

MRC

Medical Research Council

A FRAMEWORK FOR DEVELOPMENT AND EVALUATION OF RCTs FOR COMPLEX INTERVENTIONS TO IMPROVE HEALTH

This document is a discussion document drafted by members of the MRC Health Services and Public Health Research Board. It is intended to provide a framework for individuals considering the evaluation of a complex intervention. It does not set out a set of required steps in carrying out trials in this area.

April 2000

INTRODUCTION

Complex interventions in health care, whether therapeutic or preventative, comprise a number of separate elements which seem essential to the proper functioning of the intervention although the “active ingredient” of the intervention that is effective is difficult to specify. If we were to consider a randomised controlled trial of a drug vs. a placebo as being at the simplest end of the spectrum, then we might see a comparison of a stroke unit to traditional care as being at the most complex end of the spectrum. The greater the difficulty in defining precisely what, exactly, are the “active ingredients” of an intervention and how they relate to each other, the greater the likelihood that you are dealing with a complex intervention.

Why is the evaluation of complex interventions difficult?

Even seemingly straightforward interventions have inherent complexities to bedevil well-designed research. To take a hypothetical example, what is a physiotherapist’s contribution to management of knee injury? The package of care to treat a knee injury may be quite straightforward and easily definable - and therefore reproducible: “This series of exercise in this order with this frequency for this long, with the following changes at the following stages”. However, the physiotherapist may have, in addition to the exercises, a psychotherapy role in rebuilding the patient's confidence, a training role teaching their spouse how to help with care or rehabilitation, and potentially significant influence via advice on the future health behaviour of the patient. Each of these elements may be an important contribution to the effectiveness of a physiotherapy intervention. If we now hypothetically consider evaluating a specialist stroke unit, the physiotherapist is one potentially complex contribution in a larger and more complex combination of diverse health professionals’ expertise, medications, organisational arrangements and treatment protocols that constitute the intervention of that unit.

What is the purpose of this document?

In this guidebook, we hope to provide investigators with guidance in recognising the unique challenges which arise in the evaluation of complex interventions, as well as to suggest some strategies for addressing these issues in the development of their own trials. We present a step-wise approach to the evaluation of complex interventions. This step-wise approach does not imply that a literal, sequential series of studies is always necessary and it may be considered more as a framework to help investigators determine what the state of knowledge and uncertainty is regarding a complex package at a given time.

WHAT IS A COMPLEX INTERVENTION?

Complex interventions are built up from a number of components, which may act both independently and inter-dependently. The components usually include behaviours, parameters of behaviours (e.g. frequency, timing), and methods of organising and delivering those behaviours (e.g. type(s) of practitioner, setting and location). It is not easy precisely to define the “active ingredients” of a complex intervention. For example, although research suggests that stroke units work, what, exactly, is a stroke unit? What are the active ingredients that make it work? The physical set-up? The mix of care providers? The skills of the providers? The technologies available? The organisational arrangements?

Health services have to evaluate a wide array of existing and newly proposed complex packages, so that the service can learn what is effective about any given intervention so that it can be more widely applied throughout the service. Some complex interventions are intended as improvements in the form of direct interventions at the level of **individual patient care**; for example a novel form of cognitive behavioural therapy. Other interventions, although ultimately intended to improve patient care, are actually delivered in the form of an **organisational or service modification**; for example, the introduction of a physiotherapist or Parkinson’s disease nurse into primary care services. A third type of complex intervention is further removed again from individual patient care, although ultimately intended just as much to impact there, when an intervention is **targeted on the health professional**; for example, educational interventions in the form of treatment guidelines, protocols or decision-aids. Finally, many complex interventions are delivered **at a population level**; for example, in the form of media-delivered health promotion campaigns.

