

Module 14: Further issues in meta-analysis

This module deals with some important general concepts relevant in all reviews as well as some issues that not all reviewers encounter.

Learning objectives

- Know when and how sensitivity analysis can investigate the robustness of your findings
- Identify types of data not easily analysed using RevMan
- Consider perspectives on dealing with intention-to-treat analyses
- Understand problems associated with indirect comparisons

Relevant sections of the *Cochrane Handbook for Systematic Reviews of Interventions*

- Sections 9.7: Sensitivity analyses
- Chapter 16: Special topics in statistics

Where does this go in a Cochrane review?

- The protocol should outline where a sensitivity analysis might be appropriate. However, sensitivity analyses are the one form of statistical analysis that can (and should) be done *whether or not they were specified in the protocol*. They should be reported in the Results and/or Discussion section, perhaps depending on their implications
- Items in this Module should all be described in the Methods section

Sensitivity analysis

Sensitivity analyses can, and should, be done even if they weren't specified in the protocol

The process of undertaking a systematic review and meta-analysis involves many decisions. Ideally, most of these are made while designing the protocol. It is usually necessary, however, to make further decisions in order to deal with the studies subsequently identified: few reviewers correctly anticipate all problems that arise. Many questions that arise during the review process will have obvious answers. Others will not have clear answers, and in some cases our decisions may change the whole conclusion of the review. The role of a *sensitivity analysis* is to determine whether the assumptions or decisions we have made do in fact have a major effect on the results of the review. You should present your investigations of the effect your assumptions had in the Results section of your review by detailing the range of treatment estimates and confidence intervals resulting from the various sensitivity analyses.

A sensitivity analysis addresses the question 'Are the findings robust to the method used to obtain them?' Sensitivity analyses involve comparing the results of two or more meta-analyses calculated using different assumptions. Here are a few examples of the sorts of things performed as sensitivity analyses in Cochrane reviews:

- If a study is of doubtful eligibility for the systematic review, then comparing meta-analyses excluding and including that study might be undertaken as a sensitivity analysis
- Results may be calculated using all studies and then excluding poorer quality studies
- Both fixed and random effects meta-analyses might be undertaken to assess the robustness of the results to the method used
- If a study appears to be an outlier (has results very different from the rest of the studies) then its influence on a meta-analysis might be assessed by excluding it
- Where missing information is 'imputed' (brought in from another source, perhaps by estimating it) then the effect of imputed numbers should be assessed through sensitivity analysis. This would normally take the form of re-analyzing the data using several alternative imputed values. This is frequently necessary when including cross-over trials, cluster randomized trials or change-from-baseline outcomes in meta-analysis.
- To determine whether a meta-analysis result is being heavily determined by a particular trial it might be repeated excluding that trial. The largest trial or the earliest trial could be driving the result, for example.

We take a more detailed look at one particular type of sensitivity analysis below when we address missing outcome data and intention to treat analyses.

Other types of data

Additional Module 1 gives further information on continuous outcome data

At the beginning of Module 11 we listed a range of different types of outcome data. We dealt with dichotomous data in that module and in Module 12. Continuous data (including long ordinal scales) are addressed in Additional Module 1. Here we address the remaining types of data that you may plan as outcomes, or might come across in your included studies.

Counts of events

You can't treat count data directly as dichotomous data...

Count data are counts of occurrences measured on individuals. Examples include number of lesions, number of pregnancies, number of cigarettes smoked, number of strokes, or number of days in hospital. In Additional Module 2 on unit of analysis issues, you will see that one source of errors in meta-analysis is treating count data directly as dichotomous data. We cannot enter '12 strokes out of 28 people' as dichotomous data if any of the 28 people had more than one stroke.

How you might deal with count data depends on how common they are, in two respects:

- Do most participants have events? (Will most people have at least one?)
- Do participants have lots of events? (Will individual people tend to have high counts?)

Answering each question with a 'yes' or a 'no' gives us four approximate classifications. Two of them have fairly obvious solutions; the other two do not.

...but you can treat them as continuous data...

