

Meta-analysis

Objectives:

1. define and compare review, systematic review and meta-analysis
2. summarize steps required for a systematic review including:
 - framing a specific question for review
 - identifying relevant literature
 - assessing the quality of the literature
 - summarizing the evidence
3. recognize the possible bias due to publication bias and describe approach to identifying publication bias using a funnel plot
4. interpret a forest plot
5. describe benefits and limitations of a meta-analysis
6. define heterogeneity, and recognize that heterogeneity may mean a meta-analysis is not feasible/valid
7. define cumulative meta-analysis
8. define sensitivity analysis

Definitions:

Review Article (*synonym*: narrative review)

- an unstructured summary of a topic (examples include most (but not all) textbook chapters, NEJM review articles etc.) – can be very useful approach to summarizing information
- lack of structure may cause problems with validity – for example:
 - reviewers may not review original research with respect to validity
 - may focus on a small subset of the available studies
 - it is not always clear how articles cited were identified (some resources may be selectively quoted)

Systematic review

- a rigorous structured review of a specific clinical question (e.g. Cochrane library)
- ‘systematic’ because original literature is summarized
- explicit information is provided about how relevant literature was identified and analyzed, including criteria used to reject studies

Elements of a systematic review

1. **define the clinical question** – a clearly formulated and focused clinical question, including patient population, intervention and outcome measure
2. **identify all completed studies** of the question (published and unpublished)
 - available data sources include:
 - Medline database
 - Cochrane clinical trials register
 - foreign language literature
 - references in primary sources
 - experts may have access to unpublished material
 - raw data from trials (by personal communication)

3. **critically appraise studies and select studies that are valid** according to predefined eligibility criteria (for example was the study described as randomized? was analysis by intention-to-treat?)
4. **describe quality of selected studies**
5. **summarize information, draw conclusions**, may use a ‘forest plot’ to summarize studies

Meta-analysis

- requires a systematic review and then integrates results of independent studies using specific statistical analyses
- also known as a ‘quantitative systematic review’
- should be as carefully planned as any other research proposal, and requires *a priori* definition of eligible studies and a comprehensive search of the literature (as for a systematic review)

Benefits of meta-analysis

1. a clearer picture – allows integration of multiple studies, including smaller studies that may have been inconclusive
2. less bias than an unsystematic review
3. precision – effect size may be more precise when trials are integrated, a larger sample size means statistical power improved
4. transparency – it may be difficult to completely eliminate bias, however a good meta-analysis will have a clearly written protocol
5. may be faster and less expensive than undertaking a large RCT

Limitations of meta-analysis

- quality of a meta-analysis is only as good as the individual studies it is based on
- bias (including publication bias) can be introduced in the process of locating and selecting studies for inclusion
- criteria for included studies are critical to understanding the meta-analysis
- it may not be appropriate to combine studies
- as with any study you need to consider both clinical and statistical significance

Is it possible to have meta-analyses on the same topic come to different conclusions?

- *yes – there are a number of examples in the literature of discordant results*
- *most often due to differences in studies identified and included in meta-analysis*
- *for example industry supported meta-analyses may come to more favourable conclusions than Cochrane reviews (Cochrane reviews compared with industry supported meta-analyses and other meta-analyses of the same drugs: systematic review. BMJ 2006;333:782)*

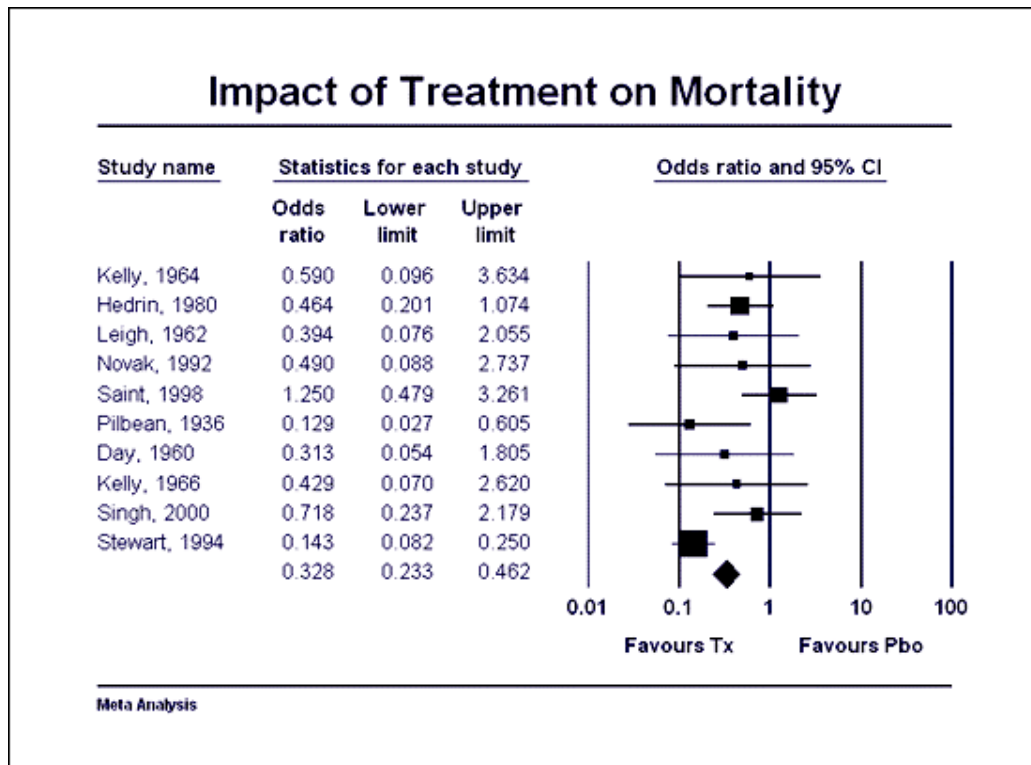
Are there examples where a meta-analysis of trials has come to a different conclusion than a subsequent large clinical trial?

