SCREENING

Removing duplicates, with a record of each one

For anyone combining database exports before screening, in a reference manager or a screening tool.

By the end you will have one record for each report, a log of every removal, and the number for the PRISMA box.

Checked on 29 Sep 2026 · Version 1.0 · CC BY 4.0

Several databases find many of the same papers, so a combined search holds copies. By the end of this guide you will have one record for each report your searches found, a log of every record you removed and why, and the number for the PRISMA 2020 box "Duplicate records removed".

Removing copies saves screening time. Removing a paper that only looked like a copy can cost you a study. The steps below keep both in view, in whichever tool you use, and leave a record that a supervisor or an examiner can follow.

Before you start

- Every database export, each with the number of records the database showed when you exported it.
- All of them in one library or one tool, so that every copy can meet every other.
- A copy of the untouched set, exported to a file and dated, before anything is removed.
- The log near the end of this guide, or a spreadsheet with the same columns.

Tell a duplicate from a second report

PRISMA 2020 draws the line in its glossary (Box 1). A record is "the title or abstract (or both) of a report indexed in a database or website". The glossary goes on: "Records that refer to the same report (such as the same journal article) are 'duplicates'; however,

records that refer to reports that are merely similar (such as a similar abstract submitted to two different conferences) should be considered unique."

A report can be "a journal article, preprint, conference abstract, study register entry, clinical study report, dissertation, unpublished manuscript, government report, or any other document providing relevant information". So a conference abstract and the full paper of one study are two reports, not a duplicate pair. Keep both, and join them to their study later. Cochrane's standard MECIR C42 asks reviewers to "Collate multiple reports of the same study, so that each study, rather than each report, is the unit of interest in the review." That joining is a later step, not de-duplication.

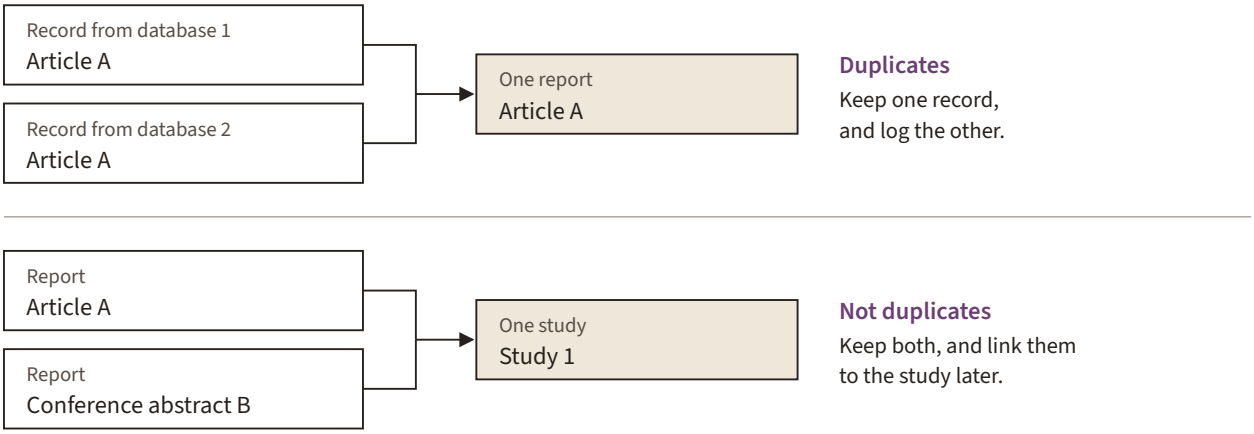

Figure 1. Two records of one report are duplicates; two reports of one study are not. Illustrative.

Table 1. Same report, or a second one? (after the PRISMA 2020 glossary, Box 1)

What you find	Same report?	What to do
One article in two databases, its fields written differently (page ranges, accents, a shortened title)	Yes	Keep one record
One article with an English title in one database and its original title in another	Yes	Keep one record
The same DOI or PubMed ID, but different titles or first pages	Look closer	Compare both before merging
A conference abstract and the full paper	No	Keep both; link them to the study later
A preprint and the published article	No	Keep both; link them later
A protocol and its results paper	No	Keep both; link them later
An article and its correction notice	No	Keep both; read them together
Similar abstracts at two conferences	No	Keep both
One study's results published twice, or in translation	No	Keep both; link them later

Balance catching copies against keeping papers apart

Studies that compare ways of removing duplicates measure two things, and each study names them its own way.

