Systematically deduplicating in EndNote

Introduction

EndNote is a reference management software designed for collecting, organising, and storing a library of references. It has the ability to insert citations and a reference list into your text document.

While not custom designed for the systematic-style review process, EndNote can be used for deduplication, screening, and reporting in accordance with the PRISMA flow diagram. Doing so involves a fairly advanced knowledge of the 'find duplicates', 'groups', and 'create from groups' features of EndNote. If you are new to EndNote, I recommend working through our EndNote 20 Basics Tutorial.

Setting up EndNote for the systematic-style review process

To systematically deduplicate records in EndNote in accordance with the PRISMA flow diagram, you will need to create a new EndNote library specifically for your systematic-style review. This is the only time we recommend having more than one Endnote library.

Firstly, you will need to create four 'group sets', and populate these group sets with a number of 'custom groups'.

In the left-hand panel, you can see that I've created four group sets:

- Search results
- Deduplication
- TI-AB: title-abstract screening
- Full-text screening.

We will not be looking at the two screening groups in this video.

Within my 'search results' group set, I have created three custom groups, which correspond with the three databases I searched. For this video, I'm keeping the numbers of results low; it's likely you will be searching more databases and working with larger sets of records. Having a

separate custom group for each database allows us to keep a record of how many results our search found in each of the databases.

You can see that I've included a number at the beginning of the custom groups' names. This is to force EndNote to display the custom groups in the order I want. By default, EndNote orders custom groups in alphabetical order. If you want to edit a group name, right-click on the group, and select 'rename group'.

Importing RIS files into EndNote

To populate each custom group, you need to import the RIS files you exported from each database. I have these saved in a folder and have renamed each RIS file the name of the database and the date of exploitation. We highly recommend that you keep a copy of these RIS files in case you need to repopulate your EndNote library.

When we open one of these files the records are available in the 'imported references' and 'unfiled' groups in EndNote; these are groups that EndNote automatically creates and populates. Control-A on the keyboard will select all these records and you can drag-and-drop them into the correct database.

EndNote display fields

Let's take a look at the information you can see here in the display window: the 'file' icon, 'author', 'year', 'title', etc. You can adjust how much space you give each of these 'display fields', and which fields you see. Simply right-click on one of the display fields and the display field menu will appear. EndNote will allow 10 display fields at one time. For the purpose of deduplication and screening the most useful display fields are:

- Author
- Year
- Title; article title
- Journal; which EndNote also calls the secondary title
- Volume
- Number

- Pages
- Abstract
- Reference type.

I also have 'file attachment' displayed, which you don't need for deduplication, but will be useful for the full-text screening later.

You can change the order of these fields appear in by clicking on the field name and dragging it to wherever you want.

Alternate RIS import method

Another way to import RIS files is by selecting 'file', then 'import', and select 'file'. This will open a pop-up box. Select the RIS file you'd like to import; in the 'import options' select 'reference manager (RIS)' as we are working with RIS files. Click 'import' and the imported references group will be updated. We'll move these records into the correct database custom group. Select all these records, again we could use the Control-A function on the keyboard, or we can select the first record, scroll down to the last record, and use the Shift button on the keyboard, and select that final record. The Shift button allows us to select all the records between each of the records we have selected.

Once all the database RIS files are imported, we have 32 records in the 1-Eric-ProQuest custom group, 50 in the 2-Scopus custom group, and 15 in the 3-PsycINFO custom group: 97 records in total.

The PRISMA flow diagram

At this point, you can populate your PRISMA flow diagram. Note: here we've only searched databases, no citation or grey literature searching. Therefore, we will use the 'PRISMA 2020 flow diagram for new systematic reviews that included searches of databases and registers only'. If we had found citations by other methods, citations searching or suggestions from experts, we would use the other PRISMA flow diagram, 'Prisma 2020 flow diagram for new systematic reviews, which included searches of databases registers and other sources'.

The first box of the PRISMA flow diagram is 'records identified from databases and registers'. Since we only searched databases, you can populate this field with 97. Our suggestion would

be to record the number of results from each database: 32 from Eric via ProQuest, 50 from Scopus, and 15 from PsycINFO; this extra level of detail can be helpful.

