Uses and abuses of meta-analysis

Matthias Egger, George Davey Smith and Jonathan A C Sterne

Matthias Egger
MD MSc MFPHM

**George Davey
Smith** MD MSc
FFPHM

**Jonathan AC
Sterne** PhD

MRC Health
Services Research
Collaboration,
Department of
Social Medicine,
University of Bristol

ABSTRACT – Meta-analysis, the statistical combination of results from several studies to produce a single estimate of the effect of a treatment, continues to attract controversy. We illustrate the potentials and pitfalls of meta-analysis of controlled clinical trials. Cumulative meta-analysis demonstrates that this technique could prevent delays in the introduction of effective treatments. Meta-analyses are, however, liable to numerous biases both at the level of the individual trial ('garbage in, garbage out') and the dissemination of trial results (publication bias). We argue that meta-analysis should be performed only within the framework of systematic reviews – that is, reviews prepared using a systematic approach to minimise bias and address the combinability of studies.

> *Many 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.*
>
> K Pearson, *British Medical Journal,* 1904[1]

> *Meta-analysis begins with scientific studies, usually performed by academics or government agencies, and sometimes incomplete or disputed. The data from the studies are then run through computer models of bewildering complexity, which produce results of implausible precision.*
>
> B Davis, *Wall Street Journal,* 1992[2]

The distinguished statistician Karl Pearson was probably the first medical researcher to perform meta-analysis, a statistical analysis of the results from independent studies to produce a single estimate of a treatment effect. The rationale for pooling studies put forward by Pearson in 1904 in his account of the preventive effect of serum inoculations against enteric fever[1] is still one of the main reasons for conducting meta-analyses today. Noticeably, Pearson's conclusions did not go unchallenged, and heated correspondence in the *British Medical Journal* followed his publications[3]. The second quote above demonstrates that meta-analysis has continued to attract controversy since Pearson's time. In this article we will give a brief overview of the potentials and pitfalls of meta-analysis of clinical trials. A comprehensive account of the issues and methods involved in systematic reviews and meta-analysis can be found elsewhere[4].

Why do we need meta-analyses of controlled trials?

A patient with myocardial infarction in 1981

A likely scenario in the early 1980s, when discussing the discharge of a patient who had suffered an uncomplicated myocardial infarction, is as follows: a keen junior doctor asks whether the patient should receive a beta-blocker for secondary prevention of a future cardiac event. After a moment of silence, the consultant states that this is a question which should be discussed in detail at the Journal Club on Thursday. The junior doctor (who now regrets that she asked the question) is told to assemble and present the relevant literature. Her MEDLINE search identifies four clinical trials (Table 1).

When reviewing the conclusions from these trials the doctor finds them to be rather confusing and contradictory. Her consultant points out that the sheer amount of research published makes it impossible to keep track of and critically appraise individual studies. He recommends a good review article. Back in the library the junior doctor finds an article which the *British Medical Journal* published in 1981 in a 'Regular Reviews' section[5]. This narrative review concluded that:

> *Thus, despite claims that they reduce arrhythmias, cardiac work, and infarct size, we still have no clear evidence that beta-blockers improve long-term survival after infarction despite almost 20 years of clinical trials.[5]*

The junior doctor is relieved. She presents the findings of the review article, the Journal Club is a success and the patient is discharged without a beta-blocker.

Limitations of a single study

Sampling variability means that treatment effect estimates will vary, even between studies performed in exactly the same way in identical populations. The smaller the study, the larger will be the sampling variability. The number of patients included in trials is often inadequate[6], so a single study often fails to detect, or exclude with certainty, a modest but important difference in the effects of two therapies. A trial may thus show no statistically significant treatment effect when in reality a clinically important effect exists – it may produce a false-negative result.

A recent examination of 1,941 trials relevant to the treatment of schizophrenia showed that only 58 (3%) studies were large enough to detect an important improvement[7]. In some cases the required sample size may be difficult to achieve. A drug that reduces the risk of death from myocardial infarction by 10% could, for example, delay many thousands of deaths each year in the UK alone. In order to detect such an effect with 90% certainty, over 10,000 patients in each treatment group would be needed[8]. The meta-analytic approach appears to be an attractive alternative to such a large, expensive and logistically problematic study. Data from patients in trials evaluating the same or a similar drug in a number of smaller, but comparable, studies are considered. In this way, the necessary number of patients may be reached, and relatively small effects can be detected or excluded with confidence.

Limitations of traditional reviews

Traditional narrative reviews have a number of disadvantages that systematic reviews may overcome. First, the conventional narrative review is subjective and therefore prone to bias and error[9]. Mulrow showed that all but one of 50 reviews published in the mid-1980s in leading general medicine journals did not specify the source of the information and failed to perform a standardised assessment of the methodological quality of studies[10]. Our junior doctor could have consulted another review of the same topic, published in the *European Heart Journal* in 1981. This review concluded that 'it seems perfectly reasonable to treat patients who have survived an infarction with timolol'[11]. Without guidance by formal rules, reviewers will inevitably disagree about issues as basic as what types of studies it is appropriate to include and how to balance the quantitative evidence they provide.

