BMC Medical Research Methodology

RESEARCH

Open Access

An exploration of available methods and tools to improve the efficiency of systematic review production: a scoping review

Lisa Affengruber[1,2*], Miriam M. van der Maten[3,11], Isa Spiero[4], Barbara Nussbaumer-Streit[1], Mersiha Mahmić-Kaknjo[5], Moriah E. Ellen[6,7,8], Käthe Goossen[9], Lucia Kantorova[10], Lotty Hooft[11], Nicoletta Riva[12], Georgios Poulentzas[13], Panagiotis Nikolaos Lalagkas[13], Anabela G. Silva[14], Michele Sassano[15,16], Raluca Sfetcu[17,18], María E. Marqués[19], Tereza Friessova[20], Eduard Baladia[19], Angelo Maria Pezzullo[15], Patricia Martinez[19,21], Gerald Gartlehner[1,22] and René Spijker[11,23]

Abstract

Background Systematic reviews (SRs) are time-consuming and labor-intensive to perform. With the growing number of scientific publications, the SR development process becomes even more laborious. This is problematic because timely SR evidence is essential for decision-making in evidence-based healthcare and policymaking. Numerous methods and tools that accelerate SR development have recently emerged. To date, no scoping review has been conducted to provide a comprehensive summary of methods and ready-to-use tools to improve efficiency in SR production.

Objective To present an overview of primary studies that evaluated the use of ready-to-use applications of tools or review methods to improve efficiency in the review process.

Methods We conducted a scoping review. An information specialist performed a systematic literature search in four databases, supplemented with citation-based and grey literature searching. We included studies reporting the performance of methods and ready-to-use tools for improving efficiency when producing or updating a SR in the health field. We performed dual, independent title and abstract screening, full-text selection, and data extraction. The results were analyzed descriptively and presented narratively.

Results We included 103 studies: 51 studies reported on methods, 54 studies on tools, and 2 studies reported on both methods and tools to make SR production more efficient. A total of 72 studies evaluated the validity ($n=69$) or usability ($n=3$) of one method ($n=33$) or tool ($n=39$), and 31 studies performed comparative analyses of different methods ($n=15$) or tools ($n=16$). 20 studies conducted prospective evaluations in real-time workflows. Most studies evaluated methods or tools that aimed at screening titles and abstracts ($n=42$) and literature searching ($n=24$), while for other steps of the SR process, only a few studies were found. Regarding the outcomes included, most studies reported on validity outcomes ($n=84$), while outcomes such as impact on results ($n=23$), time-saving ($n=24$), usability ($n=13$), and cost-saving ($n=3$) were less often evaluated.

*Correspondence:
Lisa Affengruber
lisa.affengruber@donau-uni.ac.at
Full list of author information is available at the end of the article

Conclusion For title and abstract screening and literature searching, various evaluated methods and tools are available that aim at improving the efficiency of SR production. However, only few studies have addressed the influence of these methods and tools in real-world workflows. Few studies exist that evaluate methods or tools supporting the remaining tasks. Additionally, while validity outcomes are frequently reported, there is a lack of evaluation regarding other outcomes.

Keywords Rapid review, Systematic review, Evidence synthesis, Scoping review, Method, Automation tools

Introduction

Systematic reviews (SRs) provide the most valid way of synthesizing evidence as they follow a structured, rigorous, and transparent research process. Because of their thoroughness, SRs have a long history in informing health policy decision-making, clinical guidelines, and primary research [1]. However, the standard method employed in high-quality SRs involves many steps that are predominantly conducted manually, resulting in a laborious and time-intensive process lasting an average of fifteen months [2]. With the exponential growth of scientific literature, the challenge of SR development is exacerbated. Especially, this is the case in contexts where timely evidence is imperative and decision-makers need urgent answers, as demonstrated during the coronavirus pandemic [3]. Additionally, researchers' aspirations to undertake SRs prior to initiating new primary studies is hindered by the complex and resource-intensive nature of the SR development process [4].

In response to these challenges, a surge of interest in methods to accelerate SR development occurred in recent years, leading to the emergence of "rapid reviews" (RRs). Methods used in RR development include searching a limited number of bibliographic databases, single-reviewer literature screening, or abbreviated quality assessment [5]. However, depending on various factors, the trade-off between the time saved and the potential reduction in quality and comprehensiveness is a critical issue that must be carefully weighed and discussed with stakeholders. Concurrently, efforts are underway to leverage technological innovations to expedite the research process, involving machine learning, natural language processing, active learning, or text mining to mimic human activities in SR tasks [6]. Supportive tools can offer varying levels of automation and decision-making, ranging from basic file management to fully automated decision-making, as outlined by O'Connor et al. [7]. These levels include semiautomated tools for workflow prioritization to fully automated decision-making processes [7]. Some tools offer researchers ready-to-use applications, while other algorithms are not yet developed into user-friendly tools [8]. Table 1 provides an overview of commonly used terms in this scoping review and their definitions.

While these methods and tools hold promise for enhancing the efficiency of SR production, their widespread adoption faces challenges, including limited awareness among review teams, concerns about validity, and usability issues [12]. Addressing these barriers requires evaluations to determine the validity and usability of various methods and tools across different stages of the review process [8, 13].

To bridge this gap, we conducted a scoping review to comprehensively map the landscape of methods and tools aimed at improving the efficiency of SR development, assessing their validity, resource utilization (workload/time/costs), and impact on results, as well as exploring usability for all steps of the review process. This review complements another scoping review that identified the most resource-intensive areas when conducting or updating a SR [14]. We mapped the efficiency outcomes of each method and tool against the steps of the SR process. Specifically, our scoping review aimed to answer the following key questions (KQs):

1. Which methods and tools are used to improve the efficiency of SR production?
2. How efficient are these methods or tools regarding validity, resource use, and impact on results?
3. How was the user experience when using these methods and tools?

Methods

We conducted this review as part of working group 3 of EVBRES (EVidence-Bases RESearch) COST Action CA17117 (www.ebvres.eu). We published the protocol for this scoping review on June 18, 2020 (Open Science Framework: https://osf.io/9423z).

Study design

We conducted a scoping review following the guidance of Arksey and O'Malley [15], Levac et al. [16], and Peters et al. [5]. Within EVBRES, we adopted the definition of a scoping review as "a form of knowledge synthesis that addresses an exploratory research question aimed at mapping key concepts, types of evidence, and gaps in research related to a defined area or field by systematically searching, selecting, and synthesising existing knowledge" [17]. We report our review in accordance with the PRISMA Extension for Scoping Reviews (PRISMA-ScR) [18].

Table 1 Commonly used terms

Term	Definition
Text mining	Text mining refers to the extraction of information from unstructured text data. The process of text mining involves examining large amounts of text data to discover patterns, relationships, and insightsthat aid informed decision-making. Techniques employed in text mining ecompass natural language processing, machine learning, active learning, and statistical analysis [7, 9]
Machine learning	Machine learning is a type of artificial intelligence that enables computer systems to enhance their performance on specific tasks through data-driven learning, rather than explicit programming. This approach involves creating algorithms capable of discerning patterns and relationships within datasets. These algorithms then apply the acquired knowledge to make predictions or decisions when presented with new, unseen data [7]
Active learning	Active learning is a machine learning approach where the algorithm strategically chooses which data to learn from, instead of relying on a prelabeled dataset. In this method, the algorithm iteratively identifies the most informative data points that require labeling by a human expert. These labeled data points are then used to refine and update the model. This cycle is repeated until the model reaches a satisfactory level of accuracy or until further labeling does not enhance the model's performance [7]
Fully automated	Full automated tools make decisions idenpendently without the need for human intervention. For instance, such a tool might autonomously evaluate the risk of bias in a randomized controlled trial, completing the assessment process without requiring input or oversight from a human reviewer [7, 10]
Semiautomated	Semiautomated tools provide decision-making recommendationsyet they require human verification to finalize decisions. For example, such tools indicate whether a record is likely suitable for inclusion, but human judgment is needed to make the final in- or exclusion decision. This approach is particularly beneficial in tasks like prioritizing relevant abstracts and detecting duplicates [7, 10]
Supportive tools without automation	Tools supporting the file management process, such as citation databases, reference management tools, and SR management tools, that operate without incorporating any form of automation [7]
Supportive tools	Tools employing text mining, machine and active learning with fully or semiautomated approaches, and non-automated tools that support RR production
Ready-to-use tools	Tools that offer researchers ready-to-use, fully developed applications, while other algorithms are not yet developed into user-friendly tools
Crowdsourcing	Utilizing a large group of human volunteers to perform a task online
Workload saving	Workload saving refers to the reduction in the amount of work required to complete a task or process through abbreviated methods or tools in comparison to SR methodology. In this context, workload saving refers to reducing the number of references or documents that must be screened, reviewed, or analyzed during the review process
Time-saving	Time-saving in this context refers to the reduction in the amount of time measured in hours or minutes required to complete the review process through abbreviated methods or supportive tools
Cost-saving	Cost-saving in this context refers to the reduction in financial expenditures associated with conducting the review process. It mainly encompasses personnel costs required to complete the review
Impact on results	Impact on results in this context refers to changes in statistical significance, effect estimates, or conclusions as a result of abbreviated methods or supporting tools during the review process
Validity outcomes	Validity outcomes in this context describe the degree to which the method or tool is aligned with the SR methodology. Validity is measured as the tool's precision, sensitivity, specificity, accuracy, errors made, etc., compared to SR methods done by humans
Usability	Usability refers to the extent to which end users can use the method/tool in a specific context to achieve specific goals effectively, efficiently, and satisfactorily, and how the end users experienced it [11]. In the context of SR tools, usability is often reported as evaluation results (e.g., usability scores) of the method or tool or end-user experiences including joy of use, interface design, etc

Abbreviations: RR Rapid review, *SR* Systematic review

Information sources and search

The search for this scoping review followed an iterative three-step process recommended by the Joanna Briggs Institute [19]:

1) First, an information specialist (RS) conducted a preliminary limited, focused search in Scopus in March 2020. We screened the search results and analyzed relevant studies to discover additional relevant keywords and sources.
2) Second, based on identified search terms from the included studies, the information specialist performed a comprehensive search (November 2021) in MEDLINE and Embase, both via Ovid. The comprehensive MEDLINE strategy was reviewed by another information specialist (IK) in accordance with the Peer Review of Electronic Search Strategies (PRESS) guideline [20].
3) Third, we checked the reference lists of the identified studies and background articles, conducted grey literature searches (e.g., organizations that produce SRs and RRs), and contacted experts in the field. In addition, to identify grey literature, we searched for conference proceedings covered in Embase and checked if the associated full text was available. We also searched the systematicreviewtools.com website for additional evaluation studies using the search strategy employed by colleagues working in this field [21].

We limited the database searches to articles on methodological adaptations published since 1997, as this was the first year of mention in the published literature of methods to make the review process more efficient [22]. For tools, we limited the search to articles published since 2005, as this was the first year of mention of a text mining model in the published literature, according to Jonnalagadda and Petitti [23]. The search strategies are provided in Appendix 1. Our search was updated on December 14, 2023 to include evidence published since our initial search.

Eligibility criteria

The eligibility criteria are outlined in Table 2. Our focus was on incorporating primary studies that assessed the efficacy of automated, ready-to-use applications of tools, or RR methods within the SR process. Specifically, we sought tools that demand no programming expertise, relying instead on user-friendly interfaces devoid of complex codes, syntaxes, or algorithms. We were interested in studies assessing their use within one or more of the fifteen steps of the SR process as defined by Tsafnat et al. (2014), further supplemented by the steps "critical appraisal," "grading the certainty of evidence," and "administration/project management" [24] [14] (Fig. 1).

