
Lessons to be learnt from meta-analyses of newer versus older antidepressants

Ian Anderson

Meta-analysis is the use of statistical techniques to analyse the findings of many individual analyses (Glass, 1977). It covers all aspects of the review process involving formulating relevant research questions, searching the literature, assessing the quality of studies and choosing relevant ones, extracting and combining the data (for review see Henry & Wilson, 1992; Wilson & Henry, 1992). Meta-analysis as part of a systematic review has advantages over a narrative review but there are problems in applying it in practice (Box 1).

The methodologies of meta-analyses themselves are frequently poor and should be examined critically before conclusions are accepted (Box 2). Interpretation requires an understanding of the summary measures used in order to combine studies. For continuous data a standardised difference (or effect size) can be calculated for each study (the difference in scores from the two treatments divided by the standard deviation; Hedges & Olkin, 1985). For events (e.g. response *v.* non-response) the odds ratio (OR; Altman, 1991) or, better, the risk ratio (or relative risk, RR; Sinclair & Bracken, 1994) is usually calculated. The odds is the number of occurrences divided by the number of non-occurrences, and the risk is number of occurrences divided by the total number; the latter is better understood by clinicians. The OR (RR) is the ratio of the odds (risks) in the treatment group compared with controls.

Box 1. Advantages and problems with meta-analysis

Advantages

Systematic review of evidence
Increased power to detect differences
Provides an estimate of the size of effect and confidence interval
Possibility of analysing subgroups not achievable in single study

Problems

Difficulty in identifying all relevant studies
Comparability of studies to be combined (e.g. diagnostic issues)
Quality of studies to be combined (e.g. methodology, completeness of data)
Publication bias
Statistical heterogeneity among studies (some controversy over whether pooling is then appropriate)

Meta-analyses of newer versus older antidepressants

Meta-analysis is particularly relevant in comparative studies of antidepressants in which the power to detect differences between treatments is usually low because of small patient numbers. The variety of antidepressants does, however, restrict the number of studies available for individual drugs. If groupings of antidepressants (e.g. selective serotonin reuptake inhibitors, SSRIs) are considered, then respectable numbers can be achieved but only at the expense of missing possible differences between individual drugs. The present review is restricted to double-blind,

Dr Ian Anderson is Senior Lecturer in Psychiatry at the University of Manchester and Honorary Consultant in General Adult Psychiatry at the Rawnsley Building, Manchester Royal Infirmary, Oxford Road, Manchester M13 9WL. His research interests are in psychopharmacology and the aetiology and treatment of affective disorders.

Box. 2 How to assess the quality of a meta-analysis

The question should be answerable from the studies selected

The search should be explained and comprehensive

Selection criteria for studies should be described and be appropriate to the question

The statistical method should be explained and be appropriate

Pooling should be variance-weighted (larger studies with smaller variance are given more weight)

Results should be expressed with confidence intervals

Heterogeneity of studies should be assessed and taken into account in calculating the results

A sensitivity analysis should be undertaken, if possible, by examining better quality studies separately or by using alternative methods of analysis

Publication bias should be considered, for example with a funnel plot (see Wilson & Henry, 1992)

randomised, comparative studies of antidepressants in depressive illness based on a MEDLINE search supplemented by manual searching of review articles and literature provided by drug companies. With duplicated or updated meta-analyses I considered only the most informative study.

The major groupings of antidepressants are tricyclic antidepressants (TCAs), monoamine oxidase inhibitors (MAOIs), second-generation antidepressants and SSRIs. Some drugs are not easily classified. For example, lofepramine is a TCA that is often grouped with second-generation antidepressants related to issues of tolerability and safety, and so-called 'fourth-generation' antidepressants (e.g. venlafaxine, nefazodone) share properties with earlier antidepressants. In general, meta-analyses have concentrated on comparisons between a newer drug or drug class and first-generation TCAs, either as a class or individually.

