5

Design

Once investigators have decided to carry out a cluster randomised trial, some of the basic principles of recruitment and ethics have been understood (Chapter 2) and an intervention has been developed (Chapter 3), and any piloting or feasibility work undertaken (Chapter 4), the next step is to consider in more detail how the trial should be designed. The majority of cluster randomised trials involve the comparison of outcomes in two arms of a trial concurrently; one arm which receives an intervention and the other which receives treatment as usual. In a review of cluster randomised trials in primary care, 123 out of 152 trials had this two-arm parallel design (Eldridge *et al.*, 2004). These trials can be completely randomised, or employ a method of allocation to intervention arms in which investigators try to ensure balance between arms in terms of prognostic factors (stratified, minimised or matched designs) or to fix more precisely the number of clusters in each arm for some other reason (using blocking). In Section 5.1 we cover the different methods of allocation in two-arm trials, including whether to use an equal or unequal number of clusters in each arm. If investigators wish to collect information at more points than just the primary endpoint, for example at baseline, this information can be collected from the same participants or different participants, an option unique to those conducting cluster randomised trials; we cover this in Section 5.2. For most readers of this book, these two sections contain what they need to make decisions about the design of their trial. However, in some cases, investigators may wish to have more than two arms in their trial. These types of trial are covered in Section 5.3. Further sections cover less common designs such as crossover (Section 5.4) and stepped wedge and pseudo cluster designs (Section 5.5).

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

5.1 Parallel designs with only two arms

5.1.1 Introduction

Many of the trials we have encountered so far in this book are parallel trials with only two arms; for example, Oakeshott, Kerry and Williams, 1994 (Section 1.3.1), Cappuccio *et al.*, 2006 (Section 1.2). In the review by Eldridge *et al.* (2004), approximately half of the trials with parallel designs were completely randomised, a third were stratified or minimised, and a sixth were pair matched. The following sections cover these three types of randomisation. For those unfamiliar with the general principles of these designs, a straightforward introduction is provided in Roberts and Torgerson (1998) or can be found in most standard textbooks on trials, such as those referenced at the start of Chapter 1.

5.1.2 Completely randomised designs

The principle of complete randomisation is that each cluster has a pre-specified (usually equal) chance of being included in each of the trial arms. In theory this could be achieved by tossing a coin, but this method is not recommended because of the potential for bias, for example if a coin is not perfectly balanced or some tosses of the coin are, for whatever reason, ignored. In practice, randomisation is generally carried out using computers. While many cluster randomised trials have used completely randomised designs in the past, these may become less common as the advantages of stratification (see Section 5.1.4) are being increasingly recognised. The main advantage of the completely randomised design is its simplicity. One example of such a design is a trial of an intervention designed to improve intercultural communication (Table 5.1).

Table 5.1 Educational intervention to improve intercultural communication.

Aim: To assess the effectiveness of an educational intervention on intercultural communication aimed to decrease inequalities in care provided between western and non-western patients
Location and type of cluster: Dutch general practitioners
Number of clusters and individual participants: 38 clusters recruited, and 35 analysed; 986 individual participants analysed
Interventions: (i) Control: (details not specified) (ii) Intervention: educational intervention to patients (12-minute videotape in waiting rooms) and for general practitioners (2.5 days of training spread over 2 weeks)
Primary outcome: Mutual understanding between patients and general practitioners (on a validated scale)
Randomisation: GPs randomly allocated to an intervention or a control arm

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

One disadvantage of a completely randomised design is that imbalances in certain prognostic factors may arise between the trial arms by chance. Although this does not in itself lead to bias, trial results often appear more plausible if the arms look balanced. Peto *et al.* (1977) argues that, in most individually randomised trials of a reasonable size, factors are very unlikely to be imbalanced between intervention arms and a completely randomised design will suffice; while Lasagna (1976) argues that it is worth protecting against extreme imbalance, even if this is very rare. In cluster randomised trials, the argument in favour of using a method more likely to achieve balance between arms is stronger, because there are often relatively few clusters. Imbalance in factors is then more likely by chance; results appear less plausible, and analyses which take account of baseline factors (Sections 6.3.2 and 6.3.3) are statistically less efficient if applied to imbalanced rather than balanced designs. Several methods of allocation are available to try to balance arms with respect to prognostic or other factors. These are often referred to as restricted designs. We describe these designs in Sections 5.1.4–5.1.9. In the next section we focus on how to choose which particular factors to attempt to balance.

5.1.3 Choosing factors to balance in designs that are not completely randomised

The most common factors that investigators try to balance between intervention and control arms are: (i) a feature of the cluster, for example the make-up of staff in a cluster; (ii) cluster size, although often this is not known in advance and a proxy for cluster size must be used; (iii) geographic location, for example region; (iv) a feature of the cluster population, for example a measure of socio-demographic characteristics; and (v) baseline characteristics of trial participants.

Choice of which factors to try and balance depends on which factors investigators think may be most important in influencing outcome or in influencing the effectiveness of the intervention. For an outcome measuring the process of care such as whether patients are referred from primary to secondary healthcare, features of the cluster may be most likely to affect that process.

For a clinical outcome or an outcome measuring other individual patient characteristics or behaviour, for example blood pressure or quality of life, it may be more important to balance by characteristics of patients. It is not possible to balance individual participant factors directly using usual methods of stratification, minimisation or matching within a cluster randomised trial, because in these trials any balancing factors have to be specified at the cluster level. Instead, individual-level data, either routine or collected directly from patients, are summarised at the cluster level and used as the balancing factor. When using routine data, a summary is often obtained from the cluster population or from a population located in the same geographical area as the cluster, even though not all the individuals in the cluster population or area will eventually form part of the actual cluster. For example, in Hampshire *et al.* (1999) socio-demographic characteristics of the practice area were used (Table 5.2).

Table 5.2 Action research to promote child surveillance reviews.

Aim: To assess the benefits of using action research in primary care to increase child surveillance reviews
Location and type of cluster: UK general practices
Number of clusters and individual participants: 28 clusters recruited and analysed; 2015 records of individuals analysed
Interventions: (i) Control: anonymised feedback reports on child health surveillance (ii) Intervention: anonymised reports plus visits by researchers who facilitated practice team meetings
Primary outcome: Return rate of child surveillance reviews
Randomisation: Practices were randomly allocated to the action research arm or a control arm. Stratification factors: (i) single-handed practice versus larger practices (ii) socio-demographic characteristics of the practice area

Note: action research is a process of reflection and progressive problem solving by individuals working in a particular environment to improve their working practices, often facilitated by researchers from outside that environment.
Source: Hampshire *et al.* (1999).

Table 5.3 Diabetes Manual trial: manual and structured care to improve outcomes.

Aim: To determine the effects of the Diabetes Manual on glycaemic control, diabetes-related distress and confidence to self-care of patients with type 2 diabetes
Location and type of cluster: UK general practices
Number of clusters and individual participants: 48 practices were randomised and analysed; 245 individual participants were recruited, and 202 analysed
Interventions: (i) Control: delayed intervention (ii) Intervention: education of nurses, followed by one-to-one structured education of patients via a self-completed manual and nurse support, and audiotapes
Primary outcome: HbA1c level
Randomisation: Practices were allocated in blocks to intervention or delayed intervention arms by a statistician blind to practice identity using computer-aided minimisation
Minimisation factors: (i) Mean HbA1c of consented patients (ii) Number of patients recruited per practice (iii) Practice-level Quality and Outcomes Framework aspirational points score

Source: Sturt *et al.* (2008).