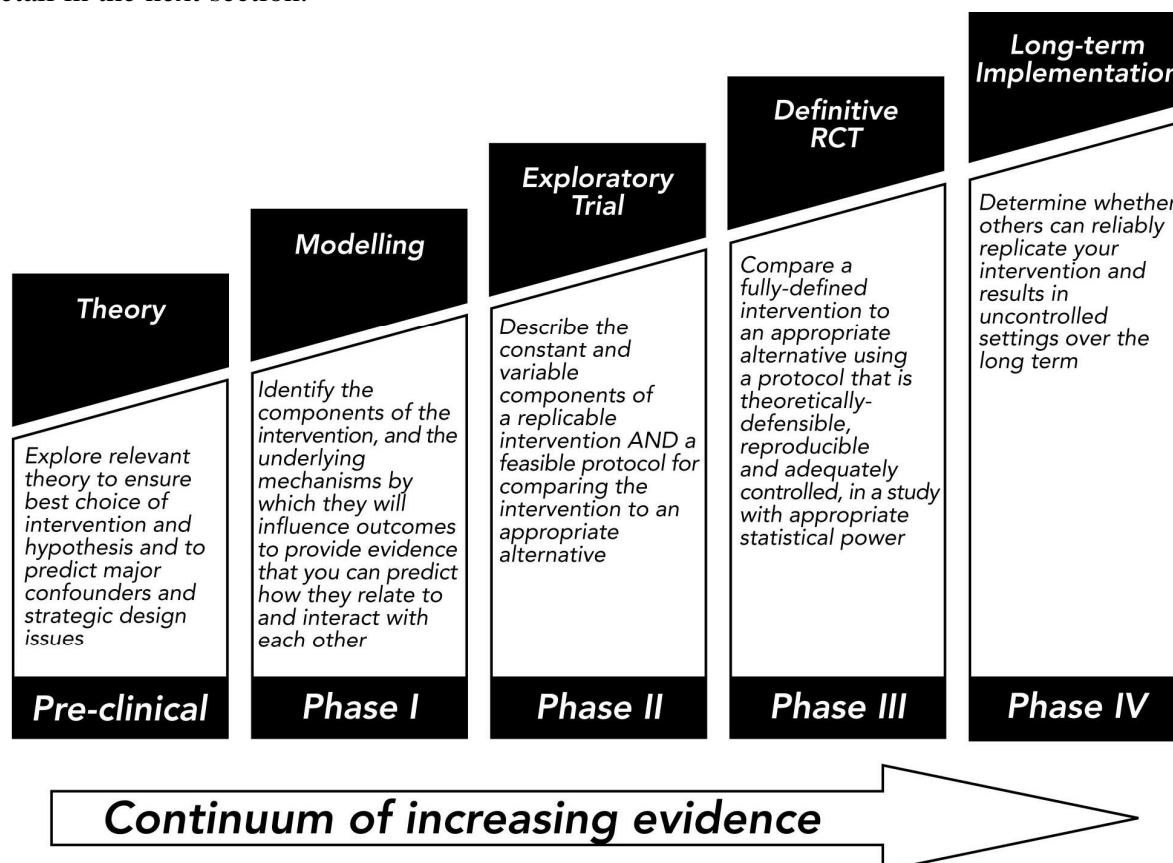
An important question needs to be addressed by investigators at an early stage. Why is it essential to know how (by what specific ingredients) a complex intervention is effective? Indeed if it is clear to those who read the results of a trial how the intervention can be transported and put into operation in other contexts, then it may not be essential to discover the precise mechanisms of action. However, more often, it is only through the researchers in the trial investigating by one of a number of research methods the possible mechanisms that it is possible for inferences to be drawn by others in terms of wider implementation of the study intervention.

It is likely that the randomised controlled trial is the optimal study design to minimise bias and provide the most accurate estimate of a complex interventions benefits. However there are circumstances where such a design is not possible. This document should provide considerations just as relevant to observational studies to evaluate complex packages.

FRAMEWORK FOR TRIALS OF COMPLEX INTERVENTIONS

The main thrust of this document provides a framework with discussion of issues that takes the form of a sequential series of phases of investigation in the evaluation of a complex intervention. There are parallels in the sequence of steps usually required in the evaluation of drugs from initial pre-clinical research through to post-marketing surveillance. However although such parallels and analogies are pursued in this framework, this document in no way intends to imply that the evaluation of a complex package is like the development and evaluation of a new drug. The sequential framework is useful because it sets out the objectives to be met at each stage prior to moving forward to the next stage. However, this framework should not be read as an inflexible “to do list”, but rather as advice to apply *to the extent to which it is relevant* at each stage of your own research. Not all research will need to begin at the beginning and work stepwise through the entire framework; in many cases, much of the preliminary evidence already exists. Depending on the quality of the existing evidence and the nature of the intervention, some researchers may find they devote significant time to certain of the preliminary phases and little to others. Sometimes the process is more iterative; sometimes stages can be addressed simultaneously; sometimes stages are less important.

The phases that we distinguish in this framework are therefore (1) ‘Pre-Clinical’ or theoretical (2) Phase I or modelling (3) Phase II or exploratory trial (4) Phase III or main trial (5) Phase IV or long term surveillance (Figure). The purpose of each phase is briefly described before each phase is considered in more detail in the next section.



