Standard statistical algorithms in Cochrane reviews

Version 5

Jon Deeks and Julian Higgins

on Behalf of the Statistical Methods Group
of The Cochrane Collaboration

April 2005

Data structure

Consider a meta-analysis of k studies. When the studies have a binary outcome the results of each study can be presented in a 2x2 table (Table 1) giving the numbers of subjects who do or do not experience the event in each of the two groups (here called intervention and control).

Table 1 Binary data

Study i	Event	No event	Total
Intervention	a_i	b_i	n_{1i}
Control	c_i	d_i	n_{2i}

If the outcome is a continuous measure, the number of subjects in each of the two groups, their mean response and the standard deviation of their responses are required to perform meta-analysis (Table 2).

Table 2 Continuous data

Study i	Group size	Mean response	Standard deviation
Intervention	n_{1i}	m_{1i}	sd_{1i}
Control	n_{2i}	m_{2i}	sd_{2i}

If the outcome is analysed by comparing observed with expected values (for example using the Peto method or a log-rank approach for time-to-event data), then 'O – E' statistics and their variances are required to perform the meta-analysis. Group sizes are also entered by the review author, but are not involved in the analysis.

Table 3 O minus E and variance

Study i	O minus E	Variance of (O minus E)	Group size (intervention)	Group size (control)
	Z_i	V_i	n_{1i}	n_{2i}

For other outcomes a generic approach can be used, the user directly specifying the values of the treatment effect and its standard error for each trial (the standard error may be calculable from a confidence interval). "Ratio" treatment effects (e.g. odds ratio, risk ratio, hazard ratio, ratio of means) will normally be expressed on a log-scale, "difference" treatment effects (e.g. risk

difference, differences in means) will normally be expressed on their natural scale. Group sizes can optionally be entered by the review author, but are not involved in the analysis.

Table 4 Generic data

Study i	Estimate of effect	Standard error of estimate	Group size (intervention)	Group size (control)
	$\hat{\theta}_i$	$\mathrm{SE}\left\{\hat{\theta}_i\right\}$	n_{1i}	n_{2i}

Formulae for individual studies

Individual study responses: binary outcomes

Peto odds ratio

For study i denote the cell counts as in Table 1, and let $n_{1i} = a_i + b_i$, $n_{2i} = c_i + d_i$, and $N_i = n_{1i} + n_{2i}$. For the Peto method the individual odds ratios are given by

$$OR_{Peto,i} = \exp\left\{\frac{Z_i}{V_i}\right\}.$$

The logarithm of the odds ratio has standard error

$$\mathrm{SE}\left\{\ln\left(OR_{Peto,i}\right)\right\} = \sqrt{\frac{1}{V_i}},$$

where Z_i is the 'O – E' statistic:

$$Z_i = a_i - \mathrm{E}\left[a_i\right],$$

with

$$\mathrm{E}\left[a_i\right] = \frac{n_{1i}\left(a_i + c_i\right)}{N_i}$$

(the expected number of events in the intervention group), and

$$V_i = \frac{n_{1i}n_{2i}\left(a_i + c_i\right)\left(b_i + d_i\right)}{N_i^2\left(N_i - 1\right)}$$

(the hypergeometric variance of a_i).

Odds ratio

For methods other than the Peto method, the odds ratio for each study is given by

$$OR_i = \frac{a_i d_i}{b_i c_i},$$

the standard error of the log odds ratio being

$$\mathrm{SE}\left\{\ln\left(OR_i\right)\right\} = \sqrt{\frac{1}{a_i} + \frac{1}{b_i} + \frac{1}{c_i} + \frac{1}{d_i}}.$$

Risk ratio

The risk ratio for each study is given by

$$RR_i = \frac{a_i / n_{1i}}{c_i / n_{2i}},$$

the standard error of the log risk ratio being

$$\mathrm{SE}\left\{\ln\left(RR_i\right)\right\} = \sqrt{\frac{1}{a_i} + \frac{1}{c_i} - \frac{1}{n_{1i}} - \frac{1}{n_{2i}}}.$$

