Value and limitation of meta-analysis

M. Verstraete, MD, PhD

Center for Molecular and Vascular Biology, University of Leuven, Campus Gasthuisberg, O & N, Herestraat 49, B-3000 Leuven, Belgium

Abstract

Meta-analysis is a valuable method of aggregating data sets of different trials which are inadequate or unconvincing on their own. There are strict rules on the retrieval and selection of data which are crucial in terms of the validity of a meta-analysis. Many so-called meta-analyses are flawed because of lack of a protocol or non adherence to it. Also underreporting of clinical trials is haunting the validity of meta-analysis. This hidden information is not a random process but is affecting mainly trials with negative results. To avoid this publication bias all clinical trials should enter an international online database.

Copyright © 2002 S. Karger AG, Basel

Meta-analysis can be defined as the combination of results from several randomized controlled trials of similar design to produce a single estimate of the effect of a treatment which is more precise than the estimate of outcome of single trials. This systematic review tends to overcome the subjective and often biased traditional narrative reviews which most often do not specify the source of information and fails to perform a standardized assessment of the methodological quality of studies (1, 2).

Meta-analysis have received a mixed reception from the outset as meta-analyses of small trials were later contradicted

by a single large randomized trial on the same topic. Well known examples are the effects of nitrates (3, 4) and magnesium (5, 6) on mortality in acute myocardial infarction, comprehensive in-hospital assessment on mortality in the elderly (7, 8) and the effect of aspirin on the risk of pre-eclampsia (9, 10).

Obviously the quality of component trials is of crucial importance: if the raw material is flawed, so will the resulting meta-analysis ("garbage in, garbage out"). Several experts reviewed the numerous biases that threaten the validity of meta-analysis (11, 12). These may relate to differences among component trials in patients' characteristics (selection bias), provided care (performance bias), assessment of outcome (detection bias) and exclusion of randomized patients (attrition bias).

Also underreporting of trials is haunting the validity of narrative reviews and meta-analysis. It appears that only about half of abstracts presented at conferences are later published in full (13). Based on research proposals approved by ethics committees of four leading medical schools, only 49 to 67% are fully published in medical journals (14, 15). Of trials funded by the National Institutes of Health 20% are still unpublished several years after completion (14, 16). This hidden information is unfortunately not a random process but is affecting mainly trials with negative results. A review of published papers could thus identify a spurious beneficial treatment or miss an important adverse effect. To avoid the publication bias (unpublished or duplicate publication) it has been proposed to enter all trials in an (inter)national online data base (17). In

addition to the Cochrane controlled register there is Pub Med Central, a free electronic archive of biomedical research launched by the US National Institutes of Health (18).

Meta-analysis is much more than pooling of data of individual trials

A meta-analysis will not provide the same quality of information about the efficacy of an intervention as a single large randomized trial if the standards are less stringent.

Current standards emphasize the importance of several qualitative features such as the development of a protocol before starting the meta-analysis, inclusion of only truly randomized studies, and the collection of complete outcome information of all randomized patients (11, 19, 20). Many so-called meta-analyses are flawed because of lack of a protocol or non adherence to it. Another important shortcoming is that meta-analyses are done when the outcome results of individual trials are known. To circumvent this problem one should, in selecting randomized trials for uptake in a meta-analysis, study each trial with concealed outcome events.

Thus, the retrieval and selection of data are crucial matters in terms of the validity of a meta analysis. The criteria set forward by Chalmers et al (19) should be strictly adhered to:

1. Proper double-blind randomization. Central procedures are important whatever the scheme, because they minimize the possibility of biased treatment allocation.
2. Complete availability of data in line with “intention to treat” principles.
3. Full consideration of any patients “lost to follow-up”, and any other missing information (e.g.: percentage of missed visits, number and reasons for withdrawals from the treatment, etc.).
4. Procedures for outcome validation.

