The maturation of randomised controlled trials in mental health

Graham Thornicroft
Health Service Research Department, Institute of Psychiatry, De Crespigny Park, King's College London, London SE5 8AF, England

Abstract
The aims of this paper are: (i) to give an overview of the use and maturation of randomised controlled trials (RCTs) in mental health services research, (ii) to indicate areas in which mental health may present particular challenges, and (iii) to outline necessary steps to strengthen the capacity to conduct better quality randomised controlled trails.

Keywords: *Randomised controlled trials, Mental health*

A useful starting point for this discussion is the view of Barker & Rose that 'Randomisation is the hallmark of an honest trial' by which they meant that the results of an RCT are more likely to give a clear and a trustworthy result to a research question than any other form of trial.[1] Put differently, "Randomised clinical trials are the sine qua non for evaluating treatment in man".[2] There is now a substantial literature which testifies to the fact that the RCT is regarded as the 'gold standard' for answering questions about treatment efficacy.[3,4,5]

In the following sections the reasons for such a robust view will be described. Certain technical issues particular to RCTs in mental health will be discussed, and an attempt to set out a balanced view of when RCTs are useful or even essential, and when they are of limited use or are even inappropriate, will be made.

To put such trials in context, Table 1 shows the number of current RCTs active in the fields of cancer, cardio-vascular and mental health research, according to two registers of trials. Although there are some relatively minor discrepancies, the results are remarkably similar from the two sources and show that the number of mental health trials is about 10% that of cancer trials and 30%-50% that of cardio-vascular trials. In addition, RCTs are a relatively recent addition to the pantheon of published research in psychiatric and mental health research and in general this design is relatively under-used compared with many other fields of biomedical science.

Table 1. Comparative estimates of the number of current mental health RCTs

	Cancer	*Cardio-vascular*	*Mental health*
ControlledTrials.com	2230	782	242
NIH Clinicaltrials.gov	1957	419	251

Correspondence:
Professor G Thornicroft, Health Service Research Department, Institute of Psychiatry, De Crespigny Park, King's College London, London SE5 8AF, England
email: g.thornicroft@iop.kcl.ac.uk

Advantages of RCTs

Pocock gives a detailed account of the use, design and analysis of RCTs.[3] This design is extremely powerful and one of its major advantages is that it controls for the many confounding variables that may exist. It also eliminates the problematic effects of spontaneous remission, regression to the mean, and the placebo effect, all of which can produce improvement that might be incorrectly attributed to the treatment. A further important advantage provided by the RCT is that, if blindness is maintained, the results are independent of any bias from the clinicians involved in giving treatment or from the researchers conducting the study. The RCT methodology in medicine as a whole, and in psychiatry in particular, has been based on evaluating the efficacy of new drugs. Leber states that in the view of the Food and Drug Administration in the USA there is simply no acceptable alternative to the randomised controlled trial in assessing drug efficacy.[5] However the methodology has been taken up enthusiastically in evaluating other treatments, for example the psychotherapies[6] and in evaluating alternatives to hospital treatment in psychiatry.

RCTs may be used not only as the basis for meta-analytic methods of summarising knowledge in a particular field of medicine, but also as a strong form of evidence in their own right. The cornerstone of mental health policy in the United Kingdom (UK), for example, is the National Service Framework for Mental Health.[7] It is based upon a foundation of supporting evidence, where the relevant research is categorised according to the following rank order of evidential strength.

- Type I at least 1 good systematic review, including at least 1 RCT
- Type II at least 1 good RCT
- Type II >1 well-designed intervention study without randomisation
- Type IV >1 well-designed observational study
- Type V expert opinion, including the opinion of service users and carers

Technical challenges in RCT design

The level of random allocation is a common question at the design phase of a study. The allocation of individual patients is the most common choice, especially for 'simple' trials of a single intervention. Allocation of clinician or practitioners may be indicated where, for example, the research question related to the use of clinical guidelines by doctors, in which case the practitioners may be studied directly, or the effects upon the clusters of patients they treat may also be investigated. The allocation of clinical or professional teams is a further option where an intervention may apply to a whole group of staff, such as a training course. Finally, larger practice or population groups, such as localities, may be the unit of random allocation for particular studies, for example where the experimental condition to be tested consists of a new provision to a whole local population, such as a public health education campaign.