(1) 'Pre-Clinical' or theoretical phase

The first step in evaluating a complex intervention is to establish the theoretical basis that suggests that your intervention should have the effect(s) you expect it to. This may be formal theory of individual or organisational behaviour or it may be informal evidence regarding organisational constraints or types of patients' or health professionals' beliefs that may promote or inhibit behavioural change. This phase of assessing theory and evidence may identify in preliminary form the kind of intervention needed and study design. If a particular intervention to be evaluated is already widely practised, a theoretical phase may well not be essential; the health service already has what is considered a clear understanding of the 'mechanisms of action' of an intervention that is nevertheless to be evaluated. It also has to be recognised that, for pragmatic reasons or urgency of answers for public policy, it is not always possible to await clarity from theory prior to evaluation. Moreover in many fields, theories of behaviour of individuals and organisations may be of limited value.

(2) Phase I or modelling

The second step in evaluating a complex intervention is to develop an understanding of your intervention and its possible effects. This involves delineating an intervention's components and how they inter-relate and how active components of a complex package may relate to either surrogate or final outcomes. 'Modelling' refers to the possibility that this phase is paper-based, for example, computer simulations, or economic modeling. It may also include qualitative testing through focus groups, preliminary surveys, case studies, or small observational studies.

(3) Phase II or Exploratory trial

The third step in evaluating complex interventions is the crucial stage prior to a main RCT. In Phase II, all the evidence gathered thus far is put to the test. In Phase II, it may be appropriate to experiment with your intervention, varying different components to see what effect each has on the intervention as a whole. Your ability to fully control the intervention in different settings can be established. Evidence can be obtained to support the theoretically expected treatment effect, to identify an appropriate control group, outcome measures, estimates of recruitment for a main trial, and other requirements of such a trial. This phase could permit testing of alternative forms ('doses') of an intervention. A key and distinctive possibility of studies at this point could therefore be the adaptive nature of interventions, study design, analyses, (being modified over time in the course of the study phase) although there may also be circumstances where phase two trials need to be precise and consistent about the intervention to inform decisions for a main trial.

(4) Phase III or main trial

The fourth and central step in the evaluation of a complex intervention is the main randomised controlled trial to evaluate a complex intervention and requires attention to standard issues of adequate power, adequate randomisation and blinding (where feasible), appropriate outcomes measures, informed consent of participants and other standard features of well designed trials.

(5) Phase IV or long term surveillance

The final step in the evaluation of a complex intervention is a separate study to establish the long-term and real-life effectiveness of the intervention. The broader applicability of an intervention outside of a research context may be tested and rare or long term adverse events identified. This stage is likely to involve an observational study.

WHAT ARE THE ISSUES TO CONSIDER AT EACH STAGE OF THIS FRAMEWORK?

The remainder of this guide considers further the methodological issues that may arise in relation to the distinct phases of work that have been distinguished above. We consider key methodological issues under three headings: (i) theoretical and modelling phases, (ii) the exploratory trial phase and (iii) the main trial phase.

METHODOLOGICAL ISSUES AT THE THEORETICAL AND MODELLING PHASES

Considerations that may arise at the theoretical and modelling phases are here considered together because they are often inter-related. As a preface it is assumed that investigators will need to have consulted data bases such as the Cochrane Library and NHS Centre for Reviews and Dissemination and carried out other literature searches of the proposed area of investigation.

We have urged investigators to expect and incorporate a theoretical phase to the development of their evaluation of a complex intervention. In this way investigators force themselves to consider underlying assumptions being made, whether at the physiological, psychological, organisational or whatever level regarding postulated mechanisms and processes in the intervention being examined. The use of the term 'theoretical' is intended to refer to formal bodies of evidence where they exist, but, as acknowledged earlier, for many areas fields may not have been developed to a stage of providing clearly delineated bodies of evidence, 'theory' may be too grand a term for the body of evidence or the field may be fluid and the choice of relevant theory very unclear for any given research problem. In some cases, the relevant body of evidence may be less formal 'theory' and more accumulating wisdom from empirical evidence.

This phase may seem rather abstract and remote. However consulting the relevant literature may highlight important issues that revise or clarify the complex intervention to be studied. A trial of behaviour change amongst patients may rely less on information-giving strategies alone if the relevant behavioural scientific literature is first consulted. A trial involving nurses providing an intervention in primary care may benefit from consulting the literature on inter-professional relations. An evaluation of peer-led health behaviour change in adolescents may benefit from consulting current evidence of the nature of peer influences. A trial to change doctors' prescribing behaviour in a particular field may be refined by considering research on factors influencing prescribing decisions. In each case, consulting available evidence may help rule out certain kinds of intervention are implausible or help tease out the likely facilitating factors or barriers in developing a specific intervention.

Complex interventions tend to be marked by a paucity of high quality literature, and thus it is not unlikely that some level of original research will have to be undertaken in the early phases of evaluation of a complex intervention. Theory and modelling were presented as two separate phases, when in fact, refining theory through modelling activities is liable to be iterative.

