1

Introduction

Cluster randomised trials are trials in which *groups (or clusters) of individuals* are randomly allocated to different forms of treatment. In healthcare, the different forms of treatment are sometimes different drugs or, more commonly, different ways of managing a disease or promoting healthy living. These trials are in contrast to conventional randomised trials which randomise *individuals* to different treatments, classically comparing new drugs with a placebo. Cluster randomised trials are common in health services research. This is an area of research concerned with the way healthcare is delivered and with measures taken to prevent ill health and encourage healthy living. It covers a broad range of topics and is an important area in maintaining high standards in a modern health service. New initiatives or interventions in health care may be evaluated by comparing health outcomes in those that are exposed to the new initiative with outcomes in those receiving usual care or an alternative intervention. Since interventions often need to be introduced to a whole organisational unit such as a general practice or geographical area, cluster randomised trials are often the best method of evaluating such interventions.

There are many books written about trials in general, which explain in detail the key features of the design, conduct and analysis of randomised trials; but these are mainly concerned with trials which randomise individual patients to different interventions (Pocock, 1983; Matthews, 2000; Torgerson and Torgerson, 2008). There are now three books that describe the design, analysis and conduct of cluster randomised trials: Murray (1998), Donner and Klar (2000) and Hayes and Moulton (2009). These books have mainly concentrated on large community trials. Hayes and Moulton have a particular emphasis on trials in low-income countries where whole communities have been randomised. Since we have extensive experience in

A Practical Guide to Cluster Randomised Trials in Health Services Research, First Edition.
Sandra Eldridge and Sally Kerry.

health services research, in this book we have focused on cluster randomised trials in this area, though we have used other examples where useful. This book is intended as a practical guide, written for researchers from the health professions, including doctors, psychologists, and allied health professionals, as well as statisticians, who are involved in the design, execution, analysis and reporting of cluster randomised trials. It is specifically written to address the issues arising from allocating groups of individuals, or clusters, to different interventions, and is primarily concerned with those aspects of cluster randomised trials which differ from randomised trials of individual subjects. Several trials are used as examples throughout the book. These are listed at the front of the book.

1.1 Introduction to randomised trials

A formal definition of a trial is given in Box 1.1. The 'gold standard' for trials is the randomised controlled trial (RCT), originally developed in order to test the efficacy of new drugs. In the earliest example of such a trial (Medical Research Council, 1948), patients were randomly allocated to treatments, each participant having an equal chance of being given the active drug or placebo. As a result any patient characteristics that might have affected the outcome of the treatment would have been randomly distributed between the intervention and control arms, and the observed difference in outcome between the arms could be attributed to the active drug.

Over the years the RCT design has been extended to many other situations: more than two different treatments; crossover trials; non-drug interventions such as surgery, physiotherapy or health education; and in health services research to assess the effectiveness of different models of care.

1.2 Explanatory or pragmatic trials

Randomised trials may be used to test causal research hypotheses. Various epidemiological studies have shown that high salt intake is associated with high blood

> Any research project that prospectively assigns human subjects to intervention and comparison groups to study the cause-and-effect relationship between a medical intervention and a health outcome. By 'medical intervention' we mean any intervention used to modify a health outcome. This definition includes drugs, surgical procedures, devices, behavioural treatments, process-of-care changes, and the like.
>
> Source: International Committee of Medical Journal Editors, 2009.

Box 1.1 Definition of a trial

Table 1.1 Kumasi trial: health education to prevent stroke.

Aim: To see if a health education programme to reduce salt intake among rural and semi-rural communities in the Ashanti region of Ghana leads to a reduction in blood pressure
Location and type of cluster: Ghana, villages of 500–2000 inhabitants
Number of clusters randomised: 12
Number of villagers randomised: 1013
Interventions: (i) Control: health education not including salt reduction (ii) Intervention: health education including salt reduction messages
Primary outcome: Reduction in systolic blood pressure after six months

Source: Cappuccio *et al.* (2006).

pressure. In order to test whether this relationship was causal, the DASH trial (Moore *et al*., 2001) recruited a carefully selected group of patients with moderately raised blood pressure and randomised them to take a low salt diet or usual American diet. All the subjects' food was provided by the trial team. Trials such as this, which seek to understand a biological process, are described as explanatory. Explanatory trials may also test the efficacy of treatments under ideal conditions (Roland and Torgerson, 1998). Cluster randomised trials rarely fall into this category.

Pragmatic trials, on the other hand, are designed to help choose between care options applied in routine clinical practice. Providing a low salt diet for people is not a practical option, except perhaps in hospitals and care homes, and a more realistic approach is to reduce dietary salt using health education for the whole community. The Kumasi trial (Table 1.1) took a whole community approach to health promotion: advice on how to reduce dietary salt was dispensed not only to the individuals participating in the trial but also to their families and neighbours, with whom they might share meals. The intervention was therefore not a 'low salt diet' but 'community education to reduce dietary salt'. Many cluster randomised trials are pragmatic trials and share common features with other individually randomised pragmatic trials (Zwarenstein *et al*., 2008; Eldridge, 2010).

1.3 How does a cluster randomised trial differ from other trials?

A cluster randomised trial is one in which groups or clusters of individuals rather than individuals themselves are randomised to intervention arm. These clusters are often social units. They can range in size from small units such as households, to much larger units such as towns or regions. Often they comprise individuals connected to particular institutions, for example patients attending particular clinics or general practices, or children in particular schools.

While whole clusters form the units of randomisation (or experimental units) in cluster randomised trials, the members of these clusters form the units of observation. These may be all the members of the cluster or a sample from each cluster. It is this distinction between units of randomisation and units of observation which distinguishes cluster randomised trials from the more usual types of randomised trial, with statistical and practical consequences. In this section we briefly describe the consequences of cluster randomisation, covering recruitment, randomisation, consent, analysis, sample size and interventions. All of these issues are dealt with more fully in later chapters.

