Research Synthesis Methods (2025), **16**: 587–600
doi:10.1017/rsm.2025.20

Research Synthesis Methods

TUTORIAL

Same, same, but different: A method to harmonise and deduplicate study records from WHO ICTRP and ClinicalTrials.gov prior to screening

Zahra Premji[1] and Chris Cooper[2]

[1]Libraries, University of Victoria, Victoria, BC, Canada
[2]University of Bristol Medical School, Bristol University, Bristol, UK

Corresponding author: Zahra Premji; Email: zahrapremji@uvic.ca

Received: 5 December 2024; **Revised:** 4 March 2025; **Accepted:** 15 March 2025; published online 25 April 2025

Keywords: clinicaltrials.gov; deduplication; EndNote; ICTRP; trials registry

Abstract

Trials registry records represent a challenge in deduplication compared to deduplicating studies reported in journals and exported from bibliographic databases such as MEDLINE. We demonstrate why this is the case and propose a method to deduplicate registry records from the WHO International Clinical Trials Registry Platform (ICTRP) and ClinicalTrials.gov (CTG) specifically in the reference management tool EndNote (desktop version). We believe that our method is not only more efficient but that it will minimise the risk of registry records being incorrectly removed as duplicates in automated deduplication. The method has seven steps and is detailed in this tutorial as a step-by-step guide.

Highlights

- We illustrate the case for a separate and stand-alone process of deduplication for study records from trials registry resources.
- We illustrate the methods for deduplication using the bibliographic tool EndNote.
- By the end of the tutorial, a reader will be able to deduplicate records from WHO ICTRP and ClinicalTrials.gov prior to merging with the separately deduplicated bibliographic search results for study selection.

What is already known

Registry records present studies and study data in a format that differs from journal article records exported from bibliographic databases such as MEDLINE. This makes the export and import into bibliographic management screening tools and the deduplication process less efficient and less effective.

What is new

With automated deduplication now common in tools such as Covidence, there is a risk that (as proven in this tutorial) registry records might be inaccurately deleted as duplicate records and significant numbers of valid duplicates might not be removed.

We present a method to deduplicate records from WHO ICTRP and ClinicalTrials.gov in EndNote to overcome these issues.

Potential impact for RSM readers

This is the first step-by-step tutorial to illustrate a new process of deduplication of WHO ICTRP and ClinicalTrials.gov records.

© The Author(s), 2025. Published by Cambridge University Press on behalf of The Society for Research Synthesis Methodology. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NoDerivatives licence (https://creativecommons.org/licenses/by-nd/4.0), which permits re-use, distribution, and reproduction in any medium, provided that no alterations are made and the original article is properly cited.

https://doi.org/10.1017/rsm.2025.20 Published online by Cambridge University Press

1. Background

After navigating the eternal challenge of searching the WHO International Clinical Trials Registry Platform (ICTRP) and ClinicalTrials.gov (CTG),[1] the researcher is faced with the grim realisation that their battle with the registers is not yet ended. The records must be exported and deduplicated.

Searching trials registers is a mandatory item (C27) in the Methodological Expectations of Cochrane Intervention Reviews (MECIR) standards.[2] The Cochrane Handbook[3] (Chapter 4.3.3) specifically recommends searching both CTG and the WHO ICTRP and further recommends that they should be searched separately, despite the overlap in records. ICTRP contains trial records from CTG and many other trial registries across the globe.[4] Further details about the value of searching trial registers and the reasons for searching CTG and ICTRP separately, including the evidence base for these recommendations, are well described in Chapter 4 of the Cochrane Handbook of Systematic Reviews of Interventions.

2. The issue at hand–why are records from trials registry resources tricky?

Registry records represent a challenge in deduplication compared to deduplicating studies reported in journals and exported from bibliographic databases such as MEDLINE. Registry records are non-bibliographic in nature—registry resources report data in different and often more/non-standard fields compared to journals—and the same registry record can be found in multiple sources yet be formatted differently and so appear 'differently' when presented for deduplication.

To demonstrate, Figure 1 shows how the same study record (NCT01777542) appears in EndNote when exported from CTG and ICTRP. This is the same study, but these two records would not be identified as duplicates because none of the relevant fields (such as title, author, etc.) have similar data.

More than one title-related metadata field exists for a given reference type. The 'Generic' reference type, which is what ICTRP records are formatted for, has the following title-related fields: title, short title, and alternate title. On the other hand, CTG records are formatted as 'Online Database' records and have the title, short title and translated title fields. Trial registry records can have more than one title, specifically, a short title or public title and a longer scientific title. For ICTRP records, when imported into EndNote, the public title (which is the short title) goes in the title field, and the scientific title ends up in the alternate title field (Figure 2). Whereas, in CTG records, the short title is found in the short title field and the scientific title in the title field. Furthermore, the URLs in these two records are not identical, even though both lead to a record in CTG (see Figure 1).

This makes it harder to identify duplicate study records from trials registry resources when using automated or even manual methods to deduplicate.

Many systematic review screening tools include a deduplication feature which may be based on the comparison of a preset combination of fields. For example, Covidence software (Covidence.org), a commonly used systematic review software, states that it deduplicates based on exact matches of the title, year, and volume, and sufficiently similar values in the author field.[5] Based on the example provided in Figures 1 and 2, it is clear that Covidence would not identify the two records as duplicates of each other.

