Department of Health Sciences

M.Sc. in Evidence Based Practice, M.Sc. in Health Services Research

Meta-analysis: method for quantitative data synthesis

Martin Bland

Professor of Health Statistics

University of York

http://martinbland.co.uk/msc/

Adapted from work by Seokyung Hahn

What is a meta-analysis?

- An optional component of a systematic review.
- A statistical technique for summarising the results of several studies into a single estimate.

What does it do?

- identifies a common effect among a set of studies,
- allows an aggregated clearer picture to emerge,
- improves the precision of an estimate by making use of all available data.

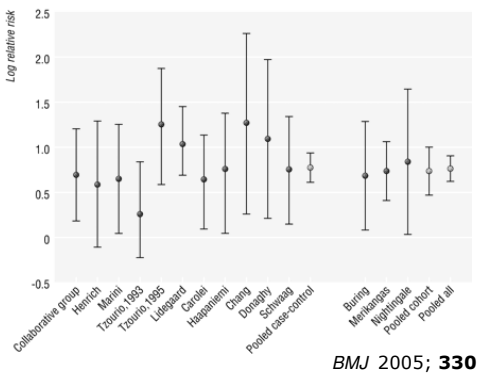
When can you do a meta-analysis?

- When more than one study has estimated the effect of an intervention or of a risk factor,
- when there are no differences in participants, interventions and settings which are likely to affect outcome substantially,
- when the outcome in the different studies has been measured in similar ways,
- when the necessary data are available.

A meta-analysis consists of three main parts:

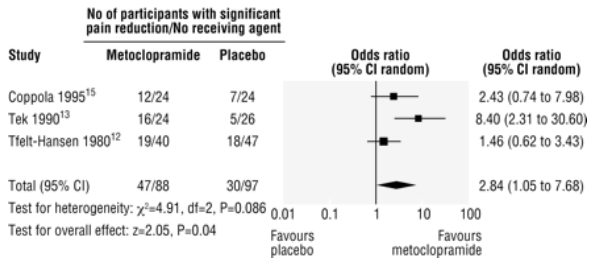
- a pooled estimate and confidence interval for the treatment effect after combining all the studies,
- a test for whether the treatment or risk factor effect is statistically significant or not (i.e. does the effect differ from no effect more than would be expected by chance?),
- a test for heterogeneity of the effect on outcome between the included studies (i.e. does the effect vary across the studies more than would be expected by chance?).

Example: migraine and ischaemic stroke



BMJ 2005; **330**: 63- .

Example: metoclopramide compared with placebo in reducing pain from acute migraine



BMJ 2004; **329**: 1369-72.

Types of meta-analysis

Meta-analysis can be done whenever we have more than one study addressing the same issue

- Interventions: usually randomised trials to give treatment effect.
- Epidemiological: usually case-control and cohort studies to give relative risk.
- Diagnostic: combined estimates of sensitivity, specificity, positive predictive value.

In this lecture I shall concentrate on clinical trials, but the principles are the same.

Summary statistics

- Calculate a summary statistic for each trial $\Rightarrow$ calculate an estimate of treatment effect for each trial
- Common effect is then calculated by averaging the individual study effects

BUT a simple average would treat all the trials as if they were of equal value.

Some trials have more information than others, e.g. are larger.

We weight the trials before we average them.

Weighted average

1. Define weights which reflect the importance of the trial.
2. E.g. weight = 1/variance of trial estimate
   = 1/standard error squared.
3. Multiply each trial difference by its weight and add.
4. Divide by sum of weights.

- **high variance ⇒ low amount of information ⇒ low weight**
- **low variance ⇒ high amount of information ⇒ high weight**

General framework for pooling results

- the pooled estimate is basically a summary measure of the results of the included trials,
- the pooled estimate is a weighted combination of the results from the individual trials,
- the weight given to each trial is the inverse of the variance of the summary measure from each of the individual trials,
- therefore, more precise estimates from larger trials with more events are given more weight.
- Then find 95% confidence interval and P value for the pooled difference.

Methods of meta-analysis

There are several different ways to produce the pooled estimate:

- inverse-variance weighting,
- Mantel-Haenszel method,
- Peto method,
- DerSimonian and Laird method.

Slightly different solutions to the same problem.

