11.5: Coding systems

In some circumstances, it may be necessary to break the intervention code for an individual. This might arise, for example, if a severe adverse event becomes manifest and the treatment for it may be influenced by knowledge of what intervention the individual received. The coding system which is used to record which individuals received which intervention should be designed, such that, if it is necessary to break the code for one individual, the blindness of the investigator, with respect to the interventions received by other trial participants, should be preserved. For example, if one intervention is coded A and the other B, breaking the code for one individual effectively breaks the code for all participants (if the investigator knows who has received A and who has received B). The use of a single code for each intervention is generally a poor design. It is better to have a unique code for each participant and to have a separate list linking participant numbers with the intervention allocated, or to have only a very small number of participants sharing the same code number. For example, in a BCG trial in South India for tuberculosis prevention, ampoules (each containing several doses of vaccine) were packed in boxes of three. Each box held three vials containing one of two different vaccine doses or a placebo preparation. The three ampoules were randomly coded 1, 2, and 3. The vaccine received by a participant was coded in the trial records by a combination of the box number and the ampoule number (Tuberculosis Prevention Trial Madras, 1979). If it had been necessary to break the vaccine code for an individual, it would only have been broken for those participants who received vaccine from the same ampoule in the same box.

The randomization list should usually be prepared in advance of the trial, and the codes assigned by someone other than the PI. If the intervention is a drug or a vaccine, the manufacturer may agree to supervise the packaging and coding, but the allocation procedure should be overseen, and the code should be held during the trial by a disinterested party. Often, the code is held by the data safety and monitoring committee (see Chapter 7, Section 4). It is also worth checking, for a random sample of the drugs or vaccines, that the codes are correct and errors have not been made in the packaging.
5.1 Individual allocations

Suppose two interventions are to be allocated between 200 individuals. A good coding scheme would be to choose 100 random numbers between 1 and 200 and allocate these codes for intervention A, say, and allocate the other 100 for intervention B (there may also be some 'blocking' within the total group of 200, say in blocks of size ten; see Section 2.2). When an intervention is allocated to the 127th patient in the trial, they would be given the drugs in envelope number 127, and this would be noted in their trial record. A master list of the interventions corresponding to each number would be kept in a secure place by a third party not directly connected with the trial. If it were necessary to break the code for an individual patient, the third party could do this without revealing any of the other codes to the investigator. Only at the end of the trial would the list be released to the investigator for the analysis of the results of the trial.

5.2 Group allocations

If a trial involves many thousands of participants, it may be logistically too complicated to allocate a separate treatment code number to each participant, though this will depend upon the circumstances, and, in some cases, having thousands of individual codes poses no problem. An alternative approach is to use a fixed, but not too small, number of codes for the different interventions. If there are $N$ participants in the trial and $C$ codes for the interventions, then breaking the code for one participant would break the codes for $N/C$ in total. For example, the coding system used for a vaccine trial in Venezuela is given in Box 11.2. In this trial, 998 different codes were used (499 for one vaccine and 499 for the other) for about 30 000 participants. Breaking the code for one individual would break it for about 30 others (Convit et al., 1992).

A simpler system might be required if participants had to be given the same intervention on a number of occasions. A method that was used in a trial of ivermectin against onchocerciasis in Sierra Leone was to allocate 20 codes for ivermectin or placebo treatments (A, B, C, D, and so on) (Whitworth et al., 1991). The drugs were taken to the field in 20 tins, with the code letters on them (ten of which contained ivermectin, and ten contained placebo tablets), and participants were allocated to one of the 20 codes at random. If a participant was allocated, say to code E, then each time they were treated, the dose was taken from tin E. About 1000 patients were included in the trial, so that breaking the code for one individual would have also broken it for $1000/20 = 50$ others. A similar system was used in a trial of a pneumococcal vaccine in The Gambia, which involved many thousands of participants, and each participant was scheduled to receive three doses of the vaccine at different times (Cutts et al., 2005).

With either individual or group allocations, it is helpful if the intervention codes are on removable sticky labels that can be affixed to an individual’s form, thus minimizing the likelihood of recording errors. Where possible, the coding system should be devised so that transcription errors in recording may be detected. How this was achieved in the leprosy vaccine trial in Venezuela is illustrated in Box 11.2. More commonly now, bar codes are used to identify interventions in trials using drugs or vaccines, and, provided that suitable computer systems are set up, this should eliminate the possibility of transcription errors.

Box 11.2 Assignment of check letter for three-digit vaccine code

The coding system described was that used in a leprosy vaccine trial conducted in Venezuela (Convit et al, 1992).
Randomization was to one of two vaccines.

The vaccine vials were labelled with a number between 1 and 998. A total of 499 of these numbers were allocated at random for one vaccine, and the other 499 for the other vaccine. A check letter was added to each number, so that transcription errors would stand a high chance of being detected. The code was devised, such that every possible permutation of the same three digits in a number had a different check letter, as illustrated:

In some countries, number 1 and number 7 are distinguished clearly when written, as it is the custom for the number 7 to have a horizontal stroke put through it. In other countries, however, this is not the custom, and there is a danger that these numbers will be confused. In such cases, it would be advisable to change the check coding system, such that, if a 1 is confused with a 7, or vice versa, the check letter will enable the error to be detected. Thus, the system outlined might be modified, as indicated:

```
001A 010B 100C
002D 020E 200F
   .   .     .
   .   .     .
009M 090N 900P
```

010B—already allocated—see line 1

```
011R 101S 110T
   .   .     .
   .   .     .
123W 132X 213Y 231A 312B 321C
124D 142E 214F 241G 412H 421J
```

```
001A 010B 100C 007D 070E 700F
002G 020H 200J
  .   .     .
  .   .     .
003K   .     .
`
Source: data from Peter Smith, (personal communication).