

Step 7: Deduplicate

	STEP
1	Formulate the review question
2	Find previous and upcoming SRs
3	Write the protocol
4	Design the systematic search
5	Design data extraction form and pilot
6	Run the systematic search
7	Deduplicate
8	Screen titles and abstracts
9	Obtain full text
10	Screen full text
11	Search and screen trial registries
12	Citation analysis
13	Screen citation analysis
14	Extract data
15	Assess risk of bias
16	Update the systematic search
17	Synthesize data
18	Write up review

Aim: Remove identical studies.

Requirements: Database access; Reference management software (EndNote).

Timeline: The deduplication process can take several hours or days (allow half a day to 3 days for deduplication).

Tools:

- Deduplication function of EndNote
- SRA – Deduplicator: sr-accelerator.com/#/help/dedupe

Step 7: Deduplicate

How to find and remove duplicate references in your EndNote library

Before you start deduplicating, first, pool all references from multiple databases/searches into a single EndNote library. A duplicate reference is an article that has been found more than once by the search strategy. Duplicates occur because there is overlap in journal coverage between databases. For example, Medline and Embase contain some of the same journals, and therefore some of the same articles.

It is useful to remove the duplicates before screening, to reduce unnecessary work in the screening process. This is usually done in EndNote, and is a manually intensive process.

Deduplication is done by performing several 'sweeps' of the duplicate references. These sweeps match references on different EndNote reference fields (such as title, author, year etc.). The duplicates can then be identified and removed from the EndNote library.

The first step in the deduplication process is to set up the display fields in your EndNote library. To set the Display Fields select Edit > Preferences then select "Display Fields" on the pop-up menu box. Select the fields Author, Year, Journal, Volume, Pages and DOI. These are the fields most useful for identifying duplicates.

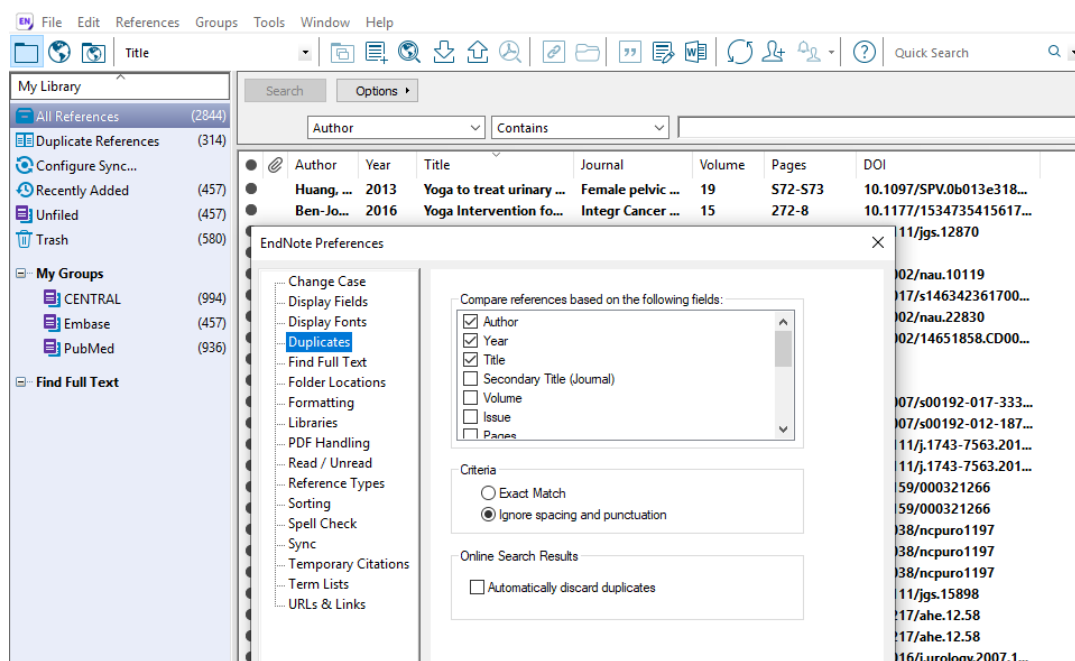
Note: Record the number of deleted duplicates identified during the screening process, these will need to be added to the PRISMA flowchart.