- **Copies caught.** McKeown and Mir (2021) report sensitivity, "the proportion of correctly identified duplicate references". Bateup and colleagues (2026) count the other side, "missed duplicates": records a tool "classified as a unique record when it was a duplicate record".
- **Papers kept apart.** McKeown and Mir report specificity, "the proportion of correctly identified non-duplicate references". Its failures are false positives, "references incorrectly identified as duplicate references and flagged for removal". Bateup and colleagues call these "singulars removed".

The two pull against each other. Looser matching catches more copies and also removes more papers that were not copies. Stricter matching keeps more papers apart and leaves more copies behind. The costs are not equal. A missed copy costs some screening time, and you can still catch it later. A paper removed as a copy never reaches screening. That is why McKeown and Mir looked closely at specificity: "removing false positives from the screening process may result in missing eligible studies and introduce bias to syntheses."

Both papers give scores for named tools. Their tests ran from December 2018 to January 2020 and from December 2023 to February 2025, and tools change, so this guide does not repeat the scores. Read them in the papers, and keep each tool's version and test date beside its score. Bateup and colleagues also timed each tool, including "any additional manual checking that is required", and judged all eight tools they tested good enough to recommend, so that people can choose by the time and funds they have.

Remove the clearest copies first

Work from strict comparisons to looser ones. Bramer and colleagues' EndNote method (2016) opens with rounds strict enough to need no checking by hand, and Rayyan's Help Center suggests starting its Auto-Resolver with exact field matching or 100% similarity, then lowering the threshold in a second pass.

1. Let the tool remove only copies that agree on every field it compares, such as the same DOI and title, or the same authors, year, title and pages.
2. Loosen the comparison one step at a time.

3. Look at each pair a looser round suggests before you merge it.
4. Write each round in the log: what was compared, and the records before, removed and after.

Look at each pair before you merge

For every pair a looser round suggests, look at seven things. The duplicate check card has them on one page.

1. **DOI and PubMed ID:** the same, different, or missing.
2. **Title,** with its subtitle. A missing subtitle or a translated title can still be the same article.
3. **Authors,** the first author above all.
4. **Year.**
5. **Journal, volume and first page.** In 2016 Bramer and colleagues noted that MEDLINE and the Cochrane Library shortened page ranges (“1008–12”) where most databases wrote them in full (“1008–1012”).
6. **Kind of publication:** article, conference abstract, preprint, protocol, correction or comment.
7. **Abstract:** the same text, or the same study told another way?

Merge when the two are the same report. Keep both when they are two reports of one study. When you cannot tell, keep both for now, note the pair, and look again when you have the full texts.

Run it in your tool

EndNote

In EndNote 2025, choose Library › Find Duplicates. By default two references are duplicates when they have the same reference type and identical Author, Year and Title fields, so copies whose fields are written differently may not be matched. To change what is compared, open the Duplicates preferences (Windows: Edit › Preferences › Duplicates; macOS: EndNote 2025 › Settings › Duplicates). There you choose the fields, and either an exact match or a comparison that ignores spacing and punctuation.

For each pair, “Keep This Record” keeps that copy and moves the other to the Trash, and “Skip” leaves both. “Cancel” closes the side-by-side view and leaves the duplicates in a temporary “Duplicate References” group.

The method of Bramer and colleagues (2016) builds on this:

1. Show page numbers in the library, and expand shortened page ranges. The paper supplies an export style and an import filter for this.

2. Run seven rounds, each with the fields set out in the paper's Table 1: set the fields, run Find Duplicates, click "Cancel", then remove the duplicates as that row of the table says.
3. The first two rounds are strict enough that the authors remove their duplicates without checking. The next three need a check of the references without page numbers, and the last two need more checking by hand.

The paper dates from 2016, so some menu names may differ in your version.

Zotero

Zotero's page on duplicate detection (last updated 25 Nov 2017) describes the "Duplicate Items" collection in your library; you can also right-click the library and choose "Show Duplicates". Zotero compares titles, DOIs and ISBNs, then years within one of each other and at least one author's last name and first initial. Select a set, choose the master item and the version of any field that differs, and click the merge button, which gives the number of items ("Merge 2 Items"). A merge keeps the collections and tags of every copy. Detection works within one library, and two or more items of the same item type can also be merged by hand with "Merge Items…". The page notes that "it is not possible to mark false positive matches as non-duplicates".

