Stata™ datasets and other additional information can be found on the book's web site:
www.systematicreviews.com

SYSTEMATIC REVIEWS IN HEALTH CARE:
META-ANALYSIS IN CONTEXT

Second edition

Edited by

Matthias Egger

Senior Lecturer in Epidemiology and Public Health Medicine, Division of Health Services Research and MRC Health Services Research Collaboration, Department of Social Medicine University of Bristol, UK

George Davey Smith

Professor of Clinical Epidemiology, Division of Epidemiology and MRC Health Services Research Collaboration, Department of Social Medicine, University of Bristol, UK

and

Douglas G Altman

Professor of Statistics in Medicine, ICRF Medical Statistics Group, Centre for Statistics in Medicine, Institute of Health Sciences, University of Oxford, UK

BMJ Books

1 Rationale, potentials, and promise of systematic reviews

MATTHIAS EGGER, GEORGE DAVEY SMITH, KEITH O'ROURKE

Summary points

- Reviews are essential tools for health care workers, researchers, consumers and policy makers who want to keep up with the evidence that is accumulating in their field.
- Systematic reviews allow for a more objective appraisal of the evidence than traditional narrative reviews and may thus contribute to resolve uncertainty when original research, reviews, and editorials disagree.
- Meta-analysis, if appropriate, will 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.
- Exploratory analyses, e.g. regarding subgroups of patients who are likely to respond particularly well to a treatment (or the reverse), may generate promising new research questions to be addressed in future studies.
- Systematic reviews may demonstrate the lack of adequate evidence and thus identify areas where further studies are needed.

The volume of data that need to be considered by practitioners and researchers is constantly expanding. In many areas it has become simply impossible for the individual to read, critically evaluate and synthesise the state of current knowledge, let alone keep updating this on a regular basis. Reviews have become essential tools for anybody who wants to keep up with the new evidence that is accumulating in his or her field of interest. Reviews are also required to identify areas where the

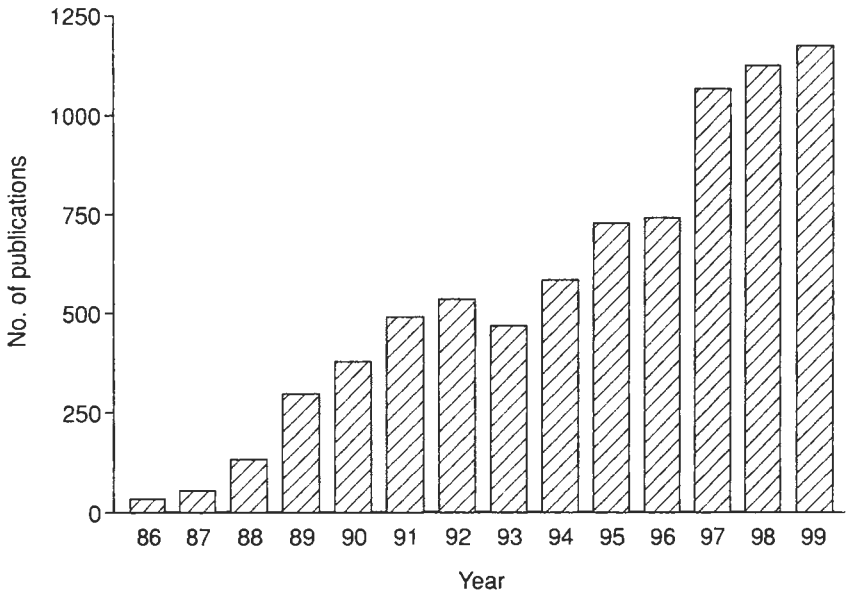


Figure 1.1 Number of publications concerning meta-analysis, 1986–1999. Results from MEDLINE search using text word and medical subject (MESH) heading "meta-analysis" and text word "systematic review".

available evidence is insufficient and further studies are required. However, since Mulrow[1] and Oxman and Guyatt[2] drew attention to the poor quality of narrative reviews it has become clear that conventional reviews are an unreliable source of information. In response to this situation there has, in recent years, been increasing focus on formal methods of systematically reviewing studies, to produce explicitly formulated, reproducible, and up-to-date summaries of the effects of health care interventions. This is illustrated by the sharp increase in the number of reviews that used formal methods to synthesise evidence (Figure 1.1).

In this chapter we will attempt to clarify terminology and scope, provide some historical background, and examine the potentials and promise of systematic reviews and meta-analysis.

