Journal of Clinical Epidemiology 152 (2022) 110−115

ORIGINAL ARTICLE

Deduplicating records in systematic reviews: there are free, accurate automated ways to do so

Nathalia Sernizon Guimarães[a,1,*], Andrêa J.F. Ferreira[b,c,1], Rita de Cássia Ribeiro Silva[b,d], Adelzon Assis de Paula[a], Cinthia Soares Lisboa[e], Laio Magno[a,f], Maria Yury Ichiara[b], Maurício Lima Barreto[a,b]

[a]Institute of Collective Health. Federal University of Bahia, Salvador, Bahia, Brazil
[b]Centre for Data and Knowledge Integration for Health (CIDACS), Oswaldo Cruz Foundation, Salvador, Bahia, Brazil
[c]The Ubuntu Center on Racism, Global Movements, and Population Health Equity, Dornsife School of Public Health, Drexel University, Philadelphia, PA, USA
[d]Department of Nutrition, School of Nutrition, Federal University of Bahia, Salvador, Bahia, Brazil
[e]Pos-graduation programme of Collective Health, State University of Feira de Santana, Feira de Santana, Bahia, Brazil
[f]Department of Life Sciences, State University of Bahia, Salvador, Bahia, Brazil

Accepted 4 October 2022; Published online 12 October 2022

Abstract

Objective: Here, we examined the accuracy measures of a set of automated deduplication tools to identify duplicate in the eligibility process of systematic reviews.

Study Design and Setting: A planned search strategy was carried out on seven electronic databases until May 31, 2021. Using manual search as the reference standard, we assessed sensibility, specificity, negative predictive value, and positive predictive value (PPV).

Results: Specificity ranged from 0.96 to 1.00. Rayyan, Mendeley, and Systematic Review Accelerator (SRA) presented high sensibility (0.98 [95% CI = 0.94−1.00]; 0.93 [95% CI = 0.88−0.97] and 0.90 [95% CI = 0.84−0.95], respectively), whereas EndNote X9 and Zotero had only fair sensitivity (0.73 [95% CI = 0.65−0.80] and 0.74 [95% CI = 0.66−0.81], respectively). Negative predictive value ranged from 0.99 to 1.00. Mendeley and SRA had good PPV (0.93 [95% CI = 0.88−0.97] and 0.99 [95% CI = 0.96−1.00], respectively). PPV was fair for EndNote X9 (0.61 [95% CI = 0.54−0.69]) and Zotero (0.62 [95% CI = 0.54−0.69]) and poor for Rayyan (0.41 [95% CI = 0.36−0.47]).

Conclusion: Choosing the most suitable tool depends on its interface's characteristics, the algorithm to identify and exclude duplicates, and the transparency of the process. Therefore, Rayyan, Mendeley, and SRA proved to be accurate enough for the systematic reviews' deduplication step.

Keywords: Accuracy; Deduplication; Systematic review; Libraries; Epidemiological research; Nutrition research methodologies

Conflicts of interest: The authors declare that the research was conducted without any commercial or financial relationship that could be considered a potential conflict of interest.

Author contributions: N.S. and A.F. contributed to the study concept and design. N.S. and A.F. perform the duplication process of all records manually and with automated tools. N.S., A.F., and A.A.P. analyzed the data. N.S. and A.F. drafted the initial manuscript. N.S., A.F., R.C., A.A.P., C.S., L.M., M.Y.I., and M.L.B. contributed to the script and read the final version of the manuscript. All authors read and approved the final version of the manuscript.

Funding: This research received no specific grant.

Declaration of interests: The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: On behalf of Andrêa Jacqueline Fortes Ferreira, Rita de Cássia Ribeiro Silva, Adelzon Assis de Paula, Cinthia Soares Lisboa, Laio Magno, Maria Yuri Ichiara, and Maurício Barreto.

[1] These authors have contributed equally to this work and share first authorship.