Recent studies exist that compare the deduplication performance of systematic review software/tools (such as Covidence) against other automated tools and reference managers (such as EndNote). In a 2021 study, comparing the deduplication efficiency of various tools for bibliographic records from

Author	Title	Year	Volume	Journal/Secondary Title	URL	Accession Number	DOI	Reference Type
	Pharmacological Treatment of Rett Sy...	2013			https://clinicaltrials.gov/study/NCT01777542	NCT01777542		Online Database
NCT01777542	Treatment of Rett Syndrome With Rec...	2013			https://clinicaltrials.gov/show/NCT01777542			Generic

Figure 1. *The same record (NCT01777542) exported from CTG and ICTRP and imported into EndNote 20.1–Windows version.*

https://doi.org/10.1017/rsm.2025.20 Published online by Cambridge University Press

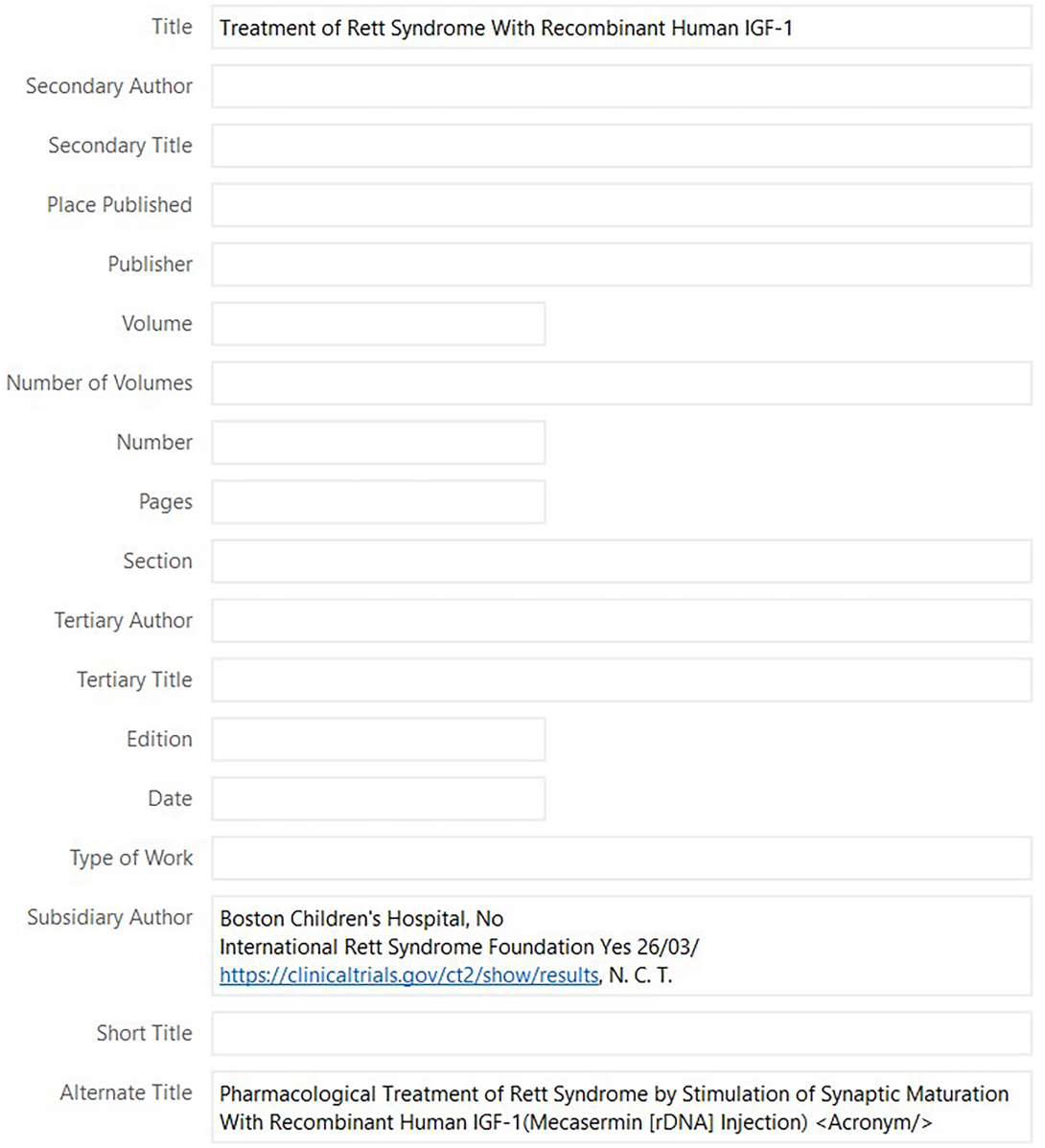


Figure 2. *Screenshot of the metadata in EndNote for NCT01777542 imported from ICTRP showing the title, short title, and alternate title field.*

health databases, Covidence scored well, with an accuracy of 96% and a sensitivity of 90%.[6] A more recent study used six datasets to test deduplication in four tools, of which four datasets included records from CTG and ICTRP.[7] Covidence demonstrated 100% precision (all duplicates identified were true duplicates) and 76.8% recall (an indicator of the number of missed duplicates). In comparing their result to the 2021 study, Janka and Metzendorf[7] specifically noted that the difference in accuracy of deduplication across the two studies may be due to the inclusion of trial registry data in the latter study's dataset.

