11.1: Introduction to randomization, blinding, and coding

As discussed in Chapter 4, the random allocation of participants in a trial to the different interventions being compared is of fundamental importance in the design of investigations that are conducted to produce the highest-quality evidence of any differences in the effects of the interventions. Only if the units to which the interventions are applied (for example, individuals, households, or communities) are randomized between the interventions under study and the study is of a sufficient size is it possible to be confident that differences in the outcome measures of the trial among those in the different intervention groups are due to the effects of the interventions, rather than to underlying differences between the groups. Randomization should ensure that any potential confounding factors, whether known or unknown, are similarly distributed in each of the intervention groups and therefore cannot bias the comparisons of outcome measures between the groups.

Randomization, if done properly, eliminates the possibility of subjective influence in the assignment of individuals to the different intervention groups. Sometimes ‘pseudo-randomization’ methods are employed in trials for reasons of convenience such as alternate assignment of the different interventions to successive trial entrants or allocation based upon the date of birth or date of entry (with, say, one intervention being assigned to those reporting on even dates and another to those reporting on odd dates). However, proper randomization is superior to any systematic method of allocation, and these other methods should be avoided, unless there are very compelling reasons for using them. With systematic allocation, it is possible for the investigator, and sometimes the participant, to know in advance the group to which a participant will be allocated, and this may introduce conscious or unconscious bias into the allocation procedure. For example, such knowledge may affect the investigator’s judgement as to whether or not an individual is eligible for entry into a particular trial. For this reason, it is essential that the randomization is done (or the randomization allocation is revealed to the investigator) only after it has been ascertained both that an individual is eligible for entry into a trial and also that he or she is prepared to participate in the trial, no matter which intervention is assigned.

pointed out, the success of randomization depends on two interrelated processes. The first entails generating a sequence by which the participants in a trial are allocated between intervention groups. To ensure unpredictability of that allocation sequence, it should be generated by a random process. The second process *allocation concealment* shields those involved in a trial from knowing upcoming assignments in advance, so that investigators cannot change who gets the next assignment, potentially making the comparison groups less equivalent and thus biasing the measurement of the effects of the intervention.

In this chapter, various ways are described in which interventions may be randomly assigned among trial participants. The simplest method, if there are two intervention groups, is by using a procedure which is equivalent to tossing a coin to decide the allocation for each individual unit. This can either be done literally, or an equivalent procedure may be simulated using a table of random numbers or by using a computer to generate random numbers, as described in Section 2.1. In large trials, the use of such a simple randomization procedure is highly likely to ensure that there are nearly equal numbers of units allocated to the different intervention groups and the distribution of potentially confounding factors will be similar in all groups. However, if the total number of units in a study is small, such an assignment procedure may result by chance in the compositions of the different intervention groups being markedly different with respect to factors that may affect the outcome measures in the trial, or markedly unequal numbers of participants may be recruited to each intervention group. Such imbalance may arise by chance as, for example, it is possible that, if a coin is tossed ten times, it will come down heads, say, only twice. In fact, the chance that it will come down exactly heads five times and tails five times is only about 25%. For trials involving several hundreds of participants or more, any such imbalance is likely to be small and can be taken into account in the analysis of the trial. In a small trial, imbalance may make the trial more difficult to interpret, and it is advisable to design the randomization procedure to ensure balance. For this purpose, ‘restricted’ or ‘blocked’ randomization (see Section 2.2) can be used to ensure balance in group sizes. Blocked randomization also helps to achieve balance on time sequence and, in multicentre trials, study site. Stratum-matched designs (see Section 2.3) can be employed to produce balance in the composition of the groups, with respect to those variables on which the matching is based.

The techniques described in Sections 2 and 3 may be used whether the intervention is assigned to communities or to individuals. However, when communities are randomized, as in cluster randomized trials, the number of randomization units (communities) may be relatively small (often 20 or less), and more sophisticated methods of randomization have been devised to reduce sources of potential bias in the allocation of interventions in such trials. These methods are summarized in Section 3.

Whenever possible, intervention studies should be both randomized and *double-blind*, i.e. neither the participants nor the investigator should know to which group each participant has been allocated. This guards against biases that may result from knowledge of the intervention affecting the way an individual behaves, is treated, or is monitored during the trial, or assessed during, or at the end of, the trial. Blinding is discussed in Section 4. In Section 5, there is a discussion of coding systems for recording intervention allocation that may be used in trials.