

The uses and abuses of meta-analysis

Bruce G Charlton

Charlton BG. The uses and abuses of meta-analysis. *Family Practice* 1996; **13**: 397–401.

Meta-analysis is a quantitative process of summary and interpretation which involves pooling information from independent studies concerning a single theme in order to draw conclusions. Greatly increased employment of meta-analysis is currently being advocated for clinical and policy decision making. However, the prestige of meta-analysis is based upon a false model of scientific practice. Interpreting empirical research is an extremely complex activity requiring clinical and scientific knowledge of the field in question; and teams of professional 'meta-analysts' with a primary skill base in information technology and biostatistics cannot take over this role. Meta-analysis is not a hypothesis-testing activity, and cannot legitimately be used to establish the reality of a putative hazard or therapy. The proper use of meta-analysis is to increase the precision of quantitative estimates of health states in populations. If used to estimate an effect, the reality of that effect should have been established by previous scientific studies. But the summary estimate from a meta-analysis can only be directly applied to a target population when the 'meta-protocol' and 'meta-population' match the target situation in all relevant particulars. These constraints can rarely be satisfied in practice, so the results of meta-analysis typically require adjustment—which is a complex, assumption-laden process that negates many of the statistical power advantages of a meta-analysis. Lacking any understanding or acknowledgement of the need for adjustment, most meta-analyses must be regarded as abuses of the technique.

Keywords. Meta-analysis, mega-trials, epidemiology, clinical science.

Introduction

Meta-analysis may conveniently be defined as a quantitative method of pooling information from independent studies concerning a single theme in order to draw conclusions. It is a two-stage process of summary and interpretation.

Opinion regarding the technique ranges between extremes of approbation and disdain. Many commentators agree with Olkin that a meta-analysis of randomized trials constitutes the best form of evidence regarding therapeutic effectiveness.¹ Others have argued that it is motivated by a quasi-alchemical urge to transmute the base metal of inadequate data into the gold (standard) of validated fact, suggested that it is mostly a rather mundane and second-rate kind of intellectual activity and undeserving of high prestige, or simply erupted "meta-analysis—schmeta-analysis!"^{2,3,4}

I will argue that the critics of meta-analysis are closer to the truth than are the evangelists. Meta-analysis has its uses, and may occasionally be valid and applicable to real clinical situations, but these circumstances are so rare that most published instances of the technique must be regarded as abuses.

Meta-analysis based on a false model of science

All commentators emphasize the difficulty of performing a valid meta-analysis, but the reasons given usually reveal a false model of scientific practice.^{5,6} Meta-analysis is often stated to be necessary due to the sheer amount of data generated by present-day research.^{1,7} Scientific practice is implied to involve a process of pooling or combining evidence from independent studies, then drawing conclusions based on the weight of evidence. If this were the case, then summarization would indeed be crucial and valid inference would become more difficult as the volume of research increased. This justification for overviews and meta-analyses is principally one of enabling increased efficiency in data assimilation.⁸ But this description of theoretical science is false.

In reality, the theoretical practice of science draws upon evidence from studies judged to be both relevant and valid—such studies are seldom common and usually well known to practitioners. This highly selected evidence is then taken into account in constructing and testing theoretical models which can be tested against experiment and observation.⁹ Most would-be evidence tends to be judged irrelevant to this process, and is deservedly ignored—certainly bad evidence is not pooled with the good.

Received 23 February 1996; Accepted 29 March 1996.
Department of Psychology, Ridley Building, University of Newcastle upon Tyne, Newcastle upon Tyne NE1 7RU, UK.

The ingredients which make up this process of qualitative judgement and inference have never adequately been described in explicit terms, and scientific practice includes much knowledge that is tacit, and implicit, learned by apprenticeship to other scientists and from experience working in the field.^{6,10} It can, however, be asserted with a high degree of certainty that the scientific process is not primarily a statistical one based upon summarization and combination of all relevant data.^{3,11}

Implicit assumptions of meta-analysis

Proponents of meta-analysis make much of the 'objectivity' of the technique, which derives from the explicit nature of its procedures when compared with most editorials, reviews and commentaries.^{6,12} The sheer quantity and range of sources of the cited literature in a meta-analysis may be very impressive. This is achievable partly because of advances in computer systems of information retrieval, but mostly by the employment of full-time research assistants whose job is to hand-search journals, network among researchers and (by other labour-intensive means) endeavour to unearth recondite and far-flung publications and projects.^{1,7,13}