Study selection

We piloted the abstract screening with 50 records and the full-text screening with five records. Following the piloting, the results were discussed with all reviewers, and the screening guidance was updated to include clarifications wherever necessary. The review team used Covidence (www.covidence.org) to dually screen titles/abstracts and full texts. We resolved conflicts throughout the screening process through re-examination of the study and subsequent discussion and, if necessary, by consulting a third reviewer.

Data charting

We developed a data extraction form and pilot-tested it before implementation using Google Forms. The data abstraction was done by one author and checked by a second author to ensure consistency and correctness in the extracted data. A third author made final decisions in cases of discrepancies. We extracted relevant study characteristics and outcomes per review step.

Data mapping

We mapped the identified methods and tools by each review step and summarized the outcomes of individual studies. As the objective of this scoping review was to descriptively map efficiency outcomes and the usability of methods and tools against SR production steps, we did not apply a formal certainty of evidence or risk of bias (RoB) assessment. Additionally, we used data mapping to identify research gaps.

Table 2 Eligibility criteria for study inclusion in the scoping review

	Inclusion criteria	Exclusion criteria
Topic	Studies reporting the performance of methods/ready-to-use tools with regard to improving efficiency when producing or updating a SR in the health field	Studies not reporting on improving efficiency Studies assessing algorithms, classifiers, or models that are not usable as tools
Concept	Addressing improvement in efficiency within one or more steps of a SR or RR (as depicted in Fig. 1) of health intervention or diagnostic or prognostic studies	Studies assessing other steps such as the dissemination of results
Outcomes	Workload saved Time saved Personnel effort saved Costs saved Impact on results, conclusions Usability Validity outcomes: for example, recall (sensitivity), precision, accuracy, specificity, false positives, false negatives, reliability, proportion of relevant studies missed, prediction performance, ordering performance	Any other outcomes
Study design/ publication type	Evaluation study of method/ready-to-use tool (e.g., method studies, RCTs, nRCTs, cohort studies,..)	Development studies of ready-to-use tools Validation studies of of ready-to-use tools Evaluation studies of tools that are no longer available
Full text availability	Full text retrievable via our libraries	Full text not retrievable via our libraries
Timing	Studies assessing methods: 1997 Studies assessing tools: 2005	
Language	All languages	

Abbreviations: nRCT Nonrandomized-controlled trial, *RCT* Randomized controlled trial, *RR* Rapid review, *SR* Systematic review

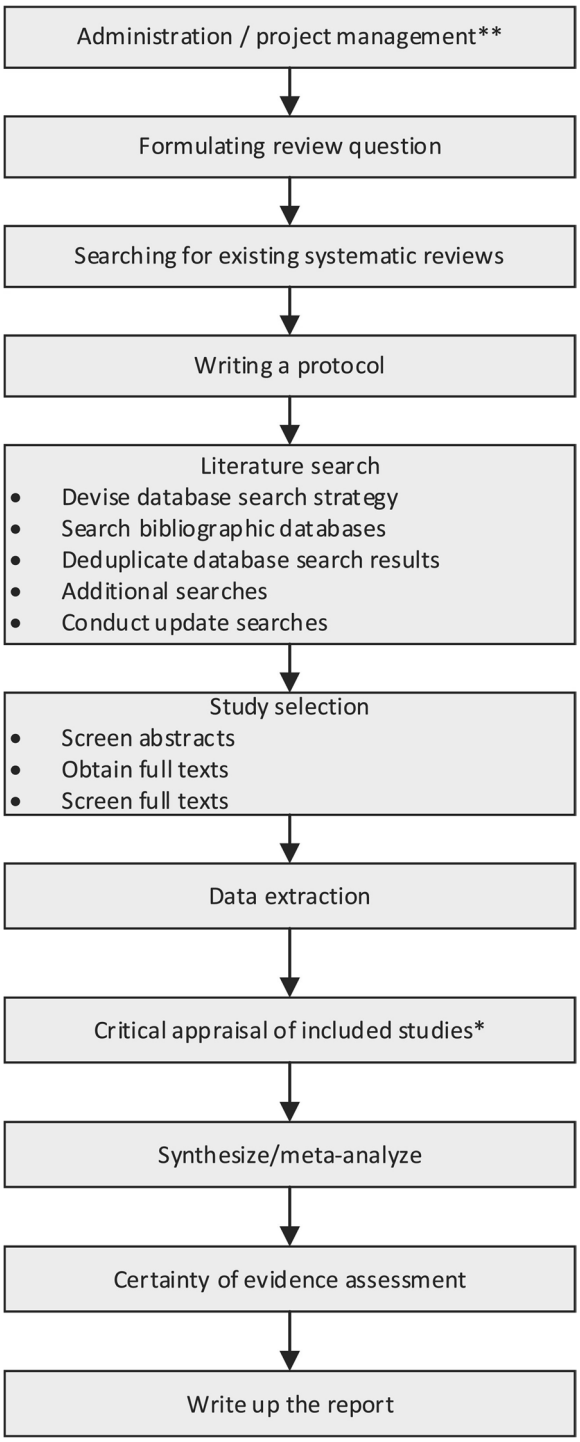

Fig. 1 Steps of the SR process. *added to Tsafnat et al.'s list [25]. ** added based on Nussbaumer-Streit et al. [14]

Results

We included 103 studies [10, 26–127] evaluating 21 methods ($n=51$) [26, 29, 30, 32, 34, 38, 40–43, 51, 53, 54, 56, 61, 62, 65–67, 73, 75, 77, 79, 80, 82, 83, 85–93, 100–102, 105, 107, 109–112, 114, 115, 117, 122, 123, 126, 127] and 35 tools ($n=54$) [10, 27, 28, 31, 33, 35–37, 39, 44–50, 52, 55, 57–60, 63, 64, 68–72, 74–76, 78, 81, 84, 94–99, 103, 104, 106, 108, 109, 113, 116, 118–121, 125, 128] (Fig. 2: PRISMA study flowchart). Table 3 provides an overview of the identified methods and tools. A total of 73 studies were validity studies ($n=70$) [26–35, 38, 40, 44–47, 49, 50, 52, 53, 57–59, 61, 63–67, 70, 71, 76, 78–80, 82–86, 88–92, 95–98, 100, 102–104, 106–113, 115–118, 120–123, 125, 126] or usability studies ($n=3$) [60, 68, 69] assessing a single method or tool, and 30 studies performed comparative analyses of different methods or tools [10, 36, 37, 39, 41–43, 48, 51, 54–56, 62, 72–75, 77, 81, 87, 93, 94, 99, 101, 105, 109, 114, 119, 127, 128]. Few studies prospectively evaluated methods or tools in a real-world workflow ($n=20$) [10, 28, 33, 36, 47, 51, 68, 69, 78, 79, 89, 91, 95, 99, 106, 109, 113, 115, 126, 128], 7 studies of those used independent testing (by a different reviewer team) with external data [10, 36, 47, 95, 99, 113, 128].

The majority of studies evaluated methods or tools for supporting the tasks of title and abstract screening ($n=42$) [33, 36, 37, 39, 44–46, 48, 49, 52, 56, 59, 60, 64, 74, 80, 83–85, 87–91, 95–97, 101, 103, 106–109, 113–115, 118–121, 126, 128] or devising the search strategy and performing the search ($n=24$) [29, 35, 38, 40, 43, 53, 54, 57, 61, 65–67, 73, 82, 92–94, 99, 100, 105, 111, 122, 123, 127] (see Fig. 3). For several steps of the SR process, only a few studies that evaluated methods or tools were identified: deduplication: $n=6$ [31, 37, 55, 58, 72, 81], additional search: $n=2$ [34, 98], update search: $n=6$ [37, 51, 62, 78, 110, 112], full-text selection: $n=4$ [86, 114, 115, 126], data extraction: $n=11$ [32, 37, 47, 68, 70, 71, 75, 104, 113, 125, 126] (one study evaluated both a method and a tool [75]); critical appraisal: $n=9$, [27, 28, 37, 50, 63, 69, 76, 102, 116], and combination of abbreviated methods/tools: $n=6$ [10, 26, 77, 79, 101, 117] (see Fig. 3). No studies were found for some steps of the SR process, such as administration/project management, formulating the review question, searching for existing reviews, writing the protocol, full-text retrieval, synthesis/meta-analysis, certainty of evidence assessment, and report preparation. In Appendix 2, we summarize the characteristics of all the included studies.

Most studies reported on validity outcomes ($n=84$, 46%) [10, 26–35, 39, 42–59, 62–64, 66, 67, 69–76, 78, 80, 81, 83–91, 94, 96, 97, 99, 101–114, 116–125, 127], while outcomes such as workload saving ($n=35$, 19%) [10, 28, 29, 32, 33, 39, 44–46, 48, 49, 52, 59, 64, 67, 84–87, 91, 95, 97, 103, 105–109, 111, 114, 115, 117, 119, 121, 127], time-saving ($n=24$, 13%) [33–35, 39, 45, 47, 48, 51, 52, 70–72, 74, 75, 87, 91, 96–99, 104, 109, 125, 126]; impact on results ($n=23$, 13%) [26, 32, 38, 40–43, 59, 61, 62, 65, 77, 79, 82, 92, 93, 100, 101, 111, 115, 117, 122, 127],

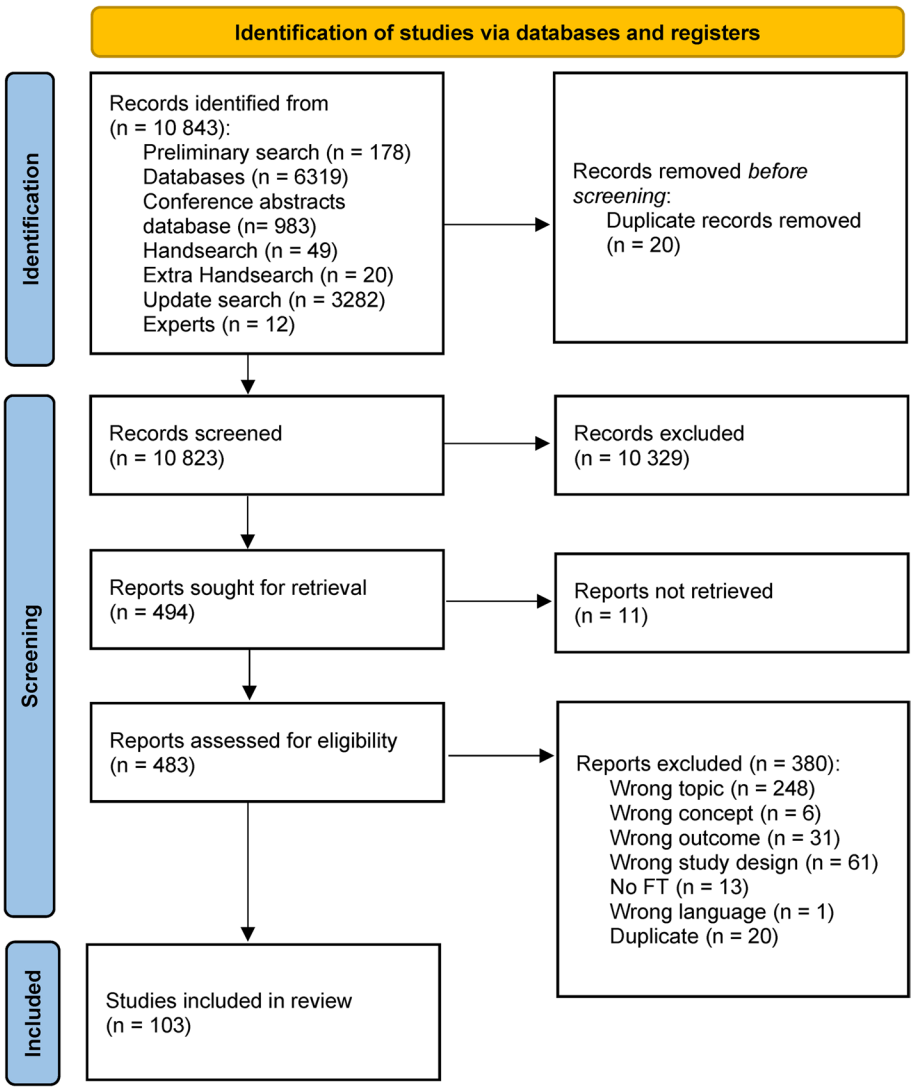

Fig. 2 PRISMA flowchart

usability ($n=13$, 7%) [36, 37, 39, 60, 68, 69, 78, 94, 95, 97, 98, 121, 125] and cost-saving ($n=3$, 2%) [33, 83, 114] were less evaluated (Fig. 4: Outcomes reported in the included studies). In Appendix 2, we map the efficiency and usability outcomes per tool and method against the review steps of the SR process.The included studies reported various validity outcomes (i.e., specificity, precision, accuracy) and time, costs, or workload savings to undertake the review. None of the studies reported the personnel effort saved.

