© International Epidemiological Association 2002 Printed in Great Britain *International Journal of Epidemiology* 2002;**31**:1–5

EDITORIAL

# Where now for meta-analysis?

**Matthias Egger, Shah Ebrahim and George Davey Smith**

In the short time since its introduction, meta-analysis, the statistical pooling of the results from independent but 'combinable' studies, has established itself as an influential branch of clinical epidemiology and health services research, with hundreds of meta-analyses published in the medical literature each year.[1] This issue of the *International Journal of Epidemiology* contains several papers[2–9] that address methodological issues in meta-analytic research, a review article on where we stand with systematic reviews in observational epidemiology[10] and three meta-analyses of observational studies.[11–13] Publication of a themed issue on meta-analysis by an epidemiological journal begs several questions: Where does meta-analysis come from? Does it deserve the attention it is currently getting? And where should it be going next?

The statistical basis of meta-analysis reaches back to the 17th century when, in astronomy, intuition and experience suggested that combinations of data might be better than attempts to select amongst them.[14] In the 20th century the distinguished statistician Karl Pearson (Figure 1), was, in 1904, probably the first medical researcher using formal techniques to combine data from different studies when examining the preventive effect of serum inoculations against enteric fever.[15] However, such techniques were not widely used in medicine for many years to come. In contrast to medicine, the social sciences and in particular psychology and educational research, demonstrated early interest in the synthesis of research findings. In 1976 the psychologist Gene Glass coined the term 'meta-analysis' in a paper entitled 'Primary, Secondary and Meta-analysis of Research'.[16] Three years later the British physician and epidemiologist Archie Cochrane drew attention to the fact that people who want to make informed decisions about health care do not have ready access to reliable reviews of the available evidence.[17] In the 1980s meta-analysis became increasingly popular in medicine, particularly in the clinical trial fields of cardiovascular disease, oncology, and perinatal care. In the 1990s the foundation of The Cochrane Collaboration,[18] an international network of health care professionals who prepare and regularly update systematic reviews ('Cochrane Reviews') facilitated the conduct of meta-analyses in all areas of health care.

The achievements of meta-analysis in the realm of clinical trial research are impressive. First, meta-analysis helped to overcome the problem first identified by Pearson, that 'any of the groups ... are far too small to allow of any definite opinion being formed at all, having regard to the size of the probable error involved'. Although the size of trials published in general health care journals has been increasing since 1948 (see the paper by McDonald *et al.*[7] in this issue), many trials fail to detect, or exclude with certainty, a modest but clinically relevant difference in the effects of two therapies. This means that the conclusions from several small trials will often be contradictory and confuse those seeking guidance. The meta-analytic approach may overcome this problem by combining trials evaluating the same intervention in a number of smaller, but comparable, trials. Early examples include meta-analyses of trials of beta-blockers in secondary prevention after myocardial infarction,[19] of a short course of corticosteroids given to women about to give birth prematurely[20] and of adjuvant tamoxifen in early breast cancer.[21] A welcome effect of the surge of systematic reviews and meta-analysis, and of evidence-based medicine in general, is the dismantling of the magnificence and splendour of the full professor, who was used to argue casually based on status and opinion, but is now confronted by well-informed junior members of staff and consumers of health services.

**Figure 1** Distinguished statistician Karl Pearson is seen as the first medical researcher to use formal techniques to combine data from different studies

Second, meta-analysis may highlight areas where there is a lack of adequate evidence and thus identify where further

Editorial Office, International Journal of Epidemiology, Department of Social Medicine, University of Bristol, UK.

Correspondence: Matthias Egger, Department of Social Medicine, Canynge Hall, Whiteladies Road, Bristol BS8 2PR, UK. E-mail: m.egger@bristol.ac.uk

studies are needed. For example, a period of starvation is common practice after gastrointestinal surgery, but a recent meta-analysis[22] of randomized controlled trials concluded that keeping patients 'nil by mouth' may do more harm than good, and that a large trial is required to clarify this issue. About half of Cochrane reviews and a fifth of meta-analyses published in medical journals conclude that the evidence is inappropriate and that a large trial is needed.[23] Indeed, as Iain Chalmers pointed out, systematic reviews of existing trials and registers of ongoing trials should be seen as prerequisites for scientific and ethical trial design.[24]

Third, meta-analyses offer a sounder basis for subgroup analyses, particularly if they are based on individual participant data.[25] For example, the meta-analysis of individual patient data from 55 trials of tamoxifen in operable breast cancer showed that the benefit of tamoxifen was much smaller and non-significant in women reported to have oestrogen receptor negative disease.[26] Based on these findings, oestrogen receptor status is now used to inform treatment decisions.

