Meta-analysis
Potentials and promise

Matthias Egger, George Davey Smith

The number of papers published on meta-analyses in medical research has increased sharply in the past 10 years (fig 1). The merits and perils of the somewhat mysterious procedure of meta-analysis, however, continue to be debated in the medical community.[1-3] What, then, is meta-analysis? A useful definition was given by Huque: "A statistical analysis that combines or integrates the results of several independent clinical trials considered by the analyst to be 'combinable.'"[4] The terminology, however, is still debated, and expressions used concurrently include "overview," "pooling," and "quantitative synthesis." We believe that the term meta-analysis should be used to describe the statistical integration of separate studies, whereas "systematic review" is most appropriate for denoting any review of a body of data that uses clearly defined methods and criteria (box). Systematic reviews can include meta-analyses, appraisals of single trials, and other sources of evidence.[5] In this article we examine the potentials and promise of meta-analysis of randomised controlled trials. In later articles of this series we will consider the practical steps involved in meta-analysis,[7] examine various extensions beyond the calculation of a combined estimate,[8] address potential biases and discuss strategies to detect and minimise the influence of these in meta-analysis of randomised trials[9] and of observational studies.[10] We will conclude with a discussion of unresolved issues and future developments.[11] Details of relevant software will appear on the *BMJ*'s website at the end of the series.

Historical notes

Efforts to pool results from separate studies are not new. In his account on the preventive effect of serum inoculations against enteric fever, statistician Karl Pearson, was in 1904 probably the first researcher reporting the use of formal techniques to combine data from different samples. The rationale put forward by Pearson for pooling studies 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
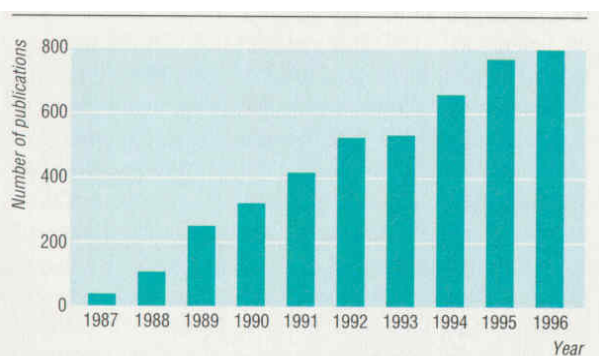
Summary points

Well conducted meta-analysis allows for a more objective appraisal of the evidence, which may lead to resolution of uncertainty and disagreement

Meta-analysis may reduce the probability of false negative results and thus prevent undue delays in the introduction of effective treatments into clinical practice

Meta-analysis of a large number of individual studies or of individual patient data allows testing of a priori hypotheses regarding treatment effects in subgroups of patients

Heterogeneity between study results may be explored and sometimes explained

Promising research questions to be addressed in future studies may be generated, and the sample size needed in future studies may be calculated accurately

opinion being formed at all, having regard to the size of the probable error involved."[12]

The first meta-analysis assessing the effect of a therapeutic intervention was published in 1955; interestingly, the treatment being evaluated was the placebo.[13] A simple average was calculated of the effectiveness of placebos in such diverse conditions as postoperative wound pain, cough, and angina pectoris: the placebo was apparently effective in 35% of patients. The development of more sophisticated statistical techniques, however, took place in the social sciences, in particular in education research, in the 1970s. The term meta-analysis was coined in 1976 by the psychologist Glass.[14] Meta-analysis was rediscovered by medical researchers to be used mainly in randomised clinical trial research, particularly in the fields of cardiovascular disease,[15] oncology,[16] and perinatal care.[17] Meta-analysis of observational studies[18] and "cross design synthesis" (the integration of observational data with the results from meta-analyses of randomised clinical trials[19 20]) have also been advocated.

