Rapid reviews: defining, evaluating methods, and reducing screening burden using artificial intelligence

Hamel, Candyce

Doctoral thesis / Disertacija

2021

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Split, School of Medicine / Sveučilište u Splitu, Medicinski fakultet**

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:171:149967

Rights / Prava: In copyright / Zaštićeno autorskim pravom.

Download date / Datum preuzimanja: **2024-03-29**

SVEUČILIŠTE U SPLITU
MEDICINSKI FAKULTET
UNIVERSITAS STUDIORUM SPALATENSIS
FACULTAS MEDICA

Repository / Repozitorij:

MEFST Repository

UNIVERSITY OF SPLIT

dabar
DIGITALNI AKADEMSKI ARHIVI I REPOZITORIJI

UNIVERSITY OF SPLIT
SCHOOL OF MEDICINE

CANDYCE HAMEL

RAPID REVIEWS: DEFINING, EVALUATING METHODS, AND REDUCING SCREENING BURDEN USING ARTIFICIAL INTELLIGENCE

DOCTORAL DISSERTATION

Split, Croatia, July 2021

University of Split, School of Medicine
Mentor: Beverley Shea, PhD
Supervisor: Adrienne Stevens, PhD

DEDICATION

There are so many people who contributed toward this PhD work, either by actively taking part in the research, by supporting me while I spent time conducting this research, or in other valuable contributory ways.

I cannot write in words my appreciation for Beverley Shea. She has contributed towards my life in so many ways, including my career, education, and personal growth. I have known Bev for over 25 years and she has been a great boss, mentor, and most of all a friend.

I have had the opportunity to work with amazing people, including Adrienne Stevens. We have had many thought-provoking conversations over the years and she has been a huge support and resource in my learning of systematic reviews, meta-analysis, and epidemiology in general. If Adrienne did not know the answer, she could dig up a paper with the answer within minutes! I would also like to thank Chantelle Garritty for introducing me to the world of rapid reviews, for her contributions to my PhD publications, and most of all, for her never-ending support. Thank you to my boss, Brian Hutton, who was very supportive of me pursuing this journey and who has always given me the flexibility in selecting my projects. Thank you Alan Michaud and Micere Thuku, who were huge contributors to the scoping reviews. Thank you to all of those who contributed to these three publications (presented alphabetically): Lisa Affengruber, Shannon Kelly, Danielle Rice, Barbara Nussbaumer-Streit, Becky Skidmore, Kednapa Thavorn, and George Wells.

I would like to acknowledge and thank my thesis committee members, Professors Ljubo Znaor (Chair), Lorainne Car Tudor, and Tina Poklepović Peričić for their time and thoughtful questions. Thank you to Dr. Livia Puljak who has offered her time and advice throughout the entire PhD process, and has made it easy to progress through the PhD program from Canada. She has provided her ongoing support and has had to answer far too many emails from me.

This PhD would not have been possible without the support of my family. I would like to thank the best husband anyone could ask for, Dana, who has always been supportive in all of my personal goals and self-imposed challenges. He is always my sounding board or just a pair of ears to listen to my thoughts and ideas. Thanks to the sweetest kids, Rylan, Connor, and Kaylee,

who have never complained once about my time away from them as I worked on these projects and for all of the hugs that kept me going. They have given me the determination to pursue my professional goals with the view of demonstrating to them that all of their future goals are also obtainable. My appreciation also goes out to my parents, who have instilled in me the value of education, and for their encouragement and support through my PhD studies. Lastly, a special thanks to my sister, Keri, who is always willing to proof-read my documents, even when she has no idea what they are about.

1. Table of Contents

1. Table of Contents 1
2. Abbreviations 2
3. Introduction 3
4. Aims of the Compiled Research Papers 6
5. Scientific Contribution of the Compiled Research Papers 7
5.1 Overview of the methodology of the joint papers 7
Publication 1: Systematic scoping review of RR definitions 7
Publication 2: Systematic scoping review of RR methods 8
Publication 3: Active machine-learning prioritization tool 9
5.2 Summary of the Results of the Joined Works 11
Publication 1: Systematic scoping review of RR definitions 11
Publication 2: Systematic scoping review of RR methods 14
Publication 3: Active machine-learning prioritization tool 16
5.3 Discussion 18
Scientific contribution 20
Future Research 21
5.4 Conclusions 22
6. Sažetak 23
7. Abstract 25
8. References 27
9. Curriculum Vitae: Candyce Hamel 31
10. Papers Combined in the Dissertation 37
Appendix 1. AI Ranking Simulation output 74

2. Abbreviations

AI	Artificial intelligence
AML	Active machine learning
CI	Confidence Interval
ID	Identification
MECIR	Methodological Expectations of Cochrane Intervention Reviews
PRISMA-ScR	Preferred Items in Systematic Reviews and Meta-Analysis extension for Scoping Reviews
RR	Rapid Review
SD	Standard deviation
SR	Systematic Review

3. Introduction

Systematic reviews (SRs) are considered the gold standard in collating available evidence related to a specific question. SRs use systematic and rigorous methods with the goal to identify all relevant research to answer a research question [1]. SRs have been used to inform policy for health care and public health since the early 1990's [2] and are considered to be essential to produce trustworthy guidelines [3]. However, they are time- and resource-intensive undertakings. An analysis of 197 reviews registered in PROSPERO reported that SRs take an average of 67.3 weeks to conduct (from registration to publication), with a range of six to 186 weeks [4]. Additionally, the team required to produce a SR may be large (mean author team size: 5 [standard deviation (SD): 3; range: 1 to 27]) [4], and should include, at a minimum, a systematic review methodologist, a clinical expert, and a statistician. Screening of the title and abstract records of possibly relevant studies is a particularly time-intensive step and it is not uncommon for a systematic search to yield a large number of records, many of which are irrelevant (i.e., low precision). A recent study by Wang *et al.* (2020) evaluated 25 SRs which included 139,467 citations (mean yield of 5579 records per review) which resulted in a final inclusion rate of 5.48% (95% confidence interval [CI]: 2.38 to 8.58%) [5]. This large number of records also introduces the opportunity for human error in the screening process. The same study by Wang *et al*. reported a total error rate (i.e., false inclusion and false exclusion) of 10.76% (95% CI: 7.43% to 14.09%) [5].

As SRs are often conducted to answer policy-related, healthcare practice, public health, and urgent clinical practice questions, the length of time taken to produce a traditional SR may not meet the timeline when urgent answers are required. Thus, the emergence of rapid reviews (RRs), which are to produce evidence reviews in a timely manner while maintaining rigorous and robust methods.

The steps taken to conduct a RR are similar, or the same, as the steps taken to conduct a SR. So, what is the difference? Cochrane, a leading organization producing high-quality SRs, describes a SR as a review that "attempts to identify, appraise, and synthesize all the empirical evidence that meets pre-specified eligibility criteria to answer a specific research question" [6]. To date, the only consensus around a definition of a RR is that a formal definition does not exist [7–9]. In 2010, Ganann *et al.* defined RRs as "literature reviews that use methods to accelerate or streamline traditional systematic review processes" [10]. Further, Tricco *et al*.

(2015) described RRs as "a type of knowledge synthesis in which components of the systematic review process are simplified or omitted to produce information in a short period of time" [8].

Although there is no universally agreed upon definition of a RR, is it important to note that the *length of time* to conduct a review cannot be the defining feature of a RR, as adding more reviewers to the conduct of a SR may result in a timely report. Likewise, a review with few or no included studies may be conducted quickly, as there is little or no requirement for data extraction, risk of bias assessment, evaluating the certainty of the evidence, and writing the results section of the report. This has led to the suggestion that RRs should instead be called 'restricted systematic reviews' to focus on the restriction of the methods, rather than the speed of conduct [11].

Methodological investigations published by Tricco *et al.* in 2015 [8] and Haby *et al*. in 2016 [12] have highlighted that a variety of methods have been used to facilitate the evaluation of studies in a RR, including limiting the scope of a review or making abbreviations or omissions across the processes of conduct. In 2018, Robson *et al.* published a SR which identified the studies that examined methods for selecting studies, abstracting data, and appraising quality in SRs [13]. However, no comprehensive review of evaluations of RR methods *abbreviations, shortcuts, or omissions* has been undertaken to: (i) reflect the totality and the more recently emerging evaluations in this area, or (ii) to identify research gaps.

Due to the growing number of research papers being published in growing numbers of journals and databases, even well-constructed literature searches often result in several thousands of records to be screened. Title and abstract screening of these records is a time- and resource-intensive stage in the conduct of a review. It has been estimated that reviewers can screen, on average, two abstracts per minute, resulting in approximately 900 records in a 7.5-hour work day. However, this estimate is highly variable and can be dependent on factors such as the complexity of the topic [14] and the skill level of the reviewers. More realistically, in factoring in breaks, meetings, and a decrease in productivity over the day, this number is likely closer to 300-500 records per day. Several methods exist to decrease or optimize the time spent screening, with varying levels of success, including:

- The use of dual-monitors for screening [15];
- Crowdsourcing, which distributes tasks to workers (with varying levels of training) via the web [16,17];

- Using participants, intervention and comparator (PICo)-based title only screening [18] (e.g., screening first based on title only, then title and abstract on the remaining records [19]);
- Single-reviewer screening [20–22];
- Liberal accelerated screening [7], in which one reviewer is required to include the record and two reviewers are required to exclude the record;
- Machine-assisted abstract screening, where humans screen a portion of the titles and abstracts to create a training set and the machine screens the remaining records [23,24], and
- Machine-assisted abstract screening through active machine-learning, in which the automation tool learns from all previous responses and prioritizes the remaining records based on likelihood of inclusion [25,26].

A systematic review by O'Mara-Eves *et al.* (2015) looked at the approaches in text mining and concluded that there is almost no replication between studies or collaboration between research teams, which makes it difficult to establish any overall conclusions about best approaches [27].

One emerging method to conduct SRs and other review types (e.g., scoping review, RRs) is the use of artificial intelligence (AI). The interest in and development of AI tools, including active machine-learning (AML) algorithms, may be due to the large screening burden while conducting reviews. AML is an iterative process whereby the accuracy of the predictions made by the algorithm is improved through interaction with reviewers as they screen additional records [27]. Several SR software exists that support title and abstract screening [28], however, not all packages include AML. Among those that do, there is variation in the level of sophistication of the machine-learning tool, the algorithms used, the cost of the software, and if and how often the software is updated and/or supported. In many cases, a barrier to uptake of AI and AML is that researchers conducting evidence reviews do not know how to optimally use the AI and AML within these software packages. There may also be optimistic trust or cautious mistrust in AI that requires the additional evaluation of these tools before adoption by the SR community. While AI might not be ready to fully replace human screeners, several studies in this area suggest that optimizing, accelerating, and reducing screening burden through the use of AI and AML might be a viable option [23–27,29–32].

4. Aims of the Compiled Research Papers

The overarching goal of this doctoral research was to identify rapid review (RR) definitions, RR methods, and evaluate an abbreviated method that could be used in RRs. The overall objectives of the research were to:

1. Identify how RRs have been defined in both RRs and RR methods literature; to perform a thematic analysis to determine key themes in definitions; and to provide a suggested definition of a RR for further discussion within the review community.
2. Identify the methods literature pertaining to RRs with a specific focus on studies that formally evaluate the performance or impact of methods shortcuts when compared to other RR or SR methods; to map these methods to key stages of review conduct to determine research gaps; and to map these methods to the Methodological Expectations of Cochrane Intervention Reviews (MECIR) criteria.
3. To assess the performance of an AI-AML tool in a SR software (DistillerSR ©); to determine the reduction of screening burden; and to estimate the potential time savings.

To address these objectives, the following three scientific papers were published:

1. Hamel C, Michaud A, Thuku M, Skidmore B, Stevens A, Nussbaumer-Streit B, Garritty C. Defining Rapid Reviews: a systematic scoping review and thematic analysis of definitions and defining characteristics of rapid reviews. *J Clin Epi* 2021 Jan; 129:74-85. (JIF 4.952), ePub: 2020 October 8. doi: 10.1016/j.jclinepi.2020.09.041.

2. 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 Epi* 2020 Oct; 126:131-140. (JIF 4.952), ePub: 2020 Jun 26. doi: 10.1016/j.jclinepi.2020.06.027.

3. 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 Medical Research Methodology* 2020; 20: 256. (JIF 3.031), ePub: 2020 Jun. doi: 10.1186/s12874-020-01129-1.

5. Scientific Contribution of the Compiled Research Papers

5.1 Overview of the methodology of the joint papers

PUBLICATION 1: SYSTEMATIC SCOPING REVIEW OF RR DEFINITIONS

This systematic scoping review was conducted following guidance from the Joanna Briggs Institute [33] and reported according to the Preferred Items in Systematic Reviews and Meta-Analysis extension for Scoping Reviews (PRISMA-ScR) [34]. The protocol for this work was registered on the Open Science Framework (OSF: https://osf.io/y5f2m/) prior to undertaking the scoping review.

Detailed methods are provided in *Appendix A of publication* [35] with a brief description (extracted from the published article) provided in ***Table 1***. Additional study details (e.g., search strategy) can be found in the appendices of the publication (doi.org/10.1016/j.jclinepi.2020.09.041).

Table 1 - Systematic scoping review methods in brief (RR definitions)

Review Stage	Method description
Eligibility criteria	▪ Published rapid reviews using 'rapid' or derivative (e.g., abbreviated) in the title or abstract ▪ Published between January 2017 and January 2019 ▪ Written in English (for feasibility)
Searching for studies	▪ Developed by an experienced information specialist with input on search terms by members of the research team ▪ Peer reviewed using the PRESS checklist [36] ▪ Search (Dec 2018): MEDLINE® ALL, Embase Classic+Embase, PsycINFO, ERIC, Cochrane Library, CINAHL, Web of Science (*Appendix B of publication*) ▪ Search strategies not restricted by language ▪ Supplemented with definitions from rapid review methods scoping review [37]
Study selection	▪ Performed in DistillerSR [38] ▪ Piloted title/abstract (n=100) and full-text screening (n=25), conflicts resolved through discussion ▪ Liberal accelerated [7] screening for titles and abstracts ▪ Dual-independent screening based on full text, with conflicts resolved through discussion
Data charting	▪ Performed in DistillerSR [38] ▪ Piloted extractions (n=5), conflicts resolved through discussion

Review Stage	Method description
	▪ One reviewer extracted the definitions verbatim and the citations of the studies that were referenced, a second reviewer verified all extracted data, conflicts resolved through discussion
Data synthesis	▪ Rapid review details and citations referenced were exported to MS Excel 2016 for quantitative analysis ▪ Definitions (including definitions from the RR methods scoping review) were imported into NVivo (version 12) for coding into themes ▪ The thematic analysis allowed for the suggestion of a preliminary definition, with additional caveats, to allow further discussion within the review community.

PUBLICATION 2: SYSTEMATIC SCOPING REVIEW OF RR METHODS

This systematic scoping review was conducted following guidance from the Joanna Briggs Institute [33] and reported according to PRISMA-ScR [34]. The protocol for this work was registered on the Open Science Framework (OSF: https://osf.io/dekx6/) prior to undertaking the scoping review.

Detailed methods are provided in *Appendix A of the publication* [37] with a brief description provided in ***Table 2.*** Additional study details (e.g., search strategies) can be found in the appendices of the publication (doi.org/10.1016/j.jclinepi.2020.06.027).

Table 2 - Systematic scoping review methods in brief (RR methods)

Review Stage	Method description
Eligibility criteria	▪ Methods studies that evaluated shortcut approaches that could be applied or related to RR stages of conduct ▪ Written in English (for feasibility) ▪ Published or identified through grey literature since 1997
Searching for studies	▪ Developed by an experienced information specialist with input on search terms by members of the research team ▪ Focus on interventional RR methods ▪ Peer reviewed using the PRESS checklist [36] ▪ Original search (Jan 2019): MEDLINE® ALL, Embase Classic+Embase, PsycINFO, ERIC, Cochrane Library, CINAHL, Web of Science, Epistemonikos (*Appendix C.1 of publication*) ▪ Supplemental search (Feb 2019): MEDLINE® ALL, Embase Classic+Embase, PsycINFO and ERIC (*Appendix C.2 of publication*) ▪ Search strategies not restricted by language

Review Stage	Method description
	▪ Additional searching: grey literature (e.g., organizations that produce RRs), bibliographies of included studies, contacting experts in the field, bibliography of Robson 2018 study [13]
Study selection	▪ Performed in stages due to large yield of first search ▪ Performed in DistillerSR [38] ▪ Piloted title/abstract and full-text screening, conflicts resolved through discussion ▪ Liberal accelerated [7] screening for titles and abstracts ▪ Dual-independent screening based on full text, with conflicts resolved through discussion ▪ Artificial intelligence tool used to help screen titles and abstract ▪ Reported in a PRISMA flow diagram [39]
Data charting (*Appendix D of publication*)	▪ Piloted extractions (n=5), conflicts resolved through discussion ▪ One reviewer extracted studies, a second reviewer verified all extracted data, conflicts resolved through discussion
Data synthesis	▪ Formal evaluative studies: - Two reviewers mapped the studies into four categories highlighting the focus or intent of the papers (partially informed by Tricco et al 2015 [8], and further adapted through discussion) - Studies that formally evaluated shortcut methods used in the RR context were mapped back to the stage of conducts to identify gaps, and are presented narratively with details provided in tables - Each shortcut was compared to the MECIR guidelines [40] for Cochrane reviews to see whether it met the MECIR criteria ▪ Other categories are narratively described with details provided in tables

PUBLICATION 3: ACTIVE MACHINE-LEARNING PRIORITIZATION TOOL

In the spring of 2020, Evidence Partners released a new version of the AI toolkit in their online SR software application, called DistillerSR©. As part of the AI toolkit, an AI simulation tool was included, which allows a retrospective evaluation of how the AML would have worked had prioritized screening been used during screening of titles and abstracts. A primary experimental design was used to test the accuracy of the AML in DistillerSR using the AI simulation tool comprised the primary study for this thesis. This was done using a convenience sample of 10 completed SRs. The protocol for this work was registered on the Open Science Framework (OSF: https://osf.io/2fgz7/) prior to undertaking this work.

The unit of analysis for this study was the unique record (i.e., the title and abstract of the primary study) being assessed for each of the included SRs. The AI simulation tool was run 10 times on each SR to account for any variation in the simulations and to introduce randomness (through shuffling the references, which is performed automatically by the simulation tool) into the initial training sets (i.e., a set of responses which inform the AML). ***Figure 1*** provides a pictorial representation of how the AI simulation tool uses the existing information (i.e., the include or exclude response) to simulate the process of screening by humans using the prioritization tool.

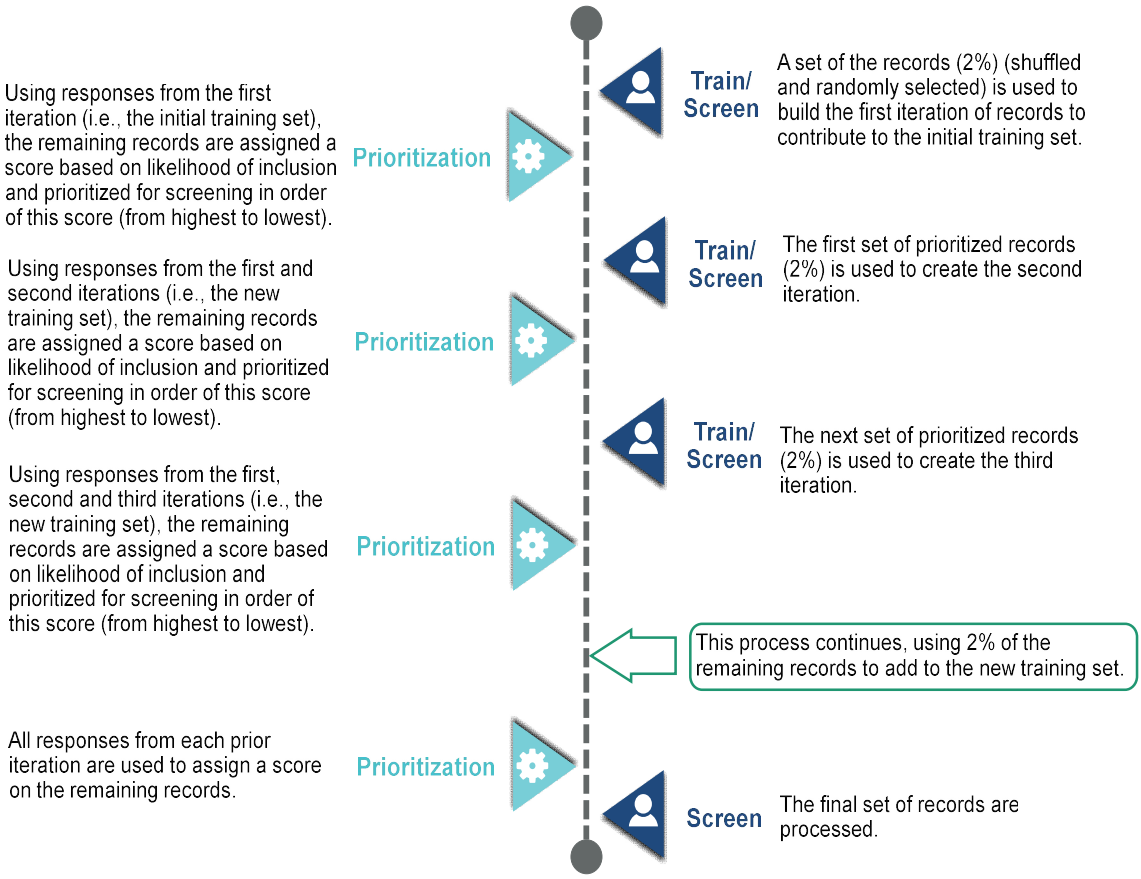


Figure 1 - AI simulation process

After each training set (i.e., 2% of the records in the database, with a minimum of 25 records and a maximum of 200 records), the AML is activated and records are assigned a score (by the software) relating to the likelihood of inclusion. References are re-ranked (i.e., prioritized) in order of this score from most likely to least likely to be relevant, and screening continues.

To evaluate the reduction in the screening burden and performance of AML using a true recall of 95% (i.e., stop screening once 95% of the studies included at title and abstract level are identified), the following information was collected:

- the total number of records screened to achieve a true recall @ 95%;
- the number of records screened that were excluded once a true recall @ 95% was achieved; and
- the reference identification (ID) numbers of the 5% of the records that were not yet identified as included records (i.e., false negatives).

Appendix 1 provides an example of the output produced by the AI Ranking Simulation tool in DistillerSR.

Means (standard deviations) and medians (ranges) were calculated to evaluate the reduction in the screening burden. This information was also used to calculate the time saved by not having to screen the least relevant records. To determine performance, the reference IDs of the false negative studies were used to determine if any were for a citation that were included in the completed SR.

5.2 Summary of the Results of the Joined Works

PUBLICATION 1: SYSTEMATIC SCOPING REVIEW OF RR DEFINITIONS

The search strategies to identify RRs resulted in 2,657 unique records, of which 422 were evaluated at full text. Two Hundred and sixteen RRs published between 2017 and January of 2019 were identified. Most of the RRs (82.5%) were from corresponding authors from the UK, Australia, the USA, and Canada. In total, 158 (73%) RRs provided a definition. Among all RRs a median of two references (range 0 to 7) were cited. Among the 90 RR method papers, 81% (73/90) provided a definition.

Terminology

For feasibility, several terms which may be used to describe a RR were not included at the title and abstract phase (*Appendix C of publication*). Among the 216 RRs that were included, 'Rapid Review' was the most often term used (n=136, 63.0%). The terms included at the title and abstract phase of this scoping review are presented in ***Table 3.***

Table 3 - Terminology used to describe the review

Terminology used (as first mentioned in the RR)	n (%) (N=216)
Rapid review	136 (63.0%)
Rapid evidence assessment	22 (10.1%)
Rapid systematic review	19 (8.8%)
Rapid evidence review; Rapid literature review	12 (5.6%) (each)
Systematic rapid evidence assessment; Systematic rapid review	2 (0.9%) (each)
Abbreviated review; Rapid appraisal; Rapid best-fit framework synthesis; Rapid-evidence based review; Rapid evidence summary; Rapid evidence synthesis; Rapid meta-review; Rapid qualitative review; Rapid response review; Rapid structured evidence review; Rapid synthesis	1 (0.5%) (each)

Thematic analysis of definitions

A total of 204 RRs and RR methods papers provided a definition that could be thematically analyzed (75 did not provide a definition and 27 RRs cited other studies with no identifiable themes). There were 79 unique citations showing the variability in definitions that are currently being cited. After a thematic analysis was performed, eight major themes in the definition were identified (***Figure 2***).

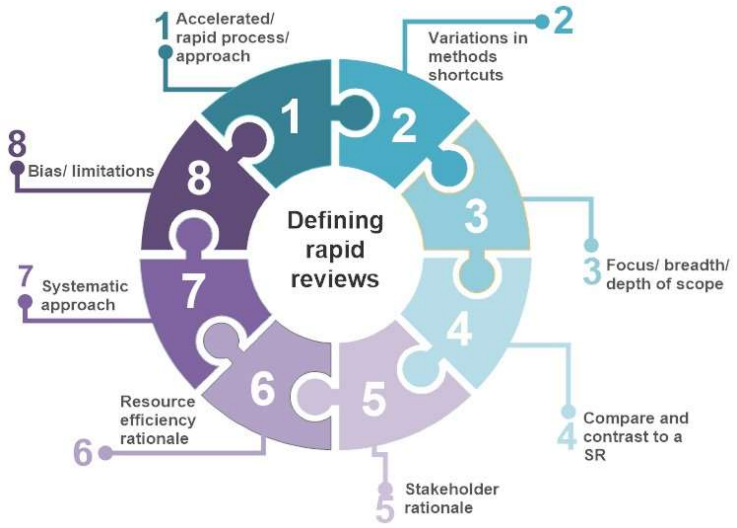


Figure 2 - Eight key themes in defining RRs

Among the reported definitions, the most common themes were Theme 4: Compare and contrast to SRs (68.1%; 139/204), Theme 2: Variation in shortcut methods (54.9%; 112/204), with Theme 1: Accelerated/rapid process and Theme 6: Resource efficiency rationale tied (48.5%; 99/204 each) (***Figure 3***). Definitions often covered more than one of these themes, with a range of 1 to 8 (median: 3; mean: 3).

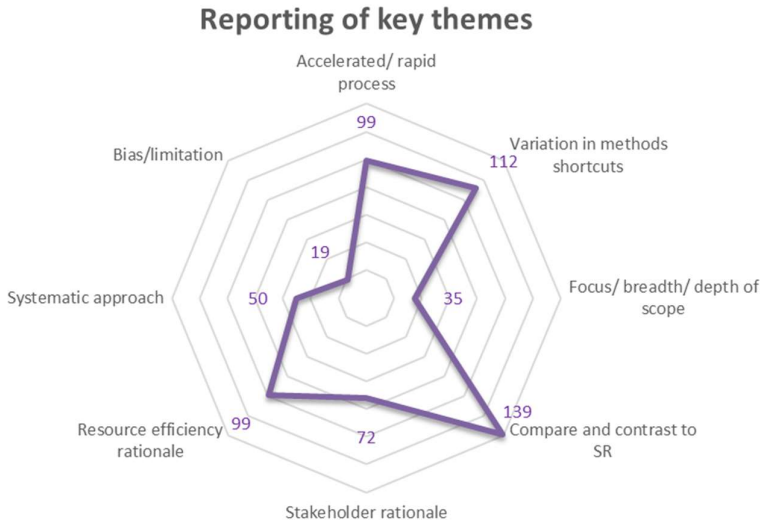


Figure 3 - Frequency of reporting of key themes

[redacted]

As there may be different requirements from stakeholders, funders and/or knowledge users of RR, there may not be one common definition for a RR. As such, we suggest the following broad definition, which meets a minimum set of requirements identified in the thematic analysis.

> "A rapid review is a form of knowledge synthesis that accelerates the process of conducting a traditional systematic review through streamlining or omitting a variety of methods to produce evidence in a resource-efficient manner."

This definition covers the most common themes (i.e., 1, 2, 4 and 6) that were identified in approximately 50% or more of the RRs and methods papers. By using broad words like resources, this definition captures the time element as well as cost and human elements. Users could then tailor this definition accordingly to best meet their individual remit and mandates for producing RRs by adding additional details covered in other themes. For example, if an organization produces RRs only when stakeholders make a request (Theme 5), it can be modified to include this requirement.

> "A rapid review is a form of knowledge synthesis that accelerates the process of conducting a traditional systematic review through streamlining or omitting a variety of methods to produce evidence for stakeholders in a resource-efficient manner."

PUBLICATION 2: SYSTEMATIC SCOPING REVIEW OF RR METHODS

The search strategies to identify studies evaluating RR methods, plus the results from grey literature searching, resulted in 1,873 unique records, of which 156 were evaluated at full text, and 90 studies were included. The majority of the studies were conducted in Canada, the UK, and Australia (66/90, 73.3%), and were published in 2014 or later (68/90, 75.6%). The majority of the formal evaluation studies have been published since 2017 (11/14, 78.6%).

Categorizing RR studies

Although the primary objective of the scoping review was to identify studies that evaluated abbreviated, shortcut, or omitted methods in RRs, to build a comprehensive repository, we also identified studies that described RR methods. Using guidance from Tricco 2015 [8] and further guided by discussions among the review group, the studies were divided into four main RR categories (***Figure 4***), with an addition six studies identified as SR surrogates in which the methods were evaluated in SRs, but could equally be applied while conducting RRs.