* Corresponding author: Nathalia Sernizon Guimarães, Federal University of Bahia, Brazil. Tel.: +55 031 997772844.
E-mail address: nasernizon@gmail.com (N.S. Guimarães).

https://doi.org/10.1016/j.jclinepi.2022.10.009

What is new?

- To compare effectiveness of different options for de-duplicating records retrieved from systematic review searches including free automated tools.

Key findings

- Rayyan, Mendeley, and Systematic Review Accelerator proved to be accurate enough for the deduplication step of systematic reviews.

What this adds to what was known?

- Our results on accuracy parameters from the available main automated tools could help researchers choose the most suitable tool based on a set of empirically defined values to conduct deduplication steps in systematic review studies.

1. Introduction

The ever-growing number of scientific publications related to Health Sciences has been evidenced over the last decades [1,2]. Hence, given the acquisition of a vast range of scientific information, health professionals struggle to manage scientific evidence for routine updates. Therefore, synthesizing scientific research evidence in the Health Sciences field is necessary to underpin better health practices and policies [1,2].

Systematic review studies are research designs that assist in analyzing and disseminating scientific evidence, using rigorous and well-defined methods to generate empirically derived answers to clear research questions [3]. Multiple database searching is essential to ensure broader coverage and identify all evidence relevant to the researched questions. In addition, searching distinct databases that use different indexing tools increases the likelihood of retrieving relevant references produced in the scientific literature [4], but unsurprisingly, duplicate records are retrieved due to an overlapping content of multiple databases [5].

Identifying and removing duplicate references is essential to ensure researchers do not waste time tracking down the same citation multiple times and preventing the inclusion of multiple records published from the same dataset, which could create bias in the conclusions of systematic review studies [3]. A range of automated tools is available for the task, whether free of charge, such as Mendeley [6] and Zotero [7], or not, such as EndNote X9 [8]. More recently, free and paid online automated deduplication tools have been developed to ease the production of systematic reviews, including Rayyan and Covidence, respectively. Covidence, for example, is made freely accessible for Cochrane authors, whereas nonauthors are granted a single, up to two reviewers free use—further utilization being charged [9,10].

Previous studies highlighted proceedings to identify and account for duplicate records by different methodologies (e.g, using data platforms as the OVID); however, independent databases have not been employed or the proper accuracy parameters were not measured [11−13]. Thus, despite the advances in the field, no analyses have been outlined to consistently determine the most suitable automated deduplication tool to be applied to independent databases, particularly the larger ones, because the manual search can be time-consuming and somewhat subjective in those scenarios. Here, we sought to examine the accuracy measures of a set of automated deduplication tools relative to the manual search to identify duplicate references in the eligibility process of systematic reviews.

2. Methods

2.1. Literature search

The search strategy used in this study was developed for the systematic review registered at PROSPERO under number CRD42021255570 [14], which aimed to answer the question "What is the impact of cash transfer programs on the health of children aged less than 5 years?"

To create a benchmark set of references, we planned a search strategy (N.S.G. and A.J.F.F.) for each of the following bibliographic databases: The Cochrane Central Register of Controlled Trials, PubMed, Embase, Latin American and Caribbean Health Sciences Literature, and SciELO; Web of Science as a citation database and Scopus as abstract and citation database and search for publication initial at 1,946 until May 31, 2021. Details of the search strategy are described in the supplementary appendix (Table S1). Manual search and automated tools were applied to discriminate duplicate an non-duplicate records in this benchmark set of of references.

2.2. Deduplication

A duplicate record was defined as the same bibliographic record (irrespective of how the citation details were reported, e.g., variations in page numbers, author details, accents used, or abridged titles) [13]. In case further reports from a single study were published, those were not classified as duplicates as they were merely multiple reports published across or within journals. A manuscript translation published in a distinct journal was not considered a duplicate record. Similarly, sets of conference abstracts, preprints, and journal articles describing the same research were not classified as duplicate records [13].

