Randomisation in clinical trials

Robert D Herbert
School of Physiotherapy, University of Sydney

Randomised trials provide a way of determining the effects of health interventions. In the simplest randomised trials, some subjects receive an intervention and others do not. The difference in outcomes of treated and untreated (control) subjects provides an estimate of the effects of the intervention.

A defining characteristic of randomised trials is that the decision about who receives the intervention is made randomly. That is, a random process is used to allocate subjects to treatment and control groups.

Random allocation implies that it is not possible to predict which group an individual subject will be allocated to. But chance is full of surprises. Paradoxically, in the long run, chance behaves predictably: the process of randomly allocating a large series of subjects generates comparable groups. Groups formed by random allocation tend to be similar with regard to the severity and chronicity of the subjects' disease, and subjects' attitudes and experiences and social backgrounds. In fact the groups tend to be similar in all (known and unknown) factors that are likely to influence outcomes. Randomisation is the only mechanism that can be relied upon to produce groups that are similar in all respects.

This similarity is important. It enables us to make inferences about the effects of intervention, uncontaminated by other factors. When treated and control groups consist of similar sorts of subjects, it is fair to compare the outcomes of the two groups. If one group is treated and another group is not, and this is the only difference between the two groups, then differences in their outcomes must be due to treatment.

Randomisation generates groups that are similar, but not identical. The degree of similarity depends on the number of subjects that are randomised. When small numbers of subjects are randomised, there is a greater probability that there will be important differences between groups. (Then we say there is an 'imbalance at baseline'.) But when large numbers of subjects are randomised to groups the differences between groups at baseline tend to be small. And, large or small, differences at baseline in randomised trials are due only to chance, which means that statistical methods can be used to deal with them.

Statistical models can be used to estimate the probability with which certain differences between groups will occur. This is exploited in the analysis of trial outcomes. The standard approach to analysis assumes that there can be only two explanations for differences in the groups' outcomes: either the difference is due solely to an imbalance of randomisation, or there is some effect of the intervention. The logic of hypothesis testing is that if the differences between groups are large enough that imbalances of that magnitude are unlikely to occur then the effect must be due to the intervention.

Methods of randomisation

Simple randomisation Randomisation can be achieved by simple processes such as coin-tossing ('heads = control, tails = treatment') or drawing lots. These methods can generate allocation schedules that are essentially random. However they are open to abuse (a researcher might decide he or she doesn't like the toss and toss the coin again), and they are difficult to audit, so they are not the best way of randomising subjects.

A better way to randomise involves using a computer to generate a random number for each subject in the trial. Then a simple rule is used to assign the subject to a group on the basis of the random number. (For example, it might be that all subjects assigned random numbers in the lower half of the random number distribution are allocated to the control group and all subjects assigned random numbers in the upper half of the random number distribution are allocated to the treatment group.)

Simple randomisation is perfectly adequate for almost all trials. However many researchers use more sophisticated methods of randomisation, some of which are described in the remainder of this Research Note.

Imbalance in sample size and blocking Simple randomisation does not ensure that the groups will be the same size. For example, in a trial of 100 subjects there is only about an 8% chance that simple randomisation will allocate exactly 50 subjects to each group, and there is about a 5% chance that the imbalance in subject numbers will be greater than 60:40 (Piantadosi 1997).

Large imbalances in sample size may appear a bit alarming, but they are of little consequence. The only real cost of an imbalance in sample size is that it produces a small reduction in statistical power. When the sample size is unbalanced, each subject effectively provides slightly less information than when the sample size is perfectly balanced. A trial with a 60:40 imbalance in sample size will have the same power as a balanced trial with 13% more subjects.

Some researchers would prefer to randomise in a way that ensures balanced sample sizes. This is called 'blocked' allocation. Blocked random allocation involves first setting up a 'block' containing equal numbers of treatment and control group allocations and then assigning each subject an allocation (treatment or control) that is randomly selected from the block. For example, in a trial of 100 subjects the researcher might divide the block of 100 allocations into 50 treatment group allocations and 50 control group allocations. Subjects would then be assigned a randomly selected allocation. This procedure will generate balanced sample sizes provided the trial terminates as planned at 100 subjects.

Blocking is analogous to drawing allocations out of a hat containing equal numbers of treatment and control group allocations without returning allocations to the hat after they are drawn. But again, generation of a blocked random allocation schedule is best done by computer.

One limitation of this approach is that sample size will only be balanced if the trial runs its full course. If for any reason it is necessary to terminate the trial early the sample sizes may be unbalanced. To protect against this contingency, some researchers randomise multiple blocks. For example, in a trial of 60 subjects there could be 10 consecutive blocks of 6, each containing 3 allocations to the control group and 3 allocations to the treatment group. If the trial is stopped early it may be possible to stop at the end of a block, in which case the sample sizes in the two groups will be balanced. Even if it is not possible to stop at the end of a block, the maximum imbalance of sample size is just half the block size.

The use of multiple blocks introduces another problem. It is important that the person who recruits subjects into the trial is not aware of the group to which the next subject is to be allocated. (This is called 'concealment'; Schulz and Grimes 2002.) But blocking introduces patterns into the allocation schedule which might, in some circumstances, allow the recruiter to predict some allocations. To avoid this problem, some researchers vary the block sizes. Varying block sizes at random ('randomly permuted blocks') reduces the predictability of the allocation schedule.

Imbalance in prognostic factors and stratification Blocking ensures balance in sample size, but it does not ensure balance in prognostic characteristics. With simple random allocation or blocked random allocation it is possible that there will be moderate differences, and occasionally even large differences, in prognostic factors between groups. Some researchers use statistical tests to determine if there are baseline differences between groups, and some use statistical techniques (such as ANCOVA) to adjust for statistically significant baseline imbalances. But it is to be expected that there will be some statistically baseline differences between groups at baseline. It is not logical to test for baseline imbalances; baseline imbalances do not produce bias, and correcting for them can introduce bias.

As Pocock et al (2002) and others have pointed out, it is illogical to use statistical tests to detect baseline imbalances in prognostic variables. Hypothesis tests tell us if the differences between groups are larger than those expected to occur by chance when we randomly allocate subjects to groups. But we *know* that baseline differences between groups in randomised trials occurred by chance, because we know subjects were randomly allocated to groups — in randomised trials all statistically significant differences at baseline must be statistical errors. (In statistical parlance, they are 'Type I errors'.) The probability of making statistical errors when testing for baseline imbalances will often be high. If, for example, a researcher tests for imbalances in 10 baseline variables, the probability of finding an imbalance can be as high as 40%.

There are good reasons to ignore baseline imbalances of prognostic variables in randomised trials. First, large and overt imbalances in one direction on one prognostic variable may be balanced by small and covert imbalances in the other direction in other variables. So apparent imbalances in prognostic factors may be illusory. Even when apparent imbalances are real (so that one group truly has a better overall prognosis than another) they need not be of concern because they are fully accounted for in the usual statistical testing procedures. Statistical tests tell us if the apparent effects of treatment are big enough that they probably would not be produced by the randomisation process alone. In fact attempts to deal with baseline imbalances, using statistical techniques such as ANCOVA, can be counter-productive. The process of identifying baseline imbalances and adjusting for them can produce (not reduce) bias, and can reduce statistical power (Schluchter and Forsythe 1985, Raab et al 2000).

The preceding paragraphs argue that, provided there was some form of random allocation of subjects to groups, there is little reason to be concerned about baseline imbalances in prognostic factors. Nonetheless there are methods of randomising that are specifically designed to minimise baseline imbalances. The advantage of minimising baseline imbalances (before they happen, rather than correcting for them after they occur) is that this can increase statistical power. The simplest way to do this is with stratification.

Stratified random allocation involves first identifying important prognostic factors and then separately randomising blocks containing different levels of the prognostic factor. The prognostic factors that are most commonly stratified are disease severity and, in multi-site, trials, the site at which the subject is treated. For example, in the trial by Bø and colleagues (1999) of interventions for stress urinary incontinence, the researchers separately randomised one block containing only subjects with mild incontinence (pad test < 20 g) and another block containing only subjects with more severe incontinence (pad tests > 20 g). The effect was to ensure balance in the stratified factor at baseline. Bø and colleagues could know that all of the experimental groups (there were three groups in this trial) contained equal numbers of subjects with mild incontinence, and all three groups contained equal numbers of subjects with more severe incontinence.

While stratified allocation potentially increases statistical power, there are some practical disadvantages. First, it is generally only possible to stratify by one (or perhaps in big trials, two or three) prognostic factors, yet there may be several important prognostic factors. It is usually not practically possible to stratify by all important prognostic factors. (This means that stratification, on its own, is a very poor way of assembling comparable groups. It is the randomisation that makes groups comparable; stratification simply augments randomisation by ensuring balance of one or a few key prognostic variables.) Another limitation is that stratification must be blocked if it is to be useful (Lavori et al 1983). It is difficult to match block sizes so that each stratum fills at approximately the same time. So there is the risk that, as the trial nears completion, the researchers will have to discontinue recruitment into one stratum while they wait for another stratum to fill up. This is particularly a problem when some strata are very small. (Small strata often occur when stratification is by site, because some sites may expect to recruit only small numbers of subjects.)

Adaptive allocation Adaptive randomisation schedules adapt as the trial proceeds. With adaptive allocation, each subject's allocation depends on the baseline imbalances that exist up to that point in the conduct of the trial; the subject's allocation is such that it is likely to reduce any imbalances that exist. The most widely used method of adaptive allocation is called

'minimisation'. The essential feature of minimisation (there are different types of minimisation) is that each subject is allocated to the group which minimises some measure of overall imbalance, so it is not strictly a randomised procedure. Minimisation produces tighter control of baseline imbalances than simple randomisation or stratified random allocation (Scott et al 2004), although it can still only balance known and measurable prognostic factors. It is a little more complex to implement than stratified random allocation, but it does not suffer from the practical disadvantages of stratified allocation that are described above.

Summary

Randomisation provides a logical basis for making inferences about effects of intervention from clinical trials. Simple randomisation procedures are sufficient for this purpose. More complex randomisation procedures may produce small increases in statistical power. Blocking ensures balance in sample sizes provided the trial is terminated at the end of a block. In addition, blocking can reduce imbalances in sample size even when the trial does not terminate at the end of a block. Stratification can reduce baseline imbalances in one or a small number of key prognostic variables, but introduces significant practical difficulties. Minimisation overcomes those difficulties, and can provide tight control of balance of multiple prognostic factors.

Acknowledgement The author thanks Dr Lisa Harvey, Mark Stewart, and Associate Professor Chris Maher for their helpful comments on a draft of this paper.

References

Bø K, Talseth T and Holme I (1999): Single blind, randomised controlled trial of pelvic floor exercises, electrical stimulation, vaginal cones, and no treatment in management of genuine stress incontinence in women. *BMJ* 318: 487–493.

Lavori PW, Louis TA, Bailar JC and Polansky M (1983): Designs for experiments — parallel comparisons of treatment. *New England Journal of Medicine* 309: 1291–1299.

Piantadosi S (1997): *Clinical Trials: A Methodological Perspective*. New York: Wiley.

Pocock SJ, Assmann SE, Enos LE and Kasten LE (2002): Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. *Statistics in Medicine* 21: 2917–2930.

Raab GM, Day S and Sales J (2000): How to select covariates to include in the analysis of a clinical trial. *Controlled Clinical Trials* 21: 330–342.

Scott NW, McPherson GC, Ramsay CR and Campbell MK (2004): The method of minimization for allocation to clinical trials: a review. *Controlled Clinical Trials* 23: 662–674.

Schluchter MD and Forsythe AB (1985): Post-hoc selection of covariates in randomized experiments. *Communications in Statistics—Theory and Methods* 14: 679-699.

Schulz KF and Grimes DA (2002): Allocation concealment in randomised trials: Defending against deciphering. *Lancet* 359: 614–618.