However, to achieve balance in participant characteristics it may be most effective to use a factor which is a cluster-level summary measure of baseline data from actual trial participants. This may be important for the face validity of the trial, but is only possible if patients are recruited prior to allocation. An example of this occurs in the Diabetes Manual trial (Table 5.3). Baseline values of HbA1c were obtained

from trial participants at recruitment; the mean HbA1c level was calculated for each cluster and then clusters were allocated to intervention or control arms using minimisation, with mean HbA1c at the cluster level as one of the minimisation factors. When there is a large pool of potential participants in each cluster, known in advance, from whom a small sample must be drawn to participate in the trial, an alternative method of balancing arms using patient characteristics is stratified random sampling of participants within clusters (Section 5.1.5).

If cluster sizes vary considerably it may be advisable to balance by cluster size, in order to prevent all large clusters being randomised into one intervention arm. Even if cluster size does not influence the outcome or intervention effectiveness, very uneven-sized intervention arms will reduce the power of the study to show an effect. If the main reason for balancing by cluster size is to ensure reasonably balanced arms in terms of numbers, the most effective way of doing this is to balance by the number of participants in each cluster, as in the Diabetes Manual trial (Table 5.3). Nevertheless, as for individual participant characteristics, unless participants are recruited before allocation to intervention arms (or there is no recruitment), the number of individual participants in each cluster may be unknown at the point of allocation. In trials where recruitment has to take place after allocation, investigators can use the number of professionals in a healthcare organisation as a proxy for cluster size, or the natural cluster size (see Section 1.3.2) if this is available.

5.1.4 Stratified designs

In stratified designs investigators choose the factors that they want to balance between the different arms of the trial and the number of levels of each factor. For example, in the trial in Table 5.2, one stratification factor was size of practice: single-handed versus larger practices. This factor had two levels. If investigators wish to balance on the basis of a continuous factor this must be categorised into a number of discrete categories first. Each of these categories then becomes a separate level. Caria *et al.* (2011) describe this process for a trial in which schools in a number of different countries were randomised. The investigators stratified by a social-status index, which they created using different social-status indicators in each country. In each country the index was divided into tertiles, so the overall index of social status was a stratification factor with three levels.

Following identification of factors and levels, clusters are then divided into strata, each containing only clusters with identical levels of each factor, and within each stratum blocking (Section 5.1.8) is used to ensure equal (or as near equal as possible) numbers of clusters from each stratum in each arm of the trial. For example, with 2 factors each with 2 levels, the number of strata needed is 4 (2×2). If one factor has 2 levels and one has 3, then 6 (2×3) strata are needed. Thus, as the number of factors or the number of levels within each factor increases, the number of strata increases. If the number of strata is large compared with the number of clusters this can result in some very small, or even empty, strata. In these circumstances, it is often difficult to achieve balance within strata and the whole point of stratification is lost. There is, however, an alternative to stratification which is better able to cope with several stratification factors: minimisation, which we describe in Section 5.1.6.

5.1.5 Stratified random sampling within clusters

This technique is unique to cluster randomised trials in which participants are sampled from a larger pool of participants in each cluster (the natural cluster –Section 1.3.2), and baseline characteristics of participants that investigators wish to balance can be easily identified in advance of recruiting participants. Such characteristics might be, for example, age, sex, or routinely recorded clinical information. The technique was used in the Kumasi trial in rural Ghana; data were available on age and sex of residents in each village prior to recruitment of residents and randomisation of clusters (Table 5.4). The investigators purposely selected, as far as possible, the same proportions of residents in particular age and sex bands from each village, in order to achieve balance on these characteristics between intervention and control arms. An additional design feature of this trial was blocked randomisation. This was used to ensure that the time between participant and village recruitment and the start of the intervention was as short as possible, and is described in more detail in Section 5.1.8.

5.1.6 Minimisation

Minimisation does something slightly different from stratification (Taves, 1974; Pocock and Simon, 1975). Important prognostic factors are identified at the start of

Table 5.4 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 and individual participants: 12 clusters were recruited and analysed; 1031 individual participants were analysed
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
Randomisation: (i) Blocks of 2 villages were formed (ii) Individuals within each village in the block were stratified by age (four levels) and sex so that as far as possible participants with the same age and sex structure were selected from each village (iii) Villages were randomly allocated to intervention or control arms in blocks of size 2, and allocation was conducted only when subject recruitment was complete in the whole block

Source: Cappuccio *et al.* (2006); Kerry *et al.* (2005).

Table 5.5 Characteristics of first 15 clusters minimised in a cluster randomised trial.

Factor	Level	Intervention arm	Control arm
Natural cluster size (dichotomised)	Large	5*	4*
	Small	3	3
Deprivation score of area in which cluster is located	Deprived	4*	2*
	Less deprived	4	5
System of referral	A	5*	3*
	B	3	4

the trial, and experimental units (in our case, clusters) are assigned sequentially. Each unit's assignment depends on the assignment of previous units and is made so that the imbalance between arms in terms of the prognostic factors is minimised. There are different methods of calculating the imbalance between the arms (Taves, 2010); here we illustrate one method.

Suppose clusters are to be minimized using three factors, each of which has two levels. Table 5.5 shows the numbers allocated to each arm by subgroup after 15 clusters have been randomised. Suppose the next cluster is large, deprived, with referral system A. We have shown the numbers already in these categories with asterisks in the table. These numbers sum to 14 in the intervention group and 9 in the control group. Placing the new cluster in the arm in which the sum is lower (control arm) will result in better balance (the new sums will be 14 and 12 respectively).

Minimisation has advantages over stratification when the number of units to be allocated is small – as in the case of a relatively small number of clusters – but the number of stratification factors is large (Pocock and Simon, 1975; Altman, 1991). Minimisation performs better than some other methods of balancing arms in these circumstances (Scott *et al.*, 2002).

Disadvantages of minimisation are that it is essentially a deterministic method, but statistical analyses of trials assume random allocation; that the allocation of an experimental unit depends on the characteristics of units already allocated and as a result the next assignment can sometimes be predicted; that it may be complex to use; and that it ensures only that *overall* the intervention arms will be balanced for each factor, not that factors in combination will be balanced (as in stratification). An adaptation of the simple deterministic approach assigns a new experimental unit to the arm which would achieve the better balance with a probability of, say, 0.75. Using this method on the example in Table 5.5, the new cluster would have a 75% chance of being allocated to the control arm and a 25% chance of being allocated to the intervention arm. This introduces an element of randomness and avoids the certain prediction of intervention allocation. Minimisation with a random element does not, however, resolve the issues of how to carry out appropriate inference, or what to do about interaction between factors. These disadvantages of minimisation

Table 5.6 ELECTRA: asthma liaison nurses to reduce unscheduled care.