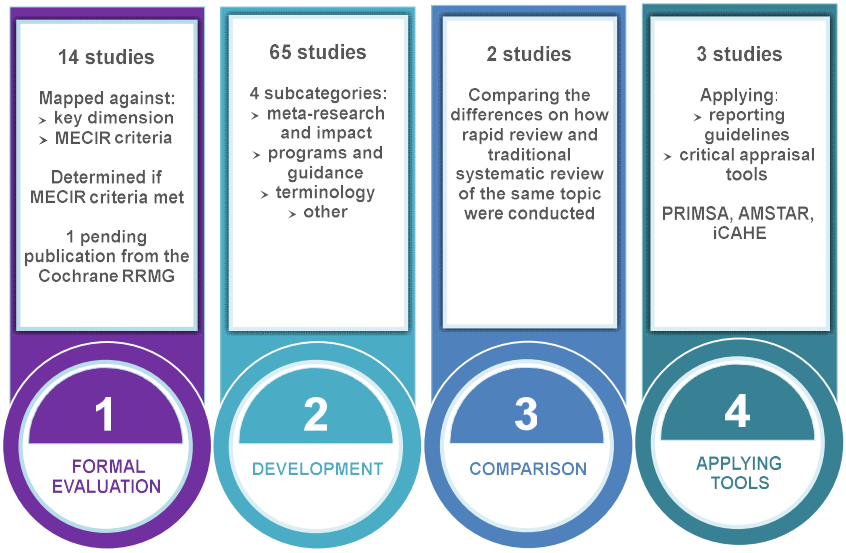


Figure 4 - RR study categories

Mapping to key dimensions

The 14 studies of formal evaluations addressed nine key dimensions related to the conduct phases of a review (*Appendix B of publication*), or "other" areas not included in this preliminary list of key dimensions (***Figure 5***).

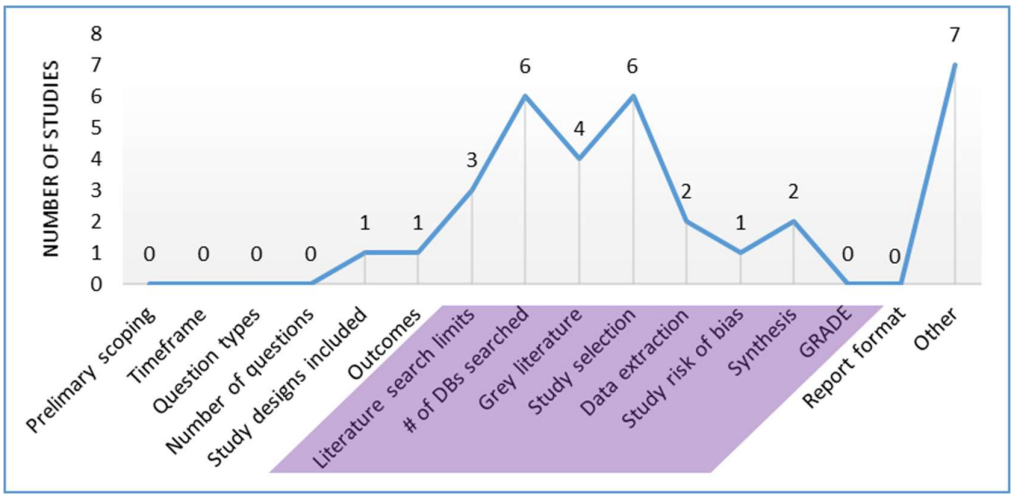


Figure 5 - Mapping to key dimensions of the review process (all evaluative studies)

Some studies evaluated more than one shortcut method, therefore, a study could have contributed to one or more key dimensions. Evaluations included:

- Assessing the impact of shortcuts within the conduct of a RR (e.g., title only screening, including only English language publications)
- Comparing different versions of the same shortcuts within the conduct of a RR (e.g., number of databases searched)
- Comparing the results/conclusions of RRs to those of SRs (e.g., including only the largest trial), or
- Evaluating the impact of including 'best-practice' methods (e.g., including stakeholders in the review process, peer-review of search strategy).

Four studies were labelled as 'composite evaluations' in which more than one methodological shortcut was taken simultaneously. Any differences in the results may be attributable to one or several of the shortcuts.

[redacted]

Only a cursory mapping to MECIR criteria was possible, as insufficient information impeded the ability to determine if criteria were met. Additionally, some of the items could not be mapped to MECIR criteria, as some are not methods performed in a traditional SR (e.g., using existing risk of bias information from a SR and performing new assessments for any studies not found in SRs), or are not currently found in MECIR (e.g., peer-reviewing the search strategy).

PUBLICATION 3: ACTIVE MACHINE-LEARNING PRIORITIZATION TOOL

Ten SRs, consisting of 69,663 records, were used in this experiment. Reviews ranged in size from 2,250 to 22,309 records to be assessed at title and abstract level, of which, based on the title and abstract, 3.0% to 39.2% (median: 16.2%) were included to be further reviewed at full text. A median of 0.6% (range 0.02 to 1.48%) of the total number of records were included in the final SRs.

This experimental study included terminology used in the areas of computer science and diagnostic test accuracy studies. To help the reader, a table with a description of the terminologies was provided in the published article as Table 1 [41].

[redacted]

Across the set of 10 SRs evaluated, the median percentage of studies required to be screened to achieve a true recall @ 95% was 47.1% (Inter Quartile Range: 37.5 to 58.0%) (***Figure 6***). The number of records that did not need to be screened (light blue portion of the bar) ranged from 30% (smoking cessation) to 72.5% (opioid use disorder). Typically, SRs with fewer studies included at the title and abstract level for further eligibility assessment based on the full-text article (dark blue portion of the bar) resulted in a larger reduction in the overall screening burden.

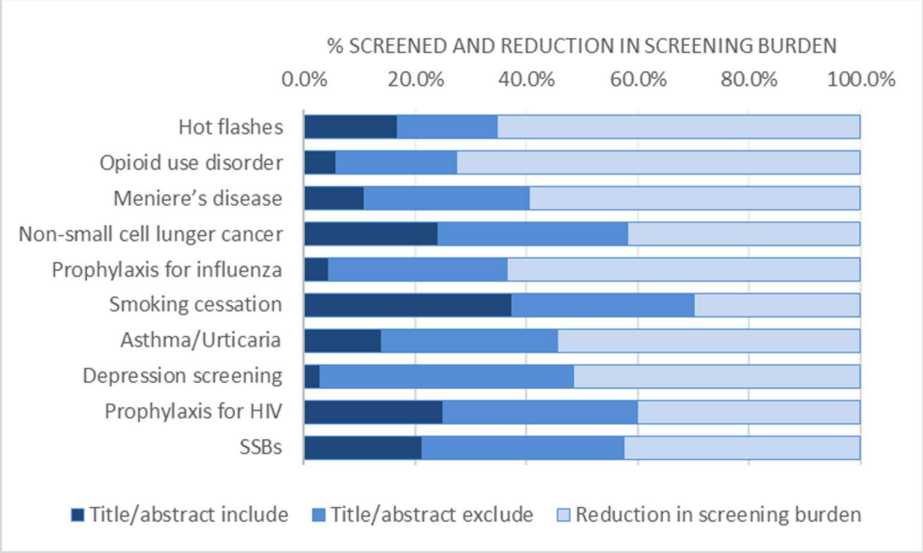


Figure 6 - Title and abstract includes, excludes and screening burden reduction

Performance of AI AML

Among the 100 iterations (10 iterations in 10 SRs), all final included studies had been identified at a true recall @ 95%. In other words, none of the last 5% of those originally included at title and abstract level (i.e., false negatives) were included in the final review.

Amount of time saved

Overall, the mean title and abstract screening hours saved when using the true recall @ 95% modified screening approach (i.e., AI would exclude all remaining references and one human reviewer would be required to screen the remaining records) was 62.8 hours (median: 29.8 hours; IQR: 28.1 to 74.7 hours), or over 1.5 weeks of dedicated screening time, though this was as high as 196.7 hours in one of the SRs (over 5 weeks of dedicated screening time) (***Table 4***).

Table 4 - Time savings (in hours)

Systematic Review	FN	Time savings (in hours)			
		Total hrs saved	Title/Abst. screening	Retrieving articles †	Full-text screening ‡
Hot flashes	19	32.4	27.9	1.3	3.2
Opioid use disorder	46	207.5	196.7	3.1	7.7
Meniere's disease	15	32.2	28.7	1.0	2.5
Non-small cell lung cancer	34	29.8	21.9	2.2	5.6
Prophylaxis for influenza	19	92.0	87.6	1.3	3.1
Smoking cessation	40	20.6	11.3	2.7	6.7
Asthma/Urticaria	23	34.9	29.6	1.5	3.8
Depression screening	6	37.2	35.8	0.4	1.0
Prophylaxis for HIV	54	42.6	30.0	3.6	9.0
SSBs	243	215.1	158.5	16.2	40.5

Abst: Abstract; FN: False negatives; HIV: Human Immunodeficiency Viruses; hrs: hours; SSB: sugar sweetened beverages

† Estimated rate of 4 minutes/article (15 articles/hour)

‡ Estimated rate of 5 minutes/article (12 article/hour). This does not factor in any time to resolve any conflicts.

5.3 Discussion

This compilation of work was undertaken to identify how RRs are being defined in the literature, to identify research in the area of evaluating the impact of abbreviated methods for conducting RRs, and last, to evaluate the AI-AML tool in an online SR software to determine if it is a viable shortcut that could be employed while conducting RRs.

Researchers conducting SRs have several guidance documents to reference, such as the Cochrane Handbook [42] and the Methodological Expectations of Cochrane Intervention Reviews (MECIR) guideline [40]. Those conducting scoping reviews may reference the Joanna Briggs Institute guidance [43]. To date, there have been several organizations who have published guidance in the area of RR, including the World Health Organization (WHO) rapid advice guidelines [44], the Canadian Agency for Drugs and Technologies in Health (CADTH) Rapid Response Service [45], and the Samueli Institute's Rapid Evidence Assessment of the Literature (REAL ©) program [46]. However, using these three examples, we can see that there are three different terms used to describe this review type (i.e., rapid advice guidelines, rapid response service, and rapid evidence assessment). Further, the descriptions of how to conduct these reviews also differ, as these organizations offer a variety of rapid products.

The two systematic scoping reviews have resulted in important research in the area of RRs, as the lack of a clear definition can result in a heterogeneous set of products under the same name, or conversely, a homogeneous set of products under different names. As shown above with the WHO, CADTH, and the REAL © program, this was further supported in the definitions scoping review, which reported 79 unique citations referenced, included RRs which used 18 different terms, and, for feasibility, had an additional 23 terms excluded at the title and abstract level. A common term for labeling these products may not (i) be feasible, as many organizations have already adopted different terms for the same types of products, or (ii) be necessary, as study design labels may be ambiguous, and a focus on the defining features of the study is more important than the label [40,47]. However, a definition with central themes, which may be modified depending on the mandate or scope of the organization producing them (several examples provided in the RR definitions paper under suggested definitions [35]), may help producers of and readers/user of these reviews to identify this research, and to differentiate these reviews from SRs or other types of review type (e.g., overview of reviews, scoping reviews).

The publication of studies that formally evaluate abbreviated or omitted methods used in RRs is increasing. Among the 14 studies identified in the RR methods scoping review, 11 were published since 2017 (78.6%). This scoping review highlighted the gap in the evidence among several areas of review conduct (e.g., number of questions included, evaluating the certainty of the evidence). Among those that were represented, most were based on case studies, which may not be generalizable to all RRs. There may not be a 'one-size-fits-all' approach to RR methodology, as omissions or abbreviations should be selected based on factors such as:

1. The requirements of the stakeholders: for example, if the stakeholder is having an annual conference in three months, which requires a cursory investigation of a particular treatment, a RR may be appropriate.
2. The availability of resources: for example, not-for-profit organizations often have limited budgets to support the conduct of a full SR, considering the time taken and the size of the team required.
3. The topic area/question of the review: for example, including only English language publications for acupuncture therapy may result in several publications published in another language (e.g., Chinese) being excluded.

Regardless of the omitted or abbreviated methods used in the conduct of a RR, the impact of these omissions or abbreviations can help inform the creation of a set of methodological standards that could be applied across RRs, taking into consideration the three points above.

The results from the primary experiment evaluating the AI simulation tool in DistillerSR are promising. In addition to a significant reduction in the screening burden, the accuracy was 100%. Studies which informed the AI with a small set of records and then assigned the AI to screen the remaining records have performed poorly [23,24]. This shows the importance of 'active machine-learning', as AI is not yet ready to take over for humans, and requires sufficient input from humans to learn [25,30].

There is currently no agreed upon stopping criteria when using prioritized screening. There are several straightforward stopping rules which may be implemented, including stopping once a certain number of irrelevant records are reviewed consecutively (i.e., a heuristic approach) and stopping at a particular point due to time constraints (i.e., pragmatic approach). However, using the AI tools that have been integrated into several systematic review software displays a graphical/numerical representation of the percentage of the predicted relevant references have been identified (e.g., a predicted recall of 95%). Although, the only way of knowing you have

in fact captured 95% of the studies is to screen all studies, resulting in no time savings. Evaluation of these prediction tools have shown that a predicted recall of 95% is usually an underestimation of the true recall, and that in fact tends to be closer to 98-100% recall [30,48].

SCIENTIFIC CONTRIBUTION

Scoping Reviews

The work undertaken as part of the systematic scoping reviews was originally performed in 2019 to help inform Cochrane's decision as to whether RRs should be a formal Cochrane product. The results from the RR methods scoping review were used to develop a survey, which was distributed to 119 representatives to 20 Cochrane entities. This survey was developed to evaluate which methods would be seen as acceptable by different producers and users of SR and RR products. However, since the emergence of COVID-19, Cochrane has been producing RRs (https://covidrapidreviews.cochrane.org/resources). As part of this initiative they have adopted the proposed definition that was created from the results of the thematic analysis. The results from the methods scoping review and survey were used to develop the Methods Guidance document on the Cochrane COVID Rapid Reviews website (https://methods.cochrane.org/rapidreviews/cochrane-rr-methods), which has been published in the *Journal of Clinical Epidemiology* [49]. We envision these documents will be useful to producers and users of RRs beyond Cochrane, as the methods are not specific to Cochrane.

In addition to this work being accepted as an abstract for the 2019 Cochrane Colloquium in Chile (cancelled due to civil unrest) and as two oral presentations for the 2020 Cochrane Colloquium in Toronto (cancelled due to COVID), I was invited to give a presentation on the RR methods scoping review for the North American Systematic Review Methods Virtual Research Day on October 30th, 2020.

Evaluation of Active Machine-Learning

We expect the results from the AI simulation project will provide the SR and RR community with an approach that will increase the confidence in using AI for screening to identify relevant citations more quickly and to reduce the screening burden. As part of this work, we also provided a step by step set of instructions (i.e., tutorial) on how to use the AI simulation tool in DistillerSR (*Additional file 2 in the publication*). We felt this was important given that part of the barrier to using new technology is not knowing how to implement the technology.

FUTURE RESEARCH

The results from the RR methods scoping review can be the catalyst for a living review to create a database of studies that evaluate RR methods. In addition to identifying these studies, a formal data extraction of the results can be performed to produce a set of data to show how the impact of the omission or abbreviation was evaluated. This can be done several ways, including for example identifying the number of studies missed, evaluating the impact on the meta-analyses, and evaluating if there would be a change in conclusions. As most of the RR methods evaluations studies have been conducted on a small number of reviews, which may not be representative of all reviews of interventions, a living database could increase the sample size of the omissions and/or abbreviations evaluated.

These results may also contribute to the development of documents to encourage the transparent reporting and conduct of RRs. For example, the PRISMA extension for RR is currently under development [50]). Other possible extensions and/or modifications to well-known and highly cited reporting and conduct tools include an extension to A MeaSurement Tool to Assess the methodological quality of systematic Reviews (AMSTAR) for RR and MECIR for RRs.

As a follow-up to the AI AML publication, which included a tutorial on navigating the prioritization tool in DistillerSR, a manuscript was recently written and submitted to *BMC Medical Research Methodology* (March 2021) to provide general guidance to integrating prioritized screening into the conduct of a review. As there are several tools that include prioritized screening (e.g., Abstrakr, DistillerSR, EPPI-Reviewer, PICO Portal, RobotAnalyst, SWIFT-Active Screener, and SWIFT-Review), this manuscript was written to be software independent. Other areas of future research include the development of a database of the results from simulations of other reviews. Our experiment included 10 SRs, with results that may or may not be representative of all SRs. We encourage other review teams to run these simulations, whose results can be added to this database, which will increase the precision in the reduction of the screening burden and accuracy of the results produced from this initial experiment.

5.4 Conclusions

The conduct of RRs has been increasing, with a large increase in those published in peer-reviewed journals in the last five to seven years. This has been even further heightened during the COVID-19 pandemic, where several researchers are undertaking RRs who had not done so prior to the pandemic. However, an agreed upon definition and set of methodological standards does not currently exist. The works undertaken as part of this PhD has provided a thematic analysis on RRs definitions, and has provided a suggested definition, with additional caveats to consider, depending on the requirements of the funders, knowledge users, and/or stakeholders. It has also provided researchers with a repository of studies that formally evaluated RR methods, and contributed to other publications which may help guide the conduct of RRs. Last, it has evaluated an AI AML tool, which displays records in prioritized order to expedite title and abstract screening, which was determined to be a viable option for the conduct of RRs.

6. Sažetak

Naslov: Brzi Pregledi Literature: Definiranje, Ocjenjivanje Metoda I Olakšavanje Probira Korištenjem Umjetne Inteligencije

Uvod: Sustavni pregledi smatraju se zlatnim standardom u prikupljanju dostupnih dokaza koji se odnose na određeno pitanje i koriste se za informiranje politika javnog zdravstva. Smatraju se ključnim u stvaranju pouzdanih smjernica. Međutim, izrada sustavnih pregleda zahtijeva vrijeme i resurse, i možda neće biti napravljeni dovoljno brzo za dionike i donositelje odluka kada su potrebni hitni odgovori. Ciljevi brzih pregleda (engl. *rapid reviews*, RR) su pravodobno izraditi preglede dokaza, uz zadržavanje rigoroznih i robusnih metoda. Međutim, do danas je jedini konsenzus oko definicije RR taj da formalna definicija ne postoji. Uz to, ne postoji standardizirani skup metoda za RR, niti postoji sveobuhvatan pregled literature koji je empirijski procijenio metode RR i procijenio učinak tih skraćenih metoda. Cilj ove doktorske disertacije bio je: (i) utvrditi kako su RR definirani u literaturi i provesti tematsku analizu tih definicija kako bi se prepoznale ključne teme; (ii) pronaći i napraviti repozitorij empirijski procijenjenih skraćenih metoda za izradu RR te prepoznati područja u kojima su potrebna nova istraživanja; i (iii) procijeniti može li se olakšati probir literature korištenjem umjetne inteligencije i aktivnog alata za strojno učenje u internetskom računalnom programu za izradu sustavnog pregleda.

Metodologija objedinjenih radova: *Definicije brzih pregleda literature*: Napravljen je pretražni sustavni pregled (engl. scoping systematic review) kojim su nađeni RR objavljeni između 2017. i siječnja 2019. godine. Definicije RR izvučene su doslovno iz tih RR i provedena je tematska analiza kako bi se utvrdile ključne teme koje bi trebale biti uključene prilikom definiranja RR.

Metode RR: Napravljen je pretražni sustavni pregled kojim su pronađene metode za skraćenje RR, u radovima objavljenim od 1997. nadalje. Kako bi se napravio opsežan repozitorij dokumenata o RR, pronađene su dodatne studije (npr. o smjernicama za provođenje brzih pregleda, rasprava o terminologiji). Sve su publikacije podijeljene u jednu od četiri glavne kategorije na temelju svrhe publikacije. Oni koji su formalno ocjenjivali skraćene metode za RR mapirane su pomoću smjernica za pisanje Cochraneovih sustavnih pregleda MECIR (engl. *Methodological Expectations of Cochrane Intervention Reviews*) kako bi se utvrdilo ispunjavaju li te kriterije.

Na koncu je napravljena eksperimentalna evaluacija u programu DistillerSR® na 10 završenih sustavnih pregleda korištenjem alata za simulaciju umjetne inteligencije kako bi se izmjerilo

smanjenje opterećenja i točnosti probira (tj. koliko je relevantnih zapisa propušteno) kada se koristi probir pomoću strojnog učenja.

Rezultati: *Definicije RR*: U 216 RR i 90 članaka o metodama RR pronađene su ukupno 204 definicije koje se mogu tematski analizirati. Definirano je osam glavnih tema, a četiri teme pronađene su u 48,5% ili više definicija: Tema 4: Usporedba i kontrast sa sustavnim pregledima (SR) (68,1%; 139/204), Tema 2: Varijacije u skraćenim metodama (54,9%; 112/204), Tema 1: Ubrzani / brzi postupak i Tema 6: Obrazloženje obrazloženja učinkovitosti resursa (48,5%; 99/204 svaka). To je dovelo do predložene definicije "RR je oblik sinteze znanja koji ubrzava postupak provođenja tradicionalnog sustavnog pregleda putem racionalizacije ili izostavljanja različitih metoda kako bi se brže došlo do potrebnih dokaza."

Metode RR: Pronađeno je devedeset radova o metodama RR, od kojih je 14 formalno ocijenilo skraćene metode RR koje se odnose na nekoliko, ali ne sve, ključnih dimenzija povezanih s provođenjem pregleda literature. Bilo je moguće samo površno mapiranje kriterija MECIR-a, budući da su nedovoljne informacije priječile mogućnost utvrđivanja jesu li kriteriji zadovoljeni.

Alat za prioritizaciju aktivnog strojnog učenja: Alat za aktivno strojno učenje, koji koristi prioritetni probir, uvelike je smanjio teret probira za 10 sustavnih pregleda koji su procijenjeni. Medijan postotka zapisa koje je trebalo pregledati kako bi se pronašlo 95% zapisa uključenih na razini naslova i sažetka iznosio je 47,1% (interkvartalni raspon: 37,5 do 58,0%). Među 5% koji nisu prepoznati kao uključivi (tj. lažno-negativni naslovi i saželi), niti jedan nije bio uključen u konačni pregled, što je dovelo do 100%-tne točnosti.

Zaključak: Pojava RR zahtijeva dosljednost u njihovom definiranju kako bi se u literaturi mogao pronaći i napraviti homogeni skup proizvoda, bez obzira na termin koji se koristi za njihovu identifikaciju. Autorima koji rade RR trebaju smjernice o tome koje se skraćene metode mogu koristiti kako bi se potencijalna pristranost svela na najmanju moguću mjeru. Na koncu, aktivno strojno učenje održiva je metoda za smanjenje opterećenja probirom literature, koja se pokazala vrlo preciznom.

7. Abstract

Rapid reviews: defining, evaluating methods, and reducing screening burden using artificial intelligence

Introduction: Systematic reviews are considered the gold standard in collating available evidence related to a specific question and are used to inform policy for health care public health. They are considered to be essential in producing trustworthy guidelines. However, they are time- and resource-intensive undertakings which may not meet the timeline of stakeholders and policy-makers when urgent answers are required. The aims of rapid reviews are to produce evidence reviews in a timely manner, while maintaining rigorous and robust methods. However, to date, the only consensus around a definition of a rapid review is that a formal definition does not exist. Additionally, there is no standardized set of methods for rapid reviews, nor is there a comprehensive review which has compiled empirically evaluated rapid review methods and evaluated the impact of these abbreviated methods. The aim of this doctoral dissertation was to: (i) identify how rapid reviews have been defined in the literature and perform a thematic analysis of these definitions to identify the key themes; (ii) identify and create a repository of empirically evaluated methods abbreviations, and identify any gaps in the research; and (iii) evaluate the reduction in the screening burden and perform of an artificial intelligence and active machine-learning tool in an online systematic review software.

Methods: *RR definitions:* A systematic scoping review identifying rapid reviews published between 2017 and January 2019 was performed. Definitions of rapid reviews were extracted verbatim from these rapid reviews and a thematic analysis was performed to identify the key themes which should be included when defining a rapid review. *RR methods:* A systematic scoping review identifying formally evaluated rapid review methods abbreviations published from 1997 onward was performed. In order to create a comprehensive repository of rapid review documents, additional studies (e.g., around guidance on conducting rapid reviews, discussing terminology) were identified. All publications were divided into one of four main categories based on the purpose of the publication. Those that formally evaluated rapid review methods abbreviations were mapped to the Methodological Expectations of Cochrane Intervention Reviews (MECIR) to determine if they met these criteria. Lastly, an experimental evaluation was conducted in DistillerSR ® on 10 completed systematic reviews, using the artificial intelligence simulation tool, to measure the reduction in the screening burden and accuracy (i.e., how many relevant records were missed) when prioritized screening using active machine-learning was employed.

Results: *RR definitions:* A total of 204 definitions that could be thematically analyzed were identified in 216 rapid reviews and 90 rapid review methods papers. Eight major themes were identified, with four themes found in 48.5% or more of the definitions: Theme 4: Compare and contrast to SRs (68.1%; 139/204), Theme 2: Variation in shortcut methods (54.9%; 112/204), with Theme 1: Accelerated/rapid process and Theme 6: Resource efficiency rationale tied (48.5%; 99/204 each). This lead to a suggested definition of "A rapid review is a form of knowledge synthesis that accelerates the process of conducting a traditional systematic review through streamlining or omitting a variety of methods to produce evidence in a resource-efficient manner." *RR methods:* Ninety rapid review methods papers were identified, of which 14 formally evaluated rapid review methods abbreviations addressing several, but not all, key dimensions related to the conduct of a review. Only a cursory mapping to MECIR criteria was possible, as insufficient information impeded the ability to determine if criteria were met. *Active machine-learning prioritization tool:* The active machine-learning tool, employing prioritized screening, greatly reduced the screening burden of the 10 systematic reviews that were evaluated. The median percentage of studies required to be screened to identify 95% of the records included at the title and abstract level (true recall @ 95%) was 47.1% (IQR: 37.5 to 58.0%). Among the 5% that were not yet identified as included (i.e., title and abstract false negatives), none were included in the final review, resulting in 100% accuracy.

Conclusion: The emergence of rapid reviews, highlighted by the ongoing COVID-19 pandemic, requires consistency in how they are defined, in order to identify and produce a homogenous set of products regardless of the term used to identify them. Producers of rapid reviews also need guidance on which abbreviated methods may be used to keep potential bias minimized. Lastly, active machine-learning is a viable method to reduce the screening burden and was shown to be very accurate.

8. References

[1] About Cochrane reviews 2020. https://www.cochranelibrary.com/about/about-cochrane-reviews (accessed January 16, 2020).

[2] Fox DM. Evidence and Health Policy: Using and Regulating Systematic Reviews. Am J Public Health 2017;107:88–92. https://doi.org/10.2105/AJPH.2016.303485.

[3] Zhang Y, Akl E, Schunemann HJ. Using systematic reviews in guideline development: The GRADE approach. Research Synthesis Methods 2019;10:312–29.

[4] 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:e012545. https://doi.org/10.1136/bmjopen-2016-012545.

[5] Wang Z, Nayfeh T, Tetzlaff J, O'Blenis P, Murad MH. Error rates of human reviewers during abstract screening in systematic reviews. PLoS ONE 2020;15:e0227742. https://doi.org/10.1371/journal.pone.0227742.

[6] Cochrane Library: Cochrane Database of Systematic Reviews. Cochrane Library n.d. https://www.cochranelibrary.com/cdsr/about-cdsr (accessed May 14, 2020).

[7] Khangura S, Konnyu K, Cushman R, Grimshaw J, Moher D. Evidence summaries: the evolution of a rapid review approach. Syst Rev 2012;1:10. https://doi.org/10.1186/2046-4053-1-10.

[8] Tricco AC, Antony J, Zarin W, Strifler L, Ghassemi M, Ivory J, et al. A scoping review of rapid review methods. BMC Med 2015;13:224. https://doi.org/10.1186/s12916-015-0465-6.

[9] Kelly SE, Moher D, Clifford TJ. DEFINING RAPID REVIEWS: A MODIFIED DELPHI CONSENSUS APPROACH. Int J Technol Assess Health Care 2016;32:265–75. https://doi.org/10.1017/S0266462316000489.

[10] Ganann R, Ciliska D, Thomas H. Expediting systematic reviews: methods and implications of rapid reviews. Implement Sci 2010;5:56. https://doi.org/10.1186/1748-5908-5-56.

[11] Pluddemann A, Aronson JK, Onakpoya I, Heneghan C, Mahtani KR. Redefining rapid reviews: a flexible framework for restricted systematic reviews. BMJ Evidence-Based Medicine 2018;23:201–3. https://doi.org/10.1136/bmjebm-2018-110990.

[12] Haby MM, Chapman E, Clark R, Barreto J, Reveiz L, Lavis JN. What are the best methodologies for rapid reviews of the research evidence for evidence-informed decision making in health policy and practice: a rapid review. Health Res Policy Syst 2016;14:83. https://doi.org/10.1186/s12961-016-0155-7.

[13] Robson RC, Pham B, Hwee J, Thomas SM, Rios P, Page MJ, et al. Few studies exist examining methods for selecting studies, abstracting data, and appraising quality in a systematic review. Journal of Clinical Epidemiology 2018;0. https://doi.org/10.1016/j.jclinepi.2018.10.003.

[14] Wallace BC, Trikalinos TA, Lau J, Brodley C, Schmid CH. Semi-automated screening of biomedical citations for systematic reviews. BMC Bioinformatics 2010;11:55. https://doi.org/10.1186/1471-2105-11-55.

[15] Wang Z, Asi N, Elraiyah TA, Abu Dabrh AM, Undavalli C, Glasziou P, et al. Dual computer monitors to increase efficiency of conducting systematic reviews. J Clin Epidemiol 2014;67:1353–7. https://doi.org/10.1016/j.jclinepi.2014.06.011.

[16] Mortensen ML, Adam GP, Trikalinos TA, Kraska T, Wallace BC. An exploration of crowdsourcing citation screening for systematic reviews. Research Synthesis Methods 2017;8:366–86. https://doi.org/10.1002/jrsm.1252.

[17] Gartlehner G, Affengruber L, Titscher V, Noel-Storr A, Dooley G, Ballarini N, et al. Single-reviewer abstract screening missed 13 percent of relevant studies: a crowd-based, randomized controlled trial. Journal of Clinical Epidemiology 2020;121:20–8. https://doi.org/10.1016/j.jclinepi.2020.01.005.

[18] Rathbone J, Albarqouni L, Bakhit M, Beller E, Byambasuren O, Hoffmann T, et al. Expediting citation screening using PICo-based title-only screening for identifying studies in scoping searches and rapid reviews. Systematic Reviews 2017;6:233.

[19] 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. https://doi.org/10.2147/CLEP.S43118.

[20] 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. Research Synthesis Methods 2016;7:433–46.

[21] Edwards P, Clarke M, DiGuiseppi C, Pratap S, Roberts I, Wentz R. Identification of randomized controlled trials in systematic reviews: accuracy and reliability of screening records. Stat Med 2002;21:1635–40. https://doi.org/10.1002/sim.1190.

[22] 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;19:132. https://doi.org/10.1186/s12874-019-0782-0.

