Designing a research project: randomised controlled trials and their principles

J M Kendall

....

The sixth paper in this series discusses the design and principles of randomised controlled trials.

....

The randomised control trial (RCT) is a trial in which subjects are randomly assigned to one of two groups: one (the experimental group) receiving the intervention that is being tested, and the other (the comparison group or control) receiving an alternative (conventional) treatment (fig 1). The two groups are then followed up to see if there are any differences between them in outcome. The results and subsequent analysis of the trial are used to assess the effectiveness of the intervention, which is the extent to which a treatment, procedure, or service does patients more good than harm. RCTs are the most stringent way of determining whether a cause-effect relation exists between the intervention and the outcome.[1]

This paper discusses various key features of RCT design, with particular emphasis on the validity of findings. There are many potential errors associated with health services research, but the main ones to be considered are bias, confounding, and chance.[2]

Bias is the deviation of results from the truth, due to systematic error in the research methodology. Bias occurs in two main forms: (a) *selection bias*, which occurs when the two groups being studied differ systematically in some way, and (b) *observer/information bias*, which occurs when there are systematic differences in the way information is being collected for the groups being studied.

A *confounding* factor is some aspect of a subject that is associated both with the outcome of interest and with the intervention of interest. For example, if older people are less likely to receive a new treatment, and are also more likely for unrelated reasons to experience the outcome of interest, (for example, admission to hospital), then any observed relation between the intervention and the likelihood of experiencing the outcome would be confounded by age.

Chance is a random error appearing to cause an association between an intervention and an outcome. The most important design strategy to minimise random error is to have a large sample size.

These errors have an important impact on the interpretation and generalisability of the results of a research project. The beauty of a well planned RCT is that these errors can all be effectively reduced or designed out (see box 1). The appropriate design strategies will be discussed below.

GETTING STARTED: DEVELOPING A PROTOCOL FROM THE INITIAL HYPOTHESIS

Analytical studies need a hypothesis that specifies an anticipated association between predictor and outcome variables (or no association, as in a *null hypothesis*), so that statistical tests of significance can be performed.[3] Good hypotheses are specific and formulated in advance of commencement (a priori) of the study. Having chosen a subject to research and specifically a hypothesis to be tested,

Correspondence to:
Dr J M Kendall, North Bristol NHS Trust, Frenchay Hospital, Frenchay Park road, Bristol BS16 1LE, UK; frenchayed@cableinet.co.uk

Figure 1 The randomised control trial.

Box 1 Features of a well designed RCT

- The sample to be studied will be appropriate to the hypothesis being tested so that any results are appropriately generalisable. The study will recruit sufficient patients to allow it to have a high probability of detecting a clinicaly important difference between treatments if a difference truly exists.
- There will be effective (concealed) randomisation of the subjects to the intervention/control groups (to eliminate selection bias and minimise confounding variables).
- Both groups will be treated identically in all respects except for the intervention being tested and to this end patients and investigators will ideally be blinded to which group an individual is assigned.
- The investigator assessing outcome will be blinded to treatment allocation.
- Patients are analysed within the group to which they were allocated, irrespective of whether they experienced the intended intervention (intention to treat analysis).
- Analysis focuses on testing the research question that initialy led to the trial (that is, according to the a priori hypothesis being tested), rather than "trawling" to find a significant difference.
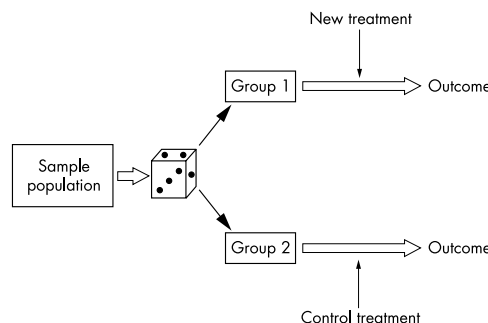
preparation should be thorough and is best documented in the form of a protocol that will outline the proposed methodology. This will start with a statement of the hypothesis to be tested, for example: "...that drug A is more efficacious in reducing the diastolic blood pressure than drug B in patients with moderate essential hypertension." An appropriate rationale for the study will follow with a relevant literature review, which is focused on any existing evidence relating to the condition or interventions to be studied.

