4.1 Randomization and ‘blindness’

Once a potential intervention has been shown to be safe and acceptable for use in humans and the dose schedule established, trials should be conducted to evaluate quantitatively the benefit attributable specifically to the intervention under trial, compared to some other intervention, while attempting to exclude the confounding effect of other variables. The best way to exclude the potential effects of other factors—both those already known to be confounders and also those that are confounders but are not known to be so—is to base allocation decisions as to which intervention is applied to a particular individual, or group, on a random process. Incorporation of randomization into the trial is an extremely important design issue (see Chapter 11).

The randomized intervention trial is as close to a rigorous scientific experimental study involving human beings as it is possible to achieve ethically. The main study design features of a randomized trial are:

1. to avoid bias in assignment to the alternative interventions, all eligible trial participants should be assigned at random to the alternative treatment groups. This involves two steps; the first is selecting participants on the basis of the pre-established criteria for eligibility, and the second is the randomization procedures should ensure that each eligible participant has the same chance of receiving a particular intervention procedure
2. to avoid bias in the assessment of the trial endpoints, whenever possible, the person(s) assessing the outcome measures should not know to which intervention group the participant was assigned (i.e. the assessor should be ‘blind’ to the intervention group)
3. to avoid bias in the behaviour or reporting by the participant, whenever possible, the participant should also be ‘blind’ (i.e. the intervention group assignment should not be known by the participant).

If neither the assessor nor the participant is aware of the intervention allocations, the trial is said to be ‘double-blind’. If only the assessors (or, more rarely, only the participants) are aware of the allocations, the trial is called ‘single-blind’. For
situations in which there is no known effective treatment or preventive method, a placebo of some sort must be used if
double-blinding is to be assured. The ‘double-blind’ approach is the key to the elimination of bias in the assessment of
the impact of an intervention, and, wherever possible, a ‘double-blind’ design should be used. Sometimes it is not
possible because of the nature of the intervention procedure, for example, where participation in health education
sessions is being compared to no intervention, or where cervical surgery is being compared to drug treatment for
cervical cancer. But even if the providers of the intervention must know the assignments, the person who assesses the
trial outcome should be kept 'blinded', if feasible. The more clearly defined and objective the outcome to be measured,
the less critical it becomes to ensure blinding of the assessor. For example, as long as there is complete ascertainment
of all deaths in all arms of the trial, blindness is unlikely to be important in a trial with mortality as the endpoint. Similarly,
the less likely a patient is to be influenced by knowledge of which intervention they have received, the less important
their blinding is.

4.2 Unit of application of the interventions

Different interventions can be applied either to an individual or groups of individuals, such as everyone in a family or
household, everyone working in a particular company, or everyone in the community. The unit for randomization should
usually vary in parallel with this. The choice of the unit for application of the intervention depends upon the nature of the
intervention, the administrative method for its application, and the purpose for which the intervention is being applied. In
statistical terms, the most efficient design, in most circumstances, is to use the individual as the unit of application, and
this should be the design of choice, unless there is good reason for household or community (group) application and
randomization. There are four main reasons for applying an intervention to a group, rather than by individual.

First, group allocation is appropriate when, by its nature, the intervention must be applied to everyone in the group such
as all those living in a geographical area, workplace, school, or community. Examples include most environmental
alterations and many vector control interventions. It also applies to many educational or health promotion interventions
which, although they can be delivered at individual level, are likely to spill over or ‘contaminate’ other individuals living in
the same community.

Second, it may be logistically easier to administer the interventions to groups, rather than on an individual basis.
Sometimes it is administratively simpler and/or more acceptable to randomize by household or village, rather than by
individual. Furthermore, with individual randomization of medications, for example, there may be a risk of individuals
sharing medications within households or villages.

Third, if the purpose of applying the intervention is to reduce transmission of infection by a parasite, for example, the
appropriate unit of application would be the ‘transmission zone’, i.e. the area in which people (and, where appropriate,
vectors and intermediate hosts) may be interacting and sharing a common pool of parasites. Factors of importance in
defining such zones may include the flight range of vectors and the movements of people, vectors, and intermediate
hosts. To reduce interchange (‘contamination’) among transmission zones, it may be useful to have intervening buffer
zones that are not involved in the trial. For many diseases, however, the size of the transmission zones may be difficult
to determine and may vary over time.

Some interventions may be applied to individuals, but with the expectation that there may be an effect on transmission,
through applying them to a high proportion of individuals in the community, that goes beyond the effect that would be
achieved directly within the individuals who received the intervention (for example, through ‘herd immunity’). The extent of coverage required to produce such effects depends upon the epidemiological circumstances, the presence of other control measures, and the type of intervention being introduced. For example, the use of a malaria vaccine to reduce the transmission of malaria in parts of Africa where the disease is ‘holoendemic’ may require so near to complete coverage that such a purpose would not be seriously considered. However, in other parts of Africa where the disease is much less prevalent, achieving high coverage with a highly effective vaccine might be sufficient to interrupt transmission.