Efficacy

The question of efficacy, central to any comparison of antidepressants, is not straightforward because

of variation in the diagnosis, nature and treatment setting of depressive illness and the analysis and definition of response. These issues are beyond the scope of this paper.

Second-generation antidepressants

The efficacy of second-generation antidepressants versus first-generation TCAs is often questioned. Unfortunately, meta-analyses addressing this question have been few and generally poorly conducted. Some studies looking at a range of comparisons have suggested that trazodone and/or mianserin may be less effective than TCAs (Moller & Haug, 1988; Kasper *et al*, 1992) whereas others have not (Davis *et al*, 1993; Workman & Short, 1993). Unfortunately little weight can be given to these analyses because of methodological problems including the lack of confidence intervals.

Patten (1992) analysed comparative trials of trazodone against imipramine and concluded equal odds of responding to treatment (OR 0.93, 95% confidence interval (CI) 0.64–1.36). Strict inclusion criteria meant he could include only six out of 25 identified studies (302 patients) and the pooling was not variance-weighted. The recalculated, variance-weighted RR is 0.86 (CI 0.70–1.05). This indicates that trazodone is unlikely to be better than imipramine but may be significantly worse as the confidence interval encompasses a 5% advantage but a 30% disadvantage. A major problem with that study is potential selection bias, which makes final interpretation difficult.

In a good-quality study Kerihuel & Dreyfus (1991) found patients responded better to lofepramine than to older TCAs and maprotiline (mostly imipramine or amitriptyline; OR 1.4, CI 1.2–1.8). They were able to use 24 out of 29 studies identified and verified the data of 602 of the total of 2040 patients using original case report forms. The recalculated RR is 1.13 (CI 1.04–1.22) suggesting that patients are 13% more likely to respond to lofepramine than to older TCAs, with a confidence interval from 4 to 22%. There are methodological caveats (e.g. pooling is not variance-weighted and selection bias is not investigated) but these data suggest lofepramine is at least as effective as older TCAs, and the possibility that it has advantages is intriguing but difficult to understand. A re-analysis using the individual studies would be valuable.

SSRIs

The comparative efficacy of SSRIs has been extensively investigated. Song *et al* (1993) calculated standardised differences using the final

Hamilton Depression Rating Scale (HAM-D) score, and assigned pooled standard deviations to incomplete studies to yield results from 49 of 63 identified studies. They do not produce a single pooled value but analyse different subsets according to the version of the HAM-D scale used. These, in general, show no significant difference in efficacy (although studies using the standard, 17-item version of the HAM-D favour comparators). A major criticism of this study is the inclusion of second-generation antidepressants. Therefore, we (Anderson & Tomenson, 1994) analysed only studies against noradrenaline reuptake inhibitors (with exclusion of TCA doses below 100 mg to lessen bias towards SSRIs). We calculated a standardised change in HAM-D scores and assigned an estimated standard deviation to incomplete studies. We were able to include 55 studies (4629 patients) with a separate analysis of studies that were placebo-controlled and those with complete data to provide quality control. We checked for selection bias and corrected for heterogeneity. The SSRIs and TCAs were significantly better than placebo (effect sizes for both about 0.4, a difference in HAM-D response of three points). The relative effect size was 0 (CI -0.08–0.08) indicating that any true difference in efficacy is small. In a tentative subgroup analysis (not corrected for multiple comparisons) individual SSRIs had effect sizes close to zero but only fluoxetine and fluvoxamine had confidence intervals that excluded a clinically important difference. In-patients fared less well on SSRIs generally and only for fluvoxamine was there confidence in equal efficacy with TCAs. Of theoretical interest, TCAs with the most potent inhibition of reuptake of serotonin in addition to noradrenaline (clomipramine and amitriptyline) were more effective than SSRIs but the confidence interval ranged from clinically important to near equivalence.

Further meta-analyses for individual SSRIs available using drug company trials databases (Bech & Cialdella, 1992; Pande & Saylor, 1993a) do not add to the conclusions above.

