Journal of Epidemiology and Community Health 1991; 45: 89–92

REVIEW ARTICLE:*Research methods in epidemiology, V*

The potential and limitations of meta-analysis

Tim D Spector, Simon G Thompson

We are currently witnessing an "epidemic" of meta-analyses and overviews in the scientific literature. This is a relatively new phenomenon and this article addresses some of the important issues raised by their increasing use. In particular the differing applications and limitations of meta-analysis are discussed, with a review of the analytic methods used and the problems and biases encountered.

What is meta-analysis?

Meta-analysis has come to refer to the combining of results from a number of experiments or studies examining the same question. Such a process is not new, and some meta-analytic studies were reported as early as 1955.[1] However, only since the term meta-analysis was first used in 1976[2] has the technique become recognised as an analytical method. Meta-analysis is a discipline that reviews critically and combines statistically the results of previous research in an attempt to summarise the totality of evidence relating to a particular medical issue. The term meta-analysis is now often used synonymously with overview.

Why use meta-analysis?

Traditionally, when seeking advice in controversial or novel areas, clinicians and scientists have relied heavily on "informed" editorials or narrative reviews. There is now good evidence to suggest that these traditional methods are subject to bias and inaccuracy.[3] Reviewers using traditional methods are less likely to detect a small but significant effect or difference compared with reviewers using formal statistical techniques.[4] In controversial topics, such as reviews of the uses of new procedures, the enthusiasm for the procedure may be associated more with the specialty of the reviewer than with the results of the trials.[5] As most current medical reviews do not use scientific methods to assess and present data, different reviewers often reach different conclusions based on the same data.[6] For these reasons some formal statistical process of review should replace the informal approach. Meta-analysis can be used to resolve uncertainty when reports, editorials or reviews disagree.

Although the randomised controlled clinical trial is now accepted as the gold standard method of assessing therapeutic regimes, individual trials may produce false positive or negative conclusions. Small numbers and the consequent lack of power of any individual study is usually the main problem area.[7] The problem of small numbers is particularly relevant when dealing with subgroup analysis, for which very often the randomised controlled trial was not designed. Combining the results of comparable trials or studies can reduce random sampling errors that may predominate in any individual study. The larger the sample size available, the more precise the estimate of the effect, and the hypothesis of subgroup effects can be more reliably investigated.

It has been suggested by some authors that *only* randomised controlled trials should be subjected to meta-analysis.[8] However this restriction is not desirable; aetiological meta-analyses (ie, of case-control or prospective studies) have recently been carried out, usually to clarify inconsistent findings or to estimate the true effect of a risk factor. However the interpretation of a meta-analysis of randomised controlled trials is usually simpler. If all relevant clinical trials are included and these are free from bias (ie, trials are randomised, all randomised individuals are included in "intention to treat" analyses, and outcome assessments are objective or blinded), a meta-analysis will give an unbiased assessment of a treatment's efficacy.[9] In observational epidemiology, potential bias in individual studies (through confounding, misclassification, or other causes) will always remain a problem, especially when effect sizes are small. If such biases are to an extent consistent over different studies, a meta-analysis will reflect both the true effect and the biases. However the increasing use of meta-analysis in observational studies should encourage the more formal reporting of aetiological studies, to facilitate the combining of such results. Indeed the direct comparison of results from meta-analyses of randomised controlled trials and of the related observational studies is a novel and informative advance.[10 11]

Examples of meta-analysis

There are now many examples of meta-analysis in a great variety of medical specialities that demonstrate their potential usefulness. One of the early important studies concerned the use of β blockers in myocardial infarction[12 13] which showed the efficacy of post-discharge treatment by combining the results of over 60 small studies. It also produced a useful framework for future studies. Another meta-analysis has concluded that steriods are of benefit in meningitis in children,[14] another that H_2 antagonists are of only minor benefit in the treatment of gastrointestinal

Department of Environmental and Preventive Medicine, St Bartholomew's Hospital Medical College, Charterhouse Square, London EC1M 6BQ
T D Spector
Medical Statistics Unit, London School of Hygiene and Tropical Medicine, Keppel St, London WC1
S G Thompson

Correspondence to:
Dr Spector

Accepted for publication
October 1990

haemorrhage, and only in gastric ulcers.[15] Although the vast majority of meta-analyses concern the assessment of therapies in randomised controlled trials, a few studies have addressed contentious aetiological issues such as the quantification of the effect of passive smoking on the risk of lung cancer,[16] alcohol in breast cancer,[17] the oral contraceptive pill in rheumatoid arthritis,[18] and leukaemia in refinery workers.[19]

Study design in meta-analysis

With the proliferation of meta-analyses, it has become apparent that their design, methods and publication should be conducted in a rigorous scientific manner, akin to that currently expected of randomised controlled trials. This is to allow critical appraisal of each individual meta-analysis in terms of its methodology and therefore the validity of its conclusions. A meta-analysis should be a research study in its own right. Specific a priori aims should be set out and a working protocol established.