Methods or tools for literature search

Search strategy and database search

Five tools (MeSH on Demand [94, 99], PubReMiner [94, 99], Polyglot Search Translator [35], Risklick search platform [57], and Yale MeSH Analyzer [94, 99]) and three methods (abbreviated search strategies for study type [53, 73, 111], topic [123], or search date [43, 127]; citation-based searching [29, 66, 67]; search restrictions for database [54, 93, 105, 122] and language (e.g. only English articles) [38, 40, 61, 65, 82, 92, 100]) were evaluated in 24 studies [29, 35, 38, 40, 43, 53, 54, 57, 61, 65–67, 73, 82, 92–94, 99, 100, 105, 111, 122, 123, 127] to support devising search strategies and/or performing literature searches.

Tools for search strategies

Using text mining tools for search strategy development (MeSH on Demand, PubReMiner, and YaleMeSH Analyzer)

Table 3 Identified methods and tools per review step

Review step	Methods	Tools
Administration/project management	-	-
Review question formulation	-	-
Search for existing reviews	-	-
Protocol preparation	-	-
Literature search		
Search strategy and database search	• Citation-based searching [29, 66, 67] • Search restrictions: database [54, 93, 105, 122], language [38, 40, 61, 65, 82, 92, 100] • Abbreviated search strategies for study type [53, 73, 111], topic [123], or search date [43, 127]	• MeSH on Demand [94, 99] • PubReMiner [94, 99] • Polyglot Search Translator [35] • Risklick search platform [57] • Yale MeSH Analyzer [94, 99]
Deduplication	-	• ASySD [58] • EBSCO [72] • EndNote [55, 58, 72, 81] • Covidence [81] • Deduklick [31] • Mendely [55, 72, 81] • OVID [72, 81] • Rayyan [55, 81] • Refworks [72] • SRA- Deduplicator [37, 55, 58] • Zotero [55, 81]
Additional search	• Scopus approach [34]	• Paperfetcher [98]
Update search	• Clinical Query search combined with PubMed-related articles search [110, 112] • Clinical Query search in MEDLINE and Embase [62] • McMaster Premium LiteratUre Service [62] • PubMed Similar Articles Search [51] • Scopus Citation Tracking [51]	• RobotReviewer LIVE [37, 78]
Study selection		
Title and abstract selection	• Crowdsourcing using different (automation) tools [85, 87–91] • Dual computer monitors [126] • Single-reviewer screening [56, 101, 114, 115] • Limited screening: o PICO-based title-only screening [107] o Review of reviews [109] • Title-first screening [80]	• AbstrackR [37, 39, 45, 46, 48, 49, 52, 60, 74, 108, 109, 119] • ASReview [84, 96, 103, 121] • ChatGPT [113] • Colandr [39] • Covidence [36, 37, 60] • DistillerSR [37, 44, 48, 59] • EPPI-reviewer [37, 60, 119, 128] • Rayyan [36, 37, 39, 60, 74, 95, 97, 120, 128] • RCT classifier [118] • Research screener [33] • RobotAnalyst [36, 37, 48, 106, 109] • SRA-helper for EndNote [36] • SWIFT-active screener [37, 64] • SWIFT-Review [74]
Full-text retrieval	-	-
Full-text selection	• Crowdsourcing using different (automation-) tools [86] • Dual computer monitors [126] • Single-reviewer screening [114, 115]	-
Data extraction	• Dual computer monitors [126] • Single-reviewer data extraction [32, 75]	• ChatGPT [113] • Data Abstraction Assistant [37, 68, 75] • Dextr [125] • ExaCT [47, 71, 104] • Plot Digitizer [70]
Critical appraisal	• Crowdsourcing with CrowdCARE [102]	• RobotReviewer [27, 28, 37, 50, 63, 69, 76, 116]
Synthesis/meta-analysis	-	-
Certainty of evidence assessment	-	-
Report preparation	-	-
Combination of abbreviated methods/tools	• Rapid review methods with multiple steps combined: abbreviated search/search limits [26, 77, 101, 117], single-reviewer screening/data extraction [26, 117], review of reviews [79], abbreviated critical appraisal [117]	Rapid review tools with multiple steps combined: SRA-Deduplicator, EndNote, Polyglot Search Translator, RobotReviewer, SRA-Helper [10]

Abbreviations: ASReview Automated Systematic Review, *ASySd* Automated Systematic Synthesis and Data extraction, *ChatGPT* Chat Generative Pre-trained Transformer (by OpenAI), *EPPI* Evidence for Policy and Practice Information and Co-ordinating Centre, *ExaCT* Extraction of Citations Tool, *MeSH* Medical Subject Headings, *PICO* Population Intervention Comparison Outcome, *PST* Polyglot Search Translator, *PLUS* McMaster Premium LiteratUre Service, *RCT* Randomized controlled trial, *SR* Systematic review, *SRA* Systematic Review Assistant, *SWIFT* Sciome Workbench for Interactive Computer-Facilitated Text-mining

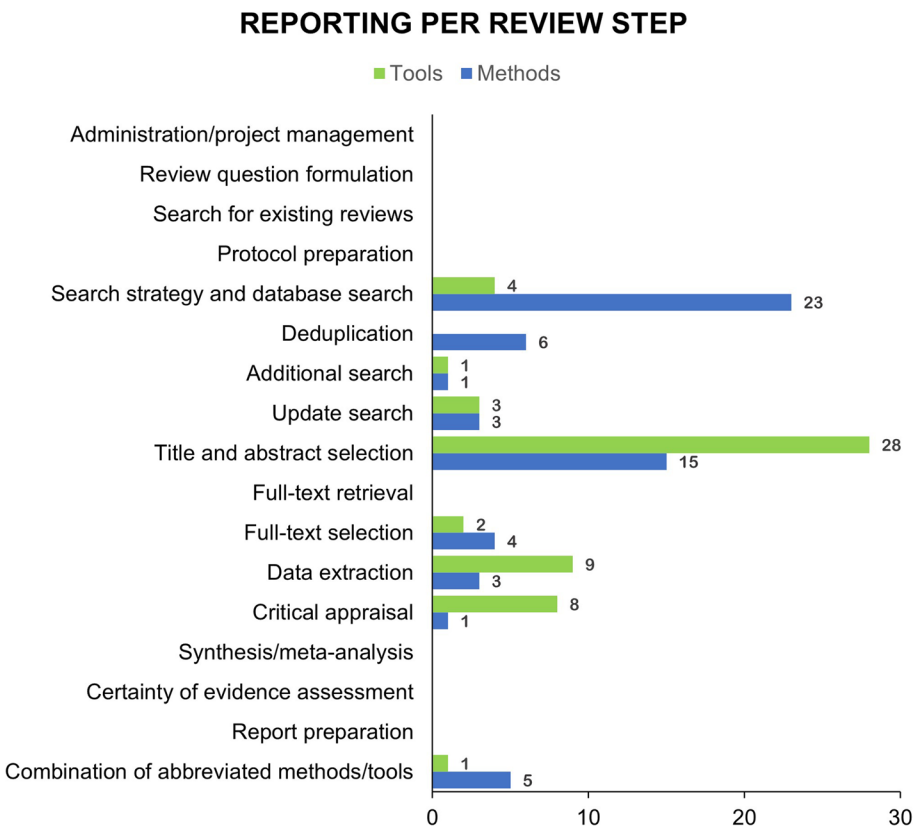

Fig. 3 The number of identified evaluation studies per review step

reduced time expenditure compared to manual searches, with tools saving over half the time required for manual searches (5 h, standard deviation [SD = 2] vs. 12 h [SD = 8]) [99]. Using supportive tools such as Polyglot Search Translator [35], MeSH on Demand, PubReMiner, and YaleMeSH Analyzer [94, 99] was less sensitive [99] and showed a slightly reduced precision compared to manual searches (11% to 14% vs. 15%) [94]. The Risklick search platform demonstrated a high precision for identifying clinical trials (96%) and COVID-19–related publications (94%) [57].

User ratings by the study authors indicated that PubReMiner and YaleMeSH Analyzer were considered "useful" or "extremely useful," while MeSH on Demand received the rating "not very useful" on a 5-point Likert scale (from extremely useful to least useful) [94].

Abbreviated search strategies for study type, topic, or search date

Two studies evaluated an abbreviated search strategy (i.e., Cochrane highly sensitive search strategy) [53] and a brief RCT strategy [111] for identifying RCTs. Both achieved high sensitivity rates of 99.5% [53] and 94% [111] while reducing the number of records requiring screening by 16% [111]. Although some RCTs were missed using abbreviated search strategies, there were no significant differences in the conclusions [111].

One study [123] assessed an abbreviated search strategy using only the generic drug name to identify drug-related RCTs, achieving high sensitivities in both MEDLINE (99%) and Embase (99.6%) [123].

Lee et al. (2012) evaluated 31 search filters for SRs and meta-analyses, with the health-evidence.ca Systematic Review search filter performing best, maintaining a high sensitivity while reducing the number of articles needing for screening (90% in MEDLINE, 88% in Embase, 90% in CINAHL) [73].

Furuya-Kanamori et al. [43] and Xu et al. [127] investigated the impact of restricting search timeframes on effect estimates and found that limiting searches to the most recent 10 to 15 years resulted in minimal changes in effect estimates (< 5%) while reducing workload by up to 45% [43, 127]. Nevertheless, this approach missed 21% to 35% of the relevant studies [43].