The subject to be addressed should be of clinical, social, or economic significance to afford relevance to the study, and the hypothesis to be evaluated must contain outcomes that can be accurately measured. The subsequent study design (population sampling, randomisation, applying the intervention, outcome measures, analysis, etc) will need to be defined to permit a true evaluation of the hypothesis being tested. In practice, this will be the best compromise between what is ideal and what is practical.

Writing a thorough and comprehensive protocol in the planning stage of the research project is essential. Peer review of a written protocol allows others to criticise the methodology constructively at a stage when appropriate modification is possible. Seeking advice from experienced researchers, particularly involving a local research and development support unit, or some other similar advisory centre, can be very beneficial. It is far better to identify and correct errors in the protocol at the design phase than to try to adjust for them in the analysis phase. Manuscripts rarely get rejected for publication because of inappropriate analysis, which is remediable, but rather because of design flaws.

There are several steps in performing an RCT, all of which need to be considered while developing a protocol. The first is to choose an appropriate (representative) sample of the population from which to recruit. Having measured relevant baseline variables, the next task is to randomise subjects into one of two (or more) groups, and subsequently to perform the intervention as appropriate to the assignment of the subject. The pre-defined outcome measures will then be recorded and the findings compared between the two groups, with appropriate quality control measures in place to assure quality data collection. Each of these steps, which can be tested in a pilot study, has implications for the design of the trial if the findings are to be valid. They will now be considered in turn.

CHOOSING THE RIGHT POPULATION

This part of the design is crucial because poor sampling will undermine the generalisability of the study or, even worse, reduce the validity if sampling bias is introduced.[4] The task begins with deciding what kind of subjects to study and how to go about recruiting them. The *target* population is that population to which it is intended to apply the results. It is important to set inclusion and exclusion criteria defining target populations that are appropriate to the research hypothesis. These criteria are also typically set to make the researchers' task realistic, for within the target population there must be an accessible/appropriate sample to recruit.

The sampling strategy used will determine whether the sample actually studied is representative of the target population. For the findings of the study to be generalisable to the population as a whole, the sample must be representative of the population from which it is drawn. The best design is *consecutive* sampling from the accessible population (taking every patient who meets the selection criteria over the specified time period). This may produce an excessively large sample from which, if necessary, a subsample can be randomly drawn. If the inclusion criteria are broad, it will be easy to recruit study subjects and the findings will be generalisable to a comparatively large population. Exclusion criteria need to be defined and will include such subjects who have conditions which may contraindicate the intervention to be tested, subjects who

Summary: population sampling

- The study sample must be representative of the target population for the findings of the study to be generalisable.
- Inclusion and exclusion criteria will determine who will be studied from within the accessible population.
- The most appropriate sampling strategy is normally consecutive sampling, although stratified sampling may legitimately be required.
- A sample size calculation and pilot study will permit appropriate planning in terms of time and money for the recruitment phase of the main study.
- Follow CONSORT guidelines on population sampling.[6]

will have difficulty complying with the required regimens, those who cannot provide informed consent, etc.

In designing the inclusion criteria, the investigator should consider the outcome to be measured; if this is comparatively rare in the population as a whole, then it would be appropriate to recruit at random or consecutively from populations at high risk of the condition in question (*stratified* sampling). The subsamples in a stratified sample will draw disproportionately from groups that are less common in the population as a whole, but of particular relevance to the investigator.

Other forms of sampling where subjects are recruited who are easily accessible or appropriate, (*convenience* or *judgmental* sampling) will have advantages in terms of cost, time, and logistics, but may produce a sample that is not representative of the target population and it is likely to be difficult to define exactly who has and has not been included.