Having defined the aims of the study, a thorough search of relevant publications needs to be performed. Computer searches have aided the inclusion of large numbers of trials in published meta-analyses. However, several studies have shown that less than two thirds of relevant trials are uncovered by computer searches.[20] Therefore computer searches should be supplemented by the bibliographies of textbooks, reviews, and the studies themselves, and information from specialists in the field. Where possible databases of ongoing clinical trials should be consulted.

In order to reduce bias, the inclusion of studies should be based on predetermined criteria. For example in clinical trials, evidence of randomisation is usually regarded as crucial[21]; in some situations a minimum study size might be desirable. Ideally all studies should be assessed in a blinded fashion by independent observers, although this is often difficult and impractical to perform. The decision to include studies should consider whether treatments, outcomes, and case definitions are similar enough to be combined. Opinions will of course differ as to how strict inclusion criteria should be. Some argue that where certain methodological differences occur it is wrong to produce summary estimates.[22] Others argue that the more varied the studies included, the more generalisable and applicable the results.[23] Differences between studies are likely to result in differences in the size, rather than the direction of the effect.[24] Peto has also pointed out the tendency for trials addressing related questions to produce answers in a similar direction, despite methodological variations.[9] It is reassuring to note however that different meta-analyses of the same subject, that differ in the number of trials included, usually reach similar conclusions.[25]

At present most meta-analyses performed do not take into account the quality of the individual studies included, and results are weighted simply in favour of the large study over the small. In principle it would seem desirable to down weight those studies of "doubtful" quality relative to "good" quality studies, because of their greater likelihood of bias. Some authors have proposed that studies can be weighted in terms of independently assessed "quality", derived from a large number of predetermined "quality" criteria.[26] The pooled estimate can then be adjusted accordingly, or else the quality score used to exclude studies. A simpler method for trials has been proposed which concentrates on three areas of potential bias, namely treatment allocation by randomisation, inclusion of all randomised individuals in analysis, and the blindness of the outcome assessments.[27] Quality assessments have also been used in epidemiological studies.[17 28] The major problem with quality weighting is that it must remain arbitrary and to an extent subjective. A single choice of weights is difficult to justify; for example, is it worse to have poor blinding or poor randomisation? Moreover the procedure goes against the general purpose of meta-analysis, that is to obtain an *objective* summary of the available evidence. Because of the time and resources needed to undertake full quality assessment, its routine use cannot be recommended unless its true worth becomes established.

Publication bias

Publication bias is a potential problem in all meta-analyses.[29 30] It arises from the fact that unpublished papers may contradict the findings of the overview due to the overrepresentation of published "positive" (ie, statistically significant) studies. There is now good evidence that negative studies in medicine are less likely to be published than positive ones.[31 32] The likelihood of this bias altering the conclusions will depend on the chances of the existence of important numbers of unpublished papers. This is less likely to occur when the result is of considerable importance (eg, vitamin supplementation and neural tube defects)[33] or when the questions can only be answered by large costly studies which are likely to reach publication (eg, trials of thrombolytics on cardiovascular mortality).

The question of publication bias needs to be addressed in all meta-analyses and its importance considered. There are now several methods of confronting the problem. One involves a simple calculation of the number of studies needed to refute the conclusions of the meta-analysis.[34] Another method is a visual one based on a "funnel plot", an example of which is given by Vandenbroucke.[35] The basis of this is that if the observed effect sizes are plotted according to sample size they should scatter around an underlying "true" value, producing a funnel pattern. Gaps in the plot indicate potential unpublished studies and the possibility of bias. Begg also produced a quantitative method of estimating the maximum potential effect of publication bias using the sample size of the study and an estimate of the size of the source population.[30 36] The problems of this method are that information is needed on specific incidence rates and the proportion of a population who would enrol in a trial, and these details are not usually available with any accuracy.