[23] Gartlehner G, Wagner G, Lux L, Affengruber L, Dobrescu A, Kaminski-Hartenthaler A, et al. Assessing the accuracy of machine-assisted abstract screening with DistillerAI: a user study. Systematic Reviews 2019;8:277. https://doi.org/10.1186/s13643-019-1221-3.

[24] Gates A, Guitard S, Pillay J, Elliott SA, Dyson MP, Newton AS, et al. Performance and usability of machine learning for screening in systematic reviews: a comparative evaluation of three tools. Systematic Reviews 2019;8:278. https://doi.org/10.1186/s13643-019-1222-2.

[25] Tsou AY, Treadwell JR, Erinoff E, Schoelles K. Machine learning for screening prioritization in systematic reviews: comparative performance of Abstrackr and EPPI-Reviewer. Systematic Reviews 2020;9:73.

[26] Howard BE, Phillips J, Miller K, Tandon A, Mav D, Shah MR, et al. SWIFT-Review: a text-mining workbench for systematic review. Systematic Reviews 2016;5:87. https://doi.org/10.1186/s13643-016-0263-z.

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

[28] 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 Medical Research Methodology 2020;20:7. https://doi.org/10.1186/s12874-020-0897-3.

[29] 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 Rev 2016;5:140. https://doi.org/10.1186/s13643-016-0315-4.

[30] Howard BE, Phillips J, Tandon A, Maharana A, Elmore R, Mav D, et al. SWIFT-Active Screener: Accelerated document screening through active learning and integrated recall estimation. Environment International 2020;138:105623. https://doi.org/10.1016/j.envint.2020.105623.

[31] 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 Medical Research Methodology 2020;20:139. https://doi.org/10.1186/s12874-020-01031-w.

[32] Giummarra MJ, Lau G, Gabbe BJ. Evaluation of Text Mining to Reduce Screening Workload for Injury-Focused Systematic Reviews. Injury Prevention 2020;26:55–60. https://doi.org/10.1136/injuryprev-2019-043247.

[33] Peters M, Godfrey C, McInerney P, Baldini Soares C, Khalil E, Parker D. Chapter 11: Scoping Reviews. Joanna Briggs Institute; 2017.

[34] Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med 2018. https://doi.org/10.7326/M18-0850.

[35] Hamel C, Michaud A, Thuku M, Skidmore B, Stevens A, Nussbaumer-Streit B, et al. Defining Rapid Reviews: a systematic scoping review and thematic analysis of definitions and defining characteristics of rapid reviews. J Clin Epidemiol 2021;129:74–85. https://doi.org/10.1016/j.jclinepi.202.09.041.

[36] McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Statement. J Clin Epidemiol 2016;75:40–6. https://doi.org/10.1016/j.jclinepi.2016.01.021.

[37] Hamel C, Michaud A, Thuku M, Affengruber L, Skidmore B, Nussbaumer-Streit B, et al. Few evaluative studies exist examining rapid review methodology across stages of conduct: a systematic scoping review. J Clin Epidemiol 2020;126:131–40. https://doi.org/doi.org/10.1016/j.jclinepi.2020.06.027.

[38] DistillerSR. Ottawa, ON, Canada: Evidence Partners; 2011.

[39] Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009;151:264–9, W64.

[40] Higgins J, Lasserson T, Chandler J, Tovey D, Thomas J, Flemyng E, et al. Methodological Expectations of Cochrane Intervention Reviews (MECIR): Standards for the conduct and reporting of new Cochrane Intervention Reviews, reporting of protocols and the planning, conduct and reporting of updates. 2019.

[41] 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 Medical Research Methodology 2020;20:256. https://doi.org/10.1186/s12874-020-01129-1.

[42] Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al. Cochrane Handbook for Systematic Reviews of Interventions version 6.2 (updated February 2021) 2021. www.training.cochrane.org/handbook (accessed April 25, 2021).

[43] Peters, MDJ, Godfrey, C, McInerney, P, Munn, Z, Tricco, A, Khalil, H. Chapter 11: Scoping Reviews. In: Aromataris E, Munn Z (Editors). JBI Manual for Evidence Synthesis. The Joanna Briggs Institute; 2020.

[44] Garritty CM. Chapter 11: Rapid advice guidelines in the setting of a public health emergency. WHO Handbook for Guideline Development. 2nd ed., World Health Organization; 2014, p. 133–56.

[45] CADTH Rapid Response Service: About the rapid response service n.d. https://www.cadth.ca/about-cadth/what-we-do/products-services/rapid-response-service (accessed November 12, 2019).

[46] Crawford C, Boyd C, Jain S, Khorsan R, Jonas W. Rapid Evidence Assessment of the Literature (REAL(©)): streamlining the systematic review process and creating utility for evidence-based health care. BMC Research Notes 2015;8:631. https://doi.org/10.1186/s13104-015-1604-z.

[47] McKenzie J, Brennan S, Ryan R, Thomson H, Johnston R, Thomas J. Chapter 3: Defining the criteria for including studies and how they will be grouped for the synthesis. Cochrane Handbook for Systematic Reviews of Interventions version 6.0, 2019.

[48] The Case for Artificial Intelligence in Systematic Reviews. Ottawa, Canda: Evidence Partners, Inc.; 2020.

[49] Garritty C, Gartlehner G, Nussbaumer-Streit B, King VJ, Hamel C, Kamel C, et al. Cochrane Rapid Reviews Methods Group offers evidence-informed guidance to conduct rapid reviews. Journal of Clinical Epidemiology 2020;0. https://doi.org/10.1016/j.jclinepi.2020.10.007.

[50] Stevens A, Garritty C, Hersi M, Moher D. Developing PRISMA-RR, a reporting guideline for rapid reviews of primary studies (Protocol). 2018.

9. Curriculum Vitae: Candyce Hamel

EDUCATION

October 2018 – present
PhD Epidemiology (candidate), University of Split, School of Medicine, Translational Research in Biomedicine (TRIBE) program
Thesis title: Rapid reviews: defining, evaluating methods, and reducing screening burden using artificial intelligence.
Mentor: Beverley Shea, Supervisor: Adrienne Stevens.

January 2007 – August 2009
Master of Science (MSc) in Electronic Business Technologies, University of Ottawa, Faculty of Graduate and Postdoctoral Studies. Thesis title: Determinants of participation in Online Communities of Practice (CoP). Supervisor: Morad Benyoucef.

January 2003 – August 2006
Bachelor of Commerce (BComm), Major in Decision Science and Management Information Systems, Concordia University, John Molson School of Business (Co-op Program). Graduated with Distinction.

September 1997 – July 1999
Diploma in Business - Information Systems, Algonquin College, School of Applied Arts and Business. Graduated with honours.

EMPLOYMENT HISTORY

February 2021 – present, ***SENIOR EPIDEMOLOGIST,*** The Canadian Association of Radiologists (CAR)

August 2013 – February 2021, ***SENIOR CLINICAL RESEARCH ASSOCIATE***, Ottawa Hospital Research Institute

August 2017 – present, ***CONSULTANT****,* Cochrane Response

May 2004 – July 2013, ***RESEARCH & DATA ANALYST****,* CIETcanada

January 2006 – May 2006, ***MARKER****, Concordia University, Decision Science and Management Information Systems Department*

January 2005 – April 2005 and September 2005 – December 2005, ***ASSISTANT ANALYST*** *(Intern),* Yellow Pages Group, Corporate Performance Department

February 2001 – May 2004, ***RESEARCH ASSISTANT***, University of Ottawa – Institute of Population Health

September 2000 – January 2001, ***INSIDE SALES REPRESENTATIVE****,* Sceptre ABS

May 2000 – August 2000, ***RESEARCH SECRETARY****,* Loeb Health Research Institute

July 1995 – August 2000, ***RESEARCH ASSISTANT****,* Ottawa Hospital

PROFESSIONAL MEMBERSHIPS

Associate Convenor, Cochrane Rapid Review Methods Group, 2018-present

Editorial Board Member, Associate Editor, BMC Medical Research Methodology, June 2020-present

ISPOR Good Practices Task Force member: Systematic Reviews with Economic Outcomes, 2018-2021

HONOURS & SCHOLARSHIPS

- *August 2009*, Dean's award. Total: $1,500
- *January 2009-July 2009,* NEAHR research award. Total: $20,500
- *January 2007-July 2009,* Admissions Scholarship, University of Ottawa. Total: $10,000
- *June 2006,* Membership in Beta Gamma Sigma, graduated top 10% of class
- *December 2003,* Golden Key Society – An academic honours society

PEER-REVIEW PUBLICATIONS

SUBMITTED

1. **Hamel C**, Hersi M, Kelly SE, Tricco AC, Straus S, Wells G, Pham B, Hutton B. Guidance for using artificial intelligence for title and abstract screening while conducting knowledge syntheses. Submitted to *BMC Medical Research Methodologies* March 2021.
2. Wieland LS, **Hamel C**, Konstantinidis M, Nourouzpour N, Shipper A, Lipski E. Zinc for prevention and treatment of the common cold. Submitted to *The Cochran Library* March 2021.
3. Rice D, Wolfe D, Hersi M, Esmaeilisaraji L, Butler C, **Hamel C**, Barbeau P, Skidmore P, Hutton B. Managing concurrent chronic pain, mental illness and substance use disorder: A Rapid Review of Guidelines and Knowledge Syntheses. Submitted to *Canadian Journal of Pain* December 2020.
4. Wolfe D, Rice D, Hersi M, Esmaeilisaraji L, Butler C, **Hamel C**, Ahmadzai N, Skidmore P, Hutton B. Reducing the risk of transition from acute to chronic pain: a rapid review of guidelines and knowledge syntheses. Submitted to *Canadian Journal of Pain* December 2020.
5. **Hamel C**, Beck A, Thuku M, Esmaeilisaraji L, Skidmore B, Colman I, Grigoriadis S, Nicholls SG, Potter BK, Ritchie K, Vasa P, Hutton B, Shea BJ, Little J, Stevens A. Screening for depression among the general adult population and in women during pregnancy or the first-year postpartum: two systematic reviews to update a guideline of the Canadian Task Force on Preventive Health Care. Submitted to *Systematic Reviews* December 2019 (currently undergoing a pre-publication update prior to resubmission).

PUBLISHED

1. **Hamel C**, Garritty C, Hersi M, Butler C, Esmaeilisaraji L, Rice D, Straus S, Skidmore B, Hutton B. Effective Models of Provider Care in Long-term Care: A Rapid Scoping Review. *PLoS ONE* 16(7): e0254527. https://doi.org/10.1371/journal.pone.0254527.
2. Cole K, Hutton B, **Hamel C**, Bourque J-M, Arnaout A, Clemons M. Breast cancer in Indigenous women living in Canada: a scoping review protocol. *JBI Evidence Synthesis* 2021 Jun 23. doi: 10.11124/JBIES-20-00522

3. Mandrik O, Severens H, Bardach A, Ghabri S, **Hamel C**, Mathes, Vale L, Wisloff T, Goldhaber-Fiebert J. Critical Appraisal of Systematic Reviews with Costs and Cost-effectiveness Outcomes: an ISPOR Good Practices Task Force Report. *Value in Health* 2021 Apr; 24(4): 463-472.
4. Yogendrakumar V, Shamy M, Dewar B, Fergusson D, Dowlatshahi D, **Hamel C**, Gocan S, Fedyk M, Mas J-L, Rothwell P, Howard V, Bereznyakova O. Identifying Sex-Specific Differences in the Carotid Revascularization Literature: Findings from a Scoping Review. *Stroke & Vascular Neurology* 2021 Mar. doi: 10.1136/svn-2020-000744
5. Beck A, LeBlanc JC, Morissette K, **Hamel C**, Skidmore B, Colquhoun H, Lang E, Moore A, Riva JJ, Thombs BD, Patten S, Bragg H, Colman I, Goldfield GS, Nicholls SG, Pajer K, Potter BK, Meeder R; Vasa P; Hutton B, Shea BJ, Graham E, Little J, Moher D, Stevens A. Screening for depression in children and adolescents: a protocol for a systematic review update. *Systematic Reviews* 2021 Jan; 10: 24.
6. Garritty C, Gartlehner G, Nussbaumer-Streit, King VJ, Kamel C, Stevens A, Devane D, **Hamel C**, Affengruber L. Rapid review methods guidance aids in Cochrane's quick response to COVID-19 crisis. In: Collaborating in response to COVID-19: editorial and methods initiatives across Cochrane. Cochrane Database of Systematic Reviews 2020; 12 Suppl 1: 28-31. 10.1002/14651858.CD202002
7. Barbeau P, Michaud A, **Hamel C**, Rice D, Skidmore B, Hutton B, Garritty C, da Silva DF, Semeniuk K, Adamo KB. Musculoskeletal Injuries among Females in the Military: A Scoping Review. *Military Medicine* 2020 Dec. doi: 10.1093/milmed/usaa555
8. Garritty C, Gartlehner G, Nussbaumer-Streit B, King V, **Hamel C**, Affengruber L, Stevens A, Kamel C. Rapid Review Methods Recommendations: Interim Guidance from the Cochrane Rapid Reviews Methods Group. *Journal of Clinical Epidemiology* 2021 Feb; 130: 13-22. ePub ahead of print: accepted October 8, 2020; available online October 15, 2020.
9. **Hamel C**, Michaud A, Thuku M, Skidmore B, Stevens A, Nussbaumer-Streit B, Garritty C. Defining Rapid Reviews: a systematic scoping review and thematic analysis of definitions and defining characteristics of rapid reviews. *Journal of Clinical Epidemiology* 2021 Jan; 129:74-85. ePub ahead of print: accepted September 29, 2020; available online October 8, 2020.
10. **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 Medical Research Methodology* 2020; 20: 256. ePub ahead of print: accepted September 22, 2020; published October 15, 2020.
11. **Hamel C**, Corace K, Hersi M, Rice D, Willows M, MacPherson P, Sproule B, Flores-Aranda J, Garber G, Esmaeilisaraji L, Skidmore B, Porath A, Ortiz Nunez R, Hutton B. Psychosocial and pharmacologic interventions for methamphetamine addiction: protocol for a scoping review of the literature. *Systematic Reviews* 2020; 9, 245.
12. **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. *Journal of Clinical Epidemiology* 2020 Oct; 126:131-140. Epub 2020 Jun 26. PMID: 32599023
13. Gardner W, Nicholls SG, Reid GJ, Hutton B, **Hamel C**, Sikora L, Salamatmanesh M, Duncan L, Georgiades K, Gilliland J. A protocol for a scoping review of equity measurement in mental health care for children and youth. *Systematic Reviews* 2020; 9:233.
14. Garritty C, **Hamel C**, Hersi M, Butler C, Monfaredi Z, Stevens A, Nussbaumer-Streit B, Cheng W, Moher D. Assessing how information is packaged in rapid reviews for

policymakers and other stakeholders: a cross-sectional study. *BMC: Health Research Policy and Systems* September 2020; 18:112.

15. **Hamel C**, Thuku M, Michaud A, Esmaeilisaraji L, Sikora L, Hutton B, Fung Kee Fung K. Prevention of Rh alloimmunization in pregnant and post-partum women: a systematic review. *PLoS ONE* 2020 15(9): e0238844.
16. Garritty C, Hersi M, **Hamel C**, Stevens A, Monfaredi Z, Butler C, Tricco AC, Hartling L, Stewart L, Welch V, Thavorn K, Cheng W, Moher D. Assessing the format and content of journal published and non-journal published rapid review reports: a comparative study. *PLOS One* 2020 Aug; 15(8): e0238025.
17. Pratt M, Garritty C, Esmaeilisaraji L, **Hamel C**, Millar K, Skidmore B, Dougan S, Armour C. Application of exome sequencing for prenatal diagnosis: a rapid scoping review. *Genetics in Medicine* 2020 Aug; 22:1925-1934.
18. **Hamel C**, Ahmadzai N, Beck A, Thuku M, Skidmore B, Pussegoda K, Bjerre L, Chatterjee A, Dennis K, Ferri L, Maziak DE, Shea BJ, Hutton B, Little J, Moher D, Stevens A. Screening for esophageal adenocarcinoma and precancerous conditions (dysplasia and Barrett's esophagus) in patients with chronic gastroesophageal reflux disease with or without other risk factors: two systematic reviews and one overview of reviews to inform a guideline of the Canadian Task Force on Preventive Health Care (CTFPHC). *Systematic Reviews* 2020 Jan; 9: 20.
19. Bereznyakova O, Dewar B, Dowlatshahi D, Howard V, **Hamel C**, Gocan S, Fedyk M, Shamy M. The benefit of carotid revascularization for women with symptomatic carotid stenosis: protocol for a systematic review. *BMJ Open* 2019 Nov;9:e032140.
20. Sigfrid L, Moore C, Salam AP, Maayan N, **Hamel C**, Garritty C, Lutje V, Buckley B, Soares-Weiser K, Marshall R, Clarke M, Horby P. A rapid research needs appraisal methodology to inform clinical research priorities in response to outbreaks - results from the Lassa fever pilot. *BMC Medicine* 2019; 17:107.
21. Garritty C, Stevens A, **Hamel C**, Golfam M, Hutton B, Wolfe D. Knowledge Synthesis in Evidence Based Medicine. *Nuclear Medicine.* 2019 Mar; 49(2):136-144. doi: 10.1053/j.semnuclmed.2018.11.006
22. Velez M, **Hamel C**, Hutton B, Gaudet L, Walker M, Thuku M, Cobey K, Pratt M, Skidmore B, Smith GN. Care plans for women pregnant using assisted reproductive technologies: A Systematic Review. *Reproductive Health.* 2019 Jan; 16(9).
23. **Hamel C**, Lang E, Morrisette K, Beck A, Stevens A, et al. Screening for depression in women during pregnancy or the first-year postpartum and in the general adult population: a protocol for two systematic reviews to update a guideline of the Canadian Task Force on Preventive Health Care. *Systematic Reviews.* 2019 Jan; 8:27.
24. Fitzpatrick T, Gocan S, Wang CQ, **Hamel C**, Bourgoin A, Dowlatshahi D, Stotts G, Shamy M. How do neurologists diagnose TIA: A systematic review. *International Journal of Stroke*. 2019 Feb; 14(2): 115-124. E-pub ahead of print 2018 Dec.
25. Wolfe D, Yazdi F, Kanji S, Burry L, Beck A, Butler C, Esmaeilisaraji L, **Hamel C**, Hersi M, Skidmore B, Moher D, Hutton B. Incidence, causes, and consequences of preventable adverse drug reactions occurring in inpatients: a systematic review of systematic reviews. *PLOS One* 2018; 13(10): e0205426.
26. Pease C, Hutton B, Yazdi F, Wolfe F, **Hamel C**, Barbeau P, Skidmore B, Alvarez GG. A systematic review of adverse events of rifapentine and isoniazid compared to other treatments for latent tuberculosis infection. *Pharmacoepidemiology and Drug Safety* 2018 Jun; 27(6): 557-566.

27. **Hamel C**, Ghannad M, McInnes MDF, Marshall J, Earnshaw JJ, Ward R, Skidmore B, Garritty C. Potential benefits and harms of offering ultrasound surveillance to men aged 65 and over with a subaneurysmal (2.5-2.9cm) infrarenal aorta. *Journal of Vascular Surgery* 2018 Apr; 67(4):1298-1307.
28. Shea BJ, Reeves BC, Wells G, Thuku M, **Hamel C**, Moran J, Moher D, Tugwell P, Welch V, Kristjansson E, Henry DA. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomized or non-randomized studies of healthcare interventions, or both. *BMJ* 2017; 358: j4008.
29. Moher D, Shamseer L, Cobey KD, Lalu MM, Galipeau J, Avey MT, Ahmadzai N, Alabousi M, Barbeau P, Beck A, Daniel R, Frank R, Ghannad M, **Hamel C**, et al. Stop this waste of people, animals and money. *Nature* 2017 Sep; 549(7670): 23-25.
30. Pease C, Hutton B, Yazdi F, Wolfe D, **Hamel C**, Quach P, Skidmore B, Moher D, Alvarez G. Efficacy and completion rates of rifapentine and isoniazid (3HP) compared to other treatment regimens for latent tuberculosis infection: a systematic review with network meta-analyses. *BMC Infectious Diseases* 2017; 17(1): 265.
31. Fitzpatrick E, **Hamel C**, Stevens A, Pratt M, Moher D, Doucet S, Neuss D, Bernstein A, Na E. Sign Language and Spoken Language for Children With Hearing Loss: A Systematic Review. *Pediatrics* 2016 Jan;137(1). doi: 10.1542/peds.2015-1974. Epub 2015 Dec 18
32. Hersi M, Stevens A, Quach P, **Hamel C**, Thavorn K, Garritty C, Skidmore B, Vallenas C, Norris SL, Egger M, Eremin S, Ferri M, Shindo N, Moher D. Effectiveness of personal protective equipment for healthcare workers caring for patients with filovirus disease: a rapid review. *PLoS One*. 2015 Oct 9;10(10): e0140290. doi: 10.1371/journal.pone.0140290. eCollection 2015
33. **Hamel C**, Enne J, Omer K, Ayara N, Yarima Y, Cockcroft A, Andersson N. Childhood Malnutrition is Associated with Maternal Care During Pregnancy and Childbirth. A Cross-Sectional Study in Bauchi and Cross River States, Nigeria. *Journal of Public Health Research* 2015 Apr 17;4(1):408. doi: 10.4081/jphr.2015.408
34. **Hamel C**, Stevens A, Singh K, Ansari M, Myers E, Ziegler P, Hutton B, Sharma A, Bjerre L, Fenton S, Lau D, O'Hara K, Reid R, Salewski E, Shrier I, Willows N, Tremblay M, Moher D. Do sugar-sweetened beverages cause adverse health outcomes in adults? A systematic review protocol. *Systematic Reviews* 2014;3; 108.
35. Stevens A, **Hamel C**, Singh K, Ansari M, Myers E, Ziegler P, Hutton B, Sharma A, Bjerre L, Fenton S, Gow R, Hadjiyannakis S, O'Hara K, Pound C, Salewski E, Shrier I, Willows N, Moher D, Tremblay M. Do sugar-sweetened beverages cause adverse health outcomes in children? A systematic review protocol. *Systematic Reviews* 2014 Sept; 3(1): 96.
36. **Hamel C**, Benyoucef M, Kuziemsky C. Determinants of Participation in an Inuit Online Community of Practice. *Knowledge Management Research & Practice* 2012; 10:41-54.
37. Cockcroft A, Khan A, Ansari NM, Omer K, **Hamel C**, Andersson N. Does contracting of health care in Afghanistan work? Public and service-users' perceptions and experience. *BMC Health Services Research* 2011; 11(Suppl 2): S11 (21 December 2011)
38. Cockcroft A, Pearson LW, **Hamel C**, Andersson N. Reproductive and sexual health in the Maldives: analysis of data from two cross-sectional surveys. *BMC Health Services Research* 2011; 11(Suppl 2):S6 (21 December 2011)
39. Andersson N, Omer K, Caldwell D, Dambam MM, Maikudi AY, Effiong B, Ikpi E, Udofia E, Khan A, Ansari U, Ansari N, **Hamel C**. Male responsibility and maternal morbidity: a cross-sectional study in two Nigerian states. *BMC Health Services Research* 2011; 11(Suppl 2):S7 (21 December 2011).

40. Nava-Aguilera E, Andersson N, Harris E, Mitchell S, **Hamel C**, Shea B, López-Vidal Y, Villegas-Arrizón A, Morales-Pérez A. Risk factors associated with recent transmission of tuberculosis: Systematic Review and Meta-Analysis. *Int J Tuberc Lung Dis* 2009 Jan; 13(1):17-26.
41. Shea BJ, **Hamel C**, Wells G, Bouter LM, Kristjansson B, Grimshaw J, Henry DA, Boers M. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. *J Clin Epidemiol* 2009; 62(10):1013-1020.
42. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, **Hamel C**, Porter AC, Tugwell P, Moher D, Bouter L. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Medical Research Methodology* 2007; 7:10.
43. Shea B, Boers M, Grimshaw GM, **Hamel CD** and Bouter LM. Does updating improve the methodological and reporting quality of Cochrane systematic reviews? *BMC Med Res Methodol* 2006 Jun 13;6(1):27.
44. Shea B, Wells G, Cranney A, Zytaruk N, Robinson V, Griffith L, **Hamel C**, Ortiz Z, Peterson J, Adachi J, Tugwell P, Guyatt G, Osteoporosis Methodology Group, Osteoporosis Research Advisory Group. Calcium supplementation on bone loss in postmenopausal women. *Cochrane Database of Systematic Reviews* 2004; (1).

Dissertation

1. **Hamel C**. Determinants of Participation in an Online Community of Practice (OCoP). MSc e-business technologies, University of Ottawa, 2009.

Presentations/Posters

1. **Hamel C**. North American Systematic Review Methods Virtual Research Day, October 30, 2020. Invited presentation.
2. **Hamel C**, Michaud A, Affengruber L, Skidmore B, Stevens A, Nussbaumer-Streit B, Garritty C. Rapid review methods: a systematic scoping review. In: Advances in Evidence Synthesis: special issue Cochrane Database of Systematic Reviews 2020;(9 Suppl 1): 431. https://doi.org/10.1002/14651858.CD202001.
3. **Hamel C**, Michaud A, Thuku M, Skidmore B, Stevens A, Nussbaumer-Streit B, Garritty C. Defining Rapid Reviews: a systematic scoping review and thematic analysis of definitions and defining characteristics of rapid reviews. In: Advances in Evidence Synthesis: special issue Cochrane Database of Systematic Reviews 2020;(9 Suppl 1): 422. https://doi.org/10.1002/14651858.CD202001.
4. **Hamel C**, Michaud A, Affengruber L, Skidmore B, Stevens A, Nussbaumer-Streit B, Garritty C. Rapid review definition and methods: two systematic scoping reviews. Cochrane Colloquium 2019 (Santiago, Chile: virtual)
5. **Hamel C**. Forum for Aboriginal Study and Research, University of Ottawa. Student Workshop, April 14, 2009. *Sources of Participation for Online Communities of Practice (CoP)*.

10. Papers Combined in the Dissertation

1. Hamel C, Michaud A, Thuku M, Skidmore B, Stevens A, Nussbaumer-Streit B, Garritty C. Defining Rapid Reviews: a systematic scoping review and thematic analysis of definitions and defining characteristics of rapid reviews. *J Clin Epi* 2021 Jan; 129:74-85. (JIF 4.952), ePub: 2020 October 8. doi: 10.1016/j.jclinepi.2020.09.041.

2. 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 Epi* 2020 Oct; 126:131-140. (JIF 4.952), ePub: 2020 Jun 26. doi: 10.1016/j.jclinepi.2020.06.027.

3. 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 Medical Research Methodology* 2020; 20: 256. (JIF 3.031), ePub: 2020 Jun. doi: 10.1186/s12874-020-01129-1.

Journal of Clinical Epidemiology 129 (2021) 74–85

Journal of Clinical Epidemiology

REVIEW

Defining rapid reviews: a systematic scoping review and thematic analysis of definitions and defining characteristics of rapid reviews

Candyce Hamel[a,b,*], Alan Michaud[a], Micere Thuku[a], Becky Skidmore[a], Adrienne Stevens[a], Barbara Nussbaumer-Streit[c], Chantelle Garritty[a,b]

[a]Ottawa Hospital Research Institute, Knowledge Synthesis Group, Ottawa, ON K1H 8L6, Canada
[b]University of Split, School of Medicine, Split, Croatia 21000
[c]Cochrane Austria, Department for Evidence-based Medicine and Evaluation, Danube University Krems, Krems, Austria

Accepted 29 September 2020; Published online 8 October 2020

Abstract

Background and Objective: Rapid reviews were first mentioned in the literature in 1997, when Best et al. described the rapid health technology assessment program in the south and west regions of England but did not provide a formal definition. More recently, the only consensus around a rapid review definition is that a formal definition does not exist. The primary aim of this work is to create a repository of existing definitions and to identify key themes, which may help the knowledge synthesis community in defining rapid review products.

Methods: A systematic scoping review was performed to identify definitions used in journal-published rapid reviews written in English between 2017 and January 2019. We searched Medline, Embase Classic + Embase, PsycINFO, ERIC, Cochrane Library, CINAHL, and Web of Science on December 21, 2018. Two reviewers performed study selection and data extraction using *a priori*–defined methods published in a protocol. Definitions from rapid review methods articles (published from 1997 onward) identified in another scoping review were added to the results, and all definitions were thematically analyzed using NVivo. A quantitative analysis was also performed around studies cited.

Results: Definitions from 216 rapid reviews and 90 rapid review methods articles were included in the thematic analysis. Eight key themes were identified: accelerated/rapid process or approach, variation in methods shortcuts, focus/depth/breadth of scope, compare and contrast to a full traditional systematic review, stakeholder rationale, resource efficiency rationale, systematic approach, bias/limitations. Secondary referencing was a common occurrence.

Conclusion: Thematic analysis performed in this systematic scoping review has allowed for the creation of a suggested definition for rapid reviews that can be used to inform the systematic review community. © 2020 Elsevier Inc. All rights reserved.

Keywords:: Scoping review; Rapid reviews; Definition; Thematic analysis

1. Introduction

A rapid review (RR) was originally mentioned in the literature in 1997, when Best et al. described the rapid health technology assessment program in the south and west regions of England [1]. Although they did not provide a definition of an RR, they described a service which produces reports within two person months. The key features of the service were to produce reports that were accurate, timely, and accessible to decision makers. More recently, the only consensus around an RR definition is that a formal definition does not exist [2–4]. Several definitions have been used in publications about RR methods, RR programs, and RRs themselves. In 2016, Kelly et al. performed a modified Delphi consensus approach and came up with a set of statements defining the characteristics of an RR [4], but did not provide a formal definition or a systematic evaluation of existing definitions.

The popularity of RRs has been increasing over the past 20 years, with various organizations developing RRs, including the World Health Organization (WHO) [5], the

Funding: This work is supported in part by funds provided by Cochrane and in part from a Canadian Institutes of Health Research grant (funding research number 142310). The funders had no role in the development of this manuscript and its protocol (registered on OSF: https://osf.io/y5f2m/). They did not have a role in the data collection, analyses, interpretation of the data, and publication of the findings.

Declaration of interest: None.

* Corresponding author. Ottawa Hospital Research Institute, 501 Smyth Road, Ottawa, Ontario, Canada K1H 8L6.
E-mail address: cahamel@ohri.ca (C. Hamel).

https://doi.org/10.1016/j.jclinepi.2020.09.041
0895-4356/© 2020 Elsevier Inc. All rights reserved.