More recently, a network of clinicians, epidemiologists, and other health professionals has been set up. The Cochrane Collaboration (named after Archie Cochrane, a pioneer in the field of evaluation of medical interventions) aims to prepare, maintain and disseminate comprehensive and systematic reviews of the effects of health care.[21 22] Since the foundation of the Cochrane Centre in Oxford in October 1992, this initiative has been growing rapidly, with the foundation of 15 further centres in Europe, North and Latin America, Africa, and

Fig 1 Number of publications about meta-analysis, 1987-96 (results from Medline search using text word and medical subject heading "meta-analysis")

What's in a name? The case for "meta-analysis"

The term meta-analysis for statistically combining and analysing data from separate studies is appropriate because:
• The term makes sense. "Meta" implies something occurring later, more comprehensive, and is often used to name a new but related discipline designated to deal critically with the original one
• The alternative terms are less specific or less poignant—for example, "overview" is also used for traditional reviews, and "pooling" incorrectly implies that the source data are merged
• "Meta-analysis" has recently been included as a Medical Subject Heading (MeSH) and publication type within the Medline indexing system of the National Library of Medicine.[5]
• "Systematic review" denotes any type of review that has been prepared using strategies to avoid bias and that which includes a material and methods section.[6] Systematic reviews may or may not include formal meta-analyses
• "Meta-analysis" is a useful term for describing a possible component of systematic reviews, and distinguishing between the two terms contributes to methodological clarity

Australia and hundreds of individuals from all over the globe collaborating in review groups.

The unacceptable face of "statisticism"?

Despite its widespread use, meta-analysis continues to be a controversial technique. While some exponents feel that meta-analysis should "replace traditional review articles of single topic issues whenever possible,"[23] others think of it as a "a new bête noir," which represents "the unacceptable face of statisticism" and "should be stifled at birth." [24] This mixed reception is not surprising. The pooling of results from a particular set of studies may be inappropriate from a clinical point of view, producing a population "average" effect, while the clinician wants to know how to best treat his or her particular patient. Meta-analyses of the same issue may reach opposite conclusions, as shown by assessments of low molecular weight heparin in the prevention of perioperative thrombosis[25][26] and of second line antirheumatic drugs in the treatment of rheumatoid arthritis.[27][28] It is nevertheless clear that for maximum benefit to be obtained from prior research, sound reviewing strategies must become more accessible and highly valued.

Narrative reviews

Traditional narrative reviews have several disadvantages that meta-analyses appear to overcome. The classic review is subjective and therefore prone to bias and error.[29] Without guidance by formal rules, reviewers can 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: 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, as it ignores sample size, effect size, and research design. It is thus hardly surprising that reviewers using traditional methods often reach opposite conclusions[32] and miss small, but poten-

tially important, differences.[33] Clinical medicine is riddled with controversies, with reviews often being commissioned to end an argument. However, in controversial areas the conclusions drawn from a given body of evidence may be associated more with the specialty of the reviewer than with the available data.[23] By integrating the actual evidence, meta-analysis allows a more objective appraisal, which can help to resolve uncertainties when the original research, classic reviews, and editorial comments disagree.

Limitations of a single study

A single study often cannot detect or exclude with certainty a modest, albeit clinically relevant, difference in the effects of two treatments. A trial may thus show no significant treatment effect when in reality such an effect exists—that is, it may produce a false negative result. In this case we are dealing with a type II error, whose probability of occurrence can be calculated for a given difference in treatment effect, study size, and significance level. Generally better recognised is the type I error—when a trial produces a significant difference due to chance—whose probability corresponds to the probability (P) value. An examination of clinical trials that reported no significant differences between experimental and control treatment has shown that type II errors in clinical research are common: for a clinically relevant difference in outcome, the a priori probability of missing this effect (given the trial size) was greater than 20% in 115 of the 136 trials examined.[34] The number of patients included in clinical trials is thus often inadequate, a situation that has changed little over recent years. In some cases, however, 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 Britain alone. To detect such an effect with 90% certainty (that is, with a type II error of no more than 10%) over 10 000 patients in each treatment group would be needed.[35]

The meta-analytic approach seems 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 several 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.

Meta-analysis can also contribute to considerations about the generalisability of study results. The findings of a particular study may 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 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 about whether an overall study result varies among subgroups—for example, among men and women, older and younger patients, or subjects with different degrees of severity of disease. As discussed later in this series,[x] these questions can be addressed in the analysis and often lead to insights beyond what is provided by the calculation of a single combined effect estimate.