We have, however, experienced some similar and some different results for registry records in terms of both precision and recall. Since trial registry records lack metadata in fields such as journal,

https://doi.org/10.1017/rsm.2025.20 Published online by Cambridge University Press

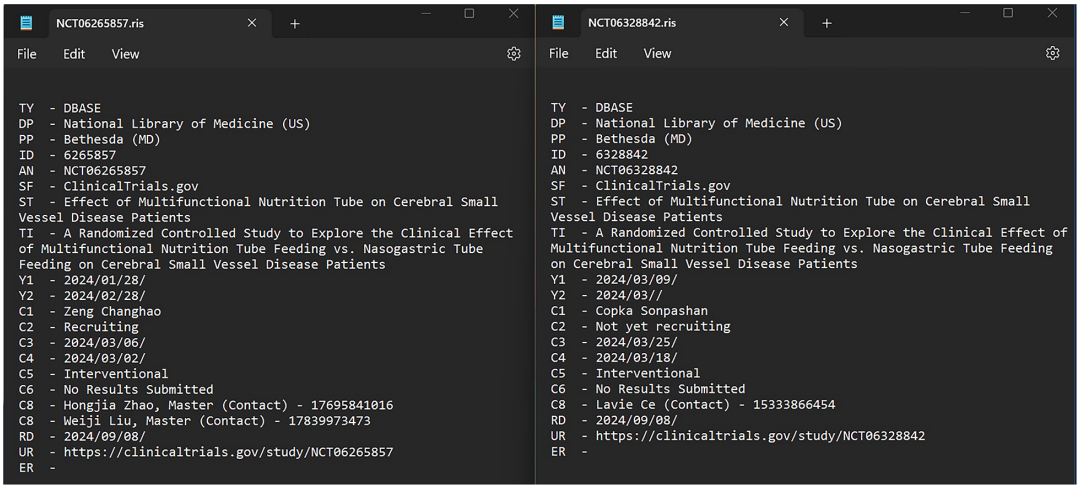


Figure 3. *Complete metadata of two records from CTG that are identified as duplicates by Covidence.*



Figure 4. *Data available in the records of NCT06265857 and NCT06328842 when imported into Covidence forcing them to be inaccurately identified as duplicates.*

volume, page numbers and DOI, when a current Research Information Systems (RIS) export from a trial registry is uploaded to Covidence, it displays only a limited number of fields, namely the title and year for both ICTRP and CTG records and the author and abstract fields for ICTRP records. Given the differences in organization of metadata across fields between CTG and ICTRP records, and with so little metadata available in the records, one risks removing false positives and including false negatives during Covidence deduplication (hence reducing the efficiency of the deduplication process). In a recent test, we came across two examples of false positives being removed; one such example is shown in Figures 3 and 4.

In Covidence, the deduplication is automatic and cannot be turned off, which is especially problematic if false positives are removed automatically. It should be noted that Covidence does have a manual review option, where researchers can look at all removed duplicate pairs individually, but a researcher would proactively have to know to look for this feature and be familiar with its use.

During the course of our work, in August 2024, comparing an original search (657 records) and an update search (817 records) in CTG, the method proposed in this tutorial accurately identified 576 duplicates. Covidence's automated performance for these records was less stellar. It mistakenly flagged

https://doi.org/10.1017/rsm.2025.20 Published online by Cambridge University Press

two records from the original search as duplicates, even though they were not (Figures 3 and 4). This left 655 records from the original search. When the new file containing 817 records was uploaded, 92 duplicates were removed, leaving 725 new records. The deduplication process resulted in a screening set containing 1,380 records, of which 2 missing records inaccurately flagged as duplicates (false positives) were removed, and 484 missed duplicates (false negatives) remained.

In a test conducted on 1 December 2024, using the same search string in both CTG and ICTRP, 68 and 22 records were retrieved, respectively (Appendix). An initial scan of the ICTRP records showed that it contained seven records from CTG, as recognised by the unique format of CTG study IDs. Importing both files into Covidence resulted in one duplicate being identified. To ensure that this wasn't only an issue in Covidence, we used the same files in another freely available deduplication tool, the SRA deduplicator,[8,9] and neither the focussed or relaxed modes found any duplicates between the two files. Following the process presented in this tutorial, however, resulted in the identification of all seven duplicates (100% precision and 100% recall), including the one found by Covidence. Analysis of the one duplicate record identified by Covidence showed that this particular record (NCT05036603) used the same title for the short/alternate title and scientific title, causing a title match between these two records and allowing this duplicate to be recognised.

Having demonstrated what the problem is, and why it happens, let us now turn our attention to the solution.

3. Our process

An external intermediary method (i.e., before Covidence, or other tools) is therefore necessary to efficiently deduplicate registry records. Since trials in registries have a unique study ID, this makes for a reliable unit for comparing duplicates. The format of the study ID varies across the different trial registries. In CTG, the study ID starts with NCT followed by eight digits (e.g., NCT05036603). In another registry, such as the Australian New Zealand Clinical Trials Registry, the study ID starts with ACTRN followed by 14 digits (e.g., ACTRN12608000601336).

Software such as EndNote (desktop version), a reference manager, which has the functionality to batch edit records and manipulate deduplication parameters is ideal for this task. Note that the web version of EndNote does not have the functions required to carry out this type of deduplication.

This method works for records from any trial registry where the unique trial number or study ID is present in a metadata field or as a fragment of the URL. We anticipate that this method will be of greatest use where multiple searches are made of multiple registry resources (e.g., CTG, ICTRP, CENTRAL) and/or where multiple searches are made of the same registry to break up complicated search syntax but for the same condition or intervention (e.g., Search 1, Search 2, Search 3, and so on). This method could also suit a model of searching which is focussed on trials reported in registers or via CENTRAL, as a separate step from the bibliographic searching.[10] An additional evaluation of this method is presented separately.[11]

Furthermore, with only the limited metadata available in screening software such as Covidence (Figure 4), screening such records would likely require accessing the registry record from the original source in order to verify eligibility criteria. This has been our experience, more often than not, where most study records have to be screened in full. The direct URL to the registry record does not appear in the Covidence record, despite being in the RIS file exported from CTG; this is due to the limited fields displayed in Covidence. This tutorial therefore also includes instructions on moving the URL for the record into the DOI field, which is more likely to be a standard displayed field in screening software.

