

Module 12: Combining studies

In the previous module we discussed choosing an effect measure for presenting the results of studies reporting dichotomous data within your review. If any of the outcomes you are planning to analyse are continuous, you should complete the additional module A1, either before or immediately after completing this module.

This module is about meta-analysis. We will look at the principles for the sensible combination of results from separate studies, and overview the methods that are most commonly used in Cochrane reviews.

Learning objectives

- Understand what a weighted average is, and how it differs from a straightforward average
- Understand the concept of standard error relating to individual trial results
- Be aware that there are alternative methods for meta-analysis (Peto, Mantel-Haenszel) which differ in the way that study weights are calculated
- Be aware that there are a group of methods known as random effects methods which take a different approach to fixed effect meta-analysis

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

- Section 9.4: Summarizing effects across studies

Where does this go in a Cochrane review?

The information in this module will be relevant to many parts of your review:

- In the data analysis part of the Methods section of a protocol or review, where you will describe what statistical techniques you are planning to use
- When (or if) you actually perform meta-analyses using the analysis part of RevMan or other software
- In the presentation of results in the Results section of the review
- In the interpretation of results, in the Discussion of your review

What is meta-analysis?

Meta-analysis combines the results of several studies

Meta-analysis is the use of statistical methods to combine results of individual studies. This allows us to make the best use of all the information we have gathered in our systematic review by increasing the power of the analysis. By statistically combining the results of similar studies we can improve the precision of our estimates of treatment effect, and assess whether treatment effects are similar in similar situations. The decision about whether or not the results of individual studies are similar enough to be combined in a meta-analysis is essential to the validity of the result, and will be covered in the next module on heterogeneity. In this module we will look at the process of combining studies and outline the various methods available.

There are many approaches to meta-analysis. We have discussed already that meta-analysis is not simply a matter of adding up numbers of participants across studies (although unfortunately some non-Cochrane reviews do this). This is the ‘pooling participants’ or ‘treat-as-one-trial’ method and we will discuss it in a little more detail now.

Pooling participants (not a valid approach to meta-analysis).

This method effectively considers the participants in all the studies as if they were part of one big study. Suppose the studies are randomised controlled trials: we could look at everyone who received the experimental intervention by adding up the experimental group events and sample sizes and compare them with everyone who received the control intervention. This is a tempting way to ‘pool results’, but let’s demonstrate how it can produce the wrong answer.

A Cochrane review of trials of daycare for pre-school children included the following two trials. For this example we will focus on the outcome of whether a child was retained in the same class after a period in either a daycare treatment group or a non-daycare control group. In the first trial (Gray 1970), the risk difference is -0.16 , so daycare looks promising:

Gray 1970	Retained	Total	Risk	<i>Risk difference</i>
Daycare	19	36	0.528	-0.16
Control	13	19	0.684	

In the second trial (Schweinhart 1993) the absolute risk of being retained in the same class is considerably lower, but the risk difference, while small, still lies on the side of a benefit of daycare:

Schweinhart 1993	Retained	Total	Risk	<i>Risk difference</i>
Daycare	6	58	0.1034	-0.004
Control	7	65	0.1077	

What would happen if we pooled all the children as if they were part of a single trial?

Pooled results	Retained	Total	Risk	<i>Risk difference</i>
Daycare	25	94	0.266	+0.03
Control	20	84	0.238	WRONG!

It suddenly looks as if daycare may be *harmful*: the risk difference is now bigger than 0. This is called Simpson's paradox (or bias), and is why we don't pool participants directly across studies. The first rule of meta-analysis is to keep participants within each study grouped together, so as to preserve the effects of randomisation and compare like with like. Therefore, we must take the comparison of risks *within* each of the two trials and somehow combine these. In practice, this means we need to calculate a single measure of treatment effect from each study before contemplating meta-analysis. For example, for a dichotomous outcome (like being retained in the same class) we calculate a risk ratio, the risk difference or the odds ratio for each study separately, then pool these estimates of effect across the studies.

We don't add up patients across trials

Simple average of treatment effects (not used in Cochrane reviews)

If we obtain a treatment effect separately from each study, what do we do with them in the meta-analysis? How about taking the average? The average of the risk differences in the two trials above is $(-0.004 - 0.16) / 2 = -0.082$. This may seem fair at first, but the second trial randomised more than twice as many children as the first, so the contribution of each randomised child in the first trial is diminished. It is not uncommon for a meta-analysis to contain trials of vastly different sizes. To give each one the same influence cannot be reasonable. So we need a better method than a simple average.

We don't use simple averages to calculate a meta-analysis

Weighted averages

The solution is to calculate a weighted average.

A weighted average is an average where the results of some of the studies make a greater contribution to the total than others. All of the methods available for conducting meta-analyses in Cochrane reviews use forms of weighted averages. The various methods do this in different ways, and we will cover these methods in this module. In all methods, the underlying principle is to give more weight to studies that give us more information about the treatment effect.