- yes – (an example is the use of magnesium post myocardial infarction where many small studies showed benefit of Mg⁺⁺ and this was confirmed in a meta-analysis however a large clinical trial (ISIS) subsequently showed no benefit – the most likely explanation is publication bias with negative Mg⁺⁺ trials not being published and therefore distorting the evidence in the literature)

'Forest' plot

- simple visual representation of multiple studies; identifies studies and effect size
- studies identified chronologically (usually)
- horizontal lines run through point estimate of study to identify 95% confidence interval; (in example below the point estimate is shown as a square and the size of the square represents the size of the study; the line is the 95% CI and an odds ratio or relative risk of 1 represents 'no effect')
- note results are usually plotted on a log-scale
- the overall estimate from the meta-analysis and the CI are at the bottom (usually as a diamond, with the horizontal edges of the diamond representing the 95% CI)

Figure 1: An example of a forest plot



from: Forest plots: trying to see the woods and the trees. BMJ 2001;322

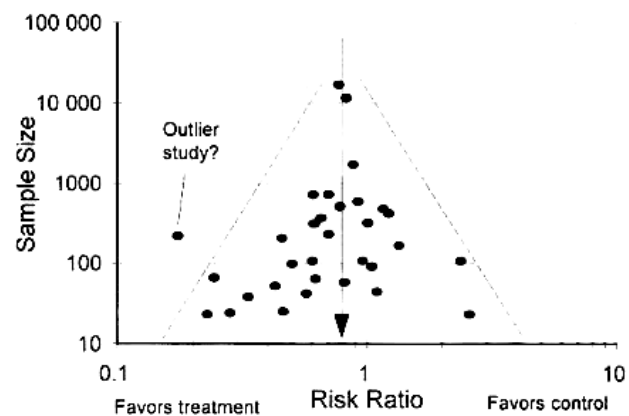
Publication bias

- publication bias occurs when the publication of research depends on the direction of the study results, and whether they are statistically significant
- publication bias includes ‘positive’ results being:
 - more likely to published
 - published rapidly
 - in English
 - in more than one source; duplication (this can cause overlap of information, making it seem as though there is more trial data than there actually is)
 - cited by others
- publication bias is a bigger danger with a review that consists of many small trials (studies with large numbers of patients are less likely to be unpublished)
- the medical literature therefore may be a selective and biased subset of studies and outcomes
- as a result of publication bias, and instances where negative trials, or trials with adverse events have not been published there is a move to a mandated trial registry (several of the core clinical journals have required trial registration as a prerequisite for publication of trials in the journal in the future)

Funnel plots

- funnel plots attempt to detect bias in the selection of studies for systematic review/meta-analysis
- results of each study are plotted against sample size
- if there is no bias then small studies should be spread out at the bottom of the graph around the true effect size
- if there is publication bias then the small studies will be clustered over to one side (‘treatment effective’) and the plot will be asymmetric (this assumes that large trials will be more likely to be published regardless of results and small trials will be more likely to be published if results are positive)
- asymmetry of the funnel plot, either visually interpreted or statistically tested, may suggest publication bias; *asymmetry however does not always accurately predict publication bias (BMJ 2006;333)*
- *another approach to assessing the impact of publication bias is to calculate how many studies of a given size with negative findings would need to be included to counterbalance a positive meta-analysis (if the amount of missing data needed to invalidate the meta-analysis is large then you will have more confidence in the findings)*