Heterogeneity

Studies differ in terms of
- Patients
- Interventions
- Outcome definitions
- Design

⇒ **Clinical heterogeneity**

- Variation in true treatment effects in magnitude or direction

⇒ **Statistical heterogeneity**

Heterogeneity

- Statistical heterogeneity may be caused by
  - clinical differences between trials
  - methodological differences between trials
  - unknown trial characteristics
- Even if studies are clinically homogeneous there may be statistical heterogeneity

Heterogeneity

How to identify statistical heterogeneity

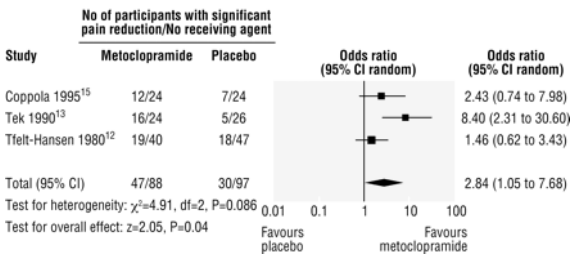
Test the null hypothesis that the trials all have the same treatment effect in the population.

The test looks at the differences between observed treatment effects for the trials and the pooled treatment effect estimate.

Square, divide by variance, sum.

This gives a chi-squared test with degrees of freedom = number of studies – 1.

Heterogeneity



Test for heterogeneity: $\chi^2 = 4.91$, df = 2, P=0.086.

Heterogeneity

Significant heterogeneity

- differences between trials exist
- it may be invalid to pool the results and generate a single summary result
- describe variation
- investigate sources of heterogeneity
- account for heterogeneity

Heterogeneity

Heterogeneity not significant

- No statistical evidence for difference between trials
- But, test for heterogeneity has low power - the number of studies is usually low - and may fail to detect heterogeneity as statistically significant when it exists.
- This cannot be interpreted as evidence of homogeneity.
- To compensate for the low power of the test a higher significance level is sometimes taken, $P < 0.1$ for statistical significance.

Types of outcome measure

Choice of measure of treatment effect depends on type of outcome variable:

Dichotomous

e.g. dead/alive, success/failure, yes/no

relative risk or risk ratio (RR), odds ratio (OR), absolute risk difference (ARD)

Continuous

e.g. weight loss, blood pressure

mean difference (MD), standardised mean difference(SMD)

Types of outcome measure

Choice of measure of treatment effect depends on type of outcome variable:

Time-to-event or survival time

e.g. time to death, time to recurrence, time to healing

Hazard ratio

Ordinal (very rare)

outcome categorised with an ordering to the categories

e.g. mild/moderate/severe, score on a scale

Dichotomise, treat as continuous, advanced methods.

Dichotomous outcome measure

Relative risk (RR), odds ratio (OR), absolute risk difference (ARD).

Relative risk and odds ratio both use logarithmic scales.

Why is this?

Example: ulcer healing (Fletcher *et al*., 1997)

elastic bandage: 31 healed out of 49 patients
inelastic bandage: 26 healed out of 52 patients.

RR = (31/49)/(26/52) = 1.27 (elastic over inelastic)

RR = (26/52)/(31/49) = 0.79 (inelastic over elastic)

We want a scale where 1.27 and 0.79 are equivalent.

Fletcher A, Nicky Cullum N, Sheldon TA. (1997) A systematic review of compression treatment for venous leg ulcers. *BMJ* 315: 576-580 .

Dichotomous outcome measure

RR = (31/49)/(26/52) = 1.273 (elastic over inelastic)

RR = (26/52)/(31/49) = 0.790 (inelastic over elastic)

We want a scale where 1.273 and 0.790 are equivalent.

Should be equally far from 1.0, the null hypothesis value.

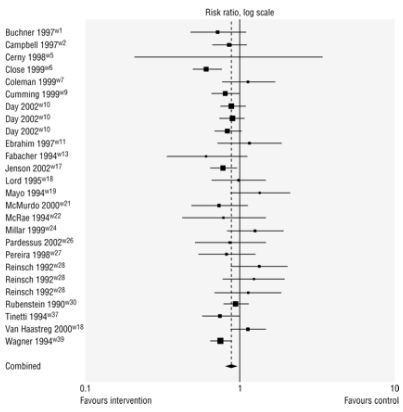
Logarithmic scale:

$\log_{10}(1.273) = 0.102$, $\log_{10}(0.790) = -0.102$

$\log_{10}(1) = 0$ (null hypothesis value)

$\log_{10}(1/2) = -0.301$, $\log_{10}(2) = +0.301$

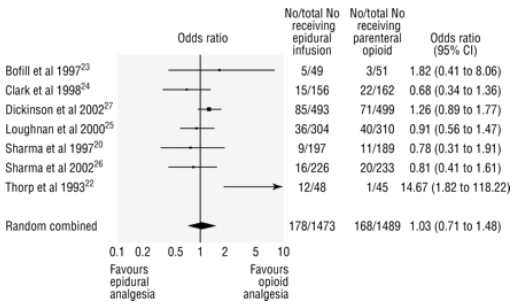
Logarithmic scale, risk ratio:



Interventions for the prevention of falls in older adults, pooled risk ratio of participants who fell at least once.