Aim: To determine whether asthma specialist nurses, using a liaison model of care, reduce unscheduled care in a deprived multiethnic setting
Location and type of cluster: UK general practices
Number of clusters and individual participants: 44 clusters were randomised and analysed; 324 participants were recruited, and 319 analysed
Interventions: (i) Control: a visit promoting standard asthma guidelines; patients were checked for inhaler technique (ii) Intervention: patient review in a nurse-led clinic, and liaison with general practitioners and practice nurses comprising educational outreach, promotion of guidelines for high risk asthma, sand ongoing clinical support
Primary outcomes: Unscheduled care for acute asthma over one year, and time to first unscheduled attendance
Randomisation: General practices were allocated to intervention and control arms using minimisation
Minimisation factors: (i) Partnership size (ii) Training practice status (iii) Hospital admission rates for asthma (iv) Employment of practice nurse (v) Whether the practice nurse was trained in asthma care

Source: Griffiths *et al.* (2004).

are partially a consequence of its development for trials in which allocation is necessarily sequential (Senn, 1997). In cluster randomised trials, the characteristics of all clusters are often known at the start of a trial and other methods of achieving balance could be used instead, although this is an area which is, as yet, relatively underexplored. Currently, it is not uncommon for minimisation to be used even when all clusters and their characteristics are known prior to randomisation. In this case the clusters should be minimised in a random order, as for example in a trial of teaching general practitioners to carry out structured assessments of their long term mentally ill patients (Kendrick, Burns and Freeling, 1995).

Minimisation was used in the ELECTRA trial (Table 5.6). Sometimes both minimisation and stratification are used in the same trial. This was the case in the IRIS trial (Table 5.7). Clusters were stratified by region and then minimised on other factors. Twenty-four clusters were recruited in each of two regions. In both trials, the MINIM program was used to perform the minimisation. This program is currently freely available over the web (Evans, Royston and Day, 2004).

A recent review has indicated that minimisation is not always used correctly and in many trials is poorly described (Taves, 2010). For example, in the ELECTRA trial (Table 5.6), a random element was used in the minimisation, but this is not reported, and the factors should have been described as 'minimisation' rather than 'stratification' factors. If using minimisation, investigators should make clear that minimisation was used and describe whether or not a random element was included. and if possible give more detail of the precise method used.

Table 5.7 IRIS: training to increase identification and referral of victims of domestic violence.

Aim: To test the effectiveness of a training and support programme for general practice teams targeting identification of women experiencing domestic violence and referral to specialist domestic violence advocates
Location and type of cluster: UK general practices
Number of clusters: 51 clusters were randomised, and 48 analysed
Interventions: (i) Control: usual care (ii) Intervention: multidisciplinary training sessions in each practice, electronic prompts in the patient record and a referral pathway to a named domestic violence advocate, as well as feedback on referrals and reinforcement over the course of a year. Posters were displayed in the practice and leaflets were available
Primary outcome: Number of recorded referrals of women aged over 16 years to advocacy services based in specialist domestic violence agencies recorded in the general practice records
Randomisation: (i) General practices were stratified by area (ii) Allocated to intervention or control arm using minimisation
Minimisation factors: (i) Proportion of whole-time-equivalent female doctors in the practice (ii) Postgraduate training status (iii) Number of patients registered with the practice (iv) Percentage of the practice population on low incomes

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

5.1.7 Other techniques for balancing factors between trial arms

There are other ways of trying to achieve balance in prognostic factors between intervention groups. Raab and Butcher (2001), for example, calculate an index of balance for every possible allocation of clusters to treatment groups. An allocation can then be randomly selected from those for which the balance is deemed acceptable. Investigators have to decide what is meant by acceptable balance; this may mean considering balance in individual factors in addition to assessing the overall index.

5.1.8 Blocking

Blocking can be used when all clusters are recruited prior to allocation. In this case all clusters are allocated at the same time. Alternatively it can be used when clusters are recruited sequentially. Here we consider both alternatives.

Using blocking, clusters are divided into blocks, and within each block equal (or as near equal as possible) numbers of clusters are randomised to the intervention and control arms. For example, with a block size of eight, four clusters in each block would be allocated to intervention and four to control. Blocking is used in a stratified design (Section 5.1.4) to ensure that equal or near equal numbers are allocated to each intervention arm from each stratum; without blocking, the advantage of stratification is lost. Often the block is the size of the stratum; if the stratum has an odd number of clusters in it then the numbers in each arm will not be exactly equal, but this will usually still be preferable to a completely randomised design. However, to use a design in which block size is equal to stratum size requires that the number in the stratum is known in advance of the blocking. When clusters are recruited and randomised sequentially this will not be the case and smaller blocks must then be used. It is considered good practice to alter the size of the blocks randomly to lessen the chance of a researcher or cluster professional involved in the trial being able to predict allocation of the next cluster through knowledge of block size and previous allocations. This random assignment of block size is often referred to as random permuted blocks.

Blocking may be used to ensure a pre-specified number of clusters in each arm even when there is no stratification for prognostic or other factors. This may be particularly important in a cluster randomised trial if the intervention being evaluated is costly or time consuming. If blocking is not used, and more clusters than expected end up in the intervention arm, this could have serious resource implications for the trial. If all clusters are to be randomised at the same time, one large block could be used, with block size equal to the number of clusters recruited. However, if recruitment and randomisation are conducted sequentially, using one large block will result in the assignment of the last cluster being predictable. To avoid this predictability, smaller block sizes can be used, and all clusters in an individual block randomised at the same time. One example of this is the Kumasi trial (Table 5.4). In this trial the block sizes were small – only two clusters per block. A pair of clusters was recruited, individual participants within each cluster were recruited, and then the clusters were randomised, one to intervention and one to control. Thus the timing of the intervention was balanced in each block and the field workers would not know the allocation of the last two villages when recruiting participants (Kerry *et al.*, 2005).

Small blocks can also ensure that the number of randomisation units being recruited into the intervention and control arms is fairly even throughout the trial. This is useful to avoid the bias that may occur if, for example, a lot more intervention units than control units are recruited early on and there is a secular trend affecting intervention effectiveness, and also facilitates trial logistics by ensuring resources required for the intervention arm can be distributed evenly over the trial duration.

Unfortunately, blocking cannot be used in conjunction with minimisation. In the OPERA trial described in Section 5.1.10, minimisation resulted in a larger number of clusters than expected in the intervention arm, and a consequent drain on the resources of the trial. Carter and Hood (2008) describe an extension of the method of Raab and Butcher (2001) (Section 5.1.7) which incorporates blocking.

5.1.9 Matched-pair designs

Matching, unlike stratification, is seldom used in individually randomised trials. It is used more often in cluster randomised trials because of the greater chance of imbalance in cluster randomised trials in which complete randomisation is used (see Section 5.1.2), and the fact that clusters can often be identified in advance of a trial so it is easier to approach or select pairs. The principle of matching is that pairs of clusters are constructed so that, within each pair, clusters are as similar as possible in relation to factors that might affect the trial outcomes. It is particularly common for clusters such as towns or communities to be matched; for example, Dietrich *et al.* (1998) used matching in a trial in which the intervention was aimed at encouraging sun protection for children, and the clusters were towns. Matching was also used in the COMMIT trial in which villages were randomised in blocks (Table 5.8). Note that the mechanism for allocation in this trial was similar to that used in the Kumasi trial (Section 5.1.5), but in the Kumasi trial blocking in blocks of two was a device to facilitate trial logistics, while in the COMMIT trial matching was used to balance trial arms in terms of prognostic factors. In the Kumasi trial the blocking was ignored in the trial analysis, but in the COMMIT trial it was not.