Having determined an appropriate sample to recruit, it is necessary to estimate the size of the sample required to allow the study to detect a clinically important difference between the groups being compared. This is performed by means of a *sample size calculation*.[5] As clinicians, we must be able to specify what we would consider to be a clinically significant difference in outcome. Given this information, or an estimate of the effect size based on previous experience (from the literature or from a pilot study), and the design of the study, a statistical adviser will be able to perform an appropriate sample size calculation. This will determine the required sample size to detect the pre-determined clinically significant difference to a certain degree of power. As previously mentioned, early involvement of an experienced researcher or research support unit in the design stage is essential in any RCT.

After deciding on the population to be studied and the sample size required, it will now be possible to plan the appropriate amount of time (and money) required to collect the data necessary. A limited pilot of the methods is essential to gauge recruitment rate and address in advance any practical issues that may arise once data collection in the definitive study is underway. Pilot studies will guide decisions about designing approaches to recruitment and outcome measurement. A limited pilot study will give the investigator an idea of what the true recruitment rate will be (not just the number of subjects available, but also their willingness to participate). It may be even more helpful in identifying any methodological issues related to applying the intervention or measuring outcome variables (see below), which can be appropriately addressed.

RANDOMISATION: THE CORNERSTONE OF THE RCT

Various baseline characteristics of the subjects recruited should be measured at the stage of initial recruitment into the trial. These will include basic demographic observations, such as name, age, sex, hospital identification, etc, but more importantly should include any important prognostic factors. It will be important at the analysis stage to show that these potential

confounding variables are equally distributed between the two groups; indeed, it is usual practice when reporting an RCT to demonstrate the integrity of the randomisation process by showing that there is no significant difference between baseline variables (following CONSORT guidelines).[6]

The random assignment of subjects to one or another of two groups (differing only by the intervention to be studied) is the basis for measuring the marginal difference between these groups in the relevant outcome. Randomisation should equally distribute any confounding variables between the two groups, although it is important to be aware that differences in confounding variables may arise through chance.

Randomisation is one of the cornerstones of the RCT[7] and a true random allocation procedure should be used. It is also essential that treatment allocations are *concealed* from the investigator until recruitment is irrevocable, so that bias (intentional or otherwise) cannot be introduced at the stage of assigning subjects to their groups.[8] The production of computer generated sets of random allocations, by a research support unit (who will not be performing data collection) in advance of the start of the study, which are then sealed in consecutively numbered opaque envelopes, is an appropriate method of randomisation. Once the patient has given consent to be included in the trial, he/she is then irreversibly randomised by opening the next sealed envelope containing his/her assignment.

An alternative method, particularly for larger, multicentre trials is to have a remote randomisation facility. The clinician contacts this facility by telephone when he is ready to randomise the next patient; the initials and study number of the patient are read to the person performing the randomisation, who records it and then reads back the randomisation for that subject.

Studies that involve small to moderate sample sizes (for example, less than 50 per group) may benefit from "blocked" and/or "stratified" randomisation techniques. These methods will balance (where chance alone might not) the groups in terms of the number of subjects they contain, and in the distribution of potential confounding variables (assuming, of course, that these variables are known before the onset of the trial). They are the design phase alternative to statistically adjusting for confounding variables in the analysis phase, and are preferred if the investigator intends to carry out subgroup analysis (on the basis of the stratification variable).

Blocked randomisation is a technique used to ensure that the number of subjects assigned to each group is equally distributed. Randomisation is set up in blocks of a pre-determined set size (for example 6, 8, 10, etc). Randomisation for a block size of 10 would proceed normally until five assignments had been made to one group, and then the remaining assignments would be to the other group until the block of 10 was complete. This means that for a sample size of 80 subjects, exactly 40 would be assigned to each group. Block size must be blinded from the investigator performing the study and, if the study is non-blinded, the block sizes should vary randomly (otherwise the last allocation(s) in a block would, in effect, be unconcealed).

Stratified randomisation is a technique for ensuring that an important baseline variable (potential confounding factor) is more evenly distributed between the two groups than chance alone might otherwise assure. In examining the effect of a treatment for cardiac failure, for example, the degree of existing cardiac failure will be a baseline variable predicting outcome, and so it is important that this is the same in the two groups. To achieve this, the sample can be stratified at baseline into patients with mild, moderate, or severe cardiac failure, and then randomisation occurs within each of these "strata". There is a limited number of baseline variables that can be balanced by stratification because the numbers of patients within a stratum are reduced. In the above example, to stratify also for age, previous infarction, and the co-existence of diabetes would be impractical.