The manual search of duplicate records considered the standard reference method of deduplication in this study and was performed by two blinded independent reviewers (A.J.F.F. and N.S.G.) using the previous benchmark set of references. To improve the agreement between blinded reviewers, both underwent concordance training before the study period to identify duplicate records using a pilot sample of references (not included in this study). Using the benchmark of references, the two reviewers independently created a list of duplicate records. Then, we compared both lists and found a perfect agreement between all the records selected (kappa = 1.0) by each reviewer. The final list was recorded on a spreadsheet and all duplicate records were considered true positives.

In this study, we used the manual search as the reference standard procedure, through which a reference dataset was assembled. For that, two blinded reviewers screened and manually deduplicated the references, which were recorded on a Microsoft Excel spreadsheet.

Five automated tools (Rayyan, Systematic Review Accelerator [SRA], EndNote X9, Mendeley, and Zotero) were used to search for duplicate records. Below are briefly described the mechanisms used by each automated tool to identify duplicates:

i. Rayyan. Upon reference exporting, this automated tool creates a separate file with potential deduplicates left to be checked manually. This automated tool automatically identifies duplicate records, which can be deleted after checking for compliance, taking into account information such as the author's name, title, and year [10].
ii. SRA. It includes a validated algorithm for faster deduplication of search results from a benchmark set of references, allowing the researcher to manually double-check duplicate records identified by the algorithm as per an intuitive interface [15].
iii. Mendeley. This automated tool automatically identifies duplicate records, which can be deleted after checking for compliance, taking into account information such as the author's name, title, year, place, and publication type and other information that can be entered into the software [16].
iv. Zotero. After importing the set of references, Zotero automatically identifies possible duplicate records in a separate file, which can be manually checked, considering some fields, such as authors name, title, date of issue, and journal [7].
v. EndNote X9. This reference manager creates a separate group for presumable duplicate records, allowing researchers to visualize and judge duplicates manually [17].

The full-text versions of the citations were consulted whenever necessary to settle doubts. In such cases, we also checked the population sizes, methodology, and outcomes to determine whether the citations were duplicates or not [12].

2.3. Analysis

We assessed the main accuracy measures for the aforementioned automated tools relative to the manual search. Confusion matrices were constructed by cross-tabulating data regarding the valid status of the records (duplicate *vs.* nonduplicate) identified through manual search (the reference standard) vis-à-vis the results from the deduplication algorithms for each automated tool.

In this study, true-positive (TP) references were those defined as a non-duplicate citations legitimately removed