What was the evidence in 1981?

What conclusions would our junior doctor have reached if she had had access to a systematic review of the beta-blocker trials? A useful way to show the accumulation of evidence over time is to perform a *cumulative meta-analysis*[12]. Cumulative meta-analysis is defined as the repeated performance of meta-analysis whenever a new relevant trial becomes available for inclusion. This allows the retrospective identification of the point in time when a treatment effect first reached conventional levels of statistical significance. Figure 1 shows mortality results from a cumulative meta-analysis of trials of beta-blockers in secondary prevention after myocardial infarction[13]. Combining the results of the 13 trials published by the end of 1981, the relative risk of mortality comparing patients treated with beta-blocker with those treated with placebo is estimated as 0.78 (95% confidence intervals (CI) 0.69–0.88, *p* <0.001). Thus, conclusive evidence of the life-saving potential of this treatment, though available, was ignored. Subsequent trials in a further 15,000 patients simply confirmed this result. This situation has been taken to suggest that further studies in large numbers of patients may be superfluous and costly, if not unethical[14] once a statistically significant

Table 1. Conclusions from four randomised controlled trials of beta-blockers in secondary prevention after myocardial infarction.

The mortality and hospital readmission rates were not significantly different in the two groups. This also applied to the incidence of cardiac failure, exertional dyspnoea, and frequency of ventricular ectopic beats.
Reynolds, 1972[52]

Until the results of further trials are reported, long-term beta-adrenoceptor blockade (possibly up to two years) is recommended after uncomplicated anterior myocardial infarction.
Multicentre International Study, 1977[53]

The trial was designed to detect a 50% reduction in mortality and this was not shown. The non-fatal reinfarction rate was similar in both groups.
Baber *et al.*, 1981[54]

We conclude that long-term treatment with timolol in patients surviving acute myocardial infarction reduces mortality and the rate of reinfarction.
The Norwegian Multicentre Study Group, 1981[55]

treatment effect is evident from meta-analysis of the existing smaller trials.

Another application of cumulative meta-analysis has been to correlate the accruing evidence with the recommendations made by experts in review articles and text books. Antman *et al.*[15] showed for thrombolytic drugs that recommendations for routine use first appeared in 1987, 14 years after a statistically significant (*p* = 0.01) beneficial effect became evident in cumulative meta-analysis. Conversely, the prophylactic use of lidocaine continued to be recommended for routine use in myocardial infarction despite the lack of evidence for any beneficial effect and the suggestion of a harmful effect when results were combined in a meta-analysis[15].

Problems and limitations

Meta-analyses have received a mixed reception from the outset. Those on the receiving end have rejected what they see as exercises in 'mega-silliness'[16], and the authors of a highly distinguished series of systematic reviews of care during pregnancy and childhood[17] have been dubbed as terrorists ('an obstetrical Baader-Meinhof gang'[18]). Some statisticians think that meta-analysis 'represents the unacceptable face of statisticism'[19] and to clinicians objecting to the findings of meta-analyses 'a tool has become a weapon'[20]. Others 'still prefer the conventional narrative review article'[21].

This mixed reception is not surprising considering that several examples exist of meta-analyses of small trials whose findings were later contradicted by a single large randomised trial (Fig 2). Also, systematic reviews addressing the same issue have reached opposite conclusions. Examples include assessments of low molecular weight heparin in the prevention of peri-operative thrombosis[22,23], second-line anti-rheumatic

Matthias Egger, George Davey Smith, Jonathan A C Sterne

Key Points

Meta-analysis of clinical trials may enhance the precision of estimates of treatment effects, leading to reduced probability of false-negative results and potentially to a more timely introduction of effective treatments

If the methodological quality of trials is inadequate and the 'raw material' therefore flawed, the findings of meta-analyses of this material will also be compromised

Publication bias can distort findings because trials with statistically significant results are more likely to get published, and more likely to be published without delay, than trials without significant results

Meta-analysis should be considered only within the framework of systematic reviews: reviews that have been prepared using a systematic approach to minimise bias and explicitly address the issues of the completeness of the evidence identified, the quality of component studies and the combinability of studies.

drugs in the treatment of rheumatoid arthritis[24,25], calcium antagonists or cholesterol-lowering interventions in hypertension and coronary heart disease, and mammography for breast cancer screening[26–28]. In the following sections potential sources of bias are discussed in some detail.

Garbage in – garbage out?

