International Journal of Technology Assessment in Health Care

cambridge.org/thc

Method

Cite this article: Sampson M, Nama N, O'Hearn K, Murto K, Nasr A, Katz SL, Macartney G, Momoli F, McNally JD (2021). Creating enriched training sets of eligible studies for large systematic reviews: the utility of PubMed's Best Match algorithm. *International Journal of Technology Assessment in Health Care* **37**, e7, 1–6. https://doi.org/10.1017/S0266462320002159

Received: 30 March 2020
Revised: 16 November 2020
Accepted: 24 November 2020

Key words:
PubMed; Relevance ranking; Systematic reviews as topic

Author for correspondence:
Margaret Sampson, E-mail: mjs.sampson@outlook.com

© The Author(s), 2020. Published by Cambridge University Press

CAMBRIDGE UNIVERSITY PRESS

Creating enriched training sets of eligible studies for large systematic reviews: the utility of PubMed's Best Match algorithm

Margaret Sampson[1], Nassr Nama[1,2,3,4], Katharine O'Hearn[1], Kimmo Murto[1,2], Ahmed Nasr[1,2], Sherri L. Katz[1,2], Gail Macartney[5], Franco Momoli[1,2] and J. Dayre McNally[1,2]

[1]Children's Hospital of Eastern Ontario, Ottawa, Ontario, Canada; [2]Faculty of Medicine, University of Ottawa, Ottawa, Ontario, Canada; [3]Faculty of Medicine, University of British Columbia, Vancouver, British Columbia, Canada; [4]Department of Pediatrics, BC Children's Hospital, Vancouver, British Columbia, Canada and [5]Faculty of Nursing, University of Prince Edward Island, Charlottetown, Prince Edward Island, Canada

Abstract

Introduction. Solutions like crowd screening and machine learning can assist systematic reviewers with heavy screening burdens but require training sets containing a mix of eligible and ineligible studies. This study explores using PubMed's Best Match algorithm to create small training sets containing at least five relevant studies.

Methods. Six systematic reviews were examined retrospectively. MEDLINE searches were converted and run in PubMed. The ranking of included studies was studied under both Best Match and Most Recent sort conditions.

Results. Retrieval sizes for the systematic reviews ranged from 151 to 5,406 records and the numbers of relevant records ranged from 8 to 763. The median ranking of relevant records was higher in Best Match for all six reviews, when compared with Most Recent sort. Best Match placed a total of thirty relevant records in the first fifty, at least one for each systematic review. Most Recent sorting placed only ten relevant records in the first fifty. Best Match sorting outperformed Most Recent in all cases and placed five or more relevant records in the first fifty in three of six cases.

Discussion. Using a predetermined set size such as fifty may not provide enough true positives for an effective systematic review training set. However, screening PubMed records ranked by Best Match and continuing until the desired number of true positives are identified is efficient and effective.

Conclusions. The Best Match sort in PubMed improves the ranking and increases the proportion of relevant records in the first fifty records relative to sorting by recency.

Introduction

PubMed is a free web-based database of citations and abstracts of the biomedical literature, produced by the National Library of Medicine (NLM). PubMed introduced the Best Match ranking algorithm in July 2017, replacing the relevance ranking system in place since 2013. In the most recent version of PubMed, launched in November 2019, Best Match has replaced "Most Recent" as the default sort order for search results. Best Match is a method based in part on machine learning. It considers factors such as nearness of the record's match to the search query, age of the record, and its past usage, including the number of times users accessed the abstract or full text (1). We explore its utility for rapidly identifying relevant studies that could be used as true positive examples in training sets for systematic review screening.

Mechanisms such as machine learning and crowd screening can assist systematic reviewers with heavy screening burdens due to large retrievals from searching. Crowdsourcing is a phenomenon where many people (the crowd) are recruited to perform small tasks. This technique has been used to distribute the screening of records to many people, thus crowd screening (2). Machine learning is a form of artificial intelligence where, following training, an algorithm performs some of the intellectual work that would otherwise be done by humans, an approach that is increasingly used in the citation screening stage of systematic reviews (3). Both approaches require training sets containing a mix of eligible and ineligible studies. In the case of machine learning, the training can happen in the course of the review, as expert reviewers identify relevant and irrelevant studies during screening and these decisions inform subsequent relevance determinations made by the system. In the case of screening by a nonexpert crowd, a training or qualification set has customarily been created based on records screened by the investigators.