Systematic review, overview or meta-analysis?

A number of terms are used concurrently to describe the process of systematically reviewing and integrating research evidence, including

"systematic review", "meta-analysis", "research synthesis", "overview" and "pooling". In the foreword to the first edition of this book, Chalmers and Altman[3] defined systematic review as a review that has been prepared using a systematic approach to minimising biases and random errors which is documented in a materials and methods section. A systematic review may, or may not, include a meta-analysis: a statistical analysis of the results from independent studies, which generally aims to produce a single estimate of a treatment effect.[4] The distinction between systematic review and meta-analysis, which will be used throughout this book, is important because 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.[5] Indeed, it is our impression that reviewers often find it hard to resist the temptation of combining studies even when such meta-analysis is questionable or clearly inappropriate.

The scope of meta-analysis

As discussed in detail in Chapter 12, a clear distinction should be made between meta-analysis of randomised 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 underlying effect in ways that may systematically differ from chance. Combining a set of epidemiological studies will thus often provide spuriously precise, but biased, estimates of associations. 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 (see Chapters 9 and 12 for examples of observational meta-analyses).

The fundamental difference that exists between observational studies and randomised controlled trials does not mean that the latter are immune to bias. Publication bias and other reporting biases (see Chapter 3) may distort the evidence from both trials and observational studies. Bias may also be introduced if the methodological quality of

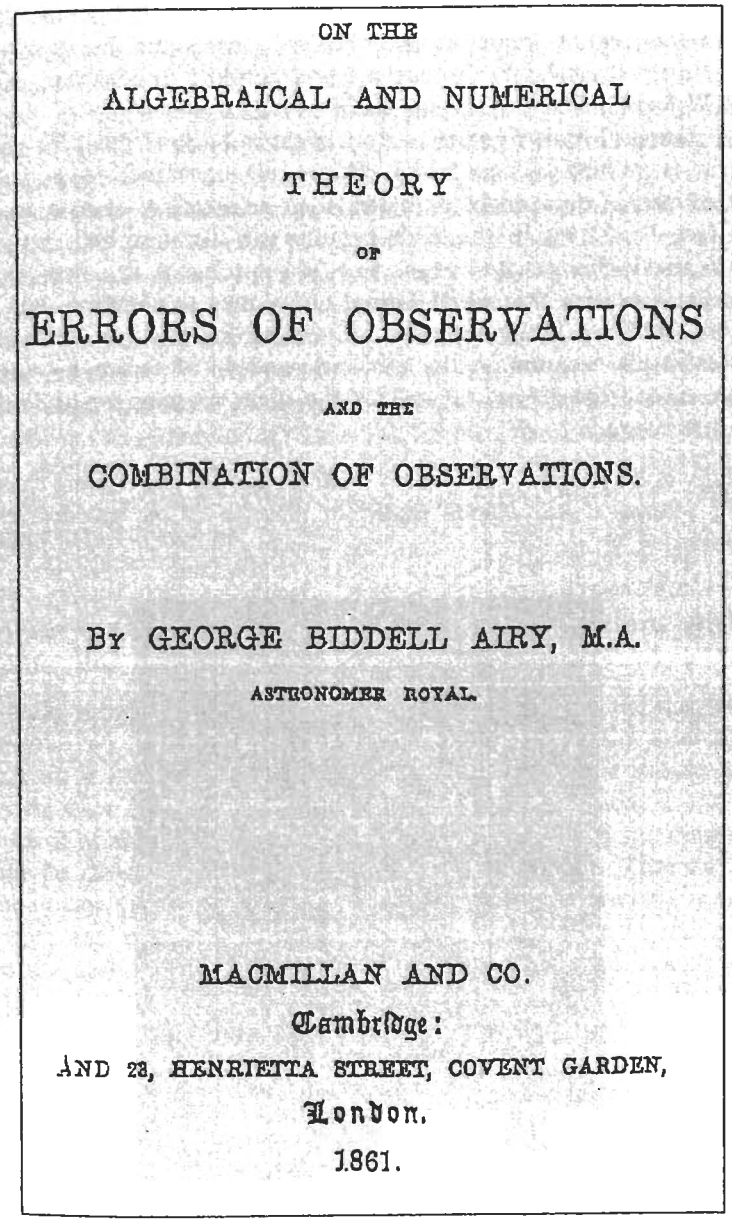
controlled trials is inadequate[6,7] (Chapter 5). It is crucial to understand the limitations of meta-analysis and the importance of exploring sources of heterogeneity and bias (Chapters 8–11), and much emphasis will be given to these issues in this book.