The accumulation of data into one place which precedes the statistical manipulations of meta-analysis is frequently unprecedented in a given field. This creation of a complete catalogue may be valuable in itself, especially if it reveals an obvious consistency or pattern to the data which was not previously noticed (although such an oversight is unlikely in a mature scientific discipline). Some authors regard this activity of 'overviewing' evidence as contributing most of the value of meta-analysis, and have suggested that analysis should not go further than identifying a qualitative consistency of results across relevant studies.¹⁴ There is no methodological objection to this kind of elaborate and expensive literature survey, but when unaccompanied by original thought it constitutes a somewhat mediocre activity which bears the same relation to creative science that an undergraduate dissertation does to a PhD thesis.³

However, the defining feature of meta-analysis is not enumeration but interpretation and proponents of meta-analysis claim that it can perform this key task of selection and analysis of independent studies by means of algorithmic procedures and statistical summarization. Meta-analysis makes the underlying assumption that when the results of relevant studies differ, the true value lies 'latent' within the existing data but concealed from investigators: firstly, by their failure to overview the whole data set (including unpublished studies); second, by excessive random error in studies examined one at a time (due to studies containing too few subjects);

and third, by the lack of an optimal arrangement of evidence. In effect, the 'scientific truth' is conceptualized as a pattern that, once revealed, is unambiguous in its relevance and applicability so that the implications of research are transparent to any observer.

Meta-analysis therefore assumes that the diversity (or 'heterogeneity') among relevant research studies is randomly distributed around the 'true' value,¹⁵ so that errors in one direction in one study will tend to be balanced by errors in the other direction in other studies and therefore that appropriate statistical pooling and averaging will tend to produce an error-free (or at least error-reduced) estimate of the underlying, unbiased, 'true' value. Meta-analysis is thus indirectly but crucially predicated on a view of scientific truth as social consensus.⁶

But real scientific practice makes no such assumption about the random distribution of error between (or within) studies.⁹ Indeed, a more plausible assumption would be that most investigators tend to make the same errors in the same direction, and only a minority of the best scientists will perform studies to the highest standard. Instead of seeking consensus, the social structures of science have the effect (albeit an imperfect one) of subjecting studies to critical appraisal by the peer group, in order to winnow the wheat from the chaff.¹⁶ The production of scientific knowledge is a process closer to 'trial by ordeal' than trial by opinion poll.

Meta-analysis usurps theoretical science

The meta-analytic view of science leads to an assertion that the relevant techniques for understanding evidence are essentially informational and statistical.^{17,18} Therefore, meta-analyses tend to be organized, performed and published by teams with disciplinary backgrounds in epidemiology, computing and biostatistics—only secondarily supplemented by advice from workers in the substantive field being over-viewed. This is in sharp contrast to the specific scientific and clinical expertise and experience considered a prerequisite for the actual performance of primary medical research.^{3,4}

The bizarre result is that meta-analysis implies that theoretical and empirical science should be done by two different sets of people with different disciplinary abilities. In effect, empirical research is to be done by scientists and clinicians, and the interpretation of this research is to be performed by the likes of epidemiologists and statisticians who will decide what inferences may be drawn from the evidence.

The above scenario would only be credible if advocates of meta-analysis could point to a successful track record of theoretical advance—which they cannot; or if the major difficulties in evaluating research were amenable to standardized evaluation of studies and

adherence to correct statistical procedures—which they are not. The massive implausibility of the biostatistical approach to interpretation should be obvious to anybody who has experienced the difficulties of learning how to become a practising scientist. Interpretation is, perhaps, the hardest of all scientific skills to master. The ability to evaluate and compare research papers, and the capacity to use this to judge the current state of knowledge and frame hypotheses for future investigation, is a skill attained—if at all—only with effort and after a prolonged apprenticeship. The skill is also relatively specific with regard to subject matter.

The notion that scientific interpretation can be reduced to statistical considerations, checklists and step-by-step flow diagrams applicable to any problem at any time^{1,8,13,17,19} would be laughable were it not becoming accepted practice in some circles. Inventories are not a substitute for substantive knowledge. Clinical experience and that partly trained, partly instinctive, understanding of causes and insight into mechanisms which comes from personally grappling with the primary process of research are both elements that have time and again proved crucial to medical science.^{3,4,20-22}

Limitations of randomized trials

The limitations of a meta-analysis are dictated by the limitations of the epidemiological studies from which it has been assembled (on the basis of ‘garbage in, garbage out’). Randomized trials are generally assumed to be the ‘best’ epidemiological evidence regarding therapeutic effectiveness, and the methodology most amenable to meta-analysis.^{1,23} Methodological constraints which apply to the randomized controlled trial (RCT) will therefore, *mutatis mutandis*, also apply to meta-analyses of other epidemiological techniques such as cohort and case-control studies, and surveys.⁹

The major limitations characteristic of ‘mega-trials’ (large, multi-centred trials analysed by ‘intention to treat’^{23,24}) derive from poor experimental control and biased recruitment.^{21,25} Mega-trials employ a deliberately simplified experimental design in order to maximize recruitment and compliance, both of subjects and of collaborating trial centres. Due to logistic and ethical constraints, trials are performed on a study population that is typically unrepresentative of any actual ‘target population’ to which their results might be applied.