In recent tests of crowd screening systematic reviews, we used sets of between fifty and hundred records that include five to ten true positive records to train or qualify crowd members, but we had prior knowledge of the relevant studies, so we could ensure some were present in the training set (4). At this time, there is no clear standard for determining optimal training set size or the optimal proportion or the minimum number of relevant records that should be included (5). Bannach-Brown, in a recent machine learning application, used three sequential training sets resulting in a cumulative 5,749 records, with an inclusion prevalence of 13.2 percent for a broad screening task (3). Clearly, that set is larger than is practical for most reviews. Larger-than-necessary training sets represent waste and may deter potential reviewers from qualifying for participation. Sets without at least a few relevant records cannot distinguish reviewers who can discern relevant records from those who cannot. This imbalance of relevant to irrelevant records is a barrier to training, often addressed through an oversampling of relevant records (6). Determining in advance which records are relevant can entail significant effort that offsets the benefit of machine or crowd assistance (7). Thus, finding ways to enrich the sample used for training with eligible citations is desirable (6).

Relevance ranking of search results may provide a solution. Biomedical databases of the type used in systematic reviews, including MEDLINE, have traditionally presented results sorted newest to oldest. However, the ability of search engines to rank systematic review search results and place relevant records relatively high in the ranking has been previously demonstrated (8). MEDLINE and PubMed are both produced by the National Library of Medicine. PubMed is a free interface that can be used to search MEDLINE, which forms the largest proportion of records available in PubMed. (For more information: https://www.nlm.nih.gov/bsd/difference.html.) MEDLINE is one of the most commonly searched databases for biomedical systematic reviews. In many cases, it contains a very high proportion of the records for eligible studies included in systematic reviews (9).

To use Best Match ranking for systematic review training sets, MEDLINE searches for systematic reviews would first need to be translated to the PubMed syntax, and the resulting query results ranked, potentially placing relevant records (those eligible for inclusion in the systematic review) nearer the top of the search result. Harvesting the first portion of that ranked retrieval could result in a training set with many relevant and near-relevant records, relative to a date-sorted set.

Objective: This study explores using PubMed's Best Match algorithm (1) to create training sets with a sufficient number of eligible studies to be informative. The proportion of MEDLINE-indexed eligible studies ranked in the top fifty under the PubMed Best Match sort and the median rank of relevant studies in Best Match sort will be determined.

Methods

A set of six systematic reviews conducted at our institution had been both investigator-screened and subsequently used to study crowd screening feasibility (4). We assessed these six for eligibility for this study.

Additional eligibility criteria were that (i) the MEDLINE search strategy had to be available so it could be translated into PubMed syntax, (ii) the full MEDLINE download from the original search had to be available, and (iii) the list of eligible studies had to be known so they could be tested for presence and position in the Best Match ranking. Eligible studies, for the purposes of this research, are defined as studies included in the systematic review. This research does not examine the ranking of studies that passed only preliminary levels of screening. Finally, the principal investigator from the original study had to permit reuse of the data for this project.

Search Preparation and Translation

All searches had originally been created in Ovid MEDLINE and required translation to PubMed syntax. Searches were in one of two styles. The first style had one search element per line, with elements in the same concept connected with Boolean OR and concepts connected with Boolean AND, resulting in searches up to hundred lines in length. Searches in the other style were developed using the method described by Bramer et al. (10), which results in a more compact search of only a few lines. Searches of the first type were edited to combine multiple elements on one line before being submitted to Polyglot Search Translator (11) for conversion to PubMed syntax. This was done to reduce the number of line number references needed in PubMed. The compressed search was tested to ensure it generated the same number of records as the original search when run in Ovid MEDLINE. The prepared Ovid search was then pasted into the Polyglot search utility. The resulting PubMed translation was reviewed and corrected when necessary.