Citation-based searching

Three studies [29, 66, 67] assessed whether citation-based searching can improve efficiency in systematic reviewing. Citation-based searching achieved a reduction in the

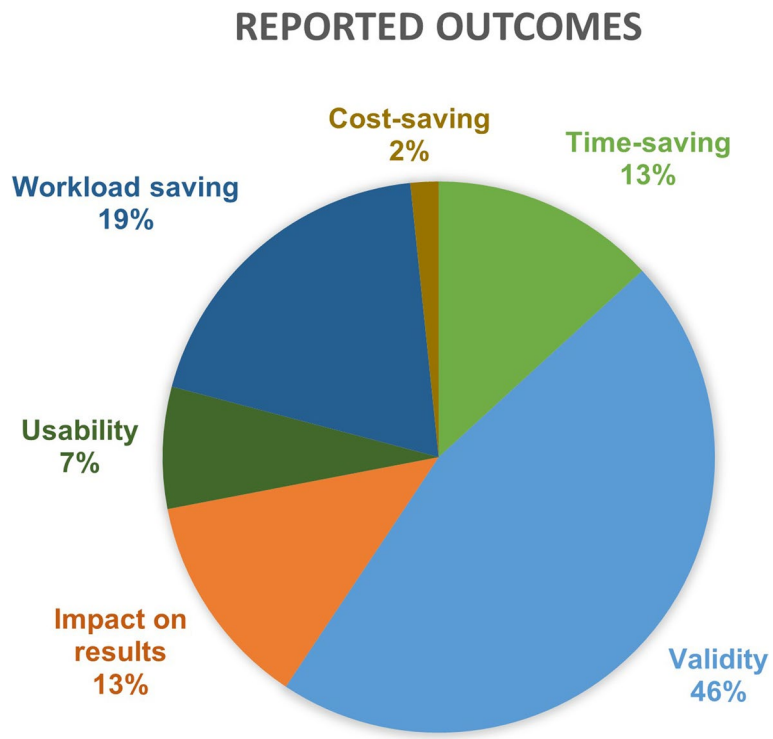

Fig. 4 Outcomes reported in the included studies

number of retrieved articles (50% to 89% fewer articles) compared to the original searches while still capturing a substantial proportion of 75% to 82% of the included articles [29, 67].

Restricted database searching

Seven studies assessed the validity of restricted database searching and suggested that searching at least two topic-related databases yielded high recall and precision for various types of studies [30, 41, 42, 54, 93, 105, 122].

Preston et al. (2015) demonstrated that searching only MEDLINE and Embase plus reference list checking identified 93% of the relevant references while saving 24% of the workload [105]. Beyer et al. (2013) emphasized the necessity of searching at least two databases along with reference list checking to retrieve all included studies [30]. Goossen et al. (2018) highlighted that combining MEDLINE with CENTRAL and hand searching was the most effective for RCTs (Recall: 99%), while for nonrandomized studies, combining MEDLINE with Web of Science yielded the highest recall (99.5%) [54]. Ewald et al. (2022) showed that searching two or more databases (MEDLINE/CENTRAL/Embase) reached a recall of ≥ 87.9% for identifying mainly RCTs [42]. Additionally, Van Enst et al. (2014) indicated that restricting searches to MEDLINE alone might slightly overestimate the results compared to broader database searches in diagnostic accuracy SRs (relative diagnostic odds ratio: 1.04; 95% confidence interval [CI], 0.95 to 1.15) [122]. Nussbaumer-Streit et al. (2018) and Ewald et al. (2020) found that combining one database with another or with searches of reference lists was noninferior to comprehensive searches (2%; 95% CI, 0% to 9%; if opposite concusion was of concern) [93] as the effect estimates were similar (ratio of odds ratios [ROR] median: 1.0 interquartile range [IQR]: 1.0–1.01) [41].

Restricted language searching

Seven studies found that excluding non-English articles to reduce workload would minimally alter the conclusions or effect estimates of the meta-analyses. Two studies found no change in the overall conclusions [61, 92], and five studies [38, 61, 65, 92, 100] reported changes in the effect estimates or statistical significance of the meta-analyses. Specifically, the statistical significance of the effect estimates changed in 3% to 12% of the meta-analyses [38, 61, 65, 92, 100].

Deduplication

Six studies [31, 37, 55, 58, 72, 81] compared eleven supportive software tools (ASySD, EBSCO, EndNote, Covidence, Deduklick, Mendeley, OVID, Rayyan, RefWorks, Systematic Review Accelerator, Zotero). Manual deduplication

took approximately 4 h 45 min, whereas using the tools reduced the time by 4 h 42 min to only 3 min [72]. False negative duplicates varied from 36 (Mendeley) to 258 (EndNote), while false positives ranged from 0 (OVID) to 43 (EBSCO) [72]. The precision was high with 99% to 100% for Deduklick and ASySD, and the sensitivity was highest for Rayyan (ranging from 99 to 100%) [55, 58, 81], followed by Covidence, OVID, Systematic Review Accelerator, Mendeley, EndNote, and Zotero [55, 58, 81]. However, Cowie et al. reported that the Systematic Review Accelerator received a low rating of 9/30 for its features and usability [37].

Additional literature search

Paperfetcher, identified as an application to automate additional searches as handsearching and citation searching, saved up to 92.0% of the time compared to manual handsearching and reference list checking, though validity outcomes for Paperfetcher were not reported [98]. Additionally, the Scopus approach, in which reviewers electronically downloaded the reference lists of relevant articles and screened only new references dually, saved approximately 62.5% of the time compared to manual checking [34].

Update literature search

We identified one tool (RobotReviewer LIVE) [37, 78] and five methods (Clinical Query search combined with PubMed-related articles search, Clinical Query search in MEDLINE and Embase, searching the McMaster Premium LiteratUre Service [PLUS], PubMed similar articles search, and Scopus citation tracking) [51, 62, 110, 112] for improving the efficiency of updating literature searches. RobotReviewer LIVE showed a precision of 55% and a high recall of 100% [78] with limitations including search restricted to MEDLINE, consideration of only RCTs, and low usability scores for features [37, 78].

The Clinical Query (CQ) search, combined with the PubMed-related articles search and the CQ search in MEDLINE and Embase, exhibited high recall rates ranging from 84 to 91% [62, 110, 112], while the PLUS database had a lower recall rate of 23% [62]. The PubMed similar articles search and Scopus citation tracking had a low sensitivity of 25% each, with time-saving percentages of 24% and 58%, respectively [51]. However, the omission of studies from searching the PLUS database only did not significantly change the effect estimates in most reviews (ROR: 0.99; 95% CI, 0.87 to 1.14) [62].

Methods or tools for study selection

Title and abstract selection

We identified 42 studies evaluating 14 supportive software tools (AbstrackR, ASReview, ChatGPT, Colandr, Covidence, DistillerSR, EPPI-reviewer, Rayyan, RCT classifier, Research screener, RobotAnalyst, SRA-Helper for EndNote, SWIFT-active screener, SWIFT-review) [33, 36, 37, 39, 44–46, 48, 49, 52, 59, 60, 64, 74, 84, 95–97, 103, 106, 108, 109, 113, 118–121, 128] using advanced text mining and machine and active learning techniques, and five methods (crowdsourcing using different [automation] tools, dual computer monitors, single-reviewer screening, PICO-based title-only screening, limited screening [review of reviews]) [56, 80, 83, 85, 87–91, 101, 107, 109, 114, 115, 126] for improving the title and abstract screening efficiency. The tested datasets ranged from 1 to 60 SRs and 148 to 190,555 records.

Tools for title and abstract selection

Various tools (e.g., EPPI-Reviewer, Covidence, DistillerSR, and Rayyan) offer collaborative online platforms for SRs, enhancing efficiency by managing and distributing screening tasks, facilitating multiuser screening, and tracking records throughout the review process [129].

In a semiautomated tool, the tool provide suggestions or probabilities regarding the eligibility of a reference for inclusion in the review, but human judgment is still required to make the final decision [7, 10]. In contrast, in a fully automated system, the tool makes the final decision without human intervention based on predetermined criteria or algorithms. Some tools provide fully automated screening options (e.g., DistillerSR), semiautomated (e.g., RobotAnalyst), or both (e.g., AbstrackR, DistillerAI) using machine learning or natural language processing methods [7, 10] (see Table 1).

Among the eleven semi- and fully -automated tools (AbstrackR [37, 39, 45, 46, 48, 49, 52, 60, 74, 108, 109, 119], ASReview [84, 96, 103, 121], ChatGPT [113], Colandr [39], DistillerSR [37, 44, 48, 59], EPPI-reviewer [37, 60, 119, 128], Rayyan [36, 37, 39, 74, 95, 97, 120, 128], RCT classifier [118], Research screener [33], RobotAnalyst [36, 37, 48, 106, 109], SRA-helper for EndNote [36], SWIFT-active screener [37, 64], SWIFT-review [74]). ASReview [84, 96, 103, 121] and Research Screener [33] demonstrated a robust performance, identifying 95% of the relevant studies while saving 37% to 92% of the workload. SWIFT-active screener [37, 64], RobotAnalyst [36, 37, 48, 106, 109], and Rayyan [36, 37, 95, 97, 120] also performed well, identifying 95% of the relevant studies with workload savings from 34 to 49%. EPPI-Reviewer identified all the relevant abstracts after screening 40% to 99% of the references across reviews [119]. DistillerAI showed substantial workload savings of 53% while identifying 95% of the relevant studies [59], with varying degrees of validity [44, 48, 59]. Colandr and SWIFT-Review exhibited sensitivity rates of 65% and 91%, respectively,

with 97% workload savings and around 400 min of time saved [39, 74]. ChatGPT's sensitivity was 100% [113] and RCT classifiers recall was 99% [118]; workload or time savings were not reported [113, 118]. AbstrackR showed a moderate performance with a potential workload savings from 4% up to 97% [39, 45, 46, 48, 49, 52, 108, 109, 119] while missing up to 44% of the relevant studies [39, 46, 49, 52, 109]. Covidence and SRA-Helper for EndNote did not report validity outcomes.

Most of the supportive software tools were easy to use or learn, suitable for collaboration, and straightforward for inexperienced users (ASReview, AbstarckR, Covidence, SRA-Helper, Rayyan, RobotAnalyst) [36, 37, 45, 60, 95, 97, 121]. Other tools were more complex regarding their usability but were useful for large and complex projects (DistillerSR, EPPI-Reviewer) [37, 48, 60]. Poor interface quality (AbstrackR) [60], issues with help section/response time (RobotAnalyst, Covidence, EPPI) [36, 60], and overloaded side panel (Rayyan) [36] were weaknesses reported in the studies.

Methods for title and abstract selection

Among the four methods identified (dual computer monitors, single-reviewer screening, crowdsourcing using different [automation] tools, and limited screening [review of reviews, PICO-based title-only screening, title-first screening]) [56, 80, 83, 85, 87–91, 101, 114, 115, 126] for supporting title and abstract screening, crowdsourcing in combination with screening platforms or machine learning tools demonstrated the most promising performance in improving efficiency. Studies by Noel-Storr et al. [83, 88–91] found that the Cochrane Crowd plus Screen4Me/RCT classifier achieved a high sensitivity ranging from 84 to 100% in abstract screening and reduced screening time. Crowdsourcing via Amazon Mechanical Turk yielded correct inclusions of 95% to 99% with a substantial cost reduction of 82% [83]. However, the sensitivity was moderate when the screening was conducted manually by medical students or on web-based platforms (47% to 67%) [87].

Single-reviewer screening missed 0% to 19% of the relevant studies [56, 101, 114, 115] while saving 50% to 58% of the time and costs, respectively [114]. The findings indicate that single-reviewer screening by less-experienced reviewers could substantially alter the results, whereas experienced reviewers had a negligible impact [101].

