

RCTs in psychiatry: challenges and the future

REBECCA WALWYN and SIMON WESSELY

INTRODUCTION

The editor of the *Lancet*, Richard Horton, an articulate critic of much of modern medicine, gave a keynote address at the 2001 *Society for Clinical Trials* annual meeting, provocatively entitled, "The clinical trial: deceitful, disputable, unbelievable, unhelpful, and shameful-what next?" Horton (2001) suggests that the clinical trial process is approaching a critical moment, with growing public scepticism already producing problems in patient recruitment for some trial organizers. Failure to recruit adequate patient numbers is a real threat to many trials and often the reason that an unknown number of trials get abandoned and left unreported. In this editorial we expand on the themes that one of us raised in a recent book (Everitt & Wessely, 2003) and look to the future and consider how the situation might be improved.

Can randomised clinical trials in psychiatry be justified? Some challenges

There have been a number of trenchant criticisms of the use of randomised controlled trials to evaluate mental health treatments. It has been argued that

- i) Psychiatric treatments are too variable and/or too complex to permit generalisations from the particular.

- ii) Psychiatric patients are too complex to permit extrapolation from one patient to the wider community.

- iii) Results from most psychiatric trials have little relevance for the day-to-day treatment of the mentally ill, i.e., the results are not generalisable.

Psychiatric treatments are too individual

Psychiatry is not 'cook book' medicine. Taking one voice from many, Silberschatz articulates the principal arguments against RCTs in psychiatry from the perspective of a psychotherapist (Persons & Silberschatz, 1998). For him the important questions are: What is bothering the patient? What do they hope to achieve? Why have they not achieved that? And so on and so forth. The argument continues that manualisation, deemed essential in psychological treatment trials to enable another clinician to be able to repeat the intervention later, and to ensure that the therapy is replicable, removes the heart of psychological treatment – empathy, therapeutic alliance and so on. What is lost in the RCT is the essential individual nature of psychological treatments. People are different, problems are different, and therefore, treatments be different.

This debate is not unique to psychiatry. Very similar questions are raised by those sceptical of the place of clinical trials in the assessment of alternative or complementary medicine. Of course people are different, but this applies across medicine. A hundred or so years of writing on the "art of medicine", the recent growth of "narrative based medicine", and the seemingly endless critiques of the limitations of narrowly orientated biomedicine, show that no one can seriously dispute the importance of understanding the individual. Nevertheless if every

Address for correspondence: Professor S. Wessely, Department of Psychological Medicine Weston Education Centre, Cutcombe Road, London SE5 9RJ (United Kingdom).

Fax: +44-(0)20-7848.5408

E-mail: S.Wessely@iop.kcl.ac.uk

Declaration of Interest: none.

patient was unique and every problem without precedent then medicine in general (and psychiatry in particular) would come to a full stop. If there were no communalities between patients, and no identifiable general patterns in particular groups of patients, then there would be no purpose in medical education, clinical experience or training. It is these shared factors that permit clinicians to draw on what they have learnt from their training and their experience, and use this acquired knowledge to assess and understand the specific patient now requiring their attention.

It is the existence of patterns of disease that make clinical trials viable. Having observed a phenomenon previously in a patient population of interest - be it a certain cancer, a particular behaviour, a biochemical abnormality, or an emotional reaction - one has something that might form the basis for a clinical trial. This systematically acquired information can be used to help future patients, without forgetting that what is truly unique about a patient (and so cannot be studied in a clinical trial) still has to be taken into account in caring for the patient. For this the treating clinician will often need large amounts of intuition, experience and empathy.

Psychiatric treatments are too complex

The argument here is that far too many things happen during even a single consultation to permit evaluation by the technology of the RCT. Certainly many of the interventions that have been developed for the treatment of mental health problems are more complex than drug treatments. Some of the characteristics of such interventions, thought by some to raise doubts about the suitability of clinical trials for their evaluation, are identified by Crawford *et al.* (2002):

- Complex interventions comprise multiple inter-connecting elements,
- Complex interventions have mechanisms of action that are difficult to identify,
- Complex interventions have effects that depend on a range of factors including the actions of the practitioners who deliver them.

No-one has ever claimed, however, that the RCT can tell you everything about psychiatric treatment. Nevertheless, so long as an intervention can be adequately described and reproduced then that intervention can be scrutinised by a clinical trial. This is more difficult for a psychological or social treatment, but certainly achievable (see MRC, 2002 for a framework for developing and evaluating RCTs for complex interventions).