Supplementary File 1 provides additional detail on the search conversion process.

Search Execution

The converted search for each review was run in PubMed. The resulting PMIDs were downloaded once using the Best Match sort order and again using the Most Recent sort order. The goal was to only assess papers that were available at the time of the original search; therefore, the download set was trimmed to remove any papers with PMIDs greater than the highest PMID from the original search result. The rank position of the included studies was noted, as were the number of included studies in ranks one to fifty under both sort conditions, Best Match and Most Recent.

Analysis

The primary outcome measure is the proportion of MEDLINE-indexed eligible studies ranked in the top fifty under the PubMed Best Match sort relative to the Most Recent sort. The secondary outcome measure is the median rank of relevant studies in Best Match sort compared with the median rank of relevant studies in date sort. As well, the first quartile (lower limit of the top 25 percent) and third quartile (lower limit of the top 75 percent) were reported to give a sense of the distribution of the relevant records within the set. *Post hoc* outcome measures include the proportion of MEDLINE-indexed eligible studies ranked in the top twenty and top hundred under the PubMed Best Match sort and the rank of the fifth eligible record under both conditions, indicating how many records would need to be included in a training set to capture five eligible records.

Results

All six systematic reviews used in the Crowd Screening pilot met inclusion criteria. These reviews covered the areas of cardiology

https://doi.org/10.1017/S0266462320002159 Published online by Cambridge University Press

Table 1. Description of included systematic reviews

Review topic	Description and last search date in the original review	Total records[a]	Eligible records[b]	Search precision
Pediatric Cardiology RCTs (12)	A scoping review of all randomized controlled trials in pediatric cardiology. Last search: April 2018	7,488	835	.11
Ambulatory Adenotonsillectomy	A systematic review of preoperative screening for factors associated with postoperative critical respiratory events in children undergoing elective adenotonsillectomy. Last Search: November 2017	5,457	129	.024
Vitamin D (13)	2017 update of a previously published systematic review on high-dose supplementation of vitamin D in children. Last search: June 2018	1,589	30	.019
Concussion Education	A systematic review of studies on concussion education and outcomes for children. Last search: April 2016	513	9	.018
CPAM Symptom Development (14)	A systematic review of studies on asymptomatic antenatal diagnoses of CPAM that describe the natural history of the disease and future symptoms. Last search: July 2016	574	16	.028
PAP Adherence (15)	A systematic review of studies on the predictors of positive airway pressure adherence at home among children with sleep-disordered breathing. Last search: June 2018	277	23	.083
Total		15,899	1,042	

[a]Total number of records identified by the search strategy of the original review from all databases searched after duplicate records were removed.
[b]Eligible records as identified by the investigators (i.e., true positives).

(12), anesthesiology, endocrinology (13), patient education, general surgery (14), and respirology (15). The characteristics of the reviews are presented in Table 1.

Search Compression

Two searches were originally in compact form, four were converted prior to being translated to PubMed syntax. Retrieval size was used to verify that the search compression was equivalent to the original search. The numbers matched in all cases.

Search Translation to PubMed

Translations resulted in similar but not identical retrieval sizes (Table 2). PubMed retrievals were generally slightly larger than the Ovid MEDLINE retrievals. Across the six reviews, 11 of 964 (1 percent) MEDLINE-indexed included studies were not retrieved by the translated search.