1.3.1 Recruitment, randomisation and consent

In these key areas, cluster randomised trials exhibit unique features not present in individually randomised trials. Consent to participate may be required from clusters, individuals or both. Even when consent is not required from participants, the methods used to select individuals on whom data will be collected need to be carefully considered in order to avoid bias. This will be discussed in more detail in Chapter 2, but here we describe a few examples to illustrate the wide variability of recruitment, randomisation and consent procedures seen in cluster randomised trials.

A simple trial of radiological guidelines to reduce unnecessary referrals for x-ray by general practitioners is described in Table 1.2. Neither practices nor individuals were asked to consent to participation. Practices regularly referring to one hospital radiology department were identified from the department's records and randomly

Table 1.2 Guidelines to reduce inappropriate referral for x-ray.

Aim: To determine whether postal distribution of radiological guidelines for x-ray referral reduces the number of x-ray referrals from primary care and inappropriate referral for x-ray
Location and type of cluster: UK General practices
Number of clusters analysed: 64
Number of individuals analysed: 2578 (different patients were included at baseline and follow up)
Interventions: (i) Control: no intervention (ii) Intervention: laminated extracts of Royal College of Radiologists' guidelines specifically produced for primary care, posted to general practitioners individually
Primary outcome: Percentage of referrals assessed as conforming to the guidelines using x-ray referral forms collected in the radiology department
Consent required from clusters: No
Consent required from patients: No
Individuals identified prior to randomisation: No, but identified and assessed blind to intervention arm

Source: Oakeshott, Kerry and Williams (1994).

Table 1.3 OPERA: physical activity in residential homes to prevent depression.

Aim: To evaluate the impact on depression of a whole-home intervention to increase physical activity among older people
Location and type of cluster: UK residential and nursing homes for older people
Number of clusters randomised: 78
Number of residents recruited: 1060
Interventions: (i) Control: depression awareness programme delivered by research nurses (ii) Intervention: depression awareness programme delivered by physiotherapists plus whole-home package to increase activity among older people, including physiotherapy assessments of individuals and activity sessions for residents
Primary outcome: Prevalence of depression (Geriatric Depression Scale) at 12 months and change in depression score
Consent required from clusters: Yes
Consent required from individuals: Consent was required separately for completion of the Geriatric Depression Scale, access to medical records, and for physiotherapy assessments. All residents were encouraged to take part in activity sessions. Residents of control homes could not access the activity sessions
Individuals consented prior to randomisation: Yes, for individuals resident in the home prior to randomisation but not for individuals moving into the home during the study

Source: Underwood *et al*. (2011).

assigned to an intervention arm or control arm. Individual general practitioners in intervention practices were sent copies of the guidelines through the post, while those in the control arm were sent nothing. Outcomes were assessed through audit of radiology request forms for individual patients held within the radiology department. Identification of the individual patients, who were the units of observation, was carried out after randomisation, but blind to whether or not their practice was in the intervention arm.

A much more complex design is described in Table 1.3. Residential homes for older people were randomised to receive an intervention aimed at reducing depression among the residents. After all residents had been asked for consent to data collection and, if agreeable, had taken part in a baseline assessment, homes were randomised to intervention or control. Part of the intervention was twice-weekly physical activity sessions run in the homes by a physiotherapist. Residents could opt out of attending specific activity sessions but, because they still belonged to a home where the staff had been trained to encourage residents to be more active, they could not opt out of the intervention entirely. Individual residents could refuse to take part in the outcome assessments or refuse to allow researchers access to their medical records. Residential homes were required to give consent and be actively involved in delivering the intervention and assisting the trial team with identification

Table 1.4 UK BEAM pilot trial: active management of back pain.

Aim: To determine whether active management of patients presenting with back pain in general practice reduces back pain disability
Location and type of cluster: UK general practices
Number of clusters recruited: 26
Number of patients recruited: 231
Interventions: (i) Control: usual care (ii) Intervention: all clinical and support staff were invited to training sessions on the active management of back pain; practices were supplied with literature to distribute during consultations and in communal areas. Patients were also randomised individually to exercise classes or spinal manipulation or neither
Primary outcome: Change in back pain disability
Consent required from clusters: Yes
Consent required from patients: Yes
Individuals consented prior to randomisation: No. Individuals were identified by the practice upon presentation to the general practitioner with back pain after randomisation. Cluster design abandoned after pilot study due to evidence of bias in recruitment

Source: Farrin *et al*. (2005).

of participants and data collection. This trial illustrates the complexities in obtaining consent that can arise in cluster randomised trials.

In a traditional RCT, consent should always take place before the allocation to intervention arm is known, thus ensuring that the decision to take part in the study is not biased by knowledge of the allocation. In cluster randomised trials such an approach can create major difficulties if the intervention is aimed at managing an acute condition or the onset of a chronic condition; the patients cannot be identified and recruited prior to randomisation, but only when they present to the general practitioner. It may therefore be necessary to allocate the clusters to intervention arms before individual cases are identified. In the UK BEAM trial pilot study (Table 1.4), 26 practices were randomised to offer active management or usual care to patients presenting with low back pain. Patients within the active management arm were also individually randomised to receive spinal manipulation, exercise classes or advice alone. After one year, practices in the control arm (traditional care) had recruited 66 patients, 54% of the number predicted based on practice list size, while those in the active management arm had recruited 165 patients, 41% more than predicted. In addition, participants from the active management arm were suffering from milder back pain than those in control practices. It is likely that the offer of exercise classes or physiotherapy made participation in the trial an attractive option for the general practitioners and their patients in the active management arm, whereas there was no such benefit for patients in the control arm. Following the pilot study, the trial was redesigned as an individually randomised trial comparing different methods of delivering active management. Here all participants, at the time of consent, would have

an equal chance of receiving an active intervention. This highlights the potential for bias that can arise if individual patients are identified or recruited after randomisation. Chapter 2 discusses identification and recruitment bias in more detail and outlines some approaches which can be used to protect against these biases.

