

The evolving randomised controlled trial in mental health: studying complexity and treatment process

Jonathan Green

Abstract As a gold-standard methodology for the testing of the effectiveness of health treatments, the randomised controlled trial (RCT) continues to evolve to meet the challenges of new contexts and areas of medicine. This article reviews two particular evolving features of RCTs that make them increasingly well adapted to testing psychological interventions in mental health. The first is a new confidence that RCTs can be successfully adapted to test the more complex (often psychosocial) health interventions and technologies. The second is an increasing emphasis on using the RCT method to explore many facets of the process of treatment as well as its outcome. These two developments should help the RCT method to come of age in mental health, increase the face validity of the RCT method for practitioners and aid the effective translation of research work into improvements in practice.

There seems every possibility that we are entering a new era of treatment trials within mental health services. There is a distinguished history of important trials, but overall dissatisfaction with their number and quality, the relative lack of trials of psychosocial treatments compared with those of drug therapy and the fact they have had insufficient impact on everyday clinical practice. A number of current developments are likely to change this situation.

Changing perceptions of clinical trials

Systematic trials in mental health have often had a negative image. Some clinicians dismiss them as reductionist or not representative of complex clinical practice and claim that they disrupt clinical decision-making or clinical engagement with patients. What is worse, after all that disruptive work their results are often thought to be not applicable – or not relevant – to real clinical practice (Hotopf, 2002). This negative image can link to more general public attitudes: there has been an implicit but perhaps pervasive social view (reinforced by notorious incidents of the political misuse of science) that the public should be

protected wherever possible from being the subjects of research in general and trials in particular – often on the basis of the ‘precautionary principle’ (Harris & Holm, 2002; Green, 2006a).

However, protecting vulnerable populations from research activity can also exclude them from its benefits. For instance, the recent debate over the prescription of selective serotonin reuptake inhibitors (SSRIs) to children and adolescents highlighted how sparse has been adequate academic research and treatment trials with medication in those under 18. Pharmaceutical companies do not need child studies to obtain product licences and it has also often been felt appropriate to ‘protect’ the young from this and other academic research (Green, 2004). This is an issue that applies to a wide range of treatment research in the paediatric population, arguably to the detriment of children’s healthcare in general.

There is a more radical alternative view. If systematic open enquiry is the contemporary guarantor of robust and stable social knowledge within a plethora of easily accessible opinion (Theodosiou & Green, 2003), it can be argued that engaging in research to help generate such knowledge is a social duty rather than something from which to be protected (Harris, 2005). It follows that it would be a professional duty

Jonathan Green is a reader in child and adolescent psychiatry at the University of Manchester and an honorary consultant child psychiatrist for Manchester Children’s Hospitals Trust. His clinical and research interests have focused on disorders of social development and the study of complex treatment interventions in child and adolescent mental health services (CAMHS). He has led cohort studies and four large randomised trials in CAMHS, including most recently the Medical Research Council’s Preschool Autism Communication Trial (PACT) of a new psychosocial intervention for autism. He has a particular interest in the measurement of process in clinical trials and has written on the therapeutic alliance.

of clinicians to advocate for more research for their patients. This changing ethical perspective coincides with recent changes in theory and practice in trials themselves and their potential place in mental health practice.

Health service prioritisation of trials research

A number of recent NHS initiatives have articulated an aspiration to an enhanced role for clinical research and trials (Department for Education and Skills *et al*, 2004; Department of Health, 2004). These include the development of the Mental Health Research Network (MHRN; <http://www.mhrn.info/dnn/>) and the new National Institute of Health Research (<http://www.nihr.ac.uk/>). Established with major Department of Health funding, the MHRN is a national coordinating and facilitating body with regional hubs. Its aim is to help the running and governance of large multi-site treatment trials and to promote the involvement of clinical services and service users in their running.

Treatment trial methodology is developing

Since the initial formulations of Bradford Hill (1955) and others, the design of randomised controlled trials (RCTs) has continued to develop to meet the challenges of new situations (Johnson, 1992). In recent years there has been increasing confidence that designs can be adapted to test more complex treatment interventions. The gap, therefore, between the treatment questions that concern clinicians in everyday practice and the study designs able to test them may be narrowing.

Seminal publications from the Medical Research Council (MRC) have presented an approach to the design of trials for complex interventions in health – outlining with useful clarity the steps that should be considered in designing them (Medical Research Council, 2000). A subsequent MRC document (Medical Research Council, 2003) further emphasised the need to adapt trials to test complex interventions used in practice, particularly in the area of psychological and psychosocial treatments. So-called platform funding for trials was introduced, which is intended to support initial proof-of-concept studies, tests of feasibility and development of designs appropriate in complexity. In addition, there was a new emphasis on a collaborative ethos both in relation to clinicians and service users, with a drive to engage the public in the methodology and practice of trials. This can be seen as part of a cultural

shift linked to the views articulated by Harris (2005), which aims to put the generation of robust science-based knowledge at the centre of cultural concern rather than as a peripheral specialised issue.

The remit of this article

In this context, this article surveys recent thinking on the modification of RCT designs in relation to two key aspects of mental health trials. First, the fact that they generally concern complex treatment interventions. Second, that – since the main agent of delivery of interventions is usually interpersonal – diverse issues of treatment process are likely to be important.

Adapting RCTs to study treatment complexity

What is a complex intervention?