Saving a Compressed library as a backup

Now your 'search results' group set is complete, it is a good time to back-up this library by saving a copy as a compressed library. To save the compressed library, navigate to the menu bar, select 'file', and select 'compressed Library .enlx'. A pop-up will appear. For this compressed library, I need 'without file attachments', then click 'next'. I will save this compressed library in my documents and give it a meaningful name: ENSsR2023- today's date - SearchResults; 'search results' to let me know which part of the process this library has completed.

Setting up the deduplication groups

Now that our search results are completed, let's move on to the deduplication process.

Within our 'Deduplication' group set, we have three custom groups:

- 'All search results' which will be a static group,
- 'Duplicates' which will be progressively populated',
- 'Undecided' which we can use for records that we're unsure about.

To populate the 'All search results' group, go to the 'All references' group at the top of the panel and select all the records within this library. Copy to the 'All search results' custom group here. We have 97 records within this custom group, which will not change for the rest of the process.

Before you start looking for duplicates, create one more group within the 'Deduplication' group set, which is called a 'create from groups'. To do this, right-click on the 'Deduplication' group set and select 'create from groups'.

This pop-up allows you to set up a 'create from groups' group. I'm going to call this 'UNIQUE RECORDS'. Tell EndNote to populate this group with all the records from the 'All search results' custom group, but not any of the records that are in the 'Duplicates' custom group, and not any of the records in the 'Undecided' custom group.

Click 'create'. You will see a new group in the left-hand panel: 'UNIQUE RECORDS' populated with the total 97 records from our 'All search results' group as we don't yet have any records in either the duplicates or undecided groups.

EndNote Preferences

Before you start the deduplication process, you need to look at how EndNote identifies a duplicate record. To do this, click 'edit' and select 'preferences'. A pop-up menu will open a list of the display options or functions that can be adjusted or customised. Select 'duplicates'. Here you can change the source elements; so you can compare and determine whether a record is a duplicate or original record.

Round One

This method of the deduplication involves eight rounds and three optional additional rounds. For the first round, you want to compare:

- Author
- Year
- Title
- Secondary title or journal name
- Volume
- Pages.

Click 'apply' and then 'okay'. Making sure you're working within your 'UNIQUE RECORDS' group, select 'library' and 'find duplicates'.

EndNote will bring up a pop-up window, comparing two records asking you to choose which record to keep. In a systematic-style review, you are usually working with a large number of records and don't have time to look at each duplicate individually. You can click 'cancel' in the upper right-hand corner of the pop-up, causing duplicates to be highlighted in green. In the left-hand panel, you can now see you're working in the temporary group called 'duplicate references'. EndNote automatically creates and populates this group with what it considers the original record because it was added to the library first, and below that original record, all the records identified as duplicates based on the criteria selected in the 'preferences' menu.

As all these duplicate records are highlighted, you can click on one and drag-and-drop them all into the 'Duplicates' group. You can see the duplicates group now has 22 records, while the 'UNIQUE RECORDS' group has 75; the 'All search results' group still has 97 records as no changes have been made to this group.

At the end of each round, you want to save a copy of the EndNote library as a compressed library to ensure that you always have a back-up and if EndNote crashes, you don't have to repeat that step. Just like before, select 'file', and 'compressed library .enlx'. Name this compressed library the same as before, but with 'Round1' at the end so you know it was created after the first round of deduplication.

Round Two

Go back to your 'UNIQUE RECORDS' group and adjust the preferences for Round 2. Again, select ‘edit’, 'preferences' and then 'duplicates'. For Round 2, compare:

- Author
- Year
- Title
- Volume
- Pages.

These are the same fields as Round 1 except you don't include the secondary title (the journal title). For each round you only change one or two of the criteria to ensure you're being systematic and thorough. Think of each round as a filter or sieve - in the early rounds, you're filtering out the big, obvious duplicates and by the end, you'll only compare a few fields, finding those less obvious, more subtle duplicates.

If you lose the highlighting of your duplicates, you can either run the 'find duplicates' function again or use the Control key on the keyboard and manually select the duplicate records, moving then into the 'Duplicates' group, which now leaves 70 records in the 'UNIQUE RECORDS' group.

Again, save the compressed library as a back-up, this time name it 'Round2'.

Round Three

For Round 3, compare:

- Author
- Year
- Secondary title or the title of the journal
- Pages.

For the number of records we're working with in this example, it's quite easy to scan the authors and titles and compare each record; this won't always be possible with a large number of records.