1.3.2 Definition of cluster size

Very often only a subset of individuals in the cluster provides data for the analysis. In this book we will refer to the number of individuals per cluster who contribute data to the analysis as the 'cluster size' and the number of patients in the larger pool from which they come as the 'natural cluster size'. In the UK BEAM trial (Table 1.4), the average cluster size was 5.1 (66 individuals from 13 practices) in the control arm and 12.7 (136 individuals from 13 practices) in the intervention arm, while the average natural cluster size was 7804 in the intervention arm and 8145 in the control arm. These averages are slightly larger than the average for all English practices, which was 6649 in 2009 (Health and Social Care Information Centre, 2011).

1.3.3 Analysis and sample size

The primary aim of a randomised trial is to compare outcome measures in different intervention arms. The simplest analysis is a *t*-test for comparing two means, or a chi-squared test for comparing two proportions. These tests assume that observations on participants can be regarded as independent of one another. However, in cluster randomised trials, members of the same cluster are more likely to have similar outcomes than a random sample from the same population, and therefore cannot be regarded as independent**.** Where outcomes relate to participants' own health or behaviour, the effect of clustering is likely to be small. Where outcomes relate directly to the behaviour of the clusters, then the effect of clustering may be much larger. For example, doctor's prescribing behaviour for a particular condition may be more dependent on the doctor's opinions, views and habits than on the patient's condition, while systolic blood pressure may have only a small tendency to be similar among patients attending the same practice. This tendency to have similar outcomes is known as within-cluster homogeneity, and needs to be taken into account in the design and analysis. An alternative expression used to describe this concept is 'between-cluster variability', and this is the term we shall use in this book. The most common measure of between-cluster variability is the intra-cluster correlation coefficient (ICC), which is described in more detail in Chapter 8.

Using analysis methods which fail to take account of clustering may lead to confidence intervals which are too narrow, and increased Type 1 error; that is, results may appear to have a higher level of statistical significance than they actually do. Chapter 6 describes in detail suitable methods to analyse cluster randomised trials.

Since correct methods for analysing cluster randomised trials lead to wider confidence intervals, the sample size also needs to be adjusted for the effect of

clustering. In order to detect the same size effect, cluster randomised trials will always require more subjects than individually randomised trials designed to answer identical research questions (assuming it is possible to randomise individuals). Where the number of subjects recruited from each cluster is small and the ICC is small, the increase in the sample size will also be fairly small. However, if the number of participants to be recruited from each cluster is large then even a small ICC may double the sample size required. The Kumasi trial (Table 1.1) used change in systolic blood pressure as an outcome and required 840 participants to be included in the final analysis; if the trial had been individually randomised it would have required less than half that number. Chapter 7 describes how to allow for clustering in sample size calculations.

1.3.4 Interventions used in cluster randomised trials

Cluster randomised trials rarely use interventions which can be delivered blind, except in the case of drugs for the treatment or control of infectious diseases. More commonly, cluster randomised trials are used to assess the effectiveness of educational interventions or management strategies aimed at the whole cluster, and it is not possible to blind the members of the cluster. Ideally the outcome should be assessed blind to the allocation. This situation is not unique to cluster randomised trials, but often presents greater challenges in these trials. If data need to be collected within the cluster it may be difficult to conceal allocation arm from any researcher entering, say, a general practice. Posters or information leaflets may be displayed on the premises, and staff aware of the intervention may inadvertently reveal the allocation. Where patients are interviewed they may be asked not to reveal the allocation of their cluster to the researcher. If an individual patient reveals the arm to which they belong and the trial is individually randomised, only the data from one individual may be compromised, but if it is a cluster randomised trial, assessors are unblinded when assessing all remaining participants from the cluster.

Many interventions used in cluster randomised trials are made up of various connecting parts and can be described as complex interventions. These can be complicated to design, to carry out and to describe. For example the Kumasi trial (Table 1.1) randomly allocated villages to receive a health education package advising villagers to reduce dietary salt in order to reduce their blood pressure. Replication of this trial would require much more detail about what the package entailed, how and when it was delivered, and what both intervention and control arms were told when consenting to take part. Many complex interventions have failed to demonstrate the desired effect of the intervention. In a drug trial, if the trial shows no evidence of benefit and is sufficiently powered, it is usually safe to conclude that the drug does not work, at least at the specified dose. In the case of complex interventions, the interpretation may be more problematic. The intervention *as delivered* has proven to be ineffective, but we need to be sure exactly what the intervention entailed and that the lack of effectiveness is not due to poor implementation, or to

changing behaviour in the control arm owing to information provided while obtaining consent. Careful consideration of how different parts of the intervention interact to bring about change in the individual is needed at the design stage. Eldridge *et al*. (2005) modelled the effect of a primary care intervention to screen older people at risk of hip fracture. This showed that the intervention was unlikely to be effective and a large expensive trial was not justified. Complex interventions are described in more detail in Chapter 3.

1.4 Between-cluster variability

In order to understand the effect of clustering on analysis and sample size, it is useful to consider why members of a cluster may be more similar in their outcomes than a random sample of individuals.

1.4.1 Factors that contribute to between-cluster variability

1.4.1.1 Geographical reasons

Most clusters have some kind of geographical basis. Patients registered with a general practice will live near the practice. Social factors such as deprivation are known to affect health outcomes and so will contribute to within-cluster homogeneity. Even stronger effects on between-cluster variance may be observed for lifestyle and behaviours such as smoking and diet.

1.4.1.2 Individuals choose the cluster to belong to

Individuals may be able to choose where they live, which general practice to attend, and which school for their children's education. These choices may be influenced by ethnic, religious or other characteristics, which may in turn influence health outcomes and behaviours, thus contributing to within-cluster homogeneity.

