A REVIEW ON USE OF AUTOMATION IN SYSTEMATIC REVIEWS FOR SCIENTIFIC EVIDENCE GENERATION

Short title: An Overview of Automation in Systematic Reviews

KRISHNA MOHAN[a], MD. AFTAB ALAM[b], RANJANA PATNAIK [a]

[a]Department of Clinical Research, School of Biosciences and Biomedical Engineering, Galgotias University, Greater Noida, Uttar Pradesh, India. [b]Department of Pharmacy, School of Medical and Allied Science, Galgotias University, Greater Noida, Uttar Pradesh, India.

Corresponding Author:

Krishna Mohan, Department of Clinical Research, School of Biosciences and Biomedical Engineering, Galgotias University, Plot No.2, Sector 17-A, Yamuna Expressway, Greater Noida, Gautam Budh Nagar, Uttar Pradesh (201310), India. Mobile No- +91-9958619443
E-mail- krishna.mohan.85@gmail.com

Author information:

Prof. (Dr.) AftabAlam

Department of Pharmacy, School of Medical and Allied Science, Galgotias University, Plot No.2, Sector 17-A, Yamuna Expressway, Greater Noida, GautamBudh Nagar, Uttar Pradesh (201310), India.
Mobile No- +91-9899554495
E-mail- aftab.alam@galgotiasuniversity.edu.in

Prof. (Dr.) Ranjana Patnaik

Department of Clinical Research, School of Biosciences and Biomedical Engineering, Galgotias University, Plot No.2, Sector 17-A, Yamuna Expressway, Greater Noida, GautamBudh Nagar, Uttar Pradesh (201310), India. Mobile No- +91-7565023852
E-mail- ranjana.patnaik@galgotiasuniversity.edu.in

***Krishna Mohan**: Conceptualization, Methodology, Writing- Original draft preparation*

***Krishna Mohan, Aftab Alam:** Visualization, Investigation, Writing- Reviewing and Editing.*

***Aftab Alam, Ranjana Patnaik:** Supervision, Validation*

DECLARATIONS OF INTEREST None

ABSTRACT:

Background: Systematic reviews are primarily literature reviews performed using systematic methods. A well-conducted review enables clinicians and policy-makers to stay updated in their respective fields of interest, and make informed decisions. Once fully automated, it will enable researchers to conduct systematic reviews efficiently, produce high-quality evidence, and contribute more to the field of evidence-based medicine. Mathematical models based on results from swiftly conducted systematic reviews may predict the future incidence or outbreak scenarios for diseases, which are public health problems.

Main text: This paper presents an exhaustive literature review on the common methods that can be deployed for automating sub-processes with-in a systematic review, their scope, current use, and limitations. A comprehensive search in PubMed and Google Scholar to identify articles or reviews describing use of existing automation tools within the systematic review process was performed. The main methods discussed include machine learning or artificial intelligence, text-mining, and text classification. Current gaps as well as opportunities to improve the quality of a systematic review and the overall evidence generation process are also reviewed.

Conclusions: Several technologies like Automatic Term Recognition (ATR), text-mining, text identification, as well as machine learning have already been incorporated to the general process of systematic reviews and so are common tools like Abstrackr, DistillerSR, and RobotAnalyst. The use of automatic classifiers, supervised classification algorithms, and natural language processing has been seen for search of pertinent literature. Harmonization of the existing tools is imperative for further development and quality evidence generation.

Keywords: Automation, Evidence-based Medicine, Machine Learning, Text-mining

1. INTRODUCTION

A Systematic Review (SR) provides a mechanism of reviewing data from research about a well-defined question using a systematic and stable approach. It helps obtain an objective and transparent overview of all available evidence surrounding that particular question. Only specific data matching pre-specified criteria are included so that the results are reliable and reproducible. Although systematic reviews may be conducted to examine diagnostic tests, interventions, adverse events, or economic evaluations, the basis of Evidence-Based Medicine (EBM) lies in a systematic review performed on Randomized Controlled Trials (RCT). [1-3] EBM has improved medical care significantly and incorporates a data-driven approach in the healthcare system. [4]

A vast amount of information exists in the form of journal articles, electronic repositories on clinical studies, pre-appraised evidence resources, and periodically updated websites readily available to clinicians and researchers. Especially for public health in third-world countries (where a lot of data exists in the form of published primary studies, case reports, thesis works, and reports, etc.) a review of pre-existing studies is more economical and a faster way of generating evidence than undertaking a new study to answer the research question. That being said, not all the evidence generation efforts currently undertaken are easy to conduct and always of high quality. [5] They are labor-intensive and time taking. Although developed for providing a thorough review of the available evidence that is methodological, comprehensive, transparent, and replicable, it is not always feasible for a full-time policymaker or clinician to perform a systematic review or meta-analysis for each and every research question that may arise. [6, 7] Similarly, for an important source of clinical evidence for policy and health decision-making, systematic reviews may take up to 67 weeks from registration to publication on an average. [8-10] It may take on an average between 2.5-6.5 years for a study to be included for analysis in a systematic review once published. [11]

The standard steps in conducting of a systematic review include 1) framing of the review question, 2) identification of relevant work through a robust literature-search using fixed inclusion criteria, 3) assessment of the quality of included studies using quality checklists and critical appraisal guidelines, 4) summarizing the evidence with or without combining the results in the form of a meta-analysis, and finally 5) interpretation of the findings. [12, 13]

