

Additional Module 2: Issues related to the unit of analysis

In this module we look at issues related to the unit of analysis, including the incorporation of crossover trials and cluster randomized trials into a Cochrane review, or meta-analysis. The main aim of the module is to help you to recognize when unit of analysis errors can occur, rather than to learn how to deal with every eventuality. Many of the methods for dealing with unit of analysis errors are tricky, and if you do have to deal with these issues, you should ensure that you have access to appropriate methodological expertise.

Learning objectives

- Identify a variety of sources of unit-of-analysis errors
- Identify a crossover trial and understand when it is an appropriate design to use
- Understand that the correct analysis of a crossover trial is a paired analysis, in which within-patient differences are the focus
- Identify a cluster randomized trial
- Understand that the appropriate analysis of a cluster randomized trial should recognize that individuals are members of clusters
- Identify repeated outcome measurements, and be aware of strategies for incorporating them into meta-analyses
- Be aware that the inverse variance method for meta-analysis can be used for crossover, cluster randomized trials and trials with other designs

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

- Section 9.3: Study designs and identifying the unit of analysis
- Section 16.3: Cluster-randomized trials
- Section 16.4: Cross-over trials

Where does this go in a Cochrane review?

- Think carefully about possible unit of analysis issues when writing your protocol. Methods for dealing with them should be described in the Methods section

For starters

The simplest trial design to analyse is called a **parallel group** design, where participants are randomized to receive one of two interventions, and outcome measurements are made at one time point at the end of the treatment period.

In this module we will look at some other types of trial design, which depart from the simple trial design above. These departures lead, among other problems, to mismatches between numbers of randomized units and the number of included participants (such as twenty participants but forty eyes) or outcome observations. These can lead to confusion, and errors. We call these errors ‘unit of analysis errors’, and the issues discussed in this module are generally known as ‘unit of analysis issues’.

To determine which issues are relevant to your review, consider whether you have, or anticipate, any trials of the following sort:



*Activity:
Think about your review and list as many potential unit-of-analysis issues as you think are likely to occur*

- Crossover trials, in which patients are randomized to a sequence of interventions. In crossover trials, all patients might receive both (or more, if there are more than two) interventions.
- Cluster randomized trials, in which groups of participants are randomized.
- ‘Body-part randomization’ designs, in which different parts of a person are randomized to be treated differently. For example, a split-mouth design in dentistry might randomize the left and right halves of the mouth to different topical applications. Or, a physiotherapy trial might randomize two limbs of the same individual to different types of exercise.

Four other related topics can occur with standard parallel group designs:

- ‘Body part analyses’ of standard trial designs, in which a person is randomized to an intervention, but outcomes are measured for several different body parts. For example, in dentistry, a person might be randomized to use a particular toothpaste, but outcomes are collected on every tooth (i.e., cavity: yes/no for each tooth). Or eczema severity may be measured separately on both arms, even though they both received the same treatment.
- Trials with three or more treatment groups. An experimental intervention may be compared with both a standard intervention and with a placebo. Alternatively, two or more experimental interventions might be tested against a standard intervention
- Trials with the same outcome assessed at several time points

- Trials where an event outcome may occur more than once in each participant. A common example is the number of adverse events – participants may have more than one adverse event. Another example, in sub-fertility studies, a woman might have more than one pregnancy during a period of treatment. Or, in cardiovascular studies, a patient might experience more than one stroke

Crossover trials

In a crossover trial, participants are randomized to a sequence of treatments. We shall consider the simplest design, in which there are two treatments. As an example, a trial was conducted in patients with asthma to compare laser acupuncture with a sham ('placebo') procedure. Patients were randomized either to acupuncture for five weeks (first period) then sham for five weeks (second period), or sham then acupuncture. One of the outcomes was a symptom score.

This gives us data for each patient both when they were on acupuncture, and when they were on sham treatment. We can compare these for each patient to assess the effect of acupuncture *within* each patient. This is a very efficient approach to analysis, because when making the comparison between treatment and control we do not have to allow for all the variation that occurs *between* patients, which we have to deal with in a parallel group trial.

In practice this means that, for the same number of participants, a crossover design is likely to be more powerful. However, crossover trials are not always appropriate. The most important consideration is whether the patients start the second period in a similar state to how they started the first period. If the characteristics of the participant have changed in some way by the time the second period starts, then the comparison of treatments is not fair, and there will be within patient variation. Some questions you should ask yourself are as follows.

- Is the condition of the patients chronic and stable? Crossover trials are common for conditions such as asthma, osteoarthritis and epilepsy. Crossover trials may not be appropriate for progressive diseases or acute conditions that will worsen or improve by the second period. If patients vary from one period to the other there is said to be a 'period effect'.
- Does the intervention provide temporary relief, and not permanent change? For example, surgical interventions are unsuitable for crossover trials if the surgery permanently alters the condition.