Finally, the realization that the results from meta-analysis are not always trustworthy[27,28] led to research into the numerous ways in which bias may be introduced, and the development of methods to detect the presence of such bias. For example, several studies have examined the influence of unpublished trials, trials published in languages other than English, and of trial quality on the results of meta-analyses of randomized controlled trials. The authors used a 'meta-epidemiological' approach[29] and considered collections of meta-analyses in which component trials had been classified according to characteristics such as publication status or study quality, thus ensuring that the treatment effects are compared only between studies in the same meta-analysis.[30] Figure 2 shows a 'meta-meta-analysis' of these studies which includes the study by Jüni *et al.* published in this issue, on language bias.[31] Combined results indicate that, on average, unpublished trials will underestimate treatment effects by about 10%, trials published in languages other than English will overestimate effects by the same amount and trials not indexed in MEDLINE will overestimate effects by about 5%. Trials with inadequate or unclear concealment of allocation and trials that are not double blind overestimate treatment effects by about 30% and 15%, respectively. The quality of trials thus appears to be a more important source of bias than the reporting and dissemination of trials. However, as pointed out by Clarke in his commentary,[32] the influence of language bias and other reporting biases may still be large in meta-analyses based on few trials. Also, the size of effects will differ across individual meta-analyses, perhaps depending on specialty, type of active and control intervention and trial design.

Considerable progress has also been made in our understanding of how best to detect bias, and deal with bias in meta-analysis, using graphical and statistical methods.[33] In this issue Higgins and Spiegelhalter[4] revisit the clinical trials of the effect of magnesium infusion in myocardial infarction, a well-known example where bias may explain the discrepancy between meta-analyses of small trials which showed a clear treatment effect[34] and the subsequent large Fourth International Study of Infarct Survival (ISIS-4)[35] which showed no effect. Using Bayesian methods the authors show how scepticism can be formally incorporated into the analysis and that such an approach would have led to appropriate caution before the results of the mega-trial became available. In his commentary,[36] Woods argues that the degree of scepticism required would have been extreme, considering the laboratory studies which made a beneficial effect of magnesium biologically plausible. Woods offers an alternative explanation for the discrepant findings between ISIS-4 and the smaller trials: myocardial protection by magnesium is abolished when treatment is delayed until after reperfusion has occurred, as was the case in ISIS-4, but not the smaller trials.

This debate must continue, but the magnesium example, and other meta-analyses that were later contradicted by single large

