ABC of Methodology

What is a forest plot?

ANDREA CIPRIANI and CORRADO BARBUI

Department of Medicine and Public Health, Section of Psychiatry and Clinical Psychology, University of Verona

KEY WORDS: forest plot, meta-analysis, methods, statistics.

In a previous note we pointed out limitations and strengths of reviewing and summarising primary research findings (Cipriani & Barbui, 2006). This note will help interpret the graphical results of meta-analyses of data extracted from primary studies. In meta-analyses, the process of combining patient data from different studies increases the precision of the estimates of treatment effects, and consequently the statistical power of the analysis (Lang & Secic, 1997). The typical graph for displaying the results of a meta-analysis is called a *forest plot*. The name *forest plot* (formerly "forrest plot" - see Lewis & Clarke, 2001) comes from the idea that this plot appears as a forest of lines. Each line stands for each study included in the meta-analysis. The forest plot allows readers to see the information from the individual studies that are included in the meta-analysis at a glance (Lewis & Clarke, 2001). It also provides a simple visual representation of the amount of variation between the results of the studies, as well as an estimate of the overall result of all the studies together (Sutton et al., 2000). Here below we report an example of meta-analytical study, and we describe step by step how to read a forest plot.

The names on the first column on the left of the plot are the names of the studies (usually the names of the first author) included in the meta-analysis (**1**). Studies can be listed according to subgroups of treatment intervention (i.e. placebo, comparator A, comparator B and so on) or according to other orders (i.e. alphabetical, by year of publication). Usually, results of these subgroups of studies can be pooled and are shown in the forest plot as subtotals [**1a**]. The overall estimate from the meta-analysis and its confidence intervals (CIs - see below for details) is reported at the bottom of the forest plot and is represented as a diamond [**1b**]. The centre of the diamond represents the pooled point estimate and its horizontal tips represent the CI. Significance is achieved at the set level if the diamond is clear of the line of no effect (that is below or above 1.0, wich corresponds to the vertical line in the graphs).

In the "number of events" column, the results of each study are pooled reporting the number of events out of the number of randomised patients (sample) for each treatment group (**2**). Accordingly, subtotals and total are also reported.

The results of each study are shown as squares centred on the point estimate (**3a**). The same results are also numerically reported as figures to make the graph clearer and more readable (**3b**). The point estimate is the best estimate of what is true for the relevant population from which the sample is taken. In other words, it corresponds to the estimate which has the biggest probability to be the most reliable estimate for that study. The most precise estimates are given the largest plotting symbols. In the graph, the horizontal line which runs through the square shows the CI (reported in brackets **3b**). The CI is the measure of the uncertainty associated with the point estimate and it is due to the play of chance. Confidence limits are the lower and upper boundaries/values of a CI, that is, the values which define the range of a CI. The width of the confidence interval gives us some idea about how uncertain we are about the unknown parameter (real estimate). The CI gives an estimated range of values which is likely to include the unknown true estimate and is expressed as percentage. CIs are usually calculated so that this percentage is 95%, but 90%, 99%, 99.9% CIs for the unknown parameter can be produced. It should be remembered that estimates with the widest CIs, which thus have the most visual impact, are the least precise and generally least influential. CIs are more informative than the simple results of hypothesis tests (p value) since they provide a range of plausible values.

In the graph the vertical line corresponds to the risk ratio of 1 (**3c**) and means that there is no difference of effect between treatment. Risk ratio can be expressed either as relative risk (RR) or as *odds* ratio (OR). The odds for a group is defined as the number of patients in the group who achieve the stated end point divided by the number of patients who do not. By contrast, *risk* is calculated as the number of patients in the group who achieve

Address for correspondence: Dr. A. Cipriani, Department of Medicine and Public Health, Section of Psychiatry and Clinical Psychology, University of Verona, Piazzale L.A. Scuro, 10 - 37134 Verona, Italy.
Fax: +39 045-585871
E-mail: andrea.cipriani@univr.it

https://doi.org/10.1017/S1121189X0000213X Published online by Cambridge University Press