The quality of component trials is of crucial importance: if the 'raw material' is flawed, the findings of reviews of this material may also be compromised. Clearly, the trials included in systematic reviews and meta-analyses should ideally be of high methodological quality and free from bias, such that the differences in outcomes observed between groups of patients can be confidently attributed to the intervention under investigation. The biases that threaten the validity of clinical trials are reviewed elsewhere[29]. These relate to systematic differences in the patients' characteristics at baseline (*selection bias*), unequal provision of care apart from the treatment under evaluation (*performance bias*), biased assessment of outcomes (*detection bias*), and bias due to exclusion of patients after they have been allocated to treatment groups (*attrition bias*)[30]. Several studies have recently attempted to quantify the impact of these biases on the results of controlled clinical trials[31]. For example, Schulz *et al.*[32] assessed the methodological quality of 250 trials from 33 meta-analyses from the Cochrane Pregnancy and Childbirth Database and examined the association between dimensions of trial quality and estimated treatment effects. Compared to trials in which authors reported adequately concealed treatment allocation, failure to prevent foreknowledge of treatment allocation or unclear concealment were associated, on average, with a 30–40% exaggeration of treatment effects. Trials that were not double-blind also yielded larger effects.
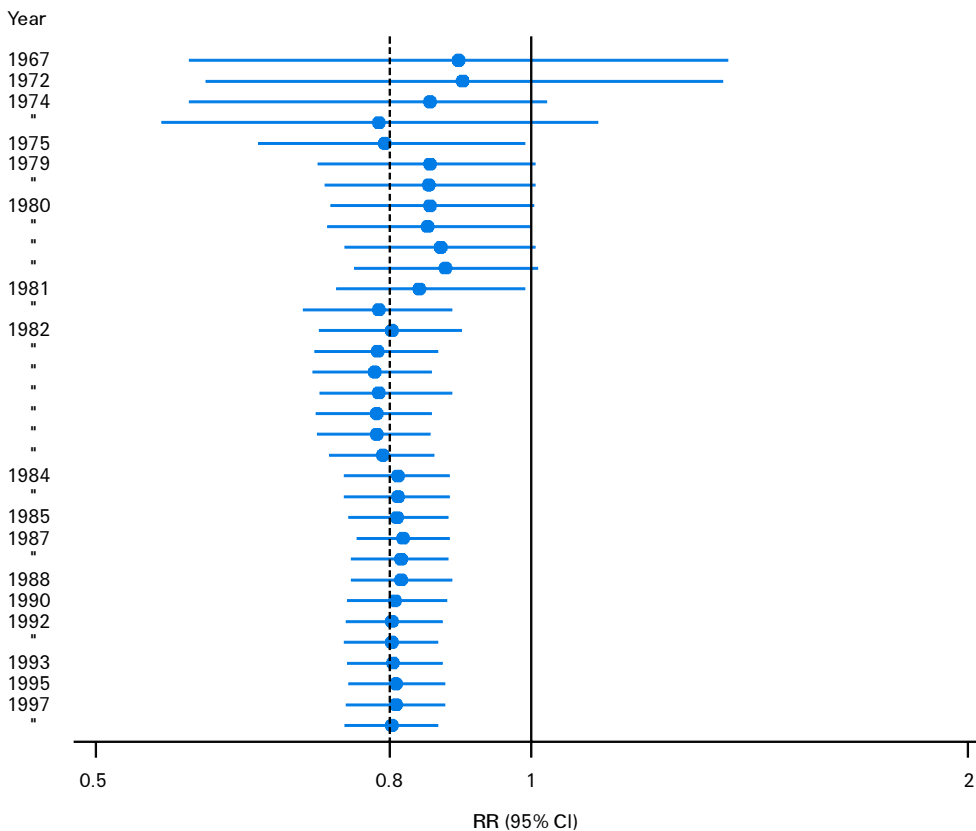

Fig 1. Cumulative meta-analysis of controlled trials of beta-blockers in secondary prevention after myocardial infarction. A clear (*p* <0.001) reduction of mortality was evident by 1981 (adapted from Freemantle *et al.*[13]).

Biased dissemination of research findings

The dissemination of research findings is a continuum ranging from the sharing of draft papers among colleagues, presentations at meetings, published abstracts, to papers in journals indexed in the major bibliographic databases[33]. It has long been recognised that only a proportion of research projects ultimately reach publication in an indexed journal and thus become easily identifiable for systematic reviews[34]. Scherer *et al.*[35] showed that only about half of abstracts presented at conferences are later published in full. Four separate studies followed up research proposals approved by ethics committees or institutional review boards in Oxford[36], Sydney[37], and at the Johns Hopkins School of Medicine[38] and School of Hygiene and Public Health, Baltimore[38]. For each cohort of research proposals the principal investigators were contacted several years later to determine the publication status of each completed study. The rates of full publication as journal articles ranged from 49 to 67%[39,40]. Similarly, 20% of trials funded by the National Institutes of Health and 45% of trials on HIV infection funded by the National Institute of Allergy and Infectious Diseases were still unpublished several years after completion[39,41,42].

