

Discrepancy corrections manual

Thank you for contributing to the systematic search for results publications of Nordic clinical trials!

TL;DR

- [Here](#) are datasets with discrepant searches. It contains the earliest publication that the 2 reviewers found for each trial, together with information on what to check. The dataset with your name is on your initiative to solve.
- [Here](#) are the raw extractions that every reviewer made.
- You *should* discuss unclear cases with the paired reviewer, but you *should NOT* discuss trivial points that you can solve yourself. The Slack channel is also here for you to consult the lead team for extra tricky cases!
- We ask you to solve the discrepancies and confirm when you're done.
- **How? Read this manual closely (~15-20 min)!**

Why discrepancy corrections?

In the first round, each trial was searched by 2 reviewers. In this second and final round, we check so that extracted data are aligned. However, not for all data!

In the first round, you searched for *all potentially eligible* publications for trials. For the current project, we want to end up with a dataset with 1 publication for each trial: the *earliest eligible* publication. Discrepancies regarding the earliest eligible publication is what we now ask you to solve.

What do you mean, we didn't extract dates! How do we know which is the earliest?

"Behind the scenes", the lead team has used R to clean the >5900 submitted forms and to automatically extract publication dates from PubMed. We have used trial identifiers (NCT ID, EudraCT ID) and publication identifiers (DOI) to automatically check whether the 2 reviewers were in agreement or not about the earliest publication for each trial. To make the discrepancy corrections as easy as possible, we provide automatic agreement variables to guide what you need to check.

Let's get started!

Materials you need

- Your discrepancy sheet (see the Google spreadsheet with your name).
Note: Please work only in the Google spreadsheet (don't download the file), since we made rules for how to edit certain cells.
- The raw extractions for each reviewer
- Protocol (and the search manual) for eligibility criteria
- Slack channel to contact your paired reviewer & to consult the lead team for particularly tricky cases
Note: Slack channel content is not saved for more than 3 months, so don't use Slack for log-keeping.

Orientation: raw extractions sheets

These are files to consult to see what extractions were made. They are NOT meant for editing or submitting any corrections.

Each row is from a submitted form, so it's either a publication or a note that no publication was found for a particular trial. It contains responses to all the questions in the form, as you entered them.

Note: it also contains automatic "agreement checks" for each publication. This can come in handy later! For example, whether the other reviewer found the particular publication or not, and how many criteria they thought it matched. They also flag publications that were published before the trial start date or completion date.

Tip: The guidance below often states "Check the original extraction sheets for the reviewers to identify the earliest publication". To do this, filter or search the trial ID column ("var3") for the trial ID. You need to download the file to filter it.

Tip: see [variable explanations](#).

Orientation: your discrepancy sheet

This is the file that shows what needs to be checked, and also the file where you enter your corrections. Please work only in the Google spreadsheet (don't download the file), since we made rules for how to edit certain cells.

Each row is for a unique trial. Only the trials with discrepancies to correct are in the file. We have divided trials between reviewers.

We ask you to take the **primary initiative** to solve discrepancies in the file with your name on it. We also ask you to collaborate on the other trials you extracted (that are in other people's files).

For each trial, you can see:

- General information
- The names of the two assigned reviewers
- Overall, trial-level, automatic agreement checks between reviewers
They answer the questions: Did one or both reviewers find *any* publications? Is the

earliest publication the same for both reviewers? Did one or both reviewers consider the publication to be a match according to all 5 criteria, or less?

- Information on the *earliest publication* as it was extracted by reviewer 1 and reviewer 2, respectively. (The lead team has automatically identified the earliest publications among your searches after extracting publication dates.) Reviewer 1 columns end with “r1” and reviewer 2 columns end with “r2”.
- Article-level agreement checks (the same that are in the raw extraction sheets!) For example, information whether the other reviewer found the particular publication or not, and how many criteria they thought it matched.
- **Correction columns to fill out**

Tip: see [variable explanations](#).

Correction algorithm

General advice: aim to spend maximally **10 minutes** per trial by yourself. If there are interpretation points you're unsure of, don't spend >10 minutes on the matter, but bring it up for discussion. Aim not to bring up trivial matters for discussion.