While formulating a clear and well-defined research question, it is recommended to follow a framework like Population, Intervention, Comparison and Outcome (PICO)to define the question scope. [14] Searches for relevant studies as per inclusion and exclusion criteria should be conducted on all databases deemed relevant for the review or at a minimum, in four of these namely Embase, Google Scholar, Medline, and Web of Science. [15] The importance of including studies with adequate intervention descriptions should be given priority, as it has been reported that the intervention details are not available in up to 60% of the trial reports. For example, an assessment on non-pharmacological stroke interventions showed that materials used, intervention procedures, fidelity, tailoring, and other details about the intervention were not reported in more than 80% of the reviews. [16] As published primary studies are amalgamated into reviews, the problem of intervention details missing from such publications is compounded. Further, the quality assessment of primary studies is also an important aspect to be considered while conducting systematic reviews. One study found that threshold level quality of the primary studies for subsequent meta-analysis was assessed in only 12.9% in a studied cross section of systematic reviews. [17]

The increasing rate of publication as well as the number of published studies is making the current practice unsustainable, with a growing workload on the reviewers. The growing need to provide clinicians or policy-makers in the field of healthcare with further and further evidence has now given rise to the concept of conducting reviews on systematic reviews. [18] Such studies are termed as a reviewor an overview of reviews, meta-review, umbrella reviews, or even as a synthesis of reviews. [19] Such kind of review is a comparatively newer means of evidence synthesis, wherein each study focusses on a broader situation or condition for which multiple interventions exist and these interventions along with their results are studies or assessed. [20] Further, reviews have become much more complicated because of the complexity of interventions being studied and the amount of data that is being produced. This issue is partly addressed by engaging librarians and other information professionals to specific roles like searching, source selection, planning, and question formulation. [21] Newer concepts like living reviews are also being conducted, which involve a periodic review of the literature and updating of the systematic review at pre-defined intervals once enough data is identified. [22, 23] However, the traditional manual method of conducting systematic reviews fails to keep up with the burden of workload being generated. According to the principles of evidence-based medicine, all the available relevant evidence should be considered at the time of clinical decisions, regardless of the sources and intended resource demands, and systematic reviews were invented to serve this purpose of enabling clinicians to use this resource. Fortunately, several aspects of the systematic review process have the potential to be automated and this has been accomplished as well, as for example in search, screening of titles and citations, and data extraction, etc. Creative tasks like the development of the question or the protocol is performed during the initial phases, and the technical tasks like the search for titles, data extraction, etc. may be performed exactly as planned in the protocol, later on. Therefore, finalization of the review question requires manual intervention for the creativity, experience, and judgment part, and then the review itself may be carried out by objectively following the standard protocol as much as possible. However, the reviewers have been reluctant to adopt these modern tools. The main reasons for this have been identified as the absence of trust in replacing manual processes by automation technology and barriers in setting up of these technologies. [24]

2. METHODS

This review was performed to understand the common automated tools and techniques that can be used by a systematic reviewer to replace the manual process wherever possible and the general advantages and drawbacks of using the same. Inclusion criteria used was a review or article highlighting the use of automation for various steps within the systematic review process. We systematically searched electronic databases PubMed and Google Scholar to identify articles or reviews describing use of existing automation tools within the systematic review process, evaluation of the applicability of these tools and their advantages or drawbacks. Keywords "*systematic reviews*", *"living reviews"*, *"machine learning"*, *"text identification or mining"*, and *"artificial intelligence"* were used. Of about 1290 citations that were retrieved, 38 were selected for full-text scanning of which, 25 were reviewed (Figure 1). For all the included articles or reviews, general usability of the technologies in elements of systematic reviews of screening, data extraction, and evidence synthesis; along with advantages and disadvantages of technique used, were evaluated. Because of a large variation in the methodology used and their analysis in reviewed literature, a meta-analysis on the technologies used for the various elements was not feasible.

3. RESULTS

For any attempt to automate, the review protocol should be developed in such a way that the steps can be implemented by a machine. [25] Tasks should be reordered in such a manner that the manual activities like planning of the protocol, inclusion criteria, and development of the question etc. are moved to the start of the review, which is then followed by the automation tasks. Tools for calculating the risk of bias (Cochrane RoB tool Version 2) and those for extracting data from the selected studies (Rayyan and Revnam) are also available and recommended. The International Collaboration for the Automation of Systematic Reviews (ICASR) is a group working on the automation of systematic reviews and evidence synthesis, has said that lack of funding has led to development of several stand-alone, isolated pieces of software. A working group formed in 2019 at the ICASR Hackathon is compiling the tools uploaded in the Systematic Review Toolbox website. [26] *Jonnalagadda et al.* found that attempts were made by researchers to automate about 48% of the items used in systematic reviews. [27] They focused their review on automation of data extraction. Three broad areas on our findings along with the respective summarized reviews or articles retrieved are provided in the following sections.

3.1. Text mining for Screening

Text mining is nothing but information retrieval in order to present to users with refined data in a concise form. Text mining includes identification of relevant literature, categorization, and its summarization. Technologies that have been reviewed the most for these purposes in systematic reviews include Automatic Term Recognition (ATR), document classification, clustering, and summarization. [28] We discuss some of the text mining and screening methods in this review that are also summarized in Table 1.