No rules to calculate the sample size of a given meta-analysis

One of the major roles of meta-analysis is to provide reliable estimates of typical treatment effects when randomized clinical trials themselves are not of sufficient size. However, how shall one define when a meta-analysis is itself of sufficient size? How should the statistical reliability of the evidence within a meta-analysis be assessed? It is reasonable to assume that the sample size should be at least as large as that of the single well-designed and optimally powered randomized controlled trial. Considering the heterogeneity in various features of study design and other possible biases in individual studies included in a meta-analysis, a larger total number of patients may be needed compared to a single randomized trial. Pogue and Yusuf (20) proposed a formula for an optimum

information size (OIS) which provides a first approximation of the minimum sample size required.

In practice, calculation for a power analysis is rarely done prospectively which may explain why the results of meta-analyses and corresponding large trials do not necessarily agree (21).

The risk of meta-analysis based on small trials

Small trials with nominally significant p values tend to overestimate the size of treatment effect, since the treatment effect must be large if statistical significance is to be reached with small sample size. This bias is amplified by the fact that negative trials are often not published. Thus a meta-analysis including mainly small trials is more likely to overestimate treatment effects. Larger trials, even if the results are negative, are known to a broad group of international investigators and are more likely to be published. Meta-analyses of larger trials should thus be less susceptible to this publication bias (20).

There is another reason why even large differences in outcome, based on a meta-analysis of small trials, should be interpreted cautiously (20). Sample variability exists even in studies performed in the same way in identical populations causing different treatment effect estimates. The smaller the study, the larger will be the sampling variability which is another argument to concentrate meta-analyses to large trials (15).

Heterogeneity of treatment effect within meta-analyses might be related to the order of publication of individual trials

Rothwell and Robertson (22) studied 26 meta-analyses of 241 trials and found that when the trials within each meta-analysis were ordered according to year of publication, there was a significant variation in the proportion of trials in which treatment was better than control with a significant excess of positive outcome in the earlier trials. This variation is independent of trial size. Early trials overestimated the treatment effect in comparison with subsequent trials in 20 of 26 meta-analyses studied.

If an initial trial is positive, it is likely to be published – and sooner – than if it is negative. However, once a treatment is considered to be effective and established, a negative trial becomes interesting. Thus meta-analyses done early in the evolution of published trials overestimate a positive treatment effect.

Can heterogeneity be avoided in meta-analysis?

Meta-analysis usually try to convince the reader that the data are homogeneous to justify combining them for a focused question. Diversity in clinical trials is unavoidable. Trials may target different populations of patients and even populations defined by the same eligibility criteria change overtime. Indeed, patients enrolled in comparable trials may belong to the same basic population, but even small differences in the criteria for diagnosis, coexisting conditions, severity of disease, and age will produce very different groups of patients. Differences in doses, time to onset, and duration of therapies can also produce substantial disparity among trials that are included in meta-analyses. The choice of concomitant treatments can also affect the results. Summarizing all the information contained in a set of trials into a single odds ratio may greatly oversimplify an extremely complex issue (23).

Can meta-analysis predict the outcome of a single large trial?

Many meta-analyses do provide a correct understanding of a given treatment or procedure. A glaring example is shocking this confidence. A meta-analysis of 1,301 patients in seven randomized trials showed that intravenous magnesium therapy reduced serious arrhythmias and death after myocardial infarction (5). A subsequent study of 2,300 patients confirmed this result (24). However, the ISIS-4 study (6) in 58,000 patients showed a slight excess of deaths in the magnesium group.

These conflicting results were tentatively explained by a publication bias of never published trials (25).

A Canadian group studied the results of 12 large (> 1,000 patients) randomized, controlled trials with the results of 19 meta-analyses published earlier on similar topics (23). The outcomes of 12 large trials were not predicted accurately 35% of the time by corresponding meta-analysis. If there had been no subsequent randomized trial, the meta-analyses would have lead to the adoption of an ineffective treatment in 32% of cases and to rejection of a useful treatment in 33% of cases.

