

# I Have the Answer, Now What's the Question?: Why Metaanalyses Do Not Provide Definitive Solutions

David L Streiner, PhD<sup>1</sup>

(Can J Psychiatry 2005;50:829–831)

It ain't so much the things we don't know that get us in trouble. It's the things we know that just ain't so.

—Artemus Ward

In a recent issue of the *British Medical Journal*, Moncrieff and Kirsch (1) concluded that “Recent meta-analyses show selective serotonin reuptake inhibitors [SSRIs] have no clinically meaningful advantage over placebo” and that “Methodological artefacts may account for the small degree of superiority shown over placebo.” Needless to say, this article generated a large number of letters to the editor, citing everything from despair about the lack of available alternatives to charges that the authors overlooked or ignored evidence about the positive effects of SSRIs, that they misinterpreted the findings and recommendations of the National Institute for Health and Clinical Excellence (NICE), that both the authors and NICE used inappropriate criteria to evaluate improvement, that the authors made erroneous assumptions about the distribution of depression, and so on.

This editorial does not aim to critique the article by Moncrieff and Kirsch. Rather, it tries to explain why different people with honourable intentions can come to different conclusions regarding metaanalyses. Suffice it to say for now that their summary and recommendations are by no means accepted by all. Other metaanalyses (for example, 2–4), including one coauthored by Moncrieff (5), have supported the use of SSRIs; this article has merely brought the controversy to a head.

The fact is that Moncrieff and Kirsch's conclusions should not come as a surprise. Fourteen years ago, Greenberg and others (6) came to similar conclusions regarding the effectiveness (or rather, the lack thereof) of the older class of tricyclic antidepressants (TCAs). What some may find strange is why this debate (and similar ones regarding the effectiveness of interventions ranging from screening for breast cancer to the use of cholinesterase inhibitors in Alzheimer's disease) is still

going on. After all, weren't we promised that metaanalyses would provide definitive answers to questions such as these? Metaanalysis is predicated on the assumption (or it may be more a belief and hope) that objectivity regarding the criteria used for conducting literature searches, selecting the articles to include or exclude, and abstracting and summarizing the findings would result in unbiased and unequivocal answers. In some hierarchies of evidence, metaanalyses are at the top, trumping even very large randomized controlled trials (7).

Since Smith and Glass's pioneering 1977 metaanalysis of psychotherapy (8), there has been an exponential explosion in the number published in the medical and psychological literature. Doing a simple Medline and PsycLit search, using just the keyword metaanalysis, I found that there were 3 published in 1981, 422 in 1991, and 1712 in 2003. Indeed, there are international organizations, such as the Cochrane Collaboration in medicine and the Campbell Collaboration in the social sciences, devoted exclusively to conducting and publishing metaanalyses. Further, there are regularly published compendia of treatment recommendations based on their results (9).

The fact is, though, that disagreements among metaanalyses of the same topic are quite common. Oxman and Guyatt found that 5 reviews about the need to treat mild hypertension all resulted in different recommendations (10), and Munsinger (11) and Kamin (12), reviewing the same articles about environmental effects on intelligence, came to diametrically opposite conclusions. Indeed, our review of the effectiveness of TCAs (13) disagreed with Greenberg and others' findings (6).

The reality is that, despite the claims of true believers, metaanalysis is neither a purely objective, mechanical process nor a panacea for answering all questions. There are 2 major reasons why metaanalyses may differ with regard to the conclusions they draw: methodological considerations and interpretation.

With respect to the first reason, metaanalysis is a complicated process comprising many different phases (I've actually described a 12-step program for metaanalysis [14], although I haven't heard of it curing anyone), and each step requires

some degree of judgment. Judgment, in turn, implies that equally competent reviewers can make decisions that affect the conclusions that are drawn. Starting at the beginning, the first steps in a metaanalysis consist of posing the question to be addressed and setting the inclusion criteria. While this may appear at first glance to be simple and straightforward, even subtle differences can lead to a search for and retrieval of different articles. For example, asking the question, “Are SSRIs more effective than placebo in treating depression?” raises several issues: Should all types of depression be included or should the analysis be limited to a specific type of depression? Must the depression be diagnosed with the use of a structured interview or is it sufficient to rely on the psychiatrists’ judgment? Should improvement be assessed by self-report or by observer evaluation? This can result in different conclusions (6). Which outcome measures will be accepted as sufficiently valid? Is there a minimum duration for the trial to ensure that the drug has had time to work, and if so, how long is it? Is there a minimum dosage of the medication, and if so, does this apply to the average for the entire group or for each individual patient? Is there a maximum number of people who can drop out of a study, and if so, how many? Will trials be included, irrespective of methodological rigour, or must they exceed some minimum score on a methodology checklist? What evidence needs to be given to ensure that there was blinding during the randomization and assessment phases of the study? How thoroughly should the “grey literature” of unpublished (and hence, unreviewed) studies be searched, especially given the well-known finding that those with negative results are less likely to be submitted or published (15)? During the data abstraction phase, a decision has to be made whether to focus on one outcome measure or to pool the results if several were used in a study. At the point of analysis, the researchers must decide whether to use a fixed-effects model (which assumes that there is one population effect size that each study approximates) or a random-effects model (which allows a range of effect sizes that vary among studies because of sampling, drug, research design features, and the like).