Historical notes

Efforts to compile summaries of research for medical practitioners who struggle with the amount of information that is relevant to medical practice are not new. Chalmers and Tröhler[8] drew attention to two journals published in the 18th century in Leipzig and Edinburgh, *Comentarii de rebus in scientia naturali et medicina gestis* and *Medical and Philosophical Commentaries*, which published critical appraisals of

Box 1.1 From Laplace and Gauss to the first textbook of meta-analysis

Astronomers long ago noticed that observations of the same objects differed even when made by the same observers under similar conditions. The calculation of the mean as a more precise value than a single measurement had appeared by the end of the 17th century.[9] By the late 1700s probability models were being used to represent the uncertainty of observations that was caused by measurement error. Laplace decided to write these models not as the probability that an observation equalled the true value plus some error but as the truth plus the "probability of some error". In doing this he recognised that as probabilities of independent errors multiply he could determine the most likely joint errors, the concept which is at the heart of maximum likelihood estimation.[10] Laplace's method of combining and quantifying uncertainty in the combination of observations required an explicit probability distribution for errors in the individual observations and no acceptable one existed. Gauss drew on empirical experience and argued that a probability distribution corresponding to what is today referred to as the Normal or Gaussian distribution would be best. This remained speculative until Laplace's formulation of the central limit theorem – that for large sample sizes the error distribution will always be close to Normally distributed. Hence, Gauss's method was more than just a good guess but justified by the central limit theorem. Most statistical techniques used today in meta-analysis follow from Gauss's and Laplace's work. Airy disseminated their work in his 1861 "textbook" on "meta-analysis" for astronomers (Figure 1.2) which included the first formulation of a random effects model to allow for heterogeneity in the results.[11] Airy offered practical advice and argued for the use of judgement to determine what type of statistical model should be used.


ON THE

ALGEBRAICAL AND NUMERICAL

THEORY

OF

ERRORS OF OBSERVATIONS

AND THE

COMBINATION OF OBSERVATIONS.

BY GEORGE BIDDELL AIRY, M.A.

ASTRONOMER ROYAL.

MACMILLAN AND CO.

Cambridge:

AND 23, HENRIETTA STREET, COVENT GARDEN,

London.

1861.

Figure 1.2 The title page of what may be seen as the first "textbook" of meta-analysis, published in 1861.

important new books in medicine, including, for example, William Withering's now classic *Account of the foxglove* (1785) on the use of digitalis for treating heart disease. These journals can be seen as the 18th century equivalents of modern day secondary publications such as the *ACP Journal Club* or *Evidence based medicine*.

The statistical basis of meta-analysis reaches back to the 17th century when in astronomy and geodesy intuition and experience suggested that combinations of data might be better than attempts to choose amongst them (see Box 1.1). In the 20th century the distinguished statistician Karl Pearson (Figure 1.3), was, in 1904, probably the first medical researcher reporting the use of formal techniques to combine data from different studies. The rationale for pooling studies put forward by Pearson in his account on the preventive effect of serum inoculations against enteric fever,[12] is still one of the main reasons for undertaking meta-analysis today:

> "*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*".[12]

Figure 1.3 Distinguished statistician Karl Pearson is seen as the first medical researcher to use formal techniques to combine data from different studies.

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, developed an early interest in the synthesis of research findings. In the 1930s, 80 experiments examining the "potency of moral instruction in modifying conduct" were systematically reviewed.[13] In 1976 the psychologist Glass coined the term "meta-analysis" in a paper entitled "Primary, secondary and meta-analysis of research".[14] 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.[15] In the 1980s meta-analysis became increasingly popular in medicine, particularly in the fields of cardiovascular disease,[16,17] oncology,[18] and perinatal care.[19] Meta-analysis of epidemiological studies[20,21] and "cross design synthesis",[22] the integration of observational data with the results from meta-analyses of randomised clinical trials was also advocated. In the 1990s the foundation of the Cochrane Collaboration (see Chapters 25 and 26) facilitated numerous developments, many of which are documented in this book.