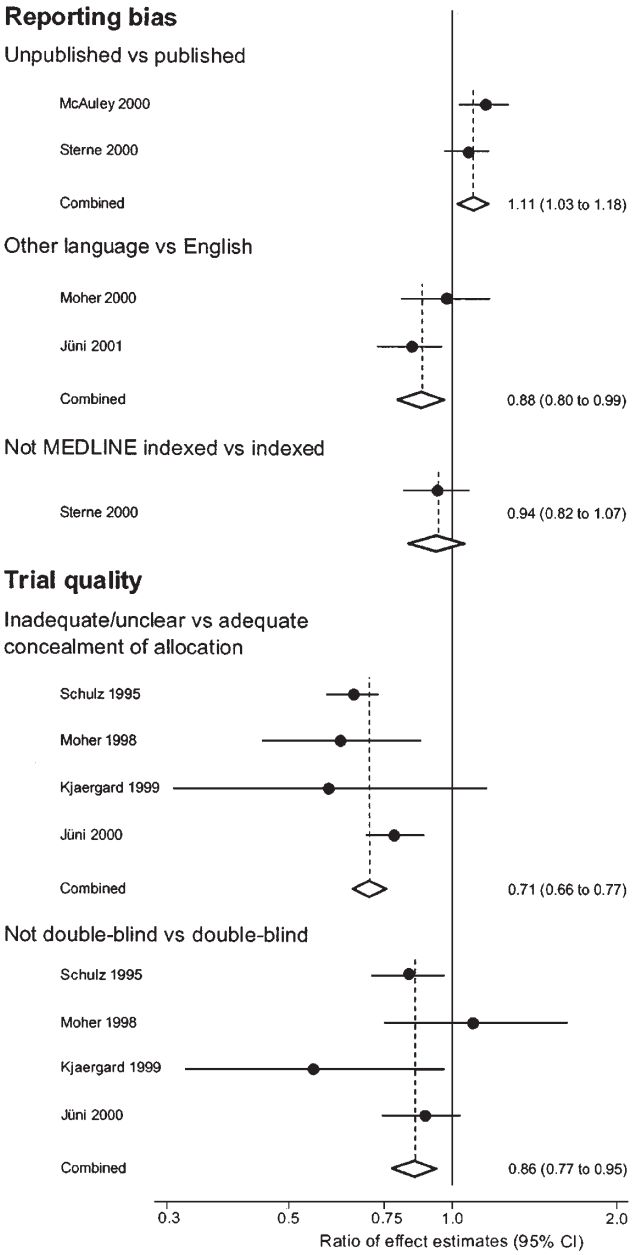


**Figure 2** Meta-analysis of empirical studies of reporting bias and trial quality. All studies compared estimates of treatment effects within a large number of meta-analyses and calculated ratios of effect estimates for this purpose. A ratio of estimates below 1 indicates that trials with the characteristic (for example published in a language other than English) showed a more beneficial treatment effect

trials,[37] has certainly demonstrated that the pooling of trials in meta-analysis may not always be appropriate. It is therefore important to distinguish between systematic reviews and meta-analysis: it is always appropriate and desirable to systematically review a body of data, but it may sometimes be inappropriate, or even misleading, to statistically pool results from separate studies.[38] Indeed, it is our impression that reviewers often find it hard to resist the temptation of combining studies when such meta-analysis is questionable or clearly inappropriate. This point is particularly pertinent to systematic reviews of observational studies. A clear distinction should be made between meta-analysis of randomized controlled trials and meta-analysis of epidemiological studies: consider a set of trials of high methodological quality that examined the same intervention in comparable patient populations: each trial will provide an unbiased estimate of the same underlying treatment effect. The variability that is observed between the trials can confidently be attributed to random variation and meta-analysis should provide an equally unbiased estimate of the treatment effect, with an increase in the precision of this estimate. A fundamentally different situation arises in the case of epidemiological studies, for example case-control studies, cross-sectional studies or cohort studies. Due to the effects of confounding and bias, such observational studies may produce estimates of associations that deviate from the true causal effects beyond what can be attributed to chance. Combining a set of epidemiological studies will thus often provide spuriously precise, but biased, estimates of associations.[39] The thorough consideration of heterogeneity between observational study results, in particular of possible sources of confounding and bias, will generally provide more insights than the mechanistic calculation of an overall measure of effect. This is illustrated by the systematic review of epidemiological studies of homocysteine and the risk of coronary heart disease published in this issue.[11] The association was weak for cohort studies (combined odds ratio [OR] = 1.06, 95% CI : 0.99–1.13), stronger for nested case-control studies (OR = 1.23, 95% CI : 1.07–1.41) and strongest for standard case-control studies (OR = 1.70, 95% CI : 1.50–1.93), as shown in the Figure in Clarke's commentary.[36] The strength of the association thus varies inversely with the strength of the study design, which surely must be taken into account when interpreting these findings.