1.4.1.3 Healthcare provided to the cluster

As well as sharing a common environment, members of a cluster will usually be treated by the same healthcare professionals. A general practice which treats hypertension more aggressively is likely to have more patients taking antihypertensive medication, and with consequently lower blood pressure, than one with a more conservative approach.

1.4.2 Measuring between-cluster variability

The variability between clusters in outcomes is often estimated by the intra-cluster correlation coefficient (ICC). This may be thought of as the ratio of the variability

between clusters to the total variability in the outcome, although there are alternative ways of defining this quantity (see Chapter 8). Much of the early work on cluster randomised trials by Donner (Donner, Birkett and Buck, 1981) used the ICC, and sample size calculations within health services research also usually use it. The ICC is the measure on which we shall concentrate in this book.

Other methods of estimating the between-cluster variation are the between-cluster variance (Cornfield, 1978), often denoted by σ_b^2, or the between-cluster coefficient of variation of the outcome (σ_b/μ) (Hayes and Bennet, 1999), where μ represents the mean outcome across all clusters. The latter is particularly useful for comparing event rates expressed as number of events per person years (Hayes and Moulton, 2009), and is described in more detail in Chapter 7.

1.5 Why carry out cluster randomised trials?

So far in this chapter we have shown that cluster randomised trials require more subjects than individually randomised trials, are harder to design, are prone to bias in ways that individually randomised trials are not, and give rise to more ethical issues, particularly with regard to informed consent. Consequently they should not be carried out without good justification. We consider seven possible reasons for undertaking cluster randomised trials.

1.5.1 The intervention necessarily acts at the cluster level

Here the intervention is directed towards the whole cluster and could not be implemented for some individuals and not others. Examples include education interventions for healthcare practitioners (Table 1.2), mass education programmes using TV, radio and posters, and changing the environment, for example fluoridation of water. In these examples the whole cluster is subject to the intervention and the intervention could not be implemented in any other way.

In the OPERA trial (Table 1.3), the intervention involved training all staff in the residential home in the importance of remaining active and ways to encourage activity among the residents, provision of activity sessions open to all residents, and assessment of individual mobility needs. The intervention aimed to change the culture within the home and therefore acted at cluster level.

1.5.2 Practical and/or ethical difficulties in randomising at individual level

A trial in Zimbabwe (Murira *et al.*, 1997) of two different antenatal systems, one an existing system in which women had 12 visits and the other a new system in which women had 6 visits during their pregnancy, would have been more difficult to organise on an individual basis. In the ObaapaVitA trial in Ghana (Table 1.5), all

Table 1.5 ObaapaVitA: vitamin A supplementation to reduce maternal and child mortality.

Aim: To see if supplementation with vitamin A would reduce maternal and child mortality in Ghana
Location and type of cluster: Ghana, geographical areas
Number of clusters randomised: 1086
Number of women randomised: 208 145
Interventions: (i) Control: placebo capsules (ii) Intervention: vitamin A capsules
Primary outcome: Pregnancy-related mortality and all-cause female mortality
Length of follow-up: 5–7 years

Source: Kirkwood *et al*. (2010).

Table 1.6 Promoting child safety by reducing baby walker use.

Aim: To evaluate the effectiveness of an educational package provided by midwives and health visitors to reduce baby walker possession and use
Location and type of cluster: UK groups of general practices sharing a health visitor (between 1 and 4 practices)
Number of clusters analysed: 46
Number of individuals analysed: 1008
Interventions: (i) Control: usual care (ii) Intervention: trained midwives and health visitors delivered an educational package to mothers to be, at 10 days postpartum and 3–4 months later, to discourage baby walker use or encourage safe use for those who already had baby walkers
Primary outcome: Possession and use of a baby walker

Source: Kendrick *et al*. (2005).

women in the same cluster, approximately 160 in number, were given identical capsules; for some clusters these contained vitamin A in peanut oil, for others peanut oil only. During monthly visits to the cluster by fieldworkers, the women were given four capsules to be taken once weekly. Fieldworkers were given only one type of capsule at a time. In this way the women could not be given the wrong capsules by mistake in this large trial in a low-income country.

1.5.3 Contamination at health professional level

In a trial of an education package to reduce the use of baby walkers by infants (Table 1.6), the intervention was delivered through midwives and health visitors during routine appointments and visits. In an individually randomised trial it would have been difficult for midwives effectively to discourage the use of baby walkers for some women and not others, and for the researchers to be sure the right women were getting the intervention.

1.5.4 Contamination between members of a cluster

In some trials where randomisation takes place at the individual level, the control arm may be partially exposed to the intervention through interaction with individuals receiving the intervention. In the baby walker trial (Table 1.6), mothers attending the same practice were more likely to use the same baby clinics and support groups, and might have interacted and discussed the use of walkers with one another. The degree of interaction between mothers is likely to be higher than between other adults attending the same general practice. The Family Heart Study (Wood *et al*., 1994) used two different control arm in a trial of cardiovascular screening among middle-aged adults. Two practices in each town were randomised so that one received cardiovascular risk screening with a risk reduction programme, and the other practice carried on with usual care. Within intervention practices, couples were randomised to the intervention or usual care. The results of the study showed little evidence of contamination between adults attending general practices for this intervention.

Torgerson (2001) has suggested that increasing the sample size of an individually randomised trial to allow for contamination between cluster members may be preferable in some circumstances to randomising clusters with all the attendant difficulties.

1.5.5 Cost or administrative convenience

In health services research, clusters are usually administrative units with participants who are located in a geographical area. Consequently there may be administrative reasons why it is easier to restrict the intervention to fewer clusters. In the SHIP trial (Jolly *et al*., 1999), specialist nurses worked with the intervention practices to help integrate primary and secondary care of patients suffering from heart attacks. It would have been much more expensive to work with all practices. Trials which involve the use of expensive equipment may also be cluster randomised to avoid having to equip all units.