Why do we need systematic reviews? 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 was 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. It is late in the evening when she makes her way to the library. The MEDLINE search identifies four clinical trials.[23–26] When reviewing the conclusions from these trials (Table 1.1) 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 *BMJ* published in 1981 in a "Regular Reviews" section.[27] This narrative review concluded:

> *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.*[27]

Table 1.1 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 and Whitlock[23]
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[24]
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.*[25]
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[26]

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

Narrative reviews

Traditional narrative reviews have a number of disadvantages that systematic reviews may overcome. First, the classical review is subjective and therefore prone to bias and error.[28] Mulrow showed that among 50 reviews published in the mid 1980s in leading general medicine journals, 49 reviews did not specify the source of the information and failed to perform a standardised assessment of the methodological quality of studies.[1] Our junior doctor could have consulted another review of the same topic, published in the *European Heart Journal* in the same year. This review concluded that "it seems perfectly reasonable to treat patients who have survived an infarction with timolol".[29] 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. Selective inclusion of studies that support the author's view is common. This is illustrated by the observation that the frequency of citation of clinical trials is related to their outcome, with studies in line with the prevailing opinion being quoted more frequently than unsupportive studies[30,31] Once a set of studies has been assembled a common way to review the results is to count the number of studies supporting various sides of an issue and to choose the view receiving the most votes. This procedure is clearly unsound, since it ignores sample size, effect size, and research design. It

is thus hardly surprising that reviewers using traditional methods often reach opposite conclusions[1] and miss small, but potentially important, differences.[32] In controversial areas the conclusions drawn from a given body of evidence may be associated more with the speciality of the reviewer than with the available data.[33] By systematically identifying, scrutinising, tabulating, and perhaps integrating all relevant studies, systematic reviews allow a more objective appraisal, which can help to resolve uncertainties when the original research, classical reviews and editorial comments disagree.

Limitations of a single study

A single study often fails to detect, or exclude with certainty, a modest, albeit relevant, difference in the effects of two therapies. A trial may thus show no statistically significant treatment effect when in reality such an effect exists – it may produce a false negative result. An examination of clinical trials which reported no statistically significant differences between experimental and control therapy has shown that false negative results in health care research are common: for a clinically relevant difference in outcome the probability of missing this effect given the trial size was greater than 20% in 115 (85%) of the 136 trials examined.[34] Similarly, a recent examination of 1941 trials relevant to the treatment of schizophrenia showed that only 58 (3%) studies were large enough to detect an important improvement.[35] The number of patients included in trials is thus often inadequate, a situation which has changed little over recent years.[34] In some cases, however, the required sample size may be difficult to achieve. A drug which 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 ten thousand patients in each treatment group would be needed.[36]

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. Methods used for meta-analysis employ a weighted average of the results in which the larger trials have more influence than the smaller ones. Comparisons are made exclusively between patients enrolled in the same study. As discussed in detail in chapter 15, there are a variety of statistical techniques available for this purpose.[37,38] In this way the necessary number of patients may be reached, and relatively small effects can be detected or excluded with confidence. Systematic reviews can also contribute to considerations regarding the applicability of study results. The findings

of a particular study might be felt to be valid only for a population of patients with the same characteristics as those investigated in the trial. If many trials exist in different groups of patients, with similar results being seen in the various trials, then it can be concluded that the effect of the intervention under study has some generality. By putting together all available data meta-analyses are also better placed than individual trials to answer questions regarding whether or not an overall study result varies among subgroups, e.g. among men and women; older and younger patients or participants with different degrees of severity of disease.

A more transparent appraisal

An important advantage of systematic reviews is that they render the review process transparent. In traditional narrative reviews it is often not clear how the conclusions follow from the data examined. In an adequately presented systematic review it should be possible for readers to replicate the quantitative component of the argument. To facilitate this, it is valuable if the data included in meta-analyses are either presented in full or made available to interested readers by the authors. The increased openness required leads to the replacement of unhelpful descriptors such as "no clear evidence", "some evidence of a trend", "a weak relationship" and "a strong relationship".[39] Furthermore, performing a meta-analysis may lead to reviewers moving beyond the conclusions authors present in the abstract of papers, to a thorough examination of the actual data.