The importance of different sources of bias will vary across different areas of epidemiological enquiry. For example, confounding and differential measurement error is a serious problem in studies of exposures that are closely linked to lifestyle, for example dietary intake of beta-carotene, but may be of considerably less relevance in genetic epidemiology.[40] Publication bias, conversely, may be a particular problem in studies of genetic factors. For example, several meta-analyses of small case-control studies found substantial associations between the angiotensin converting enzyme (ACE) insertion/deletion polymorphism and the risk of myocardial infarction.[41,42] When plotting the odds ratios from the 19 studies included in Agerholm-Larsen's[43] analysis against their standard error in a 'funnel plot', it is clear that the effect is large in small case-control studies but only modest in larger studies (Figure 3).[43] The name 'funnel plot' is based on the fact that effect estimates from small studies will scatter more widely at the bottom of the graph, with the spread narrowing among larger studies. In the absence of bias the plot will thus resemble a symmetrical inverted funnel. The degree of asymmetry observed for the ACE gene polymorphism studies, which includes the studies in Whites published up to 1998, is unlikely to be due to chance ($P$ = 0.033 by regression test[37]). Based on these findings, the results of the large ISIS genetic study,[44] which was based on 4629 myocardial infarction cases and 5934 controls and published in 2000, are hardly surprising: the estimated risk ratio was 1.10, with confidence intervals (1.00–1.21) that exclude the effects seen in the earlier meta-analyses.[41,42]

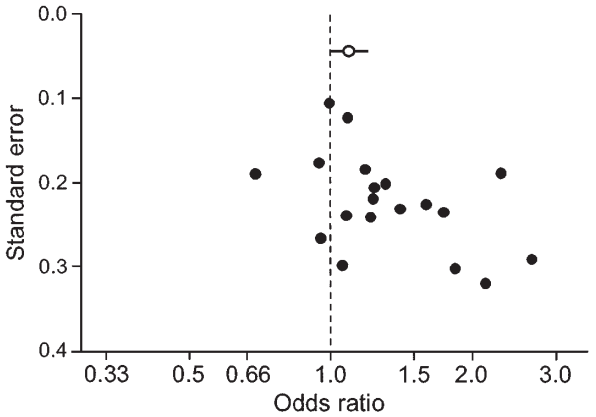


**Figure 3** Funnel plot of 19 case-control studies of angiotensin converting enzyme (ACE) gene polymorphism in coronary heart disease (from Agerholm-Larsen *et al.*[43]). Odds ratios comparing the ACE *DD* genotype with the *ID*/*II* genotype are plotted against their standard error. The open circle and horizontal line shows the point estimate and 95% CI of the ISIS genetic study (from Keavney *et al.*[44])

Where now for meta-analysis? In her review article,[10] Dickersin takes observational epidemiology to task for being far behind, and argues that methodological research is urgently required in this area. We agree, and will continue to be interested in publishing such research. The *International Journal of Epidemiology* will also participate actively in the development of reporting guidelines for epidemiological studies, similar to the Consolidated Standard for Reporting Trials (CONSORT).[45] Such guidelines are required to facilitate the assessment of the quality of epidemiological studies and will be helpful not only to systematic reviewers and meta-analysts, but also to editors of, and referees for, epidemiological journals. Dickersin's analysis of the instructions for authors on the preparation of systematic reviews made us realize that our instructions urgently need updating, and this process has now started. Finally, Dickersin is critical of journal editors who do not treat systematic reviews and meta-analyses as original research, thus depriving those who specialize in this area from a form of academic reward. We stress that at the *IJE* we do consider well-conducted systematic reviews and meta-analyses as original research and publish them as such. However, we believe that there continues to be a place for reviews that express an informed opinion, as Dickersin does in her review,[10] and (we hope) we do in this editorial.

## Acknowledgements

We thank Ken Schulz and Lise Kjaergard for kindly providing unpublished data. We are grateful to the MRC Health Services Research Collaboration for funding a workshop in November

2000, which helped identify topical issues in meta-analysis. Bristol is the lead centre of the MRC Health Services Research Collaboration.

## References


[1] Egger M, Davey Smith G, O'Rourke K. Rationale, potentials and promise of systematic reviews. In: Egger M, Davey Smith G, Altman DG (eds). *Systematic Reviews in Health Care: Meta-Analysis in Context*. London: BMJ Books, 2001, pp.23–42.

[2] Furukawa TA, Guyatt GH, Griffith LE. Can we individualize the 'number needed to treat'? An empirical study of summary effect measures in meta-analyses. *Int J Epidemiol* 2002;**31:**72–76.

[3] Song F, Khan KS, Dinnes J, Sutton AJ. Asymmetric funnel plots and publication bias in meta-analyses of diagnostic accuracy. *Int J Epidemiol* 2002;**31:**88–95.