Modelling is concerned with unravelling and distinguishing the key components in a complex intervention. A variety of methods may be used at this stage from purely paper and pencil informal modelling through more formal simulation and computer modelling through to primary data gathering via structured survey or qualitative interviews, focus groups or field work. The most challenging part of evaluating a complex intervention - and the most frequent weakness in such trials - is defining the actual intervention, that is, standardising the content and the delivery of the intervention by determining the critical components of the intervention, and how they relate to, and impact on, each other. It is essential to clarify as far as possible the important components, partly in order to devise protocols for the trial itself and partly so that readers of eventual trial results can infer from results what elements were essential and what secondary or unimportant in producing a beneficial effect.

Examples of such problems may be found in diverse fields.

- When you undertake cognitive behavioural therapy, what exactly is or are the “active ingredient(s)”? Does success depend on the personality of the therapist? The personality, health status, social status, or other characteristic of the patient? The content of the therapy? The way it is delivered? The frequency of contact? The location of contact? The duration and the timing? What other components count?
- How may researchers say what are the critical components of the successful stroke unit, so others can use published trial results as the basis to establish successful stroke units in other hospitals. Was success due to the mix of people, personalities, or expertise? The mix of skills? The location? The technologies available? Organisational culture and leadership?
- Increasingly we seek evidence of cost-effective and acceptable alternatives to in-patient hospital care, such as, for example, the community hospital or hospital at home. There are a diversity of such alternative schemes so that research evidence may need to be as explicit as possible about whether non-hospital schemes are successful because of patient selection, ready accessibility of particular technologies and professional expertise or distinctive cultural and organisational arrangements.

Informal paper modelling and simulation

A paper model of the intervention, using diagrams or flowcharts, may be a useful starting point to identifying the key relations among components, and potentially some of the vulnerabilities of stabilising the intervention. Diagramming the components (nature, timing, frequency, duration, of inputs, including skills, occupational mixes, organisational arrangements, setting) may highlight potential weak points and bottlenecks. The intervention may prove more elusive when an attempt is made to map it out in this way.

Using the diagram of inter-dependencies and interrelationships among the intervention components, potential strategies to control each of the components can be considered. Which components might be standardised or controlled? What challenges will these relationships pose to control? Which components should be varied? Can they be varied systematically?

With a developing picture of the intervention on hand, consideration should be given to the most appropriate intervention to which to compare the experimental intervention. It is not unlikely that the intervention is being compared to “standard practice” or a similar intervention; in this case, the comparison intervention is quite possibly just as complex as the test intervention, and should be diagrammed out in the same way. Considering the intervention as a whole and constituent components inevitably will prompt planning of strategies for randomisation and blinding, recruitment and continued participation, outcomes measures, and analysis.

Simulations are growing in popularity with the increasing use of RCT's to evaluate complex interventions. While paper-based diagramming is useful, it will often raise more questions than it can answer about the complex interplay between the many components of the intervention. Simulations can help explore questions such as “If I change *this* factor, what happens to all *those* factors?”. Simulations could include computer modelling, scenario building and testing, as well as mathematical and statistical modelling, and can be used both to test assumptions and provide important information regarding trial design needs. Simulation may be a useful way of incorporating thinking about relationships between intervention and surrogate or final outcomes.

Using qualitative research

Once theory and preliminary research has exhausted itself, it is quite likely that there will still be unanswered questions that may be addressed by some primary research that needs to be done before exploratory or main trials.

Consider an evaluation of a different organisation of physicians and nurses to deliver a certain type of care: success of the intervention will depend as much on inter-professional behaviours between these two groups as upon the more predictable logistics of who will do what. If you are unable to find satisfactory evidence to justify your expectations of these behaviours, it would be extremely valuable to undertake qualitative research into the determinants of collaborative behaviour between doctors and nurses, prior to finalising the definition of the intervention. Qualitative research is important to understanding why something happens: why does the intervention have therapeutically beneficial aspects? Qualitative research may be valuable in testing the underlying assumptions in relation to an intervention, to ascertain whether inappropriate beliefs and assumptions may be leading to the wrong choice of intervention or hypothesis.

Qualitative research is also helpful for identifying which are the “active ingredients” of the complex intervention, and which elements are not related to the treatment effect. Qualitative studies may be used to determine which groups of participants are most likely to respond positively to the intervention, whether the intervention must be modified in different ways for different groups, or simply not used for certain types of people. Careful pre-RCT research can mitigate against results that show no effect, which are actually a mixture of positive respondents balanced by negative respondents who are actually an identifiable group for whom this intervention simply shouldn't have been used.