- Can the outcome be repeated in the second period if it occurs in the first? For example, crossover trials are certainly unsuitable when the primary outcome is mortality, or pregnancy in infertility studies.
- Might the effect of the first intervention last into the second treatment period? In the acupuncture trial, a three-week 'washout' period was built into the trial between the two treatment periods. This is a common method to minimize 'carry-over' effects and ensure the participant is in the same state at the beginning of each period, though it is not always sufficient.
- Does the trial go on long enough for drugs to have effects and outcomes to occur? For example, a trial in epilepsy with the outcome of number of fits, needs to observe the patients for long enough to make sure that we haven't, by chance, just picked a particularly good or bad time in their illness.

If you have crossover trials in your review, you will need to decide on the following points:

- Should I include crossover trials in the review?
- Should I include crossover trials in any meta-analyses?
- Should I combine crossover trials with other types of trials?
- How should I include crossover trials in a meta-analysis?

The brief answer to the first three questions is:

- There is no reason to exclude crossover trials solely because they are crossover trials.

Of course, there may be other reasons why crossover trials might be excluded, for example, if they have treatment periods that are too short or do not have an appropriate wash out period. If you are anticipating cross over studies in your review you will need to set these inclusion criteria in your protocol.



Read more about crossover trials in Section 16.4 of the *Cochrane Handbook for Systematic Reviews of Interventions*

Analysis of crossover trials should exploit the fact that each patient acts as his or her own control. The appropriate analysis is a 'paired' analysis. See Section 16.4.5 of the *Cochrane Handbook for Systematic Reviews of Interventions* for a more extensive discussion of how you can include crossover trials in a Cochrane review or a meta-analysis. Properly analysed crossover trials may be incorporated into meta-analyses using the Generic Inverse Variance method in RevMan.

Cluster randomized trials

A cluster randomized trial is a trial in which individuals are randomized in groups (i.e. the group is randomized, not the individual). For example, in a rural area with an endemic disease, we might randomise whole villages to have the intervention or not, rather than individual people. We then say that the village is the unit of randomization. In other situations, general practices, hospitals, families or school classrooms may be randomized. Reasons for performing cluster randomized trials vary. Sometimes the intervention can only be administered to the group, for example an addition to the water supply or a public education campaign; sometimes the motivation is to avoid contamination (all participants in the trial are affected by the intervention, even if it is only given directly to some of them); sometimes the design is simply more convenient or economical.

A simple approach to dealing with cluster randomized trials is to assess outcomes only at the level of the group thereby keeping the unit of analysis the same as the unit of randomisation. One might measure a dichotomous outcome of whether the practice/classroom/village was a 'success' or a 'failure', or a continuous outcome of the percentage of individuals in the group who benefited. In this way, we obtain one outcome measurement from each randomized unit, and the analysis can proceed as if the groups were individuals – that is, using the techniques described elsewhere in Modules 11 and 12. It will probably strike you that there are limitations to this approach. First, cluster randomized trials are likely to randomize fewer groups than most simple trials randomize individuals. For example, a trial might randomise ten villages with a total of 15,000 inhabitants. Analysing by village, we would end up with only ten observations. So, we would end up with fewer data (and hence less statistical power) than a simple trial involving substantially fewer participants analysed as individuals. Second, not all groups will be the same size, and we would give the same weight to a village of 10,000 inhabitants as a village of 150 inhabitants.

An alternative possibility is to ignore the groupings and compare all the individuals in intervention groups with all those in control groups. This has been a common approach both to analysing individual cluster randomized trials and to representing them in systematic reviews. But it is problematic because it ignores the fact that individuals within a particular group tend to be more similar to each other than to members of other groups. Such analyses can spuriously overestimate the significance of differences, and should be avoided.

Think of the example where we randomize villages. Residents in one village may share the same climate, nutrition, education and health care, which make their outcomes more similar to each other than to residents in a different village. We use the term *intra-cluster correlation coefficient* to describe the extent to which two members of one cluster are more similar than two people from different clusters.



Read the section on cluster randomized trials in Section 16.3 of the *Cochrane Handbook for Systematic Reviews of Interventions*

There are statistical techniques for appropriate analyses of cluster randomized trials. We can recognize that clusters are made up of individuals and that there may be more individuals in one cluster than in another. The intra-cluster correlation coefficient plays an important role in these techniques. Further details can be found in Section 16.3.4 of the *Cochrane Handbook for Systematic Reviews of Interventions*. Cluster randomized trials may be incorporated into meta-analyses using the Generic Inverse Variance method.

