

Supported by sanofi-aventis

**Iain K Crombie**

PhD FPHM Professor of Public Health, University of Dundee

**Huw TO Davies**

PhD Professor of Health Care Policy and Management, University of St Andrews

# What is meta-analysis?

- **Meta-analysis is a statistical technique** for combining the findings from independent studies.
- Meta-analysis is most often used to assess the **clinical effectiveness of healthcare interventions**; it does this by combining data from two or more randomised control trials.
- Meta-analysis of trials provides a **precise estimate of treatment effect**, giving due weight to the size of the different studies included.
- The validity of the meta-analysis depends on the **quality of the systematic review** on which it is based.
- Good meta-analyses aim for **complete coverage of all relevant studies**, look for the **presence of heterogeneity**, and explore the robustness of the main findings using **sensitivity analysis**.

For further titles in the series, visit:  
[www.whatisseries.co.uk](http://www.whatisseries.co.uk)

## What is meta-analysis?

### Trials, systematic reviews and meta-analysis

In many medical specialties it is common to find that several trials have attempted to answer similar questions about clinical effectiveness; for example: **Does the new treatment confer significant benefits compared with the conventional treatment?** Often many of the individual trials will fail to show a statistically significant difference between the two treatments. However, when the results from the individual studies are combined using appropriate techniques (meta-analysis), significant benefits of treatment may be shown. A good example of this is a retrospective review of the evidence on the effectiveness of thrombolytic therapy for the prevention of myocardial infarction.<sup>1</sup> The study showed that had meta-analysis been conducted at an early stage, it would have demonstrated the benefits of thrombolytic therapy. Instead, experts remained unaware of its benefits for many years and patients were not given an effective therapy. Meta-analyses are now a hallmark of evidence-based medicine.

### Systematic reviews

Systematic review methodology is at the heart of meta-analysis. This stresses the need to take great care to find all the relevant studies (published and unpublished), and to assess the methodological quality of the design and execution of each study.<sup>2</sup> The objective of systematic reviews is to present a balanced and impartial summary of the existing research, enabling decisions on effectiveness to be based on all relevant studies of adequate quality. Frequently, such systematic reviews provide a quantitative (statistical) estimate of net benefit aggregated over all the included studies. **Such an approach is termed a meta-analysis.**

### Benefits of meta-analyses

Meta-analysis offers a rational and helpful way of dealing with a number of practical difficulties that beset anyone trying to make sense of effectiveness research.

### Overcoming bias

The danger of unsystematic (or narrative) reviews, with only a portion of relevant studies included, is that they could introduce bias. Certain (perhaps favourable) reports may be more likely to be included in a review than those which show no significant differences; and informal synthesis may be tainted by the prior beliefs of the reviewer. Meta-analysis carried out on a rigorous systematic review can overcome these dangers – offering an unbiased synthesis of the empirical data.

### Precision

The precision with which the size of any effect can be estimated depends to a large extent on the number of patients studied. Meta-analyses, which combine the results from many trials, have more power to detect small but clinically significant effects. Furthermore, they give more precise estimates of the size of any effects uncovered. This may be especially important when an investigator is looking for beneficial (or deleterious) effects in specific subgroups of patients. Individual studies may contain too few patients in the subgroup of interest to be informative. However, the systematic aggregation of data from many individual studies gives a clearer picture, particularly through use of the technique of meta-regression (see below).

### Transparency

It is not simply the case that meta-analyses can always exclude bias more readily than other forms of review. Their advantage also lies in the openness with which good meta-analyses reveal all the decisions that have been taken throughout the process

of achieving the final aggregate effect sizes. Thus, good meta-analyses should allow readers to determine for themselves the reasonableness of the decisions taken and their likely impact on the final estimate of effect size.

## Requirements for meta-analysis

**The main requirement for a worthwhile meta-analysis is a well-executed systematic review.**<sup>3</sup> However competent the meta-analysis, if the original review was partial, flawed or otherwise unsystematic, then the meta-analysis may provide a precise quantitative estimate that is simply wrong. The main requirement of systematic review is easier to state than to execute: a complete, unbiased collection of all the original studies of acceptable quality that examine the same therapeutic question. There are many checklists for the assessment of the quality of systematic reviews;<sup>4</sup> however, the QUOROM statement (**quality of reporting of meta-analyses**) is particularly recommended.<sup>5</sup>

## Conducting meta-analyses

### Location of studies

Meta-analysis requires a comprehensive search strategy which interrogates several electronic databases (for example, MEDLINE, EMBASE, Cochrane Central Register of Controlled Trials). Hand-searching of key journals and checking of the reference lists of papers obtained is also recommended.<sup>6</sup> The search strategy – the key terms used to search the database – needs to be developed with care. The strategy is written as a sequence of requirements: include papers with specified terms, exclude papers that do not meet certain criteria (for example, age or diagnostic group), only include studies that follow certain research designs (for example, randomised controlled trials).

