SWAR 32: Accuracy of automated deduplication tools during screening for studies to include in systematic reviews

Objective of this SWAR
To evaluate the accuracy of a set of automated deduplication tools to identify duplicates in the study screening process for systematic reviews.

Study area: evidence synthesis
Sample type: Review Authors
Estimated funding level needed: Unfunded

Background
Conducting a high-quality systematic review requires significant time and resources. One particularly time-consuming step is study screening and selection. Conducting searches across multiple databases to ensure comprehensive coverage and capture of all relevant evidence for the research questions often leads to the retrieval of duplicate records, which need to be removed before screening to avoid checking them multiple times.

Efficiently identifying and removing duplicate records can save reviewers considerable time. Moreover, it helps prevent the inclusion of multiple records for the same study, which could introduce bias into the results of a systematic review [1]. A range of automated tools are available for the task, such as Endnote, Zotero, Covidence and SRA Deduplicator. However their accuracy in identifying duplicates varies due to the different rules and logic they use. JBI SUMARI is also in the process of developing this feature and it will be available in 2024.

This Study Within a Review (SWAR) [2] will assess how accurate these tools are in identifying duplicate records. The search strategy used in this SWAR was developed for the systematic review which aimed to synthesize the experiences of informal caregivers of people with dementia in low- and middle-income countries, and is registered at PROSPERO under CRD42023453814 [3].

Interventions and Comparators
Intervention 1: Manual triage of duplicates by two independent reviewers using Microsoft Excel.
Intervention 2: Covidence: Covidence's deduplication method is based on title, year, volume and author. After importing the references, this tool automatically identifies duplicate records, which can then be checked by the reviewer.
Intervention 3: Deduplicator (Systematic Review Accelerator): Deduplicator's method is based on a relaxed search, considering title, author and year, and a focussed search based on title and year. After importing the references, it divides the uploaded references into four groups: extremely likely duplicates, highly likely duplicates, likely duplicates and non-duplicates.
Intervention 4: JBI SUMARI: JBI SUMARI's deduplication method is based on the DOI, title, author and year. After importing the references, it divides the uploaded references into three groups: exact duplicates, possible duplicates and non-duplicates.
Intervention 5: Rayyan: Rayyan identifies possible duplicates using title, authors, journal and year. It automatically deduplicates 100% exact matches and identifies possible duplicates, giving them a similarity percentage.
Intervention 6: Endnote: Endnote considers reference type, author, year and title, and can ignore spacing and punctuation.
Intervention 7: Ref works: Ref works can identify exact duplicates based on author, title and year, and close duplicates based on a combination of author, title and year.
Intervention 8: Zotero: Zotero considers title, DOI and ISBN. If these fields match (or are absent), Zotero also compares the years of publication (matching records if they are within a year of each other) and author/creator lists (if there is a match for at least one author last name plus first initial) to determine duplicates.
Intervention 9: Mendeley: Mendeley displays "sets" of documents which have potential duplicates. A "set" is any number of documents which have some identical information (not specified).

Index Type: Title

Method for Allocating to Intervention or Comparator:

Outcome Measures

Primary: According to PRISMA 2020 [4], records that refer to the same report are considered as “duplicates”. A "report" is defined as any document, whether paper or electronic, that provides details about a specific study. It could be a journal article, preprint, conference abstract, study register entry, clinical study report, dissertation, unpublished manuscript, government report or any other document providing relevant information. We will calculate the sensitivity (true positive rate) and specificity (true negative rate) and the standard error (SE) of these measures. To evaluate the proportion of true duplicates within the records removed by each tool, precision (positive predictive value) will be calculated. True-positive (TP) records are those defined as a duplicate citation legitimately removed from the combined results of database searching. A true-negative (TN) is a record that is correctly deemed to be a non-duplicate. False-negatives (FN) are records that should have been identified as duplicates but were not. Records that were removed from the combined search results but should not have been removed are false-positive (FP) duplicates.
Sensitivity = TP/(TP + FN)
Specificity = TN/(TN + FP)
Precision = TP/(TP + FP)
SE for sensitivity = sensitivity (1 − sensitivity) / (true positive + false negative)
SE for specificity = specificity (1 − specificity) / (false positive + true negative)

Secondary: Time (in minutes) required for each method. For automation tools, this will be recorded from the moment of data import to the generation of results. For manual triage of duplicates, time will be tracked from the start to the completion of screening for both independent reviewers.

Analysis Plans

Searches for the host review were conducted on 26 March 2024 in MEDLINE (Ovid), CINAHL (EBSCOhost), Embase (Ovid), PsycINFO (Ovid), AgeLine (EBSCOhost), LILACS (BVS), African Index Medicus (BVS) and Web of Science Core Collection (WOS). The results files were exported in RIS formats from all platforms with a total of 13,958 records. A manual triage and the automated tools will be applied to discriminate duplicate and non-duplicate records. The manual triage of duplicate records will be considered as the reference standard method of deduplication in this SWAR and will be performed by two independent reviewers in EndNote.

Possible Problems in Implementing This SWAR

References

1. Tramèr MR, Reynolds DJM, Moore RA, McQuay HJ. Impact of covert duplicate publication on meta-analysis: a case study. BMJ 1997;315(7109):635-40.
2. Devane D, Burke NN, Treweek S, et al. Study within a review (SWAR). Journal of Evidence Based Medicine 2022;15(4):328-32.
3. Ho HT, Jia R, Habibi N, et al. Experiences of informal caregivers of people with dementia in low-and middle-income countries: a systematic review protocol. JBI Evidence Synthesis 2024;10.11124.
4. Page MJ, Moher D, Bossuyt PM, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. BMJ 2021;372:n160.

Publications or presentations of this SWAR design

Guimarães NS, Ferreira AJF, Ribeiro Silva RC, de Paula AA, Lisboa CS, Magno L, Ichiara MY, Barreto ML. Deduplicating records in systematic reviews: there are free, accurate automated ways to do so. Journal of Clinical Epidemiology 2022;152:110-5.

Examples of the implementation of this SWAR

People to show as the source of this idea: Romy Menghao Jia, Cindy Stern, Edoardo Aromataris, Hien Ho
Contact email address: romy.jia@adelaide.edu.au
Date of idea: 01/04/2024
Revisions made by:
Date of revisions: