Randomisation in Clinical Investigations

Steven A. Julious and Jenny V. Freeman

Introduction

In previous tutorials, we have outlined methods for describing and summarising data, and the principles of hypothesis testing and estimation.[1,2] In this tutorial, we will describe the basic concepts of randomisation in investigations. We will begin by describing the background, continue by describing the rationale for randomisation, and then finally move on to some of the more advanced topics of randomisation pertinent to imaging investigations.

Background

Allocation at random has been a central tenet of clinical trial design since the first reported modern clinical trial was conducted to investigate the effect of streptomycin and bed rest compared to bed rest alone in the treatment of tuberculosis.[3-6] Randomisation is important, as it ensures that the regimen groups being investigated are objectively the same with regard to any demographic or prognostic factors. Randomisation achieves this by ensuring that each subject has a known chance of receiving a given treatment in an allocation that cannot be predicted.[7] This lack of predictability is important, as an investigator should remain masked to the order of the treatments in order to reduce the potential for bias; they only find out what regimen a patient is to be assigned to after recruiting a patient into the trial.[8]

Note that the concept of randomisation originally came from clinical trials; hence the reference to treatments. As we will describe in this note, however, randomisation is an important consideration for all types of clinical investigation. The problem of not allocating at random is evidenced by the following example.[9] A historical trial was undertaken to compare the success of a new treatment (percutaneous nephrolithomy) with that of an existing treatment (open surgery) in the removal of kidney stones. From table 1a, it appears that the new treatment is superior, with an 83% success rate compared to only 78% on the old treatment. However, when we break the table down into small (table 1b) and large (table 1c) stones, the direction of the effect first observed is reversed. The old treatment is superior for both sizes of stone. The only reason why the old treatment seemed inferior to start with was that treatment is confounded with stone. This reversal effect is known as Simpson's Paradox.[9,10]

Table 1. A comparison of the success rates of percutaneous nephrolithotomy (New) compared to open surgery (Old) in the removal of kidney stones (a) overall, (b) for stones < 2 cm and (c) for stones ≥ 2 cm

(a) Overall

Treatment	Success		Total
	Yes	No	
New	289 (83%)	61 (17%)	350
Old	273 (78%)	77 (22%)	350
Total	562	138	700

(b) Stones < 2 cm

Treatment	Success		Total
	Yes	No	
New	234 (83%)	36 (17%)	270
Old	81 (93%)	6 (7%)	87
Total	315	42	357

(c) Stones ≥2 cm

Treatment	Success		Total
	Yes	No	
New	55 (69%)	25 (31%)	80
Old	192 (73%)	71 (27%)	263
Total	247	96	343

Confounding is a statistical term for when there is a strong relationship between a third factor and both the outcome and comparison of interest. In table 1, people who had percutaneous nephrolithomy were also more likely to have small stones, and the smaller the stone, the better the prognosis. Hence, we would say that treatment is confounded with stone size.

Obviously, a bias of the magnitude observed with instances of Simpson's Paradox is rare, but randomisation protects the investigator from confounding with known and unknown prognostic factors. Therefore, wherever possible, subjects should be assigned to investigations at random. If there are known factors that could affect the outcome, such as centre, age, sex or baseline risk, then the study should be stratified to allow for these, and a block size (see later) should be set that provides balance within each stratum (see later). If there is to be a constraint in the randomisation, such as unequal allocation, then this should be allowed for in the block size and appropriate adjustment made to the sample size. Block size and strata are described now.

Mechanics of randomisation

Parallel group trials

A parallel group trial is one in which there are to be at least two arms to be investigated and subjects are to be randomised to each of these arms. It is beyond the scope of this note to describe in detail how to undertake a randomisation; however, we will give some general hints and tips (used in this context, arm is a generic term to describe the groupings in trials; subjects may be assigned to two different arms, where these arms could be treatments; assessors or imaging protocols).

When randomising subjects to the different arms in the trial, an important consideration is to maintain balance for the interventions to which subjects are being randomised. This is particularly important in small studies, where by chance there can easily be an imbalance between the numbers of subjects on the respective arms. One way to ensure that groups are balanced is to introduce 'blocks' into the randomisation. Basically, a block is a sample size after which there is balance in the randomisation. It is best to illustrate this through a simple worked example.