Simply put from a research perspective, complex interventions are those in which identification of the active ‘effective agent’ is not straightforward. In contrast to an efficacy trial of a specific drug, many treatments applied in mental health practice are multi-layered or multifaceted, or involve organisational restructuring as well as individual intervention. Moreover, it may be unclear at the outset whether, for a psychological intervention for instance, there are treatment effects from the intervention protocol itself, therapist-specific effects, effects of the environment within which the treatment is conducted or other incidental effects on the treatment process.

Of course, apparently ‘simpler’ treatments may also contain hidden complexity. The impact of so-called placebo- (non-drug- or process-) related variance in drug trials is apparently increasing (Fava *et al*, 2003) and commonly outweighs the effect of the drug itself. And the issue of treatment complexity is applicable not just to mental health interventions: many preventive programmes or complex medical interventions will have the same characteristics.

Why use randomised trials for evaluating complex interventions?

On the face of it, the rigours of the RCT design might seem to be ill-suited to studying situations of high treatment complexity. There has often been an assumption that more qualitative methods are more suitable for use in such situations. But, paradoxically, there are a number of ways in which the randomised design approach is particularly suited to testing complex treatments. This is because it is in the nature of complex treatments to have multiple

potential factors, both known and unknown, that have a bearing on outcome. Only an adequately powered randomised design technique allows for these variables to be properly controlled: those that are known and also those that are not known.

In addition, RCTs are the gold standard of treatment trial methodology, and to deprive complex (often psychosocial) interventions of their imprimatur is potentially to undervalue these areas in an evidence-based climate (Medical Research Council, 2003).

Steps to the development of randomised trials for complex interventions

In the new conceptualisation of trial methodology for complex interventions (Medical Research Council, 2000) there are three broad stages, with the actual randomised trial itself coming only at the end of important development work. These stages offer a series of fascinating intellectual and practical challenges for clinicians and researchers alike. Steps need to be taken that essentially involve a reflection on exactly what the treatment in question involves and what the active ingredients are likely to be. In this sense clinicians are being asked to address questions of the utmost interest about their practice, and questions that should logically be a prerequisite of professional activity: what exactly am I doing and how can I best model its effects?

Modelling the treatment

The first, 'pre-trial', phase is in many ways the most intellectually stimulating. It is a phase of deconstructing and modelling the treatment to be studied into researchable questions. In this crucial phase there are obviously dangers of either oversimplified reductionism in the modelling which misses key aspects, or an undersimplified re-description that does not allow research questions to be framed. This is where qualitative investigation, pencil and paper or more sophisticated modelling, and user

and clinician consultation may be of the greatest usefulness. For a particular treatment this phase may last for years while the experience is gathered and the intervention modelled in various ways. Jump too quickly past this phase and salient aspects of the intervention may be missed and thus not tested for in terms of outcome measures.

Example 1: Modelling in-patient CAMHS treatment

A sequence of clinical modelling can be illustrated in the development of a series of studies in relation to in-patient treatment in child and adolescent mental health services (CAMHS). As a group of practising clinicians in in-patient CAMHS we began to describe and model the different potential components of this highly complex intervention (Table 1). What exactly does the admission experience involve and what might be its key aspects? Which components might be essential to the treatment effect and which incidental? We considered the experience of admission itself and the fact of removal from local family and social environment; then the impact of the general ward environment or milieu, including the effect of other young people and relationships with staff. These general effects were distinguished from specific treatment programmes which might look more like out-patient work but use the ward as a base. We considered the effect of relocation to a unit school. We explored all these aspects in a descriptive way, drawing on the experience of colleagues in the discipline as well as reviewing extant models of their operation and their evidence base. This work culminated in a book that synthesised and extended these discussions (Green & Jacobs, 1998).

Finding increasing convergence about many key aspects, we refined potential models of their operation and began to see in-patient care as a series of interlocking processes, any or all (or none) of which could carry treatment effectiveness. This modelling also allowed us to develop more precise hypotheses about likely mechanisms of potential adverse effects associated with admission (Green & Jones, 1998).

We were then able to proceed to more specific consideration of the separate components: it is possible to model how familiar out-patient treatments might

Table 1 Example of the modelling of a complex intervention: components of child and adolescent in-patient admission

<i>Components</i>	<i>Potential therapeutic effects</i>	<i>Potential adverse effects</i>
Admission as relocation	Removal from (perhaps maintaining) factors in family, community, school	Removal from (perhaps hidden) supports
Provision of ward milieu	Intensive group experience for new social learning	'Contagion' of behaviours
Provision of new school environment	Intensive assessment New start and self-esteem	Generating dependency on a protected environment
Specific programmes	Intensive provision (e.g. 24h supervision for behavioural programmes)	Undermining of specific programmes by the peer group

look within the in-patient setting and be affected by some of the more non-specific aspects of milieu (Green, 2004). The ward milieu itself was studied through application of existing measures and consultation with ward staff (Imrie & Green, 1998), and a new measure was generated to try to capture 'ward atmosphere' so as to be able to test it as a variable in outcome studies. Similarly with the interpersonal relationships on the ward. We generated a new measure to try to capture the complexity of the therapeutic alliance within the in-patient unit (Kroll & Green, 1997; Green *et al*, 2001). This alliance would seem at first sight a particularly difficult phenomenon to model: the young person has relationships with the in-patient team as well as with other patients, and the parents have a largely separate set of contacts. However, it proved to be usefully measurable – in fact the child's alliance proved to be the most powerful independent predictor of health gain during treatment (Green *et al*, 2001) and emphasised the importance of measuring process variables in treatment studies. Finally, we were able to use health needs assessment methods – based largely on interviews with young people – to understand the actual experience of the intervention received during admission (and how different this might be from the formal management plans made by the team) and to evaluate how effective the different interventions were.