1.5.6 Ensuring intervention is fully implemented

In the ObaapaVitA trial (Table 1.5), the women might have been tempted to swap capsules with their neighbour in the hope of getting some benefit should they be randomised to placebo. By randomising all women in a village to the same intervention, researchers ensured that this would not matter. In other trials where new technologies are being introduced, they may have greater effect if staff become accustomed to using the new methods by treating everyone in the cluster. In the World Health Organization partograph study (World Health Organization, 1994), centres were randomised to training in the use of the partograph for monitoring progress in labour. Accordingly, midwives could become familiar with the partograph technique and it could become part of routine practice. In other cluster

randomised trials, individually administered interventions may be reinforced by publicity at the cluster level.

1.5.7 Access to routine data

In a proposed cluster randomised trial where the intervention was designed to reduce fractures amongst elderly people (Eldridge *et al*., 2001), the outcome measure chosen was the overall rate of fractured femur in the cluster. These data were easily obtainable from routine sources at the cluster level. Data direct from individuals would have been much more difficult to obtain.

1.6 Quality of evidence from cluster randomised trials

Healthcare professionals, managers and community leaders need to be informed as to which healthcare interventions are effective, and with sufficient information to be able to judge how well the evidence applies to their particular situation. Randomised trials are regarded as the best kind of evidence to inform good practice, but the strength of that evidence can be compromised in two ways. Firstly, to what extent are the results of the trial free from bias, so that the observed effect of the intervention is the result of the intervention itself and not due to characteristics of the subjects recruited to each intervention group or the way the outcome was measured? Secondly, are there key differences between the trial participants and the population to which the results are to be applied, or between the intervention as delivered in the trial and as it might be delivered in a routine setting? These two different aspects of trial quality are known as *internal validity* and *external validity* (Box 1.2). External validity is also referred to as generalisability.

Internal validity	The extent to which differences identified between randomised arms are a result of the intervention being tested.
External validity	The extent to which study results can be applied to other individuals or settings.

Box 1.2 Definition of internal and external validity

1.6.1 External validity

There have been pleas from various researchers to take the issue of external validity more seriously. For example, Glasgow, Vogt and Boles (1999) argue that much research focuses on determining efficacious interventions, thus involving trials with strong internal validity but, partially as a result, weak external validity. Rothwell

(2005) has also argued for greater consideration of external validity in the design and reporting of trials on the grounds that 'Lack of external validity is the most frequent criticism by clinicians of RCTs, systematic reviews, and guidelines, and is one explanation for the widespread underuse in routine practice of many treatments that have been shown to be beneficial in trials and are recommended in guidelines.' Cochrane (1972) and Bradford Hill (Horton, 2000) also recognised the importance of external validity several decades ago.

Table 1.7 describes the different elements of external validity based on Rothwell's paper (Rothwell, 2005) but adapted to include those aspects most relevant to cluster randomised trials; and the selection of clusters as well as participants. With the addition of 'setting', they cover the elements in the PICO (Population, Intervention, Comparator, Outcome) framework often used to define a research question in systematic reviews (Sackett *et al*., 2000).

Ideally an individually randomised trial would recruit a random sample of individuals from the target population to whom the intervention would be delivered in routine practice if shown to be effective. In practice such a situation rarely arises. Those who take part in randomised trials may be systematically different from the target population. This is known as selection bias. In a cluster randomised trial, selection bias can take place at the individual and the cluster level. This might be due to the setting in which the trial takes place, the inclusion and exclusion criteria for the trial, or the recruitment and consent process (Table 1.7).

1.6.2 Internal validity

A well conducted, double-blind, placebo-controlled trial of sufficient size is likely to have high internal validity. The blinding serves several purposes. Firstly, the subjects are allocated to active or placebo arm without either the investigator or the subject knowing which treatment the patient will receive; knowledge of such allocation may influence the researcher's assessment of the patient's suitability and the patient's decision to participate. Secondly, blinding ensures that knowledge of allocation is unlikely to affect concomitant treatments or lifestyle choices. Finally, outcomes will be assessed blind, thus avoiding any temptation on the part of the patient or the researcher to bias the results in favour of (or against) the new treatment. Blinding of participants to the intervention is uncommon in cluster randomised trials, and selection of individual participants may take place after allocation of clusters to intervention arms. Careful attention to the selection and recruitment of clusters and individual participants, as well as objective assessment of outcomes, can all help protect the trial's internal validity. We consider blinding further in Chapters 9 and 10.

1.6.3 Balancing internal validity, external validity and ethical issues

Godwin *et al*. (2003) discuss the balance between internal and external validity in pragmatic trials in primary care. They make a clear distinction between explanatory

Table 1.7 Elements of external validity.

Element	Application	Example
Setting	Setting	Healthcare system Country Primary, secondary or tertiary
	Selection of clusters	Eligibility criteria Selection process
	Characteristics of clusters	Response rates Reasons for non-response Characteristics of recruited clusters
Population	Selection of participants	Eligibility criteria Selection process
	Characteristics of clusters	Response rates Reasons for non-response Characteristics of recruited participants
Intervention	Differences between protocol and routine practice	Intervention as delivered in the trial compared to the intervention as intended to be delivered in routine care Changes in management/new guidelines/ new policies since the trial began
Control arm	Effect of trial participation	Effect of trial participation on outcomes Effect of informed consent procedures Hawthorne effect
Outcomes	Outcome measures and follow-up	Follow-up rates Relevance of outcomes Relevance of surrogate outcomes and process measures compared with patient-orientated outcomes Appropriateness of timing of follow-up
	Adverse events	Completeness of reporting of adverse events Selection of clusters on basis of skill or experience

Source: adapted from Rothwell (2005).

trials in which the primary aim is to assess efficacy, and pragmatic trials in which the primary aim is to assess effectiveness, or how interventions work in real situations (Section 1.2). For explanatory trials internal validity is paramount, but in pragmatic trials there has to be a 'creative tension' between internal and external validity. This tension exists because promoting internal and external validity requires

a certain amount of effort; investigators must judge how much work to put into ensuring each type of validity, depending on their available resources. In addition, issues around participants' consent also need to be considered. In some circumstances ethical issues may compromise either the internal or external validity, or both. In Chapter 2 we will concentrate on ethical issues, selection and recruitment. In Chapter 10 we discuss good practice in reporting internal and external validity.