Alison O'Mara-Eves et-al. have examined text mining as a potential solution to save reviewer time during the screening process in systematic reviews. [29] They postulate that reviewers aim to include all the relevant research into the systematic review in order to address the publication bias aspect. For this, they suggest a multi-layered method of searching, which includes the use of extensive Boolean searches of electronic databases (which yields three-quarters of the studies finally included), following citation trails as well as reaching out to key informants and authors at the individual level. The proposed solution for text mining is bifurcated into two methods. The first method prioritizes the list of the studies in such a way that the most relevant ones are provided at the top of the list. A classifier system is used in the second method that makes explicit include/exclude decisions. Active learning is a repetitive process in which the predictions made by the machine are habitually improved through interaction with the reviewers. An initial sample of include/exclude decisions is provided by the reviewer(s) that the machine 'learns' from. This action is repeated until a threshold is reached and the remaining decisions are generated by the machine. Both dedicated systems designed for the purpose of systematic reviews for e.g.,GAPScreener, Abstrackr, EPPI-Reviewer, and Revis, and generic software that can be readily used in a systematic review were evaluated as a part of this review, in addition to customized approaches like specialized algorithms for addressing contextual problems. Generic software applications like Pimiento and RapidMiner were found to be useful for supporting the Machine Learning (ML) aspects as well. They found that although most studies holistically suggest a workload reduction of between 30%-70% while using some form of automation, this saving was associated with about 5% of applicable studies not being picked up by the tool. This is also termed as 95% recall.

Abstackr is a software that automates the citation screening part of areview and hence expedites the process. [30] *Allison Gates et al.* have also evaluated this software for workload savings. However, the saving in the workload was associated with potentially missing of relevant records, as seen in the other studies as well. [31] *Gina Cleo et al.* and *Allison Gates et al.*aimed to examine the available automation packages for the screening part of performing systematic reviews. [32, 33] User experience was also assessed in both studies. The former assessed four software packages namely RobotAnalyst, Rayyan, SRA-Helper for EndNote, and Covidence for usability and the latter assessed Abstrackr, DistillerSR, and RobotAnalyst for automated simulation of elimination of relevant records, and for semi-automation, i.e., use of a tool to complement the work of a single reviewer in title and abstract screening.

Table 1: Methods of automation for Text mining

Sample	**Method of Automation reviewed**	**Tools**	**Advantages**	**Reference**
Literature review on 4 technologies	Automatic Term Recognition (ATR), Automated document classification, clustering, and summarization	TerMine, Automatic classifiers, Lingo3G	Quick way of identifying studies, reduces time to screen by 50%, screening prioritization by providing full text articles early on	James Thomas *et al*. [28]

44 records published after 2004, relevant to text mining in the screening phase of systematic reviews	Text mining for screening	GAPScreener, Abstrackr, EPPI-Reviewer, Revis, Pimiento and RapidMiner	Workload reduction	Alison O'Mara Eves *et al.* [29]
Systematic review of 882 citation records	Automation for screening	RobotAnalyst, Rayyan, SRA-Helper for EndNote, and Covidence	Easy to learn and use, efficient	Alison Gates *et al.* [31, 33]
3 Systematic reviews for automated and semi-automated screening simulation	Machine learning and text mining for screening title and abstract	Abstrackr, DistillerSR, and RobotAnalyst	Workload and time savings, user experience	Gina Cleo *et al.* [32]

3.2. Artificial intelligence (AI) in Screening and Data Extraction

Iain J. Marshall and Byron C. Wallace have analyzed the automation methods used to speed upvarious steps in a systematic review. They provide a practical approach to the overview of the present-day machine learning techniques that have been suggested to speed up the synthesis of evidence. [34] They further explain the two main tasks of the reviewer that can also be achieved from machine learning: text classification and data extraction, which fall in the ambit of technologies used in systematic reviews. Natural Language Processing (NLP) is at the basis of several applications that are used daily-for example translation software, search engines, and now systematic reviews. The current review suggests that while data extraction tools are still at an early phase requiring higher levels of human inputs, tools for screening are much more accessible and usable. These are summarized in Table 2.

Text classification involves computerized models that can effectively categorize papers (titles, abstracts, full-texts, etc) into prespecified categories of interest (e.g. is this an RCT or not). On the other hand, data extraction models identify the relevant text, numbers, etc. in the document and correspond this to an already specified variable (e.g., how many people were treated with a specific intervention). One of the most common examples of text classification in this area is its use in the screening of abstracts, which means deciding if a specific research paper meets the specific inclusion criteria as defined for a certain review, based on the text present in the abstracts or full text. The Machine Learning method is used as a preferred technique for the advanced methods of both text classification and data extraction. In ML, programs that specify parameterized models are written that perform certain tasks, and estimations of the parameters using large datasets are done. ML methods show resemblance to the common statistical models that are used in epidemiology. For example, the use of large datasets for estimations and logistic regression are common methods used in both areas. Examples of tools used for finding RCTs, literature exploration, screening, data extraction and bias assessment are also discussed.