What is new?

Key findings

- A repository of definitions from 216 rapid reviews and 90 rapid review methods articles was created (158 rapid reviews and 73 rapid review methods articles provided a definition).
- Among the rapid reviews, 59 unique references were cited 275 times. The top four cited authors were referenced 135 times. Among rapid review methods articles, 50 unique references were cited 179 times.
- A thematic analysis identified eight key themes in defining rapid reviews.
- Secondary referencing was common among cited articles.

What this adds to what was known?

- There is currently no consensus on what defines a rapid review.
- The four most commonly reported themes (used in ~≥50% of definitions) were used to create a preliminary definition of a rapid review. Suggestions are included on how users might tailor this definition to best meet their individual remit and mandates for producing rapid reviews.

What is the implication and what should change now?

- The preliminary definition, with caveats, presented can help the systematic review community define their review with consistency, regardless of the label used to describe it.

Samueli Institute's Rapid Evidence Assessment of the Literature (REAL©) program [6], and the Canadian Agency for Drugs and Technologies in Health Rapid Response Service [7]. The number of RRs published in the last 5 years has steadily increased. In 2013, 15 journal-published RRs were identified, growing to 52 by 2016, and 108 in 2018. Although these numbers are small, most RRs are not published in journals. For example, in 2016, 52 published RRs and over 250 unpublished RRs were identified from various health care organizations. In 2017, the Knowledge Synthesis Group at the Ottawa Hospital Research Institute identified 148 organizations globally who produced RRs [RR workshop presentation, November 27, 2019 Ottawa, Canada, with data derived from internal projects].

Some of the problems with lacking a common definition for RRs are that it makes it difficult

(i)
for researchers (e.g., building search strategies that accurately identify RR) and readers/users of results to identify RRs correctly. This is important as the line may be blurred (both in the conduct and the resulting conclusions) between systematic reviews (SRs) that do not meet a high-quality methodological conduct (e.g., low or critical risk using AMSTAR 2) and RRs that use transparent, measured abbreviated methods;
(ii)
to create and set methodological standards and apply consistent constructs (e.g., Preferred Items in Systematic Reviews and Meta-Analysis [PRISMA] for RRs, AMSTAR for RRs); and
(iii)
as it results in a heterogeneous set of products under the same name or conversely a homogeneous set of products under different names. The term 'rapid' points toward the speed at which the review is performed and not the abbreviation or omission of steps taken to conduct the review. For this reason, researchers have suggested other terms be used, for example, restricted reviews [8,9]. To date, 'rapid review' is the term that has been colloquially adopted by the research community and endorsed by various organizations, including Cochrane and the WHO. However, other organizations have chosen other terms, such as rapid evidence assessment by the UK government.

2. Objective

The objective of this systematic scoping review was to identify published RR literature to answer the question: How are RRs defined in the literature? This work will provide a summary of existing definitions identified in the literature on RRs and examine existing definitions to identify common themes across the body of literature. Creating a repository of existing definitions and developing a preliminary definition, while allowing for caveats and flexibility depending on the organizational preferences or mandate, is an important step in helping the knowledge synthesis community conduct and identify RRs. In addition, as Cochrane considers RRs an important piece in their content strategy, this makes this topic very important as these results will inform discussions within Cochrane on the utility of RRs as a product.

3. Methods

This systematic scoping review was guided by established scoping review methodology [10,11] and has been prepared in accordance with the PRISMA extension for Scoping Reviews (PRISMA-ScR) [12]. A protocol for this work was registered on the Open Science Framework

(OSF: https://osf.io/y5f2m/). Methods are briefly described in Table 1, with additional details and deviations from the protocol in Appendix A.

4. Results

The search strategies to identify RRs resulted in 2,657 unique records, of which 422 were evaluated at full text, with 216 RRs included (Figure 1). Several records were excluded at title/abstract as they did not explicitly state the term rapid or a derivative. For feasibility, only those with rapid, expedited, or abbreviated were considered for inclusion, while excluding those that described the review as focused (n = 347), targeted (n = 54), or pragmatic (n = 28). In addition, several other terms which may be considered 'rapid' derivatives were identified; however, because of the number of these records and our focus on those who self-declared as 'rapid', they were also excluded (n = 127) (Appendix C).

Among the 216 RRs, 101 were published in 2017, 106 were published in 2018, and nine were published in January of 2019 (Table 2). Most of the RRs (82.5%) were from corresponding authors from the United Kingdom (n = 82), Australia (n = 41), the United States of America (n = 31), and Canada (n = 24). Almost two-thirds (63.0%) used the term RR, with others using the terms rapid evidence assessment (10.1%) and rapid systematic review (8.8%). Nearly two-thirds (141 of 216; 65.3%) first used the term in the title, with the remaining first using the term in the abstract (Appendix D).

4.1. Definitions from published RRs

In total, 158 (73%) RRs provided a definition. Fifteen provided their own (i.e., 11 providing only their own definition and four referencing their own in addition to other authors), and one provided a definition, but the references in the publication did not line up and therefore no references were recorded [16]. Some RR authors did not provide an explicit definition, but made reference to another author or method (e.g., "We conducted a review of the literature using the rapid evidence assessment (REA) method [17,18]."") [19]. Among the 146 RRs that provided a definition citing another author, 59 unique references were cited a total of 275 times (Appendix E.1). Among all RRs, a median of two references (range 0 to 7) were cited. Furthermore, 29 articles were cited once. The top four articles cited were Khangura 2012 (n = 54) [2], Ganann 2010 (n = 42) [20], Tricco 2015 (n = 21) [3], and Grant 2009 (n = 18) [18] (Table 3).

4.2. Definitions from RR methods articles

In total, 81% (73 of 90) of the RR methods articles provided a definition. These definitions were included in the thematic analysis to supplement the definitions identified in the RRs. Briefly, methods articles were published between 1997 and 2019, with the majority of the articles published since 2014 (68 of 90 (75.6%)). A total of 200

Table 1. Methods in brief

Project stage	Method description
Eligibility criteria	- Published rapid reviews using 'rapid' or derivative (e.g., abbreviated) in the title or abstract
	- Published between January 2017 and January 2019
	- Written in English (for feasibility)
Searching for studies	- Developed by an experienced information specialist with input on search terms by members of the research team
	- Peer-reviewed using the PRESS checklist [13]
	- Search (Dec 2018): MEDLINE® ALL, Embase Classic + Embase, PsycINFO, ERIC, Cochrane Library, CINAHL, Web of Science (Appendix B)
	- Search strategies not restricted by language
	- Supplemented with definitions from rapid review methods scoping review [14]
Study selection	- Performed in DistillerSR [15]
	- Piloted title/abstract (n = 100) and full-text screening (n = 25), conflicts resolved through discussion
	- Liberal accelerated[58] screening for titles and abstracts
	- Dual-independent screening based on full text, with conflicts resolved through discussion
Data charting	- Performed in DistillerSR [15]
	- Piloted extractions (n = 5), conflicts resolved through discussion
	- One reviewer extracted studies, a second reviewer verified all extracted data, conflicts resolved through discussion
Data synthesis	- Rapid review characteristics and studies' references exported to MS Excel 2016
	- Definitions imported into NVivo (version 12) for coding into themes

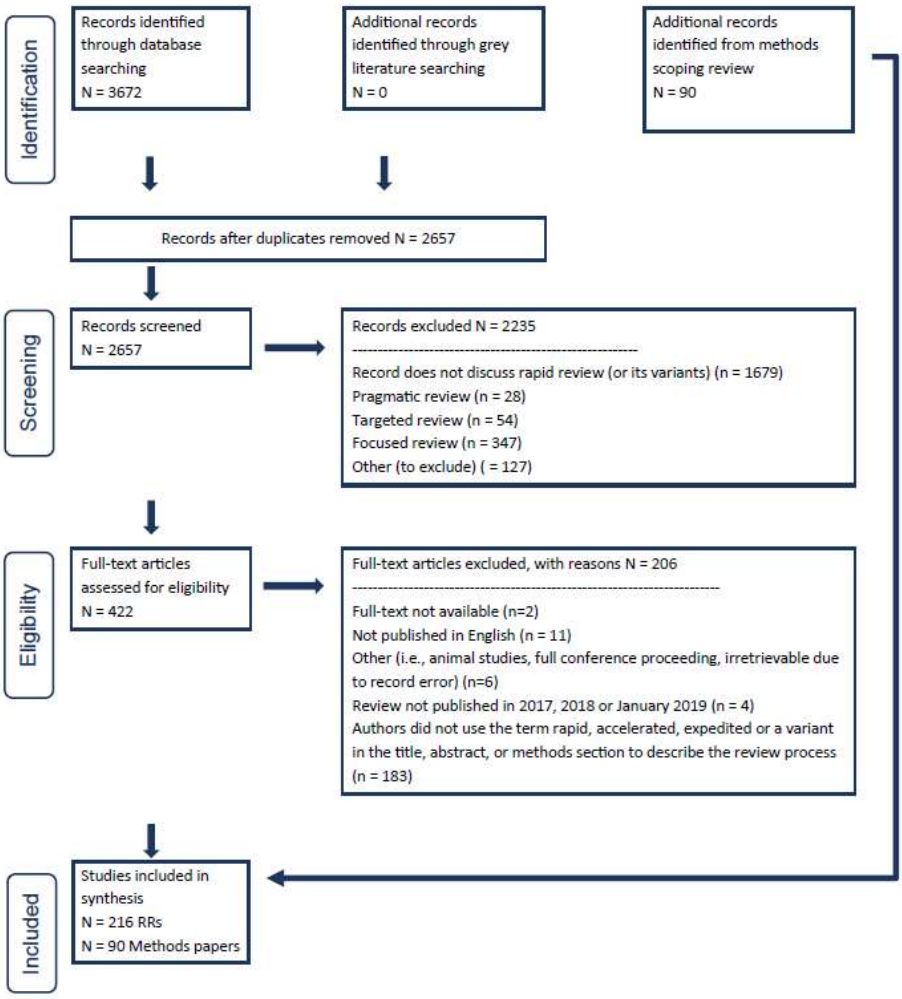


Fig. 1. PRISMA flow diagram.

definitions were cited, with 21 articles providing their own definition (with or without a reference to other articles) and 50 unique articles. Among the 21 articles that provided their own definition, 10 are those that are often referenced in the RRs [18,20–28]. Methods articles referenced an average of 2.22 references (median: 1, range: 0 to 10) (Appendix E.2). The top four articles referenced were Ganann 2010 [20] (n = 27), Khangura 2012 [2] (n = 27), Khangura 2014 [29] (n = 14), and Polisena 2015 [30] (n = 11). The other top articles in the RRs not in the top four here, Tricco 2015 [3] and Grant 2009 [18], were referenced 10 and four times, respectively.

There was overlap between the articles cited in the RRs and methods articles. Across both data sources, there were 79 unique citations, with 30 citations included in both scoping reviews, 29 unique to the RRs, and 20 unique to the methods articles. Among the citations found in only one of the two data sources, most were only referenced one or two times (highlighted in Appendix F).

4.3. Thematic analysis

All definitions from the RRs and the methods articles were thematically analyzed in NVivo. We identified eight major themes (Figure 2). Among the 204 articles that reported definitions (75 did not provide a definition and 27 RRs cited other studies with no identifiable themes), the most common themes were theme 4: *Compare and contrast*

Table 2. Rapid review characteristics

Rapid review characteristics	Rapid reviews (*N* = 216)
Year published[a]	
2017	101 (46.7%)
2018	106 (49.1%)
2019	9 (4.2%)
Countries of the corresponding author	
UK	82 (38.0%)
Australia	41 (19.0%)
USA	31 (14.4%)
Canada	24 (11.1%)
Ireland	6 (2.8%)
Italy	5 (2.3%)
Germany	4 (1.9%)
Denmark, South Africa, Switzerland	3 (1.4%) (each)
Finland, India, Spain	2 (0.9%) (each)
Japan, Korea, Nepal, Norway, Poland, Sweden, Taiwan, Thailand & UK	1 (0.5%) (each)
Terminology used (first mentioned in RR)	
Rapid review	136 (63.0%)
Rapid evidence assessment	22 (10.1%)
Rapid systematic review	19 (8.8%)
Rapid evidence review, rapid literature review (each)	12 (5.6%)
Systematic rapid evidence assessment, systematic rapid review	2 (0.9%) (each)
Abbreviated review, rapid appraisal, rapid best-fit framework synthesis, rapid evidence-based review, rapid evidence summary, rapid evidence synthesis, rapid meta-review, rapid qualitative review, rapid response review, rapid structured evidence review, rapid synthesis	1 (0.5%) (each)
Terminology first mentioned in	
Title	141 (65.3%)
Abstract	75 (34.7%)
References	
Total	290
Unique references/citations	59
Median (Range)	2 (0 to 7)
Mean	1.34
Top references	
Khangura 2012	54
Gannan 2010	42
Tricco 2015	21
Grant 2009	18
Number of references	
0	59 (27.3%)
1	84 (38.9%)
2	35 (16.2%)
3 or more	38 (17.6%)

[a] Articles may be an Epub before print with the print date after January 2019. Years published are taken as of the search date (December 20, 2018).

to SRs (68.1%; 139 of 204), theme 2: *Variation in shortcut methods* (54.9%; 112 of 204), with theme 1: *Accelerated/ rapid process* and theme 6: *Resource efficiency rationale* tied (48.5%; 99 of 204 each) (Figure 3). Definitions often covered more than one of these themes, with a range of 1 to 8 (median: 3; mean: 3).

4.3.1. Theme 1: Accelerated or rapid process/approach

The terms accelerated, streamlined, quickly or rapid were used in terms of the speed or timing for the overall approach to completing the RR. For example, "rapid reviews have been described as a streamlined alternative to standard systematic reviews [31]." [32].

Table 3. – Top four study cited and their definitions

Study	Definition	Cites
Khangura 2012[a]	"Given this lack of definition and evolving landscape, we have abstained from applying the label 'rapid review' to our KTA syntheses, and have alternatively called them 'evidence summaries'. Despite this, we consider our evidence summaries to be part of the continuum of rapid reviews, as conceptualized by Ganann and colleagues".	Ganann 2010[b]
Ganann 2010[b]	"Rapid reviews are literature reviews that use methods to accelerate or streamline traditional systematic review processes."	None
Tricco 2015[c]	"...we used the following working definition, 'a rapid review is a type of knowledge synthesis in which components of the systematic review process are simplified or omitted to produce information in a short period of time."'	Khangura 2012[a]
Grant 2009[d]	"They aim to be rigorous and explicit in method and thus systematic but make concessions to the breadth or depth of the process by limiting particular aspects of the systematic review process."	Butler 2005[e]

[a] Khangura et al. Syst Rev. 2012; 1:10.
[b] Ganann et al. 2010. Implement Sci. 2010; 5:56.
[c] Tricco et al. 2015. BMC Med. 2015; 13:224.
[d] Grant & Booth. 2009. Health Info Libr J. 2009; 26(2):91-108.
[e] Link to Butler 2005 no longer active, update Burton 2007.

4.3.2. Theme 2: Variation in methods shortcuts

There were a variety of words used to describe the shortcuts used in the methods, including streamlined, restricted, pragmatic, abbreviated, modifications, concessions, expedited, simplifying, constraints, truncated, modified or omitted steps, and limiting. The variety of words relate to the lack of a standardized approach in which steps these were applied, with some definitions providing examples on which steps of the review process these shortcuts would be applied. For example, "Major sources of streamlining can include narrowing the scope of the review questions; limiting literature search databases; the use of single (vs. dual) abstract and full-text screening; reducing the extent of data abstraction; omitting risk of bias/quality appraisal; and restricting the extent of the synthesis [33]." [34].

4.3.3. Theme 3: Focus/depth/breadth of scope

Similar to the theme 2, this was more specific to the topic, scope, or question being addressed in the RR rather than the methodology. For example, "Rapid review is an evidence synthesis methodology that applies a systematic approach to evidence identification and syntheses, but with a more limited scope than a systematic review." [35].

4.3.4. Theme 4: Compare and contrast to a full traditional systematic review

Definitions often included a comparison or related RRs to full SRs but provided an explanation in the text as to the difference in general between an SR and the RR. For example, "A rapid structured review differs from a

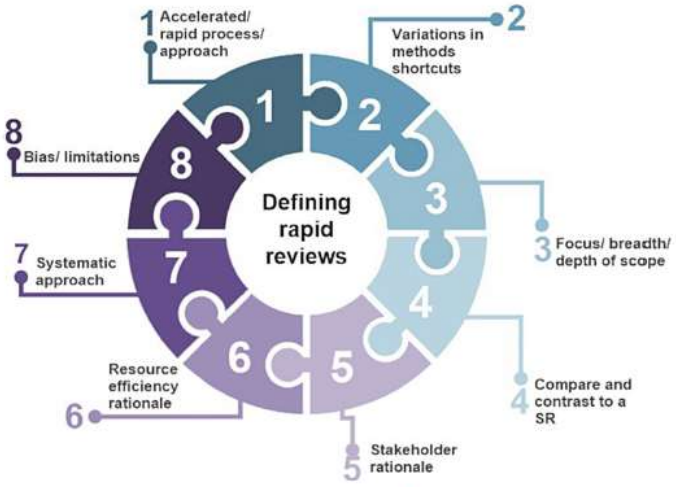


Fig. 2. Eight key themes in defining RRs.

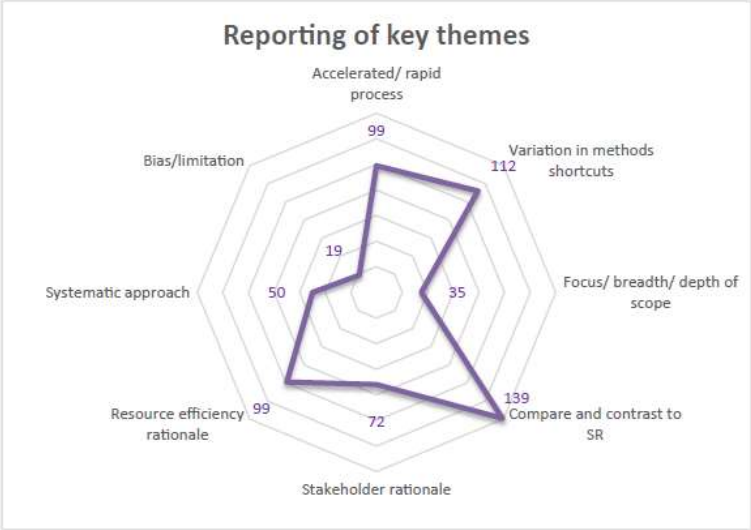


Fig. 3. Frequency of reporting of key themes.

systematic review in relation to the extensiveness of the search and methods used to undertake the analysis [36]." [37].

4.3.5. Theme 5: Stakeholder rationale

Many definitions referenced performing an RR to inform policy practice or to meet the needs of stakeholders, including decision makers (e.g., health professionals) and consumers. For example, "Rapid reviews are an emerging type of knowledge synthesis which aims 'to inform health-related policy decisions and discussions, especially when information needs are immediate' [38]." [39].

4.3.6. Theme 6: Resource efficiency rationale

Definitions often referred to RRs being performed because of resource constraints, including cost, human resources, time, and expertise. The difference between completing a review in a timely way (theme 1) and completing a review in a limited time frame is around the requirement of completing the review, rather than at the speed (e.g., rapidly, timely). For example, "Rapid reviews use systematic review methods to search and critically appraise existing research within limited resource and time constraints [40]." [41].

4.3.7. Theme 7: Systematic approach

Although RRs take shortcuts, several definitions stated that they remain systematic, transparent, rigorous, replicable, explicit, robust, using scientific methods. For example, '"Rapid reviews' are knowledge synthesis in which components of the systematic review process are simplified or omitted, to produce information in a timely manner, while retaining rigor in the selection and appraisal of studies [2,20,22]." [42].

4.3.8. Theme 8: Bias/limitations

There was some discussion around the bias that may be introduced due to shortcuts. Although there are few studies that formally evaluate RRs compared with full SRs, there is a potential for bias and limitations when using shortcuts. For example, "although potential biases related to streamlining procedures must be acknowledged [2]." [43].

4.4. Suggested definition

As there is not one common set of methods shortcuts that can be taken when conducting an RR, there may not be one common definition for an RR. As such, we suggest the following broad definition, which meets a minimum set of requirements identified in the thematic analysis, which will also be used to seek further consensus from the systematic review community.

"A rapid review is a form of knowledge synthesis that accelerates the process of conducting a traditional systematic review through streamlining or omitting a variety of methods to produce evidence in a resource-efficient manner."

This definition covers the most common themes (i.e., 1, 2, 4, and 6) that were identified in approximately 50% or more of the RRs and methods articles. By using broad words like resources, this definition captures the time element, as well as cost and human elements. Users could

then tailor this definition accordingly to best meet their individual remit and mandates for producing RRs by adding additional details covered in other themes. For example, if an organization produces RRs only when stakeholders make a request (theme 5), it can be modified to include this requirement.

"A rapid review is a form of knowledge synthesis that accelerates the process of conducting a traditional systematic review through streamlining or omitting a variety of methods to produce evidence for stakeholders in a resource-efficient manner."

Likewise, if the systematic aspect (theme 7) of RRs is important, the definition can be further modified.

"A rapid review is a rigorous and transparent form of knowledge synthesis that accelerates the process of conducting a traditional systematic review through streamlining or omitting a variety of methods to produce evidence for stakeholders in a resource-efficient manner."

4.5. Collaboration among RR definition references

It was common for RR definitions to use secondary referencing (i.e., quoting or paraphrasing from a source which is mentioned in another text) [44]. As this was not the primary objective of this scoping review, this is further discussed in Appendix G for the interested reader.

5. Discussion

To the best of our knowledge, this is the first systematically developed repository of RR definitions and an analysis of their major common themes. Eight key themes were identified, with the four most common themes being *comparing and contrasting RRs to SRs*, *variation in shortcut methods*, with *accelerated/rapid process or approach,* and *resource efficiency rationale* tied for third. As a criterion for inclusion was the use of the term rapid or derivative and the goal is to conduct the review rapidly regardless of which stages of conduct are abbreviated/omitted, it is not surprising that one of the key themes was around the accelerated or rapid approach.

As previously mentioned, some of the problems of lacking a common definition are around the difficulty in identifying RRs correctly and in having a homogenous set of products under different names. Among the RRs included in this scoping review, 18 different terms were used (Table 1), with an additional 23 terms, that may be considered derivatives, excluded (for feasibility) when screening titles/abstracts (Appendix C). Although the term 'rapid review' seems to be the generally adopted term, 'rapid' points to the speed of the process and not necessarily the methods in which this is achieved. Recently, the term 'restricted review' has been suggested to better capture the restrictions in the methods [8,9]; however, this does not relate to the speed of production. A common term for labeling these products may not be feasible, as many organizations have already adopted different terms for the same types of products. However, a definition with central tenets may help producers of these reviews to identify their research for easy identification, regardless of the term used to describe the review. The importance of defining (vs. labeling) is further supported in the Cochrane Handbook and MECIR, which state that study design labels may be ambiguous, and a focus on the defining features of the study is more important than the label [45,46].

It is better to rely on the original source of the information than to rely on the wording of another author who may impose their own interpretation or meaning [44]. Although 48 unique references were cited in the RRs, there is a high level of secondary referencing, as displayed in the collaborative map (Appendix G. Figure 1), many pointing to the same smaller set of studies. Therefore, in the context of developing a definition for RRs (and/or a minimum set or criteria/central tenets), the number of definitions used and cited may not be as extensive as what the results from this scoping review demonstrate. Using the suggested definition from this scoping review, and the key citations for additional support, may help lessen the 'noise' of what has been used and help guide future research in this area.

When comparing the key themes identified in this scoping review to related research, we see there are some similarities. Kelly et al. (2016) identified seven defining characteristics of RRs through a Delphi process [4]. However, there were some limitations to this process as only 1 reviewer selected the included studies and it is unclear how the initial survey was developed. In addition, the search was run in December 2014, and the progression in the amount of research evaluating RR methods, methodological development and guidance, and an increase in the number of published RRs has grown since this time [14]. This initial work provides a solid foundation to which this methodologically robust scoping review builds on using a more contemporary sample. We were able to map six of these seven defining characteristics to the themes we identified (Table 4). The only key theme not covered is theme 3 related to the focus/breadth or depth of the scope. The only defining characteristic of an RR, identified by Kelly et al., that could not be related to one of the key themes identified in this scoping review was that "rapid reviews have a protocol describing objectives, scope, PICO, and approach", although this is more around the process of developing an RR and less around defining it. Furthermore, Hartling et al. [47] identified 36 rapid products from 20 organizations and concluded that there is extensive variability in products labeled as RRs, but that the range of methods used in developing these products is driven by and supported by close and ongoing communication between the producers of the review and the end user, a concept captured by key themes 2 and 5.

To date, only one definition has emerged at the center [20]: "Literature reviews that use methods to accelerate

Table 4. Kelly defining characteristics compared with themes identified

Kelly et al. defining characteristics	Key theme(s)
Rapid reviews are conducted in less time than a systematic review	Theme 1: Accelerated/rapid process or approach
	Theme 4: Compare and contrast to SR
Rapid reviews use a spectrum of approaches to complete an evidence synthesis related to a defined research question(s) using the most systematic or rigorous methods as a limited time frame allows	Theme 2: Variation in methods shortcuts
	Theme 6: Resource efficiency rationale
	Theme 7: Systematic approach
Rapid reviews should have a protocol describing objectives, scope, PICO, and approach	None
Rapid reviews should tailor the explicit, reproducible methods conventionally used in a systematic review in some manner to expedite the review process	Theme 1: Accelerated/rapid process or approach
	Theme 2: Variation in methods shortcuts
	Theme 4: Compare and contrast to SR
	Theme 7: Systematic approach
Rapid reviews should transparently report methods and findings with a level of detail needed to adequately answer the research question, meet the requirements of the decision maker commissioning the review, and inform the audience for which the review is intended, while meeting a delivery time line agreed on in advance.	Theme 5: Stakeholder rationale
	Theme 6: Resource efficiency rationale
	Theme 7: Systematic approach
Rapid reviews should be considered in the context of the decision at hand when emergent or urgent decisions are required.	Theme 6: Resource efficiency rationale
Choices to adapt workflow should be balanced against the yet undetermined impact to conclusions or validity of findings, and this risk should be communicated to the end user.	Theme 8: Bias/limitations

or streamline traditional systematic review processes". However, this definition does not specifically address variances in types of RRs produced across different contexts, which are likely driven by the mandate or scope of the organization or entity producing them. In addition, when comparing this definition to the eight themes identified in the thematic analysis, it covers three of the eight key themes: *accelerated/rapid process or approach* (theme 1), *variation in methods shortcuts* (theme 2), and *compare and contrast to traditional systematic reviews* (theme 4). As this definition is from 2010, and RRs have been evolving over time, one might expect that it would not cover all key themes.

5.1. Implications for future research

Despite the increased use of RRs in policymaking [48,49], to date, there is no agreed-on definition on what constitutes a 'rapid review'. Yet, other areas of knowledge synthesis have developed definitions (e.g., what represents a systematic review update, scoping reviews) [50,51] that have been agreed on by the broader knowledge synthesis community. Several other groups and programs have developed their own definitions for RRs. For example, Crawford et al. 2015 describe the REAL© method, which "utilizes specific tools (e.g., automated online software) and standard procedures (e.g., rulebooks) to rigorously deliver more reliable, transparent and objective SRs in a streamlined fashion, without compromising quality and at a lower cost than other SR methods" [6]. The Department for International Development within the UK government has their own program and state on their website that "Rapid evidence assessments provide a more structured and rigorous search and quality assessment of the evidence than a literature review but are not as exhaustive as a systematic review". They can be used to "gain an overview of the density and quality of evidence on a particular issue, support programming decisions by providing evidence on key topics, and support the commissioning of further research by identifying evidence gaps" [17]. Based on the themes identified in this review, these definitions do not fully define RRs.

As a field of research, RRs need to at least develop a minimum set of criteria. If the concept of 'rapid review' is better defined, it will enable future studies of this methodology to be a clearly distinguishable approach, measurable to the extent possible, and understandable in terms of empirical observations. In a wider sense, researchers need to be able to describe what is and what is not a 'rapid review'. Until such a time that a general working definition is established, it may hinder efforts to promote the utility of RRs to end users who may benefit from more timely evidence to inform their decision-making. Lack of an agreed-on definition may also unfairly hamper acceptance of 'rapid reviews' by journal editors as a legitimate publication type and limit acknowledgment as a credible

academic output in terms of promotion and tenure of researchers who undertake them. It also results in authors producing a variety of products which are labeled under a wide array of names, contributing to the lack of cohesion and unity around the method. Furthermore, having a widely accepted definition may facilitate the future funding of 'rapid reviews' by granting agencies. More generally, a definition would facilitate discussion about RRs and would improve understanding by end users. Collectively, this highlights the need for an evidence-informed definition of RR which can be adopted by researchers.

5.2. Strengths and limitations

This study provides a repository of existing definitions identified in the current literature, identifies general themes, and provides a flexible working definition of RRs to be used by the wider knowledge synthesis community. In addition, through our collaborative mapping, this study has allowed us a first glance at the network of RR researchers who, through their RR and methods work, have provided and cited defining features of RRs.

However, there were some limitations. First, for feasibility, only English journal–published RRs identified in the databases that were searched were captured. The purpose of this scoping review was not to identify all RRs written in the included time period, but rather to get a sense of what definitions are currently being used. We included definitions from 216 RRs and supplemented these with the definitions from 90 RR methods articles. It is likely that RRs not captured would use definitions that would fall under the eight key themes identified. Second, as the main purpose of this review was to extract definitions verbatim from RRs, some information was not extracted (e.g., funding source of RR), as suggested by PRISMA-ScR, or was only extracted by 1 reviewer (e.g., the country of the corresponding author). In addition, in some cases, citations may not have specified a definition, but rather alluded to a component of that definition. For example, "Rapid review is a fairly new approach which has inherent strengths and limitations [2,20,28,52–54]." [55]. We did not delve into each reference to see which provided a definition and which were studies that evaluated the inherent strengths and limitations of RRs, but rather captured it in its entirety. In other cases, the reference provided was not specific to RRs but pointed to a methodology that was followed: "We conducted a rapid systematic literature review after a priori developed protocol [56]." [57]. It is therefore possible that some of the references may not actually provide a definition for RRs but instead may contain the methods of RRs or rationale as to why one might conduct an RR. Third, several terms were identified during title and abstract screening, some of which may have been RRs but were not identified as such (Appendix C). Because of the number of records with these terms, they were excluded, for feasibility. Therefore, it is possible that some reviews may have been missed that would qualify as an RR.