This modelling of the nature of the intervention and the process of treatment progressed in relation to a series of cohort studies (Green *et al*, 2001). These studies both stimulated the need for the modelling and were made possible by it. We used them to test hypotheses about the relative impact and effectiveness of the different components of care and what best predicted the outcomes of treatment. Ideally, data from these studies should now be fed back into adjustments to practice and further refinement of the modelling of the treatment and its measurement. We will then be ready to mount a systematic randomised trial against alternative interventions. To date, the process has taken 10 years of collaborative work.

The pre-trial phase must end with a robust operationalisation of what the treatment in question involves, which will usually form the basis of a manual. Such a process raises concerns about a 'cook-book' approach to therapy; a rigidity which diminishes the capacity to respond to patient individuality. But this results only if the modelling is unsophisticated. It is possible to operationalise process as much as content, and build flexibility of response into the protocol. Detailed process modelling may be more relevant for exploratory/efficacy studies; in the classic pragmatic study (see below) some of the detailed elements of the intervention can be left undefined as long as the overall approach is defined well enough to be replicable across the different intervention sites.

Confidence that a variety of treatments can be modelled in this way will be an important

counterweight to a predictable tendency otherwise for the design of new psychological treatments to follow lines most easily testable in trials – rather than those most adapted to patient need.

Constructing measurement

Another key purpose of the pre-trial phase is to define the parameters against which the treatment should be judged, strategies for deciding how to test it and, crucially, what comparison group should be chosen. Further tasks for the pre-trial phase are the testing and development of relevant measures both for process and outcome and preliminary observational studies to test various working hypotheses.

Co-construction

From a previous position where measures were solely chosen on the basis of theoretical or researcher decision, we are now entering a phase where measurement is likely to become more and more co-constructed with both fellow clinicians and service users. This is both a major challenge and an exciting opportunity. Involving users in this way should increase the face validity and external validity of trial designs, as well as form a step in the process of integrating trials into the general culture. However, clearly moves in this direction must not compromise the essential rigour of a trial. Measures must be fit for the purpose and designed to answer the primary hypothesis of the study. Measurement selection is critical and often underplayed: inadequate or superficially pragmatic measures may lead to the effort of a trial being wasted.

Example 2: Collaborative development of measures

In a new trial of an intervention for preschool children with autism, we are using initial focus groups with parents to identify what aspects of family and child functioning they think would be most relevant for a treatment to change, i.e. what are the key aspects of functioning that matter? Outcome from these groups is then refined by a process of iteration into a set of likely parameters for consideration and will be posted on the users' website for a more extensive internet-mediated consultation before a further refinement into a new quantitative measure of family functioning which will be used in the main trial.

Related questions to professionals and service users can also inform the power calculation for necessary sample size by establishing a clinically relevant 'number needed to treat' (NNT) figure. The question to professionals could be: 'For you to decide to include this new treatment into your service, what clinical effect size would be necessary,

i.e. how many cases treated to achieve one positive outcome?’ Such dialogue with professionals and service users (and commissioners) will be a key feature to integrate trials within the mainstream of clinical planning and evidence-based medicine. Clinicians will be more influenced by trials if they are involved in their design and if they see that the trial is measuring things that are relevant to them. Increasingly, the major funding bodies are requiring that such consultations have taken place to convince them of the feasibility of a new trial.

The exploratory pilot trial

Before the fully powered RCT receives funding it is usually necessary to run a preliminary exploratory or pilot trial. Here the operationalisation and manualisation of the treatment will be tested and practical matters focused on: can the treatment be delivered reliably in different sites to high enough standard? Can sufficient patient numbers be collected and how much attrition can be expected during the trial? Will the idea of the trial and its measurement be accepted by patients and practitioners? What are the correct dosage effects and should trials of different dosages of the intervention be tried? (This does not necessarily apply just to medications: ‘dosage’ might be the frequency of a psychological intervention.) What are the effect sizes shown in the measures of change and how well do they reflect the functioning of the treatment?

The main study – design variations

Exploratory *v.* pragmatic trial design

Exploratory (efficacy) trials and pragmatic (effectiveness) trials are often contrasted (Jahad, 1998; Harrington *et al.*, 2002; Box 1). This distinction derives from the classic procedure of first validating a useful treatment in controlled conditions and then studying whether such efficacious treatment will generalise effectively into routine clinical practice. This is a model suited to much pharmaceutical or laboratory-based treatment development, but may be of less conceptual value in developing and testing complex mental health interventions, where the context may be part of the object of study and where the high costs of trials may mean that it is impracticable to plan such separate stages. Nevertheless, the distinction does help clarify the issue of fitness of trial design to purpose.

Efficacy trials are organised to test mode of action as well as outcome of treatment. The design must have high internal validity; that is, it must treat the most homogeneous population group possible (to restrict variance in sampling) and must try to restrict comorbidity. From the treatment perspective it must

ensure the highest level of fidelity and consistency of treatment administration that is possible. It must address precise questions with *a priori* sub-analysis. Difficulties with efficacy designs of this kind are that they are extremely difficult to achieve in everyday psychiatric practice, as in the real world it is difficult to obtain such a pure sample or to ensure such consistency of treatment intervention. Even if such things are achieved, the efficacy trial is often compromised in terms of answering practical questions because of the lack of external validity: what is being studied in the trial bears little relationship to what happens in everyday clinical practice.