For some types of intervention procedures, when the procedure itself provides individual benefit, such as ivermectin in the treatment of onchocerciasis, a further important issue is whether reduction of transmission provides a benefit, in addition to the individual reductions of morbidity/mortality. Trial designs to demonstrate this additional benefit are likely to be complex.

A fourth reason for applying interventions to a group or community as a unit would be for trials involving an intervention of already proven efficacy in individuals, but for which the delivery may be more effectively carried out on a group or community basis. The trial might consist of a comparison of different delivery systems. Generally, the end result desired in this type of trial is based upon cost-effectiveness criteria. Here the question would be whether it is possible to achieve a greater disease reduction for a given expenditure (or alternatively the same disease reduction for less expenditure) by use of a community-based distribution system than by the usual individual distribution methods. Many types of community-based distribution systems require community participation studies. The basic principles involved in community participation studies and in cost-effectiveness studies are described in Chapters 9 and 19, respectively.

When group randomization is adopted, the efficiency of the design can be improved by ensuring that the groups allocated to the different intervention arms are as similar as possible with respect to risk factors for the outcomes of interest, in the absence of the intervention. In other words, there is ‘balance’ between the risks of the outcomes of interest between the trial arms. When there are large numbers of units to be allocated, randomization itself will ensure comparability, but usually when communities or other groups are the units to be randomized, the number of units is relatively small, and randomization may leave considerable differences between the groups in the different arms. Attempts can be made in the analysis to allow for these differences, but the persuasiveness of the results may be reduced if the conclusions depend upon extensive statistical manipulation of the trial results. A more efficient approach to increase the comparability of the groups in the different arms is to stratify the groups into ‘blocks’ having similar underlying pre-intervention risks of the disease outcome in question and to randomize within each block. Stratification should be either in terms of variables which are strongly related to the risk of the outcome under study or in terms of this risk itself. For example, in trials of interventions against malaria in which villages are to be randomized, the villages might be stratified according to their pre-trial malaria prevalence or incidence rates, if such information is available, and the randomization done within each of these strata. An extreme type of stratification is when each ‘block’ includes the same number of groups (for example, villages) as there are arms of the trial, with each village within each ‘block’ having similar malaria rates. One village in each block is then randomly allocated to each intervention (see also Chapter 11, Section 3).

An alternative to stratification, when the number of available units for simple randomization, or even for stratification, is too small, is known as ‘constrained’ or ‘restricted’ randomization. Assume there are 20 villages to be randomized. All possible combinations of ten versus ten villages are evaluated, and only those combinations with good baseline comparability between the two sets of villages are selected. Next, one of the shortlisted combinations is chosen at random, and one of the two sets of ten villages is randomly selected to become the intervention group (Moulton, 2004).
An example of the use of this approach is given in Sismanidis et al. (2008). See also Chapter 11, Section 3.3.

Often, good information on the distribution of the outcome measures will not be available in the trial population. In such circumstances, baseline studies to obtain the required information should be considered. Sometimes, as an alternative, surrogate measures must be used (i.e. measures which are thought to correlate closely with the outcome measures of principal interest). In the absence of detailed data on the population, geographical proximity and socio-economic level may be used as stratification characteristics. Thus, if a small geographical area is chosen as the randomization unit, the total trial area would be divided into regions containing a small number of relatively homogeneous units and, within each region, an equal number of units allocated to each treatment arm.

4.3 ‘Stepped wedge’ design

The issue of the ethics of randomization is presented in acute form in situations where previous studies, perhaps using short-term endpoints or a more intensive intervention than is feasible on a population basis, indicate that the intervention is likely to be beneficial. Withholding the intervention from those in one of the treatment arms for the duration of the trial may then be argued to be unacceptable. Also, some individuals or organizations have an inherent, if irrational, distrust of randomization, worrying that it is ‘experimentation’ (which of course it is!) or even ‘treating humans like laboratory animals’. Such positions can make it impossible for a straightforward RCT design to be accepted. An approach that can be adopted in this situation is the phased introduction of the intervention on a group-by-group basis, until the entire target population is covered. In order to avoid bias, the order in which the groups are given the intervention should be randomized and the number of groups should not be too small—at least six, preferably many more. This approach was first used in The Gambia to evaluate the long-term effects of vaccination against the hepatitis B virus (HBV) (The Gambia Hepatitis Study Group, 1987). A recent example of this design was a trial in Ghana to evaluate the impact on child mortality of treating fever using anti-malarials, with or without also treating with antibiotics (Chinbuah et al., 2012). Other examples are given in Brown and Lilford (2006).

The trial design is illustrated in Figure 4.1. This type of design has been called a ‘stepped wedge’ design. The power of this approach, compared to a simple allocation of groups to one or other treatment arms, is of the order of 75–80%, depending on the number of groups. The same considerations apply to stratification and blocking, as in the static allocation designs.