Epidemiology of results

Meta-analysis thus not only consists of the combination of data but includes the epidemiological exploration and evaluation of results—the "epidemiology of results," whereby the findings of an original study replace the individual as the unit of analysis.[16] New hypotheses that were not posed in the single studies can thus be tested in meta-analyses. However, although the studies included may be controlled experiments, the meta-analysis itself is subject to many biases inherent in observational studies.[17] Meta-analysis can, nevertheless, lead to the identification of the most promising or urgent research question, and may permit a more accurate calculation of the sample sizes needed in future studies. This is illustrated by an early meta-analysis of four trials that compared different methods of monitoring the fetus during labour.[18] 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.[19]

A more transparent appraisal

One benefit of meta-analysis is that it renders an important part of 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 meta-analysis readers should be able 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 by meta-analysis leads to the replacement of unhelpful descriptors such as "no relation," "some evidence of a trend," "a weak relation," and "a strong relation," with reproducible numerical values.[16] Furthermore, performing a meta-analysis may lead to reviewers moving beyond the conclusions that authors present in the abstract of papers, to a thorough examination of the actual data. As research assistants cannot be sent away with file cards to return with abridged conclusions, Rosenthal has suggested that this will lead to a "decrease in the splendid detachment of the full professor"[16]—in other words to a stronger involvement of the reviewers in the individual study results. As meta-analysis becomes a standard procedure, however, the splendid detachment may soon be restored.

Cumulative meta-analysis

Cumulative meta-analysis is defined as the repeated performance of meta-analysis whenever a new trial becomes available for inclusion. Such cumulative meta-analysis can retrospectively identify the point in time when a treatment effect first reached conventional levels of significance. For example, Lau and colleagues showed that for the trials of intravenous streptokinase in acute myocardial infarction, a significant (P = 0.01) combined difference in total mortality had been achieved by 1973[11] (fig 2). 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[12][13] and enrolled a total of 34 542 additional patients, reduced the significance level to P = 0.001 in 1979, P = 0.0001 in 1986, and to P < 0.00001 when the first very large trial appeared, narrowing the confidence intervals around an essentially unchanged 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[12] 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[13] (fig 2).

A similar picture is apparent in the case of β blockade in secondary prevention of myocardial infarction. In 1981 an influential editorial stated that "despite claims that they reduce arrhythmias, cardiac work, and infarct size, we still have no clear evidence that β blockers improve long-term survival after infarction despite almost 20 years of clinical trials."[41] Retrospective cumulative meta-analysis, however, shows that a significant beneficial effect (P = 0.02) was evident by 1977, and that the combined effect estimate was already both clinically important and highly significant (odds ratio 0.71 (95% confidence interval 0.59 to 0.84), P = 0.0001) in 1981 (fig 3). Subsequent trials in a further 13 113 patients only confirmed this result.

Another application of cumulative meta-analysis has been to correlate the accruing evidence with the recommendations made by experts in review articles and textbooks. Antman and colleagues showed for
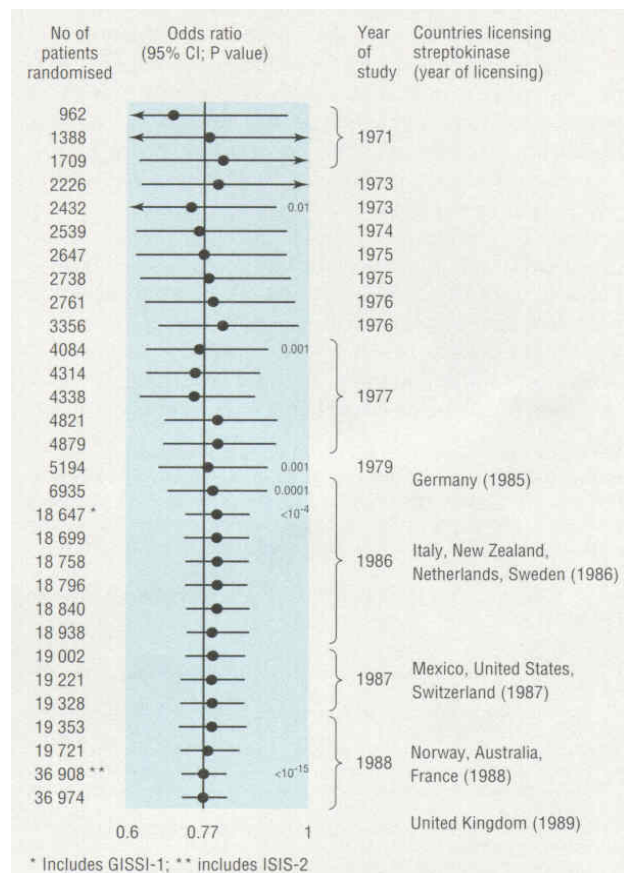