Another approach has been to seek out and include all unpublished studies performed, either from abstracts of meetings or by direct correspondence from other investigators. Although less open to publication bias, a new

problem of data quality is encountered. The decision to use abstracts or study summaries is a contentious one. Some editors have advised against their use in referencing.[37] About half of all abstracts never appear as full publications.[29] Chalmers and coworkers attempted to identify factors which determined subsequent full publication of clinical trial abstracts in the perinatal field.[38] They were unable to detect any differences in methodological quality, but did find that sample size was a significant factor in determining publication. The effect of sample size has also been shown by others.[29] However although small studies are more likely to remain unpublished, those with large effects may be preferentially published.[39] Obtaining information from the authors of unpublished studies has other inherent problems, as information obtained from an investigator may be subject to selection bias, both on the part of the meta-analyst and the original researcher. A meta-analyst thus has to weigh up the risks of including biased data (while increasing the power of the study) against the risks of publication bias.

Theoretically publication bias could be prevented or markedly reduced if researchers reported all studies undertaken and journals accepted papers based on methods rather than results. These ideals may be a long way off, and perhaps the most practical step would be the extension of clinical trial registers into other fields and disciplines.[40 41]

Statistical methods

The first step in meta-analysis should simply be to display the estimated treatment effects, together with their confidence intervals, for each study. Although the smallest and least informative studies have wide confidence intervals that tend to dominate the diagram visually, the careful inspection of such displays often prompts most of the conclusions that will emerge from a numerical analysis. There are two general philosophies for producing a combined estimate of effect and its confidence interval, the so called "fixed effect" and "random effects" methods. They differ in their assumptions about the true underlying treatment effects in the different studies.

In the *fixed effect* method, all the studies are assumed to be estimating the same underlying treatment effect. In this situation, the most precise overall average of observed treatment effects is obtained by weighting each individual treatment effect inversely according to its variance.[42] This can be applied directly, for example, to log odds ratios as summaries of each trial's observed treatment effect. Logistic regression is also sometimes used,[17] and is in fact equivalent to such an analysis. The Mantel-Haenszel method weights the odds ratios (not their logarithms) approximately inversely according to their variances[43]; in many instances the choice between odds ratios and log odds ratios is unimportant.

Peto's "observed minus expected" (O-E) method[13 44] is equivalent to the Mantel-Haenszel test. For each study, the "observed" number of events in the treated or exposed group is compared with that "expected" if the treatment or exposure had no effect. If the observed numbers (O) differ systematically from the expected numbers (E), this provides evidence of an effect of treatment. A test is provided by totalling the O-E differences, and their variances, across the studies to see if the totalled (O-E) differs more from than zero than is compatible with chance. The calculations are thus easy to perform and to present. A disadvantage for general use is that the approximation provided for the overall estimate of odds ratio is not a good one if the odds ratio is far from unity[45]; this is most unlikely to be a problem in clinical trials, but could be in meta-analyses of epidemiological studies.

The choice between these fixed effect methods would rarely materially affect the conclusions being drawn. A more important consideration is the possibility of heterogeneity between the studies, that is failure of the asssumption underlying all the fixed effect methods. The evidence for heterogeneity, ie, the systematic differences between the underlying true treatment effects in different studies, can be assessed formally using a χ^2 statistic.[46] However the test lacks power, and even in the absence of "statistically significant" heterogeneity, one may want to explore the analysis further. One approach is to attempt to "explain" the heterogeneity in terms of characteristics of the studies or the patients included. If such divisions reveal possible sources of heterogeneity, interpretation is necessarily cautious because analyses are "post hoc", that is, inspired by looking at the data.

Often the sources of any heterogeneity are intangible. If so, it may be difficult to justify a single combined estimate for all the studies. One formal approach is the *random effects* method[47] in which both a between study variance and the within study variances are taken into account in deriving the weighting given to each study. However, the method cannot be regarded as a panacea for heterogeneity. The between study variance, estimated from the χ^2 statistic for heterogeneity, is itself imprecise and, being often strongly dependent on the inclusion or exclusion of small studies, is susceptible to the effects of publication bias. Also, the representation of differences between studies by a single variance is conceptually inadequate.