The fact that a substantial proportion of studies remains unpublished even a decade after the study has been completed and analysed must be of concern, as potentially important information remains hidden from reviewers. To make things worse, the dissemination of research findings is not a random process; rather, it is strongly influenced by the nature and direction of results. Statistically significant, 'positive' results indicating that a treatment works are more likely to be published, more likely to be published rapidly, more likely to be published in English, more likely to be published more than once, and more likely to be cited by others (see Table 2 for a summary of these *reporting biases* and key references). In this situation, a meta-analysis of the published trials could identify a spurious beneficial treatment effect or miss an important adverse effect of a treatment. In the field of cancer
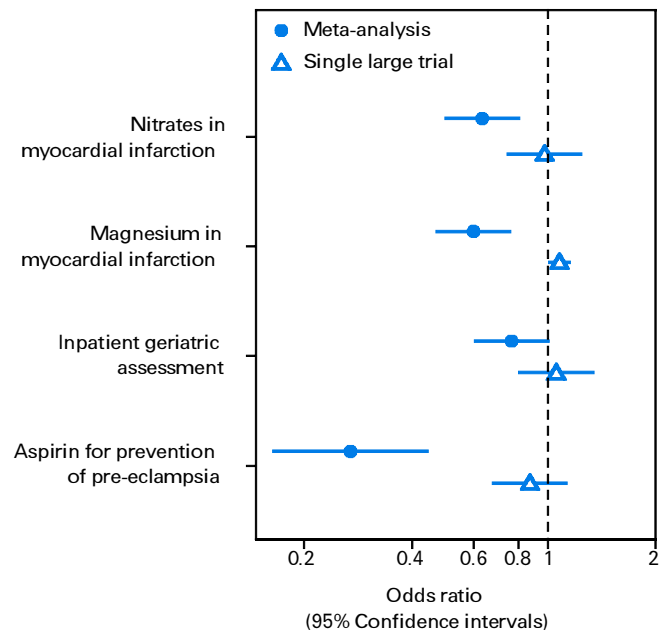

Fig 2. Results from discordant pairs of meta-analyses of small trials and single large trials. Effects of nitrates[67,68] and magnesium[69,70] on mortality in acute myocardial infarction, effect of inpatient geriatric assessment on mortality in the elderly[71,72], and effect of aspirin on the risk of pre-eclampsia[73,74].

chemotherapy, such publication bias has been demonstrated by comparing the results from studies identified in a literature search with those contained in an international trials registry[43,44]. In cardiovascular medicine, investigators in 1980 who found an increased death rate among patients with acute myocardial infarction treated with a class 1 anti-arrhythmic agent dismissed it as a chance finding and did not publish their trial at the time[45]. As discussed by Iain Chalmers, their findings would have contributed to a more timely detection of the increased mortality that has since become known to be associated with the use of class I anti-arrhythmic agents[46].

Table 2. Overview of reporting biases.

Type of reporting bias	Definition	Selected references
Publication bias	The *publication* or *non-publication* of research findings, depending on the nature and direction of the results	Begg and Berlin, 1989[56] Easterbrook *et al.* 1991[36] Dickersin *et al.* 1992, 1997[38,39]
Time lag bias	The *rapid* or *delayed* publication of research findings, depending on the nature and direction of the results	Stern and Simes, 1997[37] Ioannidis, 1998[42]
Multiple (duplicate) publication bias	The *multiple* or *singular* publication of research findings, depending on the nature and direction of the results	Gøtzsche, 1987, 1989[57,58] Tramèr *et al.* 1997[59] Huston and Moher, 1996[60]
Citation bias	The *citation* or *non-citation* of research findings, depending on the nature and direction of the results	Gøtzsche, 1987[57] Ravnskov, 1992[61]
Language bias	The publication of research findings *in a particular language*, depending on the nature and direction of the results	Grégoire *et al.* 1995[62] Egger *et al.* 1997[63]
Outcome reporting bias	The *selective reporting* of some outcomes but not others, depending on the nature and direction of the results	Hemminki, 1980[64] Pocock *et al.* 1987[65] Tannock, 1996[66]