Inherent in mega-trial design is that experimental protocols do not attempt to exclude or hold constant all known sources of bias, but instead employ randomization of large numbers of subjects to distribute these potential biases equally between comparison groups. Comparisons between allocated treatments will be unbiased but at the price of conflating several causal processes, and measuring ‘intention’ to treat rather than

the effect of treatment. For instance, if age is an important confounder, mega-trials do not control for age, but randomize large numbers of differently aged subjects. The result is that the age distribution will tend to be balanced between allocation groups; but the effects of age will be conflated with the causal variable under study. The measured association will only be directly applicable to a target population with the same age structure as the study population.^{21,25}

Mega-trials should therefore be considered as descriptive and epidemiological in nature rather than analytical and scientific.^{9,14,21} Indeed, although it is an experiment, a mega-trial can most easily be understood and interpreted as if it were a special kind of survey designed to compare the outcomes when two or more protocols are allocated to a group of subjects. Randomization ensures that the comparison groups have equivalent population characteristics, and the large number of subjects allows a high degree of precision in estimating the therapy–outcome association. Generalizing from a mega-trial also resembles generalizing from a survey because both procedures depend crucially on the study population being representative of the target population.

A mega-trial does not, as a scientific experiment would, aim to isolate and measure a single causal variable linking a therapy and an outcome; the measured relationship between therapy and outcome is therefore an estimate of the magnitude of an association, not of a causal process. Consequently, mega-trials are not hypothesis-testing studies²¹—and a secondary mathematical summarization of trials, such as a meta-analysis, cannot be hypothesis-testing either.

Meta-protocols and meta-populations

We can now begin to delineate the legitimate uses of meta-analysis. The ‘overview’ stage is neither distinctive nor sufficient to define meta-analysis—quantitative interpretation is the crucial feature. Meta-analysis is essentially a method for pooling data in order to increase the precision of estimates. The summary statistic of a meta-analysis of RCTs therefore describes the (average) outcome of allocating a meta-protocol to a meta-population. Interpreting the summary statistic of a meta-analysis (i.e. ‘applying’ the estimate of effect) involves establishing that the meta-protocol and meta-population are comparable to the proposed intervention and the target population.

The nature of a meta-protocol is defined by the methodological parameters of the pooled individual therapeutic interventions of constituent mega-trials. In other words, the meta-protocol is a ‘virtual intervention’ in an experiment whose experimental rigour is the lowest common denominator defined by the pooled deficiencies of its component studies (the level of control being defined by the lowest permitted level of control,

not the average level of control). The meta-population is defined as that virtual group of subjects which has emerged after the overview population has been pooled from the component studies (with or without statistical weighting of individual studies).

In order for the estimate of the therapeutic effect of a meta-protocol to be applicable to a target population, the meta-population must be a representative sample of the target population. This requires either that the meta-population be a randomly selected sample of the target population, or that the meta-population be created from a balanced blend of individual study populations where relevant causal variables have been measured and assembled in their proper proportion.

Clearly, the vast majority of meta-populations in published meta-analyses are not representative of the target population, or indeed of any real-world population, because meta-analyses are assembled from a group of individual RCTs the populations of which are each unrepresentative (biased) to a significant and undetermined extent.²⁵ Estimates cannot then be generalized to any actual population without adjustment. Adjustment will need to involve quantification and subtraction of biases. For instance, if an estimate has been confounded by biases in the age structure of the meta-population compared with the target population, then the magnitude of confounding by age will need to be investigated, quantified and its effects removed from the analysis.

It is insufficiently appreciated that the process of 'adjustment for confounding' is not a purely mathematical manipulation, but is a form of quantitative modelling of the consequences of uncontrolled causal influences on the study. Adjustment introduces new assumptions into the analysis—causal assumptions which require validation in independent studies. Adjustment will therefore diminish precision of the estimate, somewhat defeating the object of the meta-analysis.

Conclusion

Meta-analyses of mega-trials yield estimates that apply only to group averages, not to individual patients, due to the high level of within-group heterogeneity of subjects in mega-trials and other epidemiological studies.^{21,25} This, in itself, means that a meta-analysis does not necessarily have any relevance to clinical practice. A bad meta-analysis, like any bad piece of research, may be useless or harmful; and, unfortunately, bad research tends to be more common than is good research.