The numerical methods used in meta-analysis are therefore most reasonably based on the following sequence. A fixed effect method may be used initially, but it should be followed by an assessment of heterogeneity. The random effects method may then be useful in assessing the robustness of the initial conclusions to failure in the assumption of no heterogeneity. If the conclusions from each method agree, there is naturally greater confidence in them; if not, that the interpretation is problematic should be made explicit.

Conclusions

Meta-analysis is here to stay. Epidemiologists, statisticians, and clinicians should all be aware of the uses and limitations of the technique. A useful by product of the growing use of this form of analysis has been the greater awareness of the need for consistency in the way clinical trials and epidemiological studies are presented, so that the

results from these studies can be combined. This will undoubtedly have the effect of improving the quality of methodology, assessment, and presentation of clinical research and the availability of study data for future meta-analysis. Despite the potential problems and pitfalls we have outlined, meta-analysis should play a leading role in the review of scientific issues. This necessitates a fuller understanding of meta-analysis as a routine analytical tool, but also a wider appreciation of the issues involved.

We would like to thank Dr Kay Dickersin of the University of Maryland for her advice and comments during the preparation of this manuscript.

1 Beecher HK. The powerful placebo. *JAMA* 1955; **159**: 1602–6.
2 Glass GV. Primary, secondary and meta-analysis of research. *Educ Res* 1976; **5**: 3–8.
3 Teagarden JR. Meta-analysis: whither narrative review? *Pharmacotherapy* 1989; **9**: 274–84.
4 Cooper HM, Rosenthal R. Statistical versus traditional procedures for summarising research findings. *Psychol Bull* 1980; **87**: 442–9.
5 Chalmers TC, Frank CS, Reitman D. Minimising the three stages of publication bias. *JAMA* 1990; **263**: 1392–5.
6 Mulrow CD. The medical review article: state of the science. *Ann Intern Med* 1987; **106**: 485–8.
7 Frieman 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 randomised controlled trial: survey of 71 negative trials. *N Engl J Med* 1978; **299**: 690–4.
8 Bulpitt CJ. Meta-analysis. *Lancet* 1988; ii: 93–4.
9 Peto R. Why do we need systematic overviews of randomised trials? *Stat Med* 1987; **6**: 233–40.
10 MacMahon S, Peto R, Cutler J, *et al.* Blood pressure, stroke and coronary heart disease. Part 1, prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias. *Lancet* 1990; i: 765–73.
11 Collins R, Peto R, MacMahon S, *et al.* Blood pressure, stroke and coronary heart disease. Part 2, Short term reductions in blood pressure: overview of randomised drug trials in their epidemiological context. *Lancet* 1990; i: 827–38.
12 Yusuf S, Peto R, Lewis J, Collins R, Sleight P. Beta-blockade during and after myocardial infarction: an overview of the randomised trials. *N Engl J Med* 1987; **316**: 450–5.
13 Yusuf S, Collins R, Peto R, *et al.* Intravenous and intracoronary fibrinolytic therapy in acute myocardial infarction: Overview of results on mortality, reinfarction and side effects from 33 randomised controlled trials. *Eur Heart J* 1985; **6**: 556–85.
14 Havens PL, Wendelberger KJ, Hoffman GM, Lee MB, Chusid MJ. Corticosteriods as adjunctive therapy in bacterial meningitis. *Am J Dis Child* 1989; **143**: 1051–5.
15 Collins R, Langman M. Treatment with histamine H_2 antagonists in acute upper gastrointestinal haemorrhage. *N Engl J Med* 1985; **313**: 660–6.
16 Wald NJ, Nanchahal K, Thompson SG, Cuckle HS. Does breathing other people's tobacco smoke cause lung cancer? *BMJ* 1986; **293**: 1217–22.
17 Longnecker MP, Berlin JA, Orza MJ, Chalmers TC. A meta-analysis of alcohol consumption in relation to risk of breast cancer. *JAMA* 1988; **260**: 652–6.
18 Spector TD, Hochberg MC. A meta-analysis of the association between the oral contraceptive pill and the development of rheumatoid arthritis. *Clin Epidemiol* 1990; **43**: 1221–30.