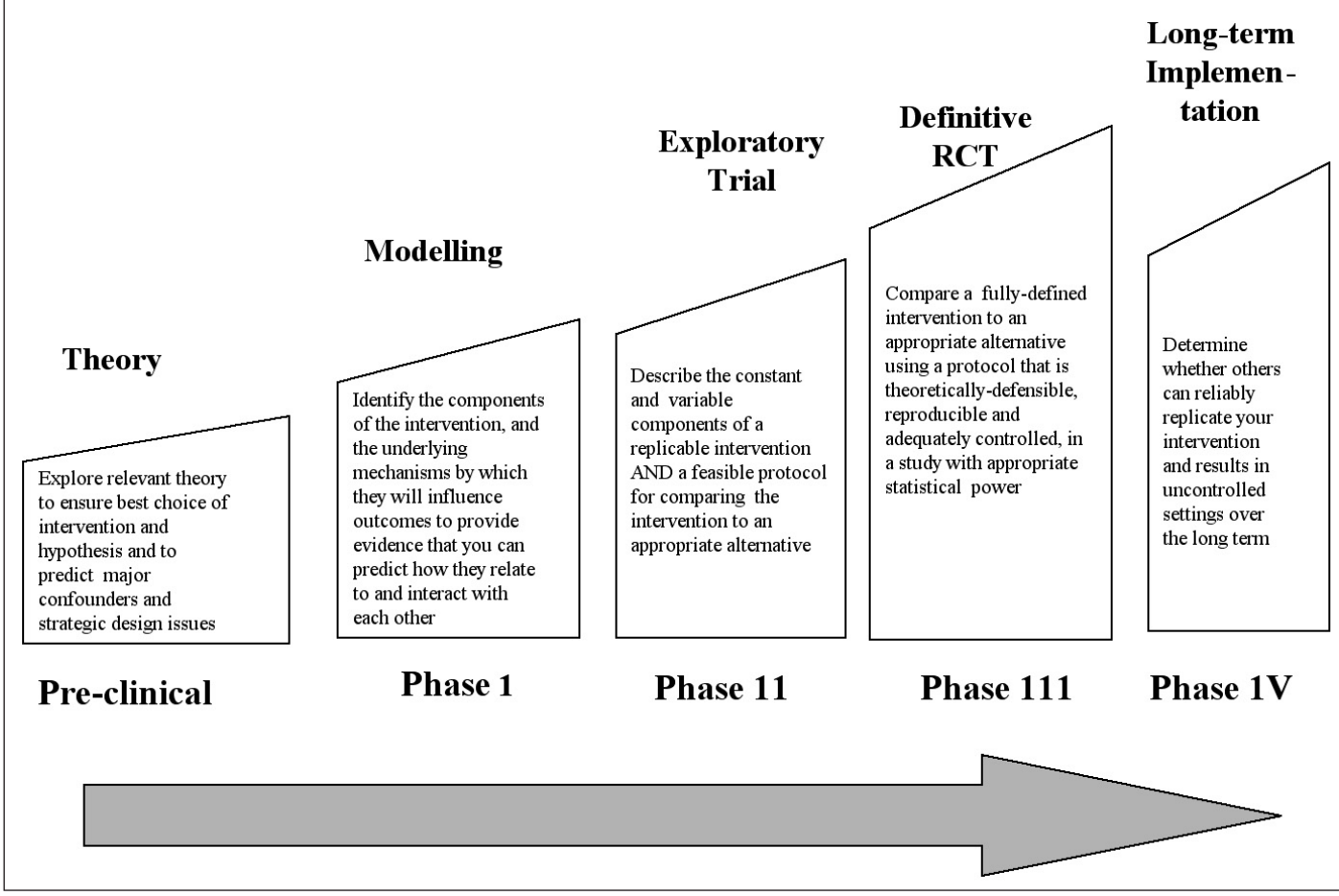
Although much attention is usually focussed in RCT design and analysis upon the experimental condition, it is vital to keep in mind that most RCT analyses report differences between experimental and control groups. Such differences rely as much upon the characteristics and outcomes of the control groups as upon those of the experimental groups. For example, the process of policy creep, in which a new policy such as the provision of mental health care in primary care is introduced into a region or country, may mean the additional advantage of any enhanced form of specialised input at the primary care level. Often such differences are masked, since the control condition may be generically referred to as 'treatment as usual', and it is relatively rare for a control condition to be described in the same level of detail as an experimental condition. This may be especially important in cross-national studies where a treatment protocol developed abroad is used, with 'fidelity', in a new setting, whereas the control conditions in the two sites show great heterogeneity.

