A brief review on methods of meta-analysis

By:

Dr. Sepideh Soheilifar

Planning the analysis

- Cochrane reviews contain analyses of the primary studies. Analyses may be narrative, such as a structured summary and discussion of the studies' characteristics and findings, or quantitative, that is involving statistical analysis. **Meta-analysis** – the statistical combination of results from two or more separate studies – is the most commonly used statistical technique. Cochrane review writing software (RevMan) can perform a variety of meta-analyses, but it must be stressed that meta-analysis is not appropriate in all Cochrane reviews.

Planning the analysis

- The contrast between the outcomes of two groups treated differently is known as the 'effect', the 'treatment effect' or the 'intervention effect'. Whether analysis of included studies is narrative or quantitative, a general framework for synthesis may be provided by considering four questions:

1. What is the direction of effect?
2. What is the size of effect?
3. Is the effect consistent across studies?
4. What is the strength of evidence for the effect?

- Meta-analysis provides a statistical method for questions 1 to 3. Assessment of question 4 relies additionally on judgements based on assessments of study design and risk of bias, as well as statistical measures of uncertainty.

Why perform a meta-analysis?

- **To increase power**. Power is the chance of detecting a real effect as statistically significant if it exists.
- **To improve precision**. The estimation of an intervention effect can be improved when it is based on more information.
- **To answer questions not posed by the individual studies**. Primary studies often involve a specific type of patient and explicitly defined interventions. A selection of studies in which these characteristics differ can allow investigation of the consistency of effect and, if relevant, allow reasons for differences in effect estimates to be investigated.
- **To settle controversies arising from apparently conflicting studies or to generate new hypotheses**. Statistical analysis of findings allows the degree of conflict to be formally assessed, and reasons for different results to be explored and quantified.

When not to use meta-analysis in a review

- A common criticism of meta-analyses is that they 'combine apples with oranges'. If studies are clinically diverse then a meta-analysis may be meaningless, and genuine differences in effects may be obscured.
- Meta-analyses of studies that are at risk of bias may be seriously misleading. If bias is present in each (or some) of the individual studies, meta-analysis will simply compound the errors, and produce a 'wrong' result that may be interpreted as having more credibility.
- Finally, meta-analyses in the presence of serious publication and/or reporting biases are likely to produce an inappropriate summary.

Types of data

- **dichotomous (or binary) data**, where each individual's outcome is one of only two possible categorical responses;
- **continuous data**, where each individual's outcome is a measurement of a numerical quantity;
- **ordinal data (including measurement scales)**, where the outcome is one of several ordered categories, or generated by scoring and summing categorical responses;
- **counts and rates** calculated from counting the number of events that each individual experiences; and
- **time-to-event (typically survival) data** that analyse the time until an event occurs, but where not all individuals in the study experience the event (censored data).

Effect measures for dichotomous outcomes

Dichotomous data

- Dichotomous (binary) outcome data arise when the outcome for every participant is one of two possibilities, for example, dead or alive, or clinical improvement or no clinical improvement. The most commonly encountered effect measures used in clinical trials with dichotomous data are:
- the risk ratio (RR) (also called the relative risk);
- the odds ratio (OR);
- the risk difference (RD) (also called the absolute risk reduction); and
- the number needed to treat (NNT).

Risk

- **Risk** is the concept more familiar to patients and health professionals. Risk describes the probability with which a health outcome (usually an adverse event) will occur. In research, risk is commonly expressed as a decimal number between 0 and 1, although it is occasionally converted into a percentage. In 'Summary of findings' tables in Cochrane reviews, it is often expressed as a number of individuals per 1000 .

Odds

- **Odds** is a concept that is more familiar to gamblers. The odds is the ratio of the probability that a particular event will occur to the probability that it will not occur, and can be any number between zero and infinity. In gambling, the odds describes the ratio of the size of the potential winnings to the gambling stake; in health care it is the ratio of the number of people with the event to the number without

$$\text{odds} = \frac{\text{risk}}{1 - \text{risk}}$$

$$\text{risk} = \frac{\text{odds}}{1 + \text{odds}}$$

Odds and risk

- The interpretation of an odds is more complicated than for a risk. The simplest way to ensure that the interpretation is correct is to first convert the odds into a risk. For example, when the odds are 1:10, or 0.1, the risk of the event is 0.1/(1+0.1) = 0.091.
- The difference between odds and risk is small when the event is rare (as illustrated in the first example above where a risk of 0.091 was seen to be similar to an odds of 0.1). When events are common, as is often the case in clinical trials, the differences between odds and risks are large. For example, a risk of 0.5 is equivalent to an odds of 1; and a risk of 0.95 is equivalent to odds of 19.