BMJ 2004; **328**: 680-3.

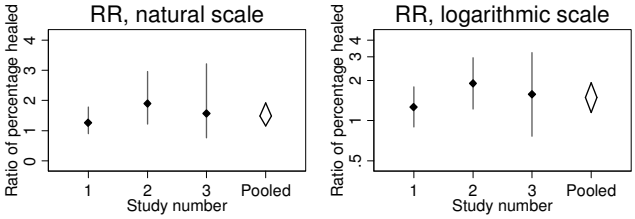
Logarithmic scale, odds ratios:



Rates of Caesarean section in trials of nulliparous women receiving epidural analgesia or parenteral opioids.
BMJ 2004; **328**: 1410-12.

Logarithmic scales:

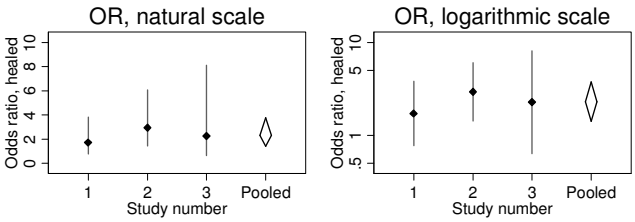
For both relative risk and odds ratio we find the standard error of the log ratio rather than the ratio. The log ratio also tends to have a Normal distribution. On the logarithmic scale, confidence intervals are symmetrical.



Data of Fletcher A, Cullum N, Sheldon TA. (1997) A systematic review of compression treatment for venous leg ulcers. *BMJ* **315**: 576-580 .

Logarithmic scales:

For both relative risk and odds ratio we find the standard error of the log ratio rather than the ratio. The log ratio also tends to have a Normal distribution. On the logarithmic scale, confidence intervals are symmetrical.



Data of Fletcher A, Cullum N, Sheldon TA. (1997) A systematic review of compression treatment for venous leg ulcers. *BMJ* **315**: 576-580 .

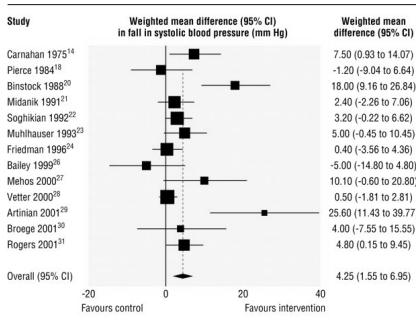
Continuous outcome variable

Measures of treatment effect for continuous outcome:

- Weighted Mean Difference:
  - -- same units as observations.
  - -- useful when the outcome is always the same measurement,
  - -- usually physical measurements.
- Standardised Mean Difference:
  - -- standard deviation units,
  - -- same as effect size,
  - -- useful when the outcome is not always the same measurement,
  - -- often psychological scales.

Continuous outcome variable

Example of weighted mean difference



Blood pressure control by home monitoring.
BMJ 2004; **329**: 145-9.

Continuous outcome variable

Data required:

mean,
standard deviation,
sample size,

for each group.

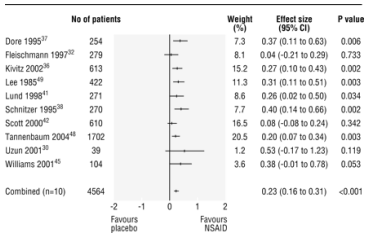
For each study, we then find the difference between means and its standard error in the usual way.

For standardised differences, we divide the difference between means by the standard deviation.

Everything is then in the same units, i.e. standard deviation units.

Continuous outcome variable

Example of standardised mean difference: pain scales



Non-steroidal anti-inflammatory drugs, including cyclo-oxygenase-2 inhibitors, in osteoarthritic knee pain.
BMJ 2004; **329**: 1317.

Continuous outcome variable

Data required: mean, standard deviation, sample size.

Unfortunately, these are not always available for all published studies.