Perform the deduplication by performing sweeps of the library as described below.

First sweep – finding duplicates by matching Author, Year and Title

This sweep will identify most of the duplicates and is relatively fast.

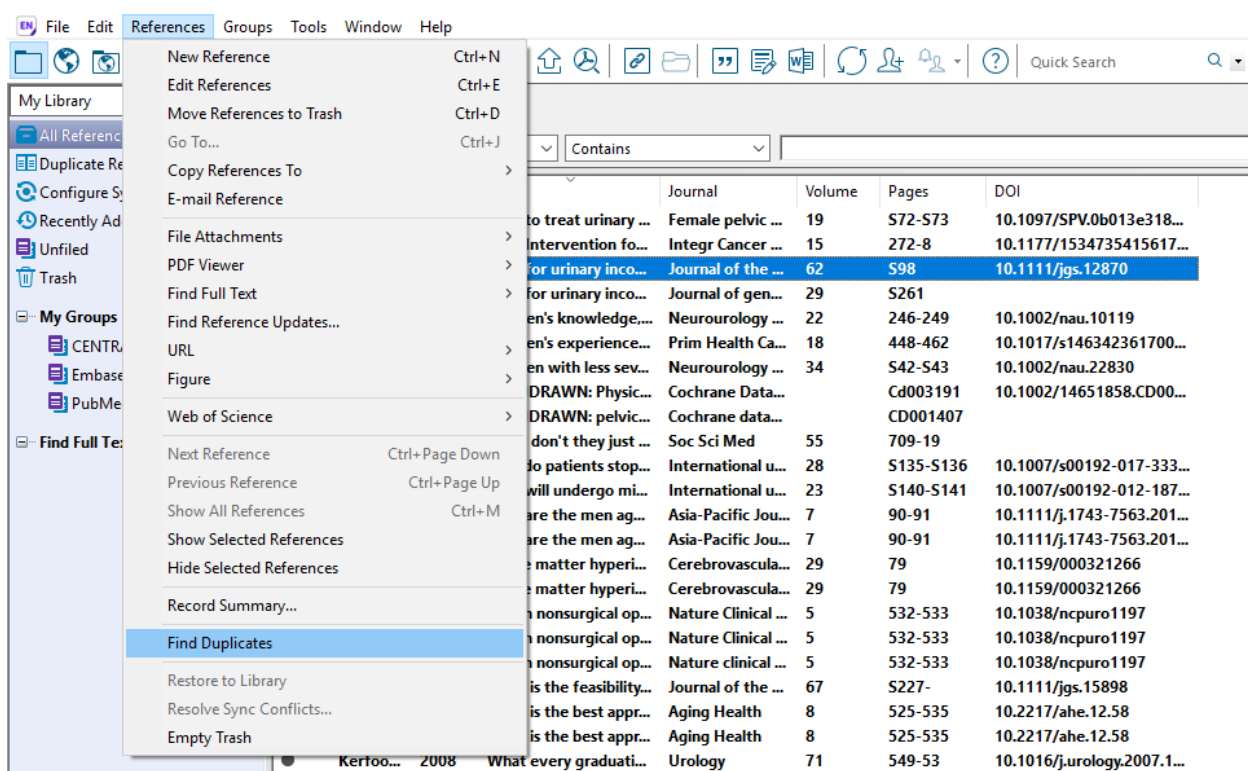
In the EndNote library select Edit > Preferences in the pop-up box, select Duplicates