Systematic Reviews are the fundamental tools for EBM. Usage of collective Bayesian classifier algorithms for example Discriminative Multinomial Naıve Bayes (DMNB) and Complement Naıve Bayes (CNB), which are statistical classifiers that can predict class-membership probabilities (such as the probability of a paper belonging to a particular pre-specified class). Performances of supervised learning models called Support Vector Machines (SVM) have been studied as machine learning text classification techniques used for the screening of pertinent literature in some reviews.[35]

A Pubmed-specific tool SWIFT-Review priority-ranks literature while conducting systematic reviews. This tool has been reported to save more than 50% of the screening effort in systematic reviews. [36] Another attempt was made by *Stan Matwin et al.* to measure the performance of another algorithm, the Factorized version of the Complement Naive Bayes (FCNB), aimed for workload reduction and a minimum 95 per cent recall. [37] Wallace and colleagues made adjustments to their algorithms using the SVM approach, for adjusting more heavily for false negatives against false positives. [38]

Table 2: Artificial Intelligence in Screening and Data Extraction

Sample	Method of Automation reviewed	Tools	Advantages	Author name
Manual screening of records in SR Toolbox (http://systematicreviewtools.com/)	Text classification and data extraction, NLP, bag of words modelling, Artificial Learning, Classification	**Search-Finding RCTs**-RCT Tagger, RobotSearch, Cochrane Register of Studies **Literature exploration-** Thalia **Screening-** Abstrackr,	Validated machine learning filters, search for concepts, automatic search retrievals, automatic extraction of data elements, automatic assessment	Iain J. Marshall *et al.* [34]

		EPPI reviewer, RobotAnalyst, SWIFT-Review, Rayyan, Colandr **Data Extraction-** ExaCT, RobotReviewer **Bias Assessment-** RobotReviewer	of Bias, etc.	
Collection of 4 systematic reviews	Machine learning for screening, text classification	Discriminative Multinomial Naïve Bayes and Complement Naïve Bayes, SVM	Performance and training time in large data sets.	Abdullah Aref and Thomas Tran [35]
21 Case Studies including 15 public datasets	Text-mining and Machine learning for screening, topic modelling	SWIFT-Review	Reduction in human screening burden and assistance in problem formulation	Brian E Howard *et al*. [36]
15 systematic drug class reviews	Machine learning	Factorized version of the Complement Naive Bayes (FCNB) classifier	Workload reduction	Stan Matwin*et al*. [37]

Dataset of four manually curated resources summarizing articles indexed in MEDLINE (3 systematic reviews and one registry summarizing published analysis)	Machine learning for updating systematic reviews	SVM Classifiers	Reduction of labor to produce and maintain systematic reviews	Bryon C Wallace *et al.* [38]

3.3. Automation of Evidence Synthesis tasks

As also shown in Table 3, *Guy Tsafnat et al.* have performed a review on the latest information systems with the potential to automate specific tasks in the process of a systematic review. [39] They identified that especially in systematic reviews of randomized controlled clinical trials, convergence of several such related projects is seen. They have described the various tasks involved in the systematic review process and how the scope of automation seen in each of these.

The current approach to extract data has 2 steps. The first uses algorithms to reduce the amount of text to be processed,while the second step associates the extracted elements with outcomes and experimental arms [40, 41]. An information-highlighting algorithm called ExaCT is also used in information extraction and screening. It classifies sentences and sometimes phrases that contain about 20 elements.

Usage of several algorithms like multi-layer perceptron (MLP), SVM, Naive Bayes (NB), J48 and Random forest (decision trees) was deployed by *Boudin et al.* to identify PICO statements in medical abstracts. [42] The PICO structure, which stands for Patient/ Population, Intervention, Comparison, Outcome is the basis of structuring a systematic review question. As several variables exist in the libraries for these 4 elements, and no search engines have detected and indexed these elements till date, retrieval using any automation tool is seldom difficult. In this study, each document was segmented into plain sentences. A feature set was created by which each sentence was converted to a feature vector. Each of these vectors was then introduced to the classifiers corresponding to each element, allowing the system to categorize the corresponding sentence. A position classifier was also included at baseline, as the PICO elements are found at specific sections of the artefact, in a specific order, where usually the population comes first and outcome comes last. Results did not show superiority of any one classifier. This experiment found that identification of PICO items was especially challenging. A high accuracy could be achieved for the detection of the P element.

Xu et al. evaluated a method to structure RCT abstracts and automatically extract patient demographics by identifying sentences containing subject demographics using a text classification method coupled with a Hidden Markov Model. He postulated that subject demographics would most likely to be present in the Methods section. An F1 measure, which is a combined measure of precision and recall was used to test the performance. [43, 44]

Christopher Norman et al. evaluated the possibility of extracting diagnostic test accuracy results from published systematic reviews. [45] They created their dataset from published review articles containing the elements in free text, HTML, data tables as well as PNG images. They not only extracted these elements, but also linked these elements together. Automated extraction methods complemented with manual extraction, verification and post-editing were used. The HTML contents were processed using the LXML Python package. Optical Character Recognition (OCR) was used for the diagnostic test results only presented in images. Such data was then extracted using Tesseract.