1. Most people have counts that are mostly high.

An example of this might be number of days on which a person with arthritis has pain, or pulse rate (number of pulses in a given time period). These count data can usually be treated as continuous data, although the distribution may be skewed. See Additional Module 1.

... or dichotomize people to turn them into dichotomous data...

2. Only some people have the event and counts are mostly low.

Admissions to hospital and strokes may come into class. A convenient way to analyze these data is to dichotomize people into those that have *at least one event* and those that have no events. The data can then be entered into RevMan. Alternatively, you might use the data as rate data (see below).

3. Most people have counts that are mostly low. Number of days off work with 'flu might fall into this class. These data can be awkward. They are most like ordinal data, so look in the section below for ideas.
4. Only some people have counts that are mostly high. The number of cigarettes smoked in a smoking cessation trial will likely be of this sort. This is another awkward type of count data. It may be wise to dichotomize people, for example, as smokers or non-smokers. If you are tempted to treat the high counts as continuous data, then remember that if substantial numbers of people have counts of zero, then the distribution of the outcome will be severely skewed.

...or analyze them as rate data



Activity:
Convert the person-months in this risk ratio calculation to person-years.
What happens?

When counts per unit time are of interest, then counts should be treated as *rate* data. A rate is simply a count per unit time, for example, 2 relapses per year. Rate data are extracted for a whole treatment group. To allow for the different times individual participants are followed up for, we come up with a total follow up time for all participants, by adding up the time for each participant. Data corresponding to the rate of 2 relapses per year might be extracted from a placebo group as '1042 relapses during 6247 person-months of follow-up'. These may be compared to '983 relapses during 6229 person-months' in an intervention group. The *rate ratio* from these numbers is

$$\frac{983/6229}{1042/6247} = \frac{0.158}{0.167} = 0.946$$

A standard error for the (log) rate ratio is available, and the Generic Inverse Variance method for meta-analysis may be used to combine rate ratios across studies.

Short ordinal scales

Disease severity is commonly classified as 'none', 'mild', 'moderate' or 'severe'. Many assessment scales have only a few categories, say a score between 1 and 5. We often refer to such data as ordinal data. There are numerous approaches to their analysis. The simplest and usually the best is to try and find a cut-point in the scale and to create a dichotomous outcome. For example, 'none or mild' versus 'moderate or severe'. Sometimes, reviewers present more than one cut point, such as also giving 'none' versus 'mild or moderate or severe' in a sensitivity analysis to investigate whether choice of cutpoint affects conclusions.

You can dichotomize ordinal data...

...or treat them as continuous data...

Another approach is to treat ordinal data as continuous data. Thus we could assign 'none' = 1, 'mild' = 2, 'moderate' = 3 and 'severe' = 4 and take the mean and standard deviation. This is rarely a reasonable approach because it assumes that the assigned numbers represent a real measure of the outcome (i.e. that the difference between mild and moderate is exactly the same as the difference between moderate and severe), when in fact they are arbitrary.

There are more sophisticated methods for analysing these data, which avoid these assumptions, but they are not widely used, and not available in RevMan.

Censored data or survival data

In many situations, health care interventions aim to affect the time until an event happens. For example, we may aim to prolong disease-free survival in cancer, or extend the time to the next fit in epilepsy, or time to heart attack or stroke in people who have just had their first heart attack. The outcome that is measured on each patient in studies of such treatments may be a *time until the event*. When interest is focussed on time to the event rather than simply whether the event happens, we have *survival data*. Although time is a continuous outcome, survival data cannot be analysed in the same way as continuous data because we usually have some patients who have not yet experienced the event by the end of the study. For example, although everybody dies eventually, many patients will survive beyond the end of follow up in a randomized trial. Patients that don't experience the event have survival times that are *censored*. All we know about these patients is that they survived at least until the time when they were last observed.

You can dichotomize survival data...

