![thebmj logo]

Editorials

Meta-analysis and the meta-epidemiology of clinical research

Meta-analysis is an important contribution to research and practice but it's not a panacea

C David Naylor, Chief executive officer[a]

[a] *Institute for Clinical Evaluative Sciences, North York, Ontario M4N 3M5, Canada*

This week's *BMJ* contains a pot-pourri of materials that deal with the research methodology of meta-analysis. Meta-analysis in clinical research is based on simple principles: systematically searching out, and, when possible, quantitatively combining the results of all studies that have addressed a similar research question. Given the information explosion in clinical research, the logic of basing research reviews on systematic searching and careful quantitative compilation of study results is incontrovertible. However, one aspect of meta-analysis as applied to randomised trials has always been controversial[1] [2] –combining data from multiple studies into single estimates of treatment effect.

In theory, aggregation of data from multiple trials should enhance the precision and accuracy of any pooled result. But combining data requires a leap of faith: it presumes that the differences among studies are primarily due to chance. In fact, differences in the direction or size of treatment effects may be caused by other factors, including subtle differences in treatments, populations, outcome measures, study design, and study quality.[3] Thus meta-analyses may generate misleading results by ignoring meaningful heterogeneity among studies, entrenching the biases in individual studies, and introducing further biases through the process of finding studies and selecting results to be pooled.

Our understanding of these limits of meta-analysis has arisen partly because a generation of investigators has stepped back from the unthinking pooling of data and begun researching clinical research itself. Those interested in the science of systematic reviews focus on trials as the unit of analysis; and along the way they have usefully shifted the goalposts for reporting on clinical research.

Publication bias

Among the surprising challenges in any systematic review is finding all the studies that have addressed the question(s) of interest. Many studies have documented publication bias favouring clinical trials that show a significant treatment effect. Stern and Simes extend these findings in their "cohort study" of a range of experimental and observational protocols submitted to a research ethics committee at an Australian teaching

hospital (p [640]).[4] Studies with statistically significant outcomes were more likely to be published than non-significant studies, including a threefold difference for randomised trials. They also showed that, even after adjustment for other factors that influenced publication, the negative studies took significantly longer to appear in print.

If trials with positive results are published more often and faster any meta-analysis based only on published trials will inevitably generate an inflated and unduly precise estimate of a given treatment's effectiveness. As Stern and Simes argue, the most practical solution is mandatory registration of all randomised trials at the time of ethics review or other regulatory approval.[4] This policy assures patients who agree to be randomised that their contribution to the betterment of medical care will not be lost.

What is a negative trial?

A step along the path to registration is the "medical editors trial amnesty" that also appears in this week's *BMJ* (p [622]).[5] Over 100 medical journals world wide are inviting readers to submit information on unpublished trials, including those published only as abstracts. Will this do the trick? I suspect not. The journal editors are offering registration, not publication, and the pay off from registration is obscure.

What is missing, moreover, is a clear definition of a negative trial. A negative trial is best defined as one in which a clinically significant effect on predefined end points was ruled out. This requires post hoc examination of the confidence intervals around the treatment effect size estimate in the trial. Editors could help their cause by reminding authors that they welcome submission of such negative studies for possible publication.

In contrast, an inconclusive trial is one in which uncertainty remains about the treatment's effectiveness owing to wide confidence intervals around the point estimate of the treatment effect size. Such inconclusive studies are most at risk of homelessness. Perhaps journal editors should annually invite researchers to submit these inconclusive trials for publication in a special electronic supplement. If, after peer review, the reason for an inconclusive result is indeed lack of statistical power rather than some other flaw, the authors could at least glean some publication credit for their troubles.

As meta-analysts seek unpublished trials and unpublished data from published trials they are often led into conversations with trialists. Such transactions are colourfully described by Roberts and Schierhout in what may be seen as qualitative research to complement the new meta-epidemiology of randomised trials (p [686]).[6] The reluctance of many investigators to provide even aggregate unpublished data makes it more remarkable that some meta-analysts have regularly succeeded in gathering individual patient data for re-analysis from trialists. Methodologists continue to debate the importance of gathering individual patient data for meta-analysis, but it does have advantages. Firstly, if errors in the results as published arise from basic programming or statistical mistakes, these can be rectified. Secondly, there can be greater standardisation, for example, in patient subgroups, follow up times, or use of an intention to treat analysis. Dilemmas over data access for meta-analysis emphasise the need for the research community to debate the conditions under which data from randomised trials should be shared.

Data excess

At times the problem for meta-analysts may not be data access but data excess. Huston and Moher have noted that a single trial of risperidone for chronic schizophrenia was reported in seven different publications with different authorship.[7] Tramèr et al provide a striking example of how duplicate data can affect a meta-analysis in this week's issue (p [635]).[8] In a systematic review of the effects of ondansetron on postoperative emesis they found that data from nine trials appeared in 23 separate publications, including four pairs of almost identical reports with completely different authors. Only one paper openly acknowledged the prior publication of the same data. The greatest duplication occurred in placebo controlled trials of a single 4 mg

intravenous dose of prophylactic ondansetron. When the overlapping publications were weeded out 6.4 patients (95% confidence interval 5.3 to 7.9) had to be treated for every episode of postoperative emesis avoided. When they were not weeded out, the number needed to treat fell to 4.9 (4.4 to 5.6). This is the flip side of publication bias. Just as negative trials are less likely to be published, so positive trials are more likely to be published more than once. The consequences for meta-analysis are similar in both cases: excessively precise and inflated effect size estimates. But, on the positive side, it is the science of systematic reviews that has highlighted this phenomenon of covert duplicate publication.