Measures of relative effect: the risk ratio and odds ratio

- Measures of relative effect express the outcome in one group relative to that in the other. The **risk ratio** (or relative risk) is the ratio of the risk of an event in the two groups, whereas the **odds ratio** is the ratio of the odds of an event

Box 9.2.a: Calculation of risk ratio (RR), odds ratio (OR) and risk difference (RD)

The results of a clinical trial can be displayed as a 2×2 table:

	Event ('Success')	No event ('Fail')	Total
Experimental intervention	S_E	F_E	N_E
Control intervention	S_C	F_C	N_C

where S_E, S_C, F_E and F_C are the numbers of participants with each outcome ('S' or 'F') in each group ('E' or 'C'). The following summary statistics can be calculated:

$$RR = \frac{\text{risk of event in experimental group}}{\text{risk of event in control group}} = \frac{S_E/N_E}{S_C/N_C}$$

$$OR = \frac{\text{odds of event in experimental group}}{\text{odds of event in control group}} = \frac{S_E/F_E}{S_C/F_C} = \frac{S_E F_C}{F_E S_C}$$

$$RD = \text{risk of event in experimental group} - \text{risk of event in control group}$$

$$= \frac{S_E}{N_E} - \frac{S_C}{N_C}$$

Measure of absolute effect: the risk difference

- The **risk difference** is the difference between the observed risks (proportions of individuals with the outcome of interest) in the two groups. The risk difference can be calculated for any study, even when there are no events in either group. The risk difference is straightforward to interpret: it describes the actual difference in the observed risk of events between experimental and control interventions; for an individual it describes the estimated difference in the probability of experiencing the event. However, the clinical importance of a risk difference may depend on the underlying risk of events.

Number needed to treat

- The **number needed to treat** (NNT) is defined as the expected number of people who need to receive the experimental rather than the comparator intervention for one additional person to either incur or avoid an event in a given time frame. Thus, for example, an NNT of 10 can be interpreted as 'it is expected that one additional (or less) person will incur an event for every 10 participants receiving the experimental intervention rather than control over a given time frame'. It is important to be clear that since the NNT is derived from the risk difference, it is still a ***comparative*** measure of effect (experimental versus a certain control) and not a general property of a single intervention; and the NNT gives an 'expected value'. For example, NNT = 10 does not imply that one additional event *will* occur in each and every group of ten people.

Number need to treat

- The number needed to treat is obtained from the risk difference. Although it is often used to summarize results of clinical trials, NNTs cannot be combined in a meta-analysis. However, odds ratios, risk ratios and risk differences may be usefully converted to NNTs and used when interpreting the results of a meta-analysis.

Computing NNT from a risk difference (RD)

- NNTs can be calculated for single studies as follows. Note that this approach, although applicable, should only very rarely be used for the results of a meta-analysis of risk differences, because meta-analyses should usually be undertaken using a relative measure of effect (RR or OR).

$$\mathrm{NNT} = \frac{1}{\text{absolute value of risk difference}} = \frac{1}{|\mathrm{RD}|}$$

Computing absolute risk reduction or NNT from a risk ratio (RR)

- To aid interpretation, review authors may wish to compute an absolute risk reduction or NNT from the results of a meta-analysis of risk ratios. In order to do this, an assumed control risk (ACR) is required. It will usually be appropriate to do this for a range of different ACRs. The computation proceeds as follows:

$$NNT = \left| \frac{1}{ACR \times (1 - RR)} \right|$$

Computing absolute risk reduction or NNT from an odds ratio (OR)

- Review authors may wish to compute an absolute risk reduction or NNT from the results of a meta-analysis of odds ratios. In order to do this, an assumed control risk (ACR) is required. It will usually be appropriate to do this for a range of different ACRs. The computation proceeds as follows:

$$NNT = \frac{1}{\left| ACR - \frac{OR \times ACR}{1 - ACR + OR \times ACR} \right|}$$

Effect measures for continuous outcomes

Continuous data

- The term ‘continuous’ in statistics conventionally refers to data that can take any value in a specified range. When dealing with numerical data, this means that any number may be measured and reported to arbitrarily many decimal places. Examples of truly continuous data are weight, area and volume. In practice, in Cochrane reviews we can use the same statistical methods for other types of data, most commonly measurement scales and counts of large numbers of events.
- Two summary statistics are commonly used for meta-analysis of continuous data: the mean difference and the standardized mean difference. These can be calculated whether the data from each individual are single assessments or change from baseline measures.