One way to deal with survival data is to select particular points in time and determine whether each participant had experienced the event by each time. The resulting data are dichotomous and can be analyzed as such. For example, many infectious diseases lead to fever, and one marker of a successful treatment is a reduction in the length of fever. Trials of such treatments tend to assess fever at a specific time point, for example after five days. This avoids the problem of censoring and the analysis is straightforward, as long as you have all the data. In longer-term trials, one might create a series of dichotomous outcomes for, say, mortality such as (i) death within 3 months; (ii) death within 1 year; (iii) death within 3 years, and so on. However, this approach can only be used when all participants have been followed up to or beyond the time point used for the analysis, i.e. all participants have been in the study for at least as long as the time point.

...or analyze them 'properly' using survival data techniques

In many specialties, the tradition is to analyze survival data using special methods that account for censoring. This is especially the case in cancer research. These methods include 'log-rank' tests, and 'proportional hazards regression' (or 'Cox regression'). These results can be used in meta-analyses. If the only data you can obtain are of this sort, then statistical expertise will be needed.

Intention to treat issues

Intention-to-treat: analyze participants in the groups to which they were randomized

Intention-to-treat (ITT) analyses are widely recommended as the preferred approach to the analysis of most clinical trials. Systematic reviewers often wish to practice this recommendation and plan to conduct meta-analyses according to the ITT principle. But what does this mean, and how might it be achieved?

The ITT principle

The basic intention-to-treat principle is that participants in trials should be analysed in the groups to which they were randomized, regardless of whether they received or adhered to the allocated intervention. Two issues are involved here. The first issue is that participants who strayed from the protocol (for example by not adhering to the prescribed intervention, or by being withdrawn from active treatment) should still be kept in the analysis. An extreme variation of this is participants who receive the treatment from the group they were not allocated to, who should be kept in their original group for the analysis. This issue causes no problems provided that, as a systematic reviewer, you can extract the appropriate data from trial reports.

The rationale for this approach is that, in the first instance, we want to estimate the effects of allocating an intervention in practice, not the effects in the subgroup of participants who adhere to it.

The second issue in ITT analyses is the problem of loss to follow-up. People are lost from clinical trials for many reasons. They may die, or move away; they may withdraw themselves or be withdrawn by their clinician, perhaps due to adverse effects of the intervention being studied.

If participants are lost to follow-up then the outcome may not be measured on them. But the strict ITT principle suggests that they should still be included in the analysis. There is an obvious problem – we often do not have the data that we need for these participants. In order to include such participants in an analysis, we must either find out whether outcome data are available for them by contacting the trialists, or we must 'impute' (i.e. make up) their outcomes. This involves making assumptions about outcomes in the 'lost' participants.

There are many ‘formal’ approaches to imputing missing outcomes in clinical trials. A review of these is beyond the scope of this course. We shall look at one particular situation that arises commonly and consider some alternative approaches that might be compared in a sensitivity analysis.

Consider the following trial of a Larium-Qinghaosu combination versus Qinghaosu alone for treating malaria that was included in a Cochrane review. Although 20 were randomized to the former and 18 to the latter, results were available only for the 34 people that did not drop out. These were the findings regarding the presence of parasitic infection after four weeks:

	Clear	Not clear	Total
Larium+Qinghaosu	17	0	17
Qinghaosu	10	7	17

Several methods of filling in missing outcomes can be compared in a sensitivity analysis

There were two other similar trials, also with missing data. In order to perform an ITT analysis, these data needed to be imputed. Four particular strategies for doing this are

1. (Assume the best) assume everybody missing was *clear* of infection
2. (Assume the worst) assume everybody missing was *not clear* of infection
3. (Best-case scenario for combination treatment) assume everybody missing on the combination treatment was *clear* and everybody missing on Qinghaosu was *not clear*
4. (Best-case scenario for single treatment, or worst-case scenario for combination treatment) assume everybody missing on the combination treatment was *not clear* and everybody missing on Qinghaosu was *clear*

Strategies 1 and 2 are attempts to fill in the missing data in a realistic manner. One of these may be an obvious contender for your situation. For example, in clinical trials of smoking-cessation treatments it may be reasonable to assume that dropouts continue to smoke.

Strategies 3 and 4 put bounds on the possible results of the trial had all participants been observed. Fortunately in the malaria review the results were not sensitive to any of these strategies. When dropout rates are higher the results may not be robust and caution may be needed in interpreting the findings.

