Transpl Int (2000) 13 [Suppl 1]: S11–S13

D. Elbourne

Statistical issues in clinical trials

D. Elbourne (✉)
Medical Statistics Unit,
London School of Hygiene & Tropical Medicine, Keppel Street,
London WC1E 7HT, UK
e-mail: d.elbourne@lshtm.ac.uk,
Tel.: + 44-171-927-2376,
Fax: + 44-171-637-2853

Abstract The care for patients having organ transplants has improved greatly. This improvement is due, in part, to the advances in knowledge gained through clinical trials. These trials are most useful when they address questions which are important (to patients, their families and their clinical care-givers), which are at their most rigorous statistically (by reducing bias and increasing precision), and which relate closely to the real world. Statisticians and clinicians need to work together to achieve these aims.

Key words Clinical trials · Randomised controlled trials · Statistics · Bias · Precision · Meta-analysis

Introduction

Clinical care for patients having organ transplants has greatly improved, at least in part due to advances gained through clinical trials. This paper considers some of the statistical issues relevant to clinical trials. In particular, it discusses methods for reducing bias, for increasing precision, and for the generalisability of the results to clinical practice. The paper starts from the premise that trials should use rigorous methodology to address questions of importance to patients, their families and their clinical care-givers.

Methods for clinical research

Interventions in health care should at least be scientifically plausible. In this respect, the role of basic science is crucial. The findings of in-vitro and animal laboratory research are not, however, always transferable directly to humans in the clinical setting. There is therefore a need for research on and with patients in the real world.

This research can take many forms, and most is currently observational (non-randomised). For example, much research is based on the early case report by an enthusiastic clinician. This may be followed up with a more systematic survey with larger numbers reported as a case series. While this method is extremely useful for alerting the clinical community to the possibility of a promising new treatment; it suffers from the major problem of lack of a control group with which to compare the findings. So-called "before-and-after" studies use historical controls, but these are problematic as other relevant factors may also be changing over time. Another common form of observational study identifies "cases" (those with some adverse outcome such as graft failure), finds "controls", and then looks back to explore whether the cases and controls had similar or different "exposures" to some factor such as their hospital of care. These retrospective case-control studies are particularly useful for aetiological studies with rare outcomes, but the choice of controls is crucial, as there may well be other differences between cases and controls (confounders). In such studies, a particular problem may be

recall bias, if exposures are remembered differently in the light of knowledge of the outcome of interest. This latter problem may be overcome if the cases and controls are identified prospectively in terms of the interventions, and then followed up to ascertain their outcomes. These cohort studies with contemporaneous controls may still face problems of selection biases whereby people at different prior risk of the outcome of interest differentially receive one or other of the forms of care being compared. Whereas non-randomised trials can purposefully balance *known* prognostic factors, only randomised controlled trials can balance *unknown* factors so that any remaining differences between randomised groups are due to chance.

Methods of randomisation

Randomisation is not the same as haphazard selection. In practice, random allocation will only reduce selection biases if the allocation is "concealed". For instance, using the patient's date of birth as the basis for treatment allocation in some quasi-random designs, the allocation is known in advance and so may be consciously or subconsciously subverted. Randomisation is more secure if the person entering the patient into the trial is distant from those generating the random allocation schedule, but even then a system such as that using consecutively numbered sealed opaque envelopes held locally, while preferable to quasi-randomised designs, may also be vulnerable to "cheating". Where possible, an ideal method would be a telephone call (or equivalent) to a central randomisation service to register patient details followed by random allocation to coded identically packaged drugs or to a particular treatment. The randomisation schedule may be based on simple randomisation in which each patient has known (often equal) chance of being given each treatment, but the treatment to be given cannot be predicted in advance for any one individual. For large samples, this is usually sufficient to ensure the overall balance of key prognostic factors. If, however, the trial is small (or analyses need to be carried out before the end of the trial), simple randomisation can be helped by blocking or stratification. Blocking can keep the numbers in each group close at all times, preferably varying block size to forestall guessing if the study is not blinded (see below), and stratification can achieve balance on a small number of main prognostic factors. The advantages of both of these can be combined using a central computer in a process known as minimization, in which the next patient to enter the trial is given (with probability > 0.5) whichever treatment would minimize the overall balance of several prognostic factors between groups at that stage.

Reduction of selection biases *after* random allocation

Random allocation is necessary but not sufficient to reduce bias. Later biases can arise if there is large loss to follow-up, especially if the loss is different in two trial arms (i.e. if patients are selectively withdrawn from analysis, and particularly if withdrawal is treatment-related), and if analyses are based on the sub-groups showing greatest effects. Solutions to these potential problems, at least in terms of hypothesis-testing, lie in basing the *primary* analysis on all patients as randomly allocated ("intention-to-treat" analysis), and restricting sub-group analyses to those pre-specified in the protocol. *Later* analyses for hypothesis-generating can be based on, for instance, a per-protocol group or the data can be "dredged", which may lead to future fruitful avenues for research.