### Quality assessment

Once all relevant studies have been identified, decisions must be taken about

which studies are sufficiently well conducted to be worth including. This process may again introduce bias, so good meta-analyses will use **explicit and objective criteria for inclusion or rejection** of studies on quality grounds.<sup>7</sup> There is a bewildering array of scales for assessing the quality of the individual clinical trials.<sup>8</sup> Two scales that are commonly used are those developed by Chalmers *et al*<sup>9</sup> and Jadad *et al*.<sup>10</sup> Perhaps more important than the scale used is whether a scale has been used at all. Once a quality score has been assigned, the impact of excluding low quality studies can be assessed by **sensitivity analysis** (see below).

### Calculating effect sizes

Clinical trials commonly present their results as the frequency of some outcome (such as a heart attack or death) in the intervention groups and the control group. For meta-analysis these are usually summarised as a ratio of the frequency of the events in the intervention to that in the control group. In the past the most common summary measure of effect size was the **odds ratio**, but now **the risk ratio (relative risk)** can be given. Although they are technically different, the odds ratios and relative risks are usually interpreted in the same way.<sup>11</sup> Thus, a ratio of 2 implies that the defined outcome happens about twice as often in the intervention group as in the control group; an odds ratio of 0.5 implies around a 50% reduction in the defined event in the treated group compared with the controls.

The findings from individual studies can be combined using an appropriate statistical method.<sup>12</sup> Separate methods are used for combining odds ratios, relative risks and other outcome measures such as risk difference or hazard ratio. The methods use a similar approach in which the estimate from each study is weighted by the precision of the estimate.

### Checking for publication bias

A key concern is publication bias, as clinical trials that obtain negative

# What is meta-analysis?

findings (that is, no benefit of treatment) are less likely to be published than those that conclude the treatment is effective.<sup>13</sup> One simple way of assessing the likely presence of publication bias is to examine a **funnel plot**.<sup>14</sup> Funnel plots display the studies included in the meta-analysis in a plot of effect size against sample size<sup>15</sup> (or some other measure of the extent to which the findings could be affected by the play of chance).<sup>16</sup> As smaller studies have more chance variability than larger studies, the expected picture is one of a symmetrical inverted funnel (Figure 1).<sup>17</sup> If the plot is asymmetric, this suggests that the meta-analysis may have missed some trials – usually smaller studies showing no effect. (Note that asymmetry could also occur if small studies tend to have larger effect size,<sup>18</sup> so the conclusion of publication bias should be a cautious one.)<sup>19</sup>

The funnel plot has some limitations; for example, it can sometimes be difficult to detect asymmetry by eye.<sup>20</sup> To help with this, formal statistical methods have been developed to test for heterogeneity. Egger's regression test<sup>14</sup> has been widely used to test for publication bias. It tests whether small studies tend to have larger effect sizes than would be expected (implying that small studies with small

effects have not been published). Another regression test, which in some circumstances may be better than Egger's test, has been proposed.<sup>15</sup> However, care is needed in the interpretation of the findings whatever test has been used. There is currently no clear direction in recent literature to indicate when to use each test.

## Sensitivity analyses

Because of the many ways in which decisions taken about selection, inclusion and aggregation of data may affect the main findings, it is usual for meta-analysts to carry out some **sensitivity analysis**. This explores the ways in which the main findings are changed by varying the approach to aggregation. A good sensitivity analysis will explore, among other things, the effect of excluding various categories of studies; for example, unpublished studies or those of poor quality. It may also examine how consistent the results are across various subgroups (perhaps defined by patient group, type of intervention or setting). In meta-analyses without sensitivity analyses, the reader has to make guesses about the likely impact of these important factors on the key findings.