But even accepting the population level of validity, a meta-analysis should be performed on independent studies each of which employs a qualitatively similar and therapeutically credible study design, and where the pooled trial population is representative of the target

population. Such a situation of between-study uniformity is extremely rare.²⁶

Furthermore, meta-analysis should not be used for testing hypotheses, but only for obtaining a more precise estimate of an effect which is already known to be present from well controlled, hypothesis-testing studies. This means that most meta-analyses are misuses of the technique. For instance, it is wrong (although common) to employ meta-analysis to determine whether a putative health risk is a genuine hazard, or whether a putative therapeutic intervention is genuinely effective.

Meta-analyses cannot make qualitative distinctions in cases where causation is doubtful. The epidemiological data from which meta-analyses are constructed measure association not causation, and are not sufficiently controlled to isolate and test hypotheses. Moreover, there are no valid, general-purpose algorithms nor statistical procedures for the interpretation of empirical research, so that most meta-analyses are underpinned by no more than the subjective opinion of investigators who are sometimes distinguished mainly by lacking the appropriate training, experience, approach and interest necessary to draw inferences from empirical research.

Meta-analysis, when all is said and done, is a technique with very restricted applicability to the clinical practice of medicine. In certain rare, well-understood and well-controlled circumstances it may provide an enhancement in the precision of estimates of group outcomes. But meta-analysis is always likely to mislead due the mismatch between its high statistical precision and low scientific validity.^{3,9}

References

- 1 Olkin I. Meta-analysis: reconciling the results of independent studies. *Stat Med* 1995; 14: 457-472.
- 2 Shapiro S. Meta-analysis/Schmeta-analysis. *Am J Epidemiol* 1993; 138: 673 (abstract).
- 3 Rosendaal FR. The emergence of a new species: the professional meta-analyst. *J Clin Epidemiol* 1994; 47: 1325-1326.
- 4 Feinstein AR. Meta-analysis: statistical alchemy for the 21st century. *J Clin Epidemiol* 1995; 48: 71-79.
- 5 Charlton BG. Management of science. *Lancet* 1993; 342: 99-100.
- 6 Charlton BG. Practice guidelines and practical judgement. *Br J Gen Pract* 1994; 44: 290-291.
- 7 Chalmers T. Haynes B. Reporting, updating and correcting systematic reviews of effects of health care. *Br Med J* 1994; 309: 862-865.
- 8 Mulrow CD. Rationale for systematic reviews. *Br Med J* 1994; 309: 597-599.
- 9 Charlton BG. The scope and nature of epidemiology. *J Clin Epidemiol* 1996 (in press).
- 10 Cromer A. *Uncommon sense: the heretical nature of science*. Oxford: Oxford University Press, 1993.
- 11 Van Valen LM. Why misunderstand the evolutionary half of biology? In Saarinen E (ed.) *Conceptual issues in ecology*. Dordrecht: The Netherlands, 1982.

- ¹² Friedenreich CM. Methods for pooled analysis of epidemiologic studies. *Epidemiology* 1993; 4: 295-302.
- ¹³ Dickerson K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. *Br Med J* 1994; 309: 1286-1291.
- ¹⁴ Thompson SG, Pocock SJ. Can meta-analyses be trusted? *Lancet* 1991; 338: 1127-1130.
- ¹⁵ Ahlbom A. Pooling epidemiological studies. *Epidemiology* 1993; 4: 283-284.
- ¹⁶ Ziman J. *Reliable knowledge: an exploration of the ground for belief in science*. Cambridge: Cambridge University Press, 1978.
- ¹⁷ Thompson SG. Why sources of heterogeneity in meta-analysis should be investigated. *Br Med J* 1994; 309: 1351-1355.
- ¹⁸ Victor N. Indications and contra-indications for meta-analysis. *J Clin Epidemiol* 1995; 48: 5-8.
- ¹⁹ Oxnam AD. Checklists for review articles. *Br Med J* 1994; 309: 648-651.
- ²⁰ Julian D. Trials and tribulations. *Cardiovasc Res* 1994; 28: 598-603.
- ²¹ Charlton BG. Mega-trials: methodological issues and clinical implications. *J R Coll Phys Lond* 1995; 29: 96-100.
- ²² Horvitz RH. A clinician's perspective on meta-analysis. *J Clin Epidemiol* 1995; 48: 41-44.
- ²³ Peto R, Collins R, Gray R. Large scale randomized evidence: large, simple trials and overviews of trials. *J Clin Epidemiol* 1995; 48: 23-40.
- ²⁴ Yusuf S, Collins R, Peto R. Why do we need some large, simple randomized trials? *Stat Med* 1984; 3: 409-420.
- ²⁵ Charlton BG. Randomized trials: the worst kind of epidemiology? *Nature Med* 1995; 1: 1101-1102.
- ²⁶ West RR. A look at the statistical overview (or meta-analysis). *J R Coll Phys Lond* 1993; 27: 111-115.