Select (tick) the following 3 fields: Author, Year, Title, click OK

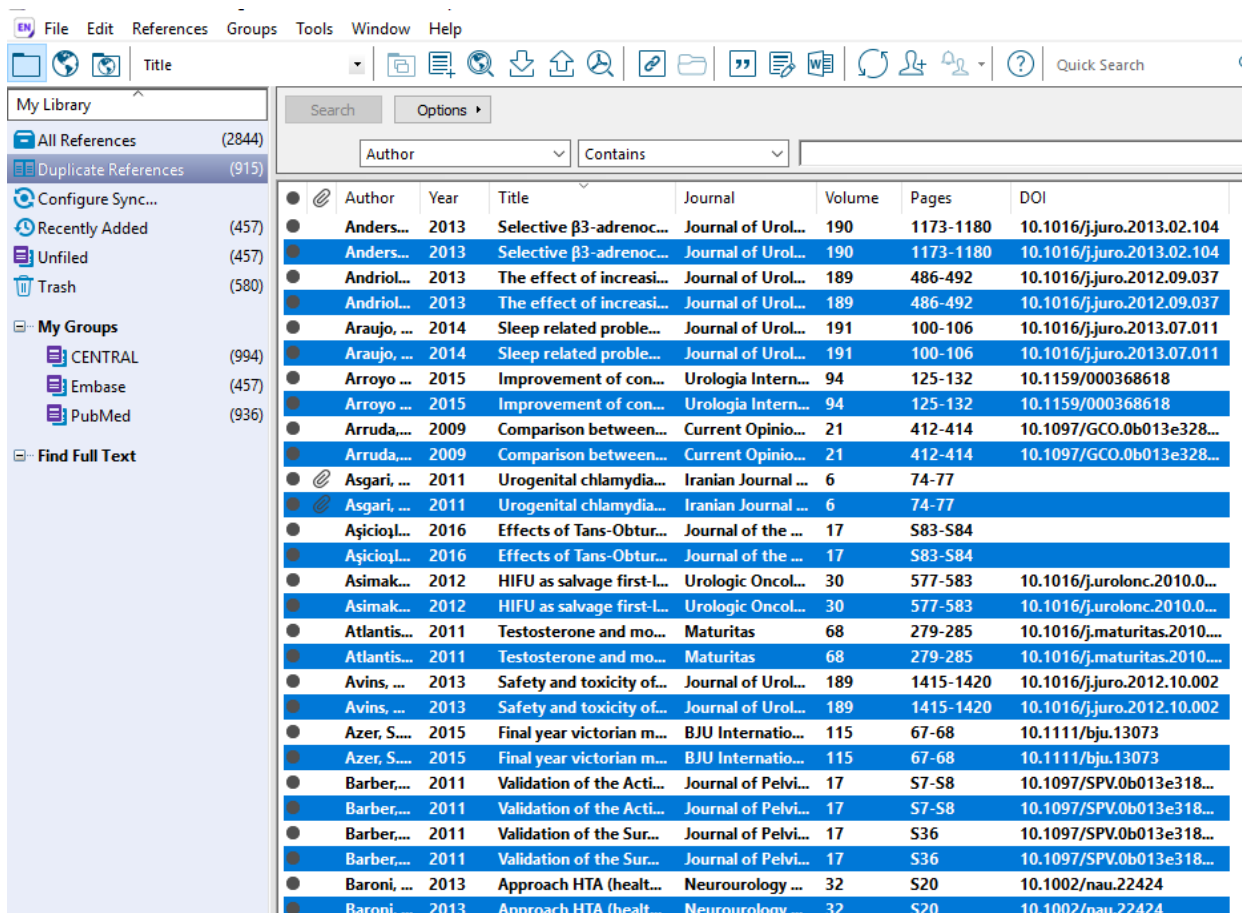
Select “All references” in the left-hand menu

Click on a reference, so one of the references is highlighted blue, then select References > Find Duplicates



The Find Duplicates window now appears, you can manually check the duplicate references to ensure they are duplicates, but if there are a large number, we recommend you continue with the following method.

Close the Find Duplicates window by clicking the Cancel tab.



The duplicates should now be displayed as blue rows (picture above). Note: if the duplicates aren't highlighted, click on one of the headings e.g. the Author heading.