Psychiatric patients are too complex

Psychiatric disorders are frequently not straightforward, and psychiatric patients often display challenging and complex behaviours that might at first sight appear incompatible with the tightly controlled demands of most clinical trials. Broad categories such as depression or schizophrenia hide several sub groups, whose boundaries are imperfectly delineated. Many (perhaps most) psychiatric patients have more than one diagnosis, something that has come to be labelled as *comorbidity*.

Complications of diagnosis and patient complexity are difficult challenges faced by psychiatric trialists, but neither provides *fundamental* objections to the use of RCTs in psychiatry. Comorbidity may, for example, affect generalisation, if the index trial was performed on an unusually "pure" subgroup of patients, but the internal validity of the data is unaffected. Trials can be (and have been) conducted in populations and situations that might seem insuperable to the faint hearted. It might have been predicted, for instance, that it would be impossible to carry out randomised trials in violent forensic patients, yet there is a seminal trial conducted in North Carolina in which 331 involuntarily hospitalised patients awaiting discharge were randomly assigned to either release or outpatient compulsory treatment (Swartz *et al.*, 2001).

Results from psychiatric trials are not generalisable

The most easily sustained criticism against the RCT in psychiatry as currently undertaken is the issue of generalisability (McKee *et al.*, 1999). But note the rider, "as currently undertaken". The fault lies not with the principles of the randomised clinical trial, but simply the way such trials are often conducted at present. The current vogue for pragmatic trials arises from the perception that many explanatory clinical trials take place in "pure" populations. For example, those free from all forms of comorbidity, with participants keen to attend follow ups, happy to take medication, and so on and so forth. The consequence of this is that the results are not considered relevant to the vast majority of the population who do suffer from comorbidity, and who might be reluctant to do any of the things mentioned. Likewise, prognostic features of patients in clinical trials may vary, even within trials, and it is certainly true that one cannot assume that because a treatment has been successful in a well conducted clinical trial, the results will apply to all patients with the same diagnosis (Rothwell, 1995). There is much merit in the arguments for more pragmatic trials in psychiatry. The answer is not for psychiatry to turn its back

on the RCT, but for trialists to push for larger, simpler trials, and to lobby against the increasing bureaucratisation of the clinical trial that stands in the way of achieving these objectives.

The future of psychiatric trials

So what does the future hold for clinical trials in psychiatry? In the coming years we anticipate that:

- The current move towards larger, simpler trials will accelerate – a positive change.
- There will be greater “consumer” involvement in setting priorities. This is to be welcomed, not least if it leads to a demystification of clinical trials, and a shared understanding of their role in protecting patients from untried and untested therapies.
- There will be a greater willingness to mix assessment methods – for example incorporating a greater use of *qualitative methods* into the experimental evaluation (Crawford *et al.*, 2002). Such methods may help to identify the cultural context, values, beliefs and community norms of target groups and thus provide the key to the design and implementation of promising interventions (see Stephenson & Imrie, 1998).
- The assessment of complex interventions, such as new services, psychotherapy and social interventions will become more prominent. More sophisticated methodologies will need to be developed to assess these within the fundamental principles of the RCT.
- More attention will be paid to issues such as the representativeness of recruitment, values and preferences of trial participants, and broader outcome measures (functioning, work and so on, rather than symptoms alone).
- Promising new interventions will appear from unexpected quarters requiring continued investment in the infrastructure and training needed to assess them.

At the top of our own particular “wish list” for the future of psychiatric trials we place size, simplicity and realism. Below we concentrate on our desire to make psychiatric trials “bigger” (larger numbers of patients) and “simpler” (fewer outcome measures, for example).

Bigger psychiatric trials

There are many reasons why clinical trials can give “wrong” answers, largely methodological, but one of the more fundamental issues is that of size. Clinical trials may give results that are later seen as flawed simply because they were too small and small trials can only detect large effects with very limited precision. Large treatment effects, unless one has stumbled on the next

penicillin, are usually *a priori* implausible. Instead most advances in medicine and psychiatry are incremental, involving small but important advances, rather than earth shattering breakthroughs. “Moderate (but worthwhile) effects on major outcomes are generally more plausible than large effects” (Collins *et al.*, 1996).

The management of depression is a fundamental question for psychiatry. At present there is no doubt that there are two classes of drugs, the tricyclics and their newer rivals, the selective serotonin reuptake inhibitors (SSRIs) that are both effective in management. But which is better? And what does “better” mean?