Limited screening methods, such as reviews of reviews (also known as umbrella reviews), exhibited a moderate sensitivity (56%) and significantly reduced the number of citations needed to screen [109]. Title-first screening and PICO-based title screening demonstrated an accurate validity, with a recall of 100% [80, 108] and a reduction in screening effort ranging from 11 to 78% [108]. However, screening with dual computer monitors did not notably improve the time saved [126].

Full-text selection

For full-text screening, we identified three methods: crowdsourcing [86], using dual computer monitors [126], and single-reviewer screening [114, 115]. Using crowdsourcing in combination with the CrowdScreenSR saved 68% of the workload [86]. With dual computer monitors, no significant difference of time taken for the full-text screening was reported [126]. Single-reviewer screening missed 7% to 12% of the relevant studies [115] while saving only 4% of the time and costs [114].

Methods or tools for data extraction

We identified 11 studies evaluating five tools (ChatGPT [113], Data Abstraction Assistant (DAA) [37, 68, 75], Dextr [125], ExaCT [47, 71, 104] and Plot Digitizer [70]) and two methods (dual computer monitors [126] and single data extraction [32, 75]) to expedite the data extraction process.

ExaCT [47, 71, 104], DAA [37, 68, 75], Dextr [125], and Plot Digitizer [70] achieved a time reduction of up to 60% [104, 125], with precision rates of 93% for ExaCT [71, 104] and 96% for Dextr and an error rate of 17% for DAA [68, 75]. Manual extraction by two reviewers and with the assistance of Plot Digitizer showed a similar agreement with the original data, showing a slightly higher agreement with the assistance of Plot Digitizer (Plot Digitizer: 73% and 75%, manual extraction: 66% and 69%) [70]. A total of 87% of manually extracted data elements matched with ExaCT, resulting in qualitatively altered meta-analysis results [104]. ChatGPT demonstrated consistent agreement with human researchers across various parameters ($\kappa = 0.79–1$) extracted from studies, such as language, targeted disease, natural language processing model, sample size, and performance parameters, and moderate to fair agreement for clinical task ($\kappa = 0.58$) and clinical implementation ($\kappa = 0.34$) [113]. Usability was assessed only for DAA and Dextr, with both tools deemed very easy to use [68, 125], although DAA scored lower on feature scores, while Dextr was noted for its flexible interface [125].

Single data extraction and dual monitors reduced the time for extracting data by 24 to 65 min per article [32, 75, 126], with similar error rates between single and dual data extraction methods (single: 16% [75], 18% [32], dual: 15% [32, 75]) and comparable pooled estimates [32, 75].

Methods or tools for critical appraisal

We identified nine studies reporting on one software tool (RobotReviewer) [27, 28, 37, 50, 63, 69, 76, 116] and one method (crowdsourcing via CrowdCARE) [102] for improving critical appraisal efficiency. Collectively, the study authors suggested that RobotReviewer can support but not replace RoB assessments by humans [27, 28, 50, 63, 69, 76] as performance varied per RoB domain [27, 50, 63, 116]. The authors reported similar completion times for RoB appraisal with and without RobotReviewer assistance [28]. Reviewers were equally as likely to accept RobotReviewer's judgments as one another's during consensus (83% for RobotReviewer, 81% for humans) [69], showing similar accuracy (RobotReviewer assisted RoB appraisal: 71%, RoB appraisal by two reviewers: 78%) [28, 76]. The reviewers generally described the tool as acceptable and useful [69], whereby collaboration with other users is not possible [37].

Combination of abbreviated methods/tools

Five studies evaluated RR methods [26, 77, 79, 101, 117], and one study evaluated various tools [10] combining multiple review steps. While two case studies found no differences in findings between RR and SR approaches [79, 117], another author/paper/study found in two of three RRs that no conclusion could be drawn due to insufficient information [26]. Additionally, in a study including three RRs, RR methods affected almost one-third of the meta-analyses with less precise pooled estimates [101]. Marshall et al. (2019) included 2,512 SRs and reported a loss of all data in 4% to 45% of the meta-analyses and changes of 7% to 39% in the statistical significance due to RR methods [77]. Automation tools (SRA- Deduplicator, EndNote, Polyglot Search Translator, RobotReviewer, SRA-Helper) reduced the person-time spent on SR tasks (42 h versus 12 h) [10]. However, error rates, falsely excluded studies, and sensitivity varied immensely across studies [26, 117].

Discussion

To the best of our knowledge, this is the first scoping review to map evaluated methods and tools for improving the efficiency of SR production across the various review steps. We conducted this scoping review to bridge the gap in understanding the validity and usability of various methods and tools across different stages of the systematic review process, addressing the challenges of limited awareness, concerns about validity, and usability issues among review teams. We describe which review steps methods and ready-to-use tools are available and have been evaluated. Additionally, we provide an overview of the contexts in which these methods and tools were evaluated, such as real-time workflow testing and the use of internal or external data.

Across all the SR review steps, most studies evaluated study selection, followed by literature searching and data extraction. Around half of the studies evaluated tools and half of the studies methods. For study selection, most of the tools offered semiautomated-assisted screening by classifying or ranking references. The methods focused mainly on limiting the review team's human resources, for example, through single-reviewer screening or by distributing the tasks to a crowd or students.

Two scoping reviews, one on tools [8] and one on methods [13] to support the SR process, are in line with this result, as the authors also identified these tasks as the most frequently evaluated in the literature [8, 13]. As shown by our scoping review and others [8], a major focus on (semi) automation tools for study selection occurred in recent years. This is important, as a recent study on resource use found that study selection and data extraction are the most resource-intensive tasks besides administration and project management as well as critical appraisal [14].

For the following tasks, we could not identify a single study evaluating a tool or method: administration/project management, formulating the review question, writing the protocol, searching for existing reviews, full-text retrieval, synthesis/meta-analysis, certainty of evidence assessment, and report preparation. However, all of these tasks are also time-consuming, especially according to the study by Nussbaumer-Streit et al. [14]. Project management requires the largest proportion of SR production time [14]. To our knowledge, tools supporting project management are already available, such as Covidence, DistillerSR, EPPI-Reviewer, or simple online platforms such as Google Forms, which can also support managing and coordinating projects. However, no evaluations of these support platforms were found. Similarly, while innovative software tools such as large language models (e.g., ChatGPT) or other technological solutions (e.g. Shiny app for producing PRISMA flow-diagrams [130]) show promise in supporting tasks such as report preparation, there is a lack of formal evaluation in this context as well. This is relevant for future research that aims to improve SR production since these tasks are extremely resource-intensive.

Our scoping review identified several research gaps. There is a lack of studies evaluating the usability of tools and methods. No study evaluated the usability of any single method. Only for the study selection task did we identify multiple studies evaluating the usability of tools. However, an important factor in adopting tools and methods is their user-friendliness [12, 131] and their fit with standard SR workflows [7, 48]. Furthermore, if

usability was considered, this was often evaluated in a nonformal or standardized way employing various scales, questions, and feedback mechanisms. To enable meaningful comparisons between different methods, there is a clear need for a formal analysis of user experience and usability. Therefore, authors and review teams would benefit from comparable usability studies on methods and tools that aim to improve the SR process's efficiency.

Few studies exist that evaluate the impact on results when using accelerated methods or tools. We identified 25 studies (13%) where the odds ratio changed by 0% to 63% depending on the method or tool [41, 61, 62, 92, 100, 122, 132]. Marshall et al. (2019) stated that there has not been a large-scale evaluation of the effects of RR methods on the number of falsely excluded studies and the consequent changes in meta-analysis results [133]. Indeed, understanding the potential impact of different methods and tools on the results is fundamental, as emphasized by Wagner et al. (2016) [134]. Wagner et al. conducted an online survey of guideline developers and policymakers, aiming to discover how much incremental uncertainty about the correctness of an answer would be acceptable in RRs. They found that participants demanded very high levels of accuracy and would tolerate a median 10% risk of wrong answers [134]. Therefore, studies focusing on the impact on results and conclusions through RR methods and tools are warranted.

The majority of studies retrospectively evaluated only a single tool or method using existing internal data, offering limited insights into real-world adoption. Prospective studies within a real-time workflow study ($n=20$) [10, 28, 33, 36, 47, 51, 68, 69, 78, 79, 89, 91, 95, 99, 106, 109, 113, 115, 126, 128], comparing several tools and methods ($n=6$) [10, 36, 51, 99, 109, 128] or from independent reviewer teams using their own dataset ($n=7$) [10, 36, 47, 95, 99, 113, 128], are scarce. However, such studies are crucial for providing valid comparative evidence on validity, workload savings, usability, impact on results, and real-world benefits. Particularly for automated title and abstract screening, where most tools function similarly, key information such as the stopping rule (indicating when screening can cease) is essential. Notably, there is limited research ($n=8$) [52, 64, 95, 96, 108, 109, 120, 128] exploring the combined effects of algorithmic re-ranking and stopping criterion determination in automated title and abstract screening. Furthermore, studies assessing the influence of automated tools (e.g., re-ranking algorithms) on human decision-making are lacking. Therefore, simulation and prospective real-time studies evaluating the workflow between manual procedures and tools with stopping criterion are warranted.

The balance between the time saved and the potential reduction in quality and comprehensiveness is influenced by various factors, including the decision-making urgency, resource availability, and the decision-makers' specific needs. However, accelerated approaches are not universally appropriate. In situations where the thoroughness and rigor of the evidence are paramount—such as in developing clinical guidelines, conducting health technology assessments, or addressing areas with significant scientific uncertainty—the risk of missing critical evidence or drawing inaccurate conclusions outweighs the benefits of speed.

Given the heterogeneity in study designs and contexts, there is a pressing need for standardized frameworks for the evaluation and reporting of tools and methods for SR production. Furthermore, researchers should be aware of the importance of testing methods and tools with their own datasets and contextual factors. Pretraining both the tools and the crowd before implementation is essential for optimizing efficiency and ensuring reliable outcomes. However, we believe our findings are generalizable to other types of evidence syntheses, such as scoping reviews, reviews of reviews, or RRs. Additionally, we highlight that a part of the Rapid Review Methods Series [129] offers guidance on the use of supportive tools, aiming to assist researchers in effectively navigating the complexities of SR and RR production.

Our scoping review has several limitations. First, our inclusion criteria focused solely on studies mentioning efficiency improvements. While this criterion aimed to strike a balance between screening workload and sensitivity, it may have inadvertently excluded relevant studies that did not explicitly highlight efficiency gains. We might have missed relevant studies. Second, the heterogeneity among the included studies poses challenges in generalizing the findings to other review teams and contexts. The narrow focus of many studies, along with their publication primarily in English and focus on specific study types, further limits the generalizability of our findings. Moreover, the limited proportion of studies (29%, 30/103) comparing different tools and methods on the same dataset within the same study accentuates the need for caution while interpreting the findings. However, we think the validity and usability outcomes reported in this scoping review provide a good orientation.

Conclusion

Based on the identified evidence, various methods and tools for literature searching and title and abstract screening are available with the aim of improving efficiency. However, only few studies have addressed the influence of these methods and tools in real-world workflows. Fewer studies evaluated methods or tools supporting the other tasks of SR production. Moreover, the reporting of the outcomes of existing evaluations varies

considerably, revealing significant research gaps, especially in assessing usability and impact on results. Future research should prioritize addressing these gaps, evaluating real-world adoption, and establishing standardized frameworks for the evaluation of methods and tools to enhance the overall effectiveness of SR development processes.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12874-024-02320-4.

Supplementary Material 1.

Supplementary Material 2.

Supplementary Material 3.