‘Body-part randomization’ and ‘body-part analysis’

We use the terms ‘body-part randomization’ and ‘body-part analysis’ to distinguish between two different types of study design involving parts of the body.

By ‘body-part randomization’ designs we mean those in which similar body parts are randomized to different interventions. For example, a person’s arms may be randomized so that each gets a different cream applied. Other examples include studies for eyes and teeth. One issue to think about is contamination – could a treatment in one part affect what happens in another? If so this raises the question of whether such designs are appropriate in the first place. You may notice a similarity between this design and the crossover design described above. In both designs, each person receives both interventions. In fact, the analysis of a body-part randomization trial should proceed in the same way as the analysis of a crossover trial, involving a paired analysis.

By ‘body-part analysis’ we mean a particular approach to the analysis of a standard parallel group design trial. Suppose a (whole) individual is randomized to receive a surgical intervention for cataracts. If he or she has cataracts in both eyes you might collect outcomes for the vision in each eye separately, and you might want the patient to therefore contribute two measurements to the data analysis. This is rather like a cluster randomized trial, where the person is the cluster and the eyes are the individuals within the cluster. In fact, the analysis of these types of trials should proceed in the same way as the analysis of a cluster randomized trial. However, if there are only one or two measurements for each individual, it may be preferable and simpler to select only one measurement per individual.

More than two treatment groups

Many trials are designed to compare more than two treatments. However, there may be two or more experimental interventions, for example a drug at different doses, or variations on a counseling intervention. Alternatively, there may be more than one control group, perhaps an established treatment and a placebo.

The most common problem reviewers experience is trials with several experimental groups. If you are comparing each of the treatment groups with placebo in two separate meta-analyses (i.e. as two separate comparisons in RevMan), then the study can be treated as two separate trials (intervention 1 versus placebo and intervention 2 versus placebo). If however, you are putting all three arms of the study into the same meta-analysis it can be tempting to, for example, enter the data as if it were two trials, one comparing high dose with placebo and one comparing low dose with placebo. But when you then pool these results in a meta-analysis you will be counting the placebo patients twice. This approach should not be followed.

There are two main approaches to dealing with trials like this. The first is to break up the control group into several parts, so that the total numbers add up to the original size of the group. The second is to group together all the experimental groups and compare them collectively with the control group. There's no single right answer, since both approaches have advantages and disadvantages.

Repeated measurements

Repeated measurements refer to measurements made at different points in time. Thus, a dietary trial may report weight loss at 4 weeks, at 8 weeks and at 6 months. We can't include all of these in the same meta-analysis since again we'd be counting the same person more than once and we'd have a unit-of-analysis error. The problem of repeated measurements can partly be overcome by specifying in the protocol which time-points are of interest, and discarding the rest. It may be helpful to classify outcomes as 'short-term', 'medium-term' and 'long-term', and to perform separate meta-analyses for these different outcomes including only one time-point from each trial in each analysis. Alternatively, you may only be interested in a single time-point, or the longest available follow-up. The disadvantages of opting for longest available follow-up are that more patients may have been lost to follow-up, and it may vary considerably between studies introducing heterogeneity.

Who is having the events?

Our final topic here can cause confusion. Consider an outcome that is an event: a stroke, for example. What if the event can happen more than once? A person can have more than one stroke during a period of follow-up. We have to be very careful that we know exactly the nature of the data being reported. For an outcome to be considered as a dichotomous outcome there must be an all-or-nothing distinction between the “yes” and “no” classifications. Thus, if we count that 245 people had *at least one stroke* and 765 people had *no strokes*, we have dichotomous data. If, on the other hand, we just know that there were 312 strokes among 1010 people then we may not know who had them. If some people could have had more than one of the strokes we do *not have dichotomous data*. We cannot treat the data as dichotomous data when entering them into RevMan, unless we can distinguish the number of people having events from the number of events.

The same difficulty occurs when we are analysing adverse events. Suppose we know that there were 120 adverse events among 250 participants in a trial. It may be that 20 people had two events each, accounting for 40 of the 120 events. If the other 80 adverse events occurred as one in each of 80 people, that would mean a total of 100 people had the 120 events. Unless we are told about how often people had different numbers of events, we cannot calculate the number of participants who had *any* adverse event.



Read 'Effect measures for counts and rates' in Section 9.2.5 of the *Cochrane Handbook for Systematic Reviews of Interventions*

So what can we do? If you think this poses a problem, consult Module 14 or section 9.2.5: Effect measures for counts and rates of the *Cochrane Handbook for Systematic Reviews of Interventions*.