6. Conclusion

Eight key themes were identified, which have been considered in developing a preliminary, broad definition of an RR. This suggested definition, with additional caveats and opportunity for flexibility, will help the systematic review community define their review with consistency, regardless of the label used to describe it. Failure to use a consistent definition, or at least a minimum set of criteria, will be a barrier to moving the science forward in this field.

CRediT authorship contribution statement

Candyce Hamel: Conceptualization, Funding acquisition, Data curation. **Alan Michaud:** Investigation, Validation, Writing - review & editing. **Micere Thuku:** Investigation, Validation, Writing - review & editing. **Becky Skidmore:** Data curation. **Adrienne Stevens:** Conceptualization, Funding acquisition, Writing - review & editing. **Barbara Nussbaumer-Streit:** Conceptualization, Funding acquisition, Writing - review & editing. **Chantelle Garritty:** Conceptualization, Data curation.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclinepi.2020.09.041.

References

[1] 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. J Clin Effectiveness 1997;2(2):51–6.
[2] Khangura S, Konnyu K, Cushman R, Grimshaw J, Moher D. Evidence summaries: the evolution of a rapid review approach. Syst Rev 2012;1:10.
[3] Tricco AC, Antony J, Zarin W, Strifler L, Ghassemi M, Ivory J, et al. A scoping review of rapid review methods. BMC Med 2015;13:224.
[4] Kelly SE, Moher D, Clifford TJ. Defining rapid reviews: a modified DELPHI consensus approach. Int J Technol Assess Health Care 2016;32(4):265–75.
[5] Tricco AC, Langlois EV, Straus SE. (editors). Rapid Reviews to Strengthen Health Policy and Systems: A Practical Guide. Geneva: World Health Organization; 2017. Available at https://www.who.int/alliance-hpsr/resources/publications/rapid-review-guide/en/. Accessed November 15, 2019.
[6] Crawford C, Boyd C, Jain S, Khorsan R, Jonas W. Rapid Evidence Assessment of the Literature (REAL(©)): streamlining the systematic review process and creating utility for evidence-based health care. BMC Res Notes 2015;8:631.
[7] CADTH Rapid Response Service: about the rapid response service. Available at https://www.cadth.ca/about-cadth/what-we-do/products-services/rapid-response-service. Accessed November 12, 2019.

[8] Aronson JK, Heneghan C, Mahtani KR, Pluddemann A. A word about evidence: "rapid reviews" or "restricted reviews"? BMJ Evidence-Based Med 2018;23(6):204−5.
[9] Pluddemann A, Aronson JK, Onakpoya I, Heneghan C, Mahtani KR. Redefining rapid reviews: a flexible framework for restricted systematic reviews. BMJ Evidence-Based Med 2018;23(6):201−3.
[10] Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005;8(1):19−32.
[11] Peters MDJ, Godfrey C, McInerney P, Baldini Soares C, Khalil H, Parker D. Chapter 11: scoping reviews. In: Aromataris E, Munn Z, editors. Joanna Briggs Institute Reviewer's Manual. https://reviewersmanual.joannabriggs.org/. Accessed October 5, 2018.
[12] Tricco AC, Lillie E, Zarin W, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Ann Intern Med* Published Online September 2018;4.
[13] McGowan J, Sampson M, Salzwdel DM, et al. PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Statement. J Clin Epidemiol 2016;75:40−6. https://doi.org/10.1016/j.jclinepi.2016.01.021.
[14] Hamel C, Michaud A, Thuku M, et al. Few evaluative studies exist examining rapid review methodology across stages of conduct: a systematic scoping review. J Clin Epidemiol 2020;126:131−40. https://doi.org/10.1016/j.jclinepi.2020.06.027.
[15] DistillerSR [Computer Program]. https://v2dis-prod.evidencepartners.com.
[16] Harrison R, Manias E, Mears S, Heslop D, Hinchcliff R, Hay L. Addressing unwarranted clinical variation: a rapid review of current evidence. J Eval Clin Pract 2019;25:53−65.
[17] Department for International Development. Rapid Evidence Assessments. GOV.UK. 2017. Available at https://www.gov.uk/government/collections/rapid-evidence-assessments. Accessed November 12, 2019.
[18] Grant MJ, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. Health Info Libr J 2009;26:91−108.
[19] Bagnasco A, Cadorin L, Barisone M, Bressan V, Iemmi M, Prandi M, et al. Ethical dimensions of paediatric nursing: a rapid evidence assessment. Nurs Ethics 2018;25(1):111−22.
[20] Ganann R, Ciliska D, Thomas H. Expediting systematic reviews: methods and implications of rapid reviews. Implement Sci 2010;5:56.
[21] Abou-Setta AM, Jeyaraman M, Attia A, Al-Inany HG, Ferri M, Ansari MT, et al. Methods for developing evidence reviews in short periods of time: a scoping review. PLOS ONE 2016;11:e0165903.
[22] Harker J, Kleijnen J. What is a rapid review? A methodological exploration of rapid reviews in Health Technology Assessments. Int J Evid Based Healthc 2012;10(4):397−410.
[23] Kelly SE, Moher D, Clifford TJ. Quality of conduct and reporting in rapid reviews: an exploration of compliance with PRISMA and AMSTAR guidelines. Syst Rev 2016;5:79.
[24] Schunemann HJ, Moja L. Reviews: Rapid! Rapid! Rapid! ... and systematic. Sys Rev 2015;4:4.
[25] Thomas J, Newman M, Oliver S. Rapid evidence assessments of research to inform social policy: taking stock and moving forward. Evid Policy 2013;9(1):5−27.
[26] Varker T, Forbes D, Dell L, Weston A, Merlin T, Hodson S, et al. Rapid evidence assessment: increasing the transparency of an emerging methodology. J Eval Clin Pract 2015;21:1199−204.
[27] Watt A, Cameron A, Sturm L, Lathlean T, Babidge W, Blamey S, et al. Rapid versus full systematic reviews: validity in clinical practice? 2008;78(11).
[28] Watt A, Cameron A, Sturm L, Lathlean T, Babidge W, Blamey S, et al. Rapid reviews versus full systematic reviews: an inventory of current methods and practice in health technology assessment. Int J Technol Assess Health Care 2008;(2):133.
[29] Khangura S, Polisena J, Clifford TJ, Farrah K, Kamel C. Rapid review: an emerging approach to evidence synthesis in health technology assessment. Int J Technol Assess Health Care 2014;30(1):20−7.
[30] Polisena J, Garritty CM, Kamel C, Stevens A, Abou-Setta AM. Rapid review programs to support health care and policy decision making: a descriptive analysis of processes and methods. Syst Rev 2015;4:26.
[31] Peterson K, Floyd N, Ferguson L, Christensen V, Helfand M. User survey finds rapid evidence reviews increased uptake of evidence by Veterans Health Administration leadership to inform fast-paced health-system decision-making. Syst Rev 2016;5.
[32] Coster JE, Turner JK, Bradbury D, Cantrell A. Why do people choose emergency and urgent care services? A rapid review utilizing a systematic literature search and narrative synthesis. Acad Emerg Med 2017;24(9):1137−49.
[33] Polisena J, Garritty C, Umscheid CA, et al. Rapid Review Summit: an overview and initiation of a research agenda. Syst Rev 2015;4:111.
[34] Patnode CD, Eder ML, Walsh ES, Viswanathan M, Lin JS. The use of rapid review methods for the U.S. Preventive services task force. Am J Prev Med 2018;54(1S1):S19−25.
[35] Chambers S, Barton KL, Albani V, Anderson AS, Wrieden WL. Identifying dietary differences between Scotland and England: a rapid review of the literature. Public Health Nutr 2017;20(14):2459−77.
[36] Centre for Reviews and Dissemination. Systematic Reviews: CRD's guidance for undertaking reviews in health care. 2009. Available at http://file:///C:/Dropbox/Dropbox/Systematic%20review/CRD%20-%20How%20to%20conduct%20a%20SR.pdf. Accessed September 25, 2018.
[37] Alison Rodriguez, Joanna Smith, Kirstine McDermid. Dignity therapy interventions for young people in palliative care: a rapid structured evidence review. Int J Palliat Nurs. (7):339.
[38] Lal S, Adair C. E-mental health: a rapid review of the literature. Psychiatr Serv 2014;65(1):24−32.
[39] Huband N, Furtado V, Schel S, et al. Characteristics and needs of long-stay forensic psychiatric inpatients: a rapid review of the literature. In: Andreasson AA, Baldwin B, Beer B, Braun B, Butwell C, et al, editors. The International Journal of Forensic Mental Health, 17 2018:45−60.
[40] Booth A, Sutton A, Papaioannou D. Systematic Approaches to a Successful Literature Review. 2nd ed. London, UK: SAGE Publications Ltd.; 2016.
[41] Clibbens N, Harrop D, Blackett S. Early discharge in acute mental health: a rapid literature review. Int J Ment Health Nurs 2018;27(5):1305−25.
[42] Naughton AM, Cowley LE, Tempest V, Maguire SA, Mann MK, Kemp AM. Ask Me! self-reported features of adolescents experiencing neglect or emotional maltreatment: a rapid systematic review. Child Care Health Dev 2017;43(3):348−60.
[43] Männistö II, Pirttimaa RA. A review of interventions to support the educational attainments of children and adolescents in foster care. Adoption & Fostering 2018;42(3):266−81.
[44] University of Leicester - Student Learning Development. 1.13 secondary reference. University of Leicester. Available at https://www2.le.ac.uk/offices/ld/resources/writing/harvard/content/1.13-secondary-reference. Accessed November 3, 2019.
[45] McKenzie J, Brennan S, Ryan R, Thomson H, Johnston R, Thomas J. Chapter 3: defining the criteria for including studies and how they will be grouped for the synthesis. Cochrane Handbook for Systematic Reviews of Interventions Version 6.0. 2019. Available at https://training.cochrane.org/handbook/current/chapter-03. Accessed July 11, 2020.
[46] Higgins J, Lasserson T, Chandler J, et al. Methodological Expectations of Cochrane Intervention Reviews (MECIR): Standards for the Conduct and Reporting of New Cochrane Intervention Reviews, Reporting of Protocols and the Planning. London, UK: Cochrane; 2019. Available at https://community.cochrane.org/mecir-manual. Accessed November 15, 2019.
[47] Hartling L, Guise J-M, Kato E, Anderson J, Belinson S, Berliner E, et al. A taxonomy of rapid reviews links report types and methods to specific decision-making contexts. J Clin Epidemiol 2015;68:1451−1462.e3.
[48] O'Leary DF, Casey M, O'Connor L, et al. Using rapid reviews: an example from a study conducted to inform policy-making. J Adv Nurs 2017;73:742−52.

[49] Lawani MA, Valera B, Fortier-Brochu E, Légaré F, Carmichael PH, Côté L, et al. Five shared decision-making tools in 5 months: use of rapid reviews to develop decision boxes for seniors living with dementia and their caregivers. Syst Rev 2017;6(1):56.
[50] Moher D, Tsertsvadze A. Systematic reviews: when is an update an update? Lancet 2006;367:881–3.
[51] Colquhoun HL, Levac D, O'Brien KK, Straus S, Tricco AC, Perrier L, et al. Scoping reviews: time for clarity in definition, methods, and reporting. J Clin Epidemiol 2014;67:1291–4.
[52] Abrami PC, Borokhovski E, Bernard RM, Wade CA, Tamim R, Persson T, et al. Issues in conducting and disseminating brief reviews of evidence. Evid Policy 2010;6(3):371–89.
[53] Australian Safety and Efficacy Register of New Interventional Procedures -Surgical. Rapid versus full systematic reviews: an inventory of current methods and practice in Health Technology Assessment. Report no 60. Stepney: Australian Safety and Efficacy Register of New Interventional Procedures -Surgical (ASERNIP-S). Report number 60 2007. Available at https://www.crd.york.ac.uk/CRDWeb/ShowRecord.asp?ID=32007000667&ID=32007000667. Accessed November 18, 2019.
[54] Gough D, Thomas J, Oliver S. Clarifying differences between review designs and methods. Syst Rev 2012;1:28.
[55] Bhardwaj K, Locke T, Biringer A, Booth A, Darling EK, Dougan S, et al. Newborn Bilirubin screening for preventing severe Hyperbilirubinemia and Bilirubin encephalopathy: a rapid review. Curr Pediatr Rev 2017;13(1):67–90.
[56] Hartling L, Guise J-M, Kato E, et al. EPC Methods: An Exploration of Methods and Context for the Production of Rapid Reviews. Rockville (MD): Agency for Healthcare Research and Quality (US); 2015.
[57] Aronow WS, Shamliyan TA. Comparative effectiveness and safety of Rivaroxaban in adults with nonvalvular atrial fibrillation. Am J Ther 2018.

ELSEVIER

Journal of Clinical Epidemiology 126 (2020) 131–140

Journal of Clinical Epidemiology

REVIEW

Few evaluative studies exist examining rapid review methodology across stages of conduct: a systematic scoping review

Candyce Hamel[a,b,*], Alan Michaud[a], Micere Thuku[a], Lisa Affengruber[c], Becky Skidmore[a], Barbara Nussbaumer-Streit[c], Adrienne Stevens[a], Chantelle Garritty[a,b]

[a]*Ottawa Hospital Research Institute, Knowledge Synthesis Group, Ottawa, Ontario K1H 8L6, Canada*
[b]*School of Medicine, University of Split, Split, Croatia 21000*
[c]*Cochrane Austria, Department for Evidence-based Medicine and Evaluation, Danube University Krems, Krems, Austria*

Accepted 23 June 2020; Published online 26 June 2020

Abstract

Objectives: The objective is to identify studies that have assessed methodological shortcuts for undertaking rapid reviews (RRs) and mapping these to review conduct stages and Methodological Expectations of Cochrane Intervention Reviews (MECIR) guidance.

Study Design and Setting: We conducted a systematic scoping review. We searched multiple databases (e.g., MEDLINE, Embase), which were supplemented by grey literature searching. Methods were defined *a priori* in a published protocol.

Results: Out of 1,873 records, 90 publications were divided into four RR categories: formal evaluation ($n = 14$), development, which included four subcategories ($n = 65$), comparison ($n = 2$), and applying reporting guidelines/critical appraisal tools ($n = 3$), and a systematic review surrogate category ($n = 6$). Four formal evaluation studies were composite evaluations, including more than one shortcut simultaneously. The remaining 10 studies evaluated viable (e.g., including English-only publications) and unviable (e.g., single-reviewer screening) shortcuts, covering five key dimensions and five 'other' (e.g., involving stakeholders) considerations while conducting a review. Because of complexities around shortcuts evaluated, only a cursory mapping to MECIR criteria was possible.

Conclusion: Some methods shortcuts may be valid in the context of RRs, but limitations in the studies may limit their applicability. The results will serve to inform discussions within Cochrane regarding possible future implementation of RRs. © 2020 Elsevier Inc. All rights reserved.

Keywords: Rapid reviews; Methodology; Shortcuts; Formal evaluations; Abbreviated methods; Scoping review

1. Introduction

Systematic reviews (SRs), considered the gold standard in collating all available evidence related to a specific question, have been used to inform policy for health care and public health since the early 1990s [1] and are considered to be essential to produce trustworthy guidelines [2]. However, SRs are time- and resource-intensive undertakings which may not meet the needs of those requesting them [3]. An analysis of 197 reviews registered in PROSPERO reported that SRs take an average of 67.3 weeks (range: 6–186) to conduct (from registration to publication) [4]. Often conducted to meet the needs of stakeholders (e.g., policy makers, health care professionals, and consumers), a rapid review (RR) is a form of knowledge synthesis that accelerates the process of conducting a traditional SR through streamlining or omitting a variety of methods to produce evidence in a timely and resource-efficient manner [5]. Length of time to conduct a review cannot be the defining feature to differentiate between an SR and an RR. An SR which yields few to no studies may be conducted in a short timeframe or a review with many reviewers may be completed quickly. Recently, some have suggested that RRs should instead be called 'restricted systematic reviews' [6], to account for the rapidity of the process and the restriction around the methods of conduct.

Funding: This work is supported in part by funds provided by Cochrane and in part from a Canadian Institutes of Health Research, Canada grant (funding research number 142310). The funders had no role in the development of this manuscript and its protocol (registered on OSF: https://osf.io/dekx6/). They did not have a role in the data collection, analyses, interpretation of the data, and publication of the findings.

Declaration of interest: None.

* Corresponding author: Candyce Hamel, Ottawa Hospital Research Institute, 501 Smyth Road, Ottawa, Ontario, Canada K1H 8L6.

E-mail address: cahamel@ohri.ca (C. Hamel).

https://doi.org/10.1016/j.jclinepi.2020.06.027
0895-4356/© 2020 Elsevier Inc. All rights reserved.

What is new?

Key findings

- Of 90 included studies, 14 formally evaluated rapid review shortcuts.
- Ten studies evaluated single shortcuts, covering five key dimensions: literature search limits, number of databases searched, gray literature, study selection/screening, and data extraction. Other areas evaluated were involvement of stakeholders, eligibility criteria, and peer review of the search strategy.

What this adds to what was known?

- Based on conclusions of the authors of the primary studies, some shortcuts (e.g., English-only publications, data extraction from existing reviews) may be viable rapid review shortcuts, whereas others (e.g., searching Embase only, single-reviewer screening) are not.
- The results from this scoping review provide a comprehensive repository on research performed in the area of rapid review shortcuts and other research that describes rapid review methods.

What is the implication and what should change now?

- Researchers should evaluate additional single shortcuts that would have the greatest impact on resource-intensive stages (e.g., title and abstract screening, data extraction, and risk of bias assessment).

Cochrane, a leading organization producing high-quality SRs, describes an SR as a review that "attempts to identify, appraise, and synthesize all the empirical evidence that meets prespecified eligibility criteria to answer a specific research question" [7]. Review authors have access to a handbook [8] on how to perform each stage of the review, as well as the Methodological Expectations of Cochrane Intervention Reviews (MECIR) guideline [9], which informs reviewers on both the mandatory and highly desirable processes when performing an SR. However, there is little evidence to support many of these suggested processes [10]. When conducting an RR, the options for shortcuts seem infinite. A recent survey conducted by the Cochrane Rapid Review Methods Group (RRMG) asked respondents for input on a variety of shortcut approaches in several areas of conduct (e.g., searching, study selection, and data extraction). In the areas of study selection, data extraction, and risk of bias alone, respondents were presented with 18 different shortcut approaches, and this list of options was by no means comprehensive (internal unpublished Cochrane RRMG report). In fact, a scoping review of RR methods identified 50 unique RR methods among 82 RRs, including omitting gray literature searching, applying language restrictions, and having one reviewer screen titles and abstracts [11].

SRs aim to minimize bias by using explicit, systematic methods [12]. The shortcuts used to produce RRs may introduce bias. However, a review published in 2016 on the methodologies for RRs concluded that the poor quality of the studies evaluating this does not allow for firm conclusions to be made [13]. Our objective was to conduct a systematic scoping review of the literature assessing one or more method(s) applicable for undertaking RRs (e.g., single reviewer screening vs. double reviewer screening) or comparing the results of RRs to those of SRs (e.g., do conclusions change?) across all stages of conduct. We provide a comprehensive summary of abbreviated methods and their validity. The results from this scoping review were used to inform the survey described above, and will guide the discussions for the operationalization of what abbreviated methods would be acceptable for use within Cochrane, map methods shortcuts identified from the studies to evaluate if RR shortcuts align with Cochrane methods guidance (e.g., MECIR), and identify gaps in knowledge.

2. Methods

This systematic scoping review was guided by established scoping review methodology [14,15]. It has been prepared in accordance with the Preferred Items in Systematic Reviews and Meta-Analysis extension for Scoping Reviews (PRISMA-ScR) [16]. A protocol for this work was registered on the Open Science Framework (OSF: https://osf.io/dekx6/). Methods are briefly described in Table 1, with additional details and deviations from the protocol in Appendix A.

3. Results

3.1. Search results

After removing duplicates from the two searches and adding results from grey literature searching, 1,873 unique references were screened based on title and abstract. Among these, 156 were further evaluated at full text and 90 studies were included (Fig. 1). Studies were primarily conducted in Canada (37.8%, 34 of 90), the United Kingdom (21.1%, 19 of 90), and Australia (14.4%, 13 of 90). The majority of the studies (75.6%, 68 of 90) were published in 2014 or later (Table 2). Most formal evaluation studies were published since 2017 (78.6%; 11 of 14).

3.2. Categorizing RR studies

Although the primary objective of this scoping review was to identify studies that evaluated shortcut methods in

Table 1. Methods (in brief)

Project stage	Method description
Eligibility criteria	• Methods studies that evaluated shortcut approaches that could be applied or related to RR stages of conduct (Appendix B) • Written in English (for feasibility) • Published or identified through grey literature since 1997
Searching for studies	• Developed by an experienced information specialist with input on search terms by members of the research team • Focus on interventional RR methods • Peer-reviewed using the PRESS checklist [17] • Original search (Jan 2019): MEDLINE® ALL, Embase Classic + Embase, PsycINFO, ERIC, Cochrane Library, CINAHL, Web of Science, Epistemonikos (Appendix C.1) • Supplemental search (Feb 2019): MEDLINE® ALL, Embase Classic + Embase, PsycINFO and ERIC (Appendix C.2 & C.3) • Search strategies not restricted by language • Additional searching: grey literature (e.g., organizations that produce RRs), bibliographies of included studies, contacting experts in the field, bibliography of Robson 2018 study [10]
Study selection	• Performed in stages due to large yield of first search • Performed in DistillerSR [18] • Piloted title/abstract and full-text screening, conflicts resolved through discussion • Liberal accelerated [19] screening for titles and abstracts • Dual-independent screening based on full text, with conflicts resolved through discussion • Artificial intelligence tool used to help screen titles and abstract
Data charting (Appendix D)	• Piloted extractions (n = 5), conflicts resolved through discussion • One reviewer extracted studies, a second reviewer verified all extracted data, conflicts resolved through discussion
Data synthesis	• Formal evaluative studies: • Two reviewers mapped the studies into four categories (partially informed by Tricco et al. 2015[11]) (Fig. 2) • Studies that formally evaluated shortcut methods used in the RR context were mapped back to the stage of conducts (Appendix B) to identify gaps (Fig. 3A and 3B), and are presented narratively and in Appendix E • Each shortcut was compared with the MECIR guidelines for Cochrane reviews to see whether it met the MECIR criteria (Appendix F) • Other categories are narratively described and presented in tables (Appendix G)

RRs, we also identified studies that described RR methods. We felt it was important to provide some detail around these studies for the purpose of building a repository of RR information. Therefore, after identifying all studies, we performed the following:

1. The RR studies were labeled as belonging to one of four categories (i.e., formal evaluation, development, comparison, applying reporting guidelines/critical appraisal tools) based on the nature of the study (Fig. 2; Table 2). It is possible that studies may have fit into more than one category, so we used the main focus of the study to assign the most appropriate category. An additional six studies were labeled as SR surrogates (i.e., studies that evaluated methods in SRs that may be transferrable to RRs), which were supplemented with those identified in the Robson review [10]. This list of surrogate studies may not be comprehensive, as it was not the purpose of the search of this scoping review.
2. The studies that formally evaluated RR methods (category 1) were i) summarized in a table (Appendix E), ii) mapped to the key dimensions (Appendix F) and then charted (Fig. 3A and 3B), iii) compared with MECIR criteria (Appendix F), and iv) narratively synthesized.

3.3. Category one: formal evaluation studies

Fourteen studies were identified [20–33] in which a formal evaluation has been performed either through evaluating the impact of shortcuts within the conduct of an RR (e.g., title only screening), comparing different versions of the same shortcuts within the conduct of an RR (e.g., number of databases searched), comparing the results/conclusions of RRs to those of SRs (e.g., including only the largest trial), or evaluating the impact of including 'best-practice' methods (e.g., peer review of search) (Appendix E.1). One additional study is being conducted by the Cochrane Austria group (not yet completed) (Appendix E.2).

The 14 studies addressed nine key dimensions related to the conduct phases of a review (Fig. 3A), in addition to some areas of evaluation not included in the preliminary list of key dimensions (Appendix B). These are presented in the "other evaluations" section below. Although there are 33 instances of evaluation, only 16 (48.5%) were from studies that evaluated the shortcut independent from other

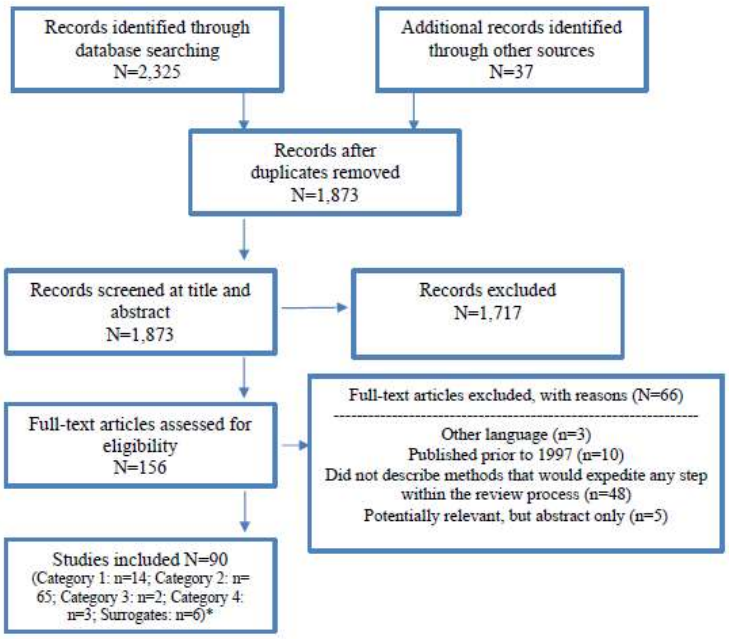


Fig. 1. PRISMA flow diagram.

shortcuts. The remaining were from studies which have been labeled as 'composite evaluations' [20,23,32,33], in which more than one methodological shortcut was taken, and any difference in the results may be attributable to one or several of the shortcuts. In other words, how these comparisons were performed would not allow determination of which shortcuts contributed to differences in the results, if any.

Studies have been categorized either as 'single' or 'composite' evaluation studies. The single evaluation studies have been separated by the key dimension (Appendix B) and are presented in Fig. 3B.

3.3.1. 'Single' evaluation studies

3.3.1.1. Literature search limits. Marshall recalculated the meta-analyses (MAs) of 2,512 SRs with a total of 16,088 included studies to measure the impact of excluding articles older than 5, 7, 10, 15, and 20 years before the search date [24]. Limiting the search to the last 5 years found 24.9% (n = 4,004) of the included studies and had the greatest impact, with 82% of the pooled effect estimates having a 5% or greater change. *Author's conclusion: All date limits evaluated led to small or greater changes exceeding the 10% tolerated error rate described in the study by Wagner et al. [34] and may not meet the level of accuracy demanded.*

3.3.1.2. Number of databases searched/gray literature. Marshall recalculated the MAs of 2,512 SRs with a total of 16,088 included studies to measure the impact of removing any studies not identified in PubMed [24]. This restriction resulted in 88.6% (n = 14,255) of the studies being identified, 19% of pooled effect estimates having a 5% or greater change, and 3.7% of the MAs losing all studies. *Author's conclusion: The PubMed-only search slightly exceeded the 10% tolerated error rate [34] (10.6% risk of 20% or greater change in results), and may be considered for scoping reviews (due to resource limitations) or where synthesis is needed urgently.*

Nussbaumer-Streit evaluated 840 abbreviated searches, combining a variety of database searches with or without gray literature searching (14 different search types on 60 reviews) and evaluated the proportion of changed conclusions [28]. Depending on the abbreviated search, the proportion of conclusions that deviated from the original conclusions ranged from 8% (MEDLINE + CENTRAL + Embase + Refs) to 27% (Embase only). *Author's conclusion: The decision on which abbreviated search to use will depend on the willingness of the decision maker to accept or not accept a lower degree of certainty when making conclusions and possibly making an opposite conclusion.*

Pham evaluated the impact of (i) including only the bibliographic database that yielded the highest number of records, plus the ancillary sources searched in the original SR/MA, and (ii) limiting the search to bibliographic databases, in three SRs. These two shortcuts were evaluated separately on how the omitted studies affected the

Table 2. Rapid review methods study characteristics

Rapid review methods study characteristics	All studies (N = 90)
Year published	
2020	2 (2.2%)[a]
2019	3 (3.3%)[b]
2018	13 (14.4%)
2017	17 (18.9%)
2016	15 (16.7%)
2015	14 (15.6%)
2014	4 (4.4%)
1997–2013	22 (24.4%)
Country of corresponding author	
Canada	34 (37.8%)
UK	19 (21.1%)
Australia	13 (14.4%)
Austria, USA	5 (5.6%) each
Germany; Norway; Canada & USA	2 (2.2%) each
Brazil; Chile; Ireland; The Netherlands; Uganda; Spain & Canada; Mexico & Australia; Canada & Switzerland	1 (1.1%) each
Category	
1. Formal evaluation	14 (15.6%)
2A. Development: metaresearch and impact	36 (40%)
2B. Development: programs and guidance	22 (24.4%)
2C. Development: terminology	2 (2.2%)
2D. Development: other	5 (5.6%)
3. Comparison	2 (2.2%)
4. Applying reporting guidelines/critical appraisal tools	3 (3.3%)
5. Surrogates	6 (6.7%)
Category 1: Formal evaluation	
	n = 14
Year published	
2020	2 (14.3%)
2019	2 (14.3%)
2018	3 (21.4%)
2017	4 (28.6%)
2016	1 (7.1%)
2010	1 (7.1%)
2008	1 (7.1%)
Country of corresponding author	
Australia	2 (14.3%)
Austria	4 (28.6%)
Canada	3 (21.4%)
Norway	1 (7.1%)
UK	4 (28.6%)

[a] These were identified by experts.
[b] Search run in January & February of 2019 and may not have captured all relevant studies published in 2019.

direction, magnitude, or precision of summary estimates [29]. Including only the highest yield database affected the highest number of MAs (n = 15), with one review missing 41.7% (15 of 36) of the studies. The omission of studies due to limiting the search to bibliographic databases resulted in less precise pooled estimates that did not differ in direction from the original estimate, when MA was still possible. *Author's conclusion: Depending on the features of the review (e.g., the specificity of the review question, population, and interventions), the impact of these shortcuts may differ.*

3.3.1.3. Screening. Gartlehner evaluated single- and dual-reviewer screening among 280 reviewers performing 24,942 screening decisions performed in two SRs [21]. Overall, single-reviewer screening missed 13.4% of the eligible studies, whereas dual-reviewer screening missed 2.5% of the studies. *Author's conclusion: Single-reviewer screening should not be used for SRs but might be a viable option for RRs.*

Gartlehner evaluated the accuracy of a machine-assisted screening, single-reviewer screening, and machine screening alone, on 2,472 abstracts [22]. Machine screening alone resulted in a sensitivity of 14% (95% confidence interval (CI) 0 to 31%). Single-reviewer and machine-assisted screening performed better, with sensitivities of 78% (95% CI 66 to 89%) and 78% (95% CI 66 to 90%), respectively. *Author's conclusion: The accuracy of the machine-assisted screening is not yet adequate to replace a human screener for SRs but might have greater utility for RRs.*

Pham evaluated the effect on the direction, magnitude, or precision of summary estimates, when title and abstract screening was performed by one reviewer in three SRs [29]. This was evaluated using two reviewers. Dependent on the reviewer, four of 21 studies were omitted and impacted four or 12 of the MAs. The omission of studies affected 39 of 143 possible MAs, of which 14 were no longer possible because of insufficient studies (<2). *Author's conclusion: Based on the number of possible missed studies, it is recommended to use two reviewers whenever possible.*

Rathbone evaluated participants, interventions and comparators-based title-only screening on screening effort and recall of relevant studies in 10 reviews [30]. The reduction in screening effort ranged from 11% to 78%, with a median reduction of 53%, and recall was 100% in 90% (9 of 10) of the reviews. In the 10th review, four of five reviewers missed the same included study. *Author's conclusion: Participants, interventions and comparators-based title-only screening reduced the workload of screening; however, it required a thorough workup to identify a list of synonyms and alternative terms.*

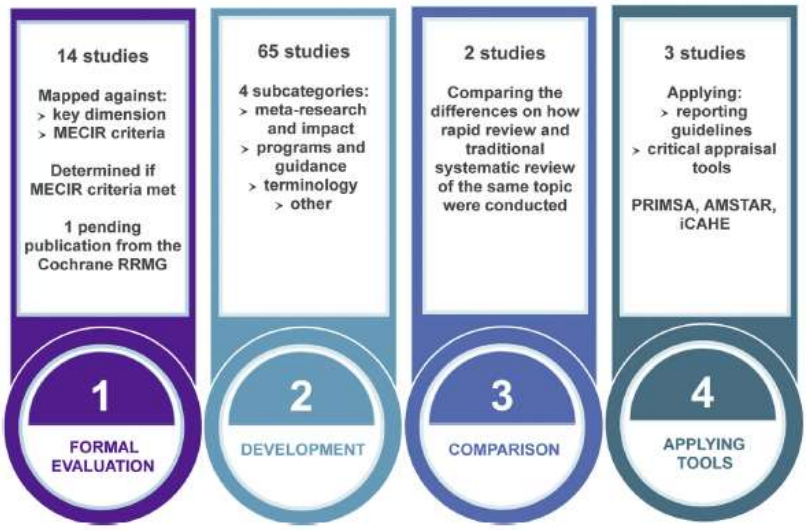


Fig. 2. Rapid review study categories.