A range of methods are available with increasingly well understood rationales, methods and forms of analysis: group interviews (focus groups), individual in-depth interviews, observational (ethnographic) research or organisational case studies. This phase of work need not rely on qualitative evidence; a quantitative survey of practitioners may also identify beliefs, behaviours or organisational factors that make an intervention more or less plausible.

METHODOLOGICAL ISSUES IN RELATION TO THE EXPLORATORY TRIAL PHASE

The exploratory trial can be used to provide vital information. It can be used to consider variants of the intervention and their possible effects on outcomes in order as clearly as possible to define the intervention in a main trial. It can also be used as a test of feasibility of key components of a larger main trial, such as recruitment, randomisation, measurement of outcome and so on. Additionally it may provide unique evidence of intervention effects for the purposes of calculating the power of a main larger trial.

Defining the intervention

While it is usually unacceptable to modify an intervention during the course of an RCT, an exploratory or pilot trial can be developed to explicitly address these goals: variations of the intervention can be tested, to see which seem to be the most appropriate for a full scale trial.

Often, quantitative questions regarding certain components of a complex intervention, such as ‘how much’ or ‘how often’, remain unanswered by the theoretical and modelling phases. The only way to obtain conclusive evidence regarding the utility, practicality or desirability of one approach over the other is to compare two or more versions of the intervention (for example, frequencies, intensities or modes of delivery of a therapy) directly in an exploratory trial. It may be possible to adapt the content of a complex intervention at this exploratory phase if there is real concern about the need to identify the optimal content of an intervention. However statistical analysis of an evolving intervention obviously is problematic and this kind of adaptive evolution of a complex intervention runs counter to other requirements investigators may have of the exploratory trial, such as calculating power for a main trial or estimating the extent to which an intervention can be standardised.

An exploratory trial provides the opportunity to determine whether you can in fact ensure that the intervention is delivered in a standardised fashion by all providers involved in the trial; it may take some practice to determine how and to what extent variables truly can be controlled. Trials with complex interventions will tend to require much more preparation and training of the practitioners involved than conventional trials, to ensure that all consistently provide as close to the same intervention as possible.

Within the trial, it could be either the *method* of delivery or the *content* of delivery that is under investigation; one or other of these may be allowed to vary by individual preferences. In such cases, acceptable limits within which practitioners can individualise interventions, either with respect to the way they deliver it or with respect to tailoring the intervention contents to the participant, must be established and checked for validity. Individualisation may be acceptable, for example, in many service or practitioner-based interventions. These issues will have to be considered in the exploratory phases of the study, and parameters for acceptable deviation established.

Another factor to consider is whether there is evidence of a learning curve effect: do the intervention providers improve over the course of the trial? Alternatively, is there evidence that variations in levels of skills across providers may affect delivery of the intervention and/ or outcomes? What kind of training will be required in a full-scale RCT? How can this be provided? Who will pay for it?

The exploratory trial offers opportunities both to seek learning curve effects, and to prepare for a full-scale

trial by using the Phase II as a training ground for future RCT participants. Additionally or alternatively, methods to adapt analysis to control for variations in skills providers where these have significant impact, may be developed. Data developed with respect to skills variations, training and learning curves will be of great value to practitioners and policy makers reviewing and potentially implementing your ultimate RCT recommendations.

Monitoring the delivery of the intervention

Once it has been determined how the intervention is to be delivered, it may be necessary to establish fairly comprehensive monitoring to ensure that they are being maintained. Interventions normally evolve over time, as providers become more experienced and individualise the intervention to meet their own styles and perceived participants needs. However, an RCT is not a "normal" treatment situation: once the RCT has begun, the intervention must not evolve, as the RCT results will be unusable if later participants experience a different intervention than earlier ones. Fidelity to treatment protocols or quality standards may be essential to effectiveness. Monitoring will be essential to ensuring that individual styles and evolution in treatment do not render results from different groups of participants incomparable or excessively reduce the effectiveness of the intervention.

A variety of methods are available. Audiotaping or videotaping of consultations can be used. Traditional supervision, 'booster' training and feedback may serve a similar purpose. Random checks are possible if acceptable to participating health professionals. Assessment by an independent reviewer can help determine where an intervention might be straying too far from the norm. At the same time, during the exploratory phase, it can help the researcher identify the ways in which personal interactions seem to be affecting outcomes, and determine which kinds of people would be best included in the RCT as deliverers of the intervention.

