

Clinical trial design: horses for courses

JOHN R. GEDDES

University Department of Psychiatry, Warneford Hospital, Oxford, OX3 7JX, UK

Fleischhacker and Goodwin contribute to the ongoing debate around the relative value of so-called “efficacy” and “effectiveness” trials. Comparisons between trials need to take into account the fact that different trials are designed to answer different clinical questions and that methodological choices inevitably involve compromises (1,2). Various approaches to describing the different priorities, and designs, of clinical trials have been suggested: e.g., explanatory vs. pragmatic (3), practical vs. large simple vs. efficacy (2) and large simple vs. small complex (4).

Explanatory (“efficacy”) trial designs tend to have a greater degree of control over internal validity and a higher signal-to-noise ratio, but will typically tend to sacrifice external validity (i.e., applicability to real world patients) (3,5). The aim of explanatory trials is to determine if the experimental intervention *can* work in controlled, optimized, circumstances (6). By contrast, the objective of more pragmatic (“effectiveness”) trials is to determine if the intervention *does* work in the real world of clinical practice, which is, almost by definition, a more noisy and less controlled environment. While pragmatic trials include less selected and more representative patients and clinical sites and use less standardized, more routine, measure of clinical outcomes, the design compromises inherent in effectiveness trials will tend to increase variability (and hence statistical noise) and, frequently, bias.

Fleischhacker and Goodwin are right, therefore, to highlight the problem that arises when there appears to be a discrepancy between the results of explanatory and more pragmatic trials. Are the explanatory or the pragmatic trial(s) intrinsically more reliable and likely to produce a closer estimate of the “true” effect of the investigational agent? This

question is chimerical. In fact, the vast majority of trials lie on a continuum between idealised explanatory and pragmatic designs. Every trial needs to be critically appraised on its own merits for likely sources of bias and noise.

Fleischhacker and Goodwin consider that randomisation with adequate concealment of allocation is the *sine qua non* of a fair comparison of two (or more) treatments. However, empirical studies suggest that other design characteristics, such as blinding, can also have substantial effects on the chances of a trial producing an unbiased result (7). Indeed, for trials with subjectively assessed outcomes, absence of blinding seems to be as important a cause of bias as inadequate allocation concealment (8). Lack of blinding can lead to both *performance bias* (knowledge of allocation leads to systematically different behaviour of physician and patient) and *ascertainment bias* (knowledge of allocation leads to systematically different assessment of outcomes between treatment and control groups). One might predict that blinding will be particularly important when both the possibility and the likelihood of these biases is high. This will be the case when behaviour and outcomes are easily modifiable and when true equipoise is absent and the investigator and/or participants have clear preferences between the compared treatments.

A good example of the need to take the designs of individual trials into account is provided by the trials comparing first generation (FGAs) with second generation antipsychotics (SGAs). By way of context, there was considerable hope that SGAs would provide a substantial step forward in the treatment of schizophrenia. This led to an early tendency to overlook the methodological limitations of the industry conducted trials (which were towards the explanatory end of the design spectrum) (9), to overrate the advantages of the SGAs (10) and for a rapid clinical shift to using SGAs in

preference to FGAs (11). Systematic reviews and meta-analyses of the industry-sponsored trials essentially found similar results (12-14), although the authors of one of the reviews drew notably more favourable conclusions concerning SGAs than the others (14,15).

A number of non-industry randomised controlled trials comparing FGAs and SGAs have now been reported (16-19). This number of independent trials is unusual in psychiatry: it is both a critically important development and a reflection of the rare degree of continuing uncertainty and importance of this issue. Taken as a group, these independent trials seem to indicate that, although there may be minor differences in efficacy between drugs, such benefits cannot be shown to be cost-effective and appear to be counterbalanced by an increased rate of certain adverse effects. However, this broad conclusion should not obscure the fact that these trials have very different designs and that they were aimed at different, although complementary questions. Space prevents a full critical appraisal of each of these trials and so I will discuss only some selected issues.

The CATIE trial, termed a “practical” trial by its designers, had some pragmatic characteristics (representative patients, variable dosing, reasonably long follow-up) but maintained blinding and high quality assessment of outcome (2). The results of CATIE were unsurprising although very valuable in that they confirmed a picture that was emerging from disparate strands of evidence, including both the meta-analyses and emerging observational data on safety (20).

CUtLASS was further along the pragmatic continuum, being unblinded and allowing choice of both SGA and FGA (18). Out of context, it inevitably remains unclear to what extent the lack of observed differences in CUtLASS mean that there were truly no differences or that the trial was too “noisy” to detect them. However, the CUtLASS cost-effectiveness findings are highly consistent with the other independent trials.

EUFEST was an ambitious trial of first episode patients, but its open design made it highly susceptible to performance bias (and consequently ascertainment

bias), which led to its authors being unable to draw any clear conclusions (19), although Fleischhacker and Goodwin seem now more prepared to do so.

Space precludes a discussion of the other effectiveness trials in other disorders discussed by Fleischhacker and Goodwin, but similar issues apply. It is clearly the case that neither “efficacy” or “effectiveness” trials are more likely to estimate the “truth”. All trials are susceptible to limitations and trial design is the art of compromise. All trials should therefore be critically appraised. Despite the various methodological shortcomings in the new generation of independent trials, their resurgence – and the willingness of government and charities to fund them – is long overdue. Many important clinical questions remain unanswered by trials designed solely to meet the narrow needs of industry and their regulatory authorities (1,21). Those designing trials with a more pragmatic focus need to make sure that important sources of bias are identified for each individual trial. It is crucial that the trial design is robust enough to make the results both credible and useful – otherwise the hard-earned results will be vulnerable to the criticisms of those who do not like them!

References

1. Geddes JR. Large simple trials in psychiatry: providing reliable answers to important clinical questions. *Epidemiol Psychiatr Soc* 2005;14:122-6.
2. Stroup TS, Geddes JR. Randomized controlled trials for schizophrenia: study designs targeted to distinct goals. *Schizophr Bull* (in press).
3. Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutic trials. *J Chronic Dis* 1967;20:637-48.
4. Peto R, Collins R, Gray R. Large-scale randomized evidence: large, simple trials and overviews of trials. *J Clin Epidemiol* 1995; 48:23-40.
5. Sackett DL. Why randomized controlled trials fail but needn't: 2. Failure to employ physiological statistics, or the only formula a clinician-trialist is ever likely to need (or understand!). *CMAJ* 2001;165:1226-37.
6. Haynes B. Can it work? Does it work? Is it worth it? *BMJ* 1999;319:652-3.
7. Schulz KF, Chalmers I, Hayes RJ et al. Empirical evidence of bias. Dimensions of

methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;273:408-12.

8. Wood L, Egger M, Gluud LL et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ* 2008;336:601-5.
9. Bagnall AM, Jones L, Ginnelly L et al. A systematic review of atypical antipsychotic drugs in schizophrenia. *Health Technol Assess* 2003;7:1-193.
10. Cipriani A, Geddes JR. Comparison of systematic and narrative reviews: the example of the atypical antipsychotics. *Epidemiol Psychiatr Soc* 2003;12:146-54.
11. Geddes J. Generating evidence to inform policy and practice: the example of the second generation “atypical” antipsychotics. *Schizophr Bull* 2003;29:105-14.
12. Geddes J, Freemantle N, Harrison P et al. Atypical antipsychotics in the treatment of schizophrenia: systematic overview and meta-regression analysis. *BMJ* 2000; 321:1371-6.
13. Leucht S, Pitschel WG, Abraham D et al. Efficacy and extrapyramidal side-effects of the new antipsychotics olanzapine, quetiapine, risperidone, and sertindole compared to conventional antipsychotics and placebo. A meta-analysis of randomized controlled trials. *Schizophr Res* 1999;35:51-68.
14. Davis JM, Chen N, Glick ID. A meta-analysis of the efficacy of second-generation antipsychotics. *Arch Gen Psychiatry* 2003; 60:553-64.
15. Lieberman JA. Comparative effectiveness of antipsychotic drugs: a commentary on Cost Utility of the Latest Antipsychotic Drugs in Schizophrenia Study (CUtLASS 1) and Clinical Antipsychotic Trials of Intervention effectiveness (CATIE). *Arch Gen Psychiatry* 2006;63:1069-72.
16. Lieberman JA, Stroup TS, McEvoy JP et al. Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *N Engl J Med* 2005;353:1209-23.
17. Rosenheck R, Perlick D, Bingham S et al. Effectiveness and cost of olanzapine and haloperidol in the treatment of schizophrenia: a randomized controlled trial. *JAMA* 2003;290:2693-702.
18. Jones PB, Barnes TRE, Davies L et al. Randomized controlled trial of the effect on quality of life of second- vs. first-generation antipsychotic drugs in schizophrenia: Cost Utility of the Latest Antipsychotic Drugs in Schizophrenia Study (CUtLASS 1). *Arch Gen Psychiatry* 2006;63:1079-87.
19. Kahn RS, Fleischhacker WW, Boter H et al. Effectiveness of antipsychotic drugs in first-episode schizophrenia and schizophreniform disorder: an open randomised clinical trial. *Lancet* 2008;371:1085-97.
20. Koro CE, Fedder DO, L'Italien GJ et al. Assessment of independent effect of olanzapine and risperidone on risk of diabetes among patients with schizophrenia: population based nested case-control study. *BMJ* 2002;325:243.
21. Geddes J, Goodwin G. Bipolar disorder: clinical uncertainty, evidence-based medicine and large-scale randomised trials. *Br J Psychiatry* 2001;178:S191-4.