Consider the case of two groups. We wish to randomly allocate individuals to either group A or group B. In this example, we could toss a coin and record either heads (H) or tails (T), so that we can then use the order to allocate individuals to groups (i.e. if heads, then group A; if tails, then group B). If we set the block size to be four, we need to ensure that after every four tosses there are two heads and two tails. Thus:

Block 1: T T (H H)

Block 2: T T (H H)

Block 3: T H T (H)

Block 4: T H H (T)

The terms in parentheses are not from tosses but entries that we were forced to enter to ensure balance. For example, in block 1 the first two tosses were tails. We thus made the next two heads, so that after 'four tosses' we had a balance. Note that after '16 tosses' by blocking, we have eight heads and eight tails.

Another important consideration is stratification. Stratification is similar to blocking, but as well as ensuring balance after a requisite block size we also ensure balance by strata. These strata are usually clinically important subgroups such as sex and age.

Again, it is best to illustrate this by example. Suppose that we are doing the same coin tossing to create a randomisation list. For this randomisation, we wish to ensure balance for a two-level stratification factor. Operationally, this would be the same as doing the coin-tossing exercise twice: once for each stratum.

Stratum 1:

Block 1: T T (H H)

Block 2: T T (H H)

Stratum 2:

Block 1: T H T (H)

Block 2: T H H (T)

Now, after '16 tosses' we have balance both in terms of heads and tails and in terms of heads and tails by strata.

A final consideration, as discussed earlier in this note, is the withholding of the randomisation until the actual allocation of subjects. Even for completely open studies, it is preferable to mask the randomisation so that investigators only find out what regimen a patient is to be assigned to after the patient has been recruited. In practice, this could be done by putting the randomisation in envelopes that are opened only after a subject has been enrolled.

Crossover trials

The distinction between parallel group designs and crossover designs is that in parallel group designs subjects are assigned at random to receive only one investigation, and as a result of the randomisation the groups are the same in all respects other than the investigation being made. However, with a crossover trial, all subjects receive all the investigations but it is the order in which subjects receive the investigations that is randomised. The big assumptions here are that prior to starting each investigation all subjects return to baseline, and that the order in which subjects have their investigation does not affect their response to the investigation.

Two period

Two-period crossover trials are the easiest to explain. In the simplest case, for a two-arm investigation (comparing A with B, say), subjects will be randomised to either A followed by B (AB) or to B followed by A (BA). AB and BA are called sequences and represent the order in which subjects receive the investigations. In practice, subjects are randomly assigned to either the sequence AB or the sequence BA, and to ensure balance, blocking can still be used.

Note that even for retrospective investigations, randomisation should be considered. For example, in a study to investigate the agreement between two image analysts, the analysts could have the images assessed randomly, with the analysts reading the images in random order much like an AB/BA design.

Multi-period

All investigations are made on all subjects

Imaging comparisons can be complicated, as there is often a finite number of subjects on whom a number of investigations are made, such as:

- a functional magnetic resonance imaging (fMRI) investigation in which subjects will receive a number of challenges
- a comparison of different imaging protocols within the same subject
- an assessment of new technology such as a comparison of 2D, 3D and 2D and 3D combined SPECT
- a comparison of several readers on the same subjects, to look at agreement
- a comparison of different therapies or different doses of the same therapy within a subject.

It is quite easy, therefore, for four or five investigations to be made on the same subject. If four investigations are made on the same subject, it will result in 24 different ways of assigning subjects to these four investigations, and hence 24 sequences. This is all very good, but what if we have only 12 subjects in the trial?

Actually, for multi-period investigations, we do not necessarily need to use all possible sequences but can form special sequences to be randomised, called Williams squares.[11]

It is again best to illustrate through example. In order to investigate an even number of investigations, we can build a Williams square from the following sequence:

0, 1, t, 2, $t - 1$, 3, $t - 2$, etc.

where t is the number of interventions minus 1. If we were to conduct four investigations, then t would be 3, and our sequences would include 0, 1, 2, 3. We build the sequences by forming the first row from the result above. We then form the second by adding 1 to this first row, but where the number is 3, the new number becomes 0 (we are adding in base 3). The calculation is simpler than the explanation.