At the other end of the spectrum, effectiveness trials should have high external validity. That is, they should test as far as possible the way treatments are actually delivered in clinical practice. This is their great strength. The reciprocal weakness is the variation in trial population and details of treatment, particularly since the definition of the treatment often has to be more flexible for a pragmatic trial, for instance including patient preference (see below). There are various ways of dealing with these problems within the trial design, but the end result often is that they need larger sample sizes to maintain statistical power to identify treatment effects. Thus, pragmatic trials tend to need large

Box 1 The spectrum of pragmatic and explanatory trials

Explanatory (efficacy) trials

- Aim to test the mode of action of a treatment
- Test efficacy – does the intervention work in individuals who receive it?
- Have a design with high internal validity and use carefully controlled conditions and restrictive inclusion criteria (to create ‘pure’ sampling)
- Emphasise treatment fidelity and mediating factors

Pragmatic (effectiveness) trials

- Compare the *policy* of delivering one intervention against another in real-world conditions
- Test effectiveness – does the intervention work overall in populations to which it is offered?
- Have a design with high external validity and non-restrictive inclusion criteria
- Place less emphasis on the details of the treatment process or mediating factors

samples with well-targeted and broad outcome measures in order to detect moderate treatment effects in practice (Hotopf, 2002; Harrington *et al*, 2002).

However, the idea is now becoming more accepted that large trials in mental health should be trying to address both pragmatic questions (Does it work? Is it cost effective?) and explanatory ones (How does it work? What components are responsible for efficacy, costs and patient-related outcomes? Can it be tailored to work more effectively or cost-effectively with particular types of patient?). The view is gaining ground that there is no reason why improving both the design and analysis of a trial to answer the explanatory questions of scientific interest should compromise its ability to answer the management-oriented pragmatic one. At its best the complex intervention trial will be a sophisticated clinical experiment designed to test the theories motivating the intervention and also help understand the underlying nature of the clinical problem being treated, in the context of patient- and service-level characteristics. It is important that these trials explicitly consider how and why the treatments work clinically and have their impact on economic outcomes (Kraemer *et al*, 2002; Kazdin & Nock, 2003; Oakley *et al*, 2006). The virtues of explanatory-type designs can be supplemented by prior consultation and co-construction of measurement; the virtues of pragmatic designs by the addition of process measurement (see below).

One outcome or many?

The classic trial discipline is to have one pre-specified primary outcome measure which tests the key hypothesis of the trial (what Bradford Hill called the essential 'precisely framed question'; Hill, 1955). Secondary outcome and intermediate measures do give an opportunity for testing wider aspects of outcome but they are used only sparingly. Although this kind of trial discipline acts against the construction of *post hoc* 'fishing expeditions' in the data, there is increasing questioning whether this kind of rigour is really appropriate for testing complex interventions where the outcomes of relevance are unlikely to be unitary or totally simple and where the intermediate effects are similarly complex. Such ambitions imply that more than just single simple outcome measures may be needed. However, the power of the study has to be adequate to carry these more complex measures.

Patient preference

Anticipated patient resistance to random allocation has led to the use of preference trial designs, which allow patients to opt for a preferred treatment rather

than be randomly allocated. This can result in a cohort study with an RCT imbedded in it (Brewin & Bradley, 1989). Variations on preference trials include Zelen's design (Zelen, 1979). Here, an identified patient group is randomised before consent is sought. Those allocated to treatment as usual never know that they are 'in a trial'. Those allocated to the experimental intervention are approached for consent. Patients who decline to participate are given the standard intervention but analysed under intention to treat as if they had had the experimental intervention. Quite apart from the (significant) ethical issues about undertaking randomisation prior to consent, it is not possible for such trials to be masked. The ethical concerns of not telling patients that they have been randomised can be met by telling each participant, after randomisation, to which group they have been allocated. They can then choose to swap to the other treatment if they wish, but are considered in the original treatment arm for the purposes of the intention-to-treat analysis. Researchers disagree about the value of preference designs. Relatively larger samples are needed to allow for statistical modelling of the outcome, and this may well make the trial impracticable. It may be that, if randomised designs gain more cultural acceptance, the need for these preference variations will disappear.

Supplementing intention-to-treat analysis

Intention-to-treat analysis is typical of pragmatic trial designs. Data on all participants recruited are analysed, whether or not they completed the trial. This is in keeping with the philosophy that the trial tests the effect of the offer of an intervention to a patient group and avoids the potential bias of only studying patients who adhere to the treatment. This analysis can be supplemented by modelling such as the complier average causal effect (CACE) analysis (Angrist *et al*, 1996), which allows for the real-life situation where a proportion of patients switch arms of the trial during the treatment phase. However, such analysis generally needs larger sample sizes.

Using trials to study development

'Hybrid' trial designs have been suggested that combine the virtues of an explanatory randomised intervention trial with a longitudinal developmental study to form a potentially powerful way of investigating the development of disorders over time (Howe *et al*, 2002). In essence, the active intervention is seen as a controlled perturbation of the development of the disorder. In so far as the intervention changes variables thought to be central in the evolution of a disorder, then – by comparing the longitudinal development of each arm using repeated measures – the trial can act as a natural experiment to test

developmental hypotheses. For such a design to work, the active intervention has to be able to make discrete changes in key developmental (mediating) variables as well as affecting target outcomes.