Systematic reviews involve important yet repetitive work. The Cochrane Collaboration, makes use of software like Review Manager or RevMan[46] and its add-on program RevManHAL. The latter is utilized in the write-up phase of systematic reviews, and abstracts various sections of the review from RevMan-generated reviews in various languages. [47]

A project on similar lines, although not related to systematic reviews and known as the Trial Bank project led by *Berry de Bruijn et al*. Inclusion of formalized trial information extracted from published randomized clinical trials into a knowledge base was developed to improve access to trial findings. A text classifier was used to extract a total of twenty-three important trial information items like inclusion and exclusion, sample size, intervention, and outcome names. Such a repository would provide decision makers and systematic reviewer with specific information published in RCT articles without having to perform any further reviews.[48]

A tool called Systematic EvidEnce Disseminator (SEED) has been developed to auto-generate a Wikipedia-compatible table of summation directly from Cochrane's RevMan files, along with the accompanying reference. [49]

Table 3: Automation in evidence-synthesis

Sample	**Method of Automation reviewed**	**Tools**	**Advantages**	**Author name**
Survey of systems that automate the processes of systematic review	Automation, data extraction	**Search-** Quick Clinical, Sherlock, Metta **Snowballing-** ParsCit **Screen-Titles and Abstracts-** Abstrackr **Extract data-** ExaCT, WebPlotDigitizer **Meta-analysis-**	Federated Meta-search engines, reference string extraction from papers, Machine learning tools, automatic re-	Guy Tsafnat*et al*. [39]

		Meta-Analyst **Write-up**- PRISMA Flow Diagram Generator Revman-Hal,	digitization of data, meta-analysis, write-ups and diagrams, etc.	
Data set of each PICO element developed for Training as well as Testing	Data Extraction, machine learning	Supervised classification algorithms, (SVM)	High accuracy for the P element achieved	Boudin*et al*. [42]
Reports of Randomized Clinical trials	Text classification and extraction	Natural language processing, text classificaition algorithms, Hidden Markov Modelling	High accuracy extraction	Xu *et al*. [44, 45]
63 Diagnostic test accuracy systematic reviews	Automated data extraction and synthesis	LXML Python package.4, optical character recognition (OCR), Tesseract	Data extraction with low error rate for diagnostic test accuracy results	Christopher Norman *et al*. [45]
RevMan-generated reviews	Automatic text generation	RevmanHal	Auto-generation of the abstract, results and discussion sections in multiple languages	Mercedes Torres *et al*. [47]

88 RCT reports	Machine learning for data extraction	Text classifier	Promising source to identify key elements published in RCT articles.	Berry de Bruijn*et al*. [48]
Cochrane's RevMan files	Data relocation	Systematic EvidEnce Disseminator (SEED)	Summary of Findings can be replicated to other information sources like Wikipedia	Lena Schmidt *et al*. [49]

4. CONCLUSION

This review describes the common methods and tools that can be used for automating the sub-processes involved in systematic reviews. Access to summarized data and information on current practices can meet the practitioner's needs and improve decision making. To address the concern of trust among the reviewers for using automation technologies, the know-how should be promoted in the form of training and secondly, the quality of the output produced by automation tools needs to be assessed for systematic reviews to be generally accepted. Reviewers would assess the scope of using each tool for their study and would prepare the protocol accordingly. Although the next logical step is a fully automated system that will deliver the best evidence in a timely fashion, this idea is especially challenging as various stand-alone systems addressing the sub-processes need to be linked with each other. A fully automated systematic review would free the systematic reviewer to concentrate on the non-tedious, automatable tasks to the creative tasks like creating the protocol, providing complex interpretations of the generated data, and ensuring the quality of the overall execution.

Use of technology as a second reviewer or "semi-automation" for citation and abstract screening have already been found effective. Innovations in the development of applications that automatically extract data from websites for the purpose of systematic reviews and meta-analysis, applications that can assist in extraction of data from graphs, and deployment of Convolutional Deep Neural Network (CDNN) in detecting patterns in medical images for the diagnosis and management of diseases have also been reported.

Computerized systems called Clinical Decision Support Systems (CDSS) designed to provide doctors and other healthcare professionals with assistance in clinical decision making tasks that are 80-90 % accurate can be alternatively used for clinical practice. Disease-specific databases or registries could also be a way of providing required information as the dataset would greatly reduce for the single disease under study and would be easy to develop and maintain. Development of novel tools should also be looked into for further advancement in the field of evidence generation. Further options for automating the evidence generation process still need to be explored.

ACKNOWLEDGEMENT

The authors are highly thankful to the Department of Clinical Research, School of Biosciences and Biomedical Engineering and Department of Pharmacy, School of Medical and Allied Science, Galgotias University to provide library facilities for literature surveys.

CONFLICT OF INTEREST

No conflicts of interest have been reported by authors.

FUNDING STATEMENT

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

ADDITIONAL INFORMATION

No additional information is available for this paper.

REFERENCES

[1] Centre for Reviews and Dissemination, University of York: York. 2008 Systematic reviews: CRD's guidance for undertaking reviews in health care (PDF). ISBN 978-1-900640-47-3.

[2] Petticrew M, Roberts H (2006). Systematic reviews in the social sciences (PDF). Wiley Blackwell. ISBN 978-1-4051-2110-1.

[3] Skelly, A. C., & Chapman, J. (2011). Evidence-based medicine (EBM): origins and modern application to spine care. Evidence-based spine-care journal, 2(1), 11–16. https://doi.org/10.1055/s-0030-1267081

[4] John T. Langell (2019) Evidence-based medicine: A data-driven approach to lean healthcare operations, International Journal of Healthcare Management, DOI: 10.1080/20479700.2019.1641650

[5] Coiera E. (1998). Information epidemics, economics, and immunity on the internet. We still know so little about the effect of information on public health. *BMJ (Clinical research ed.)*, *317*(7171), 1469–1470. https://doi.org/10.1136/bmj.317.7171.1469

[6] Centre for Evidence Based Medicine. "What is EBM?" 2009-11-20. Archived from the original on 2011-04-06.