1.7 Historical perspectives

1.7.1 Early cluster randomised trials

The first cluster randomised trials took place in the field of education, where the clusters were classes, year cohorts or schools. The need to take clustering into account in the analysis was clearly recognised as early as 1940 by Lindquist (1940), who proposed using summary statistics for each cluster as a method of analysis. This was not universally accepted among statisticians at the time, although it is now recognised as a valid method.

1.7.2 Early cluster randomised trials in health up to 2000

Early cluster randomised trials in medical research are difficult to identify as they were not usually described as 'cluster randomised'. Terms such as 'community trials' were used (Isaakidis and Ioannidis, 2003), and the first textbook of cluster randomised trials published by Murray (1998) used the term 'group randomised'. Cluster randomised trials have been used in sub-Saharan Africa since the early 1970s, and have been mainly concerned with reducing rates of infectious diseases such as malaria and sexually transmitted infections (Isaakidis and Ioannidis, 2003). Some of the earliest cluster randomised trials in health services research were in the use of computer-based, clinical decision support systems, beginning in the mid 1970s (Wexler *et al*., 1975; Chuang, Hripcsak and Jenders, 2000) and extending into screening and treatment of risk factors for coronary heart disease in the late 1970s. Cornfield's landmark paper in 1978 marked the beginning of the extensive development of methods for designing and analysing these trials (Cornfield, 1978). The mid to late 1990s saw the publication of several large cluster randomised trials in UK primary care (Wood *et al*., 1994; Feder *et al*., 1995; Kinmonth *et al*., 1998; Jolly *et al*., 1999; Feder *et al*., 1999). A series of statistics notes in the BMJ (Bland and Kerry, 1997; Kerry and Bland, 1998a, 1998b, 1998c) was published in the late 1990s, and a workshop on cluster randomised trials was held in Sheffield, UK in 1999 (Campbell, Donner and Elbourne, 2001). All of these raised the profile of cluster randomised trials and increased the awareness in the research community of the need to allow for clustering in analysis and power calculations. Two key textbooks, Murray (1998) and Donner and Klar (2000), were published at the end of the decade. Medical statistics textbooks also began to highlight the issues surrounding cluster randomised trials and to give some guidance to statisticians who might be involved in these trials (Bland, 2000).

1.7.3 Recent methodological developments

Since then the number of cluster randomised trials published has increased (Bland, 2004), and there has been a growth in awareness of the statistical issues (Eldridge *et al*., 2008) and a continuing increase in papers describing new methods of analysis and other design issues. Several reviews of methodological developments have been published in the last 10 years (e.g. Murray, Varnell and Blitstein, 2004; Campbell, Donner and Klar, 2007). In this book we describe standard older methods of design and analysis and incorporate new developments in the literature in the last decade, focusing on the following topics.

1.7.3.1 Methods of analysis

Cornfield (1978) and other early methodological papers (Kerry and Bland, 1998a; Donner and Klar, 2000) recommended using summary statistics for each cluster as a valid analysis method, which could be used by researchers with little statistical expertise on readily available software. However, since then there have been considerable advances in available statistical methods such as multilevel modelling, robust standard errors, generalised estimating equations and Bayesian hierarchical models, and their use in cluster randomised trials has been extensively reviewed. These methods are described in Donner and Klar (2000) and in Hayes and Moulton (2009). In Chapter 6 we describe these methods, focusing on the choice of analysis from a practical point of view, with a wide range of examples from the literature where different methods have been used.

1.7.3.2 Sample size

Early methodological papers on sample size assumed that clusters are the same size. This is rarely the case in health services research. In chapter 7 we describe an adaptation of the method proposed by Donner (Donner, Birkett and Buck 1981) to allow for variability in cluster size. Although we concentrate on sample size methods which use the ICC to allow for between-cluster variation but we also discuss the application of other methods to health services research.

1.7.3.3 Estimating the intra-cluster correlation coefficient

One of the key problems for sample size calculations is how to estimate the ICC. There is now much more information available to help the researcher, which will be described in Chapter 8. We describe different methods of calculating the ICC and its precision, and give some guidelines as to the likely value based on published values, type of outcome and other factors.

1.7.3.4 Reporting guidelines

As cluster randomised trials add a level of complexity to design and analysis, they need to be accurately reported. The CONSORT statement was originally developed in 1996 (Begg *et al*., 1996) to improve the standard of reporting of randomised trials. In Chapter 10 we describe how these guidelines have been extended for

cluster randomised trials (Campbell, Elbourne and Altman, 2004), alongside a consideration of the recently updated CONSORT guidelines (Moher *et al.*, 2010) for randomised trials.

1.7.3.5 Recruitment and consent

The recognition of the problems of bias in identification and recruitment of individual participants has been raised in the literature (Puffer, Torgerson and Watson, 2003; Eldridge *et al*., 2008; Eldridge, 2010) more recently than the need to adjust sample size for clustering. In Chapter 2 we describe the situations which give rise to bias in more detail and suggest some possible solutions.

1.7.3.6 Complex interventions

Many cluster randomised trials have shown no clear evidence to support the intervention being tested. This is partly because the interventions have not been well enough developed and are insufficiently intensive to be able to demonstrate a benefit. The Medical Research Council developed guidance for the development of complex interventions in 2000 (Campbell *et al*., 2000), which was updated in 2008 (Craig *et al*., 2008). We shall discuss this in detail in Chapter 3 and consider how to plan appropriate pilot studies in Chapter 4.