Although there can be gains in statistical efficiency from matching (Freedman, Green and Byar, 1990; Freedman *et al.*, 1997), there are also disadvantages. Firstly, if one cluster is lost, the matched cluster cannot be used in a matched analysis. Secondly, if the matching is not effective because the matched clusters are insufficiently similar to each other, then matching can result in decreased rather than increased power (Martin *et al.*, 1993), although this can be dealt with in the analysis by breaking the matches (see Section 6.4.1). Finally, it is not possible to calculate the ICC directly from the trial data (Klar and Donner, 1997) unless it is assumed that the intervention effect is constant across pairs – a rather strong assumption. These disadvantages led Klar and Donner to suggest that stratified rather than

Table 5.8 COMMIT: community-based intervention to increase smoking quit rates.

Aim: To assess whether a community-level, multi-channel, 4-year intervention would increase quit rates among cigarette smokers, with heavy smokers (≥25 cigarettes per day) a priority
Location and type of cluster: Communities (20 in United States, 2 in Canada)
Interventions: (i) Control: no intervention (ii) Intervention: each community formed a community board. The intervention focused on public education, healthcare providers, worksite and other organisations, and cessation resources
Primary outcome: Smoking quit rates
Randomisation: Twenty-two communities were matched (in eleven pairs)
Matching factors: (i) Geographic location (state or province) (ii) Size (iii) General socio-demographic factors

Source: COMMIT Research Group (1995).

matched designs may be preferable if investigators are not able to achieve a high degree of matching between pairs.

5.1.10 Unequal allocation to intervention arms

There are a number of reasons for unequal allocation to intervention arms in trials. In cluster randomised trials unequal allocation is not common but, when it does occur, the most common reason is to minimise cost. The implementation of an intervention in a cluster randomised trial is frequently expensive and time consuming, and cost savings can therefore be made by allocating a greater number of clusters to the control arm than to the intervention arm. This was the case in the OPERA trial (Table 5.9), in which the allocation ratio was 1.5 : 1 (control: intervention). Unfortunately, the use of minimisation resulted in an actual allocation ratio of 1.23 : 1, and consequently a greater number of homes in the intervention arm than originally intended.

5.2 Cohort versus cross-sectional designs

A cohort design differs from a cross-sectional design in the specification of which individuals are included in the trial. Most cluster randomised trials have either a

Table 5.9 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 and individual participants: 78 clusters recruited and analysed; 1060 participants recruited
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 outcomes: Prevalence of depression (Geriatric Depression Scale) at 12 months (outcome at one time point only; cross-sectional design), change in depression score at 12 months in all those present at baseline (repeated measures on same participants; cohort design), change in depression score at 6 months in those depressed at baseline (repeated measures on same participants; cohort design)
Design: Stratification by region and then minimisation of homes after recruitment of individual participants in homes at baseline using allocation ratio within the minimisation program of 1.5 : 1 (control intervention). Subsequently further individuals were permitted to join the study

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

cohort or a cross-sectional design, with the latter further divided into single and repeated cross-sectional designs, but sometimes both designs occur within the same trial as we describe later in this section.

In an individually randomised trial, investigators can take outcome measurements on each participant either at the end of the trial, or at more than one time point, for example a baseline measurement followed by the outcome measurement on each participant. The issue of *whom* to take measurements on does not usually arise: measurements are taken on trial participants.

In cluster randomised trials, measurements are taken on individual participants within the clusters recruited into the trial. However, if several measurements are taken over time these can be on the same individuals at each time point or on different individuals at each time point. If repeated measurements are taken on the same individuals at each time point, this is called a *cohort design*. This design is most useful when investigators want to determine how an intervention changes individual-level outcomes. An example of this is the diabetes care from diagnosis trial which evaluated a patient-centred approach to caring for people with diabetes (Table 5.10); consenting patients with newly diagnosed diabetes were followed up for a year, providing baseline and 12-month data on a variety of measures.

If repeated measurements are taken on different individuals at each time point, this is called a *repeated cross-sectional design*. This design is most commonly used when the aim is to determine how an intervention affects some community-level index of health. An example occurs in a trial looking at improving screening for haemoglobin disorders. In this trial a nurse worked within the practice to improve screening (Table 5.11). Screening rates were obtained at baseline and at the end of the intervention period. Clearly it would not have made sense to try and follow up the same individuals for screening after the intervention since they would already have had their screening results from the baseline screening.

Table 5.10 Diabetes care from diagnosis trial.

Aim: To assess the effect of additional training of practice staff in patient-centred care on the current well being and future risk of patients with newly diagnosed type 2 diabetes
Location and type of cluster: UK general practices
Number of clusters and individual participants: 43 clusters recruited, and 41 analysed; 250 individual participants analysed
Interventions: (i) Control: approach to care developed with practices, based on national guidelines and including patient materials (ii) Intervention: as control plus extra training on patient-centred care
Main outcome measures: Quality of life, well-being, haemoglobin A1c and lipid concentrations, blood pressure, body mass index
Data collection: Baseline and one-year data were collected on consenting participants, from clinical notes and by research nurses and project staff

Source: Kinmonth *et al.* (1998).

Table 5.11 Trial to improve screening for carriers of haemoglobin disorders.

Aim: To investigate the effectiveness of improving screening for carriers of haemoglobin disorders in general practice by using a nurse facilitator working with primary care teams and the relevant haematology laboratories
Location and type of cluster: UK general practices
Number of clusters and individual participants: 26 clusters recruited and randomised
Interventions: (i) Control: usual care (ii) Intervention: posters, leaflets, and formal education sessions
Primary outcome: Number of requests for screening tests for haemoglobin disorders
Data collection: The number of requests for screening was obtained from the laboratory at baseline and at the end of the intervention period

Source: Modell *et al.* (1998).

In some trials data may be collected using a repeated cross-sectional design, but there may be considerable overlap between those included at the different data collection time points. This may allow some outcomes to be treated as if they come from a cohort design. For example, in the OPERA trial (Table 5.9), investigators measured depression scores (the outcome) at baseline, and then at 6 months and 12 months after the introduction of the intervention. Because of a relatively high rate of death and movement out of the clusters (residential homes), the residents on whom the outcome was measured at 6 and 12 months were not exactly the same individuals as those on whom baseline measurements were taken, although there was considerable overlap. Two outcomes were prevalence of depression at 12 months and remission of depression at 6 months. For the former, a repeated cross-sectional approach to data analysis was used, with prevalence amongst all residents present in a cluster at baseline included as a covariate in the analysis. For remission of depression, however, a cohort approach to data analysis was taken in which only those who were identified as depressed at baseline *and* still present at the 6-month follow-up were included in the analysis.

In some trials some outcomes may be collected on individual participants in a cohort design and others collected at the cluster level in a repeated cross-sectional design. This was the case in the CATCH trial described in Section 3.3 of Donner and Klar (2000).

While the primary reason for selecting a cohort or cross-sectional design should be related to trial aims, there are other considerations when choosing between the two designs, such as availability of outcome data and likely attrition if a cohort design is chosen. These are discussed in Feldman and McKinlay (1994). A cohort design is potentially more statistically efficient than a cross-sectional design, but Donner and Klar (2000) point out that this advantage may be marginal or non-existent in practice and should therefore probably not be used as a reason for choosing one design over the other.