19 Wong O, Raabe GK. Critical review of cancer epidemiology in petroleum industry employees, with a quantitative meta-analysis by cancer site. *Am J Indust Med* 1989; **15**: 283–310.
20 Dickersin K, Hewitt P, Mutch L, Chalmers I, Chalmers TC. Perusing the literature: comparison of medline searching with a perinatal clinical trials database. *Controlled Clin Trials* 1985; **6**: 306–17.
21 Sacks H, Chalmers TC, Smith H. Randomised versus historical controls for clinical trials. *Am J Med* 1982; **72**: 233–4.
22 Goldsman L, Feinstein AR. Anticoagulants and myocardial infarction: the problems of pooling, drowning and floating. *Ann Intern Med* 1979; **90**: 92–4.
23 Hedges LV. Commentary. *Stat Med* 1987; **6**: 381–5.
24 Chalmers I, Hetherington J, Elbourne D, Keirse MJ, Enkin M. Materials and methods used in synthesising evidence to evaluate the effects of care during pregnancy and childbirth. In: Chalmers I, Enkin M, Keirse MJ, eds *Effective care in pregnancy and childbirth Vol 1*. Oxford: Oxford University Press, 1989: 39–65.
25 Chalmers TC, Berrier J, Sacks HS, Levin H, Reitman D, Nagalingham R. Meta-analysis of clinical trials as a scientific discipline: replicate variability and comparison of studies that agree and disagree. *Stat Med* 1987; **6**: 733–44.
26 Chalmers TC, Smith H, Blackburn B, *et al.* A method for assessing the quality of a randomised control trial. *Controlled Clin Trials* 1981; **2**: 31–49.
27 Prendiville W, Elbourne D, Chalmers I. The effects of routine oxytocic administration in the management of the third stage of labour: an overview of the evidence from controlled trials. *Br J Obstet Gynaecol* 1988; **95**: 3–16.
28 Lichtenstein MJ, Mulrow CD, Elwood PC. Guidelines for reading case-control studies. *J Chron Dis* 1987; **40**: 893–903.
29 Dickersin K. The existence of publication bias and risk factors for its occurrence. *JAMA* 1990; **263**: 1385–9.
30 Begg CB, Berlin JA. Publication bias: a problem in interpreting medical data. *J R Stat Soc* 1988; **151**: 419–63.
31 Simes RJ. Publication bias: the case for an International Registry of clinical trials. *J Clin Oncol* 1986; **4**: 1529–41.
32 Dickersin K, Chann SS, Chalmers TC, Sacks HS, Smith H. Publication bias and clinical trials. *Controlled Clin Trials* 1987; **8**: 343–53.
33 Angell M. Negative studies (Editorial). *N Engl J Med* 1989; **321**: 464–6.
34 Rosenthal R. The "file drawer problem" and tolerance for null results. *Psychol Bull* 1979; **86**: 638–41.
35 Vandenbroucke JP. Passive smoking and lung cancer: a publication bias? *BMJ* 1988; **296**: 391–2.
36 Begg CB. A measure to aid the interpretation of published clinical trials. *Stat Med* 1985; **4**: 1–9.
37 Editorial. Uniform requirements for manuscripts submitted to biomedical journals. *Lancet* 1979; i: 428–31.
38 Chalmers I, Adams M, Dickersin K, *et al.* A cohort study of summary reports of controlled trials. *JAMA* 1990; **263**: 1401–5.
39 Berlin JA, Begg CB, Louis TA. An assessment of publication bias using a sample of published clinical trials. *J Am Stat Assoc* 1989; **84**: 381–92.
40 Chalmers I, Hetherington J, Newdick M, *et al.* The Oxford database or Perinatal Trials: developing a register of published reports of controlled trials. *Controlled Clin Trials* 1986; 7: 306–25.
41 Hubbard SM, Henney JE, DeVia VT. A computer database for information on cancer treatment. *N Engl J Med* 1987; **316**: 315–8.
42 Armitage P, Berry G. *Statistical methods in medical research.* Oxford: Blackwell Scientific Publications, 1987: 194–5.
43 Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst* 1959; **29**: 719–48.
44 Collins R, Yusuf S, Peto R. Overview of randomized trials of diuretics in pregnancy. *BMJ* 1985; **290**: 17–23.
45 Greenland S, Salvan A. Bias in the one-step method for pooling study results. *Stat Med* 1990; **9**: 247–52.
46 Berlin JA, Laird NM, Sacks HS, Chalmers TC. A comparison of statistical methods for combining event rates from clinical trials. *Stat Med* 1989; **8**: 141–51.
47 DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clin Trials* 1986; **7**: 177–88.