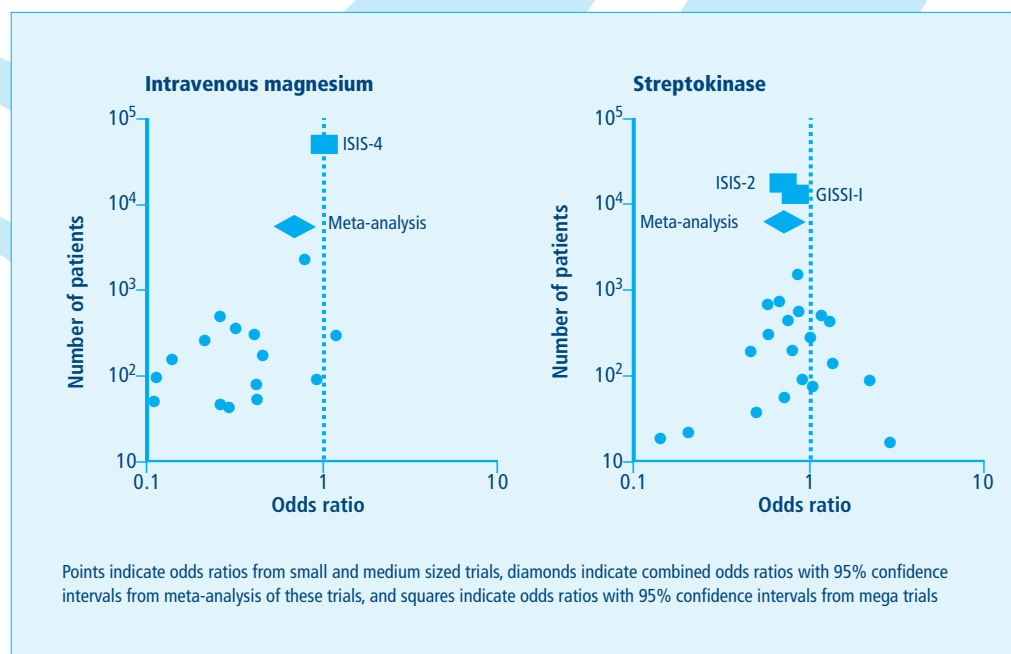


Figure 1. Funnel plots used to identify possible missing studies<sup>17</sup>

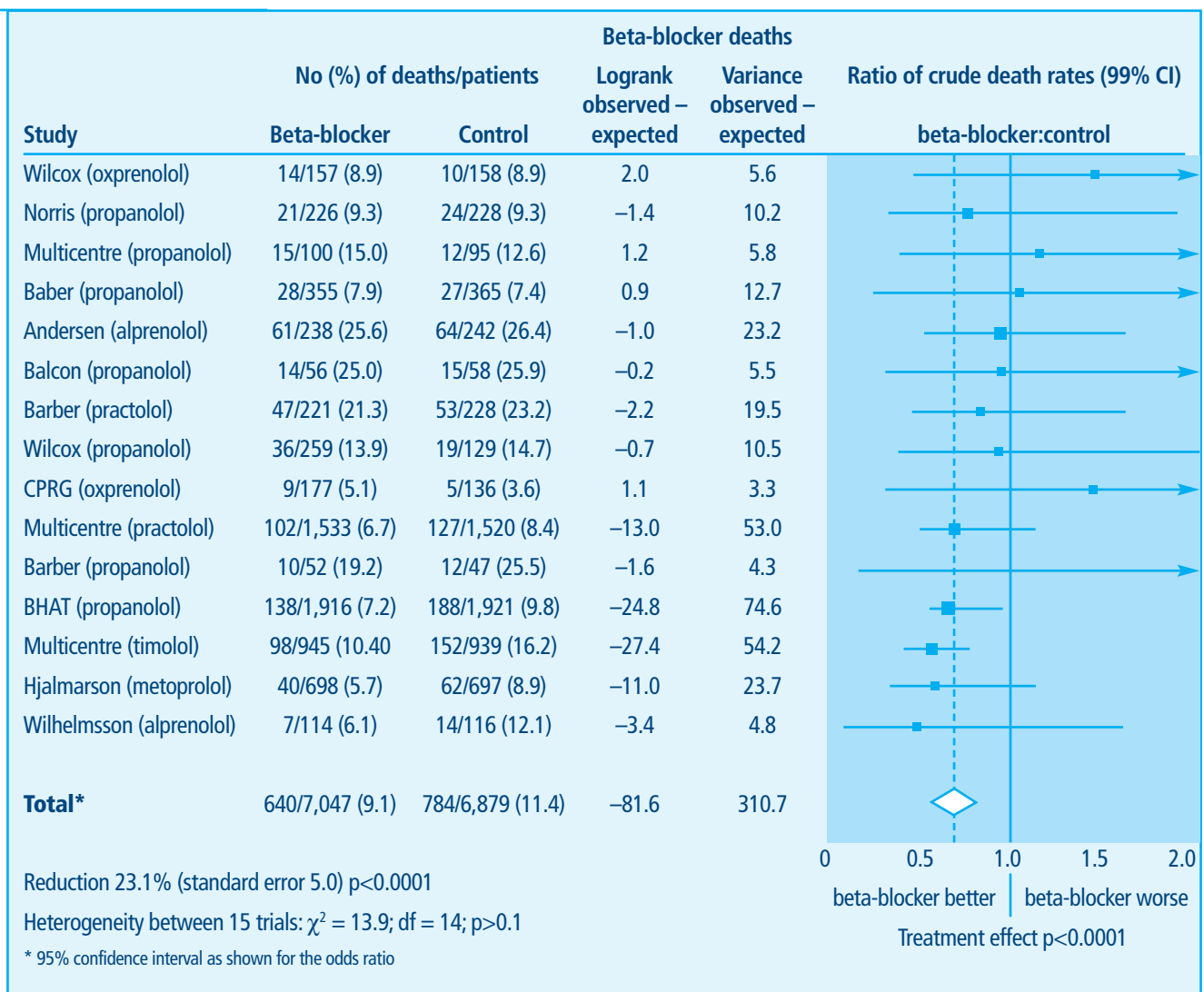


Figure 2. Presentation of the findings from a meta-analysis<sup>21</sup>

## Presenting the findings

### Forest plot

The usual way of displaying data from a meta-analysis is by a pictorial representation (sometimes known as a **Forest plot**). An example is shown in Figure 2.<sup>21</sup> This displays the findings from each individual study as a blob or square,<sup>21</sup> with squares towards the left side indicating the new treatment to be better, whereas those on the right indicate the new treatment to be less effective. The size of the blob or square is proportional to the precision of the study (roughly speaking, the sample size). A horizontal line (usually the 95% confidence interval) is drawn around each of the studies' squares to represent the uncertainty of the estimate of the treatment effect. The aggregate

effect size obtained by combining all the studies is usually displayed as a diamond.

### Heterogeneity

A major concern about meta-analyses is the extent to which they mix studies that are different in kind (heterogeneity). One widely quoted definition of meta-analysis is: **'a statistical analysis which combines or integrates the results of several independent clinical trials considered by the analyst to be "combinable"'**.<sup>22</sup> The key difficulty lies in deciding which sets of studies are 'combinable'. Clearly, to get a precise answer to a specific question, only studies that exactly match the question should be included. Unfortunately, studies can differ on the types of patient studied (disease