We might easily agree that should one class of drugs be, say 50%, better (however defined) than the other, then this group would immediately become the treatment of choice and the results would represent a dramatic breakthrough in treatment. Even a 25% improvement in outcome from one class of antidepressants over the other would be of considerable importance, and indeed still be close to being a “dramatic breakthrough”. But since depression is a very common problem worldwide (the World Bank analysis predicts that it will be the second most common cause of disability across the world by 2020) most psychiatrists would agree that even a 10% improvement produced by one class of drugs over the other would be a very worthwhile benefit.

Sadly the evidence from the literature of trials comparing tricyclics with SSRIs demonstrates that such trials were incapable of detecting any difference much smaller than the “dramatic breakthrough”. Hotopf and colleagues (Hotopf *et al.*, 1997), for example, analysed all the trials that compared tricyclics “head to head” with SSRIs (there were 121 of them at the time of the study – there are more today). Quite a few of the trials were sufficiently large to be able to detect that SSRIs were about 50% better in improving outcome than tricyclics; none of course did, and such a quantum leap in efficacy was always improbable. But we have argued that if the SSRIs were actually 20% better, then this would be real progress, and worthwhile knowing. Less than a dozen of the trials of those examined by Hotopf *et al.* (1997) could have detected such an effect. And if the differences were 10% - perhaps the most realistic possibility, then not a single trial could have come anywhere near detecting what would still be an important improvement in the management of depressed patients, although of course systematic reviews and meta-analyses could assist.

Since the publication of the Hotopf *et al.* paper, the sample size of antidepressant trials has indeed started to increase – for example Kurt Kroenke and colleagues in the United States carried out a study designed to directly

compare three antidepressants, and this time used a sample size of 573 adult depressed patients recruited from primary care (Kroenke *et al.*, 2001). They also failed to find any differences, but for the first time we can be more confident that had important differences existed, this study would have had the power to detect them.

So that it appears that trials in depression are seriously under-powered for detecting small but important differences between treatments, and the situation is no better in schizophrenia. For example, a study of over 2000 trials in schizophrenia found that the mean sample size was about 60 (Thornley & Adams, 1998). And although Johnson (1998) from a search of the four leading psychiatric journals from 1956 thorough to 1996 found that the number of patients per treatment group in psychiatric trials was indeed rising over the four decades, the increase, from 17 to 25 over 40 years, was hardly impressive! (Johnson, 1998).

Fortunately there are now some encouraging signs of change. As far as we know the largest trial yet completed in psychiatry is the Lilly sponsored study comparing olanzapine with haloperidol for the treatment of schizophrenia, which randomised 1996 patients across Europe and North America (Tollefson *et al.*, 1997). At the same time, researchers in the United Kingdom completed what may be the largest trial looking at different models of community care – the so called “UK 700” trial which actually recruited 708 patients (UK 700 Group: Burns *et al.*, 1999). And in progress as we write is what could have been the largest trial yet seen in psychiatry since the initial aim was to recruit 3,000 patients with a diagnosis of bipolar affective disorder to provide a comparison of lithium, valproate and a combination of the two (Geddes, 2002). Sadly problems with funding meant that the sample size has now been reduced to 1068, but the final study will still be a major contribution.

Simpler psychiatric trials

Large trials are urgently necessary in psychiatry but are not a panacea for all ills. They are costly in terms of resource and time. They have to be simple, yet as we have said, psychiatric interventions are not necessarily simple (it is all very well to carry out a mega trial of aspirin, but we doubt there would have been the same enthusiasm for a similarly large trial of, for example, aortic valve surgery). In most large trials the intervention itself is of relatively short duration, sometimes just a single tablet – but this is unlikely to be the case in psychiatry.

Although we accept that the complexity of some psychiatric interventions makes it more difficult for trials of

such interventions to be simple we also readily concede that in mental health we seem to have a vested interest in making things more complex than is necessary. Diagnostic issues in psychiatry can become something of a fetish, and taken to extremes can undermine the inherent simplicity of the clinical trial; few clinicians really care, for example, about the sub divisions of somatoform disorders or whether someone has dysthymia or double depression. Psychiatrists also use far too many rating scales to measure far too many things in their trials, increasing the chances of false positive findings (as the Oxford, UK, group of trialists note, “many trials would be of much greater scientific value if they collected 10 times less data on 10 times more patients”). An analysis of trials on the Cochrane Schizophrenia Database found that over 640 different rating scales had been employed (Thornley & Adams, 1998; Gilbody *et al.*, 2002). The use of a large number of outcome measures is driven by the fear of missing something that might be “clinically significant” even if that ‘something’ was not the primary reason for carrying out the study. But any advantages of such an approach are massively outweighed by the disadvantages, in particular those of multiple testing, and loss of simplicity both in analysis and in understanding of results.