Argentinian investigators evaluated the ability of meta-analysis to predict the results of a single large trial (> 1,000 patients). There were 30 meta-analyses covering 185 randomized controlled trials. The largest of these trials was then removed and compared to the recalculated relative risk of the meta-analyses (26). In 24 of 30 comparisons the meta-analysis had good predictive ability of the direction of the treatment effect of the largest trial. However, in the other six (20%) the meta-analysis failed to do this, which is casting doubt on the usefulness of this approach other than as a hypothesis – generating tool. The meta-analysis tended to demonstrate stronger protective effects than the largest trial did.

Cumulative meta-analysis

Each time a relevant trial is reported, evidence on the effects of a given intervention accumulates. In a cumulative meta-analysis a previous meta-analysis is repeated each time the results of a new randomized controlled trial is published (27, 28). The obvious goal of this process is to identify the benefit of an intervention as early as possible. This updating should not be continuous but rather periodic – for example when the increment in new information is at least 20 percent of the projected optimum information size (OIS).

A cumulative meta-analysis of controlled trials of beta-blockers in secondary prevention of myocardial infarction was done (Figure 1) (29). Combining the results of 13 trials published by the end of 1981, the relative risk of mortality in patients treated with beta-blocker versus placebo is 0.78 (95 CI 0.69-0.88, $p < 0.001$). Subsequent trials in a further 15,000 patients confirmed this result and could be considered superfluous (30). However, there are no guidelines on the interpretation and reliability of such repeated analyses but the principles of the Optimum Information Size (OIS) and application of formal monitoring have been proposed (20).

Conclusion

Assuming that there are no problems of internal or external consistency and that the pooled estimate is clinically meaningful, then:

1. If a meta-analysis of several trials of which one or more reach statistical significance leads to positive findings, the meta-analysis would strengthen the overall evidence.
2. If meta-analysis of several trials of which none reach sta-

Fig. 1. Cumulative meta-analysis of controlled trials of beta-blockers in secondary prevention after myocardial infarction. A clear ($p < 0.001$) reduction of mortality was evident by 1981 reference 29).

tistical significance leads to a positive result, it would be unadvisable to recommend a therapy on the sole basis of the meta-analysis. Larger clinical trials would probably be required.

3. If meta-analysis of several trials of which one or more reach statistical significance leads to a significant overall

result, selection criteria for the trials can be reconsidered and/or further clinical trials be undertaken.

Thus the positive results of a meta-analysis alone do not provide an absolute recommendation for a treatment, other evidence is required.