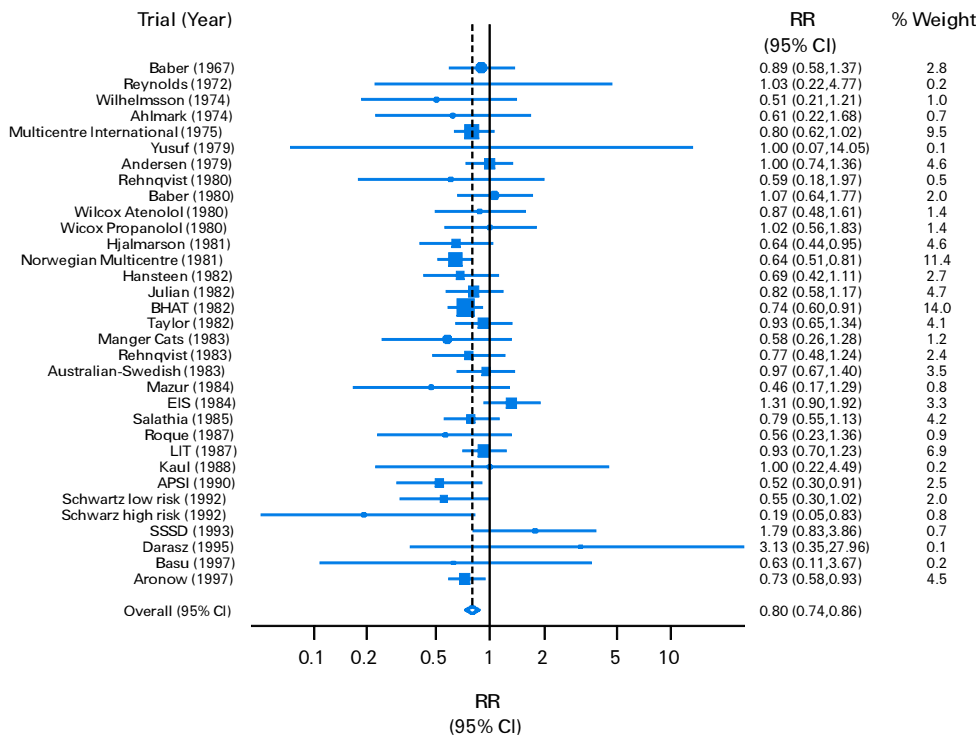Matthias Egger, George Davey Smith, Jonathan A C Sterne

Fig 3. Forest plot of controlled trials of beta-blockers in secondary prevention of mortality after myocardial infarction. The centre of the square and the horizontal line correspond to the relative risk (RR) and 95% confidence intervals (CIs). The area of the squares is proportional to the weight each trial contributes to the meta-analysis. The diamond at the bottom of the graph represents the combined RR and its 95% CI indicating a 20% reduction in the risk of death. The solid vertical line corresponds to no effect of treatment (RR 1.0), the dotted vertical line to the combined RR (0.8). The RR, 95% CI and weights are also given in tabular form (adapted from Freemantle et al.[13]).

Inadequate attention to heterogeneity

Meta-analysis of controlled trials is based on the assumption that each trial provides an unbiased estimate of the effect of an experimental treatment, with the variability of the results between the studies being attributed to random variation. The overall effect calculated from a group of sensibly combined and representative randomised trials will provide an essentially unbiased estimate of the treatment effect, with an increase in the precision of this estimate. Inspection of the 'forest plot' of the beta-blocker trials indicates that this is indeed the case: results are fairly homogeneous, clustering between a relative risk of 0.5 and 1.0, with widely overlapping CIs (Fig 3). However, there will be situations where the calculation of a combined effect estimate from trials is inappropriate or even misleading. For example, trials of BCG vaccination for prevention of tuberculosis[47] are clearly heterogeneous (Fig 4). The findings of the UK trial, which indicate substantial benefit of BCG vaccination, are not compatible with those from the Madras or Puerto Rico trials which suggest little effect or only a modest benefit. There is no overlap in the CIs of the three trials. BCG vaccination appears to be effective at higher latitudes but not in warmer regions,
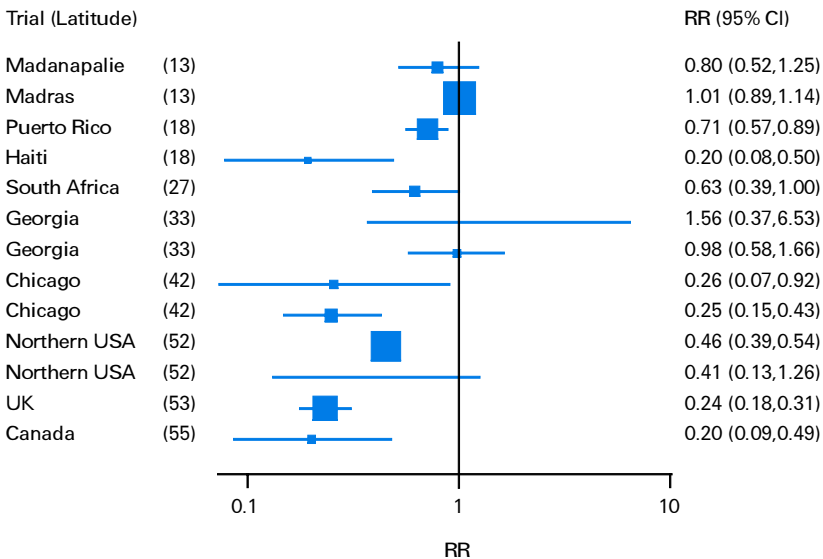