Other antidepressants

Moclobemide is a reversible inhibitor of monoamine oxidase A. A meta-analysis using individual data on 1256 patients on the drug company's database (Angst & Stabl, 1992) unfortunately has poor methodology (no examination of heterogeneity between study sites, no explanation of pooling method, lumping together all comparators and lack of confidence intervals). It is possible to calculate an overall RR of 1.03 (CI 0.95–1.12) suggesting that between 5% fewer and 12% more patients respond on moclobemide compared with comparators. However, a further meta-analysis of the use of moclobemide in 12 studies (730 patients)

against imipramine and clomipramine in in-patients is less favourable (Angst *et al.*, 1995). Individual patient data are used but the Danish University Antidepressant Group (1993) study, which found clomipramine significantly more effective, appears to be excluded on the grounds of a low fixed moclobemide dose (400 mg/day) and one study against imipramine is single-blind. The results are again unsatisfactorily presented but calculation of the RR shows that moclobemide is less effective than both TCAs (combined results 0.88 (CI 0.78–0.99)), although this is not significant taking clomipramine and imipramine individually. Excluding lower-dose moclobemide and clomipramine patients leaves the result as 0.89 (CI 0.78–1.01). Therefore, even patients on effective doses of moclobemide have up to a 22% greater risk of not responding compared with those treated with clomipramine and imipramine, although similar efficacy is not quite excluded.

Finally, a meta-analysis of the serotonin and noradrenaline reuptake inhibitor, venlafaxine, is presented as part of a cost-effectiveness analysis (Einarson *et al.*, 1995). As presented it is unacceptable because it averages response rates for individual drugs in comparative and non-comparative studies. Sound search and inclusion criteria (including drug dose), however, allow a variance-weighted RR of 1.15 (CI 1.00–1.32) to be calculated from six studies (636 patients) of venlafaxine against imipramine, trazodone, maprotiline and fluoxetine, indicating that venlafaxine is unlikely to be less effective than comparators and could be up to 32% more effective. The comparator drugs are heterogeneous and numbers are small but, given the previous finding that dual-action TCAs may be more effective than SSRIs, it deserves further investigation.

Tolerability

Tolerability has been assessed by counting drop-outs in comparative studies (which may underestimate side-effect burden), apart from one meta-analysis which considered the percentage of patients experiencing side-effects (which may overestimate significant side-effects).

Second-generation antidepressants

Kerihuel & Dreyfus (1991) found significantly fewer patients on lofepramine compared with first-generation TCAs experienced at least one side-effect (53 *v.* 64%). The OR was 0.6 (CI 0.5–0.7) and a recalculated RR gives 0.83 (0.77–0.89) indicating a

17% reduction with a range of 11–23%. This did not translate into a significant difference in the total number of patients stopping treatment, although the confidence interval is fairly wide and does not exclude a lower rate on lofepramine (RR 0.93, CI 0.79–1.09).

SSRIs

Song *et al* (1993) challenged the view that SSRIs were better tolerated than older antidepressants in their meta-analysis of 58 studies with 5518 patients. Total drop-outs (OR 0.95, CI 0.82–1.11) and those due to treatment failure were not significantly different, whereas those due to side-effects just missed significance (OR 0.81, CI 0.65–1.00). Criticisms about the inclusion of non-TCA antidepressants resulted in further meta-analyses. We (Anderson & Tomenson, 1995) analysed 64 studies with 6012 patients in which SSRIs were compared only to TCAs and found a 10% decrease in the relative overall risk of dropping out (RR 0.90, CI 0.84–0.97), with the relative risk of dropping out due to side-effects reduced by 25% (RR 0.75, CI 0.66–0.84). These results did not appear to be due to selection bias. Montgomery & Kasper (1995) analysed 67 studies of SSRIs versus TCAs (6852 patients) with the same result for drop-outs due to side-effects; unfortunately, they did not analyse total drop-outs. My conclusion is that fewer people discontinue treatment on SSRIs compared with TCAs and that this can be attributed to less severe side-effects, but the absolute risk reduction is small (between 1 and 5%) and of uncertain practical or economic significance. Although we (Anderson & Tomenson, 1995) did not find statistically significant differences between SSRIs, individual results showed that for fluoxetine and paroxetine, but not fluvoxamine, there was a significant advantage over TCAs for drop-outs due to side-effects and a similar pattern for total drop-outs (unpublished data available from the author). There were too few studies to consider citalopram and sertraline. This may not reflect a real difference but it accords with clinical impressions suggesting that some SSRIs may be better tolerated than others.