Acknowledgements
We would like to thank Hans Lund and Jos Kleijnen for their input on the manuscript.

Authors' contributions
LA, BNS, RS, and LH developed the study concept. LA wrote the protocol. RS conducted the literature searches and similar article searches. LA, MMM, IS, BNS, MMK, MEM, EB, ME, RS, PNL, GP, NR, KG, LK, MS, AMP, and PM screened the references and extracted the data. LA and MMM conducted grey literature searches, reference list checking, and hand searches. RS provided methodological input throughout the study. LA and MMM created the first draft of the manuscript, which all other authors critically revised. All authors read and approved the final version of the submitted manuscript.

Funding
This work was supported by funds from the EU funded COST Action EVBRES (CA17117), two COST STSM fundings (LA, MM), and by a scholarship for the first author (LA) of the Gesellschaft für Forschungsförderung Niederösterreich m.b.H.

Availability of data and materials
No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Cochrane Austria, Department for Evidence-Based Medicine and Clinical Epidemiology, University for Continuing Education Krems, Krems an der Donau, Austria. [2]School for Public Health and Primary Care (CAPHRI), Maastricht University, Maastricht, the Netherlands. [3]Knowledge Institute of Federation of Medical Specialists, Utrecht, The Netherlands. [4]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands. [5]Zenica Cantonal Hospital, Department for Clinical Pharmacology, Zenica, Bosnia and Herzegovina. [6]Department of Health Policy and Management, Guilford Glazer Faculty of Business and Management and Faculty of Health Sciences, Ben-Gurion University of the Negev, Beer-Sheva, Israel. [7]Institute of Health Policy Management and Evaluation, Dalla Lana School Of Public Health, University of Toronto, Toronto, Canada. [8]McMaster Health Forum, McMaster University, Hamilton, Canada. [9]Witten/Herdecke University, Institute for Research in Operative Medicine (IFOM), Cologne, Germany. [10]Czech National Centre for Evidence-Based Healthcare and Knowledge Translation (Cochrane Czech Republic, Czech CEBHC: JBI Centre of Excellence, Masaryk University GRADE Centre), Institute of Biostatistics and Analyses, Faculty of Medicine, Masaryk University, Brno, Czech Republic. [11]Cochrane Netherlands, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands. [12]Department of Pathology, Faculty of Medicine and Surgery, University of Malta, Msida, Malta. [13]Laboratory of Hygiene and Environmental Protection, Department of Medicine, Democritus University of Thrace, Alexandroupolis, Greece. [14]CINTESIS.RISE@UA, University of Aveiro, Campus Universitário de Santiago, Aveiro, Portugal. [15]Section of Hygiene, University Department of Life Sciences and Public Health, Università Cattolica del Sacro Cuore, Rome, Italy. [16]Department of Medical and Surgical Sciences, University of Bologna, Bologna, Italy. [17]National Institute for Health Services Management, Bucharest, Romania. [18]Spiru Haret University, Faculty of Psychology and Educational Sciences, Bucharest, Romania. [19]Red de Nutrición Basada en La Evidencia, Academia Española de Nutrición y Dietética, Pamplona, Spain. [20]Department of Health Sciences, Faculty of Medicine, Masaryk University, Brno, Czech Republic. [21]Techné Research Group, Department of Knowledge Engineering of the Faculty of Science, University of Granada, Granada, Spain. [22]RTI International, Center for Public Health Methods, Research Triangle Park, Durham, NC, USA. [23]Amsterdam UMC, University of Amsterdam, Medical Library, Amsterdam Public Health, Amsterdam, the Netherlands.

Received: 17 June 2024 Accepted: 26 August 2024
Published online: 18 September 2024

References

1. Oliver S, Dickson K, Bangpan M. Systematic reviews: making them policy relevant. A briefing for policy makers and systematic reviewers. London: EPPI-Centre, Social Science Research Unit, UCL Institute of Education, University College London; 2015.
2. Borah R, Brown AW, Capers PL, Kaiser KA. Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry. BMJ Open. 2017;7(2):e012545.
3. Donnelly CA, Boyd I, Campbell P, Craig C, Vallance P, Walport M, Whitty CJM, Woods E, Wormald C. Four principles to make evidence synthesis more useful for policy. Nature. 2018;558(7710):361–4.
4. Clayton GL, Smith IL, Higgins JPT, Mihaylova B, Thorpe B, Cicero R, Lokuge K, Forman JR, Tierney JF, White IR, et al. The INVEST project: investigating the use of evidence synthesis in the design and analysis of clinical trials. Trials. 2017;18(1):219.
5. Peters MDJ, Marnie C, Tricco AC, Pollock D, Munn Z, Alexander L, McInerney P, Godfrey CM, Khalil H. Updated methodological guidance for the conduct of scoping reviews. JBI Evidence Synthesis. 2020;18(10).
6. Beller E, Clark J, Tsafnat G, Adams C, Diehl H, Lund H, Ouzzani M, Thayer K, Thomas J, Turner T. Making progress with the automation of systematic reviews: principles of the International Collaboration for the Automation of Systematic Reviews (ICASR). Syst Rev. 2018;7(1):77.
7. O'Connor AM, Tsafnat G, Thomas J, Glasziou P, Gilbert SB, Hutton B. A question of trust: can we build an evidence base to gain trust in systematic review automation technologies? Syst Rev. 2019;8(1):1–8.
8. Khalil H, Ameen D, Zarnegar A. Tools to support the automation of systematic reviews: a scoping review. J Clin Epidemiol. 2022;144:22–42.
9. Khurana D, Koli A, Khatter K, Singh S. Natural language processing: state of the art, current trends and challenges. Multimedia Tools and Applications. 2023;82(3):3713–44.
10. Clark J, McFarlane C, Cleo G, Ishikawa Ramos C, Marshall S. The Impact of Systematic Review Automation Tools on Methodological Quality and Time Taken to Complete Systematic Review Tasks: Case Study. JMIR Med Educ. 2021;7(2):e24418.
11. Ergonomics of human-system interaction - Part 11: Usability: Definitions and concepts, (ISO 9241–11: 2018). ISO 9241–11:2018 - Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts.
12. Scott AM, Forbes C, Clark J, Carter M, Glasziou P, Munn Z. Systematic review automation tools improve efficiency but lack of knowledge impedes their adoption: a survey. J Clin Epidemiol. 2021;138:80–94.

13. Hamel C, Michaud A, Thuku M, Affengruber L, Skidmore B, Nussbaumer-Streit B, Stevens A, Garritty C. Few evaluative studies exist examining rapid review methodology across stages of conduct: a systematic scoping review. J Clin Epidemiol. 2020;126:131–40.
14. Nussbaumer-Streit B, Ellen M, Klerings I, Sfetcu R, Riva N, Mahmić-Kaknjo M, Poulentzas G, Martinez P, Baladia E, Ziganshina LE et al. Resource use during systematic review production varies widely: a scoping review. J Clin Epidemiol. 2021.
15. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol. 2005;8(1):19–32.
16. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci. 2010;5(1):69.
17. Colquhoun HL, Levac D, O'Brien KK, Straus S, Tricco AC, Perrier L, Kastner M, Moher D. Scoping reviews: time for clarity in definition, methods, and reporting. J Clin Epidemiol. 2014;67(12):1291–4.
18. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, Moher D, Peters MD, Horsley T, Weeks L. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med. 2018;169(7):467–73.
19. JBI Reviewer's Manual. Chapter 11.2.5. Search Strategy.
20. Peters MDJ, Godfrey C, McInerney P, Munn Z, Tricco AC, Khalil H. Chapter 11: scoping reviews (2020 version). In: Aromataris E, Munn Z, editors. JBI manual for evidence synthesis, JBI, 2020. 2020. Available from: https://wiki.jbi.global/display/MANUAL/Chapter+11%3A+Scoping+reviews.
21. Sutton A, Marshall C. PRM246 - Mapping The Systematic Review Toolbox. Value in Health. 2017;20(9):A775.
22. Best L, Stevens A, Colin-Jones D. Rapid and responsive health technology assessment: the development and evaluation process in the South and West region of England. Journal of Clinical Effectiveness. 1997.
23. Jonnalagadda S, Petitti D. A new iterative method to reduce workload in systematic review process. Int J Comput Biol Drug Des. 2013;6(1–2):5–17.
24. Higgins JPT TJ, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.2 (updated February 2021); 2021.
25. Tsafnat G, Glasziou P, Choong MK, Dunn A, Galgani F, Coiera E. Systematic review automation technologies. Syst Rev. 2014;3:74.
26. Affengruber L, Wagner G, Waffenschmidt S, Lhachimi SK, Nussbaumer-Streit B, Thaler K, Griebler U, Klerings I, Gartlehner G. Combining abbreviated literature searches with single-reviewer screening: three case studies of rapid reviews. Syst Rev. 2020;9(1):162.
27. Armijo-Olivo S, Craig R, Campbell S. Comparing machine and human reviewers to evaluate the risk of bias in randomized controlled trials. Res. 2020;11(3):484–93.
28. Arno A, Thomas J, Wallace B, Marshall IJ, McKenzie JE, Elliott JH. Accuracy and Efficiency of Machine Learning-Assisted Risk-of-Bias Assessments in "Real-World" Systematic Reviews : A Noninferiority Randomized Controlled Trial. Ann Intern Med. 2022;175(7):1001–9.
29. Belter CW. Citation analysis as a literature search method for systematic reviews. J Assoc Soc Inf Sci Technol. 2016;67(11):2766–77.
30. Beyer FR, Wright K. Can we prioritise which databases to search? A case study using a systematic review of frozen shoulder management. Health Info Libr J. 2013;30(1):49–58.
31. Borissov N, Haas Q, Minder B, Kopp-Heim D, von Gernler M, Janka H, Teodoro D, Amini P. Reducing systematic review burden using Deduklick: a novel, automated, reliable, and explainable deduplication algorithm to foster medical research. Syst. 2022;11(1):172.
32. Buscemi N, Hartling L, Vandermeer B, Tjosvold L, Klassen TP. Single data extraction generated more errors than double data extraction in systematic reviews. J Clin Epidemiol. 2006;59(7):697–703.
33. Chai KEK, Lines RLJ, Gucciardi DF, Ng L: Research Screener: a machine learning tool to semi-automate abstract screening for systematic reviews. Syst Rev.2021;10(1).
34. Chapman AL, Morgan LC, Gartlehner G. Semi-automating the manual literature search for systematic reviews increases efficiency. Health Info Libr J. 2010;27(1):22–7.
35. Clark JM, Sanders S, Carter M, Honeyman D, Cleo G, Auld Y, Booth D, Condron P, Dalais C, Bateup S, et al. Improving the translation of search strategies using the Polyglot Search Translator: a randomized controlled trial. J Med Libr Assoc. 2020;108(2):195–207.
36. Cleo G, Scott AM, Islam F, Julien B, Beller E. Usability and acceptability of four systematic review automation software packages: a mixed method design. Syst. 2019;8(1):145.
37. Cowie K, Rahmatullah A, Hardy N, Holub K, Kallmes K. Web-Based Software Tools for Systematic Literature Review in Medicine: Systematic Search and Feature Analysis. JMIR Med Inform. 2022;10(5):e33219.
38. Dechartres A, Atal I, Riveros C, Meerpohl J, Ravaud P. Association between publication characteristics and treatment effect estimates: a meta-epidemiologic study. Ann Intern Med. 2018;169(6):385–93.
39. dos Reis AHS, de Oliveira ALM, Fritsch C, Zouch J, Ferreira P, Polese JC. Usefulness of machine learning softwares to screen titles of systematic reviews: a methodological study. Syst Rev. 2023;12(1):68.
40. Egger M, Juni P, Bartlett C, Holenstein F, Sterne J. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study Health Technol Assess. 2003;7(1):1–76.
41. Ewald H, Klerings I, Wagner G, Heise TL, Dobrescu AI, Armijo-Olivo S, Stratil JM, Lhachimi SK, Mittermayr T, Gartlehner G, et al. Abbreviated and comprehensive literature searches led to identical or very similar effect estimates: a meta-epidemiological study. J Clin Epidemiol. 2020;128:1–12.
42. Ewald H, Klerings I, Wagner G, Heise TL, Stratil JM, Lhachimi SK, Hemkens LG, Gartlehner G, Armijo-Olivo S, Nussbaumer-Streit B. Searching two or more databases decreased the risk of missing relevant studies: a metaresearch study. J Clin Epidemiol. 2022;149:154–64.
43. Furuya-Kanamori L, Lin L, Kostoulas P, Clark J, Xu C. Limits in the search date for rapid reviews of diagnostic test accuracy studies. Res. 2022;13:13.
44. Gartlehner G, Wagner G, Lux L, Affengruber L, Dobrescu A, Kaminski-Hartenthaler A, Viswanathan M. Assessing the accuracy of machine-assisted abstract screening with DistillerAI: a user study. Syst. 2019;8(1):277.
45. Gates A, Gates M, DaRosa D, Elliott SA, Pillay J, Rahman S, Vandermeer B, Hartling L: Decoding semi-automated title-abstract screening: findings from a convenience sample of reviews. *Systematic Reviews* 2020, 9(1).
46. Gates A, Gates M, Sebastianski M, Guitard S, Elliott SA, Hartling L. The semi-automation of title and abstract screening: a retrospective exploration of ways to leverage Abstrackr's relevance predictions in systematic and rapid reviews. BMC Med Res Methodol. 2020;20(1):139.
47. Gates A, Gates M, Sim S, Elliott SA, Pillay J, Hartling L. Creating efficiencies in the extraction of data from randomized trials: a prospective evaluation of a machine learning and text mining tool. BMC Med Res Methodol. 2021;21(1):169.
48. Gates A, Guitard S, Pillay J, Elliott SA, Dyson MP, Newton AS, Hartling L. Performance and usability of machine learning for screening in systematic reviews: a comparative evaluation of three tools. Syst. 2019;8(1):278.
49. Gates A, Johnson C, Hartling L. Technology-assisted title and abstract screening for systematic reviews: a retrospective evaluation of the Abstrackr machine learning tool. Syst. 2018;7(1):45.
50. Gates A, Vandermeer B, Hartling L. Technology-assisted risk of bias assessment in systematic reviews: a prospective cross-sectional evaluation of the RobotReviewer machine learning tool. J Clin Epidemiol. 2018;96:54–62.
51. Gates M, Elliott SA, Gates A, Sebastianski M, Pillay J, Bialy L, Hartling L. LOCATE: a prospective evaluation of the value of Leveraging Ongoing Citation Acquisition Techniques for living Evidence syntheses. Syst. 2021;10(1):116.
52. Giummarra MJ, Lau G, Gabbe BJ. Evaluation of text mining to reduce screening workload for injury-focused systematic reviews. Inj Prev. 2020;26(1):55–60.
53. 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(2):130–6.
54. Goossen K, Tenckhoff S, Probst P, Grummich K, Mihaljevic AL, Büchler MW, Diener MK. Optimal literature search for systematic reviews in surgery. Langenbecks Arch Surg. 2018;403(1):119–29.
55. Guimaraes NS, Ferreira AJF, Ribeiro Silva RC, de Paula AA, Lisboa CS, Magno L, Ichiara MY, Barreto ML. Deduplicating records in systematic