The mean difference (or difference in means)

- The **mean difference** (more correctly, 'difference in means') is a standard statistic that measures the absolute difference between the mean value in two groups in a clinical trial. It estimates the amount by which the experimental intervention changes the outcome on average compared with the control. It can be used as a summary statistic in meta-analysis when outcome measurements in all studies are made on the same scale.

Standardized mean difference

- The **standardized mean difference** is used as a summary statistic in meta-analysis when the studies all assess the same outcome but measure it in a variety of ways (for example, all studies measure depression but they use different psychometric scales). In this circumstance it is necessary to standardize the results of the studies to a uniform scale before they can be combined. The standardized mean difference expresses the size of the intervention effect in each study relative to the variability observed in that study. (Again in reality the intervention effect is a difference in means and not a mean of differences.):

$$\text{SMD} = \frac{\text{Difference in mean outcome between groups}}{\text{Standard deviation of outcome among participants}}$$

Effect measures for ordinal outcomes and measurement scales

Ordinal data

- **Ordinal outcome data** arise when each participant is classified in a category and when the categories have a natural order. For example, a 'trichotomous' outcome with an ordering to the categories, such as the classification of disease severity into 'mild', 'moderate' or 'severe', is of ordinal type. As the number of categories increases, ordinal outcomes acquire properties similar to continuous outcomes, and probably will have been analysed as such in a clinical trial.

Ordinal data

- **Measurement scales** are one particular type of ordinal outcome frequently used to measure conditions that are difficult to quantify, such as behaviour, depression, and cognitive abilities. Measurement scales typically involve a series of questions or tasks, each of which is scored and the scores then summed to yield a total 'score'. If the items are not considered of equal importance a weighted sum may be used.

Ordinal data

- Specialist methods are available for analysing ordinal outcome data that describe effects in terms of **proportional odds ratios**, but they are not available in RevMan, and become unwieldy (and unnecessary) when the number of categories is large. In practice longer ordinal scales are often analysed in meta-analyses as continuous data, whilst shorter ordinal scales are often made into dichotomous data by combining adjacent categories together. The latter is especially appropriate if an established, defensible cut-point is available.

Effect measures for counts and rates

Count data

- Some types of event can happen to a person more than once, for example, a myocardial infarction, fracture, an adverse reaction or a hospitalization. It may be preferable, or necessary, to address the number of times these events occur rather than simply whether each person experienced any event.

Count data

- Counts of rare events are often referred to as 'Poisson data' in statistics. Analyses of rare events often focus on **rates**. Rates relate the counts to the amount of time during which they could have happened. For example, the result of one arm of a clinical trial could be that 18 myocardial infarctions (MIs) were experienced, across all participants in that arm, during a period of 314 person-years of follow-up. The rate is 0.057 per person-year or 5.7 per 100 person-years. The summary statistic usually used in meta-analysis is the **rate ratio** (also abbreviated to RR), which compares the rate of events in the two groups by dividing one by the other. It is also possible to use a difference in rates as a summary statistic, although this is much less common.

Count data

- Counts of more common events, such as counts of decayed, missing or filled teeth, may often be treated in the same way as continuous outcome data. The intervention effect used will be the mean difference which will compare the difference in the mean number of events (possibly standardized to a unit time period) experienced by participants in the intervention group compared with participants in the control group.

Effect measures for time-to-event (survival) outcomes

Time- to- event data

- **Time-to-event data** arise when interest is focused on the time elapsing before an event is experienced. They are known generically as **survival data** in statistics, since death is often the event of interest, particularly in cancer and heart disease. Time-to-event data consist of pairs of observations for each individual: (i) a length of time during which no event was observed, and (ii) an indicator of whether the end of that time period corresponds to an event or just the end of observation. Participants who contribute some period of time that does not end in an event are said to be 'censored'. Their event-free time contributes information and they are included in the analysis. Time-to-event data may be based on events other than death, such as recurrence of a disease event (for example, time to the end of a period free of epileptic fits) or discharge from hospital.

Time- to- event data

- Time-to-event data can sometimes be analysed as dichotomous data. This requires the status of all patients in a study to be known at a fixed time-point. For example, if all patients have been followed for at least 12 months, and the proportion who have incurred the event before 12 months is known for both groups, then a 2×2 table can be constructed and intervention effects expressed as risk ratios, odds ratios or risk differences.

Summarizing effects across studies

Meta- analysis

Principles of meta-analysis

1. Meta-analysis is typically a two-stage process. In the first stage, a summary statistic is calculated for each study, to describe the observed intervention effect. For example, the summary statistic may be a risk ratio if the data are dichotomous or a difference between means if the data are continuous.

