Eur Arch Psychiatry Clin Neurosci (2005) 255 : 369–370 DOI 10.1007/s00406-005-0607-5

EDITORIAL

Wolfgang Maier · Hans-Jürgen Möller

Metaanalyses – highest level of empirical evidence?

Metaanalyses have received a crucial position in evidence-based medicine (EBM). Metaanalyses promise to resolve inconsistencies between studies devoted to the same topic but with different outcomes.

Metaanalyses combine the results of all empirical studies exploring the same issue (hypotheses) in order to obtain an estimate of a global effect. For this purpose inquiries meeting a minimum of methodological standards are systematically collected; the underlying statistical method has been refined. It is now possible to identify in a global manner serious inconsistencies and biases: a putative sampling bias of studies can be detected, and it is possible to explore the between-trial heterogeneity. The metaanalysis technique, therefore, has received a high degree of acceptance and is now widely used to compile controversial results into a single indicator. The Cochrane Collaboration considers the results of metaanalyses as the highest level of evidence. However, this enthusiasm is limited. Other EBM committees allocate more weight to the outcome and design of the single studies by requiring a certain number of well conducted trials in favour of the hypotheses each of which has to be appropriate for this specific purpose; those guidelines require minimum numbers of appropriate trials in support of the recommended advice (critical review procedure). The national and cross-national drug admission authorities are also reluctant to base their decisions on metaanalyses, but prefer to consider the outcome of each single study together with its weakness and advantages separately.

This controversial, unresolved constellation divides the field of guidelines for clinical practice; there is a split between proponents of metaanalyses and proponents of systematic reviews of available studies. Although we acknowledge the persuasiveness of the advanced metaanalytic technique we argue for caution in interpreting results of metaanalyses. Here we give arguments that metaanalyses cannot provide the *highest* level of evidence:

1. Different methodological conditions can produce qualitatively different results. Metaanalyses combining differential methodological formats merge these differential effects.

   A paper by Jüni et al. (2001) illustrates these arguments. Four empirical studies relating key aspects of methodological quality of controlled studies to their effect estimates were submitted to a combined analysis. The results varied qualitatively with generation of allocation sequences, concealment of allocation as well as presence or absence of double blinding.
2. Metaanalyses consider variation across studies which are combined into a metaanalysis as random variation. However, variation across studies might reflect information on differential efficacy in different settings; therefore, variation might be informative and not just random noise.

   One example is the metaanalysis comparing the antidepressive effect of SSRIs and TCAs by Anderson (1998). Despite a broad variation of effects across various studies they concluded a weak effect in favour of TCAs. A reanalysis of the material two years later by the same group reported that this effect is exclusively due to inpatient studies whereas studies in less severe depression in other treatment settings produced conflicting results (Anderson 2000).

   Taking 1 and 2 together, variation of results and effect sizes across studies might be informative and not just random noise. Thus, metaanalyses by themselves cannot guarantee evidence. It is decisive to stratify studies by informative criteria which might guide clinical decisions.

Prof. Dr. med. Wolfgang Maier
Direktor der Klinik und Poliklinik für Psychiatrie
und Psychotherapie
Universitätsklinikum Bonn
Sigmund-Freud-Str. 25
53105 Bonn, Germany
Tel.: +49-228/287-6134

Prof. Dr. Hans-Jürgen Möller
Psychiatrische Klinik der Ludwig-Maximilians-Universität München
Nussbaumstrasse 7
80336 München, Germany
Tel.: +49-89/51605501
Fax: +49-89/51605522
E-Mail: christine.hauer@med.uni-muenchen.de

3. The results of metaanalyses are strongly dependent on the mode of how to identify the included empirical studies.

   For example, the calculated global effect (quantitative differences "drug minus placebo") for five randomly selected substances differed considerably by magnitude between published studies (detected by Medline) and submitted studies: the difference was even significant for one among these five drugs (Melander et al. 2003). Thus, it is apparently decisive to consider all available, methodologically sufficient trials (including those which were not published).

   But even if the scope of studies being considered is extended beyond the Medline search with the ambition to be as comprehensive as possible, equivocal results might still be missing. Two metaanalyses on the effect of SSRIs on suicide attempts published simultaneously in the same journal produced qualitatively different conclusions: Gunnell et al. (2005) identified all placebo-controlled informative studies in Medline and the Cochrane register, whereas Fergusson et al. (2005) identified the identical kind of studies in the "Medicines and Health Care Production Regulation Agency". Gunnell et al. found no significant difference for frequency of suicide attempts between SSRIs and placebo, whereas Fergusson et al. reported a major effect for more suicide attempts under SSRIs by considering a larger number of studies.
4. Conclusions from metaanalyses also depend strongly on the selection of the statistical technique which is used to compile the results of the available trials.