This phase is also an opportunity for determining the comparative arm of a main trial; should it be an alternative package, standard care, placebo. Choice of comparison will influence apparent effectiveness of study interventions, but also practicalities such as chances of recruiting clinicians and patients. There are secondary choices with similar consequences: what form of standard care ('usual care' or following a protocol)? This phase can also inform decisions as to how much of the control arm needs to be monitored or measured for analysis in a main trial. For example patients receiving an active placebo receive attention, support and other secondary benefits of care that may enhance the outcomes of this form of control group, compared with one comprising being on a waiting list for care.

Control groups in the form of placebo treatment can create a number of additional dilemmas, if there are ways in which patients' assignment can be guessed. The use of a control group in psychosocial and behavioural research can produce "resentful demoralisation" among the control group triggered by a perception of differential benefit for the intervention group among participants or providers. In addition, members of the control group may seek out treatment similar to the active intervention on their own, diminishing the comparability of the two trial arms.

Other purposes of the exploratory trial

The exploratory trial may be a vital source of estimates for the sample required for a main trial, an essential step since, as is frequently lamented, all too many trials prove to be under-powered. As stated earlier, the exploratory trial also provides the opportunity to "test-drive" the assumptions and strategies which will form the Phase III design; methods of recruitment, randomisation, follow-up can all be examined at this stage. Many complex interventions require a long term commitment from participants, and often necessitate significant effort from the participant to undertake major changes in their health-related behaviour: examples include long-term rehabilitation, healthy eating, or changes in health care delivery. As a result, an intervention can fail either because compliance cannot be maintained, or because compliance cannot be proven because of reduced compliance with study assessments. Either way, though the intervention may be sound, it will not be possible to establish its effectiveness using an RCT. Thus, it is important that Phase II provide adequate evidence that measurement of longer term behavioural outcomes is feasible. Innovative strategies may be needed to ensure the long-term success of the trial, and the minimisation of the dropout rate, from the development of a logo to increase identification with the project, through feedback or seasonal greetings cards to participants.

The loss of participants over the course of a trial must be carefully monitored, and minimised to the greatest extent possible because participants who drop out tend to differ in very significant ways from participants who do not. This may be particularly problematic with behavioural interventions because the same individuals who poorly adhere to behavioural recommendations may also be poor study respondents, resulting in biased assessment of study intervention.

METHODOLOGICAL ISSUES IN A MAIN TRIAL

To design and carry out a main trial to evaluate a complex intervention investigators have to make final decisions about the nature of the intervention that is to be evaluated and then address standard design issues for trials to address potential problems of bias that may reduce either the internal validity of the trial or limit its external validity (ie generalisability).

Determining the intervention

Final decisions have to be made about the content of the intervention that will be evaluated in a main trial. In one direction is the choice to tightly standardise the package so that it maximally contains the ‘active ingredients’ that may produce beneficial effects and is most easy for others to interpret when results are produced. The obvious costs of such a strategy are that it may require substantial effort and resource to produce a tightly standardised intervention and, more importantly, it is not necessarily very easy to generalise from trial results to the wider every day world of the health service, where such standardisation of treatment may not be feasible. The alternative choice is to design a very pragmatic trial where the intervention varies substantially in content from centre to centre or clinician to clinician on dimensions already discussed in this guide such as frequency, intensity and such parameters. A more flexible approach to the intervention may maximise health professionals’ ability to participate but also the ultimate ‘real-world’ generalisability of results. Hopefully work from previous phases will aid in the choice on intervention within this range of options. A key consideration will be that ultimately interventions have to be feasible within the health service. Pragmatically it will be essential to distinguish and describe what characterises the intervention, so that others can draw appropriate lessons from the trial for the health service.

In the same way, the extent to which fidelity to a particular model or protocol is monitored also has to be decided, with similar implications for resource and generalisability. It should be possible to create strategies of monitoring fidelity that are not excessively resource-intensive, for example by random monitoring. Such a strategy may also make it possible in principle to analyse the impact of additional monitoring such as reinforcing benefits of study intervention because of health professionals’ more conscientious adherence to protocol or non-specific therapeutic benefits to the patient of additional attention.

General design issues

Many research design issues for complex interventions are standard for all forms of RCT; they are intended to address either issues of internal validity (to what extent are differences between study and control intervention real rather than a product of bias?) or external validity (to what extent are trial results generalisable

to a wider population?). A number of key methodological issues need attention to address these two problems.

Randomisation

Randomisation, properly undertaken, should eliminate selection bias, but there are many ways randomisation can be unconsciously undermined. Although the results of allocation can often not be blinded in a complex intervention trial, it is critical that the process to generate the participant's allocation is fully concealed, and that the participant's involvement in the trial is fully confirmed before the allocation is made. Any predictable approach to random allocation (day of the week, participant's birthday or other identification number) is open to subversion.