Even where the RCT design is entirely appropriate, many such trials in medical research struggle to achieve their aims.[8] Barriers may exist to patient participation, especially where patients do not consent to participate, for example for fear of being unreasonably 'experimented upon'. Further, poor study design may narrow the entry criteria excessively so that fewer than expected patients can be recruited to plan. The lack of an adequate infrastructure, for example the absence of a trial manager or trial nurses, may also produce low recruitment and frustrate the completion of an RCT. In addition, insufficient expertise in RCT data analysis, for example in the handling of missing cases or in intention to treat approaches, or inexperience in reporting procedures, such as Consort guidelines, may limit the contribution of any particular study to the scientific literature.[9]

Challenges for mental health trials

While many, perhaps most, of the challenges faced by mental health trials are shared by trials in other areas of medicine, there are additional difficulties specific to the field of mental health . Until recently, the use of non-standardised outcome measure was more common.[10] In addition many mental health treatments can be considered complex treatment interventions. This, according to the recent UK Medical Research Council scheme, offers a struc-

Table 2. A Structural Framework for Development and Evaluation of RCTs for Complex Interventions to Improve Health



ture to the chain of events from initial idea to dissemination of proven intervention. It can be seen as a parallel to the five phases recognised in the development of pharmaceutical interventions, (Table 2).[11]

The question of consent may be a particular difficulty in mental health trials where the patients lack or may lack the capacity to consent. While the research guidelines governing informed written consent are now rigorous in many countries, the criteria whereby a patient may be judged to lack the capacity to give or refuse consent are far less satisfactorily elaborated, as the arrangements whereby ethically acceptable proxy consent may be sought or given for the severely or chronically incapacitous patient.[12] Interestingly, for trials where the level of allocation is for a group, team or local area, it will usually be necessary to gain the consent to participate of both the representatives of that group, as well as to seek the consent of the individuals concerned.

A further serious shortcoming of many mental health trials is that they are commonly underpowered, so that they are not of sufficient statistical power to clearly answer the question posed. Compared with many studies in the fields of HIV, communicable diseases, cancer and cardio-vascular medicine, where trials may frequently included thousands or ten of thousands of patients, to date it is uncommon to see mental health trials including more than a few hundred patients. Further, other fields of medical research often use simple dichotomous outcomes, for example, death or survival over a fixed follow-up period, which can reliably be established from administrative sources, so that completeness rates at follow-up can be expected to be very high. By comparison, death rates in the short to medium term are relatively low for most psychiatric conditions, and much more often interval or ration scales are used to rate outcome.

In addition, there may be barriers to the proper conduct of trials because of a lack of research mindedness. For example, in many social and clinical settings as well as among service user groups, RCTs are seen by staff to be unethical because patients in the control condition do not receive the experimental intervention, which is assumed a priori to be superior. Further, trials may impose time demands on staff which they refuse to accept.