1.7.3.7 Other topics

The aim of randomisation is to produce groups of participants that will be similar with respect to characteristics that might affect the outcome. This will work best when a large number of clusters are randomised, but this is not usually the case for cluster randomised trials. Consequently methods such as stratification and matching are used to improve comparability of the intervention arms. Designs which are commonly used in cluster randomised trials are discussed in Chapter 5.

Systematic reviews may include individually randomised trials and cluster randomised trials or be restricted to cluster randomised trials depending on the nature of the intervention. Chapter 9 describes how to apply the principles of systematic reviews to cluster randomised trials, and also includes sections on cost-effectiveness and process evaluation. These topics have not been included in other textbooks of cluster randomised trials.

1.8 Summary

Cluster randomised trials are trials in which groups or clusters of individuals, rather than the individuals themselves, are randomised; this makes their design, analysis and conduct more complicated. Cluster randomised trials are less powerful than individually randomised trials with the same number of individual participants, analysis must take the clustering into account to avoid spurious precision in the confidence intervals, and there is a greater potential for bias when recruiting subjects. Nevertheless, there are several good reasons for adopting this trial design to

evaluate a variety of interventions in community healthcare, and these trials are increasingly common. There is also a growing body of methodological literature focusing on these trials. In this book we bring together the literature on these recent developments, with particular emphasis on its implications for those designing and analysing these trials in health services research.

References

Begg, C., Cho, M., Eastwood, S. *et al.* (1996) Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA*, **276** (8), 637–639.

Bland, J.M. (2000) *An Introduction to Medical Statistics*, 3rd edn, Oxford University Press, Oxford.

Bland, J.M. (2004) Cluster randomised trials in the medical literature: two bibliometric surveys. *BMC Med. Res. Methodol.*, **4**, 21.

Bland, J.M. and Kerry, S.M. (1997) Statistics notes. Trials randomised in clusters. *BMJ*, **315** (7108), 600.

Campbell, M., Fitzpatrick, R., Haines, A. *et al.* (2000) Framework for design and evaluation of complex interventions to improve health. *BMJ*, **321** (7262), 694–696.

Campbell, M.J., Donner A. and Elbourne D.R. (eds) (2001) Special issue: design and analysis of cluster randomized trials. *Stat. Med.*, **20** (3), 329–496.

Campbell, M.J., Donner, A. and Klar, N. (2007) Developments in cluster randomized trials and Statistics in Medicine. *Stat. Med.*, **26**, 2–19.

Campbell, M.K., Elbourne, D.R. and Altman, D.G. (2004) CONSORT statement: extension to cluster randomised trials. *BMJ*, **328** (7441), 702–708.

Cappuccio, F.P., Kerry, S.M., Micah, F.B. *et al.* (2006) A community programme to reduce salt intake and blood pressure in Ghana [ISRCTN88789643]. *BMC Public Health*, **6**, 13.

Chuang, J.H., Hripcsak, G. and Jenders, R.A. (2000) Considering clustering: a methodological review of clinical decision support system studies. *Proc. AMIA Symp.*, **2000**, 146–150.

Cochrane, A.L. (1972) *Effectiveness and Efficiency: Random Reflections on Health Services*, Nuffield Provincial Hospitals Trust, London.

Cornfield, J. (1978) Randomization by group: a formal analysis. *Am. J. Epidemiol.*, **108**, 100–102.

Craig, P., Dieppe, P., Macintyre, S. *et al.* (2008) Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ*, **337**, a1655. doi: 10.1136/bmj.a1655

Donner, A. and Klar, N. (2000) *Design and Analysis of Cluster Randomised Trials in Health Research*, Arnold, London.

Donner, A., Birkett, N. and Buck, C. (1981) Randomization by cluster. Sample size requirements and analysis. *Am. J. Epidemiol.*, **114**, 906–914.

Eldridge, S. (2010) Pragmatic trials in primary health care: what, when and how? *Fam. Pract.*, **27**, 591–592.

Eldridge, S., Cryer, C., Feder, G. *et al.* (2001) Sample size calculations for intervention trials in primary care randomizing by primary care group: an empirical illustration from one proposed intervention trial. *Stat. Med.*, **20**, 367–376.

Eldridge, S., Spencer, A., Cryer, C. *et al.* (2005) Why modelling a complex intervention is an important precursor to trial design: lessons from studying an intervention to reduce falls-related injuries in older people. *J. Health Serv. Res. Policy*, **10** (3), 133–142.

Eldridge, S., Ashby, D., Bennett, C. *et al.* (2008) Internal and external validity of cluster randomised trials: systematic review of recent trials. *BMJ*, **336** (7649), 876–880.

Farrin, A., Russell, I., Torgerson, D. *et al.* (2005) Differential recruitment in a cluster randomized trial in primary care: the experience of the UK back pain, exercise, active management and manipulation (UK BEAM) feasibility study. *Clin. Trials*, **2** (2), 119–124.

Feder, G., Griffiths, C., Highton, C. *et al.* (1995) Do clinical guidelines introduced with practice based education improve care of asthmatic and diabetic patients? A randomised controlled trial in general practices in east London. *BMJ*, **311**, 1473–1478.

Feder, G., Griffiths, C., Eldridge, S. *et al.* (1999) Effect of postal prompts to patients and general practitioners on the quality of primary care after a coronary event (POST): randomised controlled trial. *BMJ*, **318**, 1522–1526.

Glasgow, R.E., Vogt, T.M. and Boles, S.M. (1999) Evaluating the public health impact of health promotion interventions: the RE-AIM framework. *Am. J. Public Health*, **89**, 1322–1327.

Godwin, M., Ruhland, L., Casson, I. *et al.* (2003) Pragmatic controlled clinical trials in primary care: the struggle between external and internal validity. *BMC Med. Res. Methodol.*, **3**, 28.

Hayes, J.H. and Moulton, L.H. (2009) *Cluster Randomised Trials*, Chapman & Hall.

Hayes, R.J. and Bennet, S. (1999) Simple sample size calculation for cluster-randomized trials. *Int. J. Epidemiol.*, **28** (2), 319–326.