Sample size is the main factor in determining the weight for a trial. However, the event rate also makes a difference. This is because effects are generally estimated more precisely when there are lots of events. So, trials with higher event rates get more weight. At the extreme, a trial with no events tells us nothing about the effect of the intervention, and so gets no weight at all. The exact relationship between event rates and study weights is complex, and depends on the summary statistic being used.

A statistical concept which takes into account both size of the study's population, and its event rate, is *variance*. The box below provides an outline of the concept of variance.

What is
variance?

Standard errors, confidence intervals and variances

Imagine trying to estimate the proportion of females in the population. If we took a sample of ten people from a list of all the people in a country, we may, by chance, find there were 2, or 7, or even 10 females. We wouldn't be very confident from this sample to say what the true proportion of females in the population is. Whenever we take samples from populations, there is uncertainty about the estimates we make of the true value in the whole population.

The same is true of trials – each trial involves taking a sample of the possible participants. The basic result of an individual trial is an estimate of treatment effect. The estimate is incomplete without a measure of how certain we can be about it. We'd be much more certain about an estimate from a mega-trial of tens of thousands of patients than we would an estimate from a small trial of less than a hundred. Uncertainty is often described using a *confidence interval*. For example, a 95% confidence interval gives a range within which we can be 95% confident the true effect lies.

Continued.....

Confidence intervals are calculated from a number known as a *standard error*. Standard errors are companions of all estimates. They describe the extent to which an estimate might be wrong due to random error. The smaller the standard error the more certain we are about the estimate. To get a feel for standard errors it is helpful to know that 95% confidence intervals are obtained by taking the estimate and creating limits that are 1.96 standard errors below it and 1.96 standard errors above it. Thus an estimate may be wrong by about a standard error, but to be 95% confident about where the true effect lies, we go roughly 2 standard errors either side.

Statistics as a discipline has more uses for the ‘standard error squared’ than for the ‘standard error’, so statisticians have a word for it, the variance. The *variance* of an estimate is just the square of its standard error. The standard error and variance are interchangeable in terms of the information they convey, but their numerical values are different. It’s the same as describing the size of a square: you could say either that the length of each side is 4 metres or that the area is 16 square metres; you end up with an identical shape. Quoting the length is like using the standard error; quoting the area is like using the variance.

One important point to note is that different treatment effects (OR, RR, RD) calculated for the same trial will have different variances.

We could *assume* that variance is inversely proportional to importance, i.e. the less variance in the study, the more weight it should contribute. The Inverse Variance method in RevMan, calculates study weights directly based on this assumption.

There are other methods, called Mantel-Haenszel methods, which attribute weight in a manner closely related to inverse variance. In this module we will expand a little on the various available methods and look at some of the differences between them, finishing with some guidance on which method to use in your review.

Within RevMan, the methods available are:

Type of data	Effect measure	Fixed-effect methods	Random-effects methods
Dichotomous	Odds ratio (OR)	Mantel-Haenszel (M-H)	Mantel-Haenszel (M-H)
		Inverse variance (IV)	Inverse variance (IV)
	Peto		
Risk ratio (RR)	Mantel-Haenszel (M-H)	Mantel-Haenszel (M-H)	
	Inverse variance (IV)	Inverse variance (IV)	
Risk difference (RD)	Mantel-Haenszel (M-H)	Mantel-Haenszel (M-H)	
		Inverse variance (IV)	Inverse variance (IV)
Continuous	Mean difference (MD)	Inverse variance (IV)	Inverse variance (IV)
	Standardized mean difference (SMD)	Inverse variance (IV)	Inverse variance (IV)
O – E and Variance	User-specified (default 'Peto odds ratio')	Peto	None
Generic inverse variance	User-specified	Inverse variance (IV)	Inverse variance (IV)
Other data	User-specified	None	None

Information about the statistical techniques available in RevMan is addressed in section 9.4 of the *Cochrane Handbook for Systematic Reviews of Interventions* and you should read it now.

In order to choose the method you are going to use in your meta-analysis, the first concept to understand is the difference between a fixed effect model and a random effects model.



Read [section 9.4](#) of the *Cochrane Handbook for Systematic Reviews of Interventions*

What does 'fixed effect' mean?

To come up with any statistical model, or method for meta-analysis, we first need to make some assumptions. It is these assumptions that form the differences between all the methods listed above.

A fixed effect model of meta-analysis is based on a mathematical assumption that every study is evaluating a common treatment effect. That means the effect of treatment, allowing for the play of chance, was the same in all studies. Another way of explaining this is to imagine that if all the studies were infinitely large they'd give identical results.

The summary treatment effect estimate resulting from this method of meta-analysis is this one ‘true’ or ‘fixed’ treatment effect, and the confidence interval describes how uncertain we are about the estimate.

Sometimes this underlying assumption of a fixed effect meta-analysis (i.e. that diverse studies can be estimating a single effect) is too simplistic. Therefore, the alternative approaches to meta-analysis are (i) to try to explain the variation or (ii) to use a random effects model.