Mendeley Reference Manager

Elsevier's support page (updated 5 May 2026) says that from version 2.93 a "Duplicates" smart collection groups duplicates into sets, and that duplicates are also checked when you import. You resolve a set by deleting the copies you do not want, so note each one in your log first.

Rayyan

Rayyan's Help Center (updated 20 Sep 2026) asks you to finish every import and to de-duplicate before screening starts. "Detect Duplicates", on the Overview or Review data tab, runs once per unique dataset, and again only after references are added or deleted. Each possible pair appears side by side with a confidence percentage, and you choose "Keep Left Article", "Keep Right Article" or "Keep Both Articles". The copy not kept moves to "Deleted", where it can still be seen. The count under "Possible Duplicates" › "Deleted" is the number the article gives for PRISMA.

On paid plans, the Auto-Resolver resolves every pair that meets your rules in one action. Rayyan's article says "Auto-resolution is irreversible", so export first. A deleted copy can be brought back by exporting it from "Deleted" and uploading it again.

Keep the records you removed

The removed copies are your evidence. If someone asks whether a study was lost, that file answers.

- Before anything is emptied or deleted, export the removed records to a file, with the date in its name.
- Keep one line per round in the log below.
- PRISMA-S item 16 reads: “Describe the processes and any software used to deduplicate records from multiple database searches and other information sources.” The log is that description.

Evensift guides · Worksheet

The de-duplication log

Checked on 29 Sep 2026 · Version 1.0 · CC BY 4.0
evensift.com/guides/remove-duplicates-systematic-review/

Review	Tool and version	Date started

Round	What was compared (fields, settings or method)	Records before	Removed	Records after	Every pair checked by hand?	Initials
1						
2						
3						
4						
5						
6						
7						
Found during screening						
Total						

File of removed records	Rows in it	Records identified (all databases)	Minus duplicates removed	Records left for screening

Record duplicates found during screening

Some copies surface later, when a title looks familiar. The PRISMA 2020 statement and its explanation define duplicates, but do not say where to count one found after screening has begun (checked 29 Sep 2026). Two ways keep the sums true:

- Move it to “Duplicate records removed”, and take it out of “Records screened” (and out of “Records excluded”, if you had excluded it).
- Leave it as screened, count it among “Records excluded”, and write “duplicate of” and the other record’s number in your log.

At full text, the second way becomes a reason under “Reports excluded”. Choose one way, keep to it, and say which in a note under the diagram.

Check your work

- The records from each database add up to the total you started with.
- In the log, each round’s records before, minus removed, equals records after, and each round’s after is the next round’s before.
- The total removed matches your tool’s own count, and the file of removed records has that many rows.
- Records identified, minus duplicates removed, minus anything else removed before screening, equals records screened.
- A few pairs from your loosest round, looked at again, still look like the same report.

Writing it up

Adapt this sentence for your methods:

> Records from [n] databases were combined in [software, version] and duplicates were removed on [date] by [method: for example, the staged method of Bramer et al. (2016), or automatic detection with every suggested pair checked by hand]. Different reports of the same study, such as a conference abstract and the full paper, were kept and linked to their study at [stage]. [n] duplicates were removed from [n] records, leaving [n] for screening. The removed records are listed in Supplement [n].

In the flow diagram, the number goes in “Duplicate records removed (n =)”, under “Records removed before screening”.

Reading on

- The PRISMA 2020 flow diagram: counting every box
- Exporting search results for screening, database by database
- How to screen titles and abstracts
- McKeown and Mir (2021), the evaluation quoted above (open access) (doi.org/10.1186/s13643-021-01583-y)