3.3.1.4. Data extraction. Martyn-St James evaluated the accuracy of extracting data from an existing SR compared with extracting data directly from the primary studies [25]. Data were extracted by one reviewer and numerical data were checked by a second reviewer. The data in existing reviews were highly accurate, and findings and conclusions did not differ between methods. *Author's conclusion: Rapid reviewers should consider the methodological and reporting quality of existing reviews if these are going to be used as the primary source of data extraction.*

3.3.1.5. Other evaluations

3.3.1.5.1. Involving stakeholders. Moore evaluated the effect of including knowledge brokers in the review process and how this affected the clarity (e.g., purpose, scope method, and report format) of 60 RR proposals [26]. Knowledge brokering significantly improved the scores for all six questions addressing clarity and reviewers' confidence in meeting policy makers' needs. *Author's conclusion: This model of knowledge brokering may be an effective strategy for agencies commissioning rapid reviews and the researchers performing them.*

3.3.1.5.2. Eligibility criteria (size of study). Marshall recalculated the MAs of 2,512 SRs with a total of 16,088 included studies to measure the impact of excluding trials with fewer than 50, 100, and 200 participants, and using the largest trial only [24]. Excluding studies with fewer than 200 participants resulted in 44.7% of MAs losing all studies. Including only the largest trial captured only 15.6% (n = 2,512) of the included studies, and 66% of pooled effect estimates had a 5% or greater change. *Author's conclusion: All study size limits evaluated led to small or greater changes exceeding the 10% tolerated error rate described in the study by Wagner et al. [34] and may not meet the level of accuracy demanded.*

3.3.1.5.3. Eligibility criteria (language of publication). Nussbaumer-Streit identified 29 Cochrane reviews that included 80 non-English publications and evaluated if limiting to English-only publications affected the overall conclusions [27]. For 95% (38 of 40) of the outcomes, the exclusion of non-English studies did not markedly alter the size, direction of effect estimates, or statistical significance. The proportion of changed conclusions in this sample was 0.0% (95% CI 0.0 to 0.6%). *Author's conclusion: Exclusion of non-English publications had minimal impact on overall conclusions and could be a reliable methodological shortcut.*

3.3.1.5.4. Eligibility criteria (access to publications). Pham evaluated the effect on the direction, magnitude, or precision of summary estimates, when only including studies that were available electronically in three SRs [29]. This resulted in 16.7% (3 of 18) of the MAs in one review being affected. Two MAs were no longer possible as there was only one study remaining, and the other MA resulted in a larger standardized mean difference and a wider CI. *Author's conclusions: There was a decrease in the time and cost associated with ordering paper-only archives, but e-journals only became more widely available in the 1990s and early 2000s, which could impact missing studies.*

3.3.1.5.5. Search strategy peer review. Spry investigated the impact of the peer review of search strategies of 71 Canadian Agency for Drugs and Technologies in Health

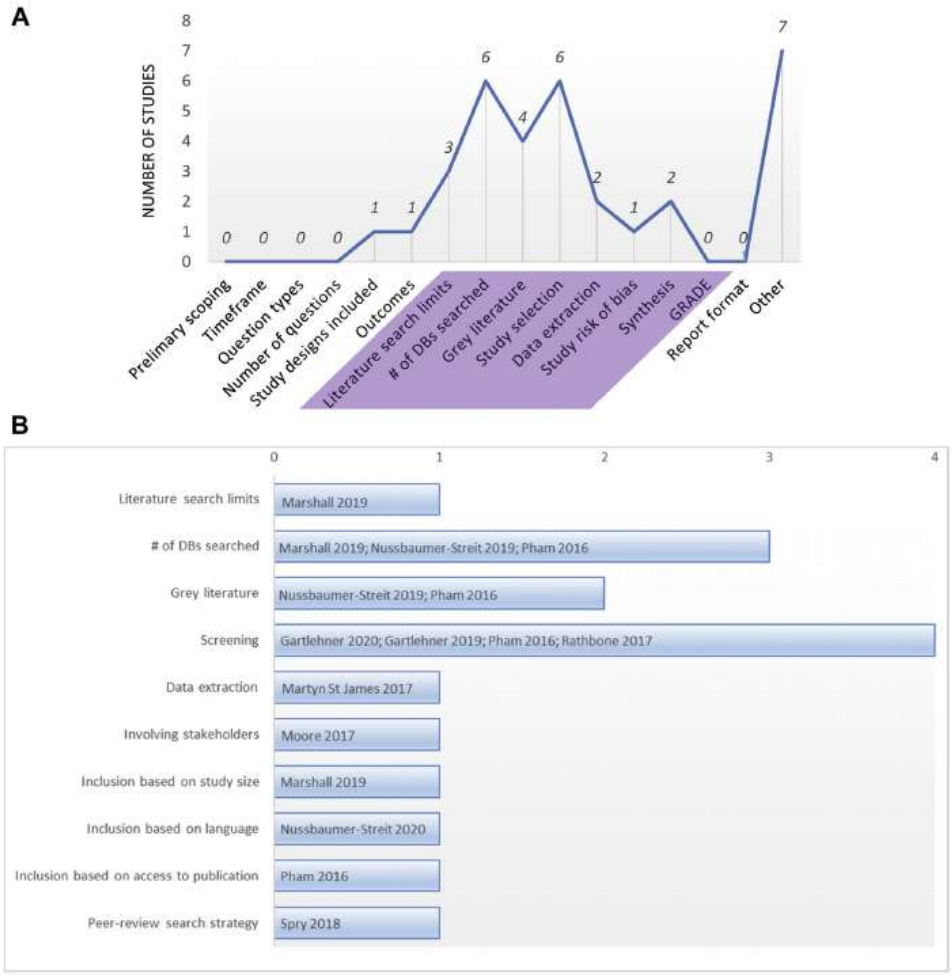


Fig. 3. A—Mapping to key dimensions of the review process (all evaluative studies). B—Mapping to key dimensions of the review process ('single' evaluative studies).

(CADTH) RRs [31]. In 30% (21 of 71) of the reviews, additional records were retrieved by the post–peer-reviewed searches, and one or more record was included in the report. However, there is a trade-off in time spent screening, as the post–peer-reviewed searches retrieved 2,507 additional records, with 4% (99 of 2507) of these records being included in the reports. *Author's conclusion: Although peer review requires more time and effort, to streamline the process, scrutiny of keywords, medical subject headings, and how these concepts are combined could be beneficial.*

3.3.2. 'Composite' evaluation studies

Four studies [20,23,32,33] compared reviews that took two or more shortcuts, labeled a rapid response report, a rapid network MA, a single-technology assessment, and a basic or enhanced rapid technology assessment, with more comprehensive reviews (e.g., SR, gold standard network meta-analysis) (Appendix E). In the less comprehensive reports, several shortcuts were used such as the number of outcomes included, number of databases searched, inclusion of gray literature, and one-reviewer–only study selection. Because of the variation in comparisons and shortcuts, we have not provided a synthesis.

3.3.3. Mapping to MECIR

The original plan was to map each shortcut to MECIR criteria and determine if the criteria would be met. However, depending on the study and the amount of information

provided around the shortcut used, it was not always possible to determine this. For example, the CADTH study evaluated two reports that both used several shortcuts. In addition, some of the items could not be mapped to existing MECIR criteria, as some are not methods performed in a traditional SR (e.g., using existing risk of bias information from an SR and performing new assessments for any studies not found in SRs, title screening), or are not currently found in MECIR although would be relevant to traditional SRs (e.g., peer-reviewing the search strategy). Therefore, only a cursory mapping to MECIR criteria was possible (Appendix F).

3.4. Other RR study categories

Three other RR categories captured all other included studies ($n = 70$) (Fig. 2) and are narratively summarized in Appendix G.

3.5. Systematic review surrogates

In addition to RR evaluations, six studies were identified that evaluated methods in SRs that may be transferrable to RRs (highlighted in Appendix H) [35–40]. This information was supplemented with the results from the Robson review [10].

4. Discussion

This scoping review identified 90 studies relating to RR methods, six of which were in the area of SRs but could be related to RR methodology. Only 14 studies formally evaluated the impact of shortcuts, of which four evaluated several shortcuts performed concurrently. The entirety of the evidence base largely comprises case studies, which may not be generalizable to all RRs. A cursory mapping exercise to the Cochrane MECIR guideline [9] resulted in a mix of shortcuts meeting and not meeting MECIR guidelines, in addition to some criteria not being covered in MECIR, or did not provide enough information to be able to make a judgment. Developing a set of standardized methods for RRs may be more difficult as there may not be one set of shortcuts (i.e., one size fits all) that should be followed. Determining which methods and shortcuts may be appropriate should be based on factors such as the topic area/question, the requirements of the stakeholders, and the availability of resources.

Several of the studies identified in this scoping review have been previously identified in other related work in this area. A scoping review by Tricco et al. (2015) [11] identified studies published between 1997 and 2013, and an RR by Haby et al. (2016) [13] identified studies published up until February 2015. As 71.1% (64 of 90) of the studies identified in this scoping review were published in 2015 onward, we have added to the growing repository of research performed in this area and have supplemented with research in the area of SR methods (Appendix H) [10].

One of the main concerns in using shortcuts in RRs is around the impact of the exclusion of relevant studies. One of the most resource-intensive stages of conducting a review is study selection (i.e., screening). Four of the formal evaluation studies evaluated shortcuts in screening. Overall, some screening shortcuts resulted in few missed studies, others resulted in mixed results (e.g., depending on the single reviewer), and some were found to perform poorly (i.e., machine screening alone). Single-reviewer screening did not perform well, although it is important to note that dual-independent screening is not without errors. A recent study by Wang et al. (2020) reported that 139,467 citations that underwent 329,332 inclusion and exclusion decisions resulted in an error rate (i.e., false inclusion or false exclusion by one reviewer) of 10.76% (95% CI 7.43 to 14.09%) [41]. The area of using artificial intelligence (AI) for automating or accelerating some of the processes of conducting reviews is a rapidly growing area of research. The International Collaboration for the Automation of Systematic Reviews was formed in 2015 and has been working in this area since [42–45]. Although completely replacing human screeners with AI performed poorly, recent research in the area of prioritization tools may be a valid option to help identify relevant studies quicker and reduce screening burden [46]. In addition, a recent SR was conducted in a 2-week period using AI to automate or accelerate screening [47]. Literature search limits and the number of databases searched were often evaluated; however, the use of AI for screening may lessen the need for decreasing the size of the search yield. As prioritization algorithms should identify the most relevant records first, those ranked low may not need to be screened, although this needs to be further evaluated. Although peer review of the search strategy would not be considered a methodological shortcut and requires additional resources, a well-developed search may optimize the results and reduce the yield, resulting in fewer citations to screen.

4.1. Implications for future research

Results from this scoping review point to several areas which may be considered by authors performing research in this area. First, four of the evaluation studies used several shortcuts concurrently, labeled as composite evaluations, which makes it difficult to assess which shortcut contributed to any difference in the results. Although it might be common practice in RRs to perform multiple shortcuts to expedite the review process, the impact of the shortcuts would be clearer if authors presented results separately, such as Pham et al. [29] Second, most of the studies identified have been classified as case studies because they were carried out in specific areas of health (e.g., breast cancer, men's sexual health), resulting in limited generalizability. Ideally, future RR methods research would crosscut multiple health areas (e.g., Nussbaumer-Streit 2019 [28] and Marshall 2019 [24]). Finally, among the 82 RRs identified

in Tricco (2015) [11], the most commonly used shortcuts (i.e., performed in ≥50 of RRs), when reported, concerned date limit (68%), data abstraction (68%), quality appraisal (57%), and title and abstract screening (54%). This may point to priority areas for evaluation. In this scoping review, date limit and data abstraction were each evaluated once, and no studies evaluated the impact of shortcuts specific to quality appraisal.

4.2. Strength and limitations

The strength of our work lies in the use of an a priori protocol, access to a collection of RR methods publications on the Cochrane RRMG website, and collaboration with the co-conveners of this Cochrane methods group. However, there were some limitations. The initial search resulted in a large yield of over 30,000 records. A specific search for records with 'rapid' in the title or abstract substantially reduced the yield and initially was not representative of the literature, as 14 of 30 studies from the initial scoping exercise were missed. Studies that did not use the word rapid (e.g., restricted) may have been missed. To reduce the chance of missing studies, a supplementary search was created using information from these missing studies. We used the AI tool within DistillerSR to exclude studies with a score of 0. This was based on a training set of 200 records at title and abstract and 25 full-text records. Any studies excluded by the AI reviewer were also screened by a human reviewer to minimize the chance of a false exclusion. Although there is a chance relevant studies may have been missed, this risk was minimized by performing gray literature and supplemental searching. In addition, we have categorized the 90 studies into four RR categories, plus SR surrogates. It is possible that a study could be placed in more than one category and does not account for overlap; therefore, categorical classifications should be considered subjective in nature and was a preliminary attempt to organize these articles, with guidance from Tricco et al. (2015) [11].

5. Conclusions

Few studies formally evaluated shortcut methods taken in RRs. There are methods shortcuts which may be useful in the context of RRs; however, there are some limitations within the included studies that may limit their applicability in the context of rapid reviews. Additional research should be performed with a focus on isolating stages of conduct in shortcuts with the greatest impact on resource-intensive stages (e.g., screening), while limiting potential bias.

CRediT authorship contribution statement

Candyce Hamel: Conceptualization, Funding acquisition, Data curation, Formal analysis, Investigation, Methodology, Project administration, Supervision, Validation, Visualization, Writing - original draft, Writing - review & editing. **Alan Michaud:** Investigation, Validation, Writing - review & editing. **Micere Thuku:** Investigation, Validation, Writing - review & editing. **Lisa Affengruber:** Investigation, Validation, Writing - review & editing. **Becky Skidmore:** Data curation, Writing - review & editing. **Barbara Nussbaumer-Streit:** Conceptualization, Funding acquisition, Writing - review & editing. **Adrienne Stevens:** Conceptualization, Funding acquisition, Writing - review & editing. **Chantelle Garritty:** Conceptualization, Data curation, Funding acquisition, Investigation, Supervision, Validation, Writing - review & editing.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclinepi.2020.06.027.

References

[1] Fox DM. Evidence and health policy: using and regulating systematic reviews. Am J Public Health 2017;107:88–92.
[2] Zhang Y, Akl E, Schunemann HJ. Using systematic reviews in guideline development: the GRADE approach. Res Synth Methods 2019; 10(3):312–29.
[3] Oliver K, Innvar S, Lorenc T, Woodman J, Thomas J. A systematic review of barriers to and facilitators of the use of evidence by policymakers. BMC Health Serv Res 2014;14:2.
[4] Borah R, Brown A, Capers P, Kaiser K. Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry. BMJ Open 2017;7: e012545.
[5] Hamel C, Michaud A, Thuku M, Skidmore B, Stevens A, Nussbaumer-Streit B, et al. Defining Rapid Reviews: a systematic scoping review and thematic analysis of definitions and defining characteristics of rapid reviews. Journal of Clinical Epidemiology; 2020: Submitted for publication.
[6] Pluddemann A, Aronson JK, Onakpoya I, Heneghan C, Mahtani KR. Redefining rapid reviews: a flexible framework for restricted systematic reviews. BMJ Evid Based Med 2018;23(6):201–3.
[7] Cochrane Library: Cochrane Database of Systematic Reviews. Cochrane Library. Available at https://www.cochranelibrary.com/cdsr/about-cdsr. Accessed May 14, 2020.
[8] Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editors. Cochrane Handbook for Systematic Reviews of Interventions version 6.0. Cochrane. 2019. Available at www.training.cochrane.org/handbook.
[9] Higgins J, Lasserson T, Chandler J, Tovey D, Thomas J, Flemyng E, et al. Methodological Expectations of Cochrane Intervention Reviews (MECIR): Standards for the Conduct and Reporting of New Cochrane Intervention Reviews, Reporting of Protocols and the Planning, Conduct and Reporting of Updates. Cochrane; 2019. Available at https://community.cochrane.org/mecir-manual. Accessed February 27, 2020.
[10] Robson RC, Pham B, Hwee J, Thomas SM, Rios P, Page MJ, et al. Few studies exist examining methods for selecting studies, abstracting data, and appraising quality in a systematic review. J Clin Epidemiol 2019;106:121–35.
[11] Tricco AC, Antony J, Zarin W, Strifler L, Ghassemi M, Ivory J, et al. A scoping review of rapid review methods. BMC Med 2015;13:224.
[12] Chandler J, Cumpston M, Thomas J, Higgins J, Deeks J, Clarke M. Chapter I: introduction. In: Cochrane Handbook for Systematic

Reviews of Interventions. 6.0. Cochrane. 2019. Available at www.training.cochrane.org/handbook. Accessed February 21, 2020.
[13] Haby MM, Chapman E, Clark R, Barreto J, Reveiz L, Lavis JN. What are the best methodologies for rapid reviews of the research evidence for evidence-informed decision making in health policy and practice: a rapid review. Health Res Policy Syst 2016;14(1):83.
[14] Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005;8(1):19−32.
[15] Peters MDJ, Godfrey C, McInerney P, Munn Z, Tricco AC, Khalil H. Chapter 11: scoping reviews. In: Aromataris E, Munn Z, editors. JBI Manual for Evidence Synthesis. The Joanna Briggs Institute; 2020. Available at https://wiki.joannabriggs.org/display/MANUAL/Chapter+11%3A+Scoping+reviews. Accessed July 14, 2020.
[16] Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018;169:467−73.
[17] McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. PRESS peer review of electronic search strategies: 2015 guideline statement. J Clin Epidemiol 2016;75:40−6.
[18] DistillerSR [Computer Program]. Ottawa, Canada 2011.
[19] Khangura S, Konnyu K, Cushman R, Grimshaw J, Moher D. Evidence summaries: the evolution of a rapid review approach. Syst Rev 2012;1:10.
[20] Eiring Ø, Brurberg KG, Nytrøen K, Nylenna M. Rapid methods including network meta-analysis to produce evidence in clinical decision support: a decision analysis. Syst Rev 2018;7(1):168.
[21] Gartlehner G, Affengruber L, Titscher V, Noel-Storr A, Dooley G, Ballarini N, et al. Single-reviewer abstract screening missed 13 percent of relevant studies: a crowd-based, randomized controlled trial. J Clin Epidemiol 2020;121:20−8.
[22] Gartlehner G, Wagner G, Lux L, Affengruber L, Dobrescu A, Kaminski-Hartenthaler A, et al. Assessing the accuracy of machine-assisted abstract screening with DistillerAI: a user study. Syst Rev 2019;8(1):277.
[23] Kaltenthaler E, Tappenden P, Booth A, Akehurst R. Comparing methods for full versus single technology appraisal: a case study of docetaxel and paclitaxel for early breast cancer. Health Policy 2008;87(3):389−400.
[24] Marshall I, Marshall R, Wallace B, Brassey J, Thomas J. Rapid reviews may produce different results to systematic reviews: a meta-epidemiological study. J Clin Epidemiol 2019;109:30−41.
[25] 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. Evid Policy 2017;13(3):517−38.
[26] Moore G, Redman S, D'Este C, Makkar S, Turner T. Does knowledge brokering improve the quality of rapid review proposals? A before and after study. Syst Rev 2017;6(1):23.
[27] Nussbaumer-Streit B, Klerings I, Dobrescu AI, Persad E, Stevens A, Garritty C, et al. Excluding non-English publications from evidence-syntheses did not change conclusions: a meta-epidemiological study. J Clin Epidemiol 2020;118:42−54.
[28] Nussbaumer-Streit B, Klerings I, Wagner G, Heise TL, Dobrescu AI, Armijo-Olivo S, et al. Abbreviated literature searches were viable alternatives to comprehensive searches: a meta-epidemiological study. J Clin Epidemiol 2018;102:1−11.
[29] Pham MT, Waddell L, Rajić 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 Synth Methods 2016;7(4):433−46.
[30] Rathbone J, Albarqouni L, Bakhit M, Beller E, Byambasuren O, Scott AM, et al. Expediting citation screening using PICo-based title-only screening for identifying studies in scoping searches and rapid reviews. Syst Rev 2017;6(1):233.
[31] Spry C, Mierzwinski-Urban M. The impact of the peer review of literature search strategies in support of rapid review reports. Res Synth Methods 2018;9(4):521−6.
[32] Taylor-Phillips S, Geppert J, Stinton C, Freeman K, Johnson S, Fraser H, et al. Comparison of a full systematic review versus rapid review approaches to assess a newborn screening test for tyrosinemia type 1. Res Synth Methods 2017;8(4):475−84.
[33] CADTH Health Technology Inquiry Service. Transcatheter Aortic Valve Implantation: A Critical Appraisal of a Health Technology Assessment and Comparison with a Rapid Review. CADTH; 2010. Available at https://www.cadth.ca/sites/default/files/pdf/I0177_transcatheter_aortic_valves_htis-2.pdf. Accessed February 27, 2020.
[34] 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.
[35] Revere D, Fuller S, Bugni PF, Martin GM. An information extraction and representation system for rapid review of the biomedical literature. Stud Health Technol Inform 2004;107:788−92.
[36] 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.
[37] Royle P, Milne R. Literature searching for randomized controlled trials used in Cochrane reviews: rapid versus exhaustive searches. Int J Technol Assess Health Care 2003;19(4):591−603.
[38] Topfer L-A, Parada A, Menon D, Noorani H, Perras C, Serra-Prat M. Comparison of literature searches on quality and costs for health technology assessment using the Medline and Embase databases. Int J Technol Assess Health Care 1999;15(2):297−303.
[39] Thomas J, McNaught J, Ananiadou S. Applications of text mining within systematic reviews. Res Synth Methods 2011;2(1):1−14.
[40] 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;19:132.
[41] Wang Z, Nayfeh T, Tetzlaff J, O'Blenis P, Murad MH. Error rates of human reviewers during abstract screening in systematic reviews. PLoS One 2020;15:e0227742.
[42] O'Connor AM, Tsafnat G, Gilbert SB, Thayer KA, Wolfe MS. Moving toward the automation of the systematic review process: a summary of discussions at the second meeting of International Collaboration for the Automation of Systematic Reviews (ICASR). Syst Rev 2018;7(1):3.
[43] Beller E, Clark J, Tsafnat G, Adams C, Diehl H, Lund H, et al. 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.
[44] O'Connor AM, Tsafnat G, Gilbert SB, Thayer KA, Shemilt I, Thomas J, et al. Still moving toward automation of the systematic review process: a summary of discussions at the third meeting of the International Collaboration for Automation of Systematic Reviews (ICASR). Syst Rev 2019;8(1):57.
[45] O'Connor AM, Glasziou P, Taylor M, Thomas J, Spijker R, Wolfe MS. A focus on cross-purpose tools, automated recognition of study design in multiple disciplines, and evaluation of automation tools: a summary of significant discussions at the fourth meeting of the International Collaboration for Automation of Systematic Reviews (ICASR). Syst Rev 2020;9(1):100.
[46] Tsou AY, Treadwell JR, Erinoff E, Schoelles K. Machine learning for screening prioritization in systematic reviews: comparative performance of Abstrackr and EPPI-Reviewer. Syst Rev 2020;9:73.
[47] Clark J, Glasziou P, Del Mar C, Bannach-Brown A, Stehlik P, Scott AM. A full systematic review was completed in 2 weeks using automation tools: a case study. J Clin Epidemiol 2020;121:81−90.

Hamel et al. BMC Medical Research Methodology (2020) 20:256
https://doi.org/10.1186/s12874-020-01129-1

BMC Medical Research Methodology

RESEARCH ARTICLE Open Access

An evaluation of DistillerSR's machine learning-based prioritization tool for title/abstract screening – impact on reviewer-relevant outcomes

C. Hamel[1,2*], S. E. Kelly[3,4], K. Thavorn[1,4], D. B. Rice[1,5], G. A. Wells[1,3,4] and B. Hutton[1,4]

Abstract

Background: Systematic reviews often require substantial resources, partially due to the large number of records identified during searching. Although artificial intelligence may not be ready to fully replace human reviewers, it may accelerate and reduce the screening burden. Using DistillerSR (May 2020 release), we evaluated the performance of the prioritization simulation tool to determine the reduction in screening burden and time savings.

Methods: Using a true recall @ 95%, response sets from 10 completed systematic reviews were used to evaluate: (i) the reduction of screening burden; (ii) the accuracy of the prioritization algorithm; and (iii) the hours saved when a modified screening approach was implemented. To account for variation in the simulations, and to introduce randomness (through shuffling the references), 10 simulations were run for each review. Means, standard deviations, medians and interquartile ranges (IQR) are presented.

Results: Among the 10 systematic reviews, using true recall @ 95% there was a median reduction in screening burden of 47.1% (IQR: 37.5 to 58.0%). A median of 41.2% (IQR: 33.4 to 46.9%) of the excluded records needed to be screened to achieve true recall @ 95%. The median title/abstract screening hours saved using a modified screening approach at a true recall @ 95% was 29.8 h (IQR: 28.1 to 74.7 h). This was increased to a median of 36 h (IQR: 32.2 to 79.7 h) when considering the time saved not retrieving and screening full texts of the remaining 5% of records not yet identified as included at title/abstract. Among the 100 simulations (10 simulations per review), none of these 5% of records were a final included study in the systematic review. The reduction in screening burden to achieve true recall @ 95% compared to @ 100% resulted in a reduced screening burden median of 40.6% (IQR: 38.3 to 54.2%).

Conclusions: The prioritization tool in DistillerSR can reduce screening burden. A modified or stop screening approach once a true recall @ 95% is achieved appears to be a valid method for rapid reviews, and perhaps systematic reviews. This needs to be further evaluated in prospective reviews using the estimated recall.

Keywords: Artificial intelligence, Systematic reviews, Rapid reviews, Prioritization, Automation, Natural language processing, Machine learning, Time savings, Efficiency, True recall

* Correspondence: cahamel@ohri.ca
[1]Clinical Epidemiology Program, Ottawa Hospital Research Institute, 501 Smyth Road, Box 201b, Ottawa, Ontario K1H 8L6, Canada
[2]Department of Medicine, University of Split, Split, Croatia
Full list of author information is available at the end of the article

© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Systematic reviews (SRs) aim to minimize bias by using systematic and rigorous methods [1]. This process, however, can require substantial resources (e.g., cost and humans), and in some cases can require more than 12 months to complete. An analysis of 195 reviews registered in PROSPERO reported a mean time (from registration to publication) of 67.3 weeks (standard deviation 31 weeks, range 6 to 186 weeks) and a mean author team of 5 people [standard deviation (SD): 3, range 1 to 27 people] [2].

It is not uncommon for a systematic search to yield a large number of records, many of which are irrelevant (i.e., low precision) [2, 3]. In a recent study, of 139,467 citations among 25 reviews, 5.48% (95% confidence interval (CI) 2.38 to 8.58%) of the citations were included in the final reviews [3]. Such volume introduces opportunity for human error in the screening process [3–5]. While screening of titles and abstracts represents only one step in the series of tasks involved in the conduct of SRs, due to the high screening burden, the resources for this step can be a large proportion of the total human resource time spent on the review [6]. Several strategies have been evaluated to decrease time spent screening titles and abstracts, including the use of dual monitors for screening [7], title only screening [8], a staged title only followed by abstract screening [6], screening by one reviewer [5, 9–12], and using artificial intelligence (AI) tools (e.g., text mining, prioritization) [11, 13–17].