5.3 Parallel designs with more than two arms

5.3.1 Introduction

Parallel designs with more than two arms are less common than those with two arms. The reason for having more than two arms is usually that investigators wish to evaluate more than one active intervention. The simplest way of including more than two arms is to allocate clusters to one more arm than there are active interventions, so that, for example, with two active interventions to evaluate, a trial would have three arms, with the third arm being the control arm. Alternatively a full factorial design can be used. In this section we discuss both options. All of the methods of allocating clusters to intervention arms described for trials with two arms are available for trials with more than two arms.

5.3.2 Trials with one more arm than there are active interventions

One example of a parallel trial with more than two arms is a trial evaluating whether a nutrition manual introduced into physician practices in the United States could enhance nutrition screening, advice/referral, and follow-up for cancer prevention. One intervention arm received a manual, a second intervention arm received the manual and an interactive tutorial, and a control arm received neither (Table 5.12). This type of design, in which one active intervention arm receives the intervention that was implemented in the other active intervention arm plus an extra element, is not uncommon. Alternatively a three-arm trial may compare three quite different

Table 5.12 Different interventions to evaluate methods of improving nutrition-related behaviour of staff.

Aim: To determine the effectiveness of two strategies for promoting the use of the (nutrition) manual in improving nutrition-related behaviour of physicians and office staff
Location and type of cluster: Family practices in the Unites States
Number of clusters and individual participants: 810 practices recruited, and 755 analysed. No information on numbers of individual participants
Interventions: (i) Usual care (ii) Mailing a manual to a physician in the practice (iii) Providing the manual, and training a physician in the practice using an interactive tutorial
Primary outcome: Adherence to recommendations in the manual
Primary comparison: 'We hypothesized that practices with a physician who participated in the in-person tutorial would engage in more nutrition-related behaviours.'

Source: Tziraki *et al.* (2000).

Table 5.13 ASSIST: different interventions to promote secondary prevention of coronary heart disease.

Aim: To assess the effectiveness of three different methods of promoting secondary prevention of coronary heart disease in primary care
Location and type of cluster: UK general practices
Number of clusters and individual participants: 21 clusters recruited and analysed; 1906 participants analysed
Interventions: (i) Audit of notes with summary feedback to primary healthcare team (audit group) (ii) Assistance with setting up a disease register and systematic recall of patients to general practitioner (GP recall group) (iii) Assistance with setting up a disease register and systematic recall of patients to a nurse-led clinic (nurse recall group)
Primary outcome: Adequate assessment of three risk factors: blood pressure, cholesterol and smoking status at follow-up

Source: Moher *et al.* (2001).

approaches to achieving changes in outcome. This was the case in the ASSIST trial (Table 5.13).

One of the issues with a three-arm parallel trial is deciding on the primary comparison. Often investigators are primarily interested in the difference between the two active interventions; this was the case in the Tziraki *et al.* (2000) trial. If there is no evidence of a difference between the active interventions, a control arm is useful to identify whether neither active intervention had an effect or whether both had similar effects. In the ASSIST trial (Table 5.13), investigators powered their trial to detect a difference between the audit arm and GP recall arm, although they were also interested in the comparison between nurse recall and audit. The rationale for including three arms in this trial was that the interventions evaluated had been tested before but not compared directly, and their cost effectiveness had not been assessed.

In addition to potentially complicating the analysis by having a variety of possible comparisons, an addition of a third arm in a cluster randomised trial will usually increase the size of the trial by about 50%. Given that these trials are often considerably larger and more complex to conduct than individually randomised trials, there must be good justification for using more than two arms, and investigators should think carefully about this if considering such a trial.

5.3.3 Full factorial designs

An alternative to a trial in which different interventions are compared in the way we have just described is a factorial design. In a factorial design randomised units are divided into arms so that each arm receives a different combination of the various potential interventions. In a full factorial design all possible combinations are

represented, but in a fractional factorial design not all possible combinations are represented. Strictly speaking, then, the three-arm trials described in Section 5.3.2 are fractional factorial designs, although this term is almost never used in this context. Recently, fractional factorial designs have been discussed in relation to designing interventions for evaluation in a full trial, and we described these in Chapter 3 (Section 3.6).

The simplest case of a full factorial design and the one most often used in trials, including cluster randomised trials, is a two by two factorial design. In this type of trial there are two interventions; each intervention can be delivered or not. This results in four arms: an arm that receives both interventions, an arm that receives only the first intervention, an arm that receives only the second intervention and an arm that receives neither. These sorts of trials were invented to achieve greater efficiency: potentially, investigators are conducting two independent trials for the price of one. If the two interventions act independently, the sample size needed is equal to that required for whichever intervention evaluation requires the largest sample size.

One example of a factorial cluster randomised trial is a trial investigating the effectiveness of two different interventions aimed at improving attendance for breast screening in general practices failing to meet national targets (Table 5.14). The investigators made a distinction between systematic interventions, such as letters from practitioners, and opportunistic interventions such as prompts to discuss screening when women attended at the general practice. They wished to evaluate the effectiveness and cost effectiveness of both types of intervention, and a factorial trial was an efficient way of doing this.

In the analysis of factorial trials (see Section 6.4.2), comparison is made between those who receive the first intervention and those who do not, ignoring whether or not participants receive the second intervention, and vice versa. The assumption is that whether or not individuals receive the first intervention does not affect the effectiveness of the second intervention and as a corollary, whether or not individuals receive the second intervention does not affect the effectiveness of the first. When this assumption holds good the design is very useful, and investigators do get two trials for the price of one. However, there are situations in which the assumption breaks down, and this can cause problems, which we now describe.

Table 5.14 Two interventions to increase breast screening.

Aim: To examine the effectiveness and cost effectiveness of two interventions based in primary care aimed at increasing uptake of breast screening
Location and type of cluster: UK general practices
Number of clusters and individual participants: 24 practices were recruited and analysed; 6133 women were randomised, and 5732 analysed
Interventions: (i) Control: usual care (ii) General practitioner letter (iii) Flag in women's notes to prompt discussion (iv) Both interventions
Primary outcome: Attendance for screening

Source: Richards *et al.* (2001).

The fundamental problem with a factorial trial occurs when there is an interaction between the interventions being evaluated. Here we illustrate the problem assuming a two by two full factorial trial, but the problem extends to more complex designs. An interaction between two interventions occurs when the effect of one intervention is dependent on whether or not individuals receive the other intervention. There are two different ways that that can happen. An antagonistic interaction means that each intervention works better without the other, so that together they may not do any better than the single interventions by themselves. Alternatively, a synergistic interaction means each intervention is more effective when combined with the other, so that when implemented together they are more effective than might be expected from the action of each separately. One example of a synergistic interaction is a trial evaluating two interventions to reduce blood pressure in Pakistan (Jafar *et al.*, 2009; Table 2.8). In this trial the clusters were communities; one intervention was delivered to households by lay workers and the other intervention was training of general practitioners. When the interventions were delivered together the effect was to decrease systolic blood pressure by 10.8 mm Hg in the combined intervention arm, but by only 5.8 mm Hg in each of the other arms, including the control arm.