Control groups

Various interesting problems arise in relation to choosing control groups. First, should the control be treatment as usual, no treatment or a contact condition in which the additional therapist time in the active arm is balanced by non-specific additional therapist time in the contact arm? In the study of complex interventions it will be necessary to decide what key variables should be considered when constructing the control group condition. In wholly pragmatic trials, there is a strong argument that the best control condition is treatment as currently practised, since this reflects the practical question at issue: Does the test treatment confer additional benefit over best current practice treatment? (Harrington *et al*, 2002). However, the limitation of such a design is that one cannot be sure whether the treatment effect found is due to the specific properties of the actual intervention or to some other, more non-specific, therapeutic effect. It is for this reason that the study of treatment process variables has become of increasing interest.

Studying the treatment process

What are process variables?

In a clinical trial process, variables may conveniently be considered in two conceptually distinct ways. First, they may be seen as specific 'mediating' mechanisms postulated for a particular treatment. These may be derived from the theory behind the treatment, may be specific to it and may explain the mechanism through which the treatment may have its effect, or they may be discovered in the course of the study.

Second, they may take the form of more general factors that have an impact on the effectiveness of a treatment, for instance the patient's pre-treatment functioning, their relationship with the therapist, their motivation or the therapist's fidelity to the treatment model. These factors are often called 'moderators'.

The terms 'mediation' and 'moderation' have been used in varying ways. An early formulation (Baron & Kenny, 1986) suggested that a mediator directly influences the treatment outcome, whereas a moderator affects the relationship between treatment and outcome. Kraemer *et al* (2002) add clarity and rigour to this definition (Box 2). Here, a moderator must be a baseline or pre-randomisation

Box 2 Mediators and moderators of treatment outcomes

Moderator

A baseline (pre-treatment) characteristic that shows statistically an interactive effect with treatment on outcome

Mediator

An event or change occurring during treatment, altering with treatment and showing statistically a main or interactive effect on outcome

(Adapted from Kraemer *et al*, 2002)

characteristic which can be shown to interact with treatment to affect outcome. A mediator of treatment has to be a change occurring during treatment which is correlated with the specific treatment chosen and has a main or interactive effect on outcome. A moderator must therefore precede the intervention in time, and be independent of an association with treatment. For instance, lack of social support before treatment is not the same as change of social support during treatment. Moderators cannot explain the overall effect of treatment but can indicate individual characteristics or circumstances associated with greater treatment effects. Mediators identify possible mechanisms through which the treatment might achieve its effect. By this strict definition, treatment alliance would not qualify as a moderator, although some baseline social competency in the patient that might be a factor in generating alliance could be a moderator.

Table 2 categorises variables as mediators, moderators or neither on the basis of the stage at which they are measured and their relationships with treatment and outcome.

Why study them?

One of the features of the development of systematic trials in mental health practice is that the intervention itself must be systematically described and the treatment trial conducted through a manualised protocol. Studies of 'treatment process' can represent a systematic approach to aspects of intervention not covered by the manual.

There are a number of reasons for studying these process variables.

- Evidence shows that process variables often account for a large part of the explanation of treatment effects, even in manualised treatments (see below).

Table 2 Identification of a variable as a mediator, a moderator or neither

<i>When measured</i>	<i>Correlation with treatment</i>	<i>Statistical relationship to outcome</i>	<i>Definition (in relation to treatment outcome)</i>
Pre-treatment	No	Interaction with the treatment effect and/or main effect	Moderator
During/after treatment	Yes	Interaction with treatment effect or main effect	Mediator
Pre-/during/after treatment	No	Main effect	Non specific predictor
During/after treatment	Yes	None	Independent outcome of treatment
Pre-/during/after treatment	No	None	Variable is irrelevant to treatment effect

Adapted from Kraemer *et al*, 2002.

- Studying process variables can help the face validity of trials by better reflecting the richness of the experience that clinicians have in a treatment. For instance over 90% of practitioners in one survey (Kazdin *et al*, 1990) cited the 'therapeutic relationship' as the most important determinant of treatment success in psychological therapy.
- Process measures also help the external validity of studies because they reflect what may happen as protocol treatments are translated into ordinary practice (Kazdin & Nock, 2003). Variation in process variables may be particularly salient in clinical practice, whereas it may be minimised in the highly controlled and organised context of efficacy trials. In psychological treatments, for example, one common issue is how much treatment effectiveness is mediated by aspects of the specific treatment protocol and how much by non-specific factors related to the interpersonal treatment alliance or the patient's baseline functioning.
- Identification of treatment moderators will help specify for whom and under what circumstances a given treatment may work (Kraemer *et al*, 2002). This may have the pragmatic value of allowing tailoring of treatments as a trial treatment is introduced into practice. There has been a limited amount of work of this kind in psychological therapies (Project MATCH Research Group, 1997) and, of course, there are many potential complex interacting variables at play (Kraemer *et al*, 2001).
- Identification of process variables can lead to a refining of hypotheses for future studies. For instance, the discovery of a strong moderating variable could lead to its inclusion as a stratification variable for randomisation in a subsequent RCT; and purposive analysis could

be designed to look at a presumed moderator-by-treatment interaction. The discovery of a strong mediator could lead to a restructuring of the treatment protocol to maximise the change in this variable. A test could then be designed for the next trial to see whether the altered treatment protocol would result in more change in this mediator and a larger overall effect size on the desired outcome.