# What is meta-analysis?

severity or co-morbidity), the nature of local healthcare facilities, the intervention given and the primary endpoint (death, disease, disability). These systematic differences between studies can influence the amount of treatment benefit (the effect size), leading to heterogeneity between studies.

Meta-analyses should test for the existence of heterogeneity. A test which was commonly used is Cochrane's Q, a statistic based on the chi-squared test.<sup>23</sup> Unfortunately, this test is thought to have low power; that is, it may sometimes fail to detect heterogeneity when it is present. To try to overcome this, a second test, the I<sup>2</sup> statistic, was developed.<sup>24</sup> This test seems attractive because it scores heterogeneity between 0% and 100%. Further, a rule of thumb was proposed, with 25% corresponding to low heterogeneity, 50% to moderate and 75% to high. Subsequent research suggests that this test may also have low power,<sup>23</sup> so it too has to be interpreted cautiously.

The presence or absence of heterogeneity influences the subsequent method of analysis. If heterogeneity is absent, then the analysis employs what is termed fixed-effects modelling. This assumes the size of treatment effect is the same (fixed) across all studies and the variation seen between studies is due only to the play of chance. Random-effects models assume that the treatment effect really does vary between studies. Such models tend to increase the variance of the summary measure, making it more difficult to obtain significant results. When the amount of heterogeneity is large, it may even be inappropriate to calculate an overall summary measure of effect size. Unfortunately, there is no reliable objective measure to decide when pooling is appropriate. Thus, a rule of thumb is given above. The technique of meta-regression is introduced because it provides one way of overcoming the problem of heterogeneity.