References

- Teagarden JR. Meta-analysis: whether narrative review? Review. *Pharmacotherapy* 1989; 9: 274-281.
- Mulrow CD. The medical review article: state of the science. *Ann Intern Med* 1987; 106: 485-488.
- Yusuf S, Collins R, MacMahon S, Peto R. Effect of intravenous nitrates on mortality in acute myocardial infarction: an overview of the randomized trials. *Lancet* 1988; I: 1088-1092.
- Gruppo Italiano per lo Studio della Streptochinasi nell'Infarto Miocardico (GISSI). GISSI-3: effects of lisinopril and transdermal glyceryl trinitrate singly and together on 6-week mortality and ventricular function after acute myocardial infarction. *Lancet* 1994; 343: 1115-1122.
- Teo KK, Yusuf S, Collins R, et al. Effects of intravenous magnesium in suspected acute myocardial infarction: overview of randomized trials. *Br Med J* 1991; 303: 1499-503.
- ISIS-4. Collaboration Group ISIS-4: a randomized factorial trial assessing early oral captopril, oral mononitrate, and intravenous magnesium in 58,050 patients with suspected acute myocardial infarction. *Lancet* 1995; 345: 669-685.
- Stuck AE, Siu AL, Wieland GD, Adams J, Rubenstein IZ. Comprehensive geriatric assessment: a meta-analysis of controlled trials. *Lancet* 1993; 342: 1032-1036.
- Reuben DB, Borok GM, Wolde-Tsadik G, Ershoff DH, et al. Randomized trial of comprehensive geriatric assessment in the care of hospitalized patients. *N Engl J Med* 1995; 332: 1345-1350.
- Imperiale TE, Petrullis AS. A meta-analysis of low-dose aspirin for the prevention of pregnancy-induced hypertensive disease. *JAMA* 1991; 266: 261-265.
- CLASP Collaborative Group. CLASP: a randomized trial of low-dose aspirin for the prevention and treatment of pre-eclampsia among 9364 pregnant women. *Lancet* 1994; 343: 619-629.
- Boissel JP, Blanchard J, Panak E, Peyrieux JC, Saks H. Considerations for the meta-analysis of randomized clinical trials. Summary of a panel discussion. *Contr Clin Trials* 1989; 10: 254-281.
- Jüni P, Altman DG, Egger RM. Assessing the quality of controlled clinical trials. In Egger M, Smith DG, Altman DG (eds). *Systematic reviews in health care: analysis in context*. London: BMJ Books 2001, 87-108.
- Scherer RW, Dickersin K, Langenberg P. Full publications of results initially presented in abstracts. A meta-analysis. *JAMA* 272, 158-162, 1994.
- Dickersin K. How important is publication bias? A synthesis of available data. *AIDS Educ Prev* 1997; 9: 15-21.
- Egger M, Smith GD, Sterne JAC. Uses and abuses of meta-analysis. *Clin Med JRCPL* 2001; 1: 478-484.
- Ioannidis JPA. Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. *JAMA* 1998; 279: 281-286.
- Horton R, Smith R. Time to register randomized trials. *Lancet* 1999; 354: 1130-1139.
- Chalmers I, Altman D. How can medical journals help prevent poor medical research? Some opportunities presented by electronic publishing. *Lancet* 1999; 353: 490-493.
- Chalmers TC, Levin H, Sacks HS, et al. Meta-analysis of clinical trials as a scientific discipline. I: control of bias and comparison with large co-operative trials. *Stat Med* 1987; 6: 315-325.
- Pogue J, Yusuf S. Overcoming the limitations of current meta-analysis of randomized controlled trials. *Lancet* 1998; 351: 47-52.
- Victor N. The challenge of meta-analysis: indications and contra-indications for meta-analysis. *J Clin Epidemiol* 1995; 48: 133-146.
- Rothwell PM, Robertson G. Meta-analysis of randomized controlled trials. *Lancet* 1997; 350: 1181-1182.
- Le Lorier J, Grégoire G, Benhaddad A, Lapierre J, Derderian F. Discrepancy between meta-analyses and subsequent large randomized, controlled trials. *NEJ M* 1997; 337: 536-541.
- Woods KL, Fletcher S, Roffe et al. Intravenous magnesium sulphate in suspected myocardial infarction: results of the Second Leicester Intravenous Magnesium Intervention Trial (LIMIT-2). *Lancet* 1992; 239: 15553-15558.
- Parmley WW. Publication bias. *JACC* 1994; 24: 1424-430.
- Villar J, Carroli G, Belizan JM. Predictive ability of meta-analyses of randomized controlled trials. *Lancet* 1995; 345: 772-776.
- 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: 248-254.
- Chalmers I. The Cochrane collaboration: preparing, maintaining, and disseminating systematic reviews of the effects of health care. *Ann NY Acad Sci* 1993; 703: 156-165.
- Freemantle N, Cleland J, Young P, Mason J, Harrison J. Beta-blockade after myocardial infarction: systemic review and metaregression analysis. *Bri Med J* 1999; 318: 1730-1737.
- Murphy DJ, Povar GJ, Pawlson LG. Setting limits in clinical medicine. *Arch Intern Med* 1994; 154: 505-512.