Summary: randomisation

- The random assignment of subjects into one of two groups is the basis for establishing a causal interpretation for an intervention.
- Effective randomisation will minimise confounding variables that exist at the time of randomisation.
- Randomisation must be concealed from the investigator.
- Blocked randomisation may be appropriate for smaller trials to ensure equal numbers in each group.
- Stratified randomisation will ensure that a potential baseline confounding variable is equally distributed between the two groups.
- Analysis of results should occur based on the initial randomisation, irrespective of what may subsequently actually have happened to the subject (that is, "intention to treat analysis").

Sample attrition ("drop outs"), once subjects have consented and been randomised, may be an important factor. Patients may refuse to continue with the trial, they may be lost to analysis for whatever reason, and there may be changes in the protocol (or mistakes) subsequent to randomisation, even resulting in the patient receiving the wrong treatment. This is, in fact, not that uncommon: a patient randomised to have a minimally invasive procedure may need to progress to an open operation, for example, or a patient assigned to medical treatment may require surgery at a later stage. In the RCT, the analysis must include an unbiased comparison of the groups produced by the process of randomisation, based on all the people who were randomised; this is known as analysis by *intention to treat*. Intention to treat analysis depends on having outcomes for all subjects, so even if patients "drop out", it is important to try to keep them in the trial if only for outcome measurement. This avoids the introduction of bias as a consequence of potentialy selectively dropping patients from previously randomised/balanced groups.

APPLYING THE INTERVENTION AND MEASURING OUTCOME: THE IMPORTANCE OF BLINDING

After randomisation there will be two (or more) groups, one of which will receive the test intervention and another (or more) which receives a standard intervention or placebo. Ideally, neither the study subjects, nor anybody performing subsequent measurements and data collection, should be aware of the study group assignment. Effective randomisation will eliminate confounding by variables that exist at the time of randomisation. Without effective blinding, if subject assignment is known by the investigator, bias can be introduced because extra attention may be given to the intervention group (intended or otherwise).[8] This would introduce variables into one group not present in the other, which may ultimately be responsible for any differences in outcome observed. Confounding can therefore also occur after randomisation. Double blinding of the investigator and patient (for example, by making the test treatment and standard/placebo treatments appear the same) will eliminate this kind of confounding, as any extra attentions should be equally spread between the two groups (with the exception, as for randomisation, of chance maldistributions).

While the ideal study design will be double blind, this is often difficult to achieve effectively, and is sometimes not possible (for example, surgical interventions). Where blinding is possible, complex (and costly) arrangements need to be made to manufacture placebo that appears similar to the test drug, to design appropriate and foolproof systems for packaging and labelling, and to have a system to permit rapid unblinding in the event of any untoward event causing the patient to become unwell. The hospital pharmacy can be invaluable in

Summary: intervention and outcome

- Blinding at the stage of applying the intervention and measuring the outcome is essential if bias (intentional or otherwise) is to be avoided.
- The subject and the investigator should ideally be blinded to the assignment (double blind), but even where this is not possible, a blinded third party can measure outcome.
- Blinding is achieved by making the intervention and the control appear similar in every respect.
- Blinding can break down for various reasons, but this can be systematically assessed.
- Continuous outcome variables have the advantage over dichotomous outcome variables of increasing the power of a study, permitting a smaller sample size.

organising these issues. Blinding may break down subsequently if the intervention has recognisable side effects. The effectiveness of the blinding can be systematically tested after the study is completed by asking investigators to guess treatment assignments; if a significant proportion are able to correctly guess the assignment, then the potential for this as a source of bias should be considered.