[4] Higgins JPT, Spiegelhalter DJ. Being sceptical about meta-analyses: a Bayesian perspective on magnesium trials in myocardial infarction. *Int J Epidemiol* 2002;**31:**96–104.

[5] Vale CL, Tierney JF, Stewart LA. Effects of adjusting for censoring on meta-analyses of time-to-event outcomes. *Int J Epidemiol* 2002;**31:** 107–11.

[6] Clark OAC, Castro AA. Searching the LILACS database improves systematic reviews. *Int J Epidemiol* 2002;**31:**112–14.

[7] McDonald S, Westby M, Clarke M, Lefebvre C and the Cochrane Centres' Working Group on 50 Years of Randomized Trials. Number and size of randomized trials reported in general health care journals from 1948 to 1997. *Int J Epidemiol* 2002;**31:**125–27.

[8] Robinson KA, Dickersin K. Development of a highly sensitive search strategy for the retrieval of reports of controlled trials using PubMed. *Int J Epidemiol* 2002;**31:**150–53.

[9] Elbourne DR, Altman DG, Higgins JPT, Curtin F, Worthington HV, Vail A. Meta-analyses involving cross-over trials: methodological issues. *Int J Epidemiol* 2002;**31:**140–49.

[10] Dickersin K. Systematic reviews in epidemiology: why are we so far behind? *Int J Epidemiol* 2002;**31:**6–12.

[11] Ford ES, Smith SJ, Stroup DF, Steinberg KK, Mueller PW, Thacker SB. Homocyst(e)ine and cardiovascular disease: a systematic review of the evidence with special emphasis on case-control studies and nested case-control studies. *Int J Epidemiol* 2002;**31:**59–70.

[12] Missmer SA, Smith-Warner SA, Spiegelman D *et al.* Meat and dairy food consumption and breast cancer: a pooled analysis of cohort studies. *Int J Epidemiol* 2002;**31:**78–85.

[13] Fischbach LA, Goodman KJ, Feldman M, Aragaki C. Sources of variation of *Helicobacter pylori* treatment success in adults worldwide: a meta-analysis. *Int J Epidemiol* 2002;**31:**128–39.

[14] Plackett RL. Studies in the history of probability and statistics: VII. The principle of the arithmetic mean. *Biometrika* 1958;**45:**130–35.

[15] Pearson K. Report on certain enteric fever inoculation statistics. *Br Med J* 1904;**3:**1243–46.

[16] Glass GV. Primary, secondary and meta-analysis of research. *Educ Res* 1976;**5:**3–8.

[17] Cochrane AL. 1931–1971: a critical review, with particular reference to the medical profession. In: *Medicines for the Year 2000*. London: Office of Health Economics, 1979.

[18] Bero L, Rennie D. The Cochrane Collaboration. Preparing, maintaining, and disseminating systematic reviews of the effects of health care. *JAMA* 1995;**274:**1935–38.

[19] Yusuf S, Peto R, Lewis J, Collins R, Sleight P. Beta blockade during and after myocardial infarction: an overview of the randomized trials. *Prog Cardiovasc Dis* 1985;**17:**335–71.

[20] Crowley P. Corticosteroids prior to preterm delivery. In: Enkin MW, Keirse MJNC, Renfrew MJ, Neilson JP (eds). *Pregnancy and Childbirth Module*. Oxford: Update Software, 1994, p.2955.

[21] Early Breast Cancer Trialists' Collaborative Group. Effects of adjuvant tamoxifen and of cytotoxic therapy on mortality in early breast cancer. An overview of 61 randomized trials among 28 896 women. *N Engl J Med* 1988;**319:**1681–92.

[22] Lewis SJ, Egger M, Sylvester PA, Thomas S. Early enteral feeding versus 'nil by mouth' after gastrointestinal surgery: systematic review and meta-analysis of controlled trials. *Br Med J* 2001;**323:**773–76.

[23] Schwarzer G, Antes G, Tallon D, Egger M. *Review Publication Bias? Matched Comparative Study of Cochrane and Journal Meta-analyses*. BMC Meeting Abstracts: 9th International Cochrane Colloquium 2001, 1: pc142.

[24] Chalmers I. Using systematic reviews and registers of ongoing trials for scientific and ethical trial design, monitoring, and reporting. In: Egger M, Davey Smith G, Altman DG (eds). *Systematic Reviews in Health Care: Meta-analysis in Context.* London: BMJ Books, 2001, pp.429–43.