- The robust study of process variables within treatment trials can sometimes be a powerful strategy for advancing basic understanding of the developmental progression of a disorder (Howe *et al*, 2002).

Shortcomings in current research on treatment process

There has been a great deal of study of certain process measures, particularly in psychotherapy research, but methodology has often been weak (Kazdin & Nock, 2003; Hill & Lambert 2004). Typical problems include the following.

- Poor conceptualisation and inaccurate measurement of the process measures to be studied.
- Poor design, which introduces rating biases. A particularly frequent and serious problem is common method or common rater variance, where the same individual is responsible for rating both the outcome and the hypothesised process. This introduces a significant biasing towards associations between process and outcome (Shirk & Karver, 2003).
- Over-interpretation of associations: causality is assumed even though the direction of association is unclear (a 'type 1 error'). A typical example of this within process measurement is the potential effect of symptom change early in

treatment on the process variable being studied. Because symptom change is usually measured at the end of treatment rather than during it, this kind of hidden or latent symptom change can confound process measures.

- Neglect in studying plausible ‘third-factor’ effects: that is, that a third, unknown, factor might explain both the process measure and the outcome measure associated with it. An example of this to be discussed further below is social functioning and alliance.
- The study is not powered to look at process-level effects. The typical power analysis in a trial addresses the primary outcome, and it is a real problem that the increased sample often needed to investigate process may be impracticable. The use of RCTs and otherwise careful measurement and design can mitigate this difficulty to some extent.
- In randomised trials, failure to prioritise the investigation of process. If the treatment process is measured at all, it is usually only in the active treatment arm, which can result in problems of hidden selection effects or confounders. In consequence, treatment process is often subjected to weaker statistical tests than those applied to testing the intervention itself, which leads to process being considered less seriously than other aspects of measurement (Dunn, 2006).

In response to these shortcomings, Kazdin & Nock (2003) propose criteria for the more rigorous establishment of the validity of process variables in a study (Box 3).

Testing for moderating and mediating effects

A number of steps are recommended to test statistically for mediating effects (Box 4). One of the advantages of the RCT design is that it allows a more powerful version of this type of analysis using linear modelling to test differences in process between the treatment group and the control group (Kraemer *et al*, 2002).

An example of a process measure: the therapeutic alliance

The therapeutic alliance (Hougaard, 1994; Green, 2006b) refers to a variety of interactional and relational factors operating between therapist and client in the delivery of treatment. Although therapeutic alliance is traditionally thought of in the context of psychodynamic therapies, there is no reason why it should

Box 3 Rigorous criteria for identifying process variables

- The candidate variable shows plausible face validity including a theoretical basis
- It shows a convincing association with outcome
- This association shows specificity (i.e. other variables are tested that do not show such an association)
- There is a positive test for statistical mediation (see Box 4)
- There is a dose–response gradient, and experimental manipulation of the proposed process variable shows expected effects
- The direction of causality between the process measure and outcome has been tested using repeated measures of proposed process variables and the outcome of interest
- There has been replication in other treatment contexts

(Adapted from Kazdin & Nock, 2003)

be confined to this form of treatment. Nor should its measurement in a trial be taken as implying a psychogenic aetiology of the condition treated. The quality of the therapeutic alliance may be part of an effective psychological treatment for disorders of all – including organic – aetiologies (Green, 2006b).

The importance of alliance relates partly to its face validity – clinicians consistently rate the therapeutic relationship as crucial to outcome (Kazdin *et al*, 1990). But there is also strong empirical evidence that the quality of alliance predicts outcome independent of other factors. Meta-analysis of studies in both adult (Martin *et al*, 2000) and child (Shirk & Carver, 2003) mental health treatment shows a consistent overall correlation of alliance with treatment outcome of about 0.2. More detailed studies within randomised designs have tended to suggest that the quality of alliance is not specific to a particular treatment style and that it is a powerful independent predictor of outcome.

For example, in one randomised trial (Krupnick *et al*, 1996) three interventions – cognitive therapy, interpersonal therapy and pharmacotherapy – along with placebo control were studied in the treatment of adult depression. Therapeutic alliance was measured through structured observations at three time points during the treatment. Results showed that the quality of patient alliance was similar across all arms of the trial and independent of baseline symptoms. Alliance showed a strong independent effect on outcome in all arms ($r=0.46$), explaining 19% of the outcome variance. Controlling for pre-treatment

Box 4 Testing a variable for statistical mediation

- 1 Demonstrate association between treatment and outcome, statistically controlling for the presumed mediator
- 2 Demonstrate association between treatment and proposed mediator
- 3 Demonstrate association between proposed mediator and outcome, statistically controlling for any effect of treatment
- 4 Simultaneously enter treatment and proposed mediator into a regression analysis with the treatment outcome as a dependent variable. If the strength of association between treatment and outcome in this analysis is lower than that in step 1, then there is evidence that the proposed variable mediates the outcome

(Adapted from Baron & Kenny, 1986)

severity, patients with a good therapeutic alliance were 17.2 times more likely to show post-treatment remission of their depression. These effect sizes were larger than those associated with the specifics of each treatment. The therapists' component of the alliance did not show much predictive effect, but this is probably because the structured protocol in the trial had trained the therapists to an extent that there was little variance in therapist effectiveness.

Similarly, the meta-analysis of trials involving children (Shirk & Karver, 2003) found important effects of alliance on outcome, particularly in externalising disorder and when the alliance was measured later in treatment by professionals (who also, however, often rated the outcome in question).