1. **Open the discrepancy sheet. Look at the column “consensus”:** What information do you need to correct?
2. See the section “Disagreement variants” below for guidance.
3. Do you need to discuss with another reviewer?
Discussion is not compulsory for each disagreement. You don't have to discuss trivial points (like obvious typos). Discussions are meant for interpretation matters such as which criteria are matching.
Preferably, ask your paired reviewer. If they are unavailable, bring the matter up in the Slack channel. If both you and your paired reviewer are puzzled, bring the matter up in the Slack channel. Be as concise as you can!
4. **Select one of the responses in the column “correction”:**
 - a. *Use “r1” entries*
(Means: Use reviewer 1 information, except any small correction, which is to be entered in the correction columns)
 - b. *Use “r2” entries*
(Means: Use reviewer 2 information, except any small correction, which is to be entered in the correction columns)
 - c. *Use new corrected entries*
This means another response than the ones already in the discrepancy file by r1 or r2. That is, another one from the raw extractions sheet, or a completely new answer. Use this if *all* information needs to be replaced. Then fill out all the correction columns. If the new corrected entry is that no publication was found, you of course only need to fill out “has_publication” (“no”).

Correction columns:

- has_publication (yes or no)

- publication_doi (DOI)
- publication_pmid (PMID)
- publication_url (URL)
- publication_date (earliest publication date available)
- identification_step (trial registration or Google search)
- publication_type (journal publication, preprint, or other)
- public_comment

Note: Please take care to avoid typos, and do not leave any comments except in the “public_comment” column, which is meant to be public.

Disagreement variants

First: a note on dates!

If the consensus column reads “implausible date” or “check date”, these points need to be taken into account when judging whether the publication is truly eligible.

If the publication date was before the start date of the trial, we have left a note: “Implausible date”. Is the publication truly eligible? Something is odd about the case. As a rule, these *should* be brought up in the Slack channel.

If it was published before the completion date of the trial, we left another note: “Check date”. Is the publication truly eligible? These are more likely to be eligible, since a trial can publish results on some outcomes before the last patient’s last visit was made.

A. Not consensus on whether publication exists

One reviewer found a potentially eligible publication and the other found none. Is the publication truly eligible? If yes, solved (select the correct reviewer response). If not:

Check the original extraction sheets for the reviewers to identify the next earliest publication. Is it truly eligible? If yes, solved (enter the publication as new corrected entries). If not:

Repeat, until there were no more potentially eligible ones (solved: select the correct reviewer response, no publication found).

B. Not consensus on earliest publication

Both reviewers found at least one publication, but the earliest one was not the same. Is the *earliest* of the two truly eligible? If yes, solved (select the correct reviewer response). If not:

Check the original extraction sheets for the reviewers to identify the next earliest publication. (It could be another than the one in the discrepancy sheet.) Is it truly eligible? If yes, solved (enter the publication as new corrected entries). If not:

Repeat, until there were no more potentially eligible ones (solved: enter as new corrected entry that no publication was found).

C. Consensus on earliest publication, but near-match

Both reviewers found this publication but at least 1 rated it as not a full match. Is it truly eligible? (Tip: we automatically checked if the other reviewer rated it as a full match or not, to guide discussions.)

If yes, solved (select either of the reviewers’ responses). If not:

Check the original extraction sheets for both reviewers to identify the next earliest publication. Is it truly eligible? If yes, solved (enter the publication as new corrected entries). If not:

Repeat, until there were no more potentially eligible ones (solved: enter as new corrected entry that no publication was found).

D. Consensus on earliest publication, but warning for publication date

Both reviewers found this publication, but it seems to be published before the trial was ended. Is it truly eligible? If yes, solved (select either of the reviewers' responses). If not:

Check the original extraction sheets for both reviewers to identify the next earliest publication. Is it truly eligible? If yes, solved (enter the publication as new corrected entries). If not:

Repeat, until there were no more potentially eligible ones (solved: enter as new corrected entry that no publication was found).