Figure 2: An example of a funnel plot showing symmetry



Heterogeneity

- patients, clinical settings, and treatment responses are expected to vary across trials that have studied the same problem; insight into reasons for the heterogeneity of trial results may often be as important as, or even more important, than obtaining aggregate results (*Ann Int Med* 1997;127:820)
1. **clinical heterogeneity** is the clinical differences in studies with respect to patients, interventions and outcomes
 - for example:
 - study location
 - age, sex, disease severity of subjects
 - other treatments subjects may be receiving
 - type of medication /dose / intensity of intervention
 - definitions of outcomes
 2. **methodological diversity (or heterogeneity)** includes differences such as type of study, study quality, length of study and approach to analysis (for example intention-to-treat)
 - because systematic reviews and meta-analysis bring together studies that are diverse both clinically and methodologically some heterogeneity in results is expected
 - it makes no sense however to pool results of very different studies ('apples and oranges')
 3. **statistical heterogeneity**
 - individual trials in a systematic review may seem to measure the same outcomes but have results that are not consistent with each other, for example some trials showing benefit and some showing harm
 - heterogeneity may be apparent when looking at the forest plot; it is possible to test for heterogeneity statistically in order to determine the extent to which trial results are (dis)similar
 - *formal statistical tests for heterogeneity:*
 - i. *null hypothesis for test of heterogeneity is that the treatment effect is the same for all the studies being analyzed*
 - *a low p value for the χ^2 test (Cochrane's Q) of heterogeneity means that chance is not likely to be responsible for the differences in the results from study to study*
 - *this means that if a low p value is found maybe the trials shouldn't be pooled together*
 - ii. *another test of heterogeneity is the I^2 statistic (BMJ 2003;327)*
 - *I^2 provides an estimate of the percentage of variability across studies that is likely due to a true difference in treatment effect, as opposed to chance*
 - *when I^2 increases (> 0.5) then heterogeneity is high*

Questions to consider:

1. Was it really a good idea to combine the trials?
2. Is there too much clinical heterogeneity for the review to make sense?
3. Do the forest plots look consistent?
4. Do statistical tests suggest that heterogeneity is a problem?

(Reference: *What is heterogeneity and is it important? BMJ 2007;334*)

Estimating a common effect:

- if the data are similar, what is the best common point estimate of a therapeutic effect and how precise is this estimate?
- pooled results have more patients and therefore a more precise estimate of treatment effect
- a type of weighted average is used to combine studies, where the results of some studies make a greater contribution to the total than others (studies with a larger sample size and higher event rates have a higher weighting); sample size is the main factor in determining the weight for a trial

Models for pooling data for meta-analysis

1. fixed-effects model:

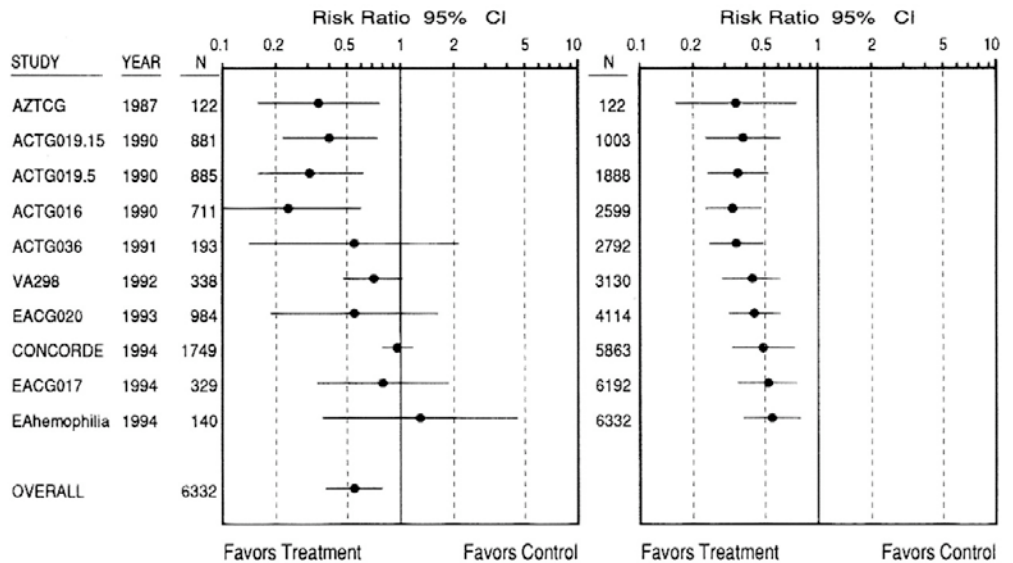
- assumes that there is a single 'true' value underlying all the study results; if all the studies were extremely large they would have identical estimates of treatment effect
- this model ignores between-study variation or heterogeneity

2. random-effects model

- assumes studies are a random sample of studies addressing related questions, and the results differ by chance AND by differences in the questions; each study will estimate a different underlying true effect
- assumes that there is no single underlying value of the effect, but there is a distribution of effects around a mean value
- will weight smaller studies higher than weighting with a fixed effects model
- tends to produce wider confidence intervals (a more conservative effect estimate)

Cumulative meta-analysis

- a new summary effect size and confidence interval is calculated each time the results of a new study become available (like a running summary of studies)
- figure shows a standard meta-analysis and on the left a cumulative meta-analysis (AnnIntMed1997)



Sensitivity analysis

- a sensitivity analysis provides information about whether assumptions and decisions made during the meta-analysis have had a major effect on the results
- involves a repetition of the analysis using different assumptions (as a quality check to make sure results are consistent)
- for example:
 - *if studies were excluded because they were judged to be of poor quality, what is the effect on the results if they are included?*
 - *if a study is an 'outlier' (ie. results are very different from remaining studies) what is the effect of excluding it?*
 - *both fixed and random-effects analyses can be done to see if results will be significantly different*