[25] Clarke MJ, Stewart LA. Systematic reviews of evaluations of prognostic variables. In: Egger M, Davey Smith G, Altman DG (eds). *Systematic Reviews in Health Care: Meta-Analysis in Context*. London: BMJ Books, 2001, pp.109–21.

[26] Early Breast Cancer Trialists' Collaborative Group. Tamoxifen for early breast cancer: an overview of the randomised trials. *Lancet* 1998;**351:** 1451–67.

[27] Thompson SG, Pocock SJ. Can meta-analysis be trusted? *Lancet* 1991; **338:**1127–30.

[28] Egger M, Davey Smith G. Misleading meta-analysis. Lessons from 'an effective, safe, simple' intervention that wasn't. *Br Med J* 1995;**310:** 752–54.

[29] Naylor CD. Meta-analysis and the meta-epidemiology of clinical research. *Br Med J* 1997;**315:**617–19.

[30] Sterne JAC, Jüni P, Schulz KF, Altman DG, Bartlett C, Egger M. Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. *Stat Med* 2002; (In press).

[31] Jüni P, Holenstein F, Sterne J, Bartlett C, Egger M. Direction and impact of language bias in meta-analyses of controlled trials: empirical study. *Int J Epidemiol* 2002;**31:**115–23.

[32] Clarke M. Searching for trials for systematic reviews: what difference does it make? *Int J Epidemiol* 2002;**31:**123–24.

[33] Sterne JA, Egger M, Davey Smith G. Systematic reviews in health care: investigating and dealing with publication and other biases in meta-analysis. *Br Med J* 2001;**323:**101–05.

[34] Teo KK, Yusuf S, Collins R, Held PH, Peto R. Effects of intravenous magnesium in suspected acute myocardial infarction: overview of randomised trials. *Br Med J* 1991;**303:**1499–503.

[35] Collaborative Group. ISIS-4: A randomised factorial trial assessing early oral captopril, oral mononitrate, and intravenous magnesium sulphate in 58 050 patients with suspected acute myocardial infarction. *Lancet* 1995;**345:**669–87.

[36] Clarke R. An updated review of the published studies of homocysteine and cardiovascular disease. *Int J Epidemiol* 2002;**31:**70–71.

[37] Egger M, Davey Smith G, Schneider M, Minder CE. Bias in meta-analysis detected by a simple, graphical test. *Br Med J* 1997;**315:** 629–34.

[38] O'Rourke K, Detsky AS. Meta-analysis in medical research: strong encouragement for higher quality in individual research efforts. *J Clin Epidemiol* 1989;**42:**1021–24.

[39] Egger M, Schneider M, Davey Smith G. Spurious precision? Meta-analysis of observational studies. *Br Med J* 1998;**316:**140–45.

[40] Clayton D, McKeigue PM. Epidemiologic methods for studying genes and environmental factors in complex diseases. *Lancet* 2001;**358:** 1356–60.

[41] Samani NJ, Thompson JR, O'Toole L, Channer K, Woods KL. A meta-analysis of the association of the deletion allele of the angiotensin-converting enzyme gene with myocardial infarction. *Circulation* 1996; **94:**708–12.

[42] Staessen JA, Wang JG, Ginocchio G *et al.* The deletion/insertion polymorphism of the angiotensin converting enzyme gene and cardiovascular-renal risk. *J Hypertens* 1997;**15:**1579–92.

[43] Agerholm-Larsen B, Nordestgaard BG, Tybjaerg-Hansen A. ACE gene polymorphism in cardiovascular disease: meta-analyses of small and large studies in whites. *Arterioscler Thromb Vasc Biol* 2000;**20:**484–92.

[44] Keavney B, McKenzie C, Parish S *et al.* Large-scale test of hypothesised associations between the angiotensin-converting-enzyme insertion/deletion polymorphism and myocardial infarction in about 5000 cases and 6000 controls. International Studies of Infarct Survival (ISIS) Collaborators. *Lancet* 2000;**355:**434–42.

[45] Altman DG, Schulz KF, Moher D *et al.* The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001;**134:**663–94.