Fig 2 Cumulative meta-analysis of total mortality results from randomised controlled trials of intravenous streptokinase in myocardial infarction

Year	Trials	Cumulative odds ratios (95% CI)
1972	A	
1974	B	
1974	C	
1977	D	P=0.02
1980	E	
1980	F	
1981	G	P=0.0001
1982	H	
1982	I	P<0.0001
1982	J	P<0.00001
1982	K	
1983	L	
1983	M	
1984	N	
1985	O	
1987	P	
1990	Q	P<0.00001

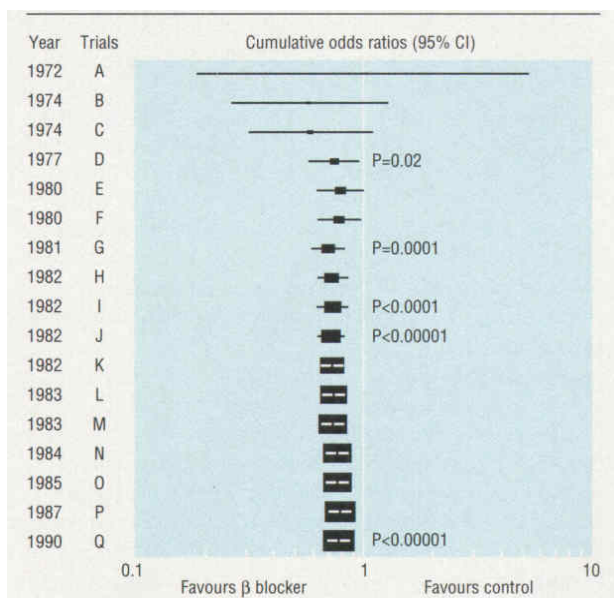0.1   1   10

Favours β blocker   Favours control

Fig 3 Cumulative meta-analysis of total mortality results from randomised controlled trials of oral β blockers after myocardial infarction. The size of the square reflects the amount of statistical information available at a given point in time

thrombolytic drugs that recommendations for routine use first appeared in 1987, 14 years after a significant (P = 0.01) beneficial effect became evident in cumulative meta-analysis.[45] 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.

These examples have been taken to suggest that further studies in large numbers of patients may be at best superfluous and costly, if not unethical,[46] once a significant treatment effect is evident from meta-analysis of the existing smaller trials. There are several other examples, however, of meta-analyses showing benefit of statistical significance and clinical importance that were later contradicted by large randomised trials.[17] Meta-analysis clearly has advantages over conventional narrative reviews and carries considerable promise as a tool in clinical research and health technology assessment. Meta-analysis is not an infallible tool, however, as will be discussed later in this series.

We thank Dr T Johansson and G Enocksson (Pharmacia, Stockholm) and Drs A Schirmer and M Thimme (Behring, Marburg) for providing data on licensing of streptokinase in different countries. The department of social medicine at the University of Bristol is part of the Medical Research Council's health services research collaboration.