Fig 4. Forest plot of trials of BCG vaccine to prevent tuberculosis. Trials are ordered according to the latitude of the study location, expressed as degrees from the equator. No meta-analysis is shown (CI = confidence intervals, RR = relative risk) (adapted from Colditz et al.[47]).

possibly because exposure to certain environmental mycobacteria acts as a 'natural' BCG inoculation in warmer regions[48]. In this situation, it is more meaningful to quantify how the effect varies according to latitude than to calculate an overall estimate of effect which will be misleading.

Conclusions: a plea for more systematic reviews and fewer meta-analyses

Can meta-analysis be trusted? Unfortunately, many published meta-analyses are not based on comprehensive literature searches and ignore the quality of the studies included in the analysis. We believe that meta-analyses should be performed only within the framework of systematic reviews: reviews that have been prepared using a systematic approach to minimise bias and explicitly address the issues of the completeness of the evidence identified, the quality of component studies and the combinability of studies. How likely is it that publication and related biases have been avoided? Is it sensible to combine the individual trials in meta-analysis or is there heterogeneity between individual study results which renders questionable the calculation of an overall estimate? Are there adequately powered trials of high methodological quality? The distinction between systematic review and meta-analysis is important because it is always appropriate and desirable systematically to review a body of data, but it may sometimes be inappropriate, or even misleading, to pool statistically results from separate studies. Systematic reviews, as performed and disseminated by the Cochrane Collaboration[49–51], are clearly superior to the narrative approach to reviewing research. In addition to providing a precise estimate of the overall treatment effect in some instances, appropriate examination of heterogeneity across individual studies can produce useful information with which to guide rational and cost-effective treatment decisions. Systematic reviews are also important to demonstrate areas where the available evidence is simply insufficient and where new trials are required.

References

1 Pearson K. Report on certain enteric fever inoculation statistics. *Br Med J* 1904;**3**:1243–6.

2 Davis B. What price safety? Risk analysis measures need for regulation, but it's no science. *Wall Street Journal* 1992;**August 6**:p 1.

3 Susser M. Judgement and causal inference: criteria in epidemiologic studies. *Am J Epidemiol* 1977;**105**:1–15.

4 Egger M, Smith DG, O'Rourke K. Rationale, potentials and promise of systematic reviews. In: Egger M, Smith DG, Altman DG (eds). *Systematic reviews in health care: meta-analysis in context.* London: BMJ Books, 2001:23–42.

5 Mitchell JRA. Timolol after myocardial infarction: an answer or a new set of questions? *Br Med J* 1981;**282**:1565–70.

6 Freiman JA, Chalmers TC, Smith H, Kuebler RR. The importance of beta, the type II error, and sample size in the design and interpretation of the randomized controlled trial. In: Bailar JC, Mosteller F (eds). *Medical uses of statistics.* Boston, MD: NEJM Books, 1992: 357–73.

7 Thornley B, Adams C. Content and quality of 2000 controlled trials in schizophrenia over 50 years. *Br Med J* 1998;**317**:1181–4.

8 Collins R, Keech A, Peto R, Sleight P, *et al.* Cholesterol and total mortality: need for larger trials. *Br Med J* 1992;**304**:1689.

9 Teagarden JR. Meta-analysis: whither narrative review? Review. *Pharmacotherapy* 1989;**9**:274–81; discussion 281–4.

10 Mulrow CD. The medical review article: state of the science. *Ann Intern Med* 1987;**106**:485–8.

11 Hampton JR. The use of beta blockers for the reduction of mortality after myocardial infarction. Review. *Eur Heart J* 1981;**2**:259–68.

12 Lau J, Antman EM, Jimenez-Silva J, Kupelnick B, Mosteller F, Chalmers TC. Cumulative meta-analysis of therapeutic trials for myocardial infarction. *N Engl J Med* 1992;**327**:248–54.

13 Freemantle N, Cleland J, Young P, Mason J, Harrison J. Beta blockade after myocardial infarction: systematic review and meta regression analysis. *Br Med J* 1999;**318**:1730–7.

14 Murphy DJ, Povar GJ, Pawlson LG. Setting limits in clinical medicine. *Arch Intern Med* 1994;**154**:505–12.

15 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. *JAMA* 1992;**268**:240–8.

16 Eysenck HJ. An exercise in mega-silliness. *Am Psychol* 1978;**33**:517.

17 Chalmers I, Enkin M, Keirse M. *Effective care during pregnancy and childbirth.* Oxford: Oxford University Press, 1989.