In the trial in The Gambia, hepatitis B vaccine was introduced into the routine child vaccination programme over a period of 4 years. The order in which the different vaccination teams (there were 17 at the time the trial was planned) began to use the vaccine was random. At the end of 4 years, there was a cohort of children who had received the vaccine and a cohort who had not. These cohorts are being followed to compare the incidence rates of liver cancer and chronic liver disease. At the end of the 4 years, all vaccination teams had started vaccinating children, so subsequent cohorts of children were vaccinated. This phased introduction of the intervention mimicked the way in which many public interventions are introduced, but the key feature of the random order of the introduction of the intervention across the ‘clusters’ (in this case, vaccination teams) brought the crucial benefit of reducing the potential for the trial producing biased results.

Figure 4.1 The ‘stepped wedge’ trial design used to evaluate the impact of hepatitis B vaccination on liver cancer rates in The Gambia.
4.4 Other approaches to allocation of the interventions

The allocation of interventions to individuals based on a random scheme is the best approach to rigorously exclude the potential biasing effects of other factors. However, non-randomized designs are often used. For example, a common approach is the ‘before–after’ or ‘pre–post’ design, in which the incidence or prevalence of the disease under study is compared before and after the intervention has been applied, and an attempt is made to attribute any difference to the effect of the intervention. This approach has important limitations as it may be wrong to assume that, in the absence of the intervention, the disease rate would have remained the same. Many diseases, and especially those of parasitic or infectious origin, vary greatly in incidence and severity from year to year and place to place, for reasons that are incompletely understood. Certainly variations in climate (for example, temperature and rainfall) can have profound effects. Some diseases show marked declines (or increases) over time in some communities (for example, TB and malaria), and sometimes these cannot be predicted in advance, or even related to any obvious specific factor. ‘Before and after’ evaluations of interventions in such situations may be very misleading. Also, it is not uncommon that the methods used to ascertain the trial outcomes change over time, either in terms of the actual data collection method or the person or organization doing the data collection changes, and the two produce systematically different results.

Another commonly employed approach is to apply an intervention in one community, and not in another, and to attribute any difference in disease rates between the two communities as being due to the intervention. This also may be very misleading, as a change may have occurred in one community, but not in the other, for reasons that had nothing to do with the intervention. Random, rather than purposive, allocation of the intervention to one of the two communities does not make any difference to this.

The commonest reason that is advanced for using a non-random allocation between intervention groups is for simplicity of design and administrative ease. Approaches like these also seem easier to explain to officials and to gain public acceptance. The rationale for randomization is difficult to communicate, even to other scientists, but the arguments in favour of randomization, as outlined in Section 4.1, are extremely strong, and failure to accept this approach has frequently led to studies from which erroneous conclusions have been drawn.
There are, however, situations in which allocation cannot be made on a randomized basis. There are occasions when the benefits of an intervention appear so clear that a properly randomized trial cannot be contemplated, or when the intervention or package has already been subjected to randomized trials and is being scaled up under routine conditions. The value of the intervention then has to be assessed by comparison of the situation before and after its introduction, or by the use of case-control studies after the intervention has been introduced (Smith, 1987). Although before vs after studies suffer from the major limitations described earlier in this section, the plausibility of the trial’s conclusions can be increased by trying to rule out alternative reasons why the changes might have occurred (Bonell et al., 2011; Victora et al., 2004). First, if possible, data should be collected on more than one occasion, both before and after the intervention is introduced (sometimes called a time-series study). This allows checking that the outcome of interest was not already declining at the same rate prior to the start of the intervention, and that any decline after the intervention was introduced was consistently present, rather than only there at one time point. Second, a comparison should be made with time trends in disease rates in neighbouring populations where the intervention or package of interventions has not been delivered, and/or in the country or region of the country as a whole. Third, the sharpness with which changes in disease rates take place should be consistent with what might be reasonable to expect from the intervention and related to the speed with which the intervention is introduced over the entire population. Fourth, knowing and recording possible confounding variables in the before and after periods or in the populations being compared in a non-randomized study may also aid interpretation of differences. For example, in a study in which an objective is to reduce transmission of lymphatic filariasis by treating the human population with antifilarial drugs, monitoring the vector population for changes in density and infectivity might be undertaken.

While acknowledging these exceptions to the use of randomization as the basis of allocation, such studies do not have the rigour of a randomized design, and any conclusions drawn from them must be viewed with some caution. It is reasonable to think of there being a hierarchy of evidence from intervention studies, with (1) well-designed and well-conducted RCTs providing the strongest evidence, followed by (2) quasi-experimental studies, in which there is a similar contemporaneous comparison group, but the receipt of the intervention has not been allocated randomly, and then (3) non-experimental designs, in which there is no similar, contemporaneous comparison group such as the before–after, time-series, or after-only designs outlined earlier in this section. Formal guidelines have been developed by the GRADE working group (Guyatt et al., 2008) (<http://www.gradeworkinggroup.org>) to rank the quality of evidence on the effect of an intervention, based on different kinds of study, ranging from the RCTs, which are judged to provide the highest quality of evidence (if properly conducted), through to other kinds of study, providing lower-quality evidence, including observational studies. The WHO has now adopted these guidelines and attempts to undertake a formal grading of the quality of the evidence, with respect to policy recommendations they make regarding specific interventions. The main focus of this book is on RCTs.