Cluster randomisation designs

There has been considerable recent debate about the importance of cluster RCT designs. Donner & Klar have argued that it may be more important to randomise organisational than treatment interventions as more possible and unknown confounders exist.[13] They offer a form of trade off between less risk of contamination at lower levels of randomisation against the disadvantage of decreasing the number of clusters. Nevertheless they may lead to increasing logistical problems at higher organisational levels, for example in attempts to randomised localities or regions, where fewer clusters run the higher risk of imbalance. Care needs to be taken to ensure that there is sample size inflation to allow for intracluster correlation as the assumption of independence between individuals is violated. In this type of trial it is necessary to specify eligibility criteria at both the individual and the cluster levels.[14] Even so, the UK Medical Research Council has recently issued guidelines which give a clear preference to RCTs using random allocation at the individual level wherever possible.[15]

Where clusters are randomised Donner and Klar have devised guidelines on the types of justification that are necessary for this choice:[13]

- justify the use of cluster randomisation
- provide a clear definition of the unit of randomization.
- indicate whether inferences are primarily directed at the cluster level or at the level of the individual subject.
- describe the process of consent used for (i) randomizing clusters, and (ii) collecting data from study participants.
- describe the experimental design (e.g. completely randomized, stratified, matched-pair) and the method of randomisation
- explain how the chosen sample size or statistical power calculations accounts for between-cluster variation.

These authors have also proposed headings to be used for proper reporting of the results of cluster randomised trials as follows:

- provide the number of clusters randomized, the average cluster size and the number of subjects selected for study from each cluster
- provide the values of the intracluster correlation coefficient as calculated for the primary outcome variables.
- compare the baseline characteristics of the intervention groups at both the individual and the cluster level
- explain how the chosen sample size or statistical power calculations accounts for between-cluster variation.
- describe how prognostically important baseline risk factors were adjusted for
- report on loss to follow-up of both individuals within clusters and entire clusters.

The maturation of RCT methodology in mental health

One way to conceive of RCT use within mental health research is in terms of a process of maturation. One can say that until recently RCTs have been in a period of infancy in which their primary characteristics have been: under sized samples, clinically unrealistic exclusion criteria, single and unrepresentative sites, and over-specified expert interventions. Arguably we have recently entered into a period of trial adolescence in which the conduct of trials is more technically adequate, for example concordance to international standards of good clinical practice, the use of multiple centres, sample sizes of sufficient power, closer collaboration with trial statisticians, some simplification and some convergence of outcome measures.

A more mature approach to trials in future can be achieved by emphasising: (i) longer term outcome studies eg early intervention for psychosis, (ii) applying trials to largely evidence-free areas of clinical practice: forensic, in-patient and out-patient care, carer interventions, learning disabilities, personality disorders, (iii) establishing clinical trials networks, (iv) considering the sociocultural context in which trials take place, (v) the increasingly sophisticated use of patient-based outcome measures, (vi) the use of a wider range of trial methodologies, such as Zelen, preference or cluster designs .[16,17,18,19,20]

Conclusion

In this paper it has been argued that until relatively recently the use of RCTs in mental health research, with the exception of trials of simple and single interventions, has shown a lack of methodological sophistication. In most research active countries there is weak and fragmented RCT capacity, while trials have more often assessed efficacy in experimental settings than effectiveness in routine care. There has been little evidence of sequential lines of enquiry leading, in the terms of the MRC Framework,

from exploratory to definitive trials. The acceleration of the judicious use of RCTs in the coming years is likely to be assisted by full participation of service users in their design, and the development of new methods of integrating qualitative and qualitative methods within research projects and programmes.[21,22]