[7] D. Gough, S. Oliver and J. Thomas. An introduction to systematic reviews. London: Sage, 2012.

[8] Garritty, C., Tsertsvadze, A., Tricco, A. C., Sampson, M., & Moher, D. (2010). Updating systematic reviews: an international survey. *PloS one*, *5*(4), e9914. https://doi.org/10.1371/journal.pone.0009914

[9] Gough , David &Elbourne, Diana. (2002). Systematic Research Synthesis to Inform Policy, Practice and Democratic Debate. Social Policy and Society. 1. 225 - 236. 10.1017/S147474640200307X.

[10] Borah R, Brown AW, Capers PL, Kaiser KA. (2017) Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry. *BMJ Open*, 7(2):e012545. Published 2017 Feb 27. doi:10.1136/bmjopen-2016-012545

[11] Elliott JH, Turner T, Clavisi O, Thomas J, Higgins JPT, Mavergames C, et al. (2014) Living Systematic Reviews: An Emerging Opportunity to Narrow the Evidence-Practice Gap. PLoS Med 11(2): e1001603. https://doi.org/10.1371/journal.pmed.1001603

[12] Khan, K. S., Kunz, R., Kleijnen, J., & Antes, G. (2003). Five steps to conducting a systematic review. *Journal of the Royal Society of Medicine*, *96*(3), 118– 121. https://doi.org/10.1258/jrsm.96.3.118

[13] Higgins J, Green S. Cochrane handbook for systematic reviews of interventions version 5.1. 0 [updated March 2011]. The Cochrane

[14] Eriksen, M., &Frandsen, T. (2018). The impact of patient, intervention, comparison, outcome (PICO) as a search strategy tool on literature search quality: a systematic review. *Journal of the Medical Library Association, 106*(4), 420–431. doi: https://doi.org/10.5195/jmla.2018.345

[15] Bramer, W. M., Rethlefsen, M. L., Kleijnen, J., & Franco, O. H. (2017). Optimal database combinations for literature searches in systematic reviews: a

prospective exploratory study. *Systematic reviews*, *6*(1), 245. https://doi.org/10.1186/s13643-017-0644-y

[16] Hoffmann, T. C., Walker, M. F., Langhorne, P., Eames, S., Thomas, E., &Glasziou, P. (2015). What's in a name? The challenge of describing interventions in systematic reviews: analysis of a random sample of reviews of non-pharmacological stroke interventions. *BMJ open*, *5*(11), e009051. https://doi.org/10.1136/bmjopen-2015-009051

[17] Seehra, J., Pandis, N., Koletsi, D., & Fleming, P. S. (2016). Use of quality assessment tools in systematic reviews was varied and inconsistent. Journal of clinical epidemiology, 69, 179–84.e5.https://doi.org/10.1016/j.jclinepi.2015.06.023

[18] Hasanpoor, E., Hallajzadeh, J., Siraneh, Y., Hasanzadeh, E., &Haghgoshayie, E. (2019). Using the Methodology of Systematic Review of Reviews for Evidence-Based Medicine. Ethiopian journal of health sciences, 29(6), 775–778. https://doi.org/10.4314/ejhs.v29i6.15

[19] Aromataris, E., Fernandez, R., Godfrey, C. M., Holly, C., Khalil, H., &Tungpunkom, P. (2015). Summarizing systematic reviews: methodological development, conduct and reporting of an umbrella review approach. International journal of evidence-based healthcare, 13(3), 132–140.https://doi.org/10.1097/XEB.0000000000000055

[20] Grant, M. J., & Booth, A. (2009). A typology of reviews: an analysis of 14 review types and associated methodologies. *Health information and libraries journal*, *26*(2), 91–108. https://doi.org/10.1111/j.1471-1842.2009.00848.x

[21] Spencer, A. J., &Eldredge, J. D. (2018). Roles for librarians in systematic reviews: a scoping review. *Journal of the Medical Library Association: JMLA*, *106*(1), 46–56. https://doi.org/10.5195/jmla.2018.82

[22] Brooker A, Synnot A, McDonald S, Elliott J, Turner T, et al.: Guidance for the production and publication of Cochrane living systematic reviews: Cochrane reviews in living mode. 2019

[23] Elliott, J. H., Synnot, A., Turner, T., Simmonds, M., Akl, E. A., McDonald, S., Salanti, G., Meerpohl, J., MacLehose, H., Hilton, J., Tovey, D., Shemilt, I., Thomas, J., & Living Systematic Review Network (2017). Living systematic review: 1. Introduction-the why, what, when, and how. *Journal of clinical epidemiology*, *91*, 23–30. https://doi.org/10.1016/j.jclinepi.2017.08.010

[24] O'Connor, A. M., Tsafnat, G., Thomas, J., Glasziou, P., Gilbert, S. B., & Hutton, B. (2019). A question of trust: can we build an evidence base to gain trust in

systematic review automation technologies?. *Systematic reviews*, *8*(1), 143. https://doi.org/10.1186/s13643-019-1062-0

[25] Tsafnat, G., Dunn, A., Glasziou, P., &Coiera, E. (2013). The automation of systematic reviews. *BMJ (Clinical research ed.)*, *346*, f139. https://doi.org/10.1136/bmj.f139

