original review teams in updating the Cochrane searches. The new relevant studies for the Cochrane reviews may have been relatively easy to find, but the pairing was also effective in the AHRQ set, and there showed better precision than is usually seen with traditional searches for systematic review of RCTs [9]. The AHRQ set was formed from the new studies identified for more complex interventions in a rigorous, well-funded study [16].

We were surprised that the related articles approach performed almost as well as SVM when used in combination with the CQ. The related article search has advantages over SVM—it requires far less data preparation, and no special software is needed for its use. This makes it useful not only in updating, but also, if appropriate seed articles can be found, in original reviews.

That both SVM and related articles sometimes showed poor recall when used alone, but consistently good recall when used with a Boolean method suggests that there is real benefit in using complementary search methods to querying MEDLINE. The related article method is not very time consuming, and easily added to other planned search efforts, while the third method used here, SVM, is more technical. The combined performance of all three methods was surprising, but such a setup might become unpractical in day-to-day use.

Other investigators have similar findings. Examination of the supplemental material presented in the Appendix C to the article by Waffenschmidt et al. [18] reveals that across the 19 reviews tested, the combination of a simple structured Boolean search (SSBS) and the first 50 retrieval of PubMed similar articles showed complete retrieval of all reference standard articles in 14 of 19 reviews and never less than 0.90 recall (Appendix C at www.jclinepi.com presents data from the table by Waffenschmidt graphically). Waffenschmidt concluded "the combination of these two search techniques that are independent of each other seems to compensate the respective weaknesses."

SSBS search by Waffenschmidt was constructed in PubMed from search terms selected for the indication and intervention with PubMed's narrow CQ filter (category: therapy). Their search using the similar articles feature (RelCits) did not use a set of seed articles; rather, the RelCits function was applied for each relevant citation previously identified in PubMed. Their test articles were the included studies in 19 systematic reviews of drugs.

Agoritsas et al. also described search construction methods for searches based on CQ and PubMed similar articles, tested for their ability to retrieve the included studies of 30 Cochrane reviews [17]. Although their search construction methods differed from those used here, both searches were derived using standard methods. Their approach to the structured Boolean search used terms from the population, intervention, and comparison with the CQ, limited to humans and English. Two clinicians selected the PubMed similar articles seeds from the initial PubMed retrieval. They noted that no one method provided consistently high retrieval [17].

Thus, the robust nature of this pairing of Boolean and non-Boolean searches has been shown in several contexts, giving support to the hypothesis that methods such as related articles and SVM can compensate for the variation inherent in selecting search terms or assigning subject headings during indexing.

Considering stability of the searches over time, the minimal change seen in the performance of the CQ is likely explained by indexing changes of a few records. The computation appears not to have changed. This suggests that indexing changes may also impact SVM results over time; however, this is likely to result in improved performance, as was seen with the CQ. The computation of similar articles is still described by National Library of Medicine as being based on the algorithm described by Kim et al. [26]. That algorithm would suggest changes over time in the nearest neighbor score as the frequency of certain terms in the MEDLINE corpus changes. A full exploration of the changes in nearest neighbor scores of similar articles over time is beyond the scope of this study. It should be noted that when the similar articles feature was initially studied, it was simple to submit seed articles, and then add limits such as date or RCT publication type. On retesting, such additional limits were more complex to apply, and the elink utility seemed to be the most practical approach to identifying the related articles [27].

The benefit of the complementary searches, relative to Boolean searching alone, was greatest for AHRQ evidence reports and less for the simpler Cochrane reviews, suggesting that this approach may be particularly useful for more complex evidence. There are examples of its utility in the literature. For example, in a realist review of a multidisciplinary body of literature identifying six domains of Clinical Practice Guideline implementability, the use of PubMed's similar articles feature as a third stage of searching identified 131 records of which 104 were relevant [28]. In a systematic review of evidence on the links between patient experience and clinical safety and effectiveness, the authors used PubMed similar articles to snowball on articles identified through an EMBASE search to overcome the limitations of predefined searches for complex evidence [29].

There is always the possibility that a search, or even a pair of searches, will fail. One method to detect such failures is to test whether the search strategies find known relevant items [30]. In the updating case, this is easily done using the included studies of the original review as a test set, allowing the review to determine the MEDLINE coverage of their particular topic at the same time.