References

1. *Barker D. and Rose G. Epidemiology in Medical Practice Churchill Livingstone, London, 1979.*
2. *Korn EL, Baumrind S. Randomised clinical trials with clinician-preferred treatment. The Lancet 1991; 337, 149-153.*
3. *Pocock SJ. Clinical Trials: A Practical Approach. New York: Wiley 1983.*
4. *Kraemer HC, Pruyn JP. The Evaluation of Different Approaches to Randomized Clinical Trials. Archives of General Psychiatry 1990; 47, 1163-1169.*
5. *Leber P. The Future of Controlled Clinical Trials. Psychopharmacology Bulletin1991; 27(1) 3-8.*
6. *Andrews, G. Evaluating Treatment Effectiveness. Australian and New Zealand Journal of Psychiatry 1989; 23, -186.*
7. *Department of Health.The National Service Framework for Mental Health. Modern Standards and Service Models. Department of Health, London, 1999. Available from: http://www.doh.gov.uk/nsf/mentalhealth.htm*
8. *Prescott R, Counsell C, Gillespie W, Grant A, Russell I, Kiauka S, Colthart I, Ross S, Shepherd S, Russell D. Factors that limit the quality, number and progress of randomised controlled trials. Health Technology Assessment 1999; 2, (15).*
9. *Newell DJ. Intention-to-Treat Analysis: Implications for Quantitative and Qualitative Research. Int J Epidemiol 1992; 21(5) 837-41.*
10. *Taylor R, Thornicroft G.Uses and limits of randomised controlled trials in mental health service research, 2001. In: Tansella M. & Thornicroft G (eds) Mental Health Outcome Measures (2nd Edition). Gaskell, Royal College of Psychiatrists, London, pp166-177.*
11. *Cambell M, Fitzpatrick R, Haines A, Kinmonth AL, Sandercock P, Spiegelhalter D, Tyrer P. Framework for design and evaluation of complex interventions to improve health. British Medical Journal 2000; 321, 694-697.*
12. *Edwards S, Lilford D, Braunholz D, Jackson J, Hewison J, Thornton J. Ethical issues in the design and conduct of randomised controlled trials. Health Technology Assessment 1998; 2, (15).*
13. *Donner A, Klar N. Cluster Randomization Trials in Health Research, Arnold, London, 2000.*
14. *Ukoumunne O, Gulliford M, Chinn S, Stern J, Burney P. Methods for evaluation area-wide and organisation-based interventions in health and health care: s systematic review. Health Technology Assessment 1999; 3, (5).*
15 *Medical Research Council,2003. http://www.mrc.ac.uk/index/publications/pdf-cluster_randomised_trials-link*
16. *Brewin C, Bradley C. patient preferences and randomised controlled trials. British Medical Journal 1989; 299, 313-315.*
17. *Pawson R, Tilley N. Realistic Evaluation. London, Sage, 1997.*
18. *Ashcroft R, Chadwick D, Clark S, Edwards R, Frith L, Hutton J. Implications of socio-cultural contexts for the ethics of clinical trials. Health Technology Assessment 1997; 1, (9).*
19. *Fitzpatrick R, Davey C, Buxton M, Jones D.Evaluation patient-based outcome measures for use in clinical trials. Health Technology Assessment 1998; 2, (14).*
20. *Jadad A. Randomised controlled trials. BMJ Books, London,1998.*
21 *Slade M, Priebe S. Are randomised controlled trials the only gold that glitters? British Journal of Psychiatry 2001; 179, 286-7.*
22. *Priebe S, Slade M. Evidence in Mental Health Care, 2002. Brunner-Routledge, Hove.*

COMMENTARY

Randomized controlled trials: still somewhat immature

Gabor Keitner
Division of Psychiatry and Human Behaviour, Brown University, Providence, RI, USA

Dr. Thornicroft's article on the maturation of randomized controlled trials (RCT) in mental health provides both good news and bad news. The good news is that there is growing awareness and acceptance of RCTs in seeking evidence based practices in psychiatric care. The bad news is that we are at an even more immature stage of methodologic development than Dr. Thornicroft suggests that we are. The obstacles to sound RCTs are many and daunting. Some of these obstacles are so central as to be unresolvable.

The conflict between the design of efficacy trials that give a reasonably sound answer to a very narrow question addressing a very limited population and the design of effectiveness trials that evaluate complex questions in a more heterogeneous and "real world" population is one example. The former provides a relatively clear answer to a question so narrow as to

Correspondence:
Professor G Keitner,Division of Psychiatry and Human Behaviour, Brown University, Providence, RI 02912-G, USA.
email: gkeitner@lifespan.org

have limited clinical utility. The latter provides more applicable and generalizable answers but without any precision as to the active ingredients.