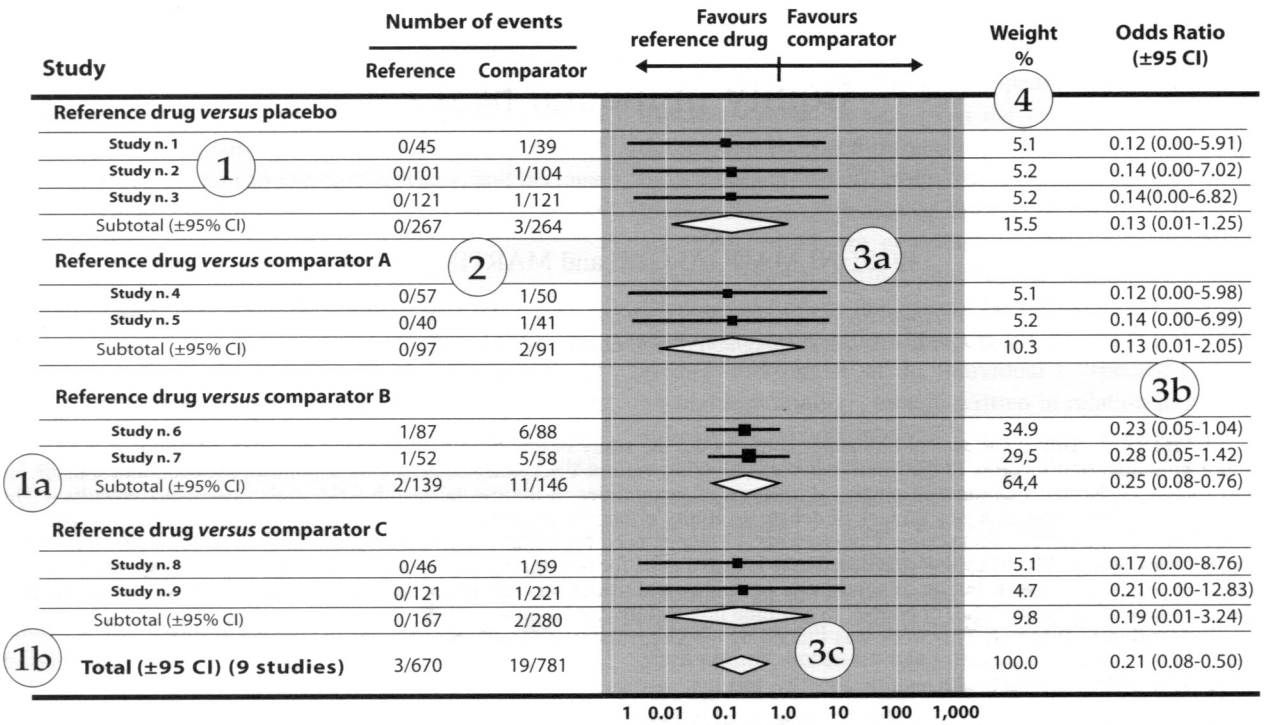


Figure 1. – *Graphic representation of the forest plot. Details and interpretation are discussed in the text.*

the stated end point divided by the total number of patients in the group. As the name implies, *OR* (or RR) is a ratio of two odds (or two risks). They are simply defined as the ratios of the odds (risks) of the treatment group to the odds (risks) of the control group. An OR or RR greater than 1 indicates increased likelihood of the stated outcome being achieved in the treatment group. If the OR or RR is less than 1, there is a decreased likelihood in the treatment group. A ratio of 1 indicates no difference - that is, the outcome is just as likely to occur in the treatment group as it is in the control group. Statistical significance refers to the likelihood that the results obtained in a study were not due to chance alone. CIs may be used to assess statistical significance. If the CI for relative risk or odds ratio for an estimate *includes* 1, then we are unable to demonstrate a statistically significant difference between the groups being compared; by contrast, if CI *does not include* 1, then there is a statistically significant difference. Clinical significance reflects the value of the results to patients and could be defined as a difference in effect size between groups that is considered to be important in clinical decision making - regardless of whether the difference is statistically significant or not.

The bottom line of the graph (**3c**) is a logarithmic scale. For the purposes of combining results of different studies, it is recommended to work with the natural logarithm of the OR, as this should provide a measure which is approximately normally distributed.

The overall treatment effect is calculated as a weighted average of the individual summary statistics (**4**). It should be noted that in meta-analysis data from individual studies are not simply combined as if they were from a single study. Each study estimate is given a weight directly proportional to its precision, that is inversely proportional to its variance and which relates closely to sample size. Greater weights are given to the results of studies that provide more information (number of events, greater sample size), because they are likely to be closer to the "true effect" we are trying to estimate.

REFERENCES

Cipriani A. & Barbui C. (2006). What is a systematic review? *Epidemiologia e Psichiatria Sociale* 15, 174-175.

Lang T.A. & Secic M. (1997). *How to Report Statistics in Medicine*. American College of Physicians: Philadelphia.

Lewis S. & Clarke M. (2001). Forest plots: trying to see the wood and the trees. *British Medical Journal* 322,1479-1480.

Sutton A.J., Abrams K.R., Jones D.R., Sheldon T.A. & Song P. (2000). *Methods for Meta-Analysis in Medical Research*. John Wiley & Sons: Chichester, Ltd.

https://doi.org/10.1017/S1121189X0000213X Published online by Cambridge University Press