reviews: there are free, accurate automated ways to do so. J Clin Epidemiol. 2022;152:110–5.
56. Gartlehner G, Affengruber L, Titscher V, Noel-Storr A, Dooley G, Ballarini N, König F. Single-reviewer abstract screening missed 13 percent of relevant studies: a crowd-based, randomized controlled trial. J Clin Epidemiol. 2020;121:20–8.
57. Haas Q, Alvarez DV, Borissov N, Ferdowsi S, von Meyenn L, Trelle S, Teodoro D, Amini P. Utilizing Artificial Intelligence to Manage COVID-19 Scientific Evidence Torrent with Risklick AI: A Critical Tool for Pharmacology and Therapy Development. Pharmacology. 2021;106(5–6):244–53.
58. Hair K, Bahor Z, Macleod M, Liao J, Sena ES. The Automated Systematic Search Deduplicator (ASySD): a rapid, open-source, interoperable tool to remove duplicate citations in biomedical systematic reviews. BMC Biol. 2023;21(1):189.
59. Hamel C, Kelly SE, Thavorn K, Rice DB, Wells GA, Hutton B. An evaluation of DistillerSR's machine learning-based prioritization tool for title/abstract screening - impact on reviewer-relevant outcomes. BMC Med Res Methodol. 2020;20(1):256.
60. Harrison H, Griffin SJ, Kuhn I, Usher-Smith JA. Software tools to support title and abstract screening for systematic reviews in healthcare: An evaluation. BMC Med Res Methodol. 2020;20(1):7.
61. Hartling L, Featherstone R, Nuspl M, Shave K, Dryden DM, Vandermeer B. Grey literature in systematic reviews: a cross-sectional study of the contribution of non-English reports, unpublished studies and dissertations to the results of meta-analyses in child-relevant reviews. BMC Med Res Methodol. 2017;17(1):64.
62. Hemens BJ, Haynes RB. McMaster Premium LiteratUre Service (PLUS) performed well for identifying new studies for updated Cochrane reviews. J Clin Epidemiol. 2012;65(1):62-72.e61.
63. Hirt J, Meichlinger J, Schumacher P, Mueller G. Agreement in Risk of Bias Assessment Between RobotReviewer and Human Reviewers: An Evaluation Study on Randomised Controlled Trials in Nursing-Related Cochrane Reviews. J Nurs Scholarsh. 2021;53(2):246–54.
64. Howard BE, Phillips J, Tandon A, Maharana A, Elmore R, Mav D, Sedykh A, Thayer K, Merrick BA, Walker V et al. SWIFT-Active Screener: Accelerated document screening through active learning and integrated recall estimation. Environ Int. 2020;138:105623.
65. Hugues A, Di Marco J, Bonan I, Rode G, Cucherat M, Gueyffier F. Publication language and the estimate of treatment effects of physical therapy on balance and postural control after stroke in meta-analyses of randomised controlled trials. PLoS ONE. 2020;15(3):e0229822.
66. Janssens A, Gwinn M, Brockman JE, Powell K, Goodman M. Novel citation-based search method for scientific literature: a validation study. BMC Med Res Methodol. 2020;20(1):25.
67. Janssens AC, Gwinn M. Novel citation-based search method for scientific literature: application to meta-analyses. BMC Med Res Methodol. 2015;15:84.
68. Jap J, Saldanha IJ, Smith BT, Lau J, Schmid CH, Li T. Investigators obotDAA: Features and functioning of Data Abstraction Assistant, a software application for data abstraction during systematic reviews. Research Synthesis Methods. 2019;10(1):2–14.
69. Jardim PSJ, Rose CJ, Ames HM, Echavez JFM, Van de Velde S, Muller AE. Automating risk of bias assessment in systematic reviews: a real-time mixed methods comparison of human researchers to a machine learning system. BMC Med Res Methodol. 2022;22(1):167.
70. Jelicic Kadic A, Vucic K, Dosenovic S, Sapunar D, Puljak L. Extracting data from figures with software was faster, with higher interrater reliability than manual extraction. J Clin Epidemiol. 2016;74:119–23.
71. Kiritchenko S, de Bruijn B, Carini S, Martin J, Sim I. ExaCT: automatic extraction of clinical trial characteristics from journal publications. BMC Med Inf Decis Mak. 2010;10:56.
72. Kwon Y, Lemieux M, McTavish J, Wathen N. Identifying and removing duplicate records from systematic review searches. Journal of the Medical Library Association : JMLA. 2015;103(4):184–8.
73. Lee E, Dobbins M, Decorby K, McRae L, Tirilis D, Husson H. An optimal search filter for retrieving systematic reviews and meta-analyses. BMC Med Res Methodol. 2012;12:51.
74. Li J, Kabouji J, Bouhadoun S, Tanveer S, Filion KB, Gore G, Josephson CB, Kwon CS, Jette N, Bauer PR, et al. Sensitivity and specificity of alternative screening methods for systematic reviews using text mining tools. J Clin Epidemiol. 2023;162:72–80.
75. Li T, Saldanha IJ, Jap J, Smith BT, Canner J, Hutfless SM, Branch V, Carini S, Chan W, de Bruijn B, et al. A randomized trial provided new evidence on the accuracy and efficiency of traditional vs. electronically annotated abstraction approaches in systematic reviews. J Clin Epidemiol. 2019;115:77–89.
76. Marshall IJ, Kuiper J, Wallace BC. RobotReviewer: evaluation of a system for automatically assessing bias in clinical trials. J Am Med Inform Assoc. 2016;23(1):193–201.
77. Marshall IJ, Marshall R, Wallace BC, Brassey J, Thomas J. Rapid reviews may produce different results to systematic reviews: a meta-epidemiological study. J Clin Epidemiol. 2019;109:30–41.
78. Marshall IJ, Trikalinos TA, Soboczenski F, Yun HS, Kell G, Marshall R, Wallace BC. In a pilot study, automated real-time systematic review updates were feasible, accurate, and work-saving. J Clin Epidemiol. 2023;153:26–33.
79. Martyn-St James M, Cooper K, Kaltenthaler E. Methods for a rapid systematic review and metaanalysis in evaluating selective serotonin reuptake inhibitors for premature ejaculation. Evidence & Policy: A Journal of Research, Debate and Practice. 2017;13(3):517–38.
80. Mateen FJ, Oh J, Tergas AI, Bhayani NH, Kamdar BB. Titles versus titles and abstracts for initial screening of articles for systematic reviews. Clin Epidemiol. 2013;5:89–95.
81. McKeown S, Mir ZM. Considerations for conducting systematic reviews: evaluating the performance of different methods for de-duplicating references. Syst. 2021;10(1):38.
82. Moher D, Klassen TP, Schulz KF, Berlin JA, Jadad AR, Liberati A. What contributions do languages other than English make on the results of meta-analyses? J Clin Epidemiol. 2000;53(9):964–72.
83. Mortensen ML, Adam GP, Trikalinos TA, Kraska T, Wallace BC. An exploration of crowdsourcing citation screening for systematic reviews. Res. 2017;8(3):366–86.
84. Muthu S. The efficiency of machine learning-assisted platform for article screening in systematic reviews in orthopaedics. Int Orthop. 2023;47(2):551–6.
85. Nama N, Iliriani K, Xia MY, Chen BP, Zhou LL, Pojsupap S, Kappel C, O'Hearn K, Sampson M, Menon K, et al. A pilot validation study of crowdsourcing systematic reviews: update of a searchable database of pediatric clinical trials of high-dose vitamin D. Transl. 2017;6(1):18–26.
86. Nama N, Sampson M, Barrowman N, Sandarage R, Menon K, Macartney G, Murto K, Vaccani JP, Katz S, Zemek R, et al. Crowdsourcing the Citation Screening Process for Systematic Reviews: Validation Study. J Med Internet Res. 2019;21(4):e12953.
87. Ng L, Pitt V, Huckvale K, Clavisi O, Turner T, Gruen R, Elliott JH. Title and Abstract Screening and Evaluation in Systematic Reviews (TASER): a pilot randomised controlled trial of title and abstract screening by medical students. Syst Rev. 2014;3:121.
88. Noel-Storr A, Dooley G, Affengruber L, Gartlehner G. Citation screening using crowdsourcing and machine learning produced accurate results: evaluation of Cochrane's modified Screen4Me service. J Clin Epidemiol. 2020;29:29.
89. Noel-Storr A, Dooley G, Elliott J, Steele E, Shemilt I, Mavergames C, Wisniewski S, McDonald S, Murano M, Glanville J, et al. An evaluation of Cochrane Crowd found that crowdsourcing produced accurate results in identifying randomized trials. J Clin Epidemiol. 2021;133:130–9.
90. Noel-Storr A, Gartlehner G, Dooley G, Persad E, Nussbaumer-Streit B. Crowdsourcing the identification of studies for COVID-19-related Cochrane Rapid Reviews. Res Synth Methods. 2022;13(5):585–94.
91. Noel-Storr AH, Redmond P, Lamé G, Liberati E, Kelly S, Miller L, Dooley G, Paterson A, Burt J. Crowdsourcing citation-screening in a mixed-studies systematic review: a feasibility study. BMC Med Res Methodol. 2021;21(1):88.
92. Nussbaumer-Streit B, Klerings I, Dobrescu AI, Persad E, Stevens A, Garritty C, Kamel C, Affengruber L, King VJ, Gartlehner G. Excluding