Several software tools exist that support title and abstract screening in SRs [18], however not all packages currently include the capacity to implement machine learning techniques for citation screening [19]. Among those that do, there is variation in the level of sophistication of the machine learning tool, the algorithms used, the cost of the software package, and if and how often it is updated and supported. The most commonly evaluated software are Abstrackr, DistillerSR, EPPI-Reviewer, RobotAnalyst, SWIFT-Active Screener, and SWIFT-Review [13–16, 20–24], with varying success depending on the size of the datasets, the machine learning algorithm, and the level of replacement of humans with AI [25]. While AI may not be ready to fully replace human screeners in the task of study selection, studies suggest that optimizing, accelerating, and reducing screening burden through the use of AI-informed screening methods represents a viable option. This includes *prioritized screening*, where the presentation of titles and abstracts to reviewers is continually adjusted, through active machine learning, based on the AI's estimated likelihood of relevance [17]. In circumstances of present day where the requestors (end users) of a particular knowledge synthesis frequently are in search of a rapidly generated synthesis of the available evidence for a research question of interest, such tools may offer attractive gains to research teams if safely implemented to minimize the risk of falsely excluding relevant evidence.

A 2015 systematic review concluded that there is almost no replication between studies or collaboration between research teams evaluating text mining methods, which makes it difficult to establish overall conclusions about best approaches [17]; this represents an especially troublesome barrier toward wider adoption of the use of such methods globally in knowledge syntheses. Another important barrier to uptake for many research teams is uncertainty as to the proper set-up and implementation, both in terms of settings within the software as well as incorporation into the well-established SR process.

Objectives

Using the AI simulation tool (which uses the prioritization algorithm) in DistillerSR, the primary objectives of this study were to:

(1) Empirically evaluate the reduction in *screening burden* (the number of records not required to be screened) once a *true recall @ 95%* was achieved (i.e., once 95% of the studies included based on the title/abstract to be further evaluated based on the full-text were identified).
(2) Evaluate the performance using a true recall @ 95%. Specifically, to identify if any of the studies that were included in the systematic review were among the 5% of *records that were not yet identified* as included based on the title/abstract [i.e., *title/abstract false negatives (FN)*].

We chose DistillerSR software (Evidence Partners Incorporated; Ottawa, Canada), as it is amongst the most widely used systematic review management software programs worldwide, and because our research teams are long-time users of this software. A list of terminology (italicized terms) used in the manuscript with descriptions are provided in Table 1.

There is currently no agreed upon *modified screening* or *stop screening* approach where a review team may decide to modify how records are being screened (e.g., changing from dual-independent screening to single-reviewer screening) or stop screening the remaining records. For the current study, we are evaluating a true recall @ 95%. In other words, once the AI simulation tool has identified 95% of the studies that were included based on the title/abstract to be further reviewed based on the full text [i.e., *title/abstract true positives (TP)*], we would assign the AI reviewer to exclude the remaining studies which would include approximately 5% of the title/abstract records that were included but not yet identified (i.e., title/abstract FP) and the title/abstract

Table 1 Terminology and descriptions

Terminology	Description
Estimated recall	The estimated percent of how many studies at title/abstract level have been identified among those that will be passed through to full-text screening. As this is calculated based on a set of records that have not been completely screened, the estimated recall may differ from the true recall.
Final include	A primary study included in the completed systematic review.
Iteration	A set of records that is used to assign a score around the likeliness of inclusion and prioritize the remaining unscreened records in order from highest relevance to lowest relevance.
Modified screening approach	An approach to modify how screening is being performed. For example, changing from: (i) dual-independent screening to liberal accelerated screening; (ii) dual-independent screening to single-reviewer screening; or (iii) assigning the remaining records to the AI reviewer to exclude, with a human reviewer(s) also screening these records as a second reviewer.
Prioritized screening	Through active machine learning, the presentation of records to reviewers is continually adjusted based on the AI's estimated likelihood of relevance. The frequency of adjustment may differ by software application.
Screening burden	The total number of records at title/abstract to be screened.
Stop screening approach	An approach to screening whereby the remaining records are not screened once a certain threshold has been achieved (e.g., estimated recall @ 95%). These records are assumed to be excluded.
Record not yet identified [i.e., title/abstract false negative (FN)]	When an estimated recall (at any %) or true recall of less than 100% is used, these are the records that would have been included based on the title/abstract to be further reviewed at full-text screening, but were not yet identified. Had these records been screened at title/abstract and further screened based on the full text, they may have been excluded or included in the final review (i.e., a final include).
Title/abstract include [i.e., title/abstract true positive (TP)]	Records included based on the title/abstract to be further reviewed based on the full text. These records may then be excluded at full-text review or included in the final review.
Training set	One or more iterations which inform the machine learning to score and prioritize the remaining unscreened records.
Title/abstract exclude [i.e., true negative (TN)]	Records considered excluded based on title/abstract screening.
True recall	This is only known once all references have been screened and includes the percentage of the actual number of records that were title/abstract includes. True recall % calculated as: [title/abstract TP / (title/abstract TP + title/abstract FN)]

excludes [i.e., *true negatives (TN)*]. This number (95%) was selected as it is a common recall number used when measuring the reduction in workload; it also approximates the level of human error in screening [3, 16, 26]. Therefore, true recall @ 95% is calculated as [title/abstract TP / (title/abstract TP + title/abstract FN)]. The distinction between true recall and *estimated recall* (as would be calculated in a prospective review) is that, as we used completed reviews, we know the actual number of studies that were included based on the title/abstract screening to be further evaluated based on the full text [23]. The findings from this study will help toward establishing the validity of this approach to citation screening as a potential additional source of time savings in the context of conducting systematic reviews and other knowledge synthesis products, including rapid reviews [27–30] and living systematic reviews [31, 32]. Furthermore, given that challenges in set-up are a known barrier amongst knowledge synthesis teams toward the decision to implement machine learning methods for their research [25], a secondary objective of the study was to provide transparent, repeatable methods for other review teams to replicate in their own research. This will allow for further testing of this process, thereby increasing the sample size and external validity of the results presented in this study.

Study methods

The protocol for this study was registered on the Open Science Framework (OSF: https://osf.io/2fgz7/) and was conducted using the AI simulation module within DistillerSR Software (May 2020 release). This version (2.31.0) of DistillerSR has fully replaced all existing AI functionality from earlier versions and includes prioritized reference screening (i.e., re-sorting records at regular screening intervals based on the AI tool's estimated probability of relevance for each remaining record) and the development of a system in which to create custom classifiers [e.g., automatically labeling randomized controlled trials (RCTs)].

This study used information from 10 previously completed SRs (i.e., responses to screening at title/abstract and the final list of included studies) that were undertaken by research teams that perform a high volume of knowledge synthesis reviews, led by our co-authors,

located at the Ottawa Hospital Research Institute and the University of Ottawa Heart Institute in Ottawa, Canada. We selected 10 reviews in this pilot experiment to capture a variety of topic areas, review sizes, and inclusion rates. An overview of the characteristics of these reviews, with brief descriptions of the objectives and PICO elements (participants, interventions, comparators, outcomes) is provided in Additional file 1.

Methods on how we implemented DistillerSR's AI simulation tool for citation screening have been described in detail in Additional file 2 for researchers who are interested in running simulations using their own review projects. In the context of the current study, DistillerSR's AI simulation tool selects a random set of records which contains 2% of the dataset (with a minimum of 25 records and a maximum of 200 records). Each set of these records is called an *iteration*. and the simulation tool uses the responses already provided (title/abstract included and excluded responses, based on our previous SRs) to build the first iteration (i.e., the initial *training set*). Subsequently, the remaining unscreened records are assigned a score (by the software) relating to the likelihood of inclusion, and references are re-ranked (i.e., prioritized) in order of this score (from most to least likely to be relevant). The next iteration (i.e., the next 2% of the records) is then run, and all remaining records are assigned an updated score based on the likelihood of inclusion estimated using the information gathered from all iterations, which creates the newest training set. This process continues until all records are screened. The AI simulation tool mimics the process of human screening. In a prospective review, responses from the reviewers would be used to build the iterations (e.g., using single reviewer, dual independent review with conflicts resolved), but would otherwise function in the same manner. Once prioritization is set up (i.e., one click when managing levels), the process of prioritization occurs automatically in the background without intervention from the reviewers, making it easy to use, and thereby providing the potential to identify relevant literature more efficiently.

Figure 1 represents how the simulation tool uses the existing information (i.e., responses) to simulate the performance of the prioritization tool.

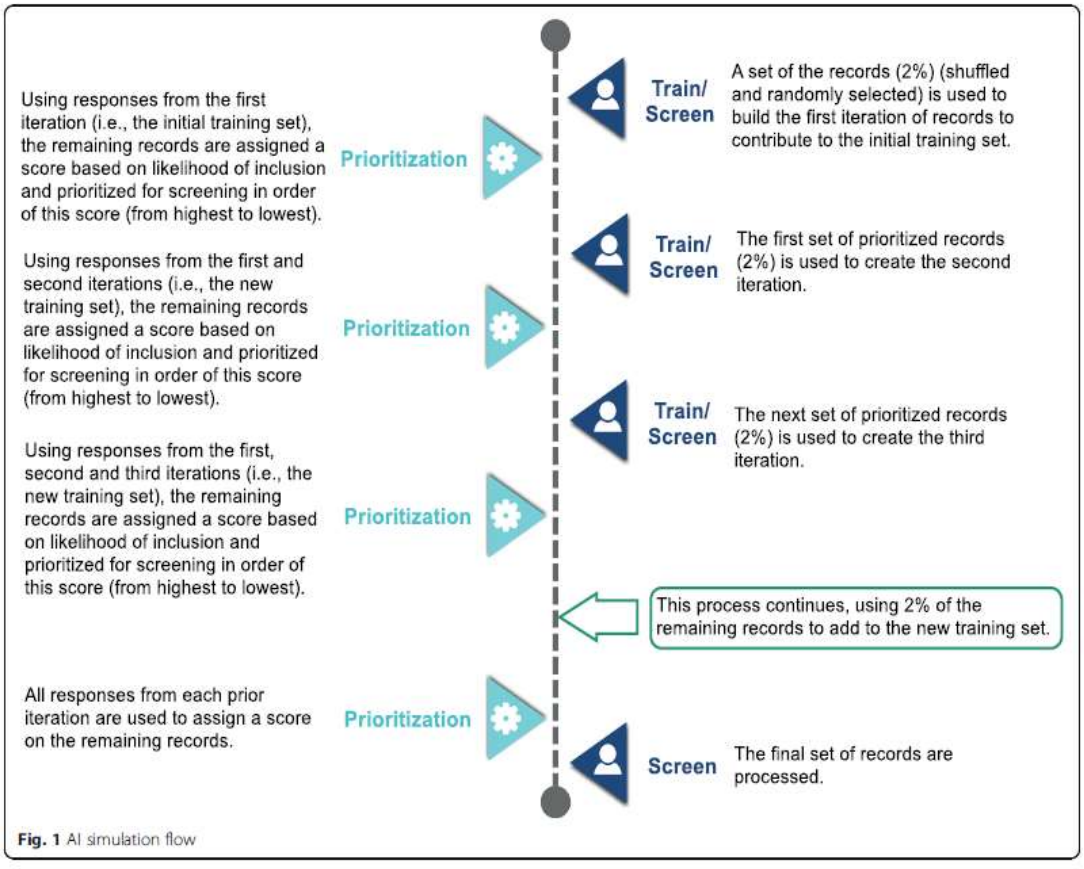


Fig. 1 AI simulation flow

Data collection

For each of the 10 SRs that served as experimental units for this work, we ran the AI simulation 10 times to account for any variation in the simulations, and to introduce randomness (through shuffling the references, which is automatically performed by the software) into the initial training sets. After each simulation was run, the following information was recorded at the first iteration that identified 95% of the studies included from title/abstract to be further evaluated at full text (i.e., true recall @ 95%):

- *The number of records per iteration and the number of iterations*. An iteration contains 2% of the total number of records, with a minimum of 25 and a maximum of 200 records per iteration. This allowed for measuring the variation within a review around the number of records at title/abstract not yet identified (i.e., title/abstract FN).
- *The total number of records screened (i.e., screening burden)*. This is composed of 95% of the title/abstract included studies and a portion of the title/abstract excluded studies.
  Calculation: (title/abstract TP + title/abstract TN).
- *The number of records included at title/abstract to be further reviewed at full-text screening once a true recall @ 95% was achieved* (title/abstract TP). This could account for slightly more than 95% of the studies, depending on the how many of these studies at title/abstract were located in the iteration which captured 95% of the title/abstract included studies.
- *The number of records screened that were excluded* (title/abstract TN). Reviews that have a large number of records that were included based on the title/abstract to be further reviewed at full text will likely have a higher rate of total number of records screened. Therefore, the number of excluded records screened was also recorded as this is the number of records that should be reduced to accurately report the reduction in screening burden.
- *The list of reference identification numbers (IDs) of the 5% of included records at title/abstract not yet identified* (title/abstract FN). This allowed for evaluation if any of these studies were on the list of final included studies in the systematic review (i.e., *final include*).

Outcomes

The combined results from the 10 simulations per SR allowed for the calculation of the mean (SD) and median (range), when reporting results for a specific review, or median [interquartile range (IQR)] when reporting results across reviews for each outcome of interest:

(1) The number and percent of records (at title/abstract) needed to screen to identify a true recall @ 95% (i.e., screening burden).
Calculation: title/abstract TP + title/abstract TN (at a true recall @ 95%)

(2) The number and percent of studies at title/abstract not yet identified at true recall @ 95% (title/abstract FN) among all studies that were included for further evaluation at full-text (title/abstract TP) at a true recall @ 100%.
Calculation: [(title/abstract TP − title/abstract FN) / title/abstract TP]. As we are using a true recall @ 95%, this should approximate 5%.

(3) The number and percent of final includes (i.e., those in the final list of included studies in the systematic review) among the title/abstract FN.

(4) Number of hours saved, which was calculated using a modified screening approach, in which the AI reviewer would exclude all remaining records and a human reviewer would review these records. The number of hours saved was calculated by multiplying the expected time to review a record (i.e., one record per minute, based on Shemilt 2016 [11] and the experience of our own research groups) by the total number of records that did not need to be screened by one reviewer (i.e., the total number of records remaining once a true recall @ 95% was achieved). As this outcome is based on true recall, rather than estimated recall, the number of hours saved is an estimate as, in a prospective review, a review team would not know for certain if the estimated 95% was in fact 95% of the studies that would have been passed through to full-text screening, as not all references would have been screened.

Deviations from the protocol

In the protocol, we stated that we would measure total cost savings as an outcome. However, the research team subsequently decided it would be of greater information and generalizability to knowledge synthesis researchers if we instead presented the number of hours saved. This would allow other researchers to calculate cost savings in different currencies at different salaries, as appropriate. Additionally, as the 95% modified screening approach resulted in a substantial number of records that did not need to be screened for some of the SRs, we performed an additional analysis to evaluate the difference in the relative screening burden when comparing how much of the total dataset was required to be screened to achieve a true recall @100% compared to a true recall @ 95%.

Results

Overview of SRs assessed

Ten SRs, consisting of 69,663 records, were used in this experiment. Four SRs included only RCTs, and the remaining SRs included both RCTs and observational studies. Using the review typology by Munn et al. (2018) [33], eight SRs were classified as effectiveness reviews [including both SR and network meta-analysis (NMAs)], and two SRs were effectiveness and etiology reviews. All SRs covered clinical areas and primarily evaluated the efficacy and safety of pharmacological, non-pharmacological (e.g., behavioural therapies), and surgical interventions. One SR each evaluated depression screening effectiveness, the use of e-cigarette for smoking cessation, and interventional/behavioural exposure to sugar sweetened beverages (SSBs) (Additional file 1). Reviews ranged in size from 2250 to 22,309 records to be assessed at title and abstract level, of which 3.0 to 39.2% (median: 16.2%) were included, based on the title/abstract, to be further reviewed at full text. A median of 0.6% (range 0.02 to 1.48%) of the total number of records were included in the final systematic reviews.

Findings: reduction in screening burden

Across the set of 10 SRs evaluated, the median percentage of studies required to be screened to achieve a true recall @ 95% was 47.1% (IQR: 37.5 to 58.0%) (Table 2 and Additional file 3: Suppl. Table 1). Four SRs [i.e., non-small cell lung cancer, smoking cessation, prophylaxis for human immunodeficiency virus (HIV), SSBs] required at least 50% of records to be screened to achieve a true recall @ 95%. All of these reviews had more than 22% of the title/abstract records passed through for full-text screening. Among all reviews, when considering only the number of excluded records required to be screened to achieve true recall @ 95%, a median of 41.2% excluded records needed to be screened (IQR: 33.4 to 46.9%) (Additional file 3: Suppl. Table 1).

Figure 2a presents the mean percentage of records that were included and excluded based on titles/abstracts, and the resulting reduction in the screening burden. The number of records that did not need to be screened (light blue portion of the bar) ranged from 30% (smoking cessation) to 72.5% (opioid use disorder). Figure 2b presents the relationship between the percentage of studies passed through to full-text screening and the mean percentage reduction in screening burden once true recall @ 95% was achieved. Typically, reviews with fewer studies passed through to full-text screening resulted in a larger reduction in the overall screening burden, as fewer excluded records would need to be screened to identify the studies requiring further review at full text.

There was little variation in the magnitude of screening burden within each of the 10 SRs among the 10 simulations. Three SRs achieve true recall @ 95% in the same number of iterations, while five SRs had a range of one iteration, and two SRs had a range of four iterations. It was common for the same references to be missed in each iteration. The difference between the total number of unique title/abstract included studies not yet identified (i.e., title/abstract FN that were listed in at least one of the ten simulations) and the largest number of title/abstract FN (i.e., the iteration with the largest number of title/abstract FNs) was 0 to 13 records [mean (SD): 5.3 records (4.03); median (IQR): 5 (2–8) records].

Figure 3 presents the variation in the number of title/abstract included studies not yet identified (i.e., title/abstract FN) the simulation with the lowest number, highest number, and overall unique number of title/abstract FN. The lower the variation between simulations, the closer the minimum, maximum and number of unique studies. In these 10 reviews, 4.8 to 6.2% of the same records were not yet identified in the 10 simulations.

Findings: amount of time saved

Overall, the mean title/abstract screening hours saved when using the true recall @ 95% modified screening approach (i.e., the AI reviewers would exclude all remaining references and one human review would be required to screen the remaining records) was 62.8 h (median: 29.8 h; IQR: 28.1 to 74.7 h). As would be expected, SRs with a larger number of records tended to result in more hours saved. SRs with fewer than 5000 records saved between 11.3 to 36 h. SRs with more than 5000 records (i.e., prophylaxis for influenza, opioid use disorder, and SSBs), saved totals of 88, 158 and 197 h (up to approximately 5 weeks of work time), respectively.

Figure 4 displays the mean hours saved per review from implementing the modified screening approach once a true recall @ 95% was achieved. The size of the bubbles represent the amount of hours saved. Reviews with fewer than 5000 records showed little variation in the total hours saved when the title/abstract true positive rate was between 10 and 30% (range 22 to 30 h, or approximately 1 day of work).

Using estimates from Shemilt et al. [11] of 4 min per person to retrieve a full text record and 5 min per person to screen a full text record, and assuming that full-text screening is done in duplicate, this would increase the total hours saved by not having to access and screen the 5% of title/abstract false negatives (Additional file 3: Suppl. Table 2). For example, in the review where AI was the least efficient in reducing the screening burden (i.e., smoking cessation), an average of 40 records did not need to be screened at title/abstract, a time savings of 11.3 h. However, adding the time to retrieve these

Table 2 Study results

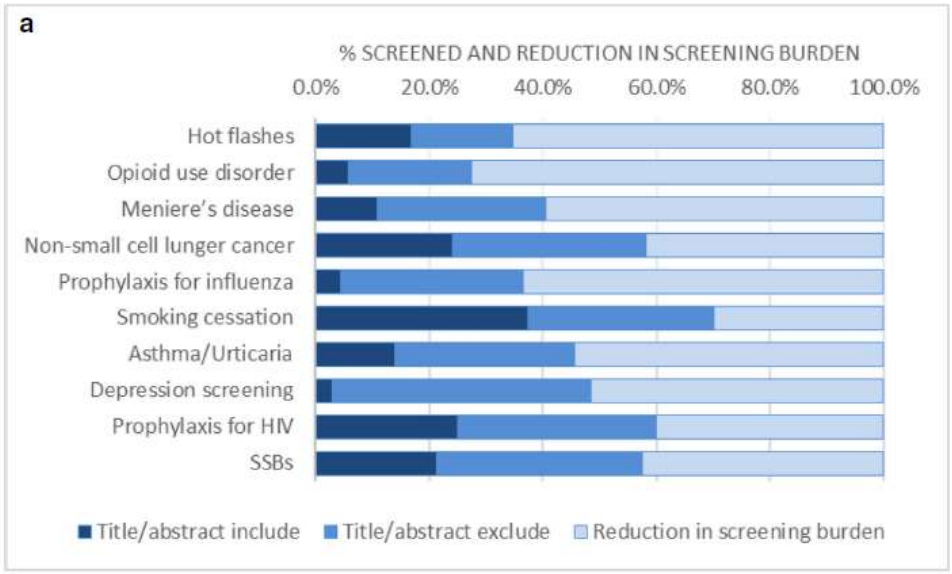
Project	Project details [a]	Iteration details	# of records needed to screen to achieve true recall @ 95% [b] Mean (%) [SD]; Median (range)	Title/abstract includes not yet identified	Hours saved at title/abstract [b c]	Final included studies missed
Hot flashes	2569; 451 (17.6%); 38 (1.48%)	51 records; 17 or 18 iterations	892.5 (34.7%) [26.9]; 892.5 (867–918)	19.2 (4.3%) [2.04]; 19 (15–22)	27.9 [0.45]; 27.9 (27.5–28.4)	0
Opioid use disorder	16,282; 984 (6.0%); 71 (0.44%)	200 records; 23 or 23 iterations	4480 (27.5%) [103.3]; 4400 (4400–4600)	46.1 (4.7%) [3.38]; 48 (41–49)	196.7 [1.72]; 198.0 (194.7–198.0)	0
Meniere's disease	2889; 332 (11.5%); 23 (0.80%)	57 records; 19–22 iterations	1168.5 (40.5%) [55.4]; 1140 (1083–1254)	15.0 (4.5%) [1.33]; 15.5 (12–16)	28.7 [0.92]; 29.2 (27.3–30.1)	0
Non-small cell lung cancer	3145; 795 (25.3%); 13 (0.40%)	62 records; 29 or 30 iterations	1829 (58.2%) [0.01]; 1829 (1798–1860)	33.7 (4.2%) [3.53]; 32.5 (29–39)	21.9 [0.54]; 21.9 (21.4–22.5)	0
Prophylaxis for influenza	8278; 395 (4.8%); 104 (1.26%)	165 records; 18 or 19 iterations	3019.5 (36.5%) [79.7]; 2970 (2970–3135)	18.8 (4.8%) [0.42]; 19 (18–19)	87.6 [1.33]; 88.5 (85.7–88.5)	0
Smoking cessation	2250; 881 (39.2%); 14 (0.62%)	45 records; 35 iterations	1575 (70.0%) [0]; 1575 (0)	39.9 (4.5%) [2.60]; 40 (34–44)	11.3 [0]; 11.3 (0)	0
Asthma/ Urticaria	3265; 482 (14.8%); 12 (0.36%)	65 records; 22 or 23 iterations	1488.5 (45.6%) [20.55]; 1495 (1430–1495)	22.5 (4.7%) [1.51]; 23 (20–24)	29.6 [0.34]; 29.5 (29.5–30.6)	0
Depression screening	4174; 126 (3.0%); 1 (0.02%)	83 records; 23–26 iterations	2025 (48.5%) [7.0]; 1992 (1909–2158)	5.8 (4.6%) [0.42]; 6 (5–6)	35.8 [1.17]; 36.4 (33.6–37.8)	0
Prophylaxis for HIV	4502; 1184 (26.4%); 46 (1.02%)	90 records; 30 iterations	2700 (60.0%) [0]; 2700 (0)	53.7 (4.5%) [1.49]; 53.5 (52–56)	30.0 [0]; 30.0 (0)	0
SSBs	22,309; 4993 (22.4%); 127 (0.57%)	200 records; 64 iterations	12,800 (57.4%) [0]; 12800 (0)	242.7 (4.9%) [2.06]; 243 (238–246)	158.5 [0]; 158.5 (0)	0

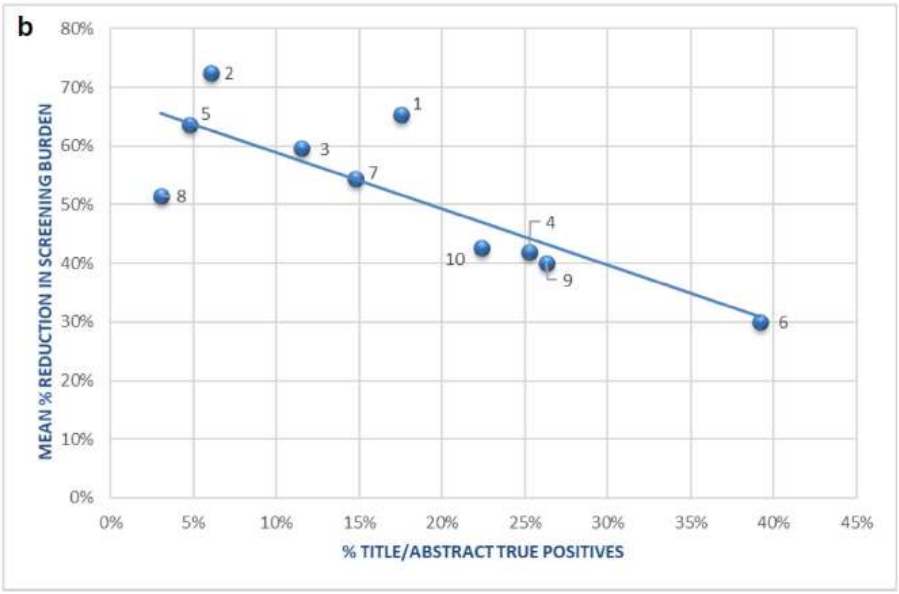
HIV Human immunodeficiency virus, *SD* Standard deviation, *SSB* Sugar sweetened beverage

[a] Total number of records; Number of includes at title/abstract (% of all records); Number of included studies in the SR (% of all records)

[b] Where there was no SD or range, this is identified by a 0

[c] Hours saved at title/abstract = [(Total records - # of records needed to screen to identify 95% of includes)/60





1 Hot flashes
2 Opioids use disorder
3 Meniere's disease
4 Non-small cell lung cancer
5 Prophylaxis for influenza
6 Smoking cessation
7 Asthma/Urticaria
8 Depression screening
9 Prophylaxis for HIV
10 SSBs

Fig. 2 a Title/abstract includes and excludes and screening burden reduction. **b** – Relationship of mean % reduction in screening burden and % of title/abstract includes

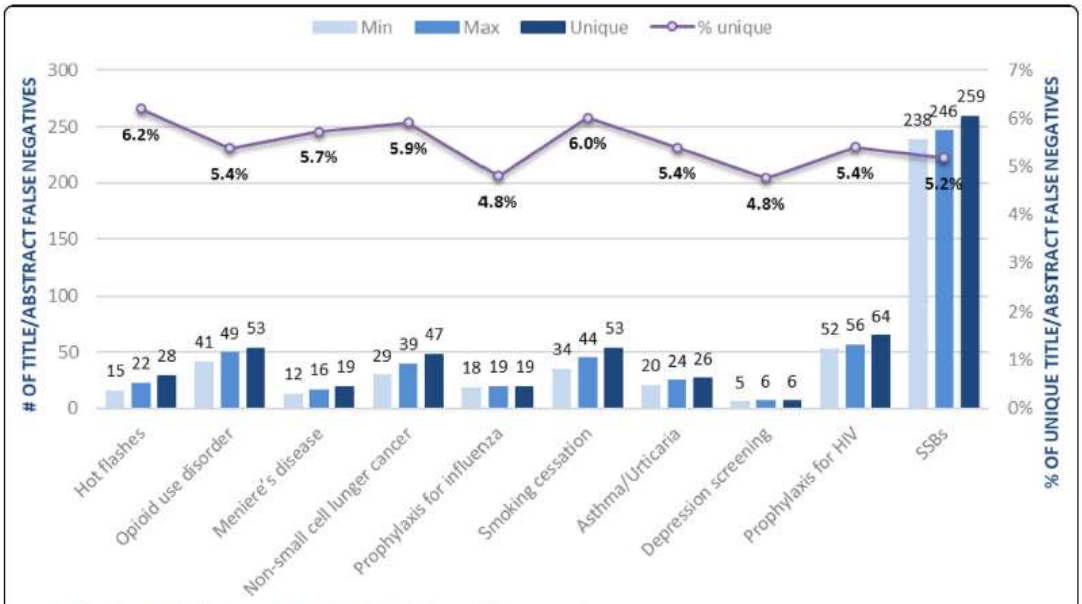


Fig. 3 Title/abstract includes not yet identified (i.e., title/abstract false negatives)

articles (40 @ 4 min/record = 2.7 h) and the time for two reviewers to screen at full text (40 @ 5 min/record × 2 = 6.7 h), this results in an additional 9.4 h of time savings, nearly doubling the time savings. The Asthma/Urticaria review (which approximated the median for total records, % of includes at title/abstract, and time savings in hours) would result in a total time savings of 35.3 h (title/abstract screening: 30 h; retrieving full texts: 1.5 h; screening full texts: 3.8 h). The largest review, SSBs, would result in a total time savings of 215.1 h (title/abstract screening: 158.5 h; retrieving full texts: 16.2 h; screening full texts: 40.5 h). These numbers do not

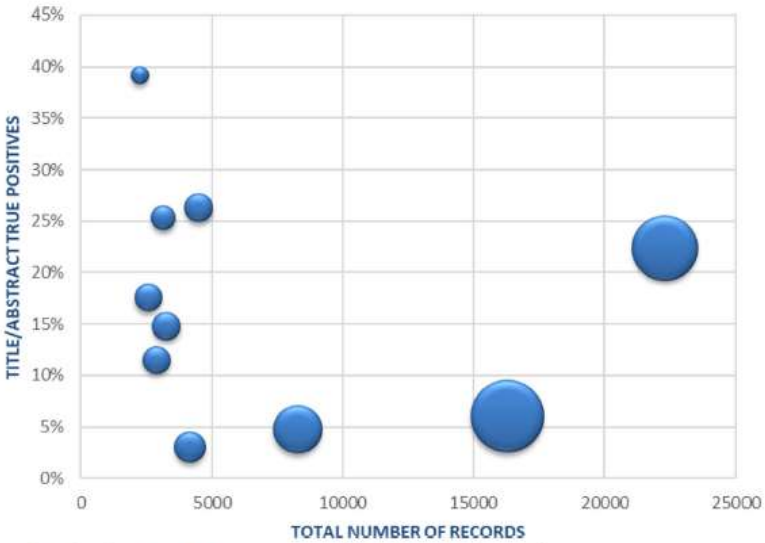


Fig. 4 Mean hours saved in title/abstract screening using a true recall @ 95% modified approach

include any ordering fees for articles not accessible without a journal subscription, plus any additional time to resolve conflicts at full text (which has been estimated to take 5 min per conflict [11]).