Nevertheless, antagonistic interactions are probably more common than synergistic interactions amongst interventions evaluated in cluster randomised trials. Unfortunately, it is antagonistic interactions that cause the more serious problem in factorial trials because they can lead to such trials being underpowered. As an illustration of the effects of an antagonistic intervention, consider a trial in which the mean outcome score for individuals receiving no intervention is 3, for individuals receiving each single intervention 5, and for individuals receiving both interventions 5 (see Table 5.15). In other words, those people who receive both interventions do not see any more improvement in their scores than the people who receive just one or the other. Those receiving neither intervention, however, have a lower outcome score. Based on these results, if investigators were to conduct two separate trials we would expect the difference between the intervention and control arms to be 2 units in each trial. But in the factorial trial, while all of those who receive the first intervention have a mean score of 5, half of those who do not receive this intervention also have a mean score of 5, so the mean difference between those who receive and do not receive the first intervention is no longer 2, but 1. A larger trial is required

Table 5.15 Mean outcome scores for individuals in each arm of hypothetical factorial trial.

	No intervention	Intervention B	Mean score across two arms
No intervention	3	5	4
Intervention A	5	5	5
Mean score across two arms	4	5	

to detect this smaller difference, so the full factorial trial loses some of its advantage. In addition, if there is an antagonistic interaction, investigators need to decide how big this might be before proceeding to calculate the sample size for the full factorial trial. Note that in this example we have illustrated an *additive* interaction on the *linear* scale; other types of interaction are possible.

In drug trials it is plausible that interactions are more common if two drugs act on the same organ in the body. In cluster randomised trials it is often less clear whether interactions might occur, and investigators need to consider this carefully on a trial by trial basis.

Investigators can design trials in which the aim is to detect whether or not there is an interaction, but this requires a much bigger sample size than a trial in which the aim is simply to detect a main effect. As for three-arm trials (Section 5.3.2), ensuring a trial is large enough to detect a realistic interaction is challenging.

Full factorial trials are not possible for all types of intervention. For example, if the aim is to compare interventions that are similar but where one is more intensive than another, it does not make sense to use a full factorial trial. The trial in Table 5.12 is an example of this.

5.3.4 Randomisation at cluster and individual level

Sometimes investigators may wish to evaluate two interventions, of which one must be evaluated using a cluster randomised design, while the other could be evaluated using an individually randomised design. One example of this is the MINT trial that evaluated treatments for whiplash (Table 5.16). An evaluation of the Whiplash Book versus usual advice was carried out using a matched cluster design. Patients with symptoms persisting two weeks after their emergency department attendance were eligible to join an individually randomised trial comparing physiotherapy with further advice. Thus, in this trial only a subset of individuals in clusters were entered into the individually randomised trial, depending on their progress as a result of the cluster-level intervention. In contrast, in some trials the cluster and individual randomisation takes place simultaneously. This was the case in a trial based in residential facilities for older people in which the investigators wished to assess the effectiveness of bright light and melatonin on cognitive and non-cognitive function (Riemersma-van der Lek *et al.*, 2008). Facilities were randomised to bright or dim light, and participants were randomised to melatonin or placebo. In the statistical literature, designs with randomisation at two different levels are referred to as split-plot designs.

5.4 Crossover designs

In a crossover trial all participants (in this case clusters) receive both the active and the control intervention; it is the order in which clusters receive the interventions that is randomised. In the analysis, comparisons are effectively within-participant comparisons; between-participant (cluster) variation is ignored, and as a result

Table 5.16 MINT: two interventions to prevent whiplash.

Aim: To investigate the effectiveness of interventions designed to prevent the chance of developing whiplash syndrome
Randomisation units: UK National Health Service Acute Trust (cluster-level); patients with persisting symptoms (individual-level)
Number of clusters: 12 clusters recruited
Interventions:
Cluster-level: (i) Control: usual advice (ii) Intervention: Whiplash Book
Individual-level : (i) Control: reinforcement of advice given in emergency department (ii) Intervention: physiotherapy
Primary outcome: Neck Disability Index
Randomisation:
Cluster-level: Trusts matched
Matching factors: (i) Number of emergency department attendances per year (ii) Star rating (iii) Ethnic composition of the surrounding area
Individual-level: Eligible individuals stratified by emergency department and allocated in intervention and control arms (members of the same household are assigned to the same intervention, to reduce contamination)

Source: Lamb *et al.* (2007).

Table 5.17 Parenting intervention to improve development in very preterm infants.

Aim: To determine the efficacy of a neonatal parenting intervention for improving development in very preterm infants
Location and type of cluster: UK neonatal centres
Number of clusters and individual participants: 6 clusters were recruited and analysed; 195 babies had data analysed
Interventions: Weekly Parent Baby Interaction Programme (PBIP) sessions during neonatal intensive care unit admission and up to six weeks after discharge
Primary outcome: Bayley scales of infant development
Randomisation: Six clusters, three from each of two regions. For four clusters, clusters (one from each region) were paired on the basis of deprivation indices, and the final pair was formed from the third cluster from each region

Source: Johnson *et al.* (2009).

overall variation is reduced and the power of the trial is increased. Thus, as for a factorial design, the idea behind these trials is to increase the efficiency of the trial or, working on the same principle, to achieve the same power with a smaller sample size. For a general introduction to crossover trials see Sibbald and Roberts (1998).

Crossover cluster randomised trials are not very common. One example is a trial by Johnson *et al.* (Table 5.17): six clinics were arranged into three pairs, and within each pair one clinic was randomised to receive the intervention first, and one to receive the control first. Essentially, then, this trial is both matched and crossover:

quite a complicated design. One of the rationales for this design was to increase precision of effect estimates given the limited number of clusters available. The intervention was aimed at parents of preterm babies.

In both periods of the trial, investigators identified mothers within the clinics and followed them up for two years to look at their outcomes. Mothers in the clinics which were receiving the intervention in that period received training.

The major disadvantage of crossover trials is the risk that whatever happens in the first trial period may carry over into the second trial period. This is referred to as the carryover effect, and it can compromise trial results because those who receive the control in the second period will be benefitting from carryover from the intervention; thus any differences between outcomes measured at the end of the control and intervention periods underestimate the effect of the intervention. This means that crossover trials are unsuitable for interventions that produce irreversible change in the first period; for example, in individually randomised trials a crossover design is no good if the intervention being evaluated cures the disease being investigated. When there is potential for short-term carryover, a washout period is introduced between the two trial periods, long enough for any short-term carryover to have disappeared by the time the second trial period begins.

In a cluster randomised trial there is potential for carryover at both cluster and individual level. Irreversible change at the cluster level can be produced in a number of different ways; for example, via education and training for professionals within the cluster. One would normally hope for at least some irreversible changes in behaviour, however small, as a result of education or training, so a crossover trial would not be suitable for evaluating health professional education. Thus, many of the trials used as examples in this chapter, including ELECTRA (Table 5.6) and the Diabetes Manual Trial (Table 5.3), could not have been designed as crossover trials. In the trial by Johnson *et al.*, carryover at the cluster level was avoided because no professionals in the cluster were involved in the intervention. The intervention was delivered direct to individual participants by research nurses who did not belong to the cluster; once they left the cluster, the intervention left with them.