4.1. Limitations

There are two limitations to our proposed strategy. Other databases should be searched in the unusual event that numerous studies, representing more than a small proportion of the total *N*, are not included in MEDLINE. Second, when it is important to find articles too new to be indexed by MEDLINE, systematic reviewers may wish to conduct a simple PubMed search limited to the nonindexed subsets [31,32].

5. Conclusion

The general approach of a Boolean plus a ranking search is effective in MEDLINE retrieval for systematic reviews. Very high levels of identification of relevant MEDLINE records, with adequate precision, are possible using a focused Boolean search complemented by a document similarity or ranking method. The efficacy of a focused Boolean search paired with a search using the PubMed similar articles feature is in agreement with previous studies [17,18], and this study shows that this complementary effect also occurs when SVM is used as the additional method. The benefit of using two complementary approaches to achieving high recall in MEDLINE is a robust effect.

PubMed-related articles is a parsimonious method, as it is readily available to all review teams without cost. The approach is robust across clinical domains, and the effect has now been demonstrated in several samples. The method may be sufficient for updating systematic reviews of interventions and may be used for new reviews of interventions when paired with a trials registry search. It is likely to work in any type of search where MEDLINE provides good subject coverage even if retrieval through traditional search methods has been challenging.

Acknowledgments

Mohamed Ansari and Jun Ji undertook screening and relevance assessment of candidate studies retrieved by the test searches, Tamara Rader developed the Boolean searches, and Raymond Daniel processed related articles searches and obtained full-text articles. David Moher cosupervised Margaret Sampson doctoral dissertation and provided invaluable advice and guidance.

Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.jclinepi.2016.03.004.

References

[1] 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(1):58.

[2] Reveiz L, Ospina E, Zorrilla AFC. Should we consider Embase in Latin America? J Clin Epidemiol 2004;57:866; author reply 867–8.

[3] Rathbone J, Carter M, Hoffmann T, Glasziou P. Better duplicate detection for systematic reviewers: evaluation of Systematic Review Assistant-Deduplication Module. Syst Rev 2015;4(1):6.

[4] Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. BMJ 1994;309:1286–91.

[5] Bramer WM, Giustini D, Kramer BM, Anderson P. The comparative recall of Google Scholar versus PubMed in identical searches for biomedical systematic reviews: a review of searches used in systematic reviews. Syst Rev 2013;2(1):115.

[6] Wieland LS, Robinson KA, Dickersin K. Understanding why evidence from randomised clinical trials may not be retrieved from Medline: comparison of indexed and non-indexed records. BMJ 2012;344:d7501.

[7] Suarez-Almazor ME, Belseck E, Homik J, Dorgan M, Ramos-Remus C. Identifying clinical trials in the medical literature with electronic databases: MEDLINE alone is not enough. Control Clin Trials 2000;21:476–87.

[8] O'Leary N, Tiernan E, Walsh D, Lucey N, Kirkova J, Davis MP. The pitfalls of a systematic MEDLINE review in palliative medicine: symptom assessment instruments. Am J Hosp Palliat Care 2007;24:181–4.

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

[10] Greenhalgh T, Peacock R. Effectiveness and efficiency of search methods in systematic reviews of complex evidence: audit of primary sources. BMJ 2005;331:1064–5.

[11] Badgett RG, Dylla DP, Megison SD, Glynn Harmon E. An experimental search strategy retrieves more precise results than PubMed and Google for questions about medical interventions. PeerJ 2015;3:e913.

[12] O'Mara-Eves A, Thomas J, McNaught J, Miwa M, Ananiadou S. Using text mining for study identification in systematic reviews: a systematic review of current approaches. Syst Rev 2015;4(1):5.

[13] Bekhuis T, Demner-Fushman D. Towards automating the initial screening phase of a systematic review. Stud Health Technol Inform [Internet] 2010;160(Pt 1):146–50. Available at http://www.ncbi.nlm.nih.gov/pubmed/20841667. Accessed January 2014.

[14] Wallace BC, Trikalinos TA, Lau J, Brodley C, Schmid CH. Semi-automated screening of biomedical citations for systematic reviews. BMC Bioinformatics 2010;11(1):55. Available at http://www.citeulike.org/user/daforerog/article/6595002. Accessed November 2010.