PubMed Ranking

An evaluation of the primary outcome determined that Best Match placed three times as many relevant records in the top fifty than Most Recent (thirty vs. ten). Best Match sorting placed at least one relevant article in the top fifty for every study (Table 3). In comparison, Most Recent sorting placed a relevant article in the top fifty in five of six studies, but in four of these cases, only a single record was ranked in the top fifty. Because the top fifty represents less than 1 percent of the total retrieval for two of the systematic reviews, we counted the number of relevant records in the top hundred positions in a *post hoc* analysis (Supplementary File 2, Table S1). The Pediatric Cardiology review, with 6,896 records in total, had eight relevant records in the top fifty and thirty-eight in the top hundred under Best Match sorting, but one in the top fifty and only three in the top hundred under Most Recent sorting. The Ambulatory Adenotonsillectomy review had two and five, respectively, under Best Match sorting and one relevant record in the top fifty but five in the top hundred under Most Recent sorting. Supplementary File 2, Table S2 shows the

Table 2. Characteristics of PubMed search translations

Search	Ovid retrieval[a]	PubMed retrieval[b]	Ovid retrieval at time of original search	Number ranked after PMID trimming	MEDLINE-indexed included studies	Found by the MEDLINE search	Found by PubMed translation
Pediatric Cardiology	5,933	8,633	5,132	6,896	763	763	758
Ambulatory Adenotonsillectomy	5,658	5,894	4,893	5,139	129	128	128
Vitamin D	1,589	1,942	1,694	1,675	30	28	26
Concussion Education	676	858	412	476	8	7	7
CPAM Symptom Development	630	645	536	563	16	16	16
PAP Adherence	180	155	174	162	18	18	18

[a]As of 12–14 June 2020.
[b]As retrieval numbers sometimes differ under Best Match sorting, the number retrieved was noted under Best Match and Most Recent sorts. No differences were found for these six reviews.

https://doi.org/10.1017/S0266462320002159 Published online by Cambridge University Press

Table 3. Placement of relevant records the top fifty by best match and most recent sort orders

Search	Trimmed PubMed retrieval	Number of relevant in MEDLINE	Relevant in the top fifty					
			Best Match sort			Most Recent sort		
			N	Recall[a]	Precision[b]	*N*	Recall[a]	Precision[b]
Pediatric Cardiology	6,896	712	8	.01	.16	1	.00	.02
Ambulatory Adenotonsillectomy	5,139	129	2	.02	.04	1	.01	.02
Vitamin D	1,675	30	6	.20	.12	0	.00	.00
Concussion Education	476	8	1	.13	.02	1	.13	.02
CPAM Symptom Development	563	16	4	.25	.08	1	.06	.02
PAP Adherence	162	18	9	.50	.18	6	.33	.12

[a]For the calculation of recall, the denominator is the number of relevant records available and varies by systematic review. The numerator is the number of relevant records placed in the top fifty.
[b]For the calculation of precision, the denominator is the set size, in this case, 50. The numerator is the number of relevant studies placed in that set. Here, precision represents the saturation of relevant studies in a set size of 50. Precision would be 1.00 when the entire set is relevant. In this study, the maximum possible precision is less than 1.00 for reviews where the number of relevant studies is less than the denominator (50), that is, for all but the Ambulatory Adenotonsillectomy and Pediatric Cardiology reviews.

number of relevant records in the top twenty, corresponding to the first two pages of search results in PubMed.

Examining the placement of relevant records in the top twenty (Supplementary File 2, Table S1), neither sorting method placed records consistently in the top twenty, although Best Match showed better performance than Most Recent sorting. The rank of the fifth highest relevant records was higher in Best Match than Most Recent in all cases, with a median position of 33 in Best Match and 199 in Most Recent (Supplementary File 2, Table S3). In the worst case (Concussion Education), reviewers would have had to screen 202 records to obtain the fifth relevant record under Best Match sorting, whereas that number was 293 for the worst case (Vitamin D) under Most Recent sorting.

Considering the distribution of relevant records through the whole retrieval, the median position of relevant studies (the secondary outcome) was higher (better) under Best Match sorting in all six systematic reviews, the first quartile positions of relevant studies were higher under Best Match sorting in five of six systematic reviews, and the third quartile position of relevant studies was higher under Best Match for all six systematic reviews (Table 4).

Discussion

Best Match ranking outperformed the traditional Most Recent sort in this set of six test systematic reviews covering a range of review sizes and topics and representing different specialties. Although the total search retrievals and the number of relevant studies varied more than 10-fold across these systematic reviews, the Best Match sort placed one or more relevant studies in the first fifty retrieved in all cases. In the best case, the Pediatric Cardiology project, Best Match enriched the eligible studies in the top fifty, eightfold over Most Recent sorting. Still, in three of six systematic reviews studied here, using the first fifty PubMed records as a training set would contain fewer than five true positive records.