The aim of the intervention in many cluster randomised trials is to produce irreversible change in individual participant behaviour, thus inducing a carryover effect if the *same* individuals are included in both intervention periods. Carryover can be avoided, however, if a repeated cross-sectional design is used (Section 5.2). This was the case in the Johnson *et al.* trial: it was not the same mothers who had outcomes measured on them during the two intervention periods. The investigators also used a three-month washout period which they describe as being instigated to reduce contamination; in other words they allowed recruits from the first period to be discharged before new recruits in the second period were involved in the trial.

Thus, using professionals external to a cluster to deliver the intervention and using a repeated cross-sectional design will reduce the likelihood of carryover in cluster randomised crossover trials. The decision about whom to use to deliver an intervention should, however, be based on an assessment of what would be most effective, and should reflect the most likely delivery strategy to be used in routine practice if the trial shows evidence of effectiveness.

5.5 Further design considerations

5.5.1 Pseudo cluster randomisation

In pseudo cluster randomised trials the majority of individuals (but not all) in intervention clusters receive the intervention, while the majority of individuals (but not all) in control clusters do not (Borm *et al.*, 2005; Teerenstra *et al.*, 2006). The idea is to reduce recruitment bias (Section 2.3) by not informing professionals within a cluster which arm of the trial they are in; while at the same time lessening the chances of contamination by ensuring that within each cluster the majority of individuals are receiving the same intervention. A pseudo cluster randomised design is only suitable for trials in which the intervention is aimed solely at the individual participants (individual-cluster, see Section 2.2.2). The design cannot be used if clusters or professionals within the clusters receive part of the intervention, because they cannot then be kept uninformed about which arm they are in. Moreover, such designs are only likely to achieve the aim of keeping professionals uninformed of which arm they are in when the cluster size is small enough to prevent them guessing by the ratio of intervention to control participants.

5.5.2 Stepped wedge designs

In stepped wedge designs the trial starts with no randomisation units (in our case clusters) in the intervention arm and ends with all units in the intervention arm. The units are gradually included in the intervention arm over time in a random order. The addition of units into the intervention arm takes place at pre-specified time points. One or more clusters may be added at each time point; usually an identical number each time. The number added will depend on the logistics of the trial. Stepped wedge designs are useful when it is necessary to roll out the intervention to all clusters involved in the trial. This may be because a policy decision has been made to do so or because it is felt that the intervention will do more good than harm so that it would be unethical to withhold it from any clusters. Brown and Lilford (2006) and Hussey and Hughes (2007) provide a good introduction to these designs. At the time of writing this is a developing area and there is currently little consensus on how best to analyse stepped wedge designs.

5.5.3 Equivalence and non-inferiority trials

Most cluster randomised trials are designed as superiority trials, to detect a difference in outcomes between trial arms. However, there are a few trials designed as equivalence or non-inferiority trials (Jaffar *et al.*, 2009; Cleveringa *et al.*, 2010). These types of trial may be particularly useful when trial investigators wish to evaluate reorganisation of services but they require larger numbers than superiority trials.

5.5.4 Delayed intervention

A number of trials in this book use a delayed intervention design, for example, the Diabetes Manual trial (Table 5.3) and the IRiS trial (Table 5.7). In these trials the control arm clusters are offered the intervention once the trial period is over. This design may enhance cluster recruitment; clusters may be more willing to participate if they know they will receive an intervention at some point. On the other hand, these designs require further resources in trials which are often already expensive.

5.6 Summary

In this chapter we have introduced a number of possible designs for cluster randomised trials. Although there are several designs to choose from, for trials in health services research the most popular design remains a two-arm parallel trial in which investigators use a method of trying to balance some specific factors between arms. Some aspects of design, such as whether to use more than two arms or whether to use a cohort or cross-sectional design, are largely dictated by the research question that the investigators wish to answer, but logistical factors and statistical considerations can also be influential. For example, unequal allocation was adopted in the OPERA trial to reduce the workload of those who had to deliver the intervention (Section 5.1.10), and the crossover design in the trial to improve development of preterm infants was adopted to increase the power of the study to detect a significant result in a situation where limited clusters were available (Section 5.4). The influence of logistical factors may be greater in cluster randomised trials, which tend to be larger and more expensive than many individually randomised trials.

References

Altman, D.G. (1991) *Practical Statistics for Medical Researchers*, Chapman and Hall, London.

Borm, G.F., Melis, R.J., Teerenstra, S. *et al.* (2005) Pseudo cluster randomization: a treatment allocation method to minimize contamination and selection bias. *Stat. Med.*, **24** (23), 3535–3547.

Brown, C.A. and Lilford, R.J. (2006) The stepped wedge trial design: a systematic review. *BMC Med. Res. Methodol.*, **6**, 54.

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.

Caria, M.P., Faggiano, F., Bellocco, R. *et al.* (2011) The influence of socioeconomic environment on the effectiveness of alcohol prevention among European students: a cluster randomized controlled trial. *BMC Public Health*, **11**, 312.

Carter, B.R. and Hood, K. (2008) Balance algorithm for cluster randomized trials. *BMC Med. Res. Methodol.*, **8**, 65.

Cleveringa, F.G., Minkman, M.H., Gorter, K.J. *et al.* (2010) Diabetes Care Protocol: effects on patient-important outcomes. A cluster randomized, non-inferiority trial in primary care. *Diabet. Med.*, **27** (4), 442–450.

COMMIT Research Group (1995) Community Intervention Trial for Smoking Cessation (COMMIT): I. cohort results from a four-year community intervention. *Am. J. Public Health*, **85** (2), 183–192.

Dietrich, A.J., Tobin, J.N., Sox, C.H. *et al.* (1998) Cancer early-detection services in community health centers for the underserved. A randomized controlled trial. *Arch. Fam. Med.*, **7**, 320–327.

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

Eldridge, S., Ashby, D., Feder, G.S. *et al.* (2004) Lessons for cluster randomised trials in the twenty-first century: a systematic review of trials in primary care. *Clin. Trials*, **1**, 80–90.

Evans, S., Royston, P. and Day, S. (2004) Minim: Allocation by Minimisation in Clinical Trials, http://www-users.york.ac.uk/~mb55/guide/minim.htm (accessed May 2011).

Feder, G., Agnew Davies, R., Baird, K. *et al.* (2011) Identification and Referral to Improve Safety (IRIS) of women experiencing domestic violence: a cluster randomised controlled trial of a primary care training and support programme. *Lancet*, Oct 12. [Epub ahead of print]

Feldman, H.A. and McKinlay, S.M. (1994) Cohort versus cross-sectional design in large field trials: precision, sample size, and a unifying model. *Stat Med.*, **13** (1), 61–78.

Freedman, L.S., Green, S.B. and Byar, D.P. (1990) Assessing the gain in efficiency due to matching in a community intervention study. *Stat. Med.*, **9** (8), 943–952.

Freedman, L.S., Gail, M.H., Green, S.B. *et al.* (1997) The efficiency of the matched-pairs design of the Community Intervention Trial for Smoking Cessation (COMMIT). *Control. Clin. Trials*, **18** (2), 131–139.