[15] Sampson M, Shojania KG, McGowan J, Daniel R, Rader T, Iansavichene AE, et al. Surveillance search techniques identified the need to update systematic reviews. J Clin Epidemiol 2008;61:755–62.

[16] Shojania KG, Sampson M, Ansari MT, Ji J, Garritty C, Doucette S, et al. Updating systematic reviews—technical review 16. Health (San Francisco). Rockville, MD: Agency for Health Care Policy and Research; 2007.

[17] Agoritsas T, Merglen A, Courvoisier DS, Combescure C, Garin N, Perrier A, et al. Sensitivity and predictive value of 15 PubMed search strategies to answer clinical questions rated against full systematic reviews. J Med Internet Res 2012;14(3):e85.

[18] Waffenschmidt S, Janzen T, Hausner E, Kaiser T. Simple search techniques in PubMed are potentially suitable for evaluating the completeness of systematic reviews. J Clin Epidemiol 2013;66:660–5.

[19] Purpose and procedure [Internet]. ACP Journal Club. [cited 2007 Feb 7]. Available at http://annals.org/SS/ACPJC_Purpose_and_Procedure.aspx.

[20] Hedges Team. Search strategies for MEDLINE in Ovid Syntax and the PubMed translation. [Internet]. Health Information Research Unit website; 2010. Available at http://hiru.mcmaster.ca/hiru/HIRU_Hedges_MEDLINE_Strategies.aspx. Accessed January 2011.

[21] Vapnik V. The nature of statistical learning theory. New York: Springer; 1995.

[22] Joachims T. Text categorization with support vector machines: learning with many relevant features. Machine Learning: ECML-98. Berlin: Springer; 1998:137–42.

[23] Joachims T. Making large-scale SVM learning practical. In: Schölkopf B, Burges CJC, Smola AJ, editors. Advances in kernel methods: support vector learning [Internet]. MIT Press; 1999. Available at http://dl.acm.org/citation.cfm?id=299094. Accessed September 2015.

[24] Glanville JM, Lefebvre C, Miles JN, Camosso-Stefinovic J. How to identify randomized controlled trials in MEDLINE: ten years on. J Med Libr Assoc 2006;94:130–6.

[25] Hulsen T, de Vlieg J, Alkema W. BioVenn - a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. BMC Genomics 2008 Jan;9(1):488.

[26] Kim W, Aronson AR, Wilbur WJ. Automatic MeSH term assignment and quality assessment. Proc AMIA Symp 2001;319–23.

[27] Sayers E. The E-utilities in-depth: parameters, syntax and more [Internet]. National Center for Biotechnology Information (US); 2015. Available at http://www.ncbi.nlm.nih.gov/books/NBK25499/. Accessed March 2015.

[28] Kastner M, Bhattacharyya O, Hayden L, Makarski J, Estey E, Durocher L, et al. Guideline uptake is influenced by six implementability domains for creating and communicating guidelines: a realist review. J Clin Epidemiol 2015;68:498–509.

[29] Doyle C, Lennox L, Bell D. A systematic review of evidence on the links between patient experience and clinical safety and effectiveness. BMJ Open [Internet] 2013;3(1). Available at http://bmjopen.bmj.com/content/3/1/e001570.abstract Accessed March 2015.

[30] Sampson M, Mcgowan J. Inquisitio validus Index Medicus: a simple method of validating MEDLINE systematic review searches. Res Synth Methods [internet] 2011;2:103–9. Available at http://onlinelibrary.wiley.com/doi/10.1002/jrsm.40, Accessed September 28, 2015.

[31] Duffy S, Misso K, Noake C, Ross J, Stirk L. Supplementary searches of PubMed to improve currency of MEDLINE and MEDLINE In-Process searches via OvidSP. UK InterTASC Information Specialists' Sub-Group Workshop, July 9. University of Exeter [Internet]; 2014. Available at http://www.systematic-reviews.com/. Accessed August 2014.

[32] Thompson JC, Quigley JM, Halfpenny NJA, Scott DA, Hawkins NS. Importance and methods of searching for E-publications ahead of print in systematic reviews. Evid Based Med 2016;21:55–9.