CONCLUSION

None of the objections raised to psychiatric trials are fundamental. Clinical trials are certainly not perfect but they remain the essential methodology in the evaluation of the effectiveness of treatments, psychiatric and non-psychiatric. No alternative is available that is more likely to lead to results that will lead to confident, reliable recommendations about treating patients that can be used to improve clinical care. While it remains important to strike a balance between the desire for large, simple trials and what is possible given probable financial and temporal constraints, as standard treatments improve, we will need to work towards detecting smaller, but still important, treatment effects using much larger, simpler and realistic trials. Quality will of course always be demanded whatever the size or complexity of the trial.

REFERENCES

- Burns T., Creed F., Fahy T., Thompson, S., Tyrer P. & White I. (1999). Intensive versus standard case management for severe psychotic illness: A randomised trial. *Lancet* 353 (9171), 2185-2189.
- Collins R., Peto R., Gray R. & Parish S. (1996). Large scale randomi-

- zed evidence: Trials and overviews. In *Oxford Textbook of Medicine*, 3rd ed. (Ed. D. Weatherall, J. Ledingham and D. Warrell), pp. 21-32. Oxford University Press: Oxford.
- Crawford M.J., Weaver T., Rutter D., Sensky T. & Tyrer P. (2002). Evaluating new treatments in psychiatry: The potential value of combining qualitative and quantitative research methods. *International Review of Psychiatry* 14(1), 6-11.
- Everitt B. & Wessely S. (2003). *Clinical Trials in Psychiatry*. Oxford University Press: Oxford.
- Geddes J. (2002). Can we conduct some large simple trials in bipolar disorder? *Bipolar Disorder* 4, Suppl 1, 62-63.
- Gilbody S., Wahlbeck K. & Adams C. (2002). Randomized controlled trials in schizophrenia: a critical perspective on the literature. *Acta Psychiatrica Scandinavica* 105(4), 243-251.
- Horton R. (2001). The clinical trial: deceitful, disputable, unbelievable, unhelpful, and shameful-what next? (Keynote Address). *Society for Clinical Trials Annual Meeting*.
- Hotopf M., Lewis G. & Normand C. (1997). Putting trials on trial - the costs and consequences of small trials in depression: a systematic review of methodology. *Journal of Epidemiology and Community Health* 51(4), 354-358.
- Johnson T. (1998). Clinical trials in psychiatry: Background and statistical perspective. *Statistical Methods in Medical Research* 7(3), 209-234.
- Kroenke K., West S.L., Swindle R., Gilsenan A., Eckert G.J., Dolor R., Stang P., Zhou X.-H., Hays R. & Weinberger M. (2001). Similar effectiveness of paroxetine, fluoxetine, and sertraline in primary care: A randomized trial. *Journal of the American Medical Association* 286(23), 2947-2955.
- McKee M., Britton A., Black N., McPherson K., Sanderson C. & Bain C. (1999). Methods in health services research: Interpreting the evidence: Choosing between randomised and non-randomised studies. *British Medical Journal* 319(7205), 312-315.
- MRC (2002). A framework for development and evaluation of RCTs for complex interventions to improve health. Retrieved May 3, 2005 from http://www.mrc.ac.uk/pdf-mrc_cpr.pdf
- Persons J.B. & Silberschatz G. (1998). Are results of randomized controlled trials useful to psychotherapists? *Journal of Consulting and Clinical Psychology* 66(1), 126-135.
- Rothwell P. (1995). Can overall results of clinical trials be applied to all patients? *Lancet* 345(8965), 1616-1619.
- Stephenson J. & Imrie J. (1998). Why do we need randomised controlled trials to assess behavioural interventions? *British Medical Journal* 316(7131), 611-613.
- Swartz M.S., Swanson J.W., Hiday V.A., Wagner H.R., Burns B.J. & Borum R. (2001). A randomized controlled trial of outpatient commitment in North Carolina. *Psychiatric Services* 52(3), 325-329.
- Thornley B. & Adams C. (1998). Content and quality of 2000 controlled trials in schizophrenia over 50 years. *British Medical Journal* 317(7167), 1181-1184.
- Tollefson G.D., Beasley C.M. Jr, Tran P.V., Street J.S., Krueger J.A., Tamura R.N., Graffeo K.A. & Thieme M.E. (1997). Olanzapine versus haloperidol in the treatment of schizophrenia and schizoaffective and schizophreniform disorders: Results of an international collaborative trial. *American Journal of Psychiatry* 154(4), 457-465.