Health and Social Care Information Centre (2011) General Practice Trends in the UK, http://www.ic.nhs.uk/webfiles/publications/TSC/General_Practice_Trends_in_the_UK.pdf (accessed 20 September 2011).

Horton, R. (2000) Common sense and figures: the rhetoric of validity in medicine (Bradford Hill memorial lecture 1999). *Stat. Med.*, **19**, 3149–3164.

International Committee of Medical Journal Editors (2009) Uniform Requirements for Manuscripts , http://www.icmje.org/publishing_10register.html (accessed 15 May 2011).

Isaakidis, P. and Ioannidis, J.P. (2003) Evaluation of cluster randomized controlled trials in sub-Saharan Africa. *Am. J. Epidemiol.*, **158**, 921–926.

Jolly, K., Bradley, F., Sharp, S. *et al.* (1999) Randomised controlled trial of follow up care in general practice of patients with myocardial infarction and angina: final results of the Southampton heart integrated care project (SHIP). *BMJ*, **318** (7185), 706–711.

Kendrick, D., Illingworth, R., Woods, A. *et al.* (2005) Promoting child safety in primary care: a cluster randomised controlled trial to reduce baby walker use. *Br. J. Gen. Pract.*, **55** (517), 582–588.

Kerry, S.M. and Bland, J.M. (1998a) Analysis of a trial randomised in clusters. *BMJ*, **316** (7124), 54.

Kerry, S.M. and Bland, J.M. (1998b) Sample size in cluster randomisation. *BMJ*, **316** (7130), 549.

Kerry, S.M. and Bland, J.M. (1998c) The intracluster correlation coefficient in cluster randomisation. *BMJ*, **316** (7142), 1455.

Kinmonth, A.L., Woodcock, A., Griffin, S. *et al.* (1998) Randomised controlled trial of patient centred care of diabetes in general practice: impact on current wellbeing and future disease risk. The Diabetes Care From Diagnosis Research Team. *BMJ*, **317** (7167), 1202–1208.

Kirkwood, B.R., Hurt, L., Amenga-Etego, S. *et al.* (2010) Effect of vitamin A supplementation in women of reproductive age on maternal survival in Ghana (ObaapaVitA): a cluster-randomised, placebo-controlled trial. *Lancet*, **375** (9726), 1640–1649.

Lindquist, E.F. (1940) *Statistical Analysis in Educational Research*, Houghton Mifflin, Boston.

Matthews, J.N.S. (2000) *An Introduction to Randomized Controlled Clinical Trials*, Arnold, London.

Medical Research Council (1948) Streptomycin treatment of pulmonary tuberculosis. *BMJ*, **2**, 769–782.

Moher, D., Hopewell, S., Schulz, K.F. *et al.* (2010) CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ*, **340**, c869.

Moore, T.J., Conlin, P.R., Ard, J. *et al.* (2001) DASH (Dietary Approaches to Stop Hypertension) diet is effective treatment for stage 1 isolated systolic hypertension. *Hypertension*, **38**, 155–158.

Murira, N., Munjanja, S.P., Zhanda, I. *et al.* (1997) Effect of a new antenatal care programme on the attitudes of pregnant women and midwives towards antenatal care in Harare. *Cent. Afr. J. Med.*, **43**, 131–135.

Murray, D.M. (1998) *Design and Analysis of Group Randomised Trials*, Oxford University Press, New York.

Murray, D.M., Varnell, S.P. and Blitstein, J.L. (2004) Design and analysis of group-randomized trials: a review of recent methodological developments. *Am. J. Public Health*, **94** (3), 423–432.

Oakeshott, P., Kerry, S.M. and Williams, J.E. (1994) Randomised controlled trial of the effect of the Royal College of Radiologists' guidelines on general practitioners' referral for radiographic examination. *Br. J. Gen. Pract.*, **44**, 197–200.

Pocock, S.J. (1983) *Clinical Trials: A Practical Approach*, John Wiley & Sons, Inc., New York.

Puffer, S., Torgerson, D. and Watson, J. (2003) Evidence for risk of bias in cluster randomised trials: review of recent trials published in three general medical journals. *BMJ*, **327** (7418), 785–789.

Roland, M.O. and Torgerson, D.J. (1998) What are pragmatic trials? *BMJ*, **316**, 285.

Rothwell, P.M. (2005) External validity of randomised controlled trials: 'To whom do the results of this trial apply?' *Lancet*, **365**, 82–93.

Sackett, D.L., Straus, S.E., Richardson, W.S. *et al.* (2000) *Evidence-Based Medicine: How to Practice and Teach*, Churchill Livingstone, London.

Torgerson, D.J. (2001) Contamination in trials: is cluster randomisation the answer? *BMJ*, **322**, 355–357.

Torgerson, D.J. and Torgerson, C.J. (2008) *Designing Randomised Trials in Health, Education and the Social Sciences*, Palgrave Macmillan, Basingstoke.

Underwood, M., Eldridge, S., Lamb, S. *et al.* (2011) The OPERA trial: protocol for a randomised trial of an exercise intervention for older people in residential and nursing accommodation. *Trials*, **12**, 27.

Wexler, J.R., Swender, P.T., Tunnessen, W.W. *et al.* (1975) Impact of a system of computer-assisted diagnosis: initial evaluation of the hospitalized patient. *AJDC*, **129**, 203–205.

Wood, D.A., Kinmonth, A.L., Davies, G.A. *et al.* (1994) Randomised controlled trial evaluating cardiovascular screening and intervention in general practice: principal results of British family heart study. *BMJ*, **308**, 313–320.

World Health Organization (1994) World Health Organization partograph in management of labour. *Lancet*, **343**, 1399–1404.

Zwarenstein, M., Treweek, S., Gagnier, J.J. *et al.* (2008) CONSORT group; Pragmatic Trials in Healthcare (Practihc) group. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. *BMJ*, **337**, a2390.