4. Tutorial aim

In this tutorial, we present a method to deduplicate trials registry records. Our aim is to improve the efficiency and effectiveness with which duplicate trial records are removed in preparation for study selection. We aim to 'clear out' some of the mess.

https://doi.org/10.1017/rsm.2025.20 Published online by Cambridge University Press

We focus on trial records from two of the most commonly searched trial registry resources: CTG and ICTRP, with initial deduplication in EndNote (any version).

Registry records from databases such as Cochrane CENTRAL could also be included in this process since they can be isolated for download from the overall search of CENTRAL. We detail the specifics for including registry records from Cochrane CENTRAL (Wiley interface only) in this process as an addendum below.

5. Who is this tutorial for?

This tutorial is intended for researchers searching trials registers for evidence synthesis reviews. We assume that in reading this tutorial, people will already be familiar with the need for and general process of deduplicating study reports for evidence synthesis. An evidence-informed summary is available in Section 4.3 of the Technical Supplement to Chapter 4 of the Cochrane Handbook,[12] so we do not elaborate further here.

6. The deduplication method

Conceptually, the deduplication process has the following steps:

1) Download records from trial registry sources.
2) Import records into EndNote.
3) Move a copy of the trial number data to the label field.
4) Edit the data for consistency (if needed).
5) Alter deduplication parameters in EndNote.
6) Execute the deduplication.
7) Celebrate with tea.

6.1. Step 1: Downloading

While both registries include downloading capabilities, they do not provide the same file export formats. ICTRP allows download into xml or csv formats. EndNote provides an import filter[13] that must be used to accurately open the xml file generated by ICTRP. CTG recently added a RIS download option. RIS files can easily be opened by EndNote using the already available RIS import option.

We suggest that registry records are exported and kept separate to any other records until after deduplication.

6.2. Step 2: Importing records

You can use groups and group sets to keep records from each source/search separate:

a) Right click on ‘My Groups’ in the left column and select ‘Create group set’. Alternatively, go to ‘Group’ in the top menu bar, and select ‘Create group set’.
b) Name your group set, as per your existing folder naming convention.
c) Right click on the newly formed group set name, and select ‘Create group’.
d) Name the group, as per your existing naming convention.

Each distinct search (be it a second trial registry source, or an update search of the same registry source) should have its own group. Repeat the third and fourth steps above as needed based on the number of distinct trial registry searches.

The order of file imports matter.

A given tool may have a set rule about which record is retained and which is removed as the duplicate. For example, in Covidence, the older record, that is, the record first in, is retained, and

https://doi.org/10.1017/rsm.2025.20 Published online by Cambridge University Press

the newer record is removed. In Endnote 21 (with the 21.3 patch), there is an option to choose which record to keep (most complete, newest, oldest), however, older versions of EndNote would, as default, select the earlier record as the one to retain when deleting duplicates in batches without reviewing each individual duplicate pair. For more information, see EndNote's help page.[14]

The metadata within records differ between ICTRP and CTG, even for the same trial. In our experience, the ICTRP records have better and more fulsome metadata so should be prioritised for retaining over the CTG records.

The following order of import is recommended:

I. For search updates:
   - When using an older version of EndNote: import the original search results first, into a separate folder (group).
   - When using EndNote 21.3, the order of import will depend on your chosen setting, but we recommend using the same approach mentioned above. That is, retain older records, and import the older search results first.

II. For deduplicating results from multiple registry sources: We recommend importing the results from ICTRP first.

ICTRP exports in XML format require a filter to import correctly into EndNote. The filter and instructions are available on the EndNote website.[13]

To import ICTRP results into a group:

a) Go to File, then select 'import', then 'file'.
b) The import file window will pop up. Click on 'choose' and then select the ICTRP XML file from your computer.
c) For 'import option', select 'WHO ICTRP' if available, otherwise select 'other filters' which will open up a new window. There, find the 'WHO ICTRP' filter and click 'Choose'. This brings you back to the original import file window.
d) Ensure that 'Duplicates' is set to 'Import all', and 'Text translation' is set to 'no translation'.
e) Click 'import'.
f) Move the newly imported records by clicking on 'Imported references', then select all (using Control + A), and then drag and drop the records to the relevant group you created earlier.

To import CTG results which are in RIS format into a group:

a) Go to File, then select 'import' then 'file'.
b) The import file window will pop up. Click on 'choose' and then select the RIS file from your computer; make sure the 'import option' is set to 'Reference Manager (RIS)'; ensure that 'Duplicates' is set to 'Import all', and 'Text translation' is set to 'no translation'.
c) Then click 'import'.
d) The newly imported records can be found in the 'Imported references' group which is a default group label available in Endnote. Click on 'Imported references', then select all (using Control + A), then drag and drop the records to the relevant group you created earlier.

6.3. Step 3: Moving a copy of the data

The objective of this step is to locate the field that contains the study ID and move a copy of the study ID into a different field as we prepare for deduplication.

The data being sought (the study ID) may be in the Author field, Accession number field, or as a fragment of the URL field depending on the source registry and date of the records.