Any user, including those doing casual searches, can appreciate a feature such as Best Match sorting that places a higher proportion of relevant records in the top fifty. As well as placing more records in the first fifty records compared with Most Recent, Best Match ranking had a lower first quartile in five of six cases and median in all six cases, suggesting that the Best Match is a useful tool in general search, placing many relevant records nearer the top than previously available sort methods.

In the training set context, several approaches to increasing the richness of the training set are possible. One approach would be to enrich the set of fifty with any relevant articles already known to the investigators. In another approach, investigators could decide how many true positives should be included in a training set and then screen the PubMed retrieval, sorted by Best Match, until sufficient true positive records were identified. Finally, a

Table 4. Median placement of relevant records by best match and most recent sort orders

Search	Trimmed retrieval	Best Match sort		Most Recent sort	
		Median	1st and 3rd quartiles	Median	1st and 3rd quartiles
Pediatric Cardiology	6,896	1,667	694, 2,805	3,829	2,153, 5,376
Ambulatory Adenotonsillectomy	5,458	800	275, 1,653	1,565	633, 2687
Vitamin D	1,675	138	67, 386	387	312, 415
Concussion Education	476	170	119, 212	205	113, 322
CPAM Symptom Development	563	84	49, 177	320	250, 391
PAP Adherence	162	53	35, 76	71	38, 88

https://doi.org/10.1017/S0266462320002159 Published online by Cambridge University Press

larger test set could be drawn. Training sets are needed most in systematic reviews with large retrievals. Smaller retrieval sets are easily screened directly by the investigators with little need of assistance through crowd screening or machine learning, although all reviews can benefit from calibration and piloting the screening criteria (16;17).

In the case of training sets for machine learning, Miwa et al. (18) demonstrated that certainty-based screening, which aims to present positive instances for screening as early as possible, were more efficient for training than attempting to present ineligible records early or random presentation. Best Match appears to be an effective and simple way to accomplish the early presentation of relevant articles.

Ineligible records that are placed highly in the Best Match sort are unlikely to be obviously irrelevant. Distinguishing them from relevant records would require a greater level of discernment. Khabsa, in describing active learning approaches for machine learning, describes these as more informative examples, and in a machine learning approach known as active learning, these examples are prioritized for assessment by the investigators (19). However, relevance determination may require information not contained in the abstract. Edinger and Cohen examined the reasons for exclusions across 6,743 Cochrane Collaboration systematic reviews (20). Almost half of the 84,229 articles were excluded for highly specific reasons that made them difficult targets, often requiring examination of the full text to determine eligibility. Thus, Best Match may be useful for identifying relevant records efficiently. The irrelevant records may not be useful as true negatives for training purposes.

Many systematic review searches use Ovid MEDLINE instead of PubMed. Converting searches from MEDLINE to PubMed in order to use the Best Match sorting feature proved feasible.

The larger size of the PubMed retrievals compared with the Ovid MEDLINE retrievals (Table 2) is likely to be due to differences between Ovid MEDLINE and PubMed in how truncation is handled and the absence of an adjacency operator in PubMed. For example, the PAP Adherence MEDLINE search had relied very heavily on adjacency (i.e., adj*n*). Most adjacency operators were replaced by phrases, reducing the retrieval size. In other cases, replacing the adjacency operator with Boolean AND seemed more suitable, potentially inflating retrieval sizes.

Although the Polyglot translation tool was useful, it does not appear to be optimized for converting between MEDLINE and PubMed. All Polyglot translations required extensive editing (see Supplementary File 1 for notes on the process). Despite these limitations, it was helpful in creating an initial rough translation. It was also useful in refining the syntax after editing. Despite the somewhat disappointing performance of Polyglot, the conversion would not be an onerous task for an experienced librarian. Translations retrieved 94 percent of eligible studies that were indexed in MEDLINE. The time spent on search conversions was not tracked. Conversion time would depend on search complexity and the searcher's familiarity with PubMed searches. Of course, this translation step would not be needed for those using PubMed rather than MEDLINE for systematic reviews.