In a recent report, for instance, it was noted that fewer than fifteen percent of patients who applied to participate in clinical trials were found to be eligible for recruitment to those studies.[1] Looking at this issue from another perspective, fewer than fifteen percent of unselected patients attending a psychiatric outpatient department clinic met eligibility criteria for inclusion into clinical trials.[2,3] Results from clinical trials are nonetheless routinely interpolated to the universe of patients with similar but not identical diagnoses and problems. A call to give greater weight to data derived from RCTs without educating the professions and the public to the significant methodologic problems and limitations of RCTs may by creating a false sense of security. By raising expectations it may also lead to unsupportable clinical practices. The recent increase in polypharmacy in the treatment of most psychiatric disorders may be, partly, a reflection of this tendency to generalize from unrepresentative studies giving rise to unrealistic expectations of effectiveness.

Dr. Thornicroft notes that clinical trials allocating individual patients to a simple trial of a single or sequential intervention is most common and presumably easier to design and implement than studies of groups and systems of care. Yet even "single" clinical trials are fraught with many variables that are difficult to control for or at worst are easily manipulated to conform to pre-existing biases.

Patients can be chosen who are more or less likely to respond to the treatments studied. Patients who have comorbidity, have an illness that is too severe and difficult to treat, who are "too young," "too old," who cannot read and cooperate with the consenting process, whose illness is characterized by frequent or infrequent recurrences can be included or excluded based on desired outcomes.

Outcome measures can also be chosen to increase or decrease the likelihood of a particular result. Many RCTs use multiple outcome measures for the same set of symptoms and focus on the ones that yield the most desirable results. Even more disturbing is the trend to altering definitions of response and remission so as to make a treatment appear better than it really is. Definitions of treatment response range from 50% to 40% to 30% of change from baseline scores.

True placebo response rates are becoming increasingly difficult to ascertain. Many RCTs have placebo response rates in the range of 30 – 70% due to a tendency to pick patients with lower severity of illness, without comorbidity, or history of treatment failures and patients who will cooperate with the treatment protocols.

Studies that attempt to evaluate the effectiveness of psychotherapies alone or in conjunction with medications face even greater hurdles. This is especially so when multiple sites are used. This is often the case as it is very difficult to obtain sufficient numbers of subjects at any one site. In these studies additional problems relate to trying to control for non-specific treatment factors, establishing and monitoring therapeutic fidelity and adherence tot he treatment studied and maintaining blindness on the part of raters to the therapies being compared.

Conducting methodologically sound studies is very expensive, adding to the difficulty of arriving at meaningful data that can be reliably accepted as guideposts for clinical practice. Because of the expense of the studies it is difficult to conduct studies with large enough numbers to have adequate power to answer the questions being asked. The source of funding for the studies also appears to have some impact on the results obtained.

None of the above concerns are meant to undermine Dr. Thorncroft's attempt to draw attention to the importance of empirical testing of treatments and his assertion that RCTs are "a gold standard for answering questions about treatment efficacy." In this he is clearly correct. He is also correct in identifying the early developmental stage that we are in in understanding and using RCTs. We need to mature both in the development of more rigorous well-designed and analyzed studies and we need to mature in our ability to recognize and accommodate to data emanating from studies that do not meet such standards.

The reality is that we are unlikely to have large numbers of well-conducted studies on large numbers of patients that address many of the daily clinical questions that practitioners need answers for in order to provide competent care. This places considerable burden on clinicians to review the methodology sections of RCTs in order to determine how much of the information presented is valid and reliable as well as relevant to their practice.

For the foreseeable future, even with the increasing recognition of the value and importance of RCTs, most clinicians will, in addition, continue to rely on their own clinical experience and the experience of their colleagues to guide their clinical decision making. The challenge will be to blend these different sources and types of information into meaningful and clinically useful guidelines.

References

1. *Keitner GI, Posternak MA, Ryan CE. How many subject with major depression meet eligibility requirements of an antidepressant efficacy trial? J Clin Psychiatry 2003; 64:1091-1093.*
2. *Zimmerman M, Mattia JI, Posternak, MA. Are subjects in pharmacological treatment trials of depression representative of patients in routine clinical practice? Am J Psychiatry 2002; 159: 469-473.*
3. *Posternak MA, Zimmerman M, Miller I, Keitner G. A Reevaluation of the Exclusion Criteria Used in Antidepressant Efficacy Trials. American Journal of Psychiatry 2002; 159:191-200.*