Once the intervention has been applied, the groups will need to be followed up and various outcome measures will be performed to evaluate the effect or otherwise of that intervention. The outcome measures to be assessed should be appropriate to the research question, and must be ones that can be measured accurately and precisely. Continuous outcome variables (quantified on an infinite arithmetic scale, for example, time) have the advantage over dichotomous outcome variables (only two categories, for example, dead or alive) of increasing the power of a study, permitting a smaller sample size. It may be desirable to have several outcome measures evaluating different aspects of the results of the intervention. It is also necessary to design outcome measures that will detect the occurrence of specified adverse effects of the intervention.

It is important to emphasise, as previously mentioned, that the person measuring the outcome variables (as well as the person applying the intervention) should be blinded to the treatment group of the subject to prevent the introduction of bias at this stage, particularly when the outcome variable requires any judgement on the part of the observer. Even if it has not been possible to blind the administration of the intervention, it should be possible to design the study so that outcome measurement is performed by someone who is blinded to the original treatment assignment.

QUALITY CONTROL

A critical aspect of clinical research is quality control. Quality control is often overlooked during data collection, a potentially tedious and repetitive phase of the study, which may lead subsequently to errors because of missing or inaccurate measurements. Essentially, quality control issues occur in clinical procedures, measuring outcomes, and handling data. Quality control begins in the design phase of the study when the protocol is being written and is first evaluated in the pilot study, which will be invaluable in testing the proposed sampling strategy, methods for data collection and subsequent data handling.

Once the methods part of the protocol is finalised, an operations manual can be written that specifically defines how to recruit subjects, perform measurements, etc. This is essential when there is more than one investigator, as it will standardise the actions of all involved. After allowing all those involved to study the operations manual, there will be the opportunity to train (and subsequently certify) investigators to perform various tasks uniformly.

Ideally, any outcome measurement taken on a patient should be precise and reproducible; it should not depend on the observer who took the measurement.[4] It is well known, for example, that some clinicians in their routine medical practice record consistently higher blood pressure values than others. Such interobserver variation in the setting of a clinical trial is clearly unacceptable and steps must be taken to avoid it. It may be possible, if the trial is not too large, for all measurements to be performed by the same observer, in which case the problem is avoided. However, it is often necessary to use multiple observers, especially in multicentre trials. Training sessions should be arranged to ensure that observers (and their equipment) can produce the same measurements in any given subject. Repeat sessions may be necessary if the trial is of long duration. You should try to use as few observers as possible without exhausting the available staff. The trial should be designed so that any interobserver variability cannot bias the results by having each observer evaluate patients in all treatment groups.

Inevitably, there will be a principal investigator; this person will be responsible for assuring the quality of data measurement through motivation, appropriate delegation of responsibility, and supervision. An investigators' meeting before the study starts and regular visits to the team members or centres by the principal investigator during data collection, permit communication, supervision, early detection of problems, feedback and are good for motivation.

Quality control of data management begins before the start of the study and continues during the study. Forms to be used for data collection should be appropriately designed to encourage the collection of good quality data. They should be user friendly, self explanatory, clearly formatted, and collect only data that is needed. They can be tested in the pilot. Data will subsequently need to be transcribed onto a computer database from these forms. The database should also be set up so that it is similar in format to the forms, allowing for easy transcription of information. The database can be pre-prepared to accept only variables within given permissible ranges and that are consistent with previous entries and to alert the user to missing values. Ideally, data should be entered in duplicate, with the database only accepting data that are concordant with the first entry; this, however, is time consuming, and it may be adequate to check randomly selected forms with a printout of the corresponding datasheet to ensure transcription error is minimal, acting appropriately if an unacceptably high number of mistakes are discovered.

Once the main phase of data collection has begun, you should try to make as few changes to the protocol as possible. In an ideal world, the pilot study will have identified any issues that will require a modification of the protocol, but inevitably some problem, minor or major, will arise once the study has begun. It is better to leave any minor alterations that are considered "desirable" but not necessary and resist the inclination to make changes. Sometimes, more substantive issues are highlighted and protocol modification is necessary to strengthen the study. These changes should be documented and disseminated to all the investigators (with appropriate changes made to the operations manual and any re-training performed as necessary). The precise date that the revision is implemented is noted, with a view to separate analysis of data collected before and after the revision, if this is considered necessary by the statistical advisor. Such revisions to the protocol should only be undertaken if, after careful consideration, it is felt that making the alteration will significantly improve the findings, or not changing the protocol will seriously jeopardise the project. These considerations have to be balanced against the statistical difficulty in analysis after protocol revision.