Risk difference

The risk difference for each study is given by

$$RD_i = \frac{a_i}{n_{1i}} - \frac{c_i}{n_{2i}},$$

with standard error

$$\mathrm{SE}\left\{RD_i\right\} = \sqrt{\frac{a_i b_i}{n_{1i}^3} + \frac{c_i d_i}{n_{2i}^3}}.$$

Empty cells

Where zeros cause problems with computation of effects or standard errors, 0.5 is added to all cells (a_i, b_i, c_i, d_i) for that study, except when $a_i = c_i = 0$ or $b_i = d_i = 0$, when the relative effect measures OR_i and RR_i are undefined.

Individual study responses: continuous outcomes

Denote the number of subjects, mean and standard deviation as in Table 2, and let

$$N_i = n_{1i} + n_{2i}$$

and

$$s_i = \sqrt{\frac{\left(n_{1i} - 1\right) sd_{1i}^2 + \left(n_{2i} - 1\right) sd_{2i}^2}{N_i - 2}}$$

be the pooled standard deviation across the two groups.

Difference in means (mean difference)

The difference in means (referred to as mean difference) is given by

$$MD_i = m_{1i} - m_{2i},$$

with standard error

$$\mathrm{SE}\left\{MD_i\right\} = \sqrt{\frac{sd_{1i}^2}{n_{1i}} + \frac{sd_{2i}^2}{n_{2i}}}.$$

Standardised difference in means (standardised mean difference)

There are several popular formulations of the standardised mean difference. The one implemented in Cochrane reviews is Hedges adjusted g, which is very similar to Cohen's d, but includes an adjustment for small sample bias

$$SMD_i = \frac{m_{1i} - m_{2i}}{s_i}\left(1 - \frac{3}{4N_i - 9}\right),$$

with standard error

$$\mathrm{SE}\left\{SMD_i\right\} = \sqrt{\frac{N_i}{n_{1i} n_{2i}} + \frac{SMD_i^2}{2\left(N_i - 3.94\right)}}.$$

Individual study responses: O – E and variance

For study i the effect estimate is given by

$$\hat{\theta}_i = \frac{Z_i}{V_i},$$

with standard error

$$\mathrm{SE}\left\{\hat{\theta}_i\right\} = \sqrt{\frac{1}{V}}.$$

The effect estimate is either of a log odds ratio or a log hazard ratio, depending on how the observed and expected values were derived.

Individual study responses: Generic method

As the user directly enters the treatment effects and their standard errors no further processing is needed. All types of treatment effects are eligible for this method, but it might be most useful when treatment effects have been calculated in a way which makes special consideration of design (e.g. cluster randomised and cross-over trials), are adjusted for other effects (adjusted effects from non-randomised studies) or are not covered by existing methods (e.g. ratios of means, relative event rates).

Pooling methods

Mantel-Haenszel methods for combining trials

Odds ratio

The Mantel-Haenszel pooled odds ratio is given by

$$OR_{MH} = \frac{\sum w_{MH,i} OR_i}{\sum w_{MH,i}},$$

where each study's odds ratio is given weight

$$w_{MH,i} = \frac{b_i c_i}{N_i}.$$

The logarithm of OR_{MH} has standard error given by

$$\mathrm{SE}\left\{\ln\left(OR_{MH}\right)\right\} = \sqrt{\frac{1}{2}\left(\frac{E}{R^2} + \frac{F+G}{RS} + \frac{H}{S^2}\right)},$$

where

$$R = \sum \frac{a_i d_i}{N_i};\ S = \sum \frac{b_i c_i}{N_i};$$

$$E = \sum \frac{\left(a_i + d_i\right) a_i d_i}{N_i^2};\ F = \sum \frac{\left(a_i + d_i\right) b_i c_i}{N_i^2};$$

$$G = \sum \frac{\left(b_i + c_i\right) a_i d_i}{N_i^2};\ H = \sum \frac{\left(b_i + c_i\right) b_i c_i}{N_i^2}.$$