Figure 5 shows that the extra time to retrieve the full text and perform full-text screening represents 4 to 45% of the estimated total time saved (median: 14%).

Findings: performance (accuracy) of the prioritization algorithm

Across the 10 SRs studied, a median of 4.57% of the records were title/abstract FN (IQR: 18.9 to 44.6). Among the 100 iterations (10 iterations in 10 SRs), no final included studies were not yet identified at a true recall @ 95% (Table 2).

A post-hoc analysis was subsequently performed to evaluate the difference in the screening burden to achieve a true recall @ 100% compared to a true recall @ 95%. In measuring this, using the mean over three simulations, this resulted in a median difference in the number required to screen of 40.6% (IQR: 38.3 to 54.2%). It is important to note that the additional screening burden to identify the last 5% of the records included at title/abstract would not have identified any final included studies in the systematic reviews, as they were all identified in the true recall @ 95%.

Figure 6 displays the reduction in screening burden over the 10 reviews at a true recall rate of 95 and 100%. Seven of the 10 reviews required over 90% of the records to be screened to achieve a true recall @ 100%. Two of these were the largest reviews (i.e., Opioids use disorder = 16,282 records, SSBs = 22,309 records).

Discussion

The new prioritization tool in DistillerSR reduced the screening burden in these 10 SRs by 30.0 to 72.5% when using a true recall @ 95% modified screening approach. Smaller studies with a high inclusion rate will take longer to identify 95% of the title/abstract includes and resulted in poorer performance for the machine learning algorithm. Although some of the larger studies had high rates of title/abstract includes, due to the size of the dataset, the reduction in screening burden would still result in a large time and potentially lead to a subsequent cost savings. A recently published study evaluated the accuracy of screening prioritization of Abstrackr and EPPI-Reviewer [15]. Screening burden to identify all title/abstract includes for the de novo review was 85% or more for seven of the nine reviews for both Abstrackr (median: 93.8%, range: 71.1 to 99.0%) and EPPI-Reviewer (median: 91.3%, range: 39.9 to 97.9%). However, six of the nine included reviews had fewer than 1000 records, thereby not starting with a particularly large screening burden. Although not a direct comparison to our experiment, as different datasets were used, identifying 100% of the title/abstract includes using

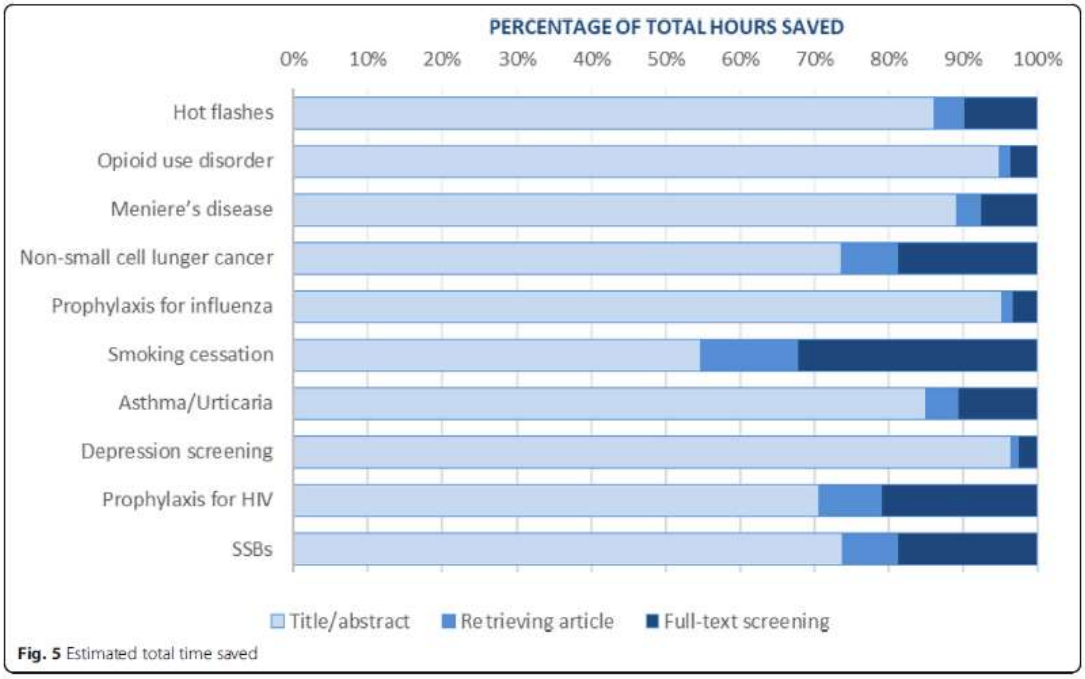


Fig. 5 Estimated total time saved

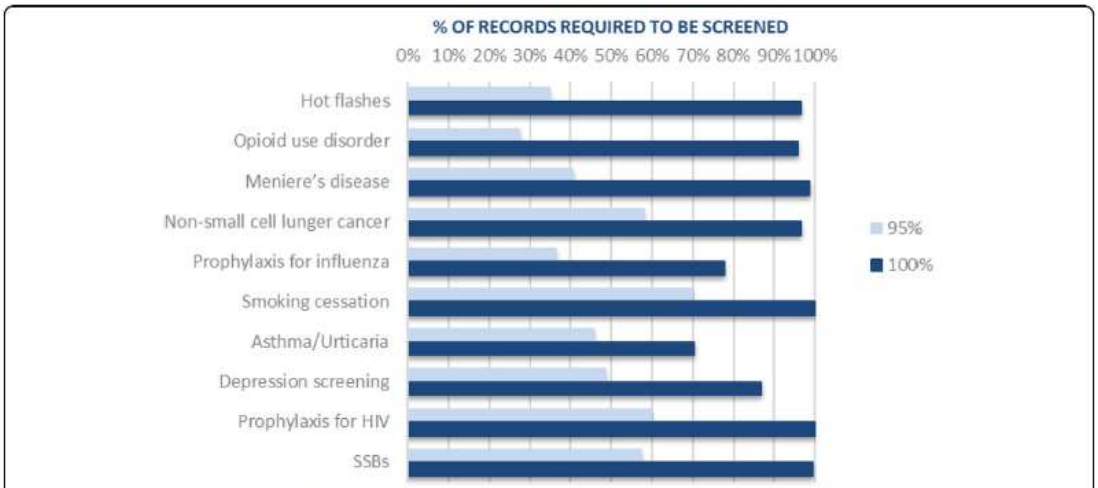


Fig. 6 Screening burden to achieve true recall @ 95% and @ 100%

DistillerSR produced similar results (median: 96.6%, range: 70.3 to 100%). As there were no final includes missed with the true recall @ 95%, the extra screening burden to identify the last 5% of studies would not have changed the final results and conclusions, and may not be worth the additional efforts. Although this could be further evaluated, this suggests this last 5% of records where passed through for full-text screening due to either human error or a tendency toward over-inclusiveness while screening titles/abstract, and/or title/abstracts that were unclear, or records with no abstract which were included based on the title only. Other research teams are encouraged to use the information we have provided in order to build the evidence base.

There are several considerations to keep in mind when deciding to use prioritized screening in prospective reviews. It is important to have a clean (e.g., all duplicates removed) dataset, as any duplicates with conflicting decisions on whether to pass through for full-text screening or exclude based on the title/abstract would confuse the machine learning algorithm. Due to the retrospective nature of this experiment, this was not checked, as the assumption was made that this was performed when the SRs were originally conducted. Second, as the success of machine learning is dependent on the quality of the training set created by human reviewers, a precise training set (i.e., correctly designating title/abstract records) is required. A 2020 study by Wang et al. reported a 10.8% (95% confidence interval 7.4 to 14.1%) error rate (i.e., incorrectly included or incorrectly excluded at title/abstract screening) among 139,467 citations that underwent 329,332 inclusion and exclusion decisions [3]. Although incorrectly excluding a record at title and abstract level is more concerning, as this record is no longer considered for inclusion, incorrectly passing a record at title and abstract for further review at full text increases screening burden at full text, in addition to the time and costs associated with retrieving the full-text articles. It is therefore important to ensure that a pilot test is first performed with conflicts resolved, that all reviewers are confident in their assessments (i.e., do not include because of uncertainty of reviewer rather than uncertainty of relevance), and that conflict resolution is performed throughout screening. Review team may also choose to set up reviewer compatibility (if the software permits), where junior reviewers are unable to screen the same references. This may decrease the number of records that are incorrectly included due to uncertainty.

Limitations

There were some limitations in the conducted study. First, screening at the title and abstract level in the set of systematic reviews we studied was performed using the liberal accelerated method [34], which requires two reviewers to exclude a reference, but requires only one reviewer to include a reference to be further evaluated at full text. Further, any conflicts resulting from the first reviewer excluding and the second reviewer including were not resolved. This presents two limitations: (i) there may be a tendency to be over-inclusive while screening titles/abstracts as only one reviewer is required to pass the reference through for further full-text screening; and (ii) by using retrospective responses, the machine-learning algorithm is not able to distinguish between records that were excluded by the first reviewer and later included by the second reviewer. These records

may be less likely to be true includes. As a training set with high accuracy (i.e., true title/abstract includes and true excludes) will result in fewer excluded references required to be screened to achieve true recall @ 95%, over-inclusiveness of records likely resulted in poorer performance of the AI tool. Second, this experiment was only conducted using DistillerSR, which might not be generalizable to all prioritization algorithms and related software.

Implications for future research

In this pilot experiment evaluating the AI simulation tool in DistillerSR, we selected 10 reviews which included a variety of review types (e.g., NMAs, SRs of RCTs, SRs including observational studies), sizes (ranging from 2250 to 22,309 records), and inclusion rates (ranging from 3.0 to 39.2% at title/abstract screening). We encourage other review teams to use the guidance provided in Additional file 2 to evaluate the AI simulation tool on their own projects. For review teams who do not have access to DistillerSR or who do not have the resources to run these experiments, the authorship team of this study plans on increasing the sample size of this experiment by asking other review teams to provide their databases so this experiment can be run. We plan to establish a website for this work that will allow for the provision of updated findings in an ongoing fashion. Offers to contribute to this initiative will be shared with other teams in the future through email, social media and other forms of communication.

In the context of rapid reviews, a form of knowledge synthesis that accelerates the process of conducting a traditional systematic review through streamlining or omitting a variety of methods to produce evidence in a timely and resource-efficient manner [27–30], identification of fewer than 95% of the title/abstract true positives may be acceptable. A survey of stakeholders (e.g., policy-makers, healthcare providers) reported that the median acceptable incremental risk of getting an incorrect answer from a rapid review is 10% (interquartile range of 5–15%) [35]. A missed study (or studies) does not imply there will be an incorrect answer, depending on the study (ies), as missed studies may not change the overall conclusion appreciably in terms of either direction or magnitude of effects studied. Therefore, the decision to stop screening or change the method of screening (e.g., from dual-independent to single screener) once another percentage of studies passed through for full-text review have been identified (e.g., 75, 85%) may be further evaluated.

As true recall can only be calculated once all records are screened, estimated recall might differ depending on how quickly relevant records (at title/abstract) are identified. For example, an estimated recall @ 95% may only be accounting for 91% of the included records if all were screened. Therefore, a review team might not be comfortable to implement a modified or stop screening approach when an estimated recall of 95% is first achieved. They may consider screening an additional set of records (e.g., two to four more iterations) to confirm no new title/abstract records are passed through for full-text screening. Estimated recall rates may be further evaluated to determine the difference between estimated and true recall rates and how many more records should be screened once a certain estimated recall threshold has been achieved.

Prospective studies using the prioritization tool should be performed that report transparent and repeatable methods. These steps might change the process by which review teams currently conduct their systematic reviews. For example, although not an option when using the AI simulation on a previously completed review, in a prospective review using prioritization, review teams are encouraged to use dual-independent screening at the title and abstract level, with conflicts resolved throughout the screening process (e.g., after every 10% of references screened, at the end of each day) to minimize over-inclusiveness and maximize the performance of the AI prioritization tool. Review teams are also encouraged to use the *Check for Error* audit throughout screening to ensure that no references are incorrectly excluded, although this should be rare when performing dual-independent screening. Prospective studies may contribute to a set of best practices for using prioritized screening, and may also help to inform a future reporting checklist for protocols and manuscripts for these types of experiments or for reviews (e.g., systematic, rapid) using AI.

Conclusion

Our findings from this study suggest that the prioritization tool in DistillerSR can reduce screening burden. Even for reviews where the tool performed less efficiently, the time savings were still appreciable. Modified or stop screening approaches once a true recall @ 95% has been achieved appears to be a valid method for rapid reviews, and perhaps systematic reviews, as it did not miss any of the final includes studies in the systematic review.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s12874-020-01129-1.

Additional file 1. Systematic review details [36–44].

Additional file 2. Steps for Testing Prioritization in DistillerSR through AI Simulation.

Additional file 3. Supplementary Tables: Table 1. Records required to screen to achieved true recall @ 95%. **Table 2.** Total hours saved.

Abbreviations

AI: Artificial intelligence; HIV: Human immunodeficiency viruses; IQR: Interquartile range; NMA: Network meta-analysis; RCT: Randomized

controlled trial; SD: Standard deviation; SR: Systematic review; SSBs: Sugar sweetened beverages

Acknowledgements
Not applicable.

Authors' contributions
CH and BH designed the experiment. CH wrote the original draft of the manuscript. All authors (CH, SEK, KT, DBR, GAW, BH) reviewed the manuscript, provided input and were involved in development of the associated funding application for this work. The author(s) read and approved the final manuscript.

Funding
This research is being funded by a bridge grant provided by the Canadian Institutes of Health Research. The funder has not had any role in study design; in the collection, analysis and interpretation of data; in the writing of the report; or in the decision to submit the article for publication.

Availability of data and materials
The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Clinical Epidemiology Program, Ottawa Hospital Research Institute, 501 Smyth Road, Box 201b, Ottawa, Ontario K1H 8L6, Canada. [2]Department of Medicine, University of Split, Split, Croatia. [3]Cardiovascular Research Methods Centre, University of Ottawa Heart Institute, Ottawa, Ontario, Canada. [4]School of Epidemiology and Public Health, University of Ottawa, Ottawa, Ontario, Canada. [5]Department of Psychology, McGill University, Montreal, Quebec, Canada.

Received: 11 June 2020 Accepted: 22 September 2020
Published online: 15 October 2020

References

1. About Cochrane reviews 2020. https://www.cochranelibrary.com/about/about-cochrane-reviews. Accessed 16 Jan 2020.
2. Borah R, Brown A, Capers P, Kaiser K. Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry. BMJ Open. 2017;7:e012545.
3. Wang Z, Nayfeh T, Tetzlaff J, O'Blenis P, Murad MH. Error rates of human reviewers during abstract screening in systematic reviews. PLoS One 2020; 15:e0227742. doi: https://doi.org/10.1371/journal.pone.0227742.
4. Taljaard M, McGowan J, Grimshaw JM, Brehaut JC, McRae A, Eccles MP, et al. Electronic search strategies to identify reports of cluster randomized trials in MEDLINE: low precision will improve with adherence to reporting standards. BMC Med Res Methodol. 2010;10:15. https://doi.org/10.1186/1471-2288-10-15.
5. 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;19:132. https://doi.org/10.1186/s12874-019-0782-0.
6. 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. https://doi.org/10.2147/CLEP.S43118.
7. Wang Z, Asi N, Elraiyah TA, Abu Dabrh AM, Undavalli C, Glasziou P, et al. Dual computer monitors to increase efficiency of conducting systematic reviews. J Clin Epidemiol. 2014;67:1353–7. https://doi.org/10.1016/j.jclinepi.2014.06.011.
8. Rathbone J, Albarqouni L, Bakhit M, Beller E, Byambasuren O, Hoffmann T, et al. Expediting citation screening using PICo-based title-only screening for identifying studies in scoping searches and rapid reviews. Syst Rev. 2017;6:233. https://doi.org/10.1186/s13643-017-0629-x.
9. Pham MT, Waddell L, Rajić 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 Synth Methods. 2016;7:433–46. https://doi.org/10.1002/jrsm.1215.
10. Edwards P, Clarke M, DiGuiseppi C, Pratap S, Roberts I, Wentz R. Identification of randomized controlled trials in systematic reviews: accuracy and reliability of screening records. Stat Med. 2002;21:1635–40. https://doi.org/10.1002/sim.1190.
11. 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 Rev. 2016;5:140. https://doi.org/10.1186/s13643-016-0315-4.
12. Gartlehner G, Affengruber L, Titscher V, Noel-Storr A, Dooley G, Ballarini N, et al. Single-reviewer abstract screening missed 13 percent of relevant studies: a crowd-based, randomized controlled trial. J Clin Epidemiol. 2020. https://doi.org/10.1016/j.jclinepi.2020.01.005.
13. Gartlehner G, Wagner G, Lux L, Affengruber L, Dobrescu A, Kaminski-Hartenthaler A, et al. Assessing the accuracy of machine-assisted abstract screening with DistillerAI: a user study. Syst Rev. 2019;8:277. https://doi.org/10.1186/s13643-019-1221-3.
14. Gates A, Guitard S, Pillay J, Elliott SA, Dyson MP, Newton AS, et al. Performance and usability of machine learning for screening in systematic reviews: a comparative evaluation of three tools. Syst Rev. 2019;8:278. https://doi.org/10.1186/s13643-019-1222-2.
15. Tsou AY, Treadwell JR, Erinoff E, Schoelles K. Machine learning for screening prioritization in systematic reviews: comparative performance of Abstrackr and EPPI-reviewer. Syst Rev. 2020;9:73.
16. Howard BE, Phillips J, Miller K, Tandon A, Mav D, Shah MR, et al. SWIFT-review: a text-mining workbench for systematic review. Syst Rev. 2016;5:87. https://doi.org/10.1186/s13643-016-0263-z.
17. O'Mara-Eves A, Thomas J, McNaught J, Miwa M, Ananiadou S. Using text mining for study identification in systematic reviews: a systematic review of current approaches. Syst Rev. 2015;4:5. https://doi.org/10.1186/2046-4053-4-5.
18. 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;20:7. doi: https://doi.org/10.1186/s12874-020-0897-3.
19. Marshall IJ, Wallace BC. Toward systematic review automation: a practical guide to using machine learning tools in research synthesis. Syst Rev. 2019; 8:163. https://doi.org/10.1186/s13643-019-1074-9.
20. Gates A, Johnson C, Hartling L. Technology-assisted Title and Abstract Screening for Systematic Reviews: A Retrospective Evaluation of the Abstrackr Machine Learning Tool. Systematic Reviews. 2018;7:45. https://doi.org/10.1186/s13643-018-0707-8.
21. Rathbone J, Hoffmann T, Glasziou P. Faster Title and Abstract Screening? Evaluating Abstrackr, a Semi-Automated Online Screening Program for Systematic Reviewers. Syst Rev 2015;4:80. doi: https://doi.org/10.1186/s13643-015-0067-6.
22. Giummarra MJ, Lau G, Gabbe BJ. Evaluation of text mining to reduce screening workload for injury-focused systematic reviews. Inj Prev. 2020;26:55–60. https://doi.org/10.1136/injuryprev-2019-043247.
23. Howard BE, Phillips J, Tandon A, Maharana A, Elmore R, Mav D, et al. SWIFT-active screener: accelerated document screening through active learning and integrated recall estimation. Environ Int. 2020;138:105623. https://doi.org/10.1016/j.envint.2020.105623.
24. 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;20:139. doi: https://doi.org/10.1186/s12874-020-01031-w.
25. 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;8:143. https://doi.org/10.1186/s13643-019-1062-0.
26. Cohen AM, Hersh WR, Peterson K, Yen P-Y. Reducing workload in systematic review preparation using automated citation classification. J Am Med Inform Assoc. 2006;13:206–19. https://doi.org/10.1197/jamia.M1929.
27. Ganann R, Ciliska D, Thomas H. Expediting systematic reviews: methods and implications of rapid reviews. Implement Sci. 2010;5:56. https://doi.org/10.1186/1748-5908-5-56.

28. Tricco AC, Antony J, Zarin W, Strifler L, Ghassemi M, Ivory J, et al. A scoping review of rapid review methods. BMC Med. 2015;13:224. https://doi.org/10.1186/s12916-015-0465-6.
29. Watt A, Cameron A, Sturm L, Lathlean T, Babidge W, Blamey S, et al. Rapid reviews versus full systematic reviews: an inventory of current methods and practice in health technology assessment. Int J Technol Assess Health Care. 2008;24:133–9.
30. Grant MJ, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. Health Inf Libr J. 2009;26:91–108. https://doi.org/10.1111/j.1471-1842.2009.00848.x.
31. Elliott JH, Synnot A, Turner T, Simmonds M, Akl EA, McDonald S, et al. Living systematic review: 1. Introduction-the why, what, when, and how. J Clin Epidemiol. 2017;91:23–30. https://doi.org/10.1016/j.jclinepi.2017.08.010.
32. Thomas J, Noel-Storr A, Marshall I, Wallace B, McDonald S, Mavergames C, et al. Living systematic reviews: 2. Combining human and machine effort. J Clin Epidemiol. 2017;91:31–7. https://doi.org/10.1016/j.jclinepi.2017.08.011.
33. Munn Z, Stern C, Aromataris E, Lockwood C, Jordan Z. What kind of systematic review should I conduct? A proposed typology and guidance for systematic reviewers in the medical and health sciences. BMC Med Res Methodol. 2018;18:5. https://doi.org/10.1186/s12874-017-0468-4.
34. Khangura S, Konnyu K, Cushman R, Grimshaw J, Moher D. Evidence summaries: the evolution of a rapid review approach. Syst Rev. 2012;1:10. https://doi.org/10.1186/2046-4053-1-10.
35. 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. https://doi.org/10.1186/s12874-017-0406-5.
36. Hutton B, Yazdi F, Bordeleau L, Morgan S, Cameron C, Kanji S, et al. Comparison of physical interventions, behavioral interventions, natural health products, and pharmacologics to manage hot flashes in patients with breast or prostate cancer: protocol for a systematic review incorporating network meta-analyses. Syst Rev. 2015;4:114. https://doi.org/10.1186/s13643-015-0099-y.
37. Rice DB, Hutton B, Poulin P, Sproule BA, Wolfe D, Moher D, et al. Evaluating comparative effectiveness of psychosocial interventions for persons receiving opioid agonist therapy for opioid use disorder: protocol for a systematic review. BMJ Open. 2018;8:e023902. https://doi.org/10.1136/bmjopen-2018-023902.
38. Ahmadzai N, Cheng W, Wolfe D, Bonaparte J, Schramm D, Fitzpatrick E, et al. Pharmacologic and surgical therapies for patients with Meniere's disease: a protocol for a systematic review and meta-analysis. Syst Rev. 2019;8:341. https://doi.org/10.1186/s13643-019-1195-1.
39. Elliott J, Bai Z, Hsieh S-C, Kelly SE, Chen L, Skidmore B, et al. ALK inhibitors for non-small cell lung cancer: a systematic review and network meta-analysis. PLoS One. 2020;15:e0229179. https://doi.org/10.1371/journal.pone.0229179.
40. Hersi M, Traversy G, Thombs BD, Beck A, Skidmore B, Groulx S, et al. Effectiveness of stop smoking interventions among adults: protocol for an overview of systematic reviews and an updated systematic review. Syst Rev. 2019;8:28. https://doi.org/10.1186/s13643-018-0928-x.
41. Johnston A, Smith C, Zheng C, Aaron SD, Kelly SE, Skidmore B, et al. Influence of prolonged treatment with omalizumab on the development of solid epithelial cancer in patients with atopic asthma and chronic idiopathic urticaria: a systematic review and meta-analysis. Clin Exp Allergy. 2019;49:1291–305. https://doi.org/10.1111/cea.13457.
42. Hamel C, Lang E, Morissette K, Beck A, Stevens A, Skidmore B, et al. Screening for depression in women during pregnancy or the first year postpartum and in the general adult population: a protocol for two systematic reviews to update a guideline of the Canadian task force on preventive health care. Syst Rev. 2019;8:27. https://doi.org/10.1186/s13643-018-0930-3.
43. Hamel C, Stevens A, Singh K, Ansari MT, Myers E, Ziegler P, et al. Do sugar-sweetened beverages cause adverse health outcomes in adults? A systematic review protocol. Syst Rev. 2014;3:108. https://doi.org/10.1186/2046-4053-3-108.
44. Stevens A, Hamel C, Singh K, Ansari MT, Myers E, Ziegler P, et al. Do sugar-sweetened beverages cause adverse health outcomes in children? A systematic review protocol. Syst Rev. 2014;3:96. https://doi.org/10.1186/2046-4053-3-96.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

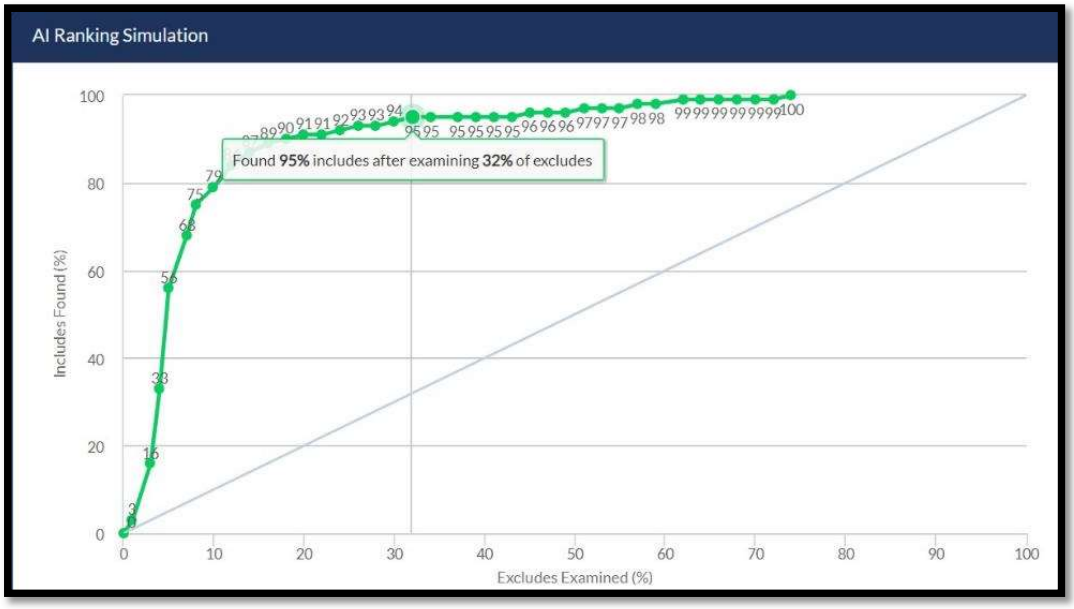
At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

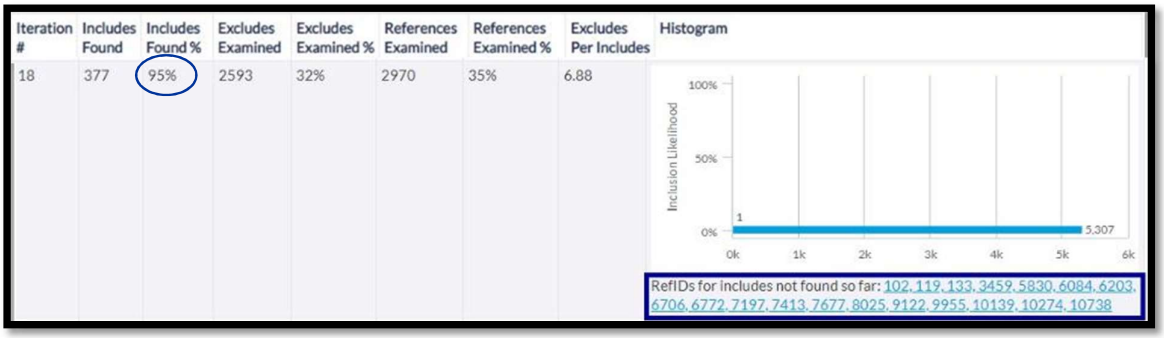


Appendix 1. AI Ranking Simulation output

The following bar chart is displayed, and updated throughout the simulation process, to show how many included studies have been identified (y-axis) in each iteration and how many excluded studies were examined (x-axis).



Below the simulation chart, a row for each iteration is provided to show the iteration number, and the numerical values for that iteration for the following: the included found, the % of the includes found, the excludes examined, the % of the excludes examine, the total number of references examined up to and including that iteration, the % of the total number of references examined, and the ratio of excludes per includes. On the far right side, a histogram is provided with the number of records yet to be 'screened' and the % value of the likelihood of inclusion. Below this histogram, it provides the reference IDs number for the 5% of the studies that were included at the title and abstract level, but had not been identified yet (i.e., title and abstract false negatives).



Iteration #	Includes Found	Includes Found %	Excludes Examined	Excludes Examined %	References Examined	References Examined %	Excludes Per Includes	Histogram
18	377	95%	2593	32%	2970	35%	6.88	

RefIDs for includes not found so far: 102, 119, 133, 3459, 5830, 6084, 6203, 6706, 6772, 7197, 7413, 7677, 8025, 9122, 9955, 10139, 10274, 10738