Randomisation depends on the statistical power of chance to evenly distribute potential confounding variables among the trial arms. Where confounders are minimal and participant numbers high, this strategy should be adequate. However, it must be remembered that it is always possible for a confounding factor, especially if it is uncommon but significant, to be more prevalent in one arm over another. This is why it is critical to identify potential confounders prior to trial design, even though randomisation will be used.

Claims that randomisation has been adequate to ensure the trial arms are equivalent will be strengthened by the provision of adequate baseline data which demonstrates that there are not meaningful differences between your arms or clusters in terms of health status and other potential confounders.

If there are a number of known confounding factors which could influence trial outcomes, and concerns that randomisation may not be adequate to keep all factors balanced in each arm of the trial, statistical and design options can be considered, such as **minimisation** or **block design**. These techniques allocate each new participant to the appropriate arm, or to a block to be randomised as a single entity once complete, based on balancing those factors determined in advanced by the trialists as potentially having a significant effect on outcomes, such as age, previous history, other health factors, etc.

For complex interventions in particular it may not be feasible to randomise at the level of individual patient and groups of patients, by doctor or by practice, for example, are randomised instead. This approach, known as **cluster randomisation**, could be used, for example, when comparing alternative strategies (guidelines versus conventional educational methods) to influence general practitioner's management of a particular problem.

There are problems with the sample size required for cluster randomisation to deal with the likelihood that patients within a practice are more likely to be similar than different; sample sizes have to increase to

address such clustering. With clusters, it is particularly important to look very carefully for any possible bias resulting from the nature or location of the practitioner/ intervention attracting certain types of participant or patient. This common challenge results in incomparable trial arms, and must be properly compensated for in the analysis. The investigator must be able to demonstrate that participants in different settings are similar, without underlying fundamental differences that would affect their response to their intervention and the outcomes of the trial.

A quite different trial design has been developed to examine the impact of interventions to implement professional behaviour change. The balanced incomplete block design randomises doctors or practices to receive different interventions in a way that targets different behaviours in different health professionals so that non-targeted behaviours provide controls. For example doctors in one group may receive an intervention such as guidelines for one health problem, say arthritis, a second group receives guidelines on asthma. The two groups of doctors constitute controls for those health problems for which they have not received guidelines. This basic version can be made into more complex factorial designs.

Blinding and preferences

The purpose of blinding is to eliminate bias resulting from the expectations of the patient or the provider regarding outcomes. The influence of patient and practitioner preferences has been frequently raised as a confounder of true randomisation. If either the patient or the provider has a treatment preference, and believe they have received their preferred intervention, they are likely to expect - and therefore achieve - more positive outcomes; where relevant, compliance is likely to be improved. Similarly, if either the patient or provider becomes convinced they did not receive the intervention of choice, outcomes may suffer commensurably.

The bias arising from patient preferences becomes especially difficult with complex interventions for a number of reasons. Participant involvement is often more active (i.e. goes well beyond passively accepting drug A or drug B), and bias increases with the increasing importance of participant behaviour on intervention success. Interventions are frequently already used in practice or are variations on known practices, and therefore people have pre-formed opinions about them. The influence of patient preference becomes even more challenging when trialists cannot blind patients and/ or practitioners, a common characteristic of complex interventions. It is not uncommon with complex interventions that one trial arm/ intervention might simply be much more convenient for the participant, involving anything from a more accessible location (community vs. hospital care) to a less invasive treatment (counselling vs. drugs, or drugs vs. surgery).

Patient preferences may result in bias if blinding is unsuccessful and also limit generalisability of results if strong preferences reduce participation in a trial. A proposed solution is that of a so called ‘preference trial’ in which those patients with no preferences are randomised as usual but those with preferences and refusing randomisation receive their preferred treatment. Analyses of such data are more complex as some account needs to be taken of there being two kinds of evidence within the trial, some randomised and some observational with potential confounding requiring adjustment.

Generalisability

Investigators need to determine how generalisable they intend the results of their proposed trial to be. This is partly set by inclusion and exclusion criteria. If they are very narrowly defined this may limit the population to which results relate. More pragmatic criteria such as inclusion through patients’ and clinicians’ uncertainty may enhance range of recruitment and expand generalisability, but leave it unclear to whom results exactly apply. Exclusion criteria such as age need to be particularly justified; for example on grounds of differential efficacy. Informal decisions about whom to invite into trials can result in substantial differences in demographic, health status and other differences between trial participants and the general population to whom results are intended to apply.

A related concern about generalisability arises when trials recruit a very low proportion of eligible patients. Recruitment into trials quite frequently falls below what was projected. There are a number of generic measures, including working with experienced trials co-ordinators, which will significantly help to ensure adequate recruitment levels. With complex interventions, recruitment can become much more challenging, especially when attempting to allocate participants to non-traditional treatment modes, such as the follow-up of breast cancer patients in primary care rather than by specialists.