Scan each article and its duplicate, paying closest attention to the fields where it is easiest to note discrepancies or non-duplicates (we recommend the Pages and DOI fields).

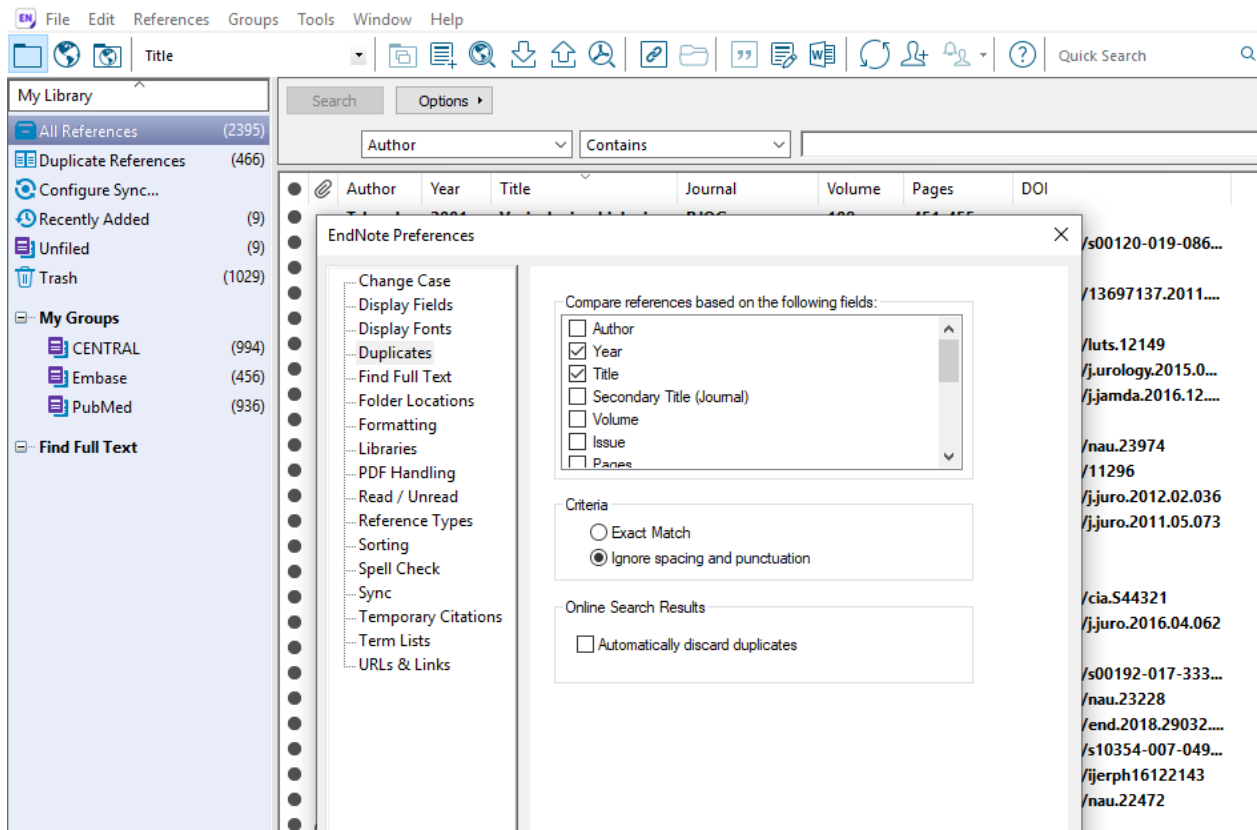
If a non-duplicate is detected, deselect it by holding down the Control (CTRL) key and selecting it. Note: this is important, if you select a reference without holding down CTRL you will deselect all references and will need to start again.

After scanning through the references, delete the duplicates by pressing the delete key.

Continue with the next sweep

Second sweep – Finding duplicates by matching references by Year and Title

Select Edit > Preferences and in the pop-up box, select Duplicates and select Year and Title.