The epidemiology of results

The tabulation, exploration and evaluation of results are important components of systematic reviews. This can be taken further to explore sources of heterogeneity and test new hypotheses that were not posed in individual studies, for example using "meta-regression" techniques (see also Chapters 8–11). This has been termed the "epidemiology of results" where the findings of an original study replace the individual as the unit of analysis.[40] However, it must be born in mind that although the studies included may be controlled experiments, the meta-analysis itself is subject to many biases inherent in observational studies.[41] Aggregation or ecological bias[42] is also a problem unless individual patient data is available (see Chapter 6). Systematic reviews can, nevertheless, lead to the identification of the most promising or the most urgent research question, and may permit a more accurate calculation of the sample sizes needed in future studies (see Chapter 24). This is illustrated by an early meta-analysis of

four trials that compared different methods of monitoring the fetus during labour.[43] The meta-analysis led to the hypothesis that, compared with intermittent auscultation, continuous fetal heart monitoring reduced the risk of neonatal seizures. This hypothesis was subsequently confirmed in a single randomised trial of almost seven times the size of the four previous studies combined.[44]

What was the evidence in 1981? Cumulative meta-analysis

What conclusions would our junior doctor have reached if she had had access to a meta-analysis? Numerous meta-analyses of trials examining the effect of beta-antagonists have been published since 1981.[17,45–48] Figure 1.4 shows the results from the most recent analysis that included 33 randomised comparisons of beta-blockers versus placebo or alternative treatment in patients who had had a myocardial infarction.[48] These trials were published between 1967 and 1997. The combined relative risk indicates that beta-blockade starting after the acute infarction reduces subsequent premature mortality by an estimated 20% (relative risk 0.80). A useful way to show the evidence that was available in 1981 and at other points in time is to perform a *cumulative* meta-analysis.[49]

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. In the case of beta-blockade in secondary prevention of myocardial infarction, a statistically significant beneficial effect ($P < 0{\cdot}05$) became evident by 1981 (Figure 1.5). 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 at best superfluous and costly, if not unethical,[50] once a statistically significant treatment effect is evident from meta-analysis of the existing smaller trials.

Similarly, Lau *et al.* showed that for the trials of intravenous streptokinase in acute myocardial infarction, a statistically significant ($P = 0{\cdot}01$) combined difference in total mortality was achieved by 1973[49] (Figure 1.6). At that time, 2432 patients had been randomised in eight small trials. The results of the subsequent 25 studies which included the large GISSI-1 and ISIS-2 trials[51,52] and enrolled a total of 34 542 additional patients reduced the significance level to $P = 0{\cdot}001$ in 1979, $P = 0{\cdot}0001$ in 1986 and to $P < 0{\cdot}00001$ when the first mega-trial appeared, narrowing the confidence intervals around an essentially unchanged