[26] Marshall C: The systematic review toolbox. 2019

[27] Jonnalagadda, S. R., Goyal, P., & Huffman, M. D. (2015). Automating data extraction in systematic reviews: a systematic review. *Systematic reviews*, *4*, 78. https://doi.org/10.1186/s13643-015-0066-7

[28] Thomas, J., McNaught, J., &Ananiadou, S. (2011). Applications of text mining within systematic reviews. *Research synthesis methods*, *2*(1), 1–14. https://doi.org/10.1002/jrsm.27

[29] O'Mara-Eves, A., Thomas, J., McNaught, J., Miwa, M., &Ananiadou, S. (2015). Using text mining for study identification in systematic reviews: a systematic review of current approaches. *Systematic reviews*, *4*(1), 5. https://doi.org/10.1186/2046-4053-4-5

[30] Byron C. Wallace, Kevin Small, Carla E. Brodley, Joseph Lau, and Thomas A. Trikalinos. 2012. Deploying an interactive machine learning system in an evidence-based practice center: abstrackr. In Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium (IHI '12). Association for Computing Machinery, New York, NY, USA, 819–824. DOI: https://doi.org/10.1145/2110363.2110464

[31] Gates, A., Johnson, C., &Hartling, L. (2018). Technology-assisted title and abstract screening for systematic reviews: a retrospective evaluation of the Abstrackr machine learning tool. *Systematic reviews*, *7*(1), 45. https://doi.org/10.1186/s13643-018-0707-8

[32] Cleo, G., Scott, A. M., Islam, F., Julien, B., &Beller, E. (2019). Usability and acceptability of four systematic review automation software packages: a mixed method design. *Systematic reviews*, *8*(1), 145. https://doi.org/10.1186/s13643-019-1069-6

[33] Gates, A., Guitard, S., Pillay, J., Elliott, S. A., Dyson, M. P., Newton, A. S., &Hartling, L. (2019). Performance and usability of machine learning for screening in systematic reviews: a comparative evaluation of three tools. *Systematic reviews*, *8*(1), 278. https://doi.org/10.1186/s13643-019-1222-2

[34] Marshall, I. J., & Wallace, B. C. (2019). Toward systematic review automation: a practical guide to using machine learning tools in research synthesis. *Systematic reviews*, *8*(1), 163. https://doi.org/10.1186/s13643-019-1074-9

[35] Aref A., Tran T. (2014) Using Ensemble of Bayesian Classifying Algorithms for Medical Systematic Reviews. In: Sokolova M., van Beek P. (eds) Advances in Artificial Intelligence. Canadian AI 2014. Lecture Notes in Computer Science, vol 8436. Springer, Cham. https://doi.org/10.1007/978-3-319-06483-3_23

[36] Howard, B. E., Phillips, J., Miller, K., Tandon, A., Mav, D., Shah, M. R., Holmgren, S., Pelch, K. E., Walker, V., Rooney, A. A., Macleod, M., Shah, R. R., & Thayer, K. (2016). SWIFT-Review: a text-mining workbench for systematic review. *Systematic reviews*, *5*, 87. https://doi.org/10.1186/s13643-016-0263-z

[37] Matwin, S., Kouznetsov, A., Inkpen, D., Frunza, O., &O'Blenis, P. (2010). A new algorithm for reducing the workload of experts in performing systematic reviews. Journal of the American Medical Informatics Association: JAMIA, 17(4), 446–453. https://doi.org/10.1136/jamia.2010.004325

[38] Wallace, B. C., Small, K., Brodley, C. E., Lau, J., Schmid, C. H., Bertram, L., Lill, C. M., Cohen, J. T., &Trikalinos, T. A. (2012). Toward modernizing the systematic review pipeline in genetics: efficient updating via data mining. *Genetics in medicine: official journal of the American College of Medical Genetics*, *14*(7), 663–669. https://doi.org/10.1038/gim.2012.7

[39] Tsafnat, G., Glasziou, P., Choong, M. K., Dunn, A., Galgani, F., &Coiera, E. (2014). Systematic review automation technologies. *Systematic reviews*, *3*, 74. https://doi.org/10.1186/2046-4053-3-74

[40] Kiritchenko, S., de Bruijn, B., Carini, S., Martin, J., & Sim, I. (2010). ExaCT: automatic extraction of clinical trial characteristics from journal publications. *BMC medical informatics and decision making*, *10*, 56. https://doi.org/10.1186/1472-6947-10-56

[41] Hsu, W., Speier, W., &Taira, R. K. (2012). Automated extraction of reported statistical analyses: towards a logical representation of clinical trial literature. *AMIA ... Annual Symposium proceedings. AMIA Symposium*, *2012*, 350–359.

[42] Boudin, F., Nie, J. Y., Bartlett, J. C., Grad, R., Pluye, P., & Dawes, M. (2010). Combining classifiers for robust PICO element detection. *BMC medical informatics and decision making*, *10*, 29. https://doi.org/10.1186/1472-6947-10-29

[43] Xu, R., Supekar, K., Huang, Y., Das, A., & Garber, A. (2006). Combining text classification and Hidden Markov Modeling techniques for categorizing sentences in

randomized clinical trial abstracts. *AMIA ... Annual Symposium proceedings. AMIA Symposium*, *2006*, 824–828.