0	1	3	2
1	2	0	3
2	3	1	0
3	0	2	1

This is known as a Latin square: each investigation appears in every row and column. The columns here would reflect different imaging sessions. A Williams square is a special form of Latin square, such that as well as being balanced for rows and columns, each investigation is preceded by each other investigation at least once, e.g. 1 is preceded by 0, 2, and 3. Here we are saying that as well as the order of investigations being important, the effect of preceding investigations is too. Hence we ensure balance for the immediately preceding investigation. This is known as first-order balance.

If we were conducting a trial in which we were undertaking four different investigations on 12 subjects, we would randomise the four sequences above so that each sequence appeared three times.

For an odd number of investigations, we need to build two Latin squares with starting sequences

0, 1, t, 2, $t - 1$, 3, $t - 2$, etc.

and

… $t - 2$, 3, $t - 1$, t, 1, 0.

With five investigations, $t = 4$, and we would therefore have

0	1	4	2	3
1	2	0	3	4
2	3	1	4	0
3	4	2	0	1
4	0	3	1	2

and

3	2	4	1	0
4	3	0	2	1
0	4	1	3	2
1	0	2	4	3
2	1	3	0	4

Not all investigations are made on all subjects

In imaging investigations, there are logistical, practical and safety considerations to be taken into account. For example, we may wish to investigate four different imaging protocols, but these must all be done in one day for each subject, and for practical reasons we can only schedule three scans in a day. Similarly, we may wish to look at four

different protocols but, for safety reasons, we may only be able to do three scans in the 24 hours we have for each subject. Although we can still construct Latin squares, we need to construct a special type of these, known as a BALANCED INCOMPLETE BLOCK. Again, we will illustrate by example.

If we could have three sessions for each subject but we have four investigations, then taking the sequences derived previously and removing the first column

0	1	3	2
1	2	0	3
2	3	1	0
3	0	2	1

and the final column

0	1	3	2
1	2	0	3
2	3	1	0
3	0	2	1

would give us eight sequences as follows:

1	3	2
2	0	3
3	1	0
0	2	1

0	1	3
1	2	0
2	3	1
3	0	2

We would hence have balance for both rows and columns, as well as first-order balance within eight sequences.

For an odd number of sequences we use a similar procedure. Using our previous example of having five investigations and only being able to do three sessions, we could delete the last two columns from the first five sequences and the first two columns from the next five sequences, i.e.

0	1	4	2	3
1	2	0	3	4
2	3	1	4	0
3	4	2	0	1
4	0	3	1	2

3	2	4	1	0
4	3	0	2	1
0	4	1	3	2
1	0	2	4	3
2	1	3	0	4

Discussion

In this tutorial we have introduced the basic concepts of randomisation, including the importance of stratification and blocking. We have described issues pertinent to imaging investigations where we may wish to perform multiple investigations on each subject or the special case where the number of investigations is greater than the number of sessions.

References

1. Freeman JV and Julious SA. Hypothesis testing and estimation. *Scope* 2006; 15(1):18-21.
2. Freeman JV and Julious SA. Describing and summarising data. *Scope* 2005; 14(3):20-23.
3. Medical Research Council. Streptomycin treatment of pulmonary tuberculosis. *BMJ* 1948; 2: 769–82.
4. Julious SA and Zariffa N. ABC of pharmaceutical trial design. *Pharmaceutical Statistics* 2002; 1(1): 45–54.
5. Yoshioka A. Use of randomisation in the Medical Research Council's clinical trials of streptomycin in pulmonary tuberculosis in the 1940s. *BMJ* 1998; 317: 1220–3.
6. Bradford Hill A. Memories of the British streptomycin trial: the first randomized clinical trial. *Controlled Clin Trials* 1990; 11: 77–9.
7. Altman DG and Bland JM. Treatment allocation in controlled trials: why randomise? *BMJ* 1999; 318: 1209.
8. Day SJ and Altman DG. Blinding in clinical trials and other studies. *BMJ* 2000; 321: 504.
9. Julious SA and Mullee MA. Confounding and Simpson's paradox. *BMJ* 1994; 308: 1480–1.
10. Simpson EH. The interpretation of interaction in contingency tables. *J R Statistic Soc* B 1951; 2: 238–41.
11. Williams EJ. Experimental designs balanced for the estimation of residual effects of treatments. *Aust J Sci Res Ser*. A 1949; 2: 149–68.