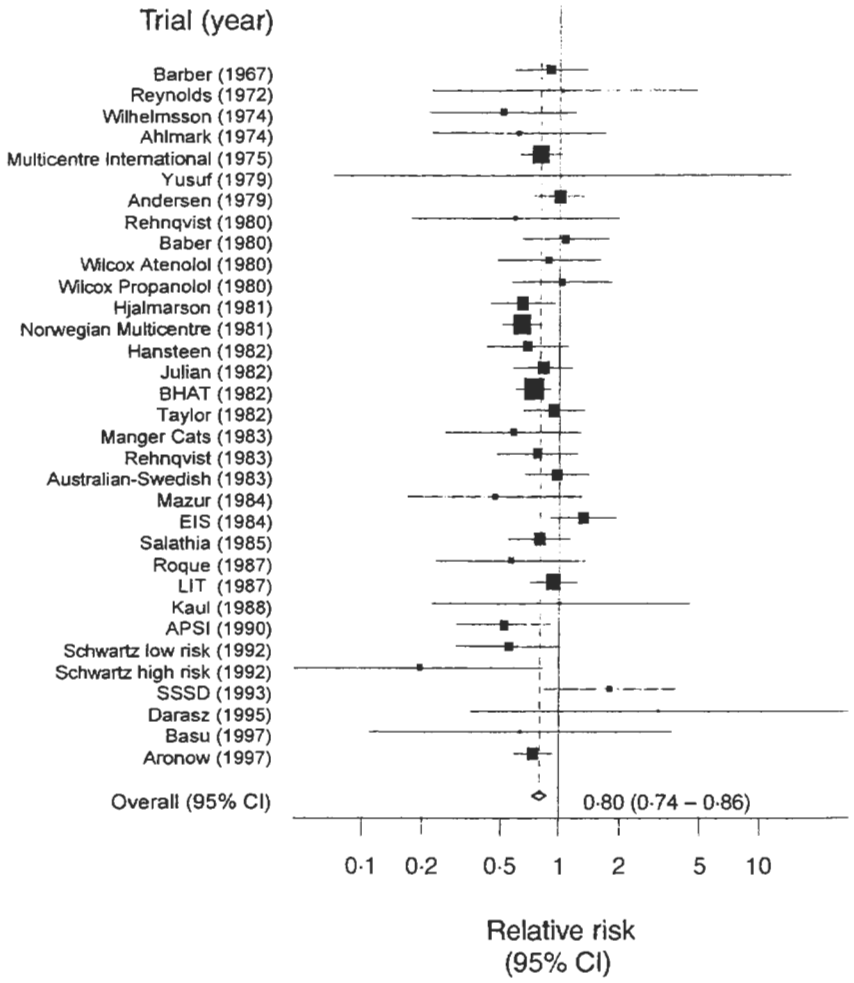


Figure 1.4 "Forest plot" showing mortality results from trials of beta-blockers in secondary prevention after myocardial infarction. Trials are ordered by year of publication. The black square and horizontal line correspond to the trials' risk ratio and 95% confidence intervals. The area of the black squares reflects the weight each trial contributes in the meta-analysis. The diamond represents the combined relative risk with its 95% confidence interval, indicating a 20% reduction in the odds of death. See Chapter 2 for a detailed description of forest plots. Adapted from Freemantle *et al.*[48]

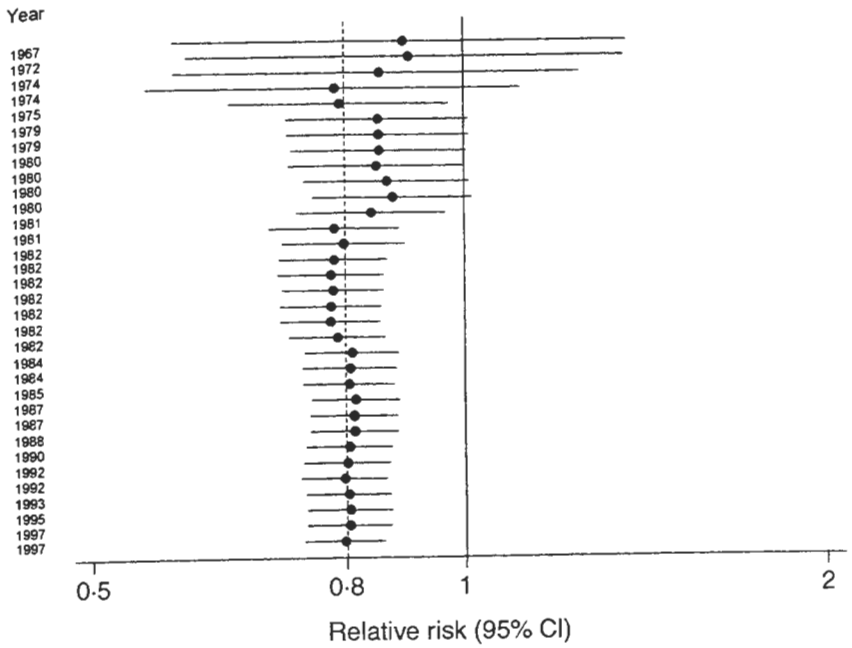


Figure 1.5 Cumulative meta-analysis of controlled trials of beta-blockers after myocardial infarction. The data correspond to Figure 1.4. A statistically significant ($P < 0{\cdot}05$) beneficial effect on mortality became evident in 1981.

estimate of about 20% reduction in the risk of death. Interestingly, at least one country licensed streptokinase for use in myocardial infarction before GISSI-1[51] was published, whereas many national authorities waited for this trial to appear and some waited a further two years for the results of ISIS-2[52] (Figure 1.6).

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.* showed for thrombolytic drugs that recommendations for routine use first appeared in 1987, 14 years after a statistically significant ($P = 0{\cdot}01$) beneficial effect became evident in cumulative meta-analysis.[53] 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 possibility of a harmful effect being evident in the meta-analysis.