Recruitment into a trial requires informed consent. However some forms of evaluation of complex interventions challenge investigators’ ability to define who may give such consent and how it may be obtained. It is far more difficult to obtain appropriate consent from participants to a population-level intervention, for example randomising a health promotion intervention between areas. However the ethical imperative still requires that the issue be addressed.

Choice of outcomes measures

Selection of outcome measures most appropriate to the evaluation of a specific complex intervention is clearly essential. For many such studies an early decision may be whether or not the outcome measure actually needs to reflect health status. For example, with studies of implementation, it may be sufficient to demonstrate that a trial brought about appropriate change in professional behaviour, provided there is clear

prior evidence that such behaviours, for example, prescribing particular treatments, are effective. However it is less clear cut in other instances. For example, there is sufficient evidence of the weak relationships between knowledge, attitudes and behaviour that it may be considered insufficient to evaluate a health promotion intervention solely by reliance on changes in knowledge as the outcome measure.

It may also be important to decide whose health is concerned. Normally it will be that of the patient but there are many health problems where an important secondary objective is to relieve distress and improve health-related quality of life of carers.

Health is multi-dimensional and several aspects may appear relevant to a given complex intervention. Nevertheless, because of errors that may arise from multiple statistical testing, and also because of problems of burden to respondents and study costs, it is essential that investigators make an explicit and strategically considered choice of outcome measures.

Increasingly it will be essential to include outcomes as assessed directly by the respondent, particularly in areas such as pain, disability, social function (areas encompassed by the term 'health-related quality of life'). Again, such measures need to be carefully selected as appropriate to the question being asked, as well as being validated. Only where there is evidence to think that respondents are sufficiently cognitively impaired is it likely that proxy respondents need to be the primary source of health-related quality of life judgements. Differential response rates by health status can be a problem for measuring these dimensions of outcome.

It was argued that for a variety of practical reasons it may be useful or necessary to use surrogate end points in the exploratory trial. However in a main trial where the primary purpose is to examine impact on health of a complex intervention, it is less defensible to measure surrogates of health status, and indeed the practice may lead to quite erroneous results.

Consideration also has to be given to assessment of outcomes not targeted by an intervention, adverse and unanticipated outcomes. Negative consequences have to be weighed against possible beneficial effects in targeted outcomes.

The issue of costs in relation to benefits also needs to be explicitly addressed in the design of any main trial intended to inform decisions for the health service. As with outcomes in terms of health status, the scale of data to assess a full range of costs to society (community as well as hospital, familial as well as personal) can challenge the feasibility of a trial so that careful strategic choice of cost-generating events requiring measurement within a trial is needed. Investigators need to consider whether costs within a trial will generalise to the health service in choosing what to measure. Costs of monitoring patients may be unusually

high in a research setting. Costs of newly established technology may rapidly change. Much evidence regarding costs may be gathered from other sources out-with the trial.

POST-MARKETING SURVEILLANCE'

Finally it is essential for investigators to think about the dissemination of their trial results. A related but distinct issue for investigators also, although not solely their responsibility, the question of whether effectiveness of an intervention is actually sustained when rolled out into wider practice.

With regard to dissemination, a prior consideration is appropriate analysis and reporting of trial results. Appropriate analysis includes adhering to a prior protocol with regards to limited sub-group analysis to avoid 'data-dredging'. Beyond that, investigators will want to take note of and follow as far as possible standards that are increasingly being disseminated for appropriate reporting of trials.

There are interesting and neglected scientific questions with regard to assessment of wider implementation of an intervention and effectiveness in its use in a broader non-experimental context. They are more likely to require observational designs that reflect regular practice. It is of course not necessarily the main trial investigators or funding sponsors who take up this next phase of investigating broader effectiveness of complex interventions.

APPENDIX

The following individuals participated in a MRC workshop to discuss issues relating to the evaluation of complex interventions:

Nicholas Black, John Bond, Senga Bond, Mary Boulton, Janet Derbyshire, Ray Fitzpatrick, Sheila Gore, Adrian Grant, Jeremy Grimshaw, Andy Haines, Martin Jarvis, Michael King, Ann-Louise Kinmonth, Peter Langhorne, Richard Lilford, David Mant, John Nicholl, Andrew Nunn, James Orford, Mark Petticrew, Martin Roland, Ros Rouse, Ian Russell, Peter Sandercock, Jan Scott, Sasha Shepperd, Bonnie Sibbald, David Spiegelhalter, Ala Szczepura, Simon Thompson, Graham Thornicroft, Peter Tyrer, Derick Wade