E. Not extracted by 2 reviewers

In our records it seems like 1 reviewer never searched this trial. Is it you or your paired reviewer?

The person with a missing trial should first try to find the search in their original extraction sheet. Is there a typo in the trial ID? Or was the only identified publication(s) not eligible matches (<4 criteria matching)? If the trial has been extracted already, look through the extractions for agreement and follow the instructions for correction cases A-D.

If the trial has been missed, the person with missing search should fill out the form for the trial and confirm to me (CA) when it's done. I'll then provide you with the automatic agreement checks so that you can solve any discrepancies.

NEW FORM for missing searches: <LINK>

F. Typo or only ineligible publications

For these cases, both reviewers made a form entry, but something was incorrect. There was either a typo in the trial ID or any identified publications were not potentially eligible (only matching 1-3 criteria).

Check the original extraction sheets to identify any pertaining entries made by the reviewer with "missing" answer. Then look through the extractions for agreement and follow the instructions for correction cases A-D.

What does “truly eligible” mean?

A publication needs to match the registration on all 5 criteria to be eligible (see the protocol and the search manual). But don't exclude a publication for “trivial” points like minor phrasing differences between the registration and publication.

If the consensus column reads “implausible date” or “check date”, these points need to be taken into account when judging whether the publication is truly eligible. The entered publication date (or PMID) can be wrong, or the publication not eligible, or other oddity.

In essence, this is what we want to answer:

- Is the publication truly for the same trial? We may need to consult publication dates and trial start/end dates.
- Is the publication capturing the trial results? Here we should necessarily be generous in our interpretation.

To find out, consult our protocol in detail. If it doesn't guide you fully, bring the case up in the Slack channel. Amendments to eligibility criteria following such case-by-case discussions will be put here below for further guidance.

Q&A

Q: What happens with all the other publications we extracted?

A: All potentially eligible publications will be kept in the dataset that we share openly to enable meta-research projects on trial reporting. The dataset will be described to contain “*potentially* eligible publications” and we will provide the double-extractions together with automatic agreement checks and information on which criteria that matched.

Q: It seems like some publications that I extracted have been missed, why?

A: First, in case there were typos in the DOI, it may be that your extraction was missed. We solved typos as best we could, but some are left unsolved.
Second, all publications that were rated to match only 1-3 criteria have been deliberately left out in the automatic identification of the earliest publication, since we don't consider them potentially eligible. However, they are still in your raw extractions sheet for your view.

Discrepancy sheet variables

proj_id: project ID, unique for each trial.

main_id: registration ID (NCT ID or EudraCT ID)

reviewer_1: Name of reviewer 1 (that's you).

reviewer_2: Name of your paired reviewer for this trial.

registry: clinicaltrials.gov or EUCTR

reviewers_done: Which reviewers that have completed searches for this trial

consensus: What needs to be checked! (Dis)agreements between reviewers for this trial

start_date_recoded: Trial start date. Sometimes missing.

completion_date_recoded: Trial completion date. Rarely missing.

has_publication_r1: Whether there is an eligible publication or not for this trial according to reviewer 1.

publication_doi_r1: DOI for this publication (selected among all identified publications as the earliest one) according to reviewer 1.

article_agreement_information_r1: Agreement information! What did the 2 different reviewers (dis)agree on regarding this particular publication?

article_found_by_r1: Which reviewers found this particular publication?

n_match_agree_r1: How many criteria did the 2 reviewers find this publication to match?

publication_pmid_r1: PMID for the publication according to reviewer 1.

publication_url_r1: URL for the publication according to reviewer 1.

publication_date_r1: Date for the publication according to reviewer 1.

identification_step_r1: How the publication was identified, according to reviewer 1.

publication_type_r1: Type of publication, according to reviewer 1.

match_design_r1: Whether the publication matched on the design criterion, according to rev. 1.

match_indication_population_r1: Whether the publication matched on the indication/population criterion, according to rev. 1.

match_intervention_r1: Whether the publication matched on the intervention criterion, according to rev. 1.

match_comparator_r1: Whether the publication matched on the comparator criterion, according to rev. 1.