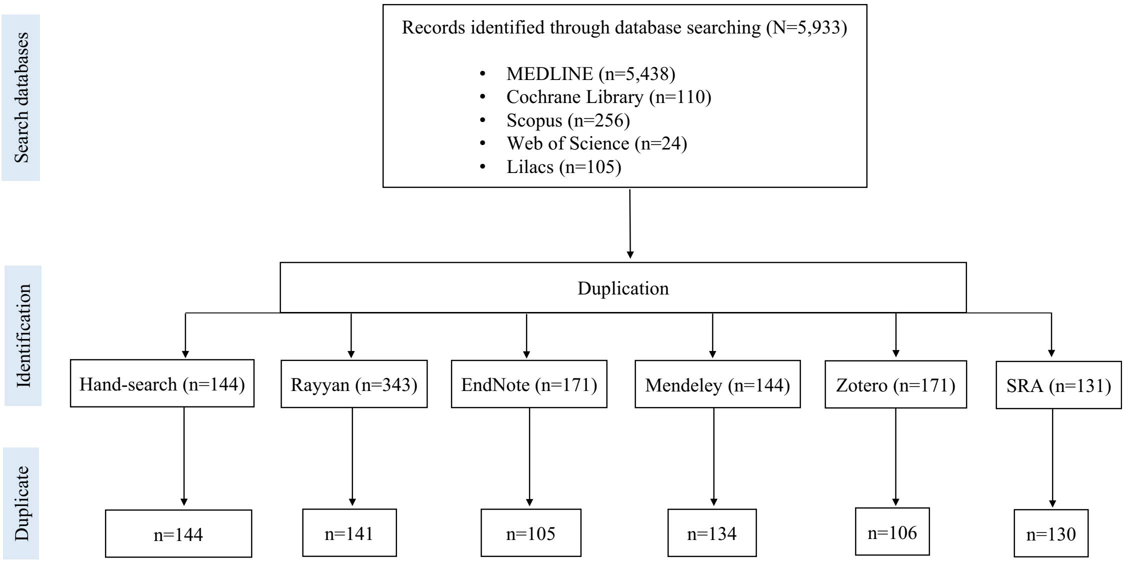

Fig. 1. Flow diagram of selected studies combining manual and automated tools of duplicate records in a set of Health databases. *n*, number of duplicates found.

from the final of database, that, is the true duplicates. On the other hand, kept in the final database, while a true-negative (TN) were the ones correctly deemed as non-duplicates − and hence kept in the database. False-negative (FN) were those that should have been tagged as duplicates and deleted from final database but were not. Conversely, references that were deleted from the databases but should not have been were classified as false-positive (FP) duplicates. Point estimates of sensitivity, specificity, and predictive values were determined as follows [18]:

Sensitivity = TP/(TP + FN).
Specificity = TN/(TN + FP).
Positive predictive value (PPV) = TP/(TP + FP).
Negative predictive value (NPV)= TN/(TN + FN).

Accuracy measures and their corresponding 95% confidence intervals (CIs) were estimated using the *caret* package [19]. We also calculated the balanced accuracy as an overall measure to account for class disproportions. All analyses were performed in the software R, version 4.1.2 (http://www.r-project.org).

3. Results

We retrieved 5,933 publications from the searched databases (Fig. 1). The manual search researchers retrieved 144 duplicated records, yielding a 0.02 proportion of duplicates. After checking the number duplicate records detected was 105, 134, 106, 141, and 130 for EndNote X9, Mendeley, Zotero, Rayyan, and SRA, respectively (Fig. 1 and Appendix 1).

The automated tools had specificity point estimates ranging from 0.96 to 1.00. Mendeley and SRA had the highest values (1.00 [95% CI = 1.00−1.00]), followed by EndNote X9 (0.99; 95% CI = 0.99−0.99), Zotero (0.99; 95% CI = 0.99−0.99), and Rayyan (0.96; 95% CI = 0.96−0.97). Rayyan, Mendeley, and SRA presented high sensitivity (0.98 [95% CI = 0.94−1.00], 0.93 [95% CI = 0.88−0.97], and 0.90 [95% CI = 0.84−0.95], respectively), whereas EndNote X9 (0.73; 95% CI = 0.65−0.80) and Zotero had only fair sensitivity (0.74; 95% CI = 0.66−0.81).

Considering a 0.02 prevalence of duplicates, negative predictive value estimation ranged from 0.99 to 1.00. Mendeley (0.93; 95% CI = 0.88−0.97) and SRA (0.99; 95% CI = 0.96−1.00) had good PPV, whereas EndNote X9 (0.61; 95% CI = 0.54−0.69) and Zotero (0.62; 95% CI = 0.54−0.69) had fair values. Only Rayyan (0.41; 95% CI = 0.36−0.47) had poor PPV. Balanced accuracy ranged from 0.86 (Zotero and EndNote X9) to 0.97 (Rayyan) (Table 1).

4. Discussion

Our findings show that Rayyan, Mendeley, and SRA are the most accurate tools for deduplication -a critical step in systematic review studies, mainly in the era of ever-growing health sciences citation databases. Our work is unique as we employed a large independent database coupled with the assessment of critical accuracy measures to single out some reliable deduplication tools to aid the initial steps of a systematic review study.