For this round, you have two duplicates.

And again, we'll save a compressed library as a back-up with the name 'Round3'.

Round Four

Now for Round 4 compare:

- Author
- Volume
- Issue
- Pages.

No duplicates are found for this round, we have no records that match for those criteria. I'm not surprised because we only have a small number of records with this example. However, we will still continue with Rounds 5 through 8 and the three extra rounds.

Since we didn't have any duplicates for this round, there are no changes to the library, so you do not need to save a compressed library.

Round Five

For Rounds 1 to 4, you haven't needed to check the results; for Rounds 5 onwards you want to give more attention to the duplicate matches. As you reduce the number of criteria to compare duplicates, you increase the risk of a false positive which can be avoid by keeping an eye on the

results. By Round 5, you have already identified the majority of your duplicate records, so it is much easier to scan the results for possible false positives.

For Round 5 compare:

- Title
- Volume
- Issue
- Pages.

For this round there is one duplicate and you can quickly compare the records. You can see that the authors' name is formatted differently for each record, which is why they haven't already been identified in an earlier round. I'm happy that this is a duplicate and I'll move it to the 'Duplicates' group.

Save a compressed library back-up named 'round5'.

Round Six

Back to the 'UNIQUE RECORDS' group and for Round 6 compare:

- Year
- Title
- Secondary title or the title of the journal
- Pages.

One duplicate has been found, move that to the 'Duplicates' group. Again, save the compressed library as a back-up, named 'Round6'.

Round Seven

For Round 7 compare:

- Author
- Year

- Title
- Secondary title or the title of the journal.

Three duplicates are found. You can see the differences and how the authors' names are formatted, how the article titles are formatted - with and without capitalisation - and differences in page numbers.

For the first example, we have pages 1 to 18 and for the duplicate 1 to 17. Looking at the other information; the article title, authors, journal title, I'm confident that this is a duplicate and the different page numbers are likely because of a discrepancy between databases or maybe a typo.

These differences come down to the metadata exported from the databases; some databases will have complete and comprehensive metadata, while others will be missing metadata or have the metadata formatted differently. When you tell EndNote how to find a duplicate; what criteria to look at, you systematically adjust the criteria many different times to capture all these differences and discrepancy. For this process, you want EndNote to do most of the work but keep control of how it does it.

Save the compressed back-up for Round 7.

Round Eight

For Round 8, compare:

- Year
- Title
- Secondary title; the journal title.

No duplicates. Since there were no duplicates, we can skip saving a compressed library and move to the extra rounds.

Extra Round One

There are three extra rounds which can be helpful for finding duplicates that are less clearly duplicates; records with authors who have common names, shorter or generic article titles, for

example. You want to be careful that you aren't removing any original records in these rounds, so manually check the results.

For Extra Round 1, compare:

- Volume
- Issue
- Pages.

No duplicates were found. It's not unusual to have very few results for these extra rounds. Again, no changes to the EndNote library, so no need to save a compressed library.

Extra Round Two

For Extra Round 2, compare:

- Author
- Secondary title; the title of the journal.

Again, no duplicates found so no need to save the compressed library.

Extra Round Three

For the final extra round, you only ask EndNote to compare the article titles. Obviously, we want to be careful with this round as it's the most likely to bring up a false positive.

There are three duplicates, and you can see that there are formatting differences like we're seen before: you can see that some records have a journal title 'Dissertation Abstracts International Section A Humanities and Social Sciences'. The reference type is different: some show as 'thesis' and others as 'journal article'. In this situation, one of the databases we searched has indexed theses, and these theses have been captured by another database and treated as journal articles. You want to look closely at the titles and the abstracts, but I'm confident that these are duplicates and will place them in the 'Duplicates' group, leaving 60 unique records and 37 duplicates.

Save the compressed library, which I will name with 'DeduplicationComplete' as this is the last round of deduplication work.

The PRISMA flow diagram

Now that you've completed the deduplication, you can populate the next box in your PRISMA flow diagram, 'records removed before screening'. For this example, populate the 'duplicate records removed' field with 37. No automation tools were used and no records were removed for other reasons. We can also populate the box 'records screened' as you know that there are 60 unique records to screen.

So, there was a total of 97 records from our three databases; 37 records were found as duplicates and removed, leaving 60 unique records to screen. The next step is to start screening the titles and abstracts of those 60 unique records.