The destination field has to be one that is available in the deduplication field options and can handle numerical data. A complete list of fields can be found in the 'Duplicates' tab of the 'Preferences' menu in EndNote. We recommend the Label field, as it is a user-defined field and does not typically contain

https://doi.org/10.1017/rsm.2025.20 Published online by Cambridge University Press

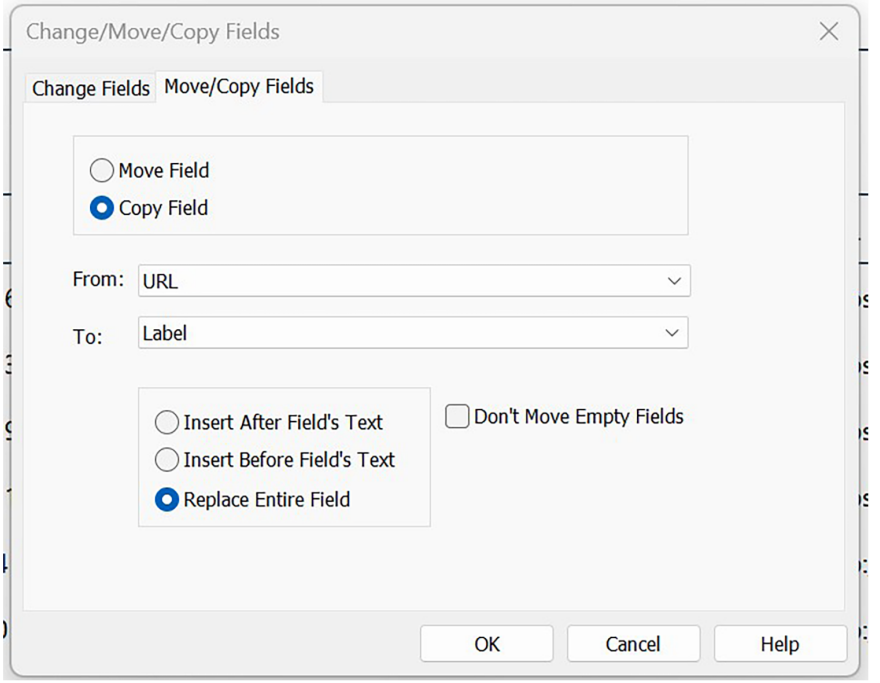


Figure 5. *Copying data from URL to Label field.*

any values from the incoming records themselves. In discussion with folks at EndNote, we understand that the Author field does not handle enumerated data, making it imperfect as the field to deduplicate against—especially when moving data between EndNote libraries (even though ICTRP output already uses this field).

Copy data in the URL, author or accession number field to the Label field, thusly:

a) Control A to select all records within the relevant group.
b) Go to: 'Library' (or 'Tools' in other EN versions) in the menu bar, and select 'Change/Move/Copy Fields'.
   - For older CTG imports that only have the study ID as part of the URL:
     i. Toggle 'Change Fields' to 'Move/Copy Fields' and then switch buttons to Copy Field'.
     ii. Toggle 'From' to URL and 'To' to Label (see Figure 5).
     iii. Select 'Replace entire field', and click 'OK'.
   - In newer CTG imports that have the study ID in the Accession Number field:
     i. Toggle 'Change Fields' to 'Move/Copy Fields' and then switch buttons to Copy Field'.
     ii. Toggle 'From' to Accession Number and 'To' to Label.
     iii. Select 'Replace entire field', and click 'OK'.
   - In ICTRP imports that have the study ID in the author field:
     i. Toggle 'Change Fields' to 'Move/Copy Fields' and then switch buttons to Copy Field'.
     ii. Toggle 'From' to Author and 'To' to Label.
     iii. Select 'Replace entire field', and click 'OK'.

It is important to *copy data* between fields rather than replace as the URL field is later used to source the registry record when selecting studies (because registry data are not bibliographic in nature, it is often necessary to 'screen' the record in full at title/abstract stage rather than waiting until full-text).

https://doi.org/10.1017/rsm.2025.20 Published online by Cambridge University Press

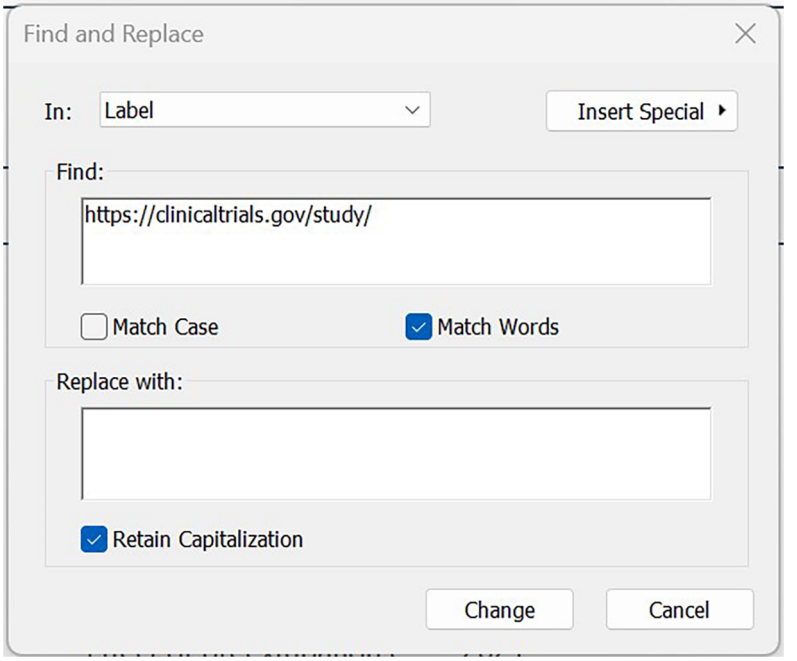


Figure 6. *Trimming the URL fragment to isolate the study ID.*

A point to flag is that the registry version of the record can change over time where the trial manager edits as the trials evolves. It is worth saving a copy of the PDF (or using other methods for archiving web content) in the study selection stage for comparison to any later trial record.