18 Mann C. Meta-analysis in the breech. *Science* 1990;**249**:476–80.

19 Oakes M. *Statistical inference: a commentary for the social and behavioural sciences.* Chichester: John Wiley & Sons, 1986.

20 Boden WE. Meta-analysis in clinical trials reporting: Has a tool become a weapon? *Am J Cardiol* 1992;**69**:681–6.

21 Bailar JC. The promise and problems of meta-analysis. *N Engl J Med* 1997;**337**:559–61.

22 Nurmohamed MT, Rosendaal FR, Bueller HR, Dekker E, *et al.* Low-molecular-weight heparin versus standard heparin in general and orthopaedic surgery: a meta-analysis. *Lancet* 1992;**340**:152–6.

23 Leizorovicz A, Haugh MC, Chapuis FR, Samama MM, Boissel JP. Low molecular weight heparin in prevention of perioperative thrombosis. *Br Med J* 1992;**305**:913–20.

24 Felson DT, Anderson JJ, Meenan RF. The comparative efficacy and toxicity of second-line drugs in rheumatoid arthritis. *Arthritis Rheum* 1990;**33**:1449–61.

25 Gøtzsche PC, Podenphant J, Olesen M, Halberg P. Meta-analysis of second-line antirheumatic drugs: sample size bias and uncertain benefit. *J Clin Epidemiol* 1992;**45**:587–94.

26 Yusuf S. Calcium antagonists in coronary heart disease and hypertension. Time for reevaluation? *Circulation* 1995;**92**:1079–82.

27 Dunnigan MG. The problem with cholesterol. No light at the end of this tunnel? *Br Med J* 1993;**306**:1355–6.

28 de Koning HJ. Assessment of nationwide cancer-screening programmes. *Lancet* 2000;**355**:80–1.

29 Jüni P, Altman DG, Egger M. Assessing the quality of controlled clinical trials. In: Egger M, Smith DG, Altman DG (eds). *Systematic reviews in health care: meta-analysis in context.* London: BMJ Books, 2001: 87–108.

30 Altman DG. *Practical statistics for medical research.* London: Chapman and Hall, 1991.

31 Jüni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. *Br Med J* 2001;**323**:42–6.

32 Schulz KF, Chalmers I, Hayes RJ, Altman D. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;**273**:408–12.

33 Smith R. What is publication? A continuum. *Br Med J* 1999;**318**:142.

34 Dickersin K. The existence of publication bias and risk factors for its occurrence. *JAMA* 1990;**263**:1385–9.

35 Scherer RW, Dickersin K, Langenberg P. Full publication of results initially presented in abstracts. A meta-analysis. *JAMA* 1994;**272**:158–62.

36 Easterbrook PJ, Berlin J, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet* 1991;**337**:867–72.

37 Stern JM, Simes RJ. Publication bias: evidence of delayed publication in a cohort study of clinical research projects. *Br Med J* 1997;**315**:640–5.

38 Dickersin K, Min YL, Meinert CL. Factors influencing publication of research results. Follow-up of applications submitted to two institutional review boards. *JAMA* 1992;**267**:374–8.

39 Dickersin K. How important is publication bias? A synthesis of available data. *AIDS Educ Prev* 1997;**9**:15–21.

40 Egger M, Dickersin K, Davey Smith G. Problems and limitations in conducting systematic reviews. In: Egger M, Smith DG, Altman DG (eds). *Systematic reviews in health care: meta-analysis in context.* London: BMJ Books, 2001:43–68.

41 Dickersin K, Min YI. NIH clinical trials and publication bias. *Online J Curr Clin Trials* 1993;**Doc No 50**:4967.

42 Ioannidis JPA. Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. *JAMA* 1998;**279**:281–6.

43 Simes RJ. Confronting publication bias: a cohort design for meta-analysis. *Stat Med* 1987;**6**:11–29.

44 Simes RJ. Publication bias: the case for an international registry of clinical trials. *J Clin Oncol* 1986;**4**:1529–41.

45 Cowley AJ, Skene A, Stainer K, Hampton JR. The effect of lorcainide on arrhythmias and survival in patients with acute myocardial infarction: an example of publication bias. *Int J Cardiol* 1993;**40**:161–6.

46 Chalmers I. Using systematic reviews and registers of ongoing trials for scientific and ethical trial design, monitoring, and reporting. In: Egger M, Smith DG, Altman DG (eds). *Systematic reviews in health care: meta-analysis in contex.* London: BMJ Books, 2001:429–43.

47 Colditz GA, Brewer TF, Berkley CS, Wilson ME, *et al*. Efficacy of BCG vaccine in the prevention of tuberculosis. *JAMA* 1994;**271**:698–702.