If your outcome is a continuous measure, imputation of missing data for ITT purposes is more difficult, as there are more than two different possibilities for each participant.

Because imputation of missing data in order to perform a full ITT analysis is controversial, it may be best to present only the results for available participants. If you do this, you should also consider the possible effects of the missing participants, either through sensitivity analyses as described here or by discussing the implications in the Discussion of your review. An alternative approach may be to only analyse the data available, but to consider drop out rate as a marker of trial quality. Whichever approach you use, ensure that it is described in the methods section of the review and that the numbers of participants with missing data are described in the results section and the characteristics of included studies table.

Indirect comparisons

We talked a little about indirect comparisons when we considered subgroup analysis in the module on heterogeneity. We frequently wish to compare interventions across studies. Here are two examples.

- In Module 13 we thought about a collection of trials comparing training of systematic reviewers with no training of systematic reviewers. Some trials used self-directed learning and others used face-to-face training. None compared the two types of training, although we really want to know which one works best.
- A Cochrane review of pharmacological interventions for dyspepsia included many comparisons between different drugs and between drugs and placebo. Among them were comparisons of histamine H2 antagonists (H2RAs) versus placebo, comparisons of prokinetics versus placebo and comparisons of H2RAs versus prokinetics. There were substantially more placebo-controlled comparisons, yet an important clinical question is whether H2RAs are more effective than prokinetics. Could the placebo-controlled studies reliably tell the reviewers more about this comparison?

It is clear that the best way to compare treatments is to seek direct comparisons in randomized trials. Sometimes there aren't enough of these studies to draw reliable conclusions, and sometimes there aren't any such studies at all. Here we address some of the issues involved in making indirect comparisons across different studies.

The first question is whether we should compare the treatment *arms* in the different studies, or compare the treatment *effect estimates* from the different studies.

Suppose we had a single trial of an H2RA versus placebo and a single trial of a prokinetic versus placebo. Can we reliably compare the patients given the H2RA in one trial with the patients given the prokinetic in the other trial? The answer is clearly no. These groups are unlikely to be comparable, because there are likely to be differences in the case-mix, outcome assessments and other aspects of trial design between the trials. Although the groups have been generated by randomisation, the randomisation acted within each trial (H2RA versus placebo, prokinetics versus placebo). Participants were not randomized to be in one trial or the other, i.e. there was no randomisation to prokinetics versus H2RA.

A more reasonable approach is to exploit the randomization within each study, and compare treatment effect estimates. This does not make the comparison randomized, but makes it considerably more reliable since we now relate the effects of the two active drugs to a common reference. For example, if we have one trial comparing Drug A to placebo, and one comparing Drug B to placebo, we could compare the treatment effects measured in the two studies as an indirect way of comparing Drug A to Drug B.

We have determined that indirect comparisons should exploit the comparisons within randomized trials. But when is it reasonable to make indirect comparisons like that suggested above? We need to ask ourselves, 'are the studies sufficiently similar to provide a meaningful result?'

Remember, this comparison between the results of two studies is still a non-randomized comparison. The problem with making indirect comparisons is that there are often systematic differences between the types of trials addressing one intervention and the types of trials addressing the other. For example, trials of self-directed learning for systematic reviewers may be undertaken in resource-limited countries, and trials of face-to-face training in resource-rich countries. There may be reasons, other than the approach to learning, for differences in the effectiveness of training in these two situations.

Methods are available for undertaking indirect comparisons *if it is reasonable to do so*. Meta-regression is one approach. Another approach is to formally compare the estimates and confidence intervals for the direct comparisons. Further information on indirect comparisons can be found in section 16.6 of the *Cochrane Handbook for Systematic Reviews of Interventions*.

If we are going to use indirect comparisons, we need to be very cautious about their interpretation.

Summary

We have covered several issues in this module.

Sensitivity analyses are a way of investigating the importance of some of the assumptions and decisions we make during a systematic review.

We have briefly covered some types of data encountered during reviews, and looked at the difficult issue of the value of indirect comparisons.

In the next modules, we start to look at some important issues in the interpretation of the data we have collected.