Limitations of This Study

All searches for the included studies were created by a single librarian and all reviews originated from a single institution, which could limit generalizability. However, reviews covered a range of specialties and a variety of interventions (surgery, patient education) and observational studies (CPAM, CPAP), and only one condition was inherently pediatric (CPAM).

Translating the search from MEDLINE to PubMed introduces some imprecision, illustrated by the few relevant studies that were retrieved by the MEDLINE search but not the PubMed search. This may be a good thing, as it demonstrated that ranking is effective even in an imprecise translation, making it useful in the real world where searchers will want to make the translation quickly. Further, the rankings are approximations. Searching retrospectively, as we have done, necessitated trimming by PubMed ID. The recency of publication and the number of times accessed are factors in the Best Match ranking. Older records may have had more access than newer records of equal importance due to their longer exposure time. This advantage would be offset by lower weighting based on recency. Thus, performance with current records might be different, although it is not clear which would have the advantage.

As well as the data underlying the rankings changing over time, changes may be introduced to the Best Match algorithms periodically. An initial set of results based on sorting as of February 2019 was updated in June 2020 (this report reflects the 2020 results throughout). Small changes in results were seen, but these did not alter the conclusions in any way.

Although this study demonstrates the utility of relevance ranking search results to more rapidly identify true positive records for training sets, the optimum training set size has not been determined, nor has an optimum ratio of relevant to irrelevant records, nor the optimum closeness of the relevant and irrelevant records. These issues require further testing in real-life situations, balancing features of the training set (length, saturation, similarity) with the correlation between reviewer performance during training and subsequent screening performance.

Conclusion

The Best Match sort option in PubMed appears effective in placing relatively more relevant articles in the first fifty records, making it useful for identifying true positive records for training sets and useful in general searching. Investigators who want to identify a certain number of true positive examples for a training set can do this efficiently by screening PubMed records ranked according to Best Match until the desired number of true positives has been identified. In many cases, this will be achieved within the first fifty records.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0266462320002159.

Acknowledgments. The authors thank Henrietta Blinder, who led the systematic review on PAP adherence.

Funding. This research received no specific funding from any agency, commercial not-for-profit sector.

Conflict of Interest. All authors have completed and signed the ICMJE Form for Disclosure of Potential Conflicts of Interest. NN reports grants from CHEO Research Institute Research Growth Award, during the conduct of the study; NN, KO, and JDM disclose that they have developed software called InsightScope that allows investigators to crowdsource systematic reviews. This software code is held under a corporation of the same name, and JDM is one of the owners.

Open Science Practices. The following are available through Open Science Framework at https://osf.io/zu3ng/. The protocol for this study, the protocol

for the search conversions, the search strategies used, results spreadsheet for each included study, a presentation to the Canadian Health Library Association 2019 annual conference, and an earlier draft of this manuscript.