2. In the second stage, a summary (pooled) intervention effect estimate is calculated as a weighted average of the intervention effects estimated in the individual studies. A weighted average is defined as

$$\text{weighted average} = \frac{\text{sum of (estimate} \times \text{weight)}}{\text{sum of weights}} = \frac{\sum Y_i W_i}{\sum W_i}$$

Principles of meta-analysis

3. The combination of intervention effect estimates across studies may optionally incorporate an assumption that the studies are not all estimating the same intervention effect, but estimate intervention effects that follow a distribution across studies. This is the basis of a **random-effects meta-analysis**. Alternatively, if it is assumed that each study is estimating exactly the same quantity a **fixed-effect meta-analysis** is performed.

4. The standard error of the summary (pooled) intervention effect can be used to derive a confidence interval, which communicates the precision (or uncertainty) of the summary estimate, and to derive a P value, which communicates the strength of the evidence against the null hypothesis of no intervention effect.

5. As well as yielding a summary quantification of the pooled effect, all methods of meta-analysis can incorporate an assessment of whether the variation among the results of the separate studies is compatible with random variation, or whether it is large enough to indicate inconsistency of intervention effects across studies

A generic inverse-variance approach to meta-analysis

A generic inverse-variance approach to meta-analysis

- The inverse variance method is so named because the weight given to each study is chosen to be the inverse of the variance of the effect estimate (i.e. one over the square of its standard error). Thus larger studies, which have smaller standard errors, are given more weight than smaller studies, which have larger standard errors. This choice of weight minimizes the imprecision (uncertainty) of the pooled effect estimate.
- A fixed-effect meta-analysis using the inverse-variance method calculates a weighted average as

$$\text{generic inverse-variance weighted average} = \frac{\sum Y_i(1/SE_i^2)}{\sum(1/SE_i^2)}$$

Random-effects (DerSimonian and Laird) method for meta-analysis

- A variation on the inverse-variance method is to incorporate an assumption that the different studies are estimating different, yet related, intervention effects. This produces a random-effects meta-analysis, and the simplest version is known as the DerSimonian and Laird method (DerSimonian 1986).

Meta-analysis of dichotomous outcomes

Meta-analysis of dichotomous outcomes

- There are four widely used methods of meta-analysis for dichotomous outcomes, three fixed-effect methods (Mantel-Haenszel, Peto and inverse variance) and one random-effects method (DerSimonian and Laird). All of these methods are available as analysis options in RevMan. The Peto method can only pool odds ratios whilst the other three methods can pool odds ratios, risk ratios and risk differences.

The Mantel-Haenszel methods

- The Mantel-Haenszel methods (Mantel 1959, Greenland 1985) are the default fixed-effect methods of meta-analysis programmed in RevMan. When data are sparse, either in terms of event rates being low or study size being small, the estimates of the standard errors of the effect estimates that are used in the inverse variance methods may be poor. Mantel-Haenszel methods use a different weighting scheme that depends upon which effect measure (e.g. risk ratio, odds ratio, risk difference) is being used. They have been shown to have better statistical properties when there are few events. As this is a common situation in Cochrane reviews, the Mantel-Haenszel method is generally preferable to the inverse variance method. In other situations the two methods give similar estimates.

Peto odds ratio method

- Peto's method (Yusuf 1985) can only be used to pool odds ratios. It uses an inverse variance approach but utilizes an approximate method of estimating the log odds ratio, and uses different weights.
- The approximation used in the computation of the log odds ratio works well when intervention effects are small (odds ratios are close to one), events are not particularly common and the studies have similar numbers in experimental and control groups. In other situations it has been shown to give biased answers. As these criteria are not always fulfilled, Peto's method is not recommended as a default approach for meta-analysis.

Random-effects method

- The random-effects method (DerSimonian 1986) incorporates an assumption that the different studies are estimating different, yet related, intervention effects. The method is based on the inverse-variance approach, making an adjustment to the study weights according to the extent of variation, or heterogeneity, among the varying intervention effects. The random-effects method and the fixed-effect method will give identical results when there is no heterogeneity among the studies. Where there is heterogeneity, confidence intervals for the average intervention effect will be wider if the random-effects method is used rather than a fixed-effect method, and corresponding claims of statistical significance will be more conservative. It is also possible that the central estimate of the intervention effect will change if there are relationships between observed intervention effects and sample sizes.

