SWAR 49: Artificial Intelligence in screening within rapid reviews: quantifying the impact on main findings, certainty of evidence and resources required

Objective of this SWAR
This Study Within a Review (SWAR) aims to explore the use of artificial intelligence (AI) to automate and expedite the rapid review process. Specifically, it aims to answer the following questions:
1. How does the use of different AI features compare to manual screening on number of missed studies, and how are overall findings of the review impacted (i.e., certainty of evidence, magnitude and direction of effect)?
2. How does the use of AI compare to manual screening on time to complete a rapid review?

Study area: Study Identification, Review resources
Sample type: Review Authors
Estimated funding level needed: Low

Background
Rapid reviews are "a form of knowledge synthesis in which components of the systematic review process are simplified or omitted to produce information within a timely manner" [1]. They have become increasingly common as a way to provide timely and relevant evidence to inform decision-making. Given that systematic reviews may take from six to 24 months to complete, rapid reviews play an important role in the evidence-informed decision-making process. The simplification or omission of components (i.e. ‘shortcuts’) can occur at any stage within the traditional systematic review process, including development of the research question or search strategy, title/abstract or full text screening, data extraction, critical appraisal and synthesis. While ‘short cuts’ help to accelerate the completion of the review (benefit), they introduce more opportunity for bias and/or human error in the review process (cost). An evidence-based understanding of both the costs and benefits (financial and opportunity) of each shortcut is therefore vital for both evidence synthesis teams undertaking rapid reviews, and for decision makers requesting an evidence synthesis [2].

Artificial Intelligence (AI) has been proposed as one way to automate and expedite the review process to maintain rapid timelines while reducing the potential bias or error introduced by omitting a step altogether. For example, AI has been used to rapidly craft a search strategy, act as a main or secondary screener, and in data extraction, critical appraisal and synthesis. While anecdotally, many evidence synthesis teams are using AI features, there is limited evidence to quantify the impact of using AI within the review process in terms of accuracy of findings and time saved. In February 2024, the Cochrane Rapid Review Methods Group published an updated guidance document with 24 recommendations for the conduct of rapid reviews [3]. None of these recommendations included guidance on the use of AI. A recent systematic review evaluating the efficacy of AI tools in automation of systematic reviews in cancer research identified five studies that evaluated four different AI tools used for what stage of review process [4]. The authors concluded that AI tools were "promising" but displayed varying levels of accuracy and efficiency. A retrospective study evaluated the use of AI in 10 historical reviews, proposing preliminary guidance for use at each step [5]. While there is evidence to suggest that screening done solely with AI is inferior to manual screening, a more nuanced approach is needed to understand the impacts of different AI features in alone or in combination with manual screening within rapid reviews.

This Study Within a Review (SWAR) [6] will be conducted within the title and abstract and full text screening phases of rapid reviews conducted within the National Collaborating Centre for Methods and Tools (NCCMT)'s Rapid Evidence Service during the study period (estimated 8-10) using DistillerSR version 2.35 [7]. Two specific AI features, Re-Rank and Check Screening Errors, will be tested and results compared to manual dual screening.

Interventions and Comparators
Intervention 1: Use of two specific AI features, Re-Rank, and Check Screening Errors during the screening process.
Intervention 2: Manual dual screening.

Index Type: Screening

Method for Allocating to Intervention or Comparator:
Not applicable

Outcome Measures
Primary: Number of missed studies, per threshold; impact of missed studies on overall evidence certainty.
Secondary: Time to complete review.

Analysis Plans
1. Thresholds for predicted relevant references (Re-Rank tool)

As reviewers screen references for inclusion (2% of set, minimum 25, maximum 200), DistillerAI assigns a prediction score indicating the likelihood the reference will be included. As references are screened, the Re-Rank tool predicts the percentage of eligible references identified; this number is constantly refined as manual screening continues. At any point, reviewers have the option to stop manually screening and allow the AI to screen the remaining set based on prediction scores. A 95% threshold has been suggested as appropriate to stop manual screening [7]. As screening occurs in each review, we will clone the project as a new file at 60%, 70%, 80%, 85%, 90% and 95% thresholds and allow AI to screen remaining references. When manual dual screening is complete in the original project, we will identify how many studies would have been missed if each threshold were used.

2. Check for screening errors (Check Screening Errors tool)

The Check Screening Errors tool is used to identify potential false excludes by manual or AI screening. The AI trains itself multiple times using random samples for accuracy and performance. Following complete screening (manually, and by AI at each of the pre-determined thresholds above) we will run the Check for Screening Errors tool to identify how many additional studies would have been included using this tool for both manual and AI-assisted screening. Standard calculations for specificity, sensitivity and accuracy of screening at each threshold, with and without the use of the Check for Screening Errors function will be calculated using formulas below. Dual manual screening will be the test standard.

Specificity = True exclude / (True exclude + False include)
Sensitivity = True include / (True include + False exclude)
Accuracy = True include + True exclude / (all references)

3. Impact on evidence certainty

We will assess whether the key findings of the review, and the certainty of evidence (assessed using the GRADE) approach would have changed if studies excluded at each of the thresholds above had been left out of the rapid review.

4. Time to complete review

Time will be tracked across each review by all reviewers at each review step using an online time tracker and team spreadsheet to calculate the time saved using each feature and compare it to the cost (of falsely identified study) per time saved.

Possible Problems in Implementing This SWAR

Consistently capturing each screening threshold (i.e., unpredictability of AI screening).

References

1. Tricco AC, Langlois EV, Straus SE (editors). Rapid reviews to strengthen health policy and systems: A practical guide. World Health Organization. 2017.
2. Hamel C, Michaud A, Thuku M, et al. Few evaluative studies exist examining rapid review methodology across stages of conduct: a systematic scoping review. Journal of Clinical Epidemiology 2020;126:131-40.
3. Garritty C, Hamel C, Trivella M, et al. Updated recommendations for the Cochrane rapid review methods guidance for rapid reviews of effectiveness. BMJ 2024;384:e076335.
4. Yao X, Kumar MV, Su E, et al. Evaluating the efficacy of artificial intelligence tools for the automation of systematic reviews in cancer research: A systematic review. Cancer Epidemiology 2024;88:102511.
5. Hamel C, Hersi M, Kelly SE, et al. Guidance for using artificial intelligence for title and abstract screening while conducting knowledge syntheses. BMC Medical Research Methodology 2021;21:285.
6. Devane D, Burke NN, Treweek S, et al. Study within a review (SWAR). Journal of Evidence-Based Medicine 2022;15(4):328-32.
7. Neil-Sztramko SE, Belita E, Traynor RL, et al. Methods to support evidence-informed decision-making in the midst of COVID-19: Creation and evolution of a rapid review service from the National Collaborating Centre for Methods and Tools. BMC Medical Research Methodology 2021;21:231.

Publications or presentations of this SWAR design

Examples of the implementation of this SWAR

People to show as the source of this idea: Dr. Sarah Neil-Sztramko
Contact email address: neilszts@mcmaster.ca
Date of idea: 07/10/2024
Revisions made by:
Date of revisions: