Intern Emerg Med (2007) 2:140–142
DOI 10.1007/s11739-007-0036-8

COCHRANE'S CORNER

Understanding systematic reviews: the meta-analysis graph (also called 'forest plot')

L. Moja • I. Moschetti • A. Liberati
G.F. Gensini • R. Gusinu

L. Moja (✉) • I. Moschetti • A. Liberati
Italian Cochrane Centre
Mario Negri Institute for Pharmacological Research
Via Eritrea 62, I-20157 Milan, Italy
e-mail: moja@marionegri.it

G.F. Gensini
Department of Critical Care Medicine and Surgery
University of Florence and Azienda Ospedaliero-Universitaria Careggi
Florence, Italy

R. Gusinu
DAI Cardiologico e dei Vasi
Azienda Ospedaliero-Universitaria Careggi
Florence, Italy

Published online: 9 July 2007

Olsen O, Gotzsche PC (2001)
Screening for breast cancer with mammography.
Cochrane Database Syst Rev (4):CD001877

Background Mammographic screening for breast cancer is controversial, as reflected in greatly varying national policies.
Objectives The objective was to assess the effect of screening for breast cancer with mammography on mortality and morbidity.
Search strategy MEDLINE (16 May 2000), The Cochrane Breast Cancer Group's trial register (24 Jan 2000) and reference lists. Letters, abstracts and unpublished trials. Authors were contacted.
Selection criteria Randomised trials comparing mammographic screening with no mammographic screening.
Data collection and analysis Data were extracted by both authors independently.
Main results Seven completed and eligible trials involving half a million women were identified. The two best trials provided medium-quality data and, when combined, yield a relative risk for overall mortality of 1.00 (95% CI 0.96-1.05) after 13 years. However, the trials are underpowered for all-cause mortality, and confidence intervals include a possible worthwhile effect as well as a possible detrimental effect. If data from all eligible trials (excluding flawed studies) are considered then the relative risk for overall mortality after 13 years is 1.01 (95% CI 0.99-1.03). The best trials failed to show a significant reduction in breast cancer mortality with a relative risk of 0.97 (95% CI 0.82-1.14). If data from all eligible trials (excluding flawed studies) are considered then the relative risk for breast cancer mortality after 13 years is 0.80 (95% CI 0.71-0.89). However, breast cancer mortality is considered to be an unreliable outcome and biased in favour of screening. Flaws are due to differential exclusion of women with breast cancer from analysis and differential misclassification of cause of death.
Reviewer's conclusions The currently available reliable evidence does not show a survival benefit of mass screening for breast cancer (and the evidence is inconclusive for breast cancer mortality). Women, clinicians and policy makers should consider these findings carefully when they decide whether or not to attend or support screening programs.

The methodologist's point of view

L. Moja, I. Moshetti, A. Liberati

As the Cochrane Corner hosts analytical comments on Cochrane systematic reviews (SRs), it is important that readers are comfortable understanding the science of combining together the results of different studies, including the most intimidating technical aspects. In this issue we start to cover a very focused range of concepts/tips related to SRs, from understanding meta-analysis graphs and all their numbers to the impact of heterogeneity on SRs.
The Cochrane Collaboration defines a SR as a review of a clearly formulated question that uses systematic and explicit methods to identify, select and critically appraise relevant research, and to collect and analyse data from the studies that are included in the review [1].
Some authors use SR and meta-analysis as interchangeable words. Following the terminology adopted by the Cochrane Collaboration, SRs and meta-analyses are two distinct entities. The term SR indicates the entire research process, while meta-analysis is the statistical technique performed in a SR to quantitatively combine the results of individual studies. A meta-analysis may or may not be part of a SR (i.e., a SR of cognitivistic theories) and may or may not be appropriate (i.e., when there is an important clinical heterogeneity between included studies). Therefore, in SRs, when appropriate, authors can perform a meta-analysis allowing a more precise estimate of the magnitude of the treatment effect. It is not to simply sum up together the results from different research and calculate a summary statistic as if it were one big study. It is to calculate a weighted average of the results across studies, in a way that ensures that studies with larger sample sizes and more events contribute more to the overall result.

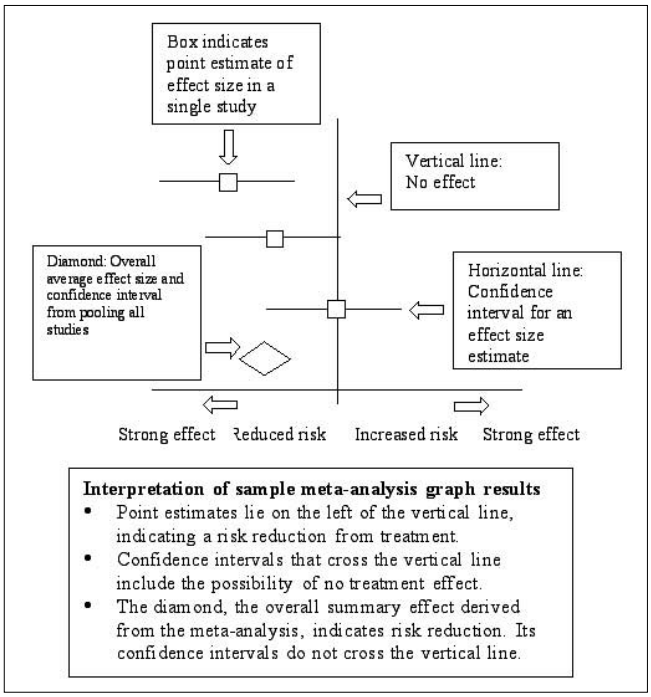

Fig. 1 Sample of a hypothetical meta-analysis graph

The most commonly used graphical presentation of SR results is the *forest plot*, so named because it helps readers to see the 'forest' of evidence while still being able to focus on the 'trees' of individual studies [2]. While less interesting, *meta-analysis graph* is a more rigorous name for it.

Basic elements

The basic elements of a meta-analysis graph are: the studies (boxes); a vertical line; and a diamond (meta-analysis result) (Fig. 1).

You are interested to know if a special treatment does more good than harm. Let us imagine that there are three studies that answered this question. Each study is indicated by a *box*. A horizontal line crosses each study box and its extremes represent the confidence interval of the study estimate, a measure of how the result of this study might vary with the play of chance. Indeed you decide to divide a blank sheet with a central *vertical line*: studies' estimates favouring the treatment over the control lie on the left side, while studies favouring the control lie on the right. Studies strongly favouring the treatment are located far from the vertical line, while studies not showing differences between treatment and control lie on the line (*no-effect line*). Estimates on the left-hand side of the vertical line do not always mean the treatment is better than the control (pay attention to the type of outcome). The pooled effect is the last element and is presented as a diamond at the bottom. The diamond horizontal ends are its confidence interval.

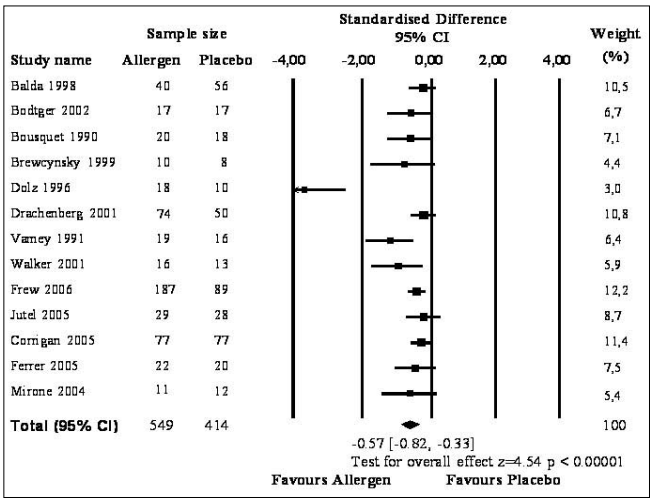

Fig. 2 *Review*: allergen injection immunotherapy for seasonal allergic rhinitis. *Comparison*: active treatment *vs*. placebo. *Outcome*: medication score. We have simplified Calderon et al.'s original analysis 01.02 [3]

A 'hay fever' example

In a recent review Calderon et al. aimed at evaluating the efficacy and safety of injection immunotherapy, compared to placebo, for reducing symptoms and medication needs in seasonal allergic rhinitis ('hay fever') due to tree, grass or weed pollens [3]. Specific allergen immunotherapy is most commonly administered as subcutaneous injections by specialists requiring a building-up period followed by a maintenance period of three to five years.

Fifty-one studies satisfied their inclusion criteria (2871 participants, 1645 in the treatment groups and 1226 in the placebo). The duration of treatment varied from three days to three years. In Figure 2 we present the results of the medication use (typically a daily score reflecting use of oral antihistamine tablets entered on a diary card and subsequently totalled and averaged). In this meta-analysis 13 studies considering concurrent use of anti-allergic medication were pooled (including 549 participants treated with active immunotherapy and 414 treated with placebo). Sample size of single studies is presented as well as the percentage weight given to the studies in the pooled analysis. The study boxes vary in size because they are proportional to the percentage weight.