match_any_outcome_r1: Whether the publication matched on the outcome criterion, according to rev. 1.

n_matching_criteria_r1: Number of matching criteria, according to rev. 1.

comment_matching_r1: Any comment rev. 1 made to this extraction regarding matching criteria.

comment_overall_r1: Any comment rev. 1 made to this extraction.

warning_cd_to_publication_r1: Flags if the publication date was before completion date of trial.

warning_start_to_publication_r1: Flags if the publication date was before start date of trial.

has_publication_r2: **For the following, see corresponding columns for reviewer 1! But this is for reviewer 2, and note that any identified publication MAY NOT BE THE SAME ONE.**

publication_doi_r2

article_agreement_information_r2

article_found_by_r2
n_match_agree_r2
publication_pmid_r2
publication_url_r2
publication_date_r2
identification_step_r2
publication_type_r2
match_design_r2
match_indication_population_r2
match_intervention_r2
match_comparator_r2
match_any_outcome_r2
n_matching_criteria_r2
comment_matching_r2
comment_overall_r2
warning_cd_to_publication_r2
warning_start_to_publication_r2

correction: Enter your response here from the drop-down list. If any items need to be corrected, or if you need to replace entirely with a new response, look also at the following columns.

has_publication_corrected: Yes or no. Enter correct response here (either if you need to correct one or few items in the selected reviewer's responses, or if you need to replace entirely with a new response)

publication_doi_corrected: DOI. Enter correct response here (either if you need to correct one or few items in the selected reviewer's responses, or if you need to replace entirely with a new response)

publication_pmid_corrected: PMID. Enter correct response here (either if you need to correct one or few items in the selected reviewer's responses, or if you need to replace entirely with a new response)

publication_url_corrected: URL. Enter correct response here (either if you need to correct one or few items in the selected reviewer's responses, or if you need to replace entirely with a new response)

publication_date_corrected: YYYY-MM-DD. Enter correct response here (either if you need to correct one or few items in the selected reviewer's responses, or if you need to replace entirely with a new response)

identification_step_corrected: Trial registration or Google search. Enter correct response here (either if you need to correct one or few items in the selected reviewer's responses, or if you need to replace entirely with a new response)

publication_type_corrected: Journal article, preprint, or other. Enter correct response here (either if you need to correct one or few items in the selected reviewer's responses, or if you need to replace entirely with a new response)

public_comment: This is the place to make any comment about the extraction. It's meant to be public!

Raw extractions variables

var1: Date when form was submitted.
var2: Name of reviewer.
var3: Trial ID as entered.
var4: Is there an eligible publication?
var5: URL
var6: DOI
var7: Is there a PMID?
var8: Identification step
var9: Publication type
var10: Matching criteria (ignore this one, see other columns below that are easier to read)
var11: Comment on matching, as entered
var12: Trial IDs (irrelevant for discrepancy corrections)
var13: Trial IDs (irrelevant for discrepancy corrections)
var14: PMID
var15: publication date if PMID was missing
var16: Overall comment as entered

Here follows data management variables that can be of help:

typo_primary_id: Whether there's a typo in the trial ID
publication_date: Publication date, automatically extracted from PubMed or manually entered by reviewer
match_design: Whether reviewer found the design to be matching
match_indication_population: Whether reviewer found the indication/population to be matching
match_intervention: Whether reviewer found the intervention to be matching
match_comparator: Whether reviewer found the comparator to be matching
match_any_outcome: Whether reviewer found the outcome to be matching
n_matching_criteria: Number of matching criteria fulfilled
registry: Name of registry
start_date_recoded: Trial start date, may be missing
completion_date_recoded: Trial end date, rarely missing
warning_cd_to_publication: Flags if publication was before trial completion date
warning_start_to_publication: Flags if publication was before trial start date
form_id: Ignore this one, just a row number

Automatic agreement checks that can be of help:

n_match_agree: Number of matching criteria fulfilled, as rated by both reviewers
article_found_by: Reviewers that found this particular publication for the particular trial
article_agreement_information: How many reviewers found this particular publication and how they rated it regarding matching criteria