Sources

1. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71. Box 1 (glossary) and the flow diagram's box labels. https://pmc.ncbi.nlm.nih.gov/articles/PMC8005924/ (checked 29 Sep 2026). CC BY 4.0.
2. 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. https://pmc.ncbi.nlm.nih.gov/articles/PMC8005925/ (checked 29 Sep 2026).
3. Rethlefsen ML, Kirtley S, Waffenschmidt S, et al. PRISMA-S: an extension to the PRISMA statement for reporting literature searches in systematic reviews. *Syst Rev* 2021;10:39. Item 16. https://pmc.ncbi.nlm.nih.gov/articles/PMC7839230/ (checked 29 Sep 2026).
4. Cochrane. MECIR standard C42, "Collating multiple reports". https://www.cochrane.org/authors/handbooks-and-manuals/mecir-manual/standards-conduct-new-cochrane-intervention-reviews-c1-c75/performing-review-c24-c75/selecting-studies-include-review-c39-c42 (checked 29 Sep 2026).
5. McKeown S, Mir ZM. Considerations for conducting systematic reviews: evaluating the performance of different methods for de-duplicating references. *Syst Rev* 2021;10:38. https://pmc.ncbi.nlm.nih.gov/articles/PMC7827976/ (checked 29 Sep 2026).
6. Bateup S, Fulbright H, Moberg K, et al. Evaluating the accuracy and speed of eight deduplication tools: a comparative study. *Res Synth Methods* 2026, published online 17 Jun 2026. https://doi.org/10.1017/rsm.2026.10100 (checked 29 Sep 2026).
7. Bramer WM, Giustini D, de Jonge GB, Holland L, Bekhuis T. De-duplication of database search results for systematic reviews in EndNote. *J Med Libr Assoc* 2016;104(3):240–243. https://pmc.ncbi.nlm.nih.gov/articles/PMC4915647/ (checked 29 Sep 2026).
8. Clarivate. EndNote 2025 help: "Finding Duplicate References" and "Duplicates Preferences", for Windows and macOS. https://docs.endnote.com/docs/endnote/2025/v1/windows/en/content/04search/finding_duplicaterefs.htm (checked 29 Sep 2026).
9. Zotero. Duplicate Detection (last updated 25 Nov 2017). https://www.zotero.org/support/duplicate_detection (checked 29 Sep 2026).
10. Elsevier Support. How do I check for duplicates in Mendeley Reference Manager? (updated 5 May 2026). https://www.elsevier.support/mendeley/answer/how-do-i-check-for-duplicates-in-mendeley-reference-manager (checked 29 Sep 2026).
11. Rayyan Help Center. How to Detect and Manage Duplicates in Rayyan (updated 20 Sep 2026), https://help.rayyan.ai/hc/en-us/articles/46779581890961 ; How to Use the Auto-Resolver in Rayyan (updated 16 Jul 2026), https://help.rayyan.ai/hc/en-us/articles/34563050458257 ; How to Recover a Deleted Duplicate in Rayyan (updated 13 Aug 2026), https://help.rayyan.ai/hc/en-us/articles/43744776756369 (all checked 29 Sep 2026).

EndNote, Zotero, Mendeley, Rayyan, MEDLINE and Cochrane Library are trade marks of their respective owners. This guide is independent: Evensift is not affiliated with, sponsored or endorsed by any of them.

Cite this guide. Evensift. Removing duplicates, with a record of each one. Evensift guides, version 1.0, 29 Sep 2026. https://evensift.com/guides/remove-duplicates-systematic-review/

Licence. © 2026 Evensift. This guide is licensed under CC BY 4.0 (creativecommons.org/licenses/by/4.0/). You may copy it, adapt it and host it, including in a library guide, if you credit it, link to the licence and to this page, and say what you changed. The Evensift name and mark are not covered by the licence. Quoted material keeps its own terms: PRISMA 2020 items and box labels are CC BY 4.0 (Page et al. 2021); quotations from the Cochrane Handbook and the JBI Manual are short and attributed, and are not relicensed here.

The closing line about Evensift and the independence paragraph are not covered by this licence; adaptations should leave them out. The method content is CC BY 4.0.

Version 1.0 · 29 Sep 2026 · What changed: First publication. The latest version is always at https://evensift.com/guides/remove-duplicates-systematic-review/. If something here no longer matches what you see, write to hello@evensift.com. Corrections are made within 7 days, with a dated note at the foot of this page.

Evensift merges near-exact copies by itself, asks about look-alikes one group at a time with the reason they look alike, and can take any merge apart again. Its matching rules have not been independently evaluated, so check the groups as you would in any tool. It is free and needs no account.