...SOME FINAL THOUGHTS

A well designed, methodologically sound RCT evaluating an intervention provides strong evidence of a cause-effect

Summary: quality control

- An inadequate approach to quality control will lead to potentially significant errors due to missing or inaccurate results.
- An operations manual will allow standardisation of all procedures to be performed.
- To reduce interobserver variability in outcome measurement, training can be provided to standardise procedures in accordance with the operations manual.
- Data collection forms should be user friendly, self explanatory, and clearly formatted, with only truly relevant data being collected.
- Subsequent data transfer onto a computerised database can be safe guarded with various measures to reduce transcription errors.
- Protocol revisions after study has started should be avoided if at all possible, but, if necessary, should be appropriately documented and dated to permit separate analysis.

relation if one exists; it is therefore powerful in changing practice to improve patient outcome, this being the ultimate goal of research on therapeutic effectiveness. Conversely, poorly designed studies are dangerous because of their potential to influence practice based on flawed methodology. As discussed above, the validity and generalisability of the findings are dependent on the study design.

Early involvement of the local research support unit is essential in developing a protocol. Subsequent peer review and ethical committee review will ensure that it is well designed, and a successful pilot will ensure that the research goals are practical and achievable.

Delegate tasks to those who have the expertise; for example, allow the research support unit to perform the randomisation, leave the statistical analysis to a statistician, and let a health economist advise on any cost analysis. Networking with the relevant experts is invaluable in the design phase and will contribute considerably to the final credence of the findings.

Finally, dissemination of the findings through publication is the final peer review process and is vital to help others act on the available evidence. Writing up the RCT at completion, like developing the protocol at inception, should be thorough and detailed[9] (following CONSORT guidelines[6]), with emphasis not just on findings, but also on methodology. Potential limitations or sources of error should be discussed so that the readership can judge for themselves the validity and generalisability of the research.[10]

Further reading

Sackett DL, Haynes RB, Guyatt GH, *et al*. *Clinical epidemiology: a basic science for clinical medicine.* 2nd edn. Toronto: Little, Brown, 1991.

Sackett DL, Richardson WS, Rosenberg W, *et al*. *Evidence-based medicine: how to practice and teach EBM.* Edinburgh: Churchill Livingstone, 1997.

Polgar S. *Introduction to research in health sciences.* 2nd edn. Edinburgh: Churchill Livingstone, 1991.

Bland M. *An introduction to medical statistics.* Oxford: Oxford Medical Publications, 1987.

REFERENCES

1 **Sibbald B**, Roland M. Understanding controlled trials: Why are randomised controlled trials important? *BMJ* 1998;**316**:201.
2 **Pocock SJ**. *Clinical trials: a practical approach.* Chichester: Wiley, 1984.
3 **Hulley SB**, Cunnings SR. *Designing clinical research—an epidemiological approach.* Chicago: Williams and Wilkins, 1988.
4 **Bowling A**. *Researching methods in health: investigating health and health services.* Buckingham: Open University Press, 1997.
5 **Lowe D**. *Planning for medical research: a practical guide to research methods.* Middlesborough: Astraglobe, 1993.
6 **Begg C**, Cho M, Eastwood S, *et al*. Improving the quality of reporting of randomised controlled trials: the CONSORT statement. *JAMA* 1996;**276**:637–9.
7 **Altman DG**. Randomisation. *BMJ* 1991;**302**:1481–2.
8 **Schultz KF**, Chalmers I, Hayes RJ, *et al*. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;**273**:408–12.
9 **The Standards of Reporting Trials Group**. A proposal for structured reporting of randomised controlled trials. *JAMA* 1994;**272**:1926–31.
10 **Altman DG**. Better reporting of randomised controlled trials: the CONSORT statement. *BMJ* 1996;**313**:570–1.