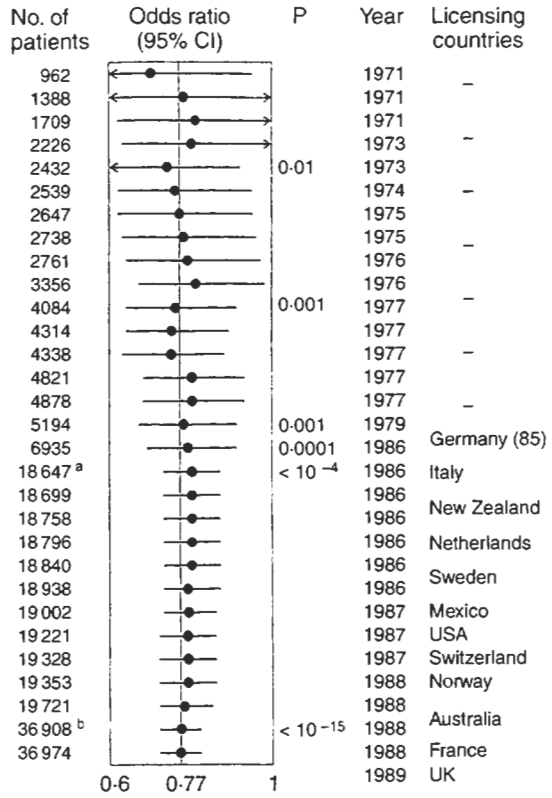


Figure 1.6 Cumulative meta-analysis of randomised controlled trials of intravenous streptokinase in myocardial infarction. The number of patients randomised in a total of 33 trials, and national authorities licensing streptokinase for use in myocardial infarction are also shown. [a]Includes GISSI-1; [b]ISIS-2.

Conclusions

Systematic review including, if appropriate, a formal meta-analysis is 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 insufficient and where new, adequately sized trials are required.

Acknowledgements

We are grateful to Sir David Cox for providing key references to early statistical work and to Iain Chalmers for his comments on an earlier draft of this chapter. We thank Dr T Johansson and G Enocksson (Pharmacia AB, Stockholm) and Dr A Schirmer and Dr M Thimme (Behring AG, Marburg) for providing data on licensing of streptokinase in different countries. This chapter draws on material published earlier in the *BMJ*.[54]

1 Mulrow CD. The medical review article: state of the science. *Ann Intern Med* 1987;**106**:485–8.
2 Oxman AD, Guyatt GH. Guidelines for reading literature reviews. *Can Med Assoc J* 1988;**138**:697–703.
3 Chalmers I, Altman D (eds). *Systematic reviews*. London: BMJ Publishing Group, 1995.
4 Huque MF. Experiences with meta-analysis in NDA submissions. *Proc Biopharmac Sec Am Stat Assoc* 1988;**2**:28–33.
5 O'Rourke K, Detsky AS. Meta-analysis in medical research: strong encouragement for higher quality in individual research efforts. *J Clin Epidemiol* 1989;**42**:1021–4.
6 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.
7 Moher D, Pham B, Jones A, *et al.* Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998;**352**:609–13.
8 Chalmers I, Tröhler U. Helping physicians to keep abreast of the medical literature: Medical and philosophical commentaries, 1773–1795. *Ann Intern Med* 2000; **133**: 238–43.
9 Plackett RL. Studies in the history of probability and statistics: VII. The principle of the arithmetic mean. *Biometrika* 1958;**1958**:130–5.
10 Stigler SM. *The history of statistics. The measurement of uncertainty before 1900.* Cambridge, MA: The Belknap Press of Harvard University Press, 1990.
11 Airy GB. *On the algebraical and numerical theory of errors of observations and the combinations of observations*. London: Macmillan, 1861.
12 Pearson K. Report on certain enteric fever inoculation statistics. *BMJ* 1904;**3**:1243–6.
13 Peters CC. Summary of the Penn State experiments on the influence of instruction in character education. *J Educat Sociol* 1933;7:269–72.
14 Glass GV. Primary, secondary and meta-analysis of research. *Educat Res* 1976;**5**:3–8.
15 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, 1979:1–11.
16 Baber NS, Lewis JA. Confidence in results of beta-blocker postinfarction trials. *BMJ* 1982; **284**:1749–50.
17 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.
18 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.
19 Chalmers I, Enkin M, Keirse M. *Effective care during pregnancy and childbirth*. Oxford: Oxford University Press, 1989.
20 Greenland S. Quantitative methods in the review of epidemiologic literature. *Epidemiol Rev* 1987;**9**:1–30.
21 Friedenreich CM. Methods for pooled analyses of epidemiologic studies. *Epidemiology* 1993;**4**:295–302.