As distinct citation databases often index overlapping content, automated tools are needed to rule out duplicated records accurately and speed up the literature search while systematic reviews are conducted. Recent manuscripts described the use of automated tools for deduplication as responsible for transforming this step as one of the least time-consuming tasks required to complete a systematic review [15,20]. SRA, Rayyan and Mendeley are tools that organize the possible duplicates in terms of matching level, that is, all possible duplicate records are organized into three levels of agreements such as higher, medium, and low) and ask for manual checking as the last level. However, SRA and Rayyan have an advantage as they do not automatically exclude studies, as with Mendeley.

Similar to our results, Kwon et al. [12] described Mendeley as the more accurate automated tool compared to EndNote X9. Although the authors did not calculate the accuracy parameters, they showed that EndNote X9 produced 7.2 times more FNs and 0.5 times more FPs records than Mendeley.

Besides, some studies with small databases also showed the highest accuracy of Mendeley contrasted to EndNote

Table 1. Accuracy parameters with 95% confidence interval for automated tools applied to deduplication records relative to the manual search

Accuracy parameters	Rayyan	EndNote X9	Mendeley	Zotero	SRA
Sensitivity	0.98 (0.94, 1.00)	0.73 (0.65, 0.80)	0.93 (0.88, 0.97)	0.74 (0.66, 0.81)	0.90 (0.84, 0.95)
Specificity	0.96 (0.96, 0.97)	0.99 (0.99, 0.99)	1.00 (1.00, 1.00)	0.99 (0.99, 0.99)	1.00 (1.00, 1.00)
PPV[a]	0.41 (0.36, 0.47)	0.61 (0.54, 0.69)	0.93 (0.88, 0.97)	0.62 (0.54, 0.69)	0.99 (0.96, 1.00)
NPV[a]	1.00 (1.00, 1.00)	0.99 (0.99, 1.00)	1.00 (1.00, 1.00)	0.99 (0.99, 1.00)	1.00 (1.00, 1.00)
Balanced accuracy[b]	0.97	0.86	0.96	0.86	0.95

PPV, positive predictive value; NPV, negative predictive value; SRA, systematic review accelerator

[a] Considering a 0.02 prevalence of duplicates.

[b] The *caret* package does not provide a 95% confidence interval for balanced measures [19].

X9 to deal with FP records in the deduplication step [4,13]. Corroborating our findings, a previous study has shown Zotero and EndNote X9 as having poorer accuracy than the manual search among a set of automated tools [11].

The choice of a particular automated tool should be informed by distinct criteria, such as the number of records to be reviewed, the completeness of documentation of the citations in the database, and accuracy parameters, mainly sensitivity, specificity, and predictive values [4,11,13,21]. In this study, we also provided balanced accuracies as an intuitive summary measure to aid researchers in their decisions. As we could only provide point estimates for such a measure, one should be cautious when interpreting its meaning.

Accuracy measures encompass quite useful metrics that formally describe how trustful a given diagnostic instrument is in discriminating entities with a particular attribute and those without [18,22]. Here, sensitivity was used to inform the proportion of FN duplicates relative to manually searched duplicates. In a complementary way, specificity provides the proportion of FP duplicates.

The PPV measures the probability of positive results from a diagnostic instrument (in our case, citations classified as duplicates by a deduplication tool) translates into the presence of the attribute assessed (i.e., the occurrence of duplicates). NPV, conversely, provides the probability of negative results (citations classified as nonduplicates) and translates into the absence of the attribute (the occurrence of nonduplicates).

Although deemed the most informative measure, predictive values are not intrinsic to the diagnostic instrument because they are heavily dependent on the prevalence of the attribute being evaluated [18,22]. Therefore, as far as the proportion of duplicated citations is not previously known, those measures are often disregarded and misinterpreted. Moreover, we highlight that manual searching of duplicate records is not devoid of limitations. Manual deduplication can also yield FP results [11,12,23], which undermines the specificity and, therefore, the PPV [18], despite being time-consuming.