6.4. Step 4: Editing data

The objective of this step is to ensure that the 'Label' field has only the study ID. ICTRP and newer CTG records do not need to be edited, as the study ID has been copied and is already available in the Label field.

This step therefore applies to cases where the study ID is only available in the URL field of CTG records. The URL data are now in both the URL and Label fields. It is therefore time to edit the data in the Label field for deduplication.

a) Control + A to select all records within the relevant group.
b) Copy (Control + C) the following, standardised text: https://clinicaltrials.gov/study/.
c) Go to: 'Library' (or 'Edit' in other EN versions), then 'Find and Replace'.
d) Toggle 'in' to Label.
e) Populate the 'find' box with the text you copied earlier (via Control + C).
f) Leave 'Replace with' blank, and
g) Click the 'Change' button (Figure 6).

This removes the URL text preceding the unique NCT number, leaving only the unique NCT number per record in the Label field.

6.5. Step 5: Altering deduplication parameters

In the EndNote preferences, in the top left of the EndNote screen, change the default settings:

a) Go to 'Edit' (in the top menu bar), then 'Preferences'.
b) In the window that pops up, click on the 'Duplicates' tab.

https://doi.org/10.1017/rsm.2025.20 Published online by Cambridge University Press

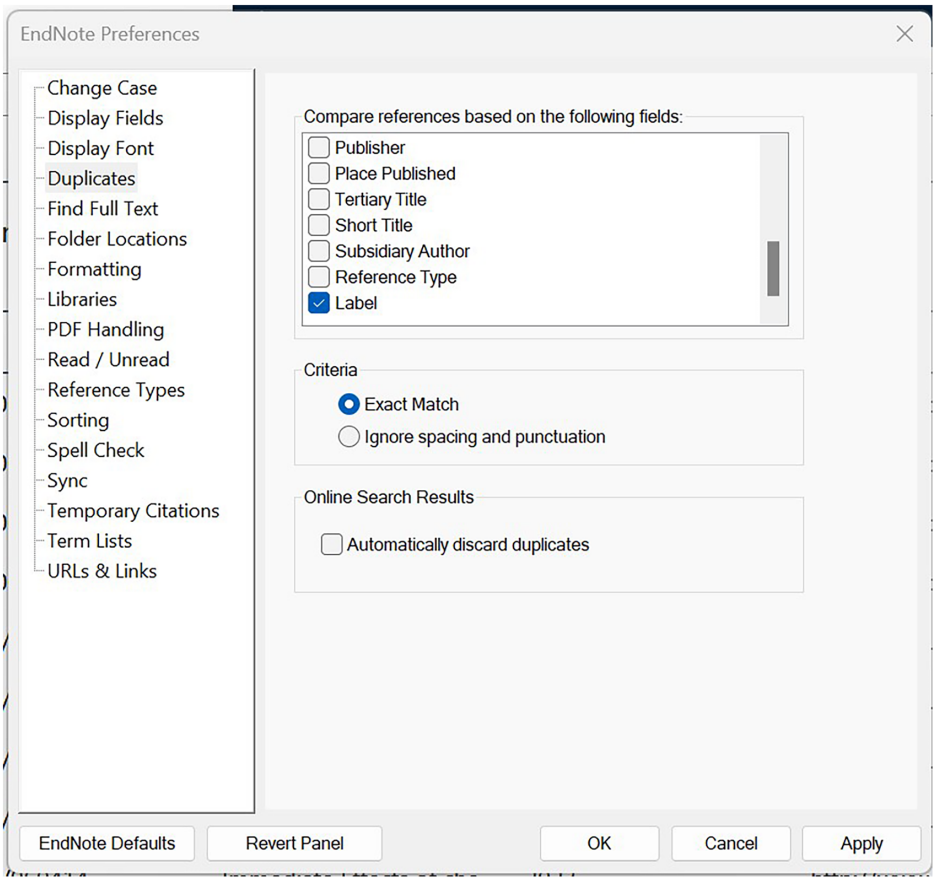


Figure 7. *Changing the deduplication parameters.*

c) In 'Compare references based on the following fields', select 'Label' (chose only this field and no other).
d) select 'Exact Match', and
e) Click 'Apply' (or 'Save' in other EN versions) and then 'OK' (See Figure 7).

6.6. Step 6: Deduplicate

Run the deduplication (which works, as the number in the label field is the unique study ID per record).

a) First select the group set by clicking on the name of the group set.
b) Then, go to 'Library' (or 'References' in other EN versions), and select 'Find duplicates'.
c) EndNote will give you the option to review each duplicate pair. To delete all duplicates in one batch, click 'Cancel' which will close the box and select all duplicate records. You can then use the 'delete' button on your keyboard to send these references to the trash.