Given these potential biases, the question remains: how often does meta-analysis mislead rather than guide therapeutic decision making? What can be done to detect misleading meta-analyses? *BMJ* readers will find this issue illuminating, but perhaps not reassuring.

For example, more and more meta-analyses with conflicting conclusions are dotting the literature. Petticrew and Kennedy invoke Sherlock Holmes to make sense of over 20 systematic reviews that have addressed surgical thromboprophylaxis, many with apparently disparate results (p 665).[9] Holmes's bottom line is that surgeons should use mechanical methods rather than heparins, aspirin, or warfarin. Unfortunately, the process whereby the great detective reaches this conclusion is not particularly transparent.

The correspondence columns this week will also reinforce readers' wariness of meta-analysis, as six letters[10] criticise the results of a meta-analysis that purported to show an absence of cardioprotective effect from hormone replacement therapy in postmenopausal and perimenopausal women (p 676).[11] For one, I shall continue to tell my patients that hormone replacement therapy is likely to help prevent coronary disease.

So, how often are meta-analyses wrong? Villar et al examined 30 meta-analyses in perinatal medicine, comparing the results of a meta-analysis of several small trials with a single large trial addressing the same topic.[12] Directionally, 80% of meta-analyses agreed with the results from the larger trial, although concordance for statistically significant findings was much less. Cappelleri et al reviewed 79 meta-analyses and also found about 80% directional agreement.[13]

Very recently LeLorier et al arrived at a more pessimistic assessment.[14] Comparing 12 definitive randomised trials to 19 previous meta-analyses, they claimed the meta-analyses would have led to the adoption of an ineffective treatment in 32% of cases and rejection of a useful treatment in 33%. However, their definition of positive and negative trials was simplistically based on the presence or absence of a statistically significant treatment effect. Directional congruence of point estimates of effectiveness occurred for 80% of the outcomes assessed in the trials and meta-analyses—a result similar to those of the previous studies. The credibility of this work is also undermined by oversights. The authors cite apparent discordance between the 1993 results of the EMERAS trial [15] and a 1985 meta-analysis of thrombolysis for acute myocardial infarction.[16] But they ignore both the findings of ISIS-2,[17] which constituted a more definitive test of the hypotheses generated by the 1985 meta-analysis, and a 1994 meta-analysis that used individual patient data from all trials of thrombolysis for acute myocardial infarction that randomised more than 1000 patients.[18] Conversely, they find concordance between the results of the LIMIT-2 trial[19] and an overview of magnesium for acute myocardial infarction by Teo et al,[20] overlooking the results of ISIS-4[21] and the controversy about magnesium and meta-analysis that has followed.[22 23 24]

A magic method?

Such discrepancies nevertheless lead one to ask: is there a magic method of determining when a meta-analysis is likely to be misleading? The short answer is no. But in this issue Egger et al do describe a graphical method that may help (p 629).[25] Funnel plots show sample sizes against the point estimate of treatment effectiveness generated in individual studies. A symmetrical funnel shaped plot is expected because of greater

scatter in treatment effect estimates for smaller trials, with convergence among larger trials. Egger et al argue that asymmetry in the funnel plot suggests bias in a meta-analysis and propose a statistical method to measure the degree of asymmetry. In reviewing 75 meta-analyses from leading journals and the *Cochrane Database of Systematic Reviews*, they found 19 reviews with significant funnel plot asymmetry.

This ingenious approach has limitations. For validation the authors show funnel plot asymmetry in three of four cases where meta-analyses of multiple small trials disagreed with subsequent large trials but not in four other cases where the meta-analysis and trials were concordant. That is not a statistically convincing number of test cases. Simulated data with computer intensive methods may provide a complementary approach to test this concept. Secondly, the unit of analysis is the randomised trial, not its patients; and the method's power is limited when only a few trials are included. It is probably prudent to pay more attention to the shape of the plot than to any statistical measures of asymmetry. Above all, even dramatic funnel plot asymmetry does not tell readers what type of bias (if any) is occurring. It must therefore be viewed as a non-specific and partially validated screening test for bias in meta-analysis.

In sum, meta-analysis has made and continues to make major contributions to medical research, clinical decision making, and standards of research reportage. However, it is no panacea. Readers need to examine any meta-analyses critically to see whether researchers have overlooked important sources of clinical heterogeneity among the included trials. They should demand evidence that the authors undertook a comprehensive search, avoiding covert duplicate data and unearthing unpublished trials and data. Lastly, readers and researchers alike need to appreciate that not every systematic review should lead to an actual meta-analysis of data with aggregate effect size estimates.[25] If the process of pooling data inadvertently drowns clinically important evidence from individual studies, then a meta-analysis can do more harm than good.