Steps to improve the study of alliance

In line with the suggestions in Box 3, Kazdin & Nock (2003) have made a number of recommendations for better testing of the therapeutic alliance in treatment trials:

- better operationalisation and modelling, to account for current treatment relationships (see also Green, 2006b);
- avoidance of common method confounds by using objective ratings of outcome and observer ratings of alliance; the use of the same rater for alliance and outcome is no longer acceptable;

- testing for candidate third-factor variables that may explain variation in both treatment alliance and treatment outcome, for example pre-treatment social functioning;
- testing for direction of effects between alliance process and outcome using repeated measure designs to measure alliance and symptom change serially though the treatment process;
- testing of the alliance across all arms of the trial.

One study testing for the direction of effects between the alliance process and outcome has been undertaken in adults (Barber *et al*, 2000). This did show some ongoing reciprocal effect of early symptom change on evolving alliance. However, when they controlled for this there was still a remaining overall effect of early alliance on eventual treatment-term outcome.

Conclusions

Although technically demanding and increasing somewhat the burden on participants, the inclusion of sophisticated process measures in randomised trials clearly has the potential to greatly improve the practical benefit flowing from them. Given the intensive resources that it takes to mount such a trial this must be a good thing.

Furthermore, inclusion of process measures immediately increases the face validity and reality of treatment trials for clinicians and other practical consumers of the research. Process measures usually tap the clinical 'feel' of what a study is testing, reducing the sense that an RCT is a rather artificial design.

Enthusiasts who have promoted the values of the RCT within mental health research have long felt that it has particular qualities to illuminate the complex processes involved in mental health interventions. These modern developments in RCT design, including the measurement of process, may make it more likely that clinicians will agree.

Declaration of interest

None.

References

- Angrist, J. D., Imbens, G. W. & Rubin, D. B. (1996) Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, **91**, 444–455.
- Barber, J. P., Connolly, M. B., Crits-Christoph, P., *et al* (2000) Alliance predicts patients' outcome beyond in-treatment change in symptoms. *Journal of Consulting and Clinical Psychology*, **68**, 1027–1032.