## Meta-regression

When heterogeneity is detected, it is important to investigate what may have

caused it. Meta-regression is a technique which allows researchers to explore which types of patient-specific factors or study design factors contribute to the heterogeneity. The simplest type of meta-regression uses summary data from each trial, such as the average effect size, average disease severity at baseline, and average length of follow-up. This approach is valuable, but it has only limited ability to identify important factors. In particular, it struggles to identify which patient features are related to the size of treatment effect.<sup>25</sup> Fortunately, another approach, using individual patient data, will give answers to the important question: what types of patients are most likely to benefit from this treatment? Using individual patient data allows much greater flexibility for the analysis, and issues can be explored that were not covered in the published trials. However, obtaining the original patient data from each of the trials is challenging.

## Limitations

Assessments of the quality of systematic reviews and meta-analysis often identify limitations in the ways they were conducted.<sup>26,27</sup> Flaws in meta-analysis can arise through failure to conduct any of the steps in data collection, analysis and presentation described above. To summarise:

- Was the search strategy comprehensive and likely to avoid bias in the studies identified for inclusion?
- Was publication bias assessed?
- Was the quality of the individual studies assessed using an appropriate checklist of criteria?
- Was combined effect size calculated using appropriate statistical methods?
- Was heterogeneity considered and tested for?

## Conflict with new experimental data

Meta-analyses seek new knowledge from existing data. One test of the validity of this new knowledge is to compare the results from meta-analyses with

subsequent findings from large-scale, well-conducted, randomised controlled trials (so-called '**mega trials**'). The results of such comparisons have, so far, been mixed – good agreement in the majority of cases but some discrepancies in others.<sup>28,29</sup> For example, one such exercise led to publication of a paper subtitled 'Lessons from an "effective, safe, simple intervention" that wasn't' (use of intravenous magnesium after heart attacks).<sup>17</sup> With the benefit of hindsight, the flaws in meta-analyses that have been subsequently contradicted by data from mega trials can often be uncovered. Such post-mortems have led to a number of methodological improvements (such as funnel plots) and a greater understanding of the pitfalls outlined above.

## Conclusion

Meta-analyses offer a systematic and quantitative approach to synthesising evidence to answer important therapeutic questions. Nonetheless, pitfalls abound in the execution of meta-analyses and they are fundamentally limited by the quality of the underlying studies (the so-called GIGO principle of '**garbage in, garbage out**'). For healthcare managers and clinicians, careful reviewing of published meta-analyses and a balanced assessment of their deficiencies is likely to become an increasingly important way of resolving therapeutic uncertainty.

The field of meta-analysis is in a period of rapid development, with theoretical and methodological advances as well as the findings from empirical research. However, two recent books provide excellent reviews of current knowledge and offer far more detail than can be presented in this short briefing document. See below for further reading.