Pande & Saylor (1993b) looked at drop-outs in comparative trials with fluoxetine from the drug company database and found similar results.

As a twist in the tail, Donovan (1993) listed seven double-blind trials of dothiepin against SSRIs involving 619 patients. He provided sufficient data to calculate the RRs for drop-outs due to side-effects (0.48, CI 0.26–0.89), showing a large advantage to dothiepin, and for total drop-outs (0.77, CI 0.57–1.04), which also favours dothiepin. These studies

are mostly small with no guarantee that they are without selection bias, but question whether all TCAs should be assumed to be equally likely to produce side-effects.

Suicide risk

An important issue was addressed by Beasley *et al* (1991) following case reports that fluoxetine might cause patients to become suicidal. From drug company data they analysed 12 studies (1450 patients), used sound methodology and reported on the incidence (risk) difference (differences between drugs in absolute percentages). Suicidal acts (0.3%) and the emergence of substantial suicidal ideation (3%) were infrequent and there were no differences between fluoxetine and either placebo or TCAs. The narrow confidence intervals make it unlikely that fluoxetine promotes suicide although they do not rule out that up to 3–4% fewer patients develop substantial suicidal ideation on fluoxetine compared with placebo (incidence difference –1.5, CI –3.3–0.3) or TCAs (–1.8, CI –4.0–0.4). A limitation is the exclusion of patients with an initial substantial suicide risk from most studies.

Lessons to be learnt from these meta-analyses

General points (Box 3) are: first, distrust claims for any antidepressant based on one or two studies (unless it is a 'mega-trial') as selection bias is likely; second, unless there are confidence intervals around any outcome measure the clinical relevance is uncertain.

Meta-analysis has shed little light on the efficacy or tolerability of second-generation antidepressants (Box 4). A meta-analysis of trazodone allows only modest confidence that it is as effective as imipramine. In contrast, lofepramine is at least equally as effective as older TCAs and is less likely to cause side-effects, although it remains uncertain that this results in better overall compliance.

For the SSRIs taken as a group there have been sufficient studies to conclude that they are as effective as older TCAs in treating depression. The question of reduced efficacy in particular subgroups remains open. With regard to tolerability, significantly fewer patients stop treatment because of side-effects with SSRIs than with older TCAs. This probably accounts for fewer overall drop-outs

on SSRIs, but only by three or four patients in 100 treated. Cost-effectiveness analyses must not overestimate the advantage to SSRIs.

Group findings for SSRIs and for TCAs may not hold for individual compounds. Assertions that fluvoxamine is less well tolerated than other SSRIs and dothiepin is tolerated at least as well as SSRIs are not excluded by these meta-analyses. As this accords with clinical impression, further investigation of these issues would be useful.

Fluoxetine and TCAs do not promote suicide in the short term in depressed patients who start with low suicide risk. Exceptions in individual cases are, however, not ruled out.

These meta-analyses are consistent with the interesting possibility that drugs inhibiting both serotonin and noradrenaline reuptake (e.g. clomipramine and venlafaxine) may have superior efficacy to single reuptake inhibitors.

There is less confidence that SSRIs and moclobemide are as effective as the older TCAs in in-patients and TCAs probably remain first-line treatment for these patients.