Being accurate enough, the marginally lower specificity of Rayyan may yield somewhat poorer PPV compared to other automated tools, mainly at the lower prevalence of duplicates. In contrast to the other deduplication tools, Rayyan was developed to expedite the initial screening of abstracts and titles in systematic review studies through a friendly interface as an alternative to Covidence from Cochrane Foundation [9]. Therefore, despite having been used for deduplication [10,11], this was not the primary goal of Rayyan, which could explain, at least in part, the lower specificity and PPV found. Furthermore, Rayyan did not allow reviewers to choose specific fields to compare potential duplicate records (for instance, date of issue, title, author's name, and journal) as Mendeley and SRA do [6,15], which reduces its deduplication accuracy [24].

As per the Cochrane Library, the accurate identification and exclusion of duplicate records is an initial and mandatory step of research targeting synthesizing the scientific literature systematically and reproducibly, which must be thoroughly conducted [25]. Hence, our results on accuracy parameters from the available main automated tools can help researchers choose the most suitable tool based on a set of empirically defined values to conduct deduplication steps in systematic review studies.

Identifying duplicate records aims to reduce reviewers' workload associated with screening. However, this process should be accurate enough to avoid excluding FP records, which can be a source of bias in systematic review studies because the maximum recall in retrieval is desirable [11,12,23]. The exclusion of FP records can negatively impact the synthesis, direction, and quality of the evidence produced by systematic reviews, which can affect health decisions and other activities related to evidence-based decisions [1,2].

This study's strengths include using a large set of records retrieved directly from free access research databases (Cochrane Library, MEDLINE, Web of Science, Embase, Scopus, Latin American and Caribbean Health Sciences Literature, and SciELO) instead of automated paid databases, such as OVID or ProQuest. The latter may not be accessible for low-income and middle-income settings. In addition, we used updated versions of the automated tools assessed [12,13].

This study has some limitations. First, we evaluated only results from a particular health research field, namely, Nutrition Epidemiology, but it is reasonable to speculate that the quality of information for other fields may have some difference. Second, we evaluate only a set of automated tools; thus, we have no information on the accuracy of other tools, such as the Reference Manager and the Bramer method [4]. These approaches have faced criticism, given that their mechanism of duplicate exclusion is not transparent [4,11]. Another limitation refers to the absence of verification of the search strategy by the PRESS Peer Review of Electronic Search Strategies [26]. Finally, as there are no a priori data on the average proportion of duplicated citations, the prevalence of duplicates in our database (0.02) was chosen as an educated guess. Hence, our values regarding predictive values are solely intended for comparative purposes because our primary purpose was not to establish individual accuracies but to provide a comparative assessment among different tools. Therefore, our comparative findings are expected to hold even if distinct scenarios of duplicate records proportion arise.

Finally, our study makes a meaningful contribution to the field of systematic review studies because using accurate automated tools can reduce subjectivity and help researchers save time searching for answers to specific guiding questions and ultimately underpin clinical and policy decisions.

5. Conclusion

A set of accuracy measures can guide researchers in conducting systematic reviews using the most reliable automated deduplication tool. By doing so, researchers can speed up the work process of systematic reviews and improve the overall quality of their analyses. Contrasted to the manual search, Rayyan, Mendeley, and SRA proved to be accurate enough to be applied in the deduplication step of systematic reviews.

Moreover, choosing the most suitable tool depends on its interface's characteristics, the algorithm to identify and exclude duplicates, and the transparency of the process. Therefore, SRA seems to be a tool that best gathers the points mentioned previously. In addition to a user-friendly interface, SRA brought the possibility of calibrating the algorithm of deduplication as per the size of the base (small and large) and organized duplicates in three levels of confidence (high, medium, and low). It allows the researcher to manually confirm, quickly and transparently, whether the reference, particularly those belonging to the low confidence category, is a TP or TN record. Although it is also feasible for Mendeley to present the degree of agreement between possible duplicate references, this software uses an automatic duplication mechanism, which is not transparent and deserves the attention of researchers, particularly those who are a beginner in this tool.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclinepi.2022.10.009.

References

[1] Horne JR. Are we losing sight of the meaning of "evidence-based nutrition?". Int J Public Health 2020;65(5):513–4.