Random effects meta-analyses (DerSimonian and Laird)

As we discussed above, fixed effect meta-analysis assumes that there is one identical true treatment effect common to every study.

The random effects model of meta-analysis is an alternative approach to meta-analysis that does not assume that a common (‘fixed’) treatment effect exists. The random effects model assumes that the true treatment effects in the individual studies may be different *from* each other. That means there is no single number to estimate in the meta-analysis, but a distribution of numbers. The most common random effects model also assumes that these different true effects are normally distributed. The meta-analysis therefore estimates the *mean* and *standard deviation* of the different effects.

By selecting ‘random effects’ in the analysis part of RevMan you can calculate an odds ratio, risk ratio or a risk difference based on this approach.

The Mantel-Haenszel approach

The Mantel-Haenszel approach was developed by Mantel and Haenszel over 40 years ago to analyse odds ratios, and has been extended by others to analyse risk ratios and risk differences. It is unnecessary to understand all the details, but is sufficient to say that the Mantel-Haenszel method assumes a fixed effect and combines studies using a method similar to inverse variance approaches to determine the weight given to each study.

The Peto method

The Peto method works for odds ratios only. Focus is placed on the observed number of events in the experimental intervention. We call this O for ‘observed’ number of events, and compare this with E , the ‘expected’ number of events. Hence an alternative name for this method is the ‘ $O - E$ ’ method. The expected number is calculated using the overall event rate in both the experimental and control groups. Because of the way the Peto method calculates odds ratios, it is appropriate when trials have roughly equal number of participants in each group and treatment effects are small. Indeed, it was developed for use in mega-trials in cancer and heart disease where small effects are likely, yet very important.

The Peto method is better than the other approaches at estimating odds ratios when there are lots of trials with no events in one or both arms. It is the best method to use with rare outcomes of this type.

The Peto method is generally less useful in Cochrane reviews, where trials are often small and some treatment effects may be large.

Which method should I use in my review?

We have talked about three methods of combining the results of trials included in a meta-analysis, the fixed effect method (the Mantel-Haenszel approach, which is more specific to dichotomous data and weights studies in a slightly different way), the Peto method which is useful for rare events, and the random effects model, which assumes that all studies are estimating their own true effect, and these effects are normally distributed. The analysis program within RevMan allows us to choose which of these models we want to use in our meta-analysis, and the results of our review will be slightly different depending on which method we use. The table below summarises the summary effect and confidence intervals resulting from selecting each of these methods in the daycare example we used earlier.

<i>Method</i>	<i>RR</i> (95% CI)	<i>OR</i> (95% CI)	<i>RD</i> (95% CI)
Mantel-Haenszel	0.64 (0.49, 0.82)	0.47 (0.31, 0.73)	-0.14 (-0.23, -0.06)
Peto Method		0.47 (0.30, 0.72)	
Random effect inverse-variance	0.64 (0.50, 0.82)	0.47 (0.31, 0.72)	-0.13 (-0.25, -0.01)

As you can see from the table, there is little difference in the results regardless of which method you choose, and the conclusions of your review would certainly not change. So why do we devote so much energy to selecting the summary statistic and devising methods with varying assumptions? There are some cases or circumstances where one method performs better than the others, and if any of these circumstances fit your review you may need to think carefully about which statistic you use.

Some general points about the performance of the various statistics

- i. The Mantel-Haenszel methods have been shown to be more reliable when there are not many data (small trials and not many of them). This is why they have been selected as the principle method of meta-analysis in the Cochrane Collaboration. This method (which can be used for OR, RR and RD) is the most appropriate for many Cochrane reviews, and many Cochrane review groups use it as standard. But it should not be used in reviews with sparse data, where lots of trials have zero events in treatment or control groups or both. The choice between OR, RR or RD should be based on the information covered in Module 11.
- ii. The Peto method performs well with sparse data and is then the best choice, but when events are common there is usually no preference to use it over the other methods. It is **not** a good idea to use the Peto method when the treatment effect is very large, as the result may be misleading. This method is also unsuitable if there are large imbalances in the size of groups within trials.
- iii. A random effects model may be better when there is statistical heterogeneity between the studies in your review (we will discuss this further in Module 13 on Heterogeneity).

Summary

From this module and the one preceding it we can see that there are many choices the reviewer has to make about the way they analyse their dichotomous data in a review. Firstly decisions about which *within trial* statistic to use (OR, RR or RD) need to be made, and then the method of combining the trial data, or *meta-analysis*, needs to be chosen. While there is not consensus about which is the best approach, and you will need to check your review group policy, following the principles set out here should help you make your decision. If you are concerned that your review falls into one of the categories where there are special considerations (rare events with zero cells, very large treatment effects, or large variation in the control event rates between the trials included in your review) you may want to seek the advice of a statistician, and your review group can help you with this.