Griffiths, C., Foster, G., Barnes, N. *et al.* (2004) Specialist nurse intervention to reduce unscheduled asthma care in a deprived multiethnic area: the east London randomised controlled trial for high risk asthma (ELECTRA). *BMJ*, **328** (7432), 144.

Hampshire, A., Blair, M., Crown, N. *et al.* (1999) Action research: a useful method of promoting change in primary care? *Fam. Pract.*, **16** (3), 305–311.

Harmsen, H., Bernsen, R., Meeuwesen, L. *et al.* (2005) The effect of educational intervention on intercultural communication: results of a randomised controlled trial. *Br. J. Gen. Pract.*, **55** (514), 343–350.

Hussey, M.A. and Hughes, J.P. (2007) Design and analysis of stepped wedge cluster randomized trials. *Contemp. Clin. Trials*, **28** (2), 182–191.

Jafar, T.H., Hatcher, J., Poulter, N. *et al.* (2009) Community-based interventions to promote blood pressure control in a developing country: a cluster randomized trial. *Ann. Intern. Med.*, **151** (9), 593–601.

Jaffar, S., Amuron, B., Foster, S. *et al.* (2009) Rates of virological failure in patients treated in a home-based versus a facility-based HIV-care model in Jinja, southeast Uganda: a cluster-randomised equivalence trial. *Lancet*, **374** (9707), 2080–2089.

Johnson, S., Whitelaw, A., Glazebrook, C. *et al.* (2009) Randomized trial of a parenting intervention for very preterm infants: outcome at 2 years. *J. Pediatr.*, **155** (4), 488–494.

Kendrick, T., Burns, T. and Freeling, P. (1995) Randomised controlled trial of teaching general practitioners to carry out structured assessments of their long term mentally ill patients. *BMJ*, **311** (6997), 93–98.

Kerry, S.M., Cappuccio, F.P., Emmett, L. *et al.* (2005) Reducing selection bias in a cluster randomized trial in West African villages. *Clin. Trials*, **2** (2), 125–129.

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. *BMJ*, **317** (7167), 1202–1208.

Klar, N. and Donner, A. (1997) The merits of matching in community intervention trials: a cautionary tale. *Stat. Med.*, **16** (15), 1753–1764.

Lamb, S.E., Gates, S., Underwood, M.R. *et al.* (2007) Managing Injuries of the Neck Trial (MINT): design of a randomised controlled trial of treatments for whiplash associated disorders. *BMC Musculoskelet. Disord.*, **8**, 7.

Lasagna, L. (1976) Randomized clinical trials. *N. Engl. J. Med.*, **295**, 1086–1087.

Martin, D.C., Diehr, P., Perrin, E.B. *et al.* (1993) The effect of matching on the power of randomized community intervention studies. *Stat. Med.*, **12** (3–4), 329–338.

Modell, M., Wonke, B., Anionwu, E. *et al.* (1998) A multidisciplinary approach for improving services in primary care: randomised controlled trial of screening for haemoglobin disorders. *BMJ*, **317** (7161), 788–791.

Moher, M., Yudkin, P., Wright, L. *et al.* (2001) Cluster randomised controlled trial to compare three methods of promoting secondary prevention of coronary heart disease in primary care. *BMJ*, **322** (7298), 1338.

Oakeshott, P., Kerry, S.M. and Williams, J.E. (1994) Randomized controlled trial of the effect of the Royal College of Radiologists' guidelines on general practitioners' referrals for radiographic examination. *Br. J. Gen. Pract.*, **44** (382), 197–200.

Peto, R., Pike, M.C., Armitage, P. *et al.* (1977) Design and analysis of randomized clinical trials requiring prolonged observation of each patient. II analysis and examples. *Br. J. Cancer*, **35**, 1–39.

Pocock, S.J. and Simon, R. (1975) Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics*, **31**, 103–115.

Raab, G.M. and Butcher, I. (2001) Balance in cluster randomized trials. *Stat. Med.*, **20**, 351–365.

Richards, S.H., Bankhead, C., Peters, T.J. *et al.* (2001) Cluster randomised controlled trial comparing the effectiveness and cost-effectiveness of two primary care interventions aimed at improving attendance for breast screening. *J. Med. Screen.*, **8** (2), 91–98.

Riemersma-van der Lek, R.F., Swaab, D.F., Twisk, J. *et al.* (2008) Effect of bright light and melatonin on cognitive and noncognitive function in elderly residents of group care facilities: a randomized controlled trial. *JAMA*, **299** (22), 2642–2655.

Roberts, C. and Torgerson, D. (1998) Randomisation methods in controlled trials. *BMJ*, **317** (7168), 130.

Scott, N.W., McPherson, G.C., Ramsay, C.R. *et al.* (2002) The method of minimization for allocation to clinical trials: a review. *Control. Clin. Trials*, **23**, 662–674.

Senn, S. (1997) *Statistical Issues in Drug Development*, John Wiley & Sons, Ltd, Chichester.

Sibbald, B. and Roberts, C. (1998) Understanding controlled trials. Cross-over trials. *BMJ*, **316** (7146), 1719–1720.

Sturt, J.A., Whitlock, S., Fox, C. *et al.* (2008) Effects of the Diabetes Manual 1:1 structured education in primary care. *Diabet. Med.*, **25** (6), 722–731.

Taves, D.R. (1974) Minimization: a new method of assigning patients to treatment and control groups. *Clin. Pharmacol. Ther*, **15**, 443–453.

Taves, D.R. (2010) The use of minimization in clinical trials. *Contemp. Clin. Trials*, **31** (2), 180–184.

Teerenstra, S., Melis, R.J., Peer, P.G. *et al.* (2006) Pseudo cluster randomization dealt with selection bias and contamination in clinical trials. *J. Clin. Epidemiol.*, **59** (4), 381–386.

Tziraki, C., Graubard, B.I., Manley, M. *et al.* (2000) Effect of training on adoption of cancer prevention nutrition-related activities by primary care practices: results of a randomized, controlled study. *J. Gen. Intern. Med.*, **15**, 155–162.

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.

6

Analysis

In Chapter 5 we introduced a range of possible designs for cluster randomised trials. In this chapter we discuss analysis options for these designs. In Chapter 1 we outlined the importance of accounting for the clustered nature of the data in an analysis of a cluster randomised trial. Most analyses of cluster randomised trials now do this. This reflects an improvement over time: in the past many trials were analysed without accounting for clustering, and even now there is substantial variation between disciplines (Figure 6.1). This chapter focuses on analysis, but we begin with a section on data collection and management, an important precursor to any analysis (Section 6.1). We follow this with an illustration of the consequences of not accounting for clustering in the analysis (Section 6.2). Most of the rest of the chapter (Section 6.3) is concerned with the analysis of two-arm parallel trials which are completely randomised, stratified or minimised; as discussed in Chapter 5, most cluster randomised trials in health services research adopt one of these designs. In Section 6.4 we describe analysis options for other designs, in Section 6.5, avoiding bias in analysis of cluster randomised trials by using intention to treat principles (this is not as straightforward as it is in individually randomised trials) and finally Section 6.6 deals with planning analyses for these trials.

6.1 Data collection and management

For cluster randomised trials, data can be collected at the individual level or the cluster level. Consider, for example, a trial in which clusters are general practices and the outcome measure is prescription rates. In the UK, routine data can provide

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