[2] Reddy KR, Freeman AM, Esselstyn CB. An urgent need to incorporate evidence-based nutrition and lifestyle medicine into medical training. Am J Lifestyle Med 2019;13(1):40–1.

[3] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372(71):1–9.

[4] Bramer WM. Reference checking for systematic reviews using Endnote. J Med Libr Assoc 2018;106:542–6.

[5] Sievert MC, Andrews MJ. Indexing consistency in information science abstracts. J Am Soc Inf Sci 1991;42(1):1–6.

[6] Mendeley. Getting started with Mendeley Desktop. Elsevier; 2021. Available at https://www.mendeley.com/reference-management/mendeley-cite.

[7] Zotero Zotero. Corporation for digital scholarship. 2020. Available at https://www.zotero.org/. Accessed August 25, 2022.

[8] Clarivate. Endnote. 2022. Available at https://endnote.com/. Accessed August 25, 2022.

[9] Community C. Covidence. 2021. Available at https://community.cochrane.org/help/tools-and-software/covidence. Accessed August 25, 2022.

[10] Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan - a web and mobile app for systematic reviews. Syst Rev 2016;5(210):1–10.

[11] McKeown S, Mir ZM. Considerations for conducting systematic reviews: evaluating the performance of different methods for deduplicating references. Syst Rev 2021;10(38):1–8.

[12] Kwon Y, Lemieux M, McTavish J, Wathen N. Identifying and removing duplicate records from systematic review searches. J Med Libr Assoc 2015;103(4):184–8.

[13] Rathbone J, Carter M, Hoffmann T, Glasziou P. Better duplicate detection for systematic reviewers: evaluation of Systematic Review Assistant-Deduplication Module. Syst Rev 2015;4(1):1–6.

[14] Lisboa CS, Guimarães NS, Falcão RI, Alves FJO. Impact of cash transfer programs on child health outcomes: a systematic review. PROSPERO; 2021. Available at https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42021255570. Accessed August 25, 2022.

[15] 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.

[16] AUo Beirute. Remove duplicates. 2021. Available at https://aub.edu.lb.libguides.com/Mendeley/RemoveDuplicates. Accessed August 25, 2022.

[17] University M. Systematic reviews, scoping reviews, and other knowledge syntheses. 2021. Available at https://libraryguides.mcgill.ca/knowledge-syntheses/deduplicating. Accessed August 25, 2022.

[18] Zhou XH, McClish DK, Obuchowski NA. Statistical methods in diagnostic medicine. John Wiley & Sons; 2009.

[19] Kuhn M. Building predictive models in R using the caret package. J Stat Softw 2008;28(5):1–26.

[20] Nussbaumer-Streit B, Ellen M, Klerings I, Sfetcu R, Riva N, Mahmić-Kaknjo M, et al. Resource use during systematic review production varies widely: a scoping review. J Clin Epidemiol 2021;139:287–96.

[21] García-Pérez MA. Accuracy and completeness of publication and citation records in the Web of Science, PsycINFO, and Google Scholar: a case study for the computation of h indices in Psychology. J Am Soc Inf Sci Technology 2010;61(10):2070–85.

[22] Alberg AJ, Park JW, Hager BW, Brock MV, Diener-West M. The use of "overall accuracy" to evaluate the validity of screening or diagnostic tests. J Gen Intern Med 2004;19:460–5.

[23] Gaudino M, Robinson NB, Audisio K, Rahouma M, Benedetto U, Kurlansky P, et al. Trends and characteristics of retracted articles in the biomedical literature, 1971 to 2020. JAMA Intern Med 2021;181(8):1118–21.

[24] Kellermeyer L, Harnke B, Knight S. Covidence and rayyan. J Med Libr Assoc 2018 Oct;106:580–3.

[25] Higgins. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editors. Cochrane Handbook for Systematic Reviews of Interventions version 6.3 (updated February 2022). Cochrane; 2022.

[26] 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.