- Baron, R. M. & Kenny, D. A. (1986) The moderator/mediator variable distinction in social psychological research: conceptual, strategic and statistical issues. *Journal of Personality and Social Psychology*, **51**, 1173–1182.
- Brewin, C. R. & Bradley, C. (1989) Patient preferences and randomised clinical trials. *BMJ*, **299**, 684–685.
- Department for Education and Skills, Department of Trade and Industry & Her Majesty's Treasury (2004) *Science and Innovation Investment Framework: 2004–2014*. London: The Stationery Office (TSO). http://www.hm-treasury.gov.uk/media/95846/spend04_sciencedoc_1_090704.pdf
- Department of Health (2004) *Research for Patient Benefit Working Party. Final Report*. London: Department of Health. http://www.dh.gov.uk/PolicyAndGuidance/ResearchAndDevelopment/ResearchAndDevelopmentAZ/PrioritiesForResearch/fs/en?CONTENT_ID=4082668&chk=xUzx/B
- Dunn, G. (2006) Psychotherapy for depression. In *Textbook of Clinical Trials* (2nd edn) (eds S. Day, S. Green & D. Machin). London: John Wiley & Sons. In press.
- Fava, M., Deen Evins, A., Dorer, D. J., et al (2003) The problem of the placebo response in clinical trials for psychiatric disorders: culprits, possible remedies and a novel study design approach. *Psychotherapy and Psychosomatics*, **72**, 115–127.
- Green, J. M. (2004) The SSRI debate and the evidence base in child and adolescent psychiatry. *Current Opinion in Psychiatry*, **17**, 233–235.
- Green, J. (2006a) Editorial: Avoiding a spiral of precaution in mental health care. *Advances in Psychiatric Treatment*, **12**, 1–4.
- Green, J. (2006b) The therapeutic alliance – a significant but neglected variable in child mental health treatment studies. *Journal of Child Psychology and Psychiatry*, **47**, 425 (DOI 10.1111/j.1469-7610.2005.01516.x).
- Green, J. M. & Jacobs, B. (eds) (1998) *In-patient Child Psychiatry: Modern Practice Research and the Future*. London: Routledge.
- Green, J. M. & Jones, D. (1998) Unwanted effects of in-patient treatment: anticipation, prevention and repair. In *In-patient Child Psychiatry: Modern Practice Research and the Future* (eds J. M. Green & B. Jacobs), pp. 212–220. London: Routledge.
- Green, J. M., Kroll, I., Imre, D., et al (2001) Health gain and predictors of outcome in in-patient and daypatient child psychiatry treatment. *Journal of the American Academy of Child and Adolescent Psychiatry*, **40**, 325–332.
- Harrington, R. C., Cartwright-Hatton, S. & Stein, A. (2002) Annotation: randomised trials. *Journal of Child Psychology and Psychiatry*, **43**, 695–704.
- Harris, J. (2005) Scientific research is a moral duty. *Journal of Medical Ethics*, **31**, 242–248.
- Harris, J. & Holm, S. (2002) Extended lifespan and the paradox of precaution. *Journal of Medicine and Philosophy*, **27**, 355–369.
- Hill, A. B. (1955) *An Introduction to Medical Statistics* (5th edn). London: Academic Press.
- Hill, C. E. & Lambert, M. (2004) Methodological issues in studying psychotherapy processes and outcomes. In *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change* (ed. M. J. Lambert), pp. 84–135. New York: John Wiley & Sons.
- Hotopf, M. (2002) The pragmatic randomised controlled trial. *Advances in Psychiatric Treatment*, **8**, 326–333.
- Hougaard, E. (1994) The therapeutic alliance: a conceptual analysis. *Scandinavian Journal of Psychology*, **35**, 67–85.
- Howe, G. W., Reiss, D. & Yuh, J. (2002) Can prevention trials test theories of etiology? *Development and Psychopathology*, **14**, 673–694.
- Imrie, D. & Green, J. M. (1998) Research into efficacy and process of in-patient treatment. In *In-patient Child Psychiatry: Modern Practice Research and the Future* (eds J. M. Green & B. Jacobs), pp. 333–339. London: Routledge.
- Jahad, A. (1998) *Randomised Controlled Trials*. London: BMJ Books.
- Johnson, T. (1992) Statistical methods and clinical trials. In *Research Methods in Psychiatry: A Beginner's Guide* (2nd edn) (eds C. Freeman & P. Tyrer), pp. 24–61. London: Gaskell.
- Kazdin, A. E. & Nock, M. K. (2003) Delineating mechanisms of change in child and adolescent therapy: methodological issues and research recommendations. *Journal of Child Psychology and Psychiatry*, **44**, 1116–1129.
- Kazdin, A. E., Siegal, T. C. & Bass, D. (1990) Drawing upon clinical practice to inform research in child and adolescent psychotherapy: a survey of practitioners. *Professional Psychology: Research and Practice*, **21**, 189–198.
- Kraemer, H. C., Wilson, G. T., Fairburn, C. G., et al (2002) Mediators and moderators of treatment effects in randomized clinical trials. *Archives of General Psychiatry*, **59**, 877–883.
- Kraemer, H. C., Stice, E., Kazdin, A., et al (2001) How do risk factors work together? Mediators, moderators, and independent, overlapping, and proxy risk factors. *American Journal of Psychiatry*, **158**, 848–856.
- Kroll, L. & Green, J. M. (1997) Therapeutic alliance in in-patient child psychiatry. Development and initial validation of the Family Engagement Questionnaire. *Clinical Child Psychology and Psychiatry*, **2**, 431–447.
- Krupnick, J. L., Sotsky, S. M., Simmens, S., et al (1996) The role of the therapeutic alliance in psychotherapy and pharmacotherapy outcome: findings in the National Institute of Mental Health Treatment of Depression Collaborative Research Programme. *Journal of Consulting and Clinical Psychology*, **64**, 532–539.
- Martin, D. J., Garske, J. P. & Davies, M. K. (2000) Relation of therapeutic alliance with outcome and other variables: a meta-analytic review. *Journal of Consulting and Clinical Psychology*, **68**, 438–450.
- Medical Research Council (2000) *A Framework for the Development and Evaluation of RCTs for Complex Interventions to Improve Health*. London: MRC. http://www.mrc.ac.uk/pdf-mrc_cpr.pdf
- Medical Research Council (2003) *Clinical Trials for Tomorrow. An MRC Review of Randomised Controlled Trials*. London: MRC. <http://www.mrc.ac.uk/pdf-clinical-trials-for-tomorrow.pdf>
- Oakley, A., Strange, V., Bonell, C., et al (2006) Process evaluation in randomised controlled trials of complex interventions. *BMJ*, **332**, 413–416.
- Project MATCH Research Group (1997) Matching Alcohol and Treatment to Client Heterogeneity: Project MATCH posttreatment drinking outcomes. *Journal of Studies of Alcohol*, **58**, 7–29.
- Shirk, S. R. & Karver, M. (2003) Prediction of treatment outcome from relationship variables in child and adolescent therapy: a meta-analytic review. *Journal of Consulting and Clinical Psychology*, **71**, 452–464.
- Theodosiou, L. & Green, J. (2003) Emerging challenges in using health information from the internet. *Advances in Psychiatric Treatment*, **9**, 387–396.
- Zelen, M. (1979) A new design for randomised clinical trials. *New England Journal of Medicine*, **300**, 1242–1245.

MCQs

1 A moderating variable in a treatment trial:

- refers to the quality of the treatment manual
- is a variable altered by the treatment that affects outcome
- is a variable independent of treatment which alters the effect of treatment on outcome
- is a variable measured after the treatment as an indicator of how the treatment has moderated the local environment
- is a factor that always makes the treatment effect smaller than it would have been otherwise.

2 As regards the therapeutic alliance:

- it has no association with pre-treatment variables
- it has been shown to vary with latent symptom change early in treatment
- it has been shown to have an impact in both psychological and drug treatments
- it is an example of a moderating variable as strictly defined
- its positive effects are most clearly seen in the treatment of internalising disorders in children.

3 As regards testing for statistical mediation:

- a a proposed mediator must not show an independent effect on outcome
- b a mediator reduces the measured effect of treatment on outcome in a regression analysis
- c a mediating variable shows an association with treatment effect
- d it can only be done in a randomised controlled trial
- e it gives us information about how a treatment has its effect.

4 As regards pragmatic trials:

- a it is essential to have modelled all the details of the intervention beforehand
- b the main aim is to tell us how a treatment works
- c they need to have high external validity

- d they are generally less valid than explanatory trials
- e they can have less exclusive referral criteria than explanatory trials.

MCQ answers

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