Conclusions

Meta-analysis is useful and can raise new research questions as well as addressing current ones. It has been applied in only a limited fashion to comparative studies of antidepressants. Proposed short-term advantages to individual drugs (e.g. response of anxiety symptoms to SSRIs, speed of onset of

Box 3. Lessons to be learned from the meta-analyses of newer versus older antidepressants

Distrust conclusions based on non-systematic reviews or single small studies

Clinical relevance can only be assessed if confidence intervals are given

SSRIs and lofepramine are as effective as older TCAs in the short-term treatment of depression

SSRIs and lofepramine have side-effect advantages over older TCAs but overall short-term compliance may only be marginally better

Fluoxetine is no more likely than TCAs or placebo to cause suicidal ideation or actions in short-term use

Box 4. Controversial issues

Efficacy of second-generation antidepressants (except lofepramine) compared with older TCAs

Relative tolerability of individual SSRIs (is fluvoxamine worse?) and older TCAs (is dothiepin better?)

Enhanced efficacy of dual serotonin and noradrenaline reuptake inhibitors (clomipramine, venlafaxine)

Efficacy of newer antidepressants in in-patients

antidepressant response with venlafaxine) and issues related to longer-term treatment such as relapse prevention, tolerability and compliance should be subjected to meta-analysis. To achieve this, registration and comprehensive data recording from randomised controlled studies such as by the Cochrane Collaboration (Godlee, 1994) is to be encouraged.

References

- Altman, D. G. (1991) *Practical Statistics for Medical Research*. London & New York: Chapman and Hall.
- Anderson, I. M. & Tomenson, B. M. (1994) The efficacy of selective serotonin reuptake inhibitors in depression: A meta-analysis of studies against tricyclic antidepressants. *Journal of Psychopharmacology*, **8**, 238–249.
- & — (1995) Treatment discontinuation with selective serotonin reuptake inhibitors compared to tricyclic antidepressants: A meta-analysis. *British Medical Journal*, **310**, 1433–1438.
- Angst, J. & Stabl, M. (1992) Efficacy of moclobemide in different patient groups: A meta-analysis of studies. *Psychopharmacology*, **suppl. 106**, S109–S113.
- , Amrein, R. & Stabl, M. (1995) Moclobemide and tricyclic antidepressants in severe depression: Meta-analysis and prospective studies. *Journal of Clinical Psychopharmacology*, **suppl. 15**, 16S–23S.
- Beasley, C. M., Dornseif, B. E., Bosomworth, J. C., *et al* (1991) Fluoxetine and suicide: A meta-analysis of controlled trials of treatment for depression. *British Medical Journal*, **303**, 685–692.
- Bech, P. & Cialdella, P. (1992) Citalopram in depression – meta-analysis of intended and unintended effects. *International Clinical Psychopharmacology*, **6** (suppl. 5), 45–54.
- Danish University Antidepressant Group (1993) Moclobemide. A reversible MAO-A-inhibitor showing weaker antidepressant effect than clomipramine in a controlled multicentre study. *Journal of Affective Disorders*, **28**, 105–116.
- Davis, J. M., Wang, Z. & Janicak, P. G. (1993) A quantitative analysis of clinical drug trials for the treatment of affective disorders. *Psychopharmacology Bulletin*, **29**, 175–181.
- Donovan, S. (1993) The efficacy and tolerability of dothiepin versus serotonin specific reuptake inhibitors in the treatment of depression. *European Neuropsychopharmacology*, **3**, 331–332.
- Einarson, T. R., Arikian, S., Sweeney, S., *et al* (1995) A model to evaluate the cost-effectiveness of oral therapies in the management of patients with major depression. *Clinical Therapeutics*, **17**, 136–153.