Trials sometimes report different measure of variation:

- standard errors
- confidence intervals
- reference ranges
- interquartile ranges
- range
- significance test
- P value
- 'Not significant' or 'P<0.05'.

Continuous outcome variable

Extracting the standard deviation:

- standard errors — straightforward
- confidence intervals — straightforward
- reference ranges — straightforward
- interquartile ranges — needs assumption about distribution
- range — estimates unstable and affected by outliers
- significance test — can work back from a t value
- P value — can work back to a t value hence to SD.
- 'Not significant' or 'P<0.05' — hopeless.

Time to event outcome variables

Time-to-event data arise whenever we have subjects followed over time until some event takes place.

Often called survival data.

Techniques also used for:

- time to recurrence of disease,
- time to discharge from hospital,
- time to readmission to hospital,
- time to conception,
- time to fracture,
- etc.

Time to event outcome variables

Time-to-event data arise whenever we have subjects followed over time until some event takes place.

Problem: not all subjects have an event.

We know only that they were observed to be event-free up to some point, but not beyond it.

Usually some of those observed not to have an event were observed for a shorter time than some of those who did have an event.

Statistical techniques: survival analysis.

Time to event outcome variables

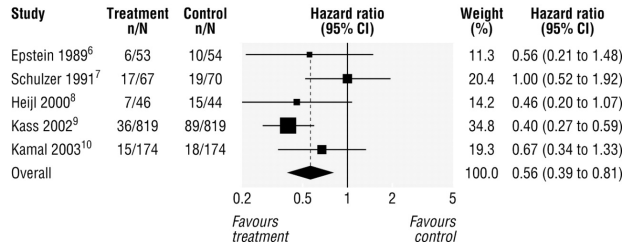
The main effect measure is the **hazard ratio**.

Standard outcome measure in survival analysis.

The ratio of the risk of having an event at any given time in one group divided by the risk of an event in the other.

Time to event outcome variables

Example: time to visual field loss or deterioration of the optic disc, or both, in patients with ocular hypertension



Hazard ratio = 1.0 represents no difference between the groups.

(BMJ 2005; **331**: 134.)

Time to event outcome variables

Hazard ratio is active treatment divided by no treatment, so if the hazard ratio is less than one, this means that the risk of visual field loss is less for patients given pressure lowering treatment.

As for risk ratios and odds ratios, hazard ratios are analysed by taking the log and the results are shown on a logarithmic scale.

Individual patient data meta-analysis

In this kind of meta-analysis, we get the raw data from each trial.

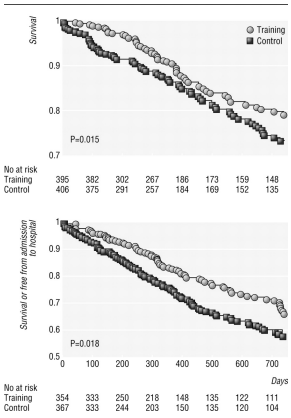
We can combine them into a single data set. We then analyse them like a single, multicentre clinical trial.

Alternatively, we may use the individual data to extract the corresponding summary statistics from each study then proceed as we would using summary statistics from published reports.

Example: Exercise training meta-analysis of trials in patients with chronic heart failure (ExTraMATCH) (*BMJ* 2004; **328**: 189).

Nine trials identified. Principal investigators provided a minimum data set in electronic form.

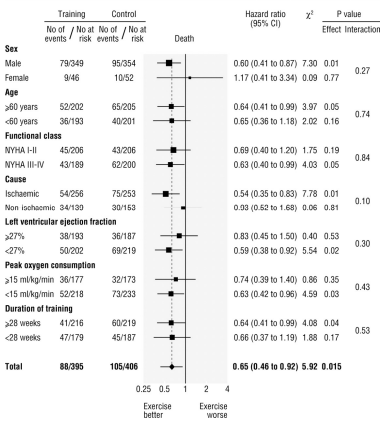
Individual patient data meta-analysis



Example: Exercise training meta-analysis of trials in patients with chronic heart failure --- survival curves

The Kaplan Meier survival curve shows the estimated proportion of subjects who have not yet experienced the event at each time.

More results from ExTraMATCH: outcome variable time to death



Individual study results are not given.

This plot shows only the effects of prognostic variables.

As it is a survival analysis, the effect is presented as a hazard ratio.

Log scale is used.

And finally,

Meta-analysis is straightforward if the data are straightforward and all available.

It depends crucially on the data quality and the completeness of the study ascertainment.