Record the number of duplicates removed for reporting in the PRISMA flow diagram.[15]

6.7. Step 7: Time for tea

A celebratory beverage is in order. Make tea or similar.

https://doi.org/10.1017/rsm.2025.20 Published online by Cambridge University Press

7. Deduplicating Cochrane CENTRAL registry records as part of this process

The process for deduplicating trial registry records from CENTRAL (Wiley interface only) against CTG and ICTRP records requires some minor modifications. Not all steps require modifications. Those that do are detailed below. All other steps of the tutorial can proceed as written.

Step 1: Records can be exported in RIS format, and we recommend exporting separately the records from CTG and ICTRP (i.e., separately from each other) in CENTRAL.

Step 2: When importing the records into EndNote, we recommend using separate 'Groups' for each. For the order of import into EndNote, we recommend importing ICTRP records first, followed by CENTRAL records, and importing CTG records last.

Step 3: The study ID (along with the registry name) can be found in the 'custom 3' field in CENTRAL records. Therefore, step 3b) ii will change to the following: Toggle 'From' to custom 3 and 'To' to Label.

Step 4: The name of the registry will need to be removed from the 'Label' field. Find and replace is done in step 4b in the tutorial. But, the text to be replaced will differ for ICTRP versus CTG records in CENTRAL as follows.

I. For ICTRP records in CENTRAL, copy (Control + C) the following text that is enclosed in quotation marks (note that the inclusion of the space is deliberate): 'ICTRP'.
II. For CTG records in CENTRAL, copy (Control + C) the following text that is enclosed in quotation marks (note that the inclusion of the space is deliberate): 'CTgov'.

8. Additional useful modifications

If the intention is to take the resulting deduplicated trial registry records and screen them alongside bibliographic records using systematic review software (such as Covidence), then it is helpful to make some additional modifications to the records.

8.1. Add-on modification 1– Move the URL to the DOI field

Covidence does not display the URL field even if the records being imported as RIS contain data in the URL field. One workaround is to move a copy of the URL data to the DOI field, which is displayed by Covidence during both stages of screening. The DOI field is typically empty in registry records, so no data will be overwritten. This would ensure that screeners have the ability to click through to the record on the trial registry website, should it be required, as it often is. The instructions for this step are similar to step 3 above.

For both CTG and ICTRP records (in their respective groups):

a) Select all records, using Control + A.
b) Go to Library, select 'Change/Move/Copy Fields'.
i. Toggle 'Change Fields' to 'Move/Copy Fields' and then switch buttons to Copy Field'.
ii. Then, toggle 'From' to URL and 'To' to DOI.
iii. Select 'Replace entire field' and click 'OK'.

8.2. Add-on modification 2–Make sure the title field contains the scientific title

To increase consistency of the title field, one could move the short title from the title field in ICTRP records to the short title field. And move the scientific title from the Alternate title field to the title field. This will ensure that both CTG and ICTRP records have the longer scientific title in the title field during

https://doi.org/10.1017/rsm.2025.20 Published online by Cambridge University Press

screening. This may be helpful as the scientific title is longer and more descriptive. To do this (only to the ICTRP group set):

a) Select all records in the ICTRP group, using Control + A.
b) To move the short title from the 'Title' to the 'Short title' field, go to 'Library', select 'Change/Move/Copy Fields'.
i. Toggle 'Change Fields' to 'Move/Copy Fields' and then switch buttons to 'Move Field'.
ii. Then, toggle 'From' to Title and 'To' to Short title.
iii. Select 'Replace entire field' and click 'OK'.
c) To move the scientific title from the 'Alternate title' to the 'Title' field, go to Library, select 'Change/Move/Copy Fields'.
i. Toggle 'Change Fields' to 'Move/Copy Fields' and then switch buttons to 'Move Field'.
ii. Then, toggle 'From' to Alternate Title and 'To' to Title
iii. Select 'Replace entire field' and click 'OK'.

9. Concluding remarks

The method presented will remove duplicate study records under the same study ID. We note that registry records are editable study reports, updated as the study in question evolves. Once eligible studies have been identified during the search stage, the trials registry record should be regularly checked for any update on study progress.

Any update searches made, using this method of deduplication, will remove a duplicative study record which might mean that a researcher will miss a study update associated with the study ID. Therefore, we recommend that researchers re-check study records for studies that were not already 'complete' when screening records during a search update.

Using the method demonstrated in this tutorial, the deduplicated records can be incorporated for study selection with the search results from bibliographic databases and other search methods.

Author contributions. The deduplication method described was conceived by Chris who wrote the first draft of this article whilst on a train in Italy. Zahra wrote the second draft, including the analysis in Covidence to illustrate the need for the method, whilst at her desk in Victoria. Both authors then developed the article for submission during the autumn of 2024. Many teas were consumed during this process.

Competing interest statement. The authors declare that no competing interests exist.

Data availability statement. Study Data not available–Data related to the August 2024 CTG search updates were gathered in the course of a project, and not an evaluation study, and can therefore not be shared. All other data are reported in the article.

Funding statement. The authors received no funding for any part of this work.