References

1.Chalmers TC, Mattra RJ, Smith H, Kunzler AM.Evidence favouring the use of anticoagulants in the hospital phase of acute myocardial infarction.*N Engl J Med* 1977;**297**: 1091–6.

2.Goldman L, Feinstein AR.Anticoagulants and myocardial infarction: the problems of pooling, drowning, and floating.*Ann Intern Med* 1979;**90**: 92–4.

3.Naylor CD.Two cheers for meta-analysis: problems and opportunities in aggregating results of clinical trials.*Can Med Assoc J* 1988;**138**: 891–5.

4.Stern JM, Simes RJ.Publication bias: evidence of delayed publication in a cohort study of clinical research projects.*BMJ* 1997;**315**: 640–5.

5.Smith R, Roberts I.An amnesty for unpublished trials.*BMJ* 1997;**315**: 622.

6.Roberts I, Schierhout G.The private life of systematic reviews.*BMJ* 1997;**315**: 686–7.

7.Huston P, Moher D.Redundancy, disaggregation, and the integrity of medical research.*Lancet* 1996;**347**: 1024–6.

8.Tramèr MR, Reynolds DJM, Moor RA, McQuay HJ.Impact of covert duplicate publication on meta-analysis: a case study.*BMJ* 1997;**315**: 635–40.

9.Petticrew M, Kennedy SC.Detecting the effects of thromboprophylaxis: the case of the rogue reviews.*BMJ* 1997;**315**: 665–7.

10.Singleton S, Bailey K, Shah S, Rhodes L, Seagroatt V, Sundkvist T, Al-Azzawi F, et al., Col NF, et al., Hemminki E [letters].Impact of postmenopausal hormone therapy on cardiovascular events and cancer.*BMJ* 1997;**315**: 676–9.

11.Hemminki E, McPherson K.Impact of postmenopausal hormonal therapy on cardiovascular events and cancer: pooled data from clinical trials.*BMJ* 1997;**315**: 149–53.

12.Villar J, Carroli G, Belizan JM.Predictive ability of meta-analyses of randomised controlled trials.*Lancet* 1995;**345**: 772–6.

13.Cappelleri JC, Ioannidis JPA, Schmid CH, de Ferranti SD, Aubert M, Chalmers TC, et al. Large trials versus meta-analysis of smaller trials.*How do their results compare? JAMA* 1996;**276**: 1332–8.

14.LeLorier J, Gregorie G, Benhaddad A, Lapierre J, Derderian F.Discrepancies betweeen meta-analyses and subsequent large randomised controlled trials.*N Engl J Med* 1997;**337**: 536–42.

15.EMERAS (Estudio Multicentro Estreptoquinasa Republicas de America del Sur) Collaborative Group.Randomised trial of late thrombolysis in patients with suspected acute myocardial infarction.*Lancet* 1993;**342**: 767–72.

16.Yusuf S, Collins R, Peto R, Furberg C, Stampfer MJ, Goldhaber SZ, 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.

17.ISIS-2 (Second International Study of Infarct Survival) 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.

18.FTT Collaborative Group.Indications for fibrinolytic therapy in suspected acute myocardial infarction: collaborative overview of early mortality and major morbidity results from all randomised trials of more than 1000 patients.*Lancet* 1994;**343**: 311–22.

19.Woods KL, Fletcher S, Roffe C, Haider Y.Intravenous magnesium sulphate in suspected acute myocardial infarction: results of the second Leicester Intravenous Magnesium Intervention Trial (LIMIT-2).*Lancet* 1992;**339**: 1553–8.

20.Teo KK, Yusuf S, Collins R, Held PH, Peto R.Effects of intravenous magnesium in suspected acute myocardial infarction: overview of randomised trials.*BMJ* 1991;**303**: 1499–1503.

21.ISIS-4 (Fourth International Study of Infarct Survival) Collaborative Group.ISIS-4: a randomised factorial trial assessing early oral captopril, oral mononitrate, and intravenous magnesium sulphate in 58 050 patients with suspected acute myocardial infarction.*Lancet* 1995;**345**: 669–85.

22.Yusuf S, Flather M.Magnesium in cute myocardial infarction. ISIS-4 provides no grounds for its routine use.*BMJ* 1995;**310**: 751–2.

23.Egger M, Davey Smith G.Misleading meta-analysis. Lessons from "an effective, safe, simple" intervention that wasn't.*BMJ* 1995;**310**: 752–4.

24.Collins R, Peto R.Magnesium in acute myocardial infarction.*Lancet* 1997;**349**: 282–3.

25.Egger M, Davey Smith G, Schneider M, Minder C.Bias in meta-analysis detected by a simple, graphical test.*BMJ* 1997;**315**: 629–34.

26.Naylor CD.The case for failed meta-analyses.*J Eval Clin Pract* 1995;**1**: 127–30.