1 Naylor CD. Meta-analysis and the meta-epidemiology of clinical research. BMJ 1997;315:617-9.
2 Bailar JC. The promise and problems of meta-analysis [editorial]. N Engl J Med 1997;337:559-61.
3 Meta-analysis under scrutiny [editorial]. Lancet 1997;350:675.
4 Huque MF. Experiences with meta-analysis in NDA submissions. Proceedings of the Biopharmaceutical Section of the American Statistical Association 1988;2:28-33.
5 Dickersin K, Berlin JA. Meta-analysis: state-of-the-science. Epidemiol Rev 1992;14:154-76.
6 Chalmers I, Altman DG. Foreword. In: Chalmers I, Altman DG, eds. Systematic reviews. London: BMJ Publishing, 1995.
7 Egger M, Davey Smith G. Phillips AN. Meta-analysis: principles and procedures. BMJ 1997 (in press).
8 Davey Smith G, Egger M, Phillips AN. Meta-analysis: beyond the grand mean? BMJ 1997 (in press).
9 Egger M, Davey Smith G. Meta-analysis: bias in location and selection of studies. BMJ 1997 (in press).
10 Egger M, Schneider M, Davey Smith G. Meta-analysis: spurious precision? Meta-analysis of observational studies. BMJ 1997 (in press).
11 Davey Smith G, Egger M. Meta-analysis: unresolved issues and future developments. BMJ 1997 (in press).
12 Pearson K. Report on certain enteric fever inoculation statistics. BMJ 1904;3:1243-6.
13 Beecher HK. The powerful placebo. JAMA 1955;159:1602-6.
14 Glass G. Primary, secondary and meta-analysis of research. Educ Res 1976;5:3-8.
15 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.
16 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.
17 Chalmers I, Enkin M, Keirse M. Effective care during pregnancy and childbirth. Oxford: Oxford University Press, 1989.
18 Greenland S. Quantitative methods in the review of epidemiologic literature. Epidemiologic Reviews 1987; 9:1-30.
19 General Accounting Office. Cross design synthesis: a new strategy for medical effectiveness research. Washington D.C. G.O.A. 1992.
20 Cross design synthesis: a new strategy for studying medical outcomes [editorial]? Lancet 1992;340:944-6.
21 Chalmers I, Dickersin K, Chalmers TC. Getting to grips with Archie Cochrane's agenda. BMJ 1992;305:786-8.
22 Bero L, Rennie D. The Cochrane Collaboration. Preparing, maintaining, and disseminating systematic reviews of the effects of health care. JAMA 1995;274:1935-8.
23 Chalmers TC, Frank CS, Reitman D. Minimizing the three stages of publication bias. JAMA 1990;263:1392-5.
24 Oakes M. Statistical inference: a commmentary for the social and behavioural sciences. Chichester: Wiley, 1986.
25 Nurmohamed MT, Rosendaal FR, Bueller HR, Dekker E, Hommes DW, Vandenbroucke JP, et al. Low-molecular-weight heparin versus standard heparin in general and orthopaedic surgery: a meta-analysis. Lancet 1992;340:152-6.
26 Leizorovicz A, Haugh MC, Chapuis F-R, Samama MM. Boissel J-P. Low molecular weight heparin in prevention of perioperative thrombosis. BMJ 1992;305:913-20.
27 Felson DT, Anderson JJ, Meenan RF. The comparative efficacy and toxicity of second-line drugs in rheumatoid arthritis. Arthritis Rheum 1990;33:1449-61.
28 Götzsche 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.
29 Teagarden JR. Meta-analysis: whither narrative review? Pharmacotherapy 1989;9:274-84.
30 Ravnskov U. Cholesterol lowering trials in coronary heart disease: frequency of citation and outcome. BMJ 1992;305:15-9.
31 Götzsche PC. Reference bias in reports of drug trials. BMJ 1987;295:654-6.
32 Mulrow CD. The medical review article: state of the science. Ann Intern Med 1987;106:485-8.
33 Cooper HM, Rosenthal R. Statistical versus traditional procedures for summarising research findings. Psychol Bull 1980;87:442-9.
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.
35 Collins R, Keech A, Peto R, Sleight P, Kjekshus J, Wilhelmsen L, et al. Cholesterol and total mortality: need for larger trials. BMJ 1992;304:1689.
36 Jenicek M. Meta-analysis in medicine. Where we are and where we want to go. J Clin Epidemiol 1989;42:35-44.
37 Gelber RD, Goldhirsch A. Interpretation of results from subset analyses within overviews of randomized clinical trials. Stat Med 1987;6:371-8.
38 Chalmers I. Randomised controlled trials of fetal monitoring 1973-1977. In: Thalhammer O, Baumgarten K, Pollak A, eds. Perinatal medicine. Stuttgart: Thieme, 1979:260.
39 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.
40 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.
41 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.
42 Gruppo Italiano per lo Studio della Streptochinasi nell'Infarto Miocardico (GISSI). Effectiveness of intravenous thrombolytic treatment in acute myocardial infarction. Lancet 1986;397-402.
43 ISIS-2 Collaborative Group. Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17,187 cases of suspected acute myocardial infarction: ISIS-2. Lancet 1988;ii:349-60.
44 Mitchell JRA. Timolol after myocardial infarction: an answer or a new set of questions? BMJ 1981;282:1565-70.
45 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.
46 Murphy DJ, Povar GJ, Pawlson LG. Setting limits in clinical medicine. Arch Intern Med 1994;154:505-12.
47 Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997;315:629-34.