Select “All references” in the left-hand menu

Click on a reference, so one of the references is highlighted blue, then select References > Find Duplicates

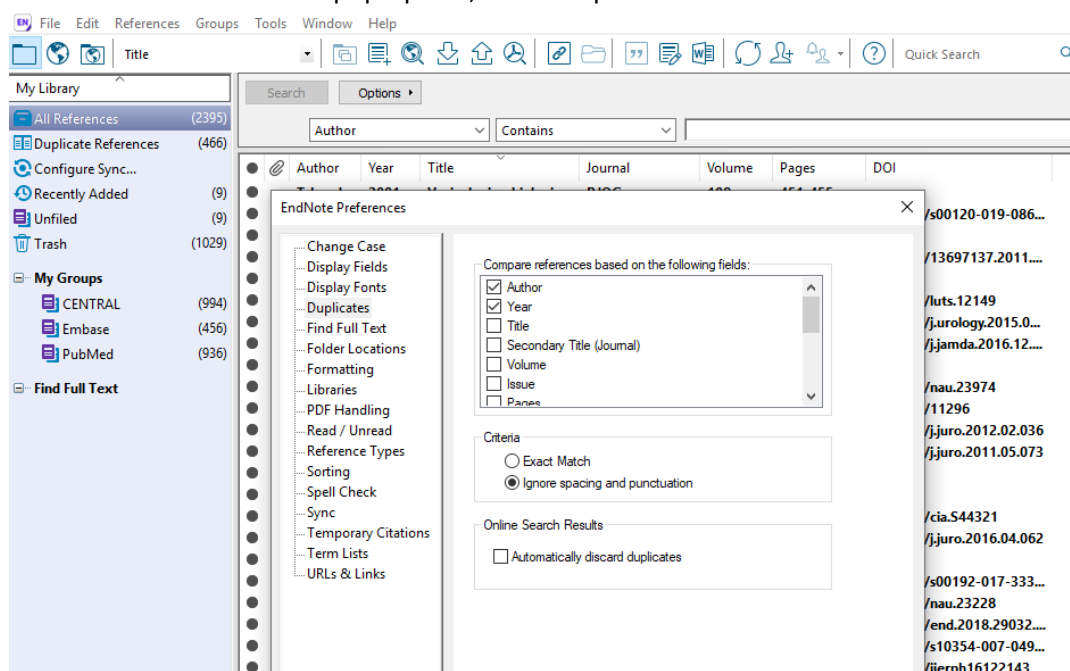
Select References > Find Duplicates

Close the Find Duplicates pop-up window by clicking the Cancel tab.

The rest of the process is identical to Sweep 1, except that the potential of there being a non-duplicate in the results is higher, therefore we recommend scanning through the results more slowly, and being more careful before deleting the duplicates.

Third sweep – Finding duplicates by matching references by Author and Year

Select Edit > Preferences and in the pop-up box, select Duplicates and select Author and Year.



Select “All references” in the left-hand menu

Click on a reference, so one of the references is highlighted blue, then select References > Find Duplicates

Close the Find Duplicates pop-up window by clicking the Cancel tab.

The rest of the process is identical to that in Sweep 1 and 2, except that there will almost certainly be some non-duplicates in the results. We recommend scanning through the results very carefully and deselecting (by holding down the CTRL key) all non-duplicates before deleting the duplicates.

Fourth sweep – no fields used

For the final sweep, sort the EndNote library by the title field (click Title at the top of the column), manually look through the EndNote library to identify any remaining duplicates.

Either delete these duplicates one at a time when you identify them, or select multiple by holding down the CTRL key. NOTE: Do not be concerned if you have missed some duplicates as any remainder will be identified during the screening process.

SRA Deduplicator

SRA deduplicator is another tool that could assist you in deduplicating your library. It requires you to upload your library to the SRA deduplicator. Step by step instructions are available at sr-accelerator.com/#/help/dedupe