Risk ratio

The Mantel-Haenszel pooled risk ratio is given by

$$RR_{MH} = \frac{\sum w_{MH,i} RR_i}{\sum w_{MH,i}},$$

where each study's risk ratio is given weight

$$w_{MH,i} = \frac{c_i \left(a_i + b_i \right)}{N_i}.$$

The logarithm of RR_{MH} has standard error given by

$$\mathrm{SE}\left\{ \ln\left(RR_{MH} \right) \right\} = \sqrt{\frac{P}{RS}},$$

where

$$P = \sum \frac{n_{1i} n_{2i} \left(a_i + c_i \right) - a_i c_i N_i}{N_i^2};\ R = \sum \frac{a_i n_{2i}}{N_i};\ S = \sum \frac{c_i n_{1i}}{N_i}.$$

Risk difference

The Mantel-Haenszel pooled risk difference is given by

$$RD_{MH} = \frac{\sum w_{MH,i} RD_i}{\sum w_{MH,i}},$$

where each study's risk difference is given weight

$$w_{MH,i} = \frac{n_{1i} n_{2i}}{N_i}.$$

RD_{MH} has standard error given by

$$\mathrm{SE}\left\{ RD_{MH} \right\} = \sqrt{\frac{J}{K^2}},$$

where

$$J = \sum \frac{a_i b_i n_{2i}^3 + c_i d_i n_{1i}^3}{n_{1i} n_{2i} N_i^2};\ K = \sum \frac{n_{1i} n_{2i}}{N_i}.$$

Test for heterogeneity

The heterogeneity statistic is given by

$$Q_{MH} = \sum w_i \left(\hat{\theta}_i - \hat{\theta}_{MH} \right)^2,$$

where $\hat{\theta}$ represents a log odds ratio, log risk ratio or risk difference and the w_i are the weights calculated as $1 \Big/ \mathrm{SE}\left\{ \hat{\theta}_i \right\}^2$ rather than the weights used for the Mantel-Haenszel meta-analyses. Under the null hypothesis that there are no differences in treatment effect among trials this follows a chi-squared distribution with $k-1$ degrees of freedom (where k is the number of studies contributing to the meta-analysis).

The statistic I^2 is calculated as

$$I^2 = \max\left\{100\% \times \frac{Q_{MH} - (k-1)}{Q_{MH}}, 0\right\}$$

This measures the extent of inconsistency among the studies' results, and is interpreted as approximately the proportion of total variation in study estimates that is due to heterogeneity rather than sampling error.

Inverse-variance methods for combining trials

Inverse-variance methods are used to pool log odds ratios, log risk ratios and risk differences as one of the analysis options for binary data, to pool all mean differences and standardised mean differences for continuous data, and also for combining treatment effects in the generic method. In the general formula the treatment effect is denoted by $\hat{\theta}_i$ which is the trial's log odds ratio, log risk ratio, risk difference, mean difference or standardised mean difference, or the estimate of treatment effect in the generic method. The individual effect sizes are weighted according to the reciprocal of their variance (calculated as the square of the standard error given in the individual study section above) giving

$$w_i = \frac{1}{\left(\mathrm{SE}\left\{\hat{\theta}_i\right\}\right)^2}.$$

These are combined to give a pooled estimate

$$\hat{\theta}_{IV} = \frac{\sum w_i \hat{\theta}_i}{\sum w_i},$$

with

$$\mathrm{SE}\left\{\hat{\theta}_{IV}\right\} = \frac{1}{\sqrt{\sum w_i}}.$$

The heterogeneity statistic is given by a similar formula as for the Mantel-Haenszel method, using the inverse variance form of the weights, w_i

$$Q_{IV} = \sum w_i \left(\hat{\theta}_i - \hat{\theta}_{IV}\right)^2.$$

Under the null hypothesis that there are no differences in treatment effect among trials this follows a chi-squared distribution with $k-1$ degrees of freedom (where k is the number of studies contributing to the meta-analysis). I^2 is calculated as

$$I^2 = \max\left\{100\% \times \frac{Q_{IV} - (k-1)}{Q_{IV}}, 0\right\}.$$

Peto's method for combining trials

Here, the overall odds ratio is given by

$$OR_{Peto} = \exp\left\{\frac{\sum V_i \ln\left(OR_{Peto,i}\right)}{\sum V_i}\right\},$$

where the odds ratio $OR_{Peto,i}$ is calculated using the approximate method described in the individual trial section, and V_i are the hypergeometric variances.