References

[1] Cooper C, Court R, Kotas E, Schauberger U. A technical review of three clinical trials register resources indicates where improvements to the search interfaces are needed. *Res Synth Methods.* 2021;12(3): 384–393. https://doi.org/10.1002/jrsm.1477.
[2] Higgins J, Lasserson T, Thomas J, Flemyng E, Churchill R. Methodological expectations of Cochrane intervention reviews (MECIR). Cochrane; 2023. https://community.cochrane.org/mecir-manual.
[3] Lefebvre C, Glanville J, Briscoe S, et al. Searching for and selecting studies. In: Higgins JPT, Thomas J, Chandler J, et al., eds. *Cochrane Handbook for Systematic Reviews of Interventions. 6.5*. Cochrane; 2024: Chapter 4. Accessed July 29, 2024. https://training.cochrane.org/handbook/current/chapter-04#section-4-4.
[4] World Health Organization. International clinical trials registry platform data providers. https://www.who.int/clinical-trials-registry-platform/network/data-providers.
[5] Veritas Health Innovation. FAQ: How does Covidence detect duplicates? Covidence systematic review software. July 19, 2023. https://support.covidence.org/help/how-does-covidence-detect-duplicates.
[6] McKeown S, Mir ZM. Considerations for conducting systematic reviews: Evaluating the performance of different methods for de-duplicating references. *Syst Rev.* 2021;10(1): 38. https://doi.org/10.1186/s13643-021-01583-y.

https://doi.org/10.1017/rsm.2025.20 Published online by Cambridge University Press

[7] Janka H, Metzendorf MI. High precision but variable recall – comparing the performance of five deduplication tools. *J Eur Assoc Health Info Libr.* 2024;20(1): 12–17. https://doi.org/10.32384/jeahil20607.
[8] Forbes C, Greenwood H, Carter M, Clark J. Automation of duplicate record detection for systematic reviews: Deduplicator. *Syst Rev.* 2024;13(1): 206. https://doi.org/10.1186/s13643-024-02619-9.
[9] Deduplicator. Accessed December 1, 2024. https://sr-accelerator.com/#/deduplicator.
[10] Cooper C, Premji Z, Worsley C, Tomlinson E, Dawson S, Prentice E. SA80 a new process model for study identification in systematic review: Separating studies from reports. *Value Health.* 2024;27(12): S630. https://doi.org/10.1016/j.jval.2024.10.3161.
[11] Zabzuni E, Anzola D, Almuallem L, Worsley C, Cruz F, Smith S, Cooper C. Evaluation of a novel approach to deduplication of trial registry records in systematic reviews. Accepted poster to be presented at: ISPOR 2025; May 13, 2025; Montreal, QC.
[12] Lefebvre C, Glanville J, Briscoe S, et al. Technical supplement to Chapter 4: Searching for and selecting studies. In: Higgins JPT, Thomas J, Chandler J, et al., eds. *Cochrane Handbook for Systematic Reviews of Interventions Version 6.4*. 2024. https://training.cochrane.org/chapter04-tech-supplonlinepdfv64-final-200224.
[13] WHO International Clinical Trials Registry Platform (ICTRP). Import Filter for EndNote. https://endnote.com/downloads/filters/who-international-clinical-trials-registry-platform-ictrp/.
[14] Clarivate. EndNote 21 & 20: Steps to remove duplicates from your EndNote library. https://support.clarivate.com/Endnote/s/article/EndNote-20-Steps-to-remove-duplicates-from-your-EndNote-library?language=en_US.
[15] Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *J Clin Epidemiol*. Published online 2021. https://doi.org/10.1016/j.jclinepi.2021.03.001.

Appendix

The test example below is provided for readers for the purposes of testing and working through the tutorial method.

As of 1 December 2024, the test example produces seven duplicates: NCT00884429, NCT02041676, NCT02458300, NCT02853838, NCT03753802, NCT04689386, and NCT05036603.

https://doi.org/10.1017/rsm.2025.20 Published online by Cambridge University Press

Table A1. *Search queries used in a comparison test and numbers of results found in ICTRP and CTG.*

Source	Search query	Settings	Results on 01/12/2024
CTG	((babe OR babes OR baby OR babies OR 'gestational age' OR 'gestational ages' OR infant OR infants OR infantile OR infancy OR 'low birth weight' OR 'low birthweight' OR neonate OR neonatal OR 'neo nate' OR 'neo natal' OR newborn OR 'new born' OR newborns OR 'new borns' OR 'newly born' OR premature OR 'pre mature' OR 'pre matures' OR prematures OR prematurity OR 'pre maturity' OR preterm OR preterms OR 'pre term' OR preemie OR preemies OR premies OR premie OR VLBW OR VLBWI OR 'VLBW I' OR VLBWs OR LBW OR LBWI OR LBWs OR ELBW OR ELBWI OR ELBWs OR NICU OR NICUs) AND ('Chest physiotherapy' OR 'Chest Therapy'))	Searched in the 'Other Terms' field	68
ICTRP	((babe OR babes OR baby OR babies OR 'gestational age' OR 'gestational ages' OR infant OR infants OR infantile OR infancy OR 'low birth weight' OR 'low birthweight' OR neonate OR neonatal OR 'neo nate' OR 'neo natal' OR newborn OR 'new born' OR newborns OR 'new borns' OR 'newly born' OR premature OR 'pre mature' OR 'pre matures' OR prematures OR prematurity OR 'pre maturity' OR preterm OR preterms OR 'pre term' OR preemie OR preemies OR premies OR premie OR VLBW OR VLBWI OR 'VLBW I' OR VLBWs OR LBW OR LBWI OR LBWs OR ELBW OR ELBWI OR ELBWs OR NICU OR NICUs) AND ('Chest physiotherapy' OR 'Chest Therapy'))	Searched in the main search box on the homepage	22

Cite this article: Premji, Z, Cooper, C. Same, same, but different: A method to harmonise and deduplicate study records from WHO ICTRP and ClinicalTrials.gov prior to screening. *Research Synthesis Methods*. 2025;16: 587–600. https://doi.org/10.1017/rsm.2025.20

https://doi.org/10.1017/rsm.2025.20 Published online by Cambridge University Press