As long as this list appears, it by no means exhausts the questions that must be addressed by the metaanalyst. Each question demands an answer, but there are no correct ones; different people can make different decisions and likely provide equally convincing reasons. Compounding the problem even further, we found that only a small minority of published articles contained all the information needed to summarize the results for inclusion in our metaanalysis (16), and unfortunately, our experience was far from unique (17–21). The solution we took was to fill in the missing information about the outcome measures (for example, standard deviations) from different trials and to extrapolate group means from graphs

when necessary. Others, faced with the same problem, may choose to reject articles that have insufficient information. Again, compelling arguments can be made for both approaches.

The second reason why conclusions of metaanalyses may differ regards the interpretation they place on the results, and this seems to be the heart of the controversy regarding Moncrieff and Kirsch (1). Their findings are consistent with previous metaanalyses and with the NICE report (22), in that there is a statistically significant effect of SSRIs, compared with placebo, with a standardized mean difference (SMD) of 0.34. Where Moncrieff and Kirsch part company from the NICE report regards the clinical importance placed on this SMD. Moncrieff and Kirsch feel that it is a trivial difference, whereas the NICE authors (and others, judging from the letters to the editor) believe it is a clinically important one. This is not an issue that can be resolved through statistical argument or recourse to picking nits about methodology; rather, it rests with the judgment of clinicians.

The unfortunate result of the necessity to make decisions at each stage of the process and when interpreting the findings is that there are, and will always be, differences among metaanalyses of the same topic. Further, the article by Moncrieff and Kirsch (1) and the numerous letters to the editor which followed highlight a “meta” metaanalytic problem—that narrative summaries of previous metaanalyses may themselves be selective and one-sided. The good news is that metaanalysts will never be out of a job. The bad news is that readers cannot assume that any metaanalysis provides the last word regarding the effectiveness of an intervention, and their own judgments will always play a role.

## References

1. Moncrieff J, Kirsch I. Efficacy of antidepressants in adults. *BMJ* 2005;331:155–7.
2. Gijssman HJ, Geddes JR, Rendell JM, Nolen WA, Goodwin GM. Antidepressants for bipolar depression: a systematic review of randomized, controlled trials. *Am J Psychiatry* 2004;161:1537–47.
3. Wilson K, Mottram P, Sivananthan A, Nightingale A. Antidepressants versus placebo for the depressed elderly. *Cochrane Depression, Anxiety and Neurosis Group. Cochrane Database Syst Rev* 2005;3.
4. Gill D, Hatcher S. Antidepressants for depression in medical illness. *Cochrane Depression, Anxiety and Neurosis Group. Cochrane Database Syst Rev* 2005;3.
5. Lima MS, Moncrieff J. Drugs versus placebo for dysthymia. *Cochrane Depression, Anxiety and Neurosis Group. Cochrane Database Syst Rev* 2005;3.
6. Greenberg RP, Bornstein RF, Greenberg MD, Fisher S. A meta-analysis of antidepressant outcome under “blinder” conditions. *J Consult Clin Psychol* 1992;60:664–9.
7. Harbour R, Miller J. A new system for grading recommendations in evidence based guidelines. *BMJ* 2001;323:334–6.
8. Smith ML, Glass GV. Meta-analysis of psychotherapy outcome studies. *Am Psychol* 1977;32:752–60.
9. Clinical evidenc. 2005. Available: [www.clinicalevidence.com/ceweb/conditions/index.jsp](http://www.clinicalevidence.com/ceweb/conditions/index.jsp). Accessed 2005 Oct 18.
10. Oxman AD, Guyatt GH. Guidelines for reading literature reviews. *CMAJ* 1988;138:697–703.
11. Munsinger H. The adopted child’s IQ: a critical review. *Psychol Bull* 1975;82:623–59.
12. Kamin LJ. Comment on Munsinger’s review of adoption studies. *Psychol Bull* 1980;88:359–69.

13. Joffe R, Sokolov S, Streiner D. Antidepressant treatment of depression: a metaanalysis. *Can J Psychiatry* 1996;41:613–6.
  14. Streiner DL. Meta-analysis: a 12-step program. *Electronic J Gambling Issues*. Available: [www.camh.net/egambling/issue9/feature/](http://www.camh.net/egambling/issue9/feature/). Accessed 2005 Oct 7.
  15. Greenwald A. Consequences of prejudices against the null hypothesis. *Psychol Bull* 1975;82:1–20.
  16. Streiner DL, Joffe R. The adequacy of reporting randomized, controlled trials in the evaluation of antidepressants. *Can J Psychiatry* 1998;43:1026–30.
  17. Göttsche PC. Methodology and overt and hidden bias in reports of 196 double-blind trials of nonsteroidal antiinflammatory drugs in rheumatoid arthritis. *Control Clin Trials* 1989;10:31–56.
  18. Fletcher RH, Fletcher SW. Clinical research in general medical journals: a 30-year perspective. *N Engl J Med* 1979;301:180–3.
  19. Altman DG. The scandal of poor medical research. *BMJ* 1994;308:283–4.
  20. Pocock SJ, Hughes MD, Lee RJ. Statistical problems in the reporting of clinical trials: a survey of three medical journals. *N Eng J Med* 1987;317:426–32.
  21. Gardner MJ, Bond J. An exploratory study of statistical assessment of papers published in the British Medical Journal. *JAMA* 1990;263:1355–7.
  22. National Institute for Health and Clinical Excellence. Depression. [www.nice.org.uk/page.aspx?o=235367](http://www.nice.org.uk/page.aspx?o=235367). Accessed 2005 Sept 8.
- 

<sup>1</sup>Director, Kunin-Lunenfeld Applied Research Unit, Baycrest Centre for Geriatric Care, Toronto, Ontario.