The logarithm of the odds ratio has standard error

$$\mathrm{SE}\left\{\ln\left(OR_{Peto}\right)\right\} = \frac{1}{\sqrt{\sum V_i}}.$$

The heterogeneity statistic is given by

$$Q_{Peto} = \sum V_i \left\{ \left(\ln OR_{Peto,i} \right)^2 - \left(\ln OR_{Peto} \right)^2 \right\}.$$

Under the null hypothesis that there are no differences in treatment effect among trials this follows a chi-squared distribution with $k-1$ degrees of freedom (where k is the number of studies contributing to the meta-analysis). I^2 is calculated as

$$I^2 = \max \left\{ 100\% \times \frac{Q_{Peto} - (k-1)}{Q_{Peto}}, 0 \right\}.$$

O – E and variance method for combining trials

This is an implementation of the Peto method, which allows its application to time-to-event data as well as binary data. The overall effect estimate is given by

$$\hat{\theta} = \exp \left\{ \frac{\sum V_i \hat{\theta}_i}{\sum V_i} \right\},$$

where the estimate, $\hat{\theta}_i$, from study i is calculated from Z_i and V_i as for individual studies. The overall effect is either a log odds ratio or a log hazard ratio (the user should specify which).

The logarithm of the effect estimate has standard error

$$\mathrm{SE}\left\{ \hat{\theta} \right\} = \frac{1}{\sqrt{\sum V_i}}.$$

The heterogeneity statistic is given by

$$Q_{Peto} = \sum V_i \left(\hat{\theta}_i^{\,2} - \hat{\theta}^2 \right).$$

Under the null hypothesis that there are no differences in treatment effect among trials this follows a chi-squared distribution with $k-1$ degrees of freedom (where k is the number of studies contributing to the meta-analysis). I^2 is calculated as

$$I^2 = \max \left\{ 100\% \times \frac{Q_{Peto} - (k-1)}{Q_{Peto}}, 0 \right\}.$$

DerSimonian and Laird random-effects models

Under the random-effects model, the assumption of a common treatment effect is relaxed, and the effect sizes are assumed to have a distribution

$$\theta_i \sim N\left(\theta, \tau^2 \right).$$

The estimate of τ^2 is given by

$$\hat{\tau}^2 = \max \left\{ \frac{Q - (k-1)}{\sum w_i - \left(\sum w_i^2 \right) / \sum w_i}, 0 \right\},$$

where the w_i are the inverse-variance weights, calculated as

$$w_i = \frac{1}{\mathrm{SE}\left\{ \hat{\theta}_i \right\}^2},$$

for log odds ratio, log risk ratio, risk difference, mean difference, standardised mean difference, or for the treatment effect in the generic method, as appropriate.

For continuous data and for the generic method, Q is Q_{IV}. For binary data, either Q_{IV} or Q_{MH} may be taken. Both are implemented in RevMan 5 (and this is the only difference between random-effects methods under 'Mantel-Haenszel' and 'inverse-variance' options). Again, for odds ratios, risk ratios and other ratio effects, the effect size is taken on the natural logarithmic scale.

Each study's effect size is given weight

$$w_i' = \frac{1}{\mathrm{SE}\left\{\hat{\theta}_i\right\}^2 + \hat{\tau}^2}.$$

The pooled effect size is given by

$$\hat{\theta}_{DL} = \frac{\sum w_i' \hat{\theta}_i}{\sum w_i'},$$

and

$$\mathrm{SE}\left\{\hat{\theta}_{DL}\right\} = \frac{1}{\sqrt{\sum w_i'}}.$$

Note that in the case where the heterogeneity statistic Q is less than or equal to its degrees of freedom $(k-1)$, the estimate of the between trial variation, $\hat{\tau}^2$, is zero, and the weights coincide with those given by the inverse-variance method.