non-English publications from evidence-syntheses did not change conclusions: a meta-epidemiological study. J Clin Epidemiol. 2020;118:42–54.
93. Nussbaumer-Streit B, Klerings I, Wagner G, Heise TL, Dobrescu AI, Armijo-Olivo S, Stratil JM, Persad E, Lhachimi SK, Van Noord MG, et al. Abbreviated literature searches were viable alternatives to comprehensive searches: a meta-epidemiological study. J Clin Epidemiol. 2018;102:1–11.
94. O'Keefe H, Rankin J, Wallace SA, Beyer F. Investigation of text-mining methodologies to aid the construction of search strategies in systematic reviews of diagnostic test accuracy-a case study. Res. 2023;14(1):79–98.
95. Olofsson H, Brolund A, Hellberg C, Silverstein R, Stenstrom K, Osterberg M, Dagerhamn J. Can abstract screening workload be reduced using text mining? User experiences of the tool Rayyan. Res. 2017;8(3):275–80.
96. Oude Wolcherink MJ, Pouwels X, van Dijk SHB, Doggen CJM, Koffijberg H. Can artificial intelligence separate the wheat from the chaff in systematic reviews of health economic articles? Expert Rev Pharmacoecon Outcomes Res. 2023;23(9):1049–56.
97. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst. 2016;5(1):210.
98. Pallath A, Zhang Q. Paperfetcher: A tool to automate handsearching and citation searching for systematic reviews. Res. 2022;19:19.
99. Paynter RA, Featherstone R, Stoeger E, Fiordalisi C, Voisin C, Adam GP. A prospective comparison of evidence synthesis search strategies developed with and without text-mining tools. J Clin Epidemiol. 2021;139:350–60.
100. Pham B, Klassen TP, Lawson ML, Moher D. Language of publication restrictions in systematic reviews gave different results depending on whether the intervention was conventional or complementary. J Clin Epidemiol. 2005;58(8):769–76.
101. Pham MT, Waddell L, Rajic A, Sargeant JM, Papadopoulos A, McEwen SA. Implications of applying methodological shortcuts to expedite systematic reviews: three case studies using systematic reviews from agri-food public health. Res. 2016;7(4):433–46.
102. Pianta MJ, Makrai E, Verspoor KM, Cohn TA, Downie LE. Crowdsourcing critical appraisal of research evidence (CrowdCARE) was found to be a valid approach to assessing clinical research quality. J Clin Epidemiol. 2018;104:8–14.
103. Pijls BG. Machine Learning assisted systematic reviewing in orthopaedics. J Orthop. 2024;48:103–6.
104. Pradhan R, Hoaglin DC, Cornell M, Liu W, Wang V, Yu H. Automatic extraction of quantitative data from ClinicalTrials.gov to conduct meta-analyses. J Clin Epidemiol. 2019;105:92–100.
105. Preston L, Carroll C, Gardois P, Paisley S, Kaltenthaler E. Improving search efficiency for systematic reviews of diagnostic test accuracy: An exploratory study to assess the viability of limiting to MEDLINE, EMBASE and reference checking. Syst Rev. 2015;4(1).
106. Przybyla P, Brockmeier AJ, Kontonatsios G, Le Pogam MA, McNaught J, von Elm E, Nolan K, Ananiadou S. Prioritising references for systematic reviews with RobotAnalyst: A user study. Res. 2018;9(3):470–88.
107. Rathbone J, Albarqouni L, Bakhit M, Beller E, Byambasuren O, Hoffmann T, Scott AM, Glasziou P. Expediting citation screening using PICo-based title-only screening for identifying studies in scoping searches and rapid reviews. Syst. 2017;6(1):233.
108. Rathbone J, Hoffmann T, Glasziou P. Faster title and abstract screening? Evaluating Abstrackr, a semi-automated online screening program for systematic reviewers. Syst. 2015;4:80.
109. Reddy SM, Patel S, Weyrich M, Fenton J, Viswanathan M. Comparison of a traditional systematic review approach with review-of-reviews and semi-automation as strategies to update the evidence. Syst. 2020;9(1):243.
110. Rice M, Ali MU, Fitzpatrick-Lewis D, Kenny M, Raina P, Sherifali D. Testing the effectiveness of simplified search strategies for updating systematic reviews. J Clin Epidemiol. 2017;88:148–53.
111. Royle P, Waugh N. A simplified search strategy for identifying randomised controlled trials for systematic reviews of health care interventions: a comparison with more exhaustive strategies. BMC Med Res Methodol. 2005;5:23.
112. 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.
113. Schopow N, Osterhoff G, Baur D. Applications of the Natural Language Processing Tool ChatGPT in Clinical Practice: Comparative Study and Augmented Systematic Review. JMIR Med Inform. 2023;11:e48933.
114. Shemilt I, Khan N, Park S, Thomas J. Use of cost-effectiveness analysis to compare the efficiency of study identification methods in systematic reviews. Syst. 2016;5(1):140.
115. Stoll CRT, Izadi S, Fowler S, Green P, Suls J, Colditz GA. The value of a second reviewer for study selection in systematic reviews. Res. 2019;10(4):539–45.
116. Šuster S, Baldwin T, Verspoor K. Analysis of predictive performance and reliability of classifiers for quality assessment of medical evidence revealed important variation by medical area. J Clin Epidemiol. 2023;159:58–69.
117. Taylor-Phillips S, Geppert J, Stinton C, Freeman K, Johnson S, Fraser H, Sutcliffe P, Clarke A. Comparison of a full systematic review versus rapid review approaches to assess a newborn screening test for tyrosinemia type 1. Res. 2017;8(4):475–84.
118. Thomas J, McDonald S, Noel-Storr A, Shemilt I, Elliott J, Mavergames C, Marshall IJ. Machine learning reduced workload with minimal risk of missing studies: development and evaluation of a randomized controlled trial classifier for Cochrane Reviews. J Clin Epidemiol. 2021;133:140–51.
119. Tsou AY, Treadwell JR, Erinoff E, Schoelles K. Machine learning for screening prioritization in systematic reviews: comparative performance of Abstrackr and EPPI-Reviewer. Syst. 2020;9(1):73.
120. Valizadeh A, Moassefi M, Nakhostin-Ansari A, Hosseini Asl SH, Saghab Torbati M, Aghajani R, Maleki Ghorbani Z, Faghani S. Abstract screening using the automated tool Rayyan: results of effectiveness in three diagnostic test accuracy systematic reviews. BMC Med Res Methodol. 2022;22(1):160.
121. van de Schoot R, de Bruin J, Schram R, Zahedi P, de Boer J, Weijdema F, Kramer B, Huijts M, Hoogerwerf M, Ferdinands G, et al. An open source machine learning framework for efficient and transparent systematic reviews. Nature Machine Intelligence. 2021;3(2):125–33.
122. Van Enst WA, Scholten RJPM, Whiting P, Zwinderman AH, Hooft L. Meta-epidemiologic analysis indicates that MEDLINE searches are sufficient for diagnostic test accuracy systematic reviews. J Clin Epidemiol. 2014;67(11):1192–9.
123. Waffenschmidt S, Guddat C. Searches for randomized controlled trials of drugs in MEDLINE and EMBASE using only generic drug names compared with searches applied in current practice in systematic reviews. Res. 2015;6(2):188–94.
124. Waffenschmidt S, Knelangen M, Sieben W, Bühn S, Pieper D. Single screening versus conventional double screening for study selection in systematic reviews: a methodological systematic review. BMC Med Res Methodol. 2019;19(1):132.
125. Walker VR, Rooney AA. CEC02-02 Automated and Semi-Automated Approaches for Literature Searching, Screening, and Data Extraction for Systematic Reviews in Environmental Health. Toxicol Lett. 2021;350(Supplement):S4.
126. Wang Z, Asi N, Elraiyah TA, Abu Dabrh AM, Undavalli C, Glasziou P, Montori V, Murad MH. Dual computer monitors to increase efficiency of conducting systematic reviews. J Clin Epidemiol. 2014;67(12):1353–7.
127. Xu C, Ju K, Lin L, Jia P, Kwong JSW, Syed A, Furuya-Kanamori L. Rapid evidence synthesis approach for limits on the search date: How rapid could it be? Res. 2022;13(1):68–76.
128. Waffenschmidt S, Sieben W, Jakubeit T, Knelangen M, Overesch I, Bühn S, Pieper D, Skoetz N, Hausner E. Increasing the efficiency of study selection for systematic reviews using prioritization tools and a single-screening approach. Syst Rev. 2023;12(1):161.
129. Affengruber L, Nussbaumer-Streit B, Hamel C, Maten MVd, Thomas J, Mavergames C, Spijker R, Gartlehner G.Rapid review methods series: Guidance on the use of supportive software. BMJ Evidence-Based Medicine. 2024;29(4):264–71.
130. Haddaway NR, Page MJ, Pritchard CC, McGuinness LA. PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant

flow diagrams, with interactivity for optimised digital transparency and Open Synthesis. Campbell Syst Rev. 2022;18(2):e1230.
131. van Altena AJ, Spijker R, Olabarriaga SD. Usage of automation tools in systematic reviews. Research Synthesis Methods. 2019;10(1):72–82.
132. Egger M. Ji¿½ni P, Bartlett C, Holenstein F, Sterne J: How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical Study. 2003;7(1):1–76.
133. Marshall IJ, Wallace BC. Toward systematic review automation: a practical guide to using machine learning tools in research synthesis. Syst Rev. 2019;8(1):163.
134. Wagner G, Nussbaumer-Streit B, Greimel J, Ciapponi A, Gartlehner G. Trading certainty for speed - how much uncertainty are decisionmakers and guideline developers willing to accept when using rapid reviews: an international survey. BMC Med Res Methodol. 2017;17:121.

Publisher's Note