   For example, Leucht et al. (2003) performed a metaanalysis based on studies comparing second versus first generation of antipsychotics. The global heterogeneity test and metaregression on dosage did not reveal inconsistencies. The conclusion was: "Potential advantages in efficacy of the new generation drugs should be a factor in clinical treatment decisions to use these rather than conventional drugs." This statement contradicted the main message extracted from a previous metaanalysis by Geddes et al. (2000) on the same issue. Therefore, Geddes et al. (2003) reanalysed the study material of Leucht et al. again also by performing a metaanalysis. Surprisingly, they did not observe an advantage for atypicals and could, therefore, not support Leucht's et al. conclusions. The critical difference between both conflicting metaanalyses was the selection of different statistics: risk-difference quotient versus log-odds ratio. This technical difference between both analyses of the same material produced qualitative differences.

Thus, the method of metaanalysis is not as robust as it should be for providing a highest level of evidence. Evidence can only be provided by appropriately designed controlled trials and by replications. However, metaanalyses are not based on a well designed controlled trial but present only a summary of statistics afflicted with limited validity. This conclusion is in agreement with many arguments in the biostatistical literature for only limited conceptual validity of the method of metaanalysis (e. g. Feinstein 1995).

What then might be the status of metaanalyses in evidence-based medicine? Metaanalyses are clearly instrumental to explain heterogeneity in treatment effects. They can identify the reasons of variability in outcome measures between different studies. For this purpose study-level variables can be entered in metaregression models; clinically relevant determinants for differential outcomes will emerge from the analyses (Sterne et al. 2000). They can be formulated and consequently tested in subsequent, well designed trials. The clinician might draw from those analyses conclusions on differential indication of treatment trials. The reanalyses of the metaanalysis comparing antidepressant effects of SSRIs and TCAs is an illuminating example. Justification of recommendations of clinical guidelines in evidence-based medicine, however, requires the critical appreciation of each relevant study which is informative for the topic under study. Conclusions on efficacy, effectiveness and practical utility are possible without relying on metaanalyses as it is documented by well accepted and approved clinical guidelines using the critical review procedure.

References

1. Anderson IM (1998) SSRIs versus tricyclic antidepressants in depressed inpatients: a meta-analysis of efficacy and tolerability. Depress Anxiety 7(Suppl 1):11–17
2. Anderson IM (2000) Selective serotonin reuptake inhibitors versus tricyclic antidepressants: a meta-analysis of efficacy and tolerability. J Affect Disord 58:19–36
3. Feinstein AR (1995) Meta-analysis: statistical alchemy for the 21st century. J Clin Epidemiol 48:71–79
4. Fergusson D, Doucette S, Glass KC, Shapiro S, Healy D, Hebert P, Hutton B (2005) Association between suicide attempts and selective serotonin reuptake inhibitors: systematic review of randomised controlled trials. BMJ 330:396–402
5. Geddes J, Freemantle N, Harrison P, Bebbington P (2000) Atypical antipsychotics in the treatment of schizophrenia: systematic overview and meta-regression analysis. BMJ 321:1371–1376
6. Geddes JR, Harrison P, Freemantle N (2003) New generation versus conventional antipsychotics. Lancet 362:404
7. Gunnell D, Saperia J, Ashby D (2005) Selective serotonin reuptake inhibitors (SSRIs) and suicide in adults: meta-analysis of drug company data from placebo controlled, randomised controlled trials submitted to the MHRA's safety review. BMJ 330:385–389
8. Jüni P, Altman DG, Egger M (2001) Assessing the quality of controlled clinical trials. BMJ 323:42–46
9. Leucht S, Wahlbeck K, Hamann J, Kissling W (2003) New generation antipsychotics versus low-potency conventional antipsychotics: a systematic review and meta-analysis. Lancet 361:1581–1589
10. Melander H, Ahlqvist-Rastad J, Meijer G, Beermann B (2003) Evidence b(i)ased medicine – selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications. BMJ 326:1171–1173
11. Sterne JA, Gavaghan DJ, Egger M (2000) Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. J Clin Epidemiol 53:1119–1129