

# Deduplication of cross-registrations manual

*TL;DR*

- [Here](#) is a dataset with our trial sample.
- Trials marked “Complement” may have cross-registrations in the other registry that haven’t already been identified automatically or in publications.
- We are a small subset of the team who will follow a predefined algorithm to search for cross-registrations and verify if they are true matches.
- We will all work in the same Google spreadsheet as per this manual.
- If you have questions, just reach out. Thank you for contributing!

## Orientation: the spreadsheet

The spreadsheet can be found [here](#).

Each row is a unique trial and the sheet contains all trials (eligible and not) in my original registry searches.

Variable descriptions can be found last in this document.

## Search algorithm

**Aim for the total process for 1 trial (search + verification) to take <5, and absolutely <10, minutes.** From the first batch of searches, it seems to me that either a match is quickly identified (and it’s likely to be a true one), or there is none. In about 10-20% of cases there is a partly overlapping registration. In those cases, try to identify what’s going on and leave a comment, but feel free to leave the final decision open (on whether to count it as a true match or not).

### Identification

A. For a row in the spreadsheet marked “Complement” without a name in column R, start by writing your name in column R.

B. Use the trial ID in column B (primary\_id) to find in the main registry (column E, registry) the original trial registration.

C. Open the OTHER registry’s search page (e.g., for project ID 1777 the main registry as per col. E is ct.gov, and so, the other registry is eucrr).

D. Use the following algorithm to search in the trial registry.

(Protocol:)

*For all EUCTR trials without an automatically identified match, and for ClinicalTrials.gov trials on medicinal products without an automatically identified match, we will manually search the opposite registry website with the following term combinations, one by one:*

*1) Trial IDs (one at a time; first identifying these manually at the original registration or in a publication)*

We only use column B (primary\_id), you don't need to look for other IDs.

*2) Trial name (e.g., a specific acronym)*

Column F in the spreadsheet

*3) Trial title*

Column G in the spreadsheet

*4) Location (country) and a combination of terms for indication/disease and intervention*

Combine terms according to your best judgment.

See the original registration to find the terms.

#### Verification of match

E. If you at any of these points find a potential match, check it against the original registration as follows to see if it's a true match.

(Protocol:)

*We define a cross-registration as an entry in one registry that matches an entry in the other registry (or its results publication) on all the following criteria:*

*1. Trial identifier (ID) or title or sponsor*

*2. (a) study design (intervention model, masking, randomization), (b) indication/population, c) intervention, (d) comparator (if applicable), and (e) outcome.*

If you need to discuss the match, note the potential cross-registration ID in column P (crossreg\_id\_team), leave your comment in column Q, and contact another reviewer. For example, it can happen that a registration in one registry has several cross-registrations in the other (partly overlapping).

If you find *no* true match, continue to F.

If you find a true match, continue to G.

#### Record-keeping

F. If you find no true match, make the following notes:

- Change "Complement" in column N to "None found".
- Your name should be in column R.

(Search done)

G. If you find a true match, make the following notes:

- Write Yes in column O (crossreg\_ver\_team).
- Write the cross-registration ID in column P (i.e., not the same as in primary\_id column, but the other ID).
- Make any comment in column Q.
- Your name should be in column R.
- Change “Complement” in column N to “Done”.

H. Importantly: search the Google spreadsheet with the cross-registration ID to see if it’s present. If it is present, and column N says “Complement”, leave your name in column R and change column N to “Duplicate”. If column N says something else than “Complement”, don’t change it.

*For example, my original trial was 2013-001240-64, and I identified a true match that is NCT02333591. I searched the Google spreadsheet for NCT02333591. In this case there was a row with that ID (proj\_id 531). I left my name and marked it “Duplicate”.*

I. Finally, please help by checking whether the cross-registration states that it’s withdrawn without enrollment. If so, please make this note in the beginning of column Q: “WITHDRAWN WITHOUT ENROLLMENT”.

(Search done)

## Variables in the spreadsheet

### Variable: Description

proj\_id: ignore

primary\_id: The trial ID in the main registry

eudract\_id: ignore

nct\_id: ignore

registry: The main registry

short\_title: Used in search step D

long\_title: Used in search step D

intervention\_type

crossreg: Ignore, comes from automatic de-cross-registration in R

crossreg\_publication: Ignore, comes from cross-registrations found through publications

crossreg\_id: Ignore

crossreg\_ver: Ignore

crossreg\_ver\_comment: Ignore

complement\_crossreg: Shows the status of the search for this record: e.g., “Complement”, “Done”, “None found”

crossreg\_ver\_manual: States “Yes” in case of a verified true cross-registration

crossreg\_id\_manual: Gives the cross-registration ID

crossreg\_manual\_comment: Here you can leave any comment about the search/matching

crossreg\_manual\_extractor: Name of reviewer who made the cross-reg search