48 Fine PEM. Variation in protection by BCG: implications of and for heterologous immunity. *Lancet* 1995;**346**:1339–45.

49 Chalmers I, Haynes B. Reporting, updating and correcting systematic reviews of the effects of health care. *Br Med J* 1994;**309**:862–5.

50 Bero L, Rennie D. The Cochrane Collaboration. Preparing, maintaining, and disseminating systematic reviews of the effects of health care. *JAMA* 1995;**274**:1935–8.

51 Dickersin K, Manheimer E. The Cochrane Collaboration: evaluation of health care and services using systematic reviews of the results of randomized controlled trials. *Clin Obstet Gynecol* 1998;**41**:315–331.

52 Reynolds JL, Whitlock RML. Effects of a beta-adrenergic receptor blocker in myocardial infarction treated for one year from onset. *Br Heart J* 1972;**34**:252–9.

53 Multicentre International Study: supplementary report. Reduction in mortality after myocardial infarction with long-term beta-adrenoceptor blockade. *Br Med J* 1977;**2**:419–21.

54 Baber NS, Wainwright Evans D, Howitt G, Thomas M, *et al*. Multicentre post-infarction trial of propranolol in 49 hospitals in the United Kingdom, Italy and Yugoslavia. *Br Heart J* 1980;**44**:96–100.

55 The Norwegian Multicenter Study Group. Timolol-induced reduction in mortality and reinfarction in patients surviving acute myocardial infarction. *N Engl J Med* 1981;**304**:801–7.

56 Begg CB, Berlin J. Publication bias and dissemination of clinical research. *J Natl Cancer Inst* 1989;**81**:107–15.

57 Gøtzsche PC. Reference bias in reports of drug trials. *Br Med J* 1987;**295**:654–6.

58 Gøtzsche PC. Multiple publication of reports of drug trials. *Eur J Clin Pharmacol* 1989;**36**:429–32.

59 Tramèr MR, Reynolds DJM, Moore RA, McQuay HJ. Impact of covert duplicate publication on meta-analysis: a case study. *Br Med J* 1997;**315**:635–40.

60 Huston P, Moher D. Redundancy, disaggregation, and the integrity of medical research. *Lancet* 1996;**347**:1024–6.

61 Ravnskov U. Cholesterol lowering trials in coronary heart disease: frequency of citation and outcome. *Br Med J* 1992;**305**:15–9.

62 Grégoire G, Derderian F, LeLorier J. Selecting the language of the publications included in a meta-analysis: is there a Tower of Babel bias? *J Clin Epidemiol* 1995;**48**:159–63.

63 Egger M, Zellweger-Zähner T, Schneider M, Junker C, *et al*. Language bias in randomised controlled trials published in English and German. *Lancet* 1997;**350**:326–9.

64 Hemminki E. Study of information submitted by drug companies to licensing authorities. *Br Med J* 1980;**280**:833–6.

65 Pocock S, Hughes MD, Lee RJ. Statistical problems in the reporting of clinical trials. A survey of three medical journals. *N Engl J Med* 1987;**317**:426–32.

66 Tannock IF. False-positive results in clinical trials: multiple significance tests and the problem of unreported comparisons. *J Natl Cancer Inst* 1996;**88**:206–7.

67 Yusuf S, Collins R, MacMahon S, Peto R. Effect of intravenous nitrates on mortality in acute myocardial infarction: an overview of the randomised trials. *Lancet* 1988;**i**:1088–92.

68 Gruppo Italiano per lo Studio della Streptochinasi nell'Infarto Miocardico (GISSI). GISSI-3: effects of lisinopril and transdermal glyceryl trinitrate singly and together on 6-week mortality and ventricular function after acute myocardial infarction. *Lancet* 1994;**343**:1115–22.

69 Teo KK, Yusuf S. Role of magnesium in reducing mortality in acute myocardial infarction. A review of the evidence. *Drugs* 1993;**46**:347–59.

70 ISIS-4 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.

71 Stuck AE, Siu AL, Wieland GD, Adams J, Rubenstein LZ. Comprehensive geriatric assessment: a meta-analysis of controlled trials. *Lancet* 1993;**342**:1032–6.

72 Reuben DB, Borok GM, Wolde-Tsadik G, Ershoff DH, *et al*. Randomized trial of comprehensive geriatric assessment in the care of hospitalized patients. *N Engl J Med* 1995;**332**:1345–50.

73 Imperiale TF, Petrullis AS. A meta-analysis of low-dose aspirin for the prevention of pregnancy-induced hypertensive disease. *JAMA* 1991;**266**:261–5.

74 CLASP Collaborative Group. CLASP: a randomized trial of low-dose aspirin for the prevention and treatment of pre-eclampsia among 9364 pregnant women. *Lancet* 1994;**343**:619–29.

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