The summary statistic is the standardised difference, which is a difference in means used when the trials all assess the same outcome (medication score) using different scales. The diamond at the bottom of Figure 2 is centred on the midpoint of the area of overlap of confidence intervals around the estimates of the individual trials. In this example the primary aim of a meta-analysis becomes clear: to include enough studies (not all significant by themselves) to narrow the confidence interval around the pooled estimate. Its confidence interval line does not cross the vertical line. Thus we can be confi-

dent of the benefit for our patients treated with active immunotherapy. In numbers the standardised difference following immunotherapy was –0.57 (95% CI –0.82 to –0.33), indicating a significant reduction (test for overall effect $p<0.00001$) in medication scores.

This review has shown that injection immunotherapy in suitably selected patients with hay fever results in significant reductions in medication use and symptom scores (data not presented) [1]. Injection immunotherapy has a known and relatively low risk of severe adverse events. This review found that the treatment was safe, with serious adverse reactions to the therapy occurring in only four patients (one of whom had been given a placebo) and adrenaline use only in 19 patients out of 14 085 injections. This low risk of life-threatening adverse reactions makes this treatment feasible only in hospitals that have full resuscitation back-up. The risk of an adverse reaction is increased in asthma sufferers, particularly if poorly controlled. For these reasons, many countries do not recommend immunotherapy in people with asthma, which excludes a large proportion of the population that might otherwise benefit.

The meta-analysis graph is a simple and clear way to give readers a visual assessment of results within studies and an overall pooled estimate. Although the meta-analysis graph helps to summarise the relative variability observed across trials, it may be still problematic to interpret the overall average effect size, because of two reasons. First, we may approach a meta-analysis in which the differences in the participants, interventions or outcomes (i.e., measurement scales) among studies are so important that a combined estimate is not a meaningful description of the set of studies. These differences are referred to as heterogeneity and will be covered in the next issue of this journal. Second, the meta-analysis graph does not solve difficulties in interpreting statistical measures: in our case the standardised difference reports the results in units of standard deviation rather than in units of any of the measurement scales used in the primary trials, leaving to the reader the mysterious extent of the magnitude of the effect.

For readers who are interested in learning more about systematic reviews, we suggest to freely access the Cochrane Open Learning Material at http://www.cochrane-net.org/openlearning [4].

The clinician's point of view

G.F. Gensini, R. Gusinu

Meta-analysis is widely used in basic research to evaluate the evidence in different areas, by pharmaceutical companies to gain approval for new drugs, by clinicians to determine the most effective course of treatment and by applied researchers in health and social sciences to plan and justify new studies.

However, although a meta-analysis can have mathematical precision, in relation to any biases that arise from the study selection process, it may produce a mathematically precise, but clinically misleading, result.

Readers of the medical literature must be conscious of the points of strength and weakness of this technique, considering the criteria to estimate the methodological quality of meta-analysis listed below:

- the attention to tracing all the studies, consulting the data banks of scientific literature;
- the explicitness of the criteria of inclusion of the studies;
- the appraisal of the methodological quality of the studies according to reproducibility and defined criteria;
- the homogeneity of the studies; and
- the choice of the outcomes.

In fact the vulnerability of meta-analysis is represented by:

- publication bias;
- selection bias; and
- heterogeneity.

However, although publication bias has to be considered of special importance, perhaps as the greatest threat to the validity of this method, this problem is not an argument against its use, because such biases exist in the literature irrespective of whether systematic review or other methodology is used to summarise research findings.

References

1. Green S, Higgins J (eds) Glossary. Cochrane Handbook for Systematic Reviews of Interventions, version 4.2.5. http://www.cochrane.org/resources/glossary.htm (accessed 4 March 2007)
2. Lewis S, Clarke M (2001) Forest plots: trying to see the wood and the trees. BMJ 322:1479–1480
3. Calderon MA, Alves B, Jacobson M et al (2007) Allergen injection immunotherapy for seasonal allergic rhinitis. Cochrane Database of Systematic Reviews Issue 1. Art. No.: CD001936. DOI: 10.1002/14651858.CD001936.pub2
4. Alderson P, Green S (eds) The Cochrane Open Learning Material. The Cochrane Collaboration 2002, version 1.1. http://www.cochrane-net.org/openlearning (accessed 4 March 2007)