## References

1. Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. Treatments for myocardial infarction. *JAMA* 1992; **268**: 240–248.
2. Mulrow CD. Rationale for systematic reviews. *BMJ* 1994; **309**: 597–599.
3. Bailar JC 3rd. The promise and problems of meta-analysis. *N Engl J Med* 1997; **337**: 559–561.
4. Shea B, Dubé C, Moher D. Assessing the quality of reports of systematic reviews: the QUOROM statement compared to other tools. In: Egger M, Smith GD, Altman DG (eds). *Systematic Reviews in Health Care: Meta-analysis in context*, 2nd edn. London: BMJ Publishing Group, 2001.
5. Moher D, Cook DJ, Eastwood S *et al*. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses. *Lancet* 1999; **354**: 1896–1900.
6. Higgins JPT, Green S. *Cochrane Handbook for Systematic Reviews of Interventions* 4.2.6. <http://cochrane.org/resources/handbook/Handbook4.2.6Sep2006.pdf> (last accessed 28 October 2008)
7. Cook DJ, Sackett DL, Spitzer WO. Methodologic guidelines for systematic reviews of randomized control trials in health care from the Potsdam Consultation on Meta-Analysis. *J Clin Epidemiol* 1995; **48**: 167–171.
8. Jüni P, Altman DG, Egger M. Assessing the quality of randomised controlled trials. In: Egger M, Smith GD, Altman DG (eds). *Systematic Reviews in Health Care: Meta-analysis in context*, 2nd edn. London: BMJ Publishing Group, 2001.
9. Chalmers TC, Smith H Jr, Blackburn B *et al*. A method for assessing the quality of a randomized control trial. *Control Clin Trials* 1981; **2**: 31–49.
10. Jadad AR, Moore RA, Carroll D *et al*. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials* 1996; **17**: 1–12.
11. Davies HT, Crombie IK, Tavakoli M. When can odds ratios mislead? *BMJ* 1998; **316**: 989–991.
12. Deeks JJ, Altman DG, Bradburn MJ. Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. In: Egger M, Smith GD, Altman DG (eds). *Systematic Reviews in Health Care: Meta-analysis in context*, 2nd edn. London: BMJ Publishing Group, 2001.
13. Thornton A, Lee P. Publication bias in meta-analysis: its causes and consequences. *J Clin Epidemiol* 2000; **53**: 207–216.
14. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997; **315**: 629–634.
15. Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. Comparison of two methods to detect publication bias in meta-analysis. *JAMA* 2006; **295**: 676–680.
16. Sterne JA, Egger M. Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. *J Clin Epidemiol* 2001; **54**: 1046–1055.
17. Egger M, Smith GD. Misleading meta-analysis. *BMJ* 1995; **311**: 752–754.
18. Sterne JA, Egger M, Smith GD. Systematic reviews in health care: Investigating and dealing with publication and other biases in meta-analysis. *BMJ* 2001; **323**: 101–105.
19. Lau J, Ioannidis JP, Terrin N, Schmid CH, Olkin I. The case of the misleading funnel plot. *BMJ* 2006; **333**: 597–600.
20. Terrin N, Schmid CH, Lau J. In an empirical evaluation of the funnel plot, researchers could not visually identify publication bias. *J Clin Epidemiol* 2005; **58**: 894–901.
21. Lewis S, Clarke M. Forest plots: trying to see the wood and the trees. *BMJ* 2001; **322**: 1479–1480.
22. Huque MF. Experiences with meta-analysis in NDA submissions. *Proc Biopharm Section Am Stat Assoc* 1988; **2**: 28–33.
23. Ioannidis JP, Patsopoulos NA, Evangelou E. Uncertainty in heterogeneity estimates in meta-analyses. *BMJ* 2007; **335**: 914–916.
24. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003; **327**: 557–560.
25. Schmid CH, Stark PC, Berlin JA, Landais P, Lau J. Meta-regression detected associations between heterogeneous treatment effects and study-level, but not patient-level, factors. *J Clin Epidemiol* 2004; **57**: 683–697.
26. Biondi-Zoccai GG, Lotriente M, Abbate A *et al*. Compliance with QUOROM and quality of reporting of overlapping meta-analyses on the role of acetylcysteine in the prevention of contrast associated nephropathy: case study. *BMJ* 2006; **332**: 202–209.
27. Moja LP, Telaro E, D'Amico R *et al*. Assessment of methodological quality of primary studies by systematic reviews: results of the metaquality cross sectional study. *BMJ* 2005; **330**: 1053.
28. LeLorier J, Gregoire G, Benhaddad A, Lapierre J, Derderian F. Discrepancies between meta-analyses and subsequent large randomized, controlled trials. *N Engl J Med* 1997; **337**: 536–542.
29. Villar J, Carroli G, Belizan JM. Predictive ability of meta-analyses of randomised controlled trials. *Lancet* 1995; **345**: 772–776.

## Further reading

1. Egger M, Smith GD, Altman DG (eds). *Systematic Reviews in Health Care: Meta-analysis in context*, 2nd edn. London: BMJ Publishing Group, 2001.
2. Petticrew M, Roberts H. *Systematic Reviews in the Social Sciences: A practical guide*. Oxford: Blackwell Publishing, 2006.

# What is meta-analysis?

First edition published 2001  
Authors: Iain K Crombie and  
Huw TO Davies

This publication, along with  
the others in the series, is  
available on the internet at  
[www.whatisseries.co.uk](http://www.whatisseries.co.uk)  
The data, opinions and statements  
appearing in the article(s) herein  
are those of the contributor(s)  
concerned. Accordingly, the  
sponsor and publisher, and their  
respective employees, officers  
and agents, accept no liability  
for the consequences of any such  
inaccurate or misleading data,  
opinion or statement.

Published by Hayward Medical  
Communications, a division of  
Hayward Group Ltd.  
Copyright © 2009 Hayward  
Group Ltd.  
All rights reserved.



Supported by *sanofi-aventis*