[44] Xu, R., Garten, Y., Supekar, K. S., Das, A. K., Altman, R. B., & Garber, A. M. (2007). Extracting subject demographic information from abstracts of randomized clinical trial reports. *Studies in health technology and informatics*, *129*(Pt 1), 550–554.

[45] Norman, C., Leeflang, M., &Névéol, A. (2018). Data Extraction and Synthesis in Systematic Reviews of Diagnostic Test Accuracy: A Corpus for Automating and Evaluating the Process. *AMIA ... Annual Symposium proceedings. AMIA Symposium*, *2018*, 817–826.

[46] Cochrane Informatics and Knowledge Management Department. RevMan 5.3 [Internet]. Available from: http://tech.cochrane.org/Revman.

[47] Torres Torres, M., & Adams, C. E. (2017). RevManHAL: towards automatic text generation in systematic reviews. *Systematic reviews*, *6*(1), 27. https://doi.org/10.1186/s13643-017-0421-y

[48] deBruijn, B., Carini, S., Kiritchenko, S., Martin, J., & Sim, I. (2008). Automated information extraction of key trial design elements from clinical trial publications. *AMIA ... Annual Symposium proceedings. AMIA Symposium*, *2008*, 141–145.

[49] Schmidt, L., Friedel, J., & Adams, C. E. (2017). SEED: a tool for disseminating systematic review data into Wikipedia. Systematic reviews, 6(1), 206. https://doi.org/10.1186/s13643-017-0607-3

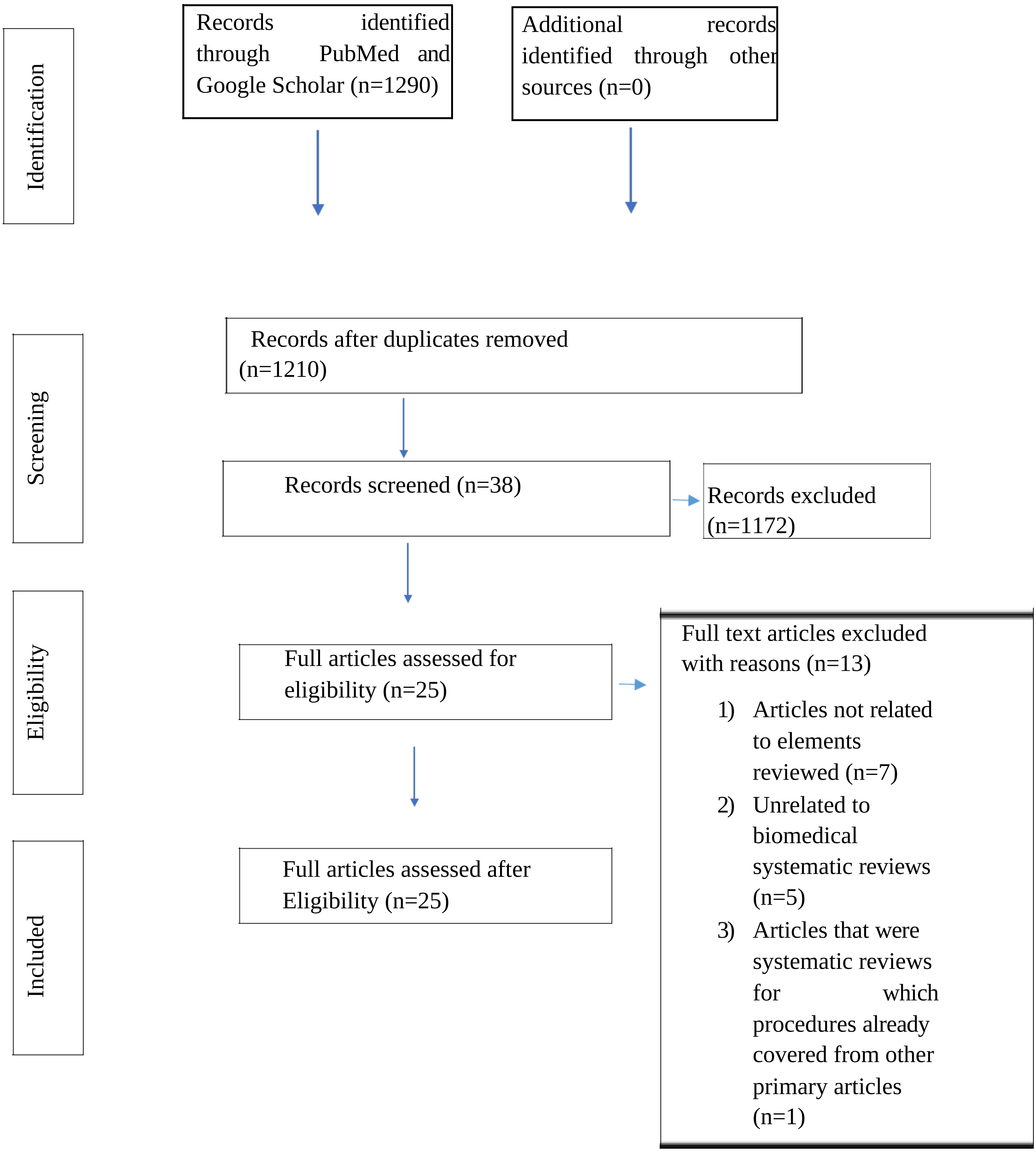

Figure 1: Process of screening of the articles for literature review.