References

1. **Fiorini N, Canese K, Starchenko G, Kireev E, Kim W, Miller V et al.** Best match: New relevance search for PubMed. *PLoS Biol*. 2018;**16**: e2005343.
2. **Mortensen ML, Adam GP, Trikalinos TA, Kraska T, Wallace BC**. An exploration of crowdsourcing citation screening for systematic reviews. *Res Synth Methods*. 2017;**8**:366–86.
3. **Bannach-Brown A, Przybyła P, Thomas J, Rice ASC, Ananiadou S, Liao J et al.** Machine learning algorithms for systematic review: Reducing workload in a preclinical review of animal studies and reducing human screening error. *Syst Rev*. 2019;**8**. doi:10.1186/s13643-019-0942-7
4. **Nama N, Sampson M, Barrowman N, Sandarage R, Menon K, Macartney G et al.** Crowdsourcing the citation screening process for systematic reviews: Validation study. *J Med Internet Res*. 2019;**21**:e12953.
5. **Nama N, Barrowman N, O'Hearn K, Sampson M, Zemek R, McNally JD**. Quality control for crowdsourcing citation screening: The importance of assessment number and qualification set size. *J Clin Epidemiol*. 2020;**122**:160–2.
6. **Olorisade BK, Brereton P, Andras P**. The use of bibliography enriched features for automatic citation screening. *J Biomed Inform*. 2019;**94**: 103202.
7. **Sampson M, Tetzlaff J, Urquhart C**. Precision of healthcare systematic review searches in a cross-sectional sample. *Res Synth Methods*. 2011;**2**: 119–25.
8. **Sampson M, Barrowman NJ, Moher D, Clifford TJ, Platt RW, Morrison A et al.** Can electronic search engines optimize screening of search results in systematic reviews: An empirical study. *BMC Med Res Methodol*. 2006;**6**:7.
9. **Sampson M, de Bruijn B, Urquhart C, Shojania K**. Complementary approaches to searching MEDLINE may be sufficient for updating systematic reviews. *J Clin Epidemiol*. 2016;**78**:108–15.
10. **Bramer WM, De Jonge GB, Rethlefsen ML, Mast F, Kleijnen J**. A systematic approach to searching: An efficient and complete method to develop literature searches. *J Med Libr Assoc*. 2018;**106**:531–41.
11. **Clark J, Carter M, Honeyman D, Cleo G, Auld Y, Booth D et al.** The polyglot search translator (PST): Evaluation of a tool for improving searching in systematic reviews: A randomised cross-over trial. In: The 25th Cochrane Colloquium; 2018 Sep 16–18; Cochrane: Edinburgh, UK.
12. **Ashkanase J, Nama N, Sandarage RV, Penslar J, Gupta R, Ly S et al.** Identification and evaluation of controlled trials in Pediatric Cardiology: Crowdsourced scoping review and creation of accessible searchable database. *Can J Cardiol*. 2020;**36**(11):1795–1804. doi: 10.1016/j.cjca.2020.01.028.
13. **Nama N, Menon K, Iliriani K, Pojsupap S, Sampson M, O'Hearn K et al.** A systematic review of pediatric clinical trials of high dose vitamin D. *PeerJ*. 2016;**4**:e1701. doi: 10.7717/peerj.1701.
14. **Kantor N, Wayne C, Nasr A**. Symptom development in originally asymptomatic CPAM diagnosed prenatally: A systematic review. *Pediatr Surg Int*. 2018;**34**:613–20.
15. **Blinder H, Momoli F, Bokhaut J, Bacal V, Goldberg R, Radhakrishnan D et al.** Predictors of adherence to positive airway pressure therapy in children: A systematic review and meta-analysis. *Sleep Med*. 2020;**69**:19–33.
16. **Clark J, Glasziou P, Del Mar C, Bannach-brown A, Scott AM**. How to complete a full systematic review in 2 weeks: processes, facilitators and barriers. *J Clin Epidemiol*. 2020. doi:10.1016/j.jclinepi.2020.01.008
17. **Higgins JP, Thomas J**, editors. Searching for and selecting studies. *Cochrane handbook for systematic reviews of interventions*. Version 6 2019 [cited 2020 Feb 25]. Available from: https://training.cochrane.org/handbook/current/chapter-04#section-4-6..
18. **Miwa M, Thomas J, O'Mara-Eves A, Ananiadou S**. Reducing systematic review workload through certainty-based screening. *J Biomed Inform*. 2014;**51**:242–53.
19. **Khabsa M, Elmagarmid A, Ilyas I, Hammady H, Ouzzani M**. Learning to identify relevant studies for systematic reviews using random forest and external information. *Mach Learn*. 2016;**102**:465–82.
20. **Edinger T, Cohen AM**. A large-scale analysis of the reasons given for excluding articles that are retrieved by literature search during systematic review. *AMIA Annu Symp Proc*. 2013;**2013**:379–87.

https://doi.org/10.1017/S0266462320002159 Published online by Cambridge University Press