Confidence intervals

The $100(1-\alpha)\%$ confidence interval for $\hat{\theta}$ is given by

$$\hat{\theta} - \mathrm{SE}\left\{\hat{\theta}\right\}\Phi\left(1-\alpha/2\right) \quad \text{to} \quad \hat{\theta} + \mathrm{SE}\left\{\hat{\theta}\right\}\Phi\left(1-\alpha/2\right),$$

where $\hat{\theta}$ is the log odds ratio, log risk ratio, risk difference, mean difference, standardised mean difference or generic treatment effect and Φ is the standard normal deviate. For log odds ratios, log risk ratios and generic treatment effects entered on the log scale (and identified as such by the review author), the point estimate and confidence interval limits are exponentiated for presentation.

Test statistics

Test for presence of an overall treatment effect

In all cases, the test statistic is given by

$$Z = \frac{\hat{\theta}}{\mathrm{SE}\left(\hat{\theta}\right)},$$

where the odds ratio, risk ratio and other ratio treatment effects are again considered on the log scale. Under the null hypothesis that there is no overall effect of treatment effect this follows a standard normal distribution.

Test for comparison of subgroups

The test is valid for all methods except the Mantel-Haenszel methods for binary data. The Q statistic defined by either Q_{IV} or Q_{Peto} is calculated separately for each of the S subgroups and for the totality of studies, yielding statistics Q_1, …, Q_S and Q_{tot}. The test statistic is given by

$$Q_{\mathrm{int}} = Q_{\mathrm{tot}} - \sum_{j=1}^{S} Q_j .$$

Under the null hypothesis that there are no differences in treatment effect among subgroups this follows a chi-squared distribution with $S-1$ degrees of freedom (where S is the number of subgroups).

The statistic I^2 is calculated as

$$I^2 = \max\left\{100\% \times \frac{Q_{\text{int}} - (S-1)}{Q_{\text{int}}}, 0\right\}.$$

This measures the extent of inconsistency among the subgroups' results, and is interpreted as approximately the proportion of total variation in subgroup estimates that is due to genuine variation across subgroups rather than sampling error.

Bibliography

Breslow NE, Day NE. Combination of results from a series of 2x2 tables; control of confounding. In: Statistical Methods in Cancer Research, Volume 1: The analysis of case-control data. IARC Scientific Publications No.32. Lyon: International Agency for Health Research on Cancer, 1980.

Deeks JJ, Altman DG, Bradburn MJ. Statistical methods for examining heterogeneity and combining results from several studies in a meta-analysis. In: Egger M, Davey Smith G, Altman DG. Systematic Reviewes and Healthcare: meta-analysis in context. BMJ Publications (in press).

DerSimonian R, Laird N. Meta-analysis in clinical trials. Controlled Clinical Trials 1986; 7: 177-188.

Greenland S, Robins J. Estimation of a common effect parameter from sparse follow-up data. Biometrics 1985;41: 55-68.

Greenland S, Salvan A. Bias in the one-step method for pooling study results. Statistics in Medicine 1990; 9:247-252.

Hedges LV, Olkin I. Statistical Methods for Meta-analysis. San Diego: Academic Press 1985. Chapter 5.

Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analysis. BMJ 2003; 327: 557-560.

Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. Journal of the National Cancer Institute 1959;22: 719-748.

Robins J, Greenland S, Breslow NE. A general estimator for the variance of the Mantel-Haenszel odds ratio. American Journal of Epidemiolgy 1986; 124:719-723.

Rosenthal R. Parametric measures of effect size. In: Cooper H, Hedges LV (eds.). The Handbook of Research Synthesis. New York: Russell Sage Foundation, 1994.

Sinclair JC, Bracken MB. Effective Care of the Newborn infant.Oxford: Oxford University Press 1992.Chapter 2.

Yusuf S, Peto R, Lewis J, Collins R, Sleight P. Beta blockade during and after myocardial infarction: an overview of the randomized trials. Progress in Cardiovascular Diseases 1985;27:335-371.