Meta-analysis of continuous outcomes

Meta-analysis of continuous outcomes

- Two methods of analysis are available in RevMan for meta-analysis of continuous data: the inverse-variance fixed-effect method and the inverse-variance random-effects method. The methods will give exactly the same answers when there is no heterogeneity. Where there is heterogeneity, confidence intervals for the average intervention effect will be wider if the random-effects method is used rather than a fixed-effect method, and corresponding P values will be less significant.

Heterogeneity

heterogeneity

- Inevitably, studies brought together in a systematic review will differ. Any kind of variability among studies in a systematic review may be termed heterogeneity.
- Variability in the participants, interventions and outcomes studied may be described as **clinical diversity** (sometimes called clinical heterogeneity)
- Variability in study design and risk of bias may be described as **methodological diversity** (sometimes called methodological heterogeneity)
- Variability in the intervention effects being evaluated in the different studies is known as **statistical heterogeneity**, and is a consequence of clinical or methodological diversity, or both, among the studies. Statistical heterogeneity manifests itself in the observed intervention effects being more different from each other than one would expect due to random error (chance) alone. We will follow convention and refer to **statistical heterogeneity** simply as **heterogeneity**

Identifying and measuring heterogeneity

- It is important to consider to what extent the results of studies are consistent. If confidence intervals for the results of individual studies (generally depicted graphically using horizontal lines) have poor overlap, this generally indicates the presence of statistical heterogeneity. More formally, a statistical test for heterogeneity is available. This chi-squared (χ^2, or Chi^2) test is included in the forest plots in Cochrane reviews. It assesses whether observed differences in results are compatible with chance alone. A low P value (or a large chi-squared statistic relative to its degree of freedom) provides evidence of heterogeneity of intervention effects (variation in effect estimates beyond chance).

Identifying and measuring heterogeneity

- Thresholds for the interpretation of I^2 can be misleading, since the importance of inconsistency depends on several factors. A rough guide to interpretation is as follows:
- 0% to 40%: might not be important;
- 30% to 60%: may represent moderate heterogeneity*;
- 50% to 90%: may represent substantial heterogeneity*;
- 75% to 100%: considerable heterogeneity*.

$$I^2 = \left(\frac{Q - df}{Q} \right) \times 100\%$$

Strategies for addressing heterogeneity

1. Check again that the data are correct
2. Do not do a meta-analysis
3. Explore heterogeneity (subgroup analysis- meta-regression)
4. Ignore heterogeneity
5. Perform a random-effects meta-analysis
6. Change the effect measure
7. Exclude studies

Subgroup analysis

- Subgroup analyses involve splitting all the participant data into subgroups, often so as to make comparisons between them. Subgroup analyses may be done for subsets of participants (such as males and females), or for subsets of studies (such as different geographical locations). Subgroup analyses may be done as a means of investigating heterogeneous results, or to answer specific questions about particular patient groups, types of intervention or types of study.

Meta-regression

- If studies are divided into subgroups, this may be viewed as an investigation of how a categorical study characteristic is associated with the intervention effects in the meta-analysis. For example, studies in which allocation sequence concealment was adequate may yield different results from those in which it was inadequate. Here, allocation sequence concealment, being either adequate or inadequate, is a categorical characteristic at the study level. Meta-regression is an extension to subgroup analyses that allows the effect of continuous, as well as categorical, characteristics to be investigated, and in principle allows the effects of multiple factors to be investigated simultaneously (although this is rarely possible due to inadequate numbers of studies) (Thompson 2002). Meta-regression should generally not be considered when there are fewer than ten studies in a meta-analysis.

Incorporating assessment of risk of bias into analysis

- Formal comparisons of intervention effects according to risk of bias can be done using meta-regression. For studies with dichotomous outcomes, results of meta-regression analyses are most usefully expressed as ratios of odds ratios (or risk ratios) comparing results of studies at high or unclear risk of bias with those of studies at low risk of bias.

$$\text{Ratio of odds ratios} = \frac{\text{Intervention odds ratio in studies at high or unclear risk of bias}}{\text{Intervention odds ratio in studies at low risk of bias}}$$

Incorporating assessment of risk of bias into analysis

- Alternatively, separate comparisons of high versus low and unclear versus low can be made. For studies with continuous outcomes (e.g. blood pressure), intervention effects are expressed as mean differences between intervention groups, and results of meta-regression analyses correspond to differences of mean differences.
- If the estimated effect of the intervention is the same in studies at high and unclear risk of bias as in studies at low risk of bias then the ratio of odds ratios (or risk ratios) equals 1, while the difference between mean differences will equal zero.

Thanks for your kind attention