22 General Accounting Office. *Cross design synthesis: a new strategy for medical effectiveness research*. Washington, DC: GAO, 1992.
23 Reynolds JL, Whitlock RML. Effects of a beta-adrenergic receptor blocker in myocardial infarctation treated for one year from onset. *Br Heart J* 1972;**34**:252–9.
24 Multicentre International Study: supplementary report. Reduction in mortality after myocardial infarction with long-term beta-adrenoceptor blockade. *BMJ* 1977;**2**:419–21.
25 Baber NS, Wainwright Evans D, Howitt G, *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.
26 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.
27 Mitchell JRA. Timolol after myocardial infarction: an answer or a new set of questions? *BMJ* 1981;**282**:1565–70.
28 Teagarden JR. Meta-analysis: whither narrative review? *Pharmacotherapy* 1989;**9**:274–84.
29 Hampton JR. The use of beta blockers for the reduction of mortality after myocardial infarction. *Eur Heart J* 1981;**2**:259–68.
30 Ravnskov U. Cholesterol lowering trials in coronary heart disease: frequency of citation and outcome. *BMJ* 1992;**305**:15–19.
31 Gøtzsche PC. Reference bias in reports of drug trials. *BMJ* 1987;**295**:654–6.
32 Cooper H, Rosenthal R. Statistical versus traditional procedures for summarising research findings. *Psychol Bull* 1980;**87**:442–9.
33 Chalmers TC, Frank CS, Reitman D. Minimizing the three stages of publication bias. *JAMA* 1990;**263**:1392–5.
34 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, MA: NEJM Books, 1992:357–73.
35 Thornley B, Adams C. Content and quality of 2000 controlled trials in schizophrenia over 50 years. *BMJ* 1998;**317**:1181–4.
36 Collins R, Keech A, Peto R, *et al.* Cholesterol and total mortality: need for larger trials. *BMJ* 1992;**304**:1689.
37 Berlin J, Laird NM, Sacks HS, Chalmers TC. A comparison of statistical methods for combining event rates from clinical trials. *Stat Med* 1989;**8**:141–51.
38 Fleiss JL. The statistical basis of meta-analysis. *Stat Meth Med Res* 1993;**2**:121–45.
39 Rosenthal R. An evaluation of procedures and results. In: Wachter KW, Straf ML, eds. *The future of meta-analysis*. New York: Russel Sage Foundation, 1990:123–33.
40 Jenicek M. Meta-analysis in medicine. Where we are and where we want to go. *J Clin Epidemiol* 1989;**42**:35–44.
41 Gelber RD, Goldhirsch A. Interpretation of results from subset analyses within overviews of randomized clinical trials. *Stat Med* 1987;**6**:371–8.
42 Piantadosi S, Byar DP, Green SB. The ecological fallacy. *Am J Epidemiol* 1988;**127**:893–904.
43 Chalmers I. Randomised controlled trials of fetal monitoring 1973–1977. In: Thalhammer O, Baumgarten K, Pollak A, eds. *Perinatal medicine*. Stuttgart: Thieme, 1979:260–5.
44 MacDonald D, Grant A, Sheridan-Pereira M, Boylan P, Chalmers I. The Dublin randomised controlled trial of intrapartum fetal heart rate monitoring. *Am J Obstet Gynecol* 1985;**152**:524–39.
45 Beta-Blocker Pooling Project Research Group. The beta-blocker pooling project (BBPP): subgroup findings from randomized trials in post-infarction trials. *Eur Heart J* 1988;**9**:8–16.
46 Goldstein S. Review of beta blocker myocardial infarction trials. *Clin Cardiol* 1989;**12**:54–7.
47 Soriano JB, Hoes AW, Meems L, Grobbee DE. Increased survival with beta-blockers: importance of ancillary properties. *Prog Cardiovasc Dis* 1997;**39**:445–56.
48 Freemantle N, Cleland J, Young P, Mason J, Harrison J. Beta blockade after myocardial infarction: systematic review and meta regression analysis. *BMJ* 1999;**318**:1730–7.
49 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.
50 Murphy DJ, Povar GJ, Pawlson LG. Setting limits in clinical medicine. *Arch Intern Med* 1994;**154**:505–12.
51 Gruppo Italiano per lo Studio della Streptochinasi nell'Infarto Miocardico (GISSI). Effectiveness of intravenous thrombolytic treatment in acute myocardial infarction. *Lancet* 1986;**i**:397–402.
52 ISIS-2 Collaborative Group. Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17187 cases of suspected acute myocardial infarction: ISIS-2. *Lancet* 1988;**ii**:349–60.
53 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.
54 Egger M, Davey Smith G. Meta-analysis: potentials and promise. *BMJ* 1997;**315**:1371–4.

Part I: Systematic reviews of controlled trials