- Glass, G. V. (1977) Integrating findings: The meta-analysis of research. *Review of Research in Education*, 5, 351–379.
- Godlee, F. (1994) The Cochrane Collaboration. *British Medical Journal*, 309, 969–970.
- Hedges, L. V. & Olkin, I. (1985) *Statistical Methods for Meta-Analysis*. Orlando, FL: Academic Press.
- Henry, D. A. & Wilson, A. (1992) Meta-analysis. Part 1: An assessment of its aims, validity and reliability. *Medical Journal of Australia*, 156, 31–38.
- Kasper, S., Fuger, J. & Moller, H.-J. (1992) Comparative efficacy of antidepressants. *Drugs*, 43, 11–23.
- Kerihuel, J. C. & Dreyfus, J. F. (1991) Meta-analyses of the efficacy and tolerability of the tricyclic antidepressant lofepramine. *Journal of International Medical Research*, 19, 183–201.
- Moller, H. J. & Haug, G. (1988) Secondary and meta-analysis of the efficacy of non-tricyclic antidepressants. *Pharmacopsychiatry*, 21, 363–364.
- Montgomery, S. A. & Kasper, S. (1995) Comparison of compliance between serotonin reuptake inhibitors and tricyclic antidepressants: A meta-analysis. *International Clinical Psychopharmacology*, 9 (suppl. 4), 33–40.
- Pande, A. C. & Saylor, M. E. (1993a) Fluoxetine – a clinical and research update: Severity of depression and response to fluoxetine. *International Clinical Psychopharmacology*, 8, 243–245.
- & — (1993b) Adverse events and treatment discontinuations in fluoxetine clinical trials. *International Clinical Psychopharmacology*, 8, 267–269.
- Patten, S. B. (1992) The comparative efficacy of trazodone and imipramine in the treatment of depression. *Canadian Medical Association Journal*, 146, 1177–1182.
- Sinclair, J. C. & Bracken, M. B. (1994) Clinically useful measures of effect in binary analyses of randomized trials. *Journal of Clinical Epidemiology*, 47, 881–889.
- Song, F., Freemantle, N., Sheldon, T. A., et al (1993) Selective serotonin reuptake inhibitors: Meta-analysis of efficacy and acceptability. *British Medical Journal*, 306, 683–687.
- Wilson, A. & Henry, D. A. (1992) Meta-analysis. Part 2: Assessing the quality of published meta-analyses. *Medical Journal of Australia*, 156, 173–187.
- Workman, E. A. & Short, D. D. (1993) Atypical antidepressants versus imipramine in the treatment of major depression: A meta-analysis. *Journal of Clinical Psychiatry*, 54, 5–12.

Multiple choice questions

- 1 Meta-analysis:
 - a involves the statistical pooling of the results of individual studies
 - b is not dependent on the quality of individual studies
 - c is not subject to bias in the selection of studies
 - d increases the power to detect small or moderate effects of clinical importance
 - e provides an estimate of the size of a treatment effect.
- 2 In a methodologically sound meta-analysis:
 - a studies should not be excluded on the basis of quality
 - b the search procedure should be explained
 - c selection criteria for studies should be explicit
 - d it is not necessary to check for publication bias
 - e statistical pooling should take into account the size or variance of individual studies.
- 3 Meta-analyses of comparative antidepressant studies:
 - a should include single blind studies
 - b do not require patients to be randomly allocated to treatment groups
 - c should give confidence intervals for the difference in effect
 - d can provide confidence that treatments are equally clinically effective
 - e cannot assess the tolerability of individual drugs.
- 4 Meta-analysis has shown that:
 - a lofepramine is less effective than older TCAs
 - b fluoxetine treatment protects against the emergence of suicidal ideas in depressed patients
 - c SSRIs are as effective as TCAs in the short-term treatment of depression
 - d trazodone is less effective than imipramine in the treatment of depression
 - e SSRIs cause fewer drop-outs attributable to side-effects than older TCAs.
- 5 The following are true:
 - a meta-analysis has been widely used in psychiatry
 - b the quality of meta-analyses of comparative antidepressant trials is good
 - c meta-analyses are only as good as the studies they use
 - d meta-analysis may identify new research questions
 - e the results of meta-analysis can be accepted uncritically.

MCQ answers

1	2	3	4	5
a T	a F	a F	a F	a F
b F	b T	b F	b F	b F
c F	c T	c T	c T	c T
d T	d F	d T	d F	d T
e T	e T	e F	e T	e F