Other sources of post-randomisation bias may be concomitant treatment bias and assessment bias. The former arises when a trial compares two treatments, one of which leads to different other management(s). This mainly applies in unblinded trials (see below). It is important to record these data, as they form part of the implications of particular treatment(s). The major problem may lie in the dilution of the main treatment effect, which may necessitate larger sample sizes. Assessment bias arises if the assessment of outcome is affected by knowledge of the allocated treatment. The effects can be reduced if outcomes of interest are "hard" (such as death, retransplantation) which are less susceptible to bias. An alternative or additional option is blinding (masking), which is often associated with placebo controls, especially in drug trials. In single blinding, only one type of participant, usually the patient, is unaware of treatment allocation. In double blinding, neither the patient nor the doctor or other carer is aware. Finally, triple blinding is when neither the patient nor the carer nor the evaluator is aware. It can be instructive to assess the extent to which participants are *actually* unaware.

Precision

The reduction of biases is necessary but not sufficient. In addition, we want the unbiased "answer"to have a small *P*-value (high statistical significance) and narrow confidence interval (tight estimation of likely size of effect). This is because the play of chance (random error) may be greater than effects of treatment (if any), and so to be confident of not missing a real effect if one exists, one needs appropriately large sample sizes (and/or meta-analyses). The factors which go into a statistical sample size calculation include the incidence of the primary outcome measure; how small an effect of treatment is considered clinically important to detect (δ); how much confidence is needed that the trial is not de-

tecting a "significant" effect when there really is no effect (α); and how much confidence is needed that the trial is not failing to detect a real effect if it exists (β), where β is the "power" to detect a specified δ. For example, in a trial of two immunosuppressants following liver transplantation, if the incidence of the primary outcome of death or retransplantation with microemulsified cyclosporin is assumed to be 30%; a reduction to 20% with tacrolimus is considered a clinically important difference; and α is set at 0.05 with β at 0.2 (i.e. power of 80%), a sample size of 300 in each group would be needed. In general, sample sizes need to be greater when δ, α and/or β are smaller. Large sample sizes for relatively rare conditions usually need multi-centre, often international, collaborations.

Generalisability

The external validity of a trial may be potentially compromised by differences:between people receiving care and the and the nature of that care *within* trials and *outside* trials. Even if these differences exist, however, it does not necessarily mean that there are *differential* effects of different treatments. Trials which have tight entry criteria and are looking to test efficacy (i.e. *could* a treatment work? – explanatory trials) are usually less generalizable than those with broader entry criteria testing effectiveness (i.e. *does* a treatment work in practice? – pragmatic trials).

Structured reviews and meta-analysis

No trial can stand alone, but must be seen within the context of other relevant evidence. This overview is more rigorous if subjected to the same scrutiny as that of a single study protocol. Examples of structured reviews can be found in the Cochrane Database of Systematic Reviews. Meta-analyses involve statistical combinations of results of several similar trials to produce an overall estimate of effect size with tighter confidence intervals (CI). For instance, the review by Couchoud [1] considered the question of cytomegalovirus (CMV) prophylaxis with antiviral agents in the context of solid organ transplantation. The comprehensive searching strategy identified 78 reports. Data were extracted by two reviewers independently, and further information was sought from authors. Sixty-five of the reports were then excluded because they had no control group or were not randomized, leaving 13 trials (1138 patients) for meta-analysis. Although the methodological quality was heterogeneous, the review showed that antiviral prophylaxis reduced CMV disease [relative risk (RR) 0.43; 95% CI 0.34–0.54]; CMV infection [RR 0.62; 95% CI 0.53–0.73]; with trends in favour of antiviral prophylaxis for death [RR 0.71; 95% CI 0.44–1.16]; graft loss [RR 0.80; 95% CI 0.51–1.27]; and acute rejection [RR 0.92; 95% CI 0.79–1.07]. These conclusions were not altered in sub-group analyses by type of antiviral (acyclovir/ganciclovir), and by organ (kidney, liver, heart). Such a review allows readers to see the "wood from the trees". The increase in effective sample size allows exploration of heterogeneity and sub-groups, and can guide future research.

Conclusions

Not all interventions need trials, and sometimes trials are not feasible. Nor is the appropriate timing always obvious. Some proponents argue for randomizing the first patients, while others point to the need to first optimize the intervention (especially in the context of steep learning curves for surgical procedures). Good trials are hard to do, so patients and families should collaborate with clinicians to focus on which questions (conduct meta-analyses first) and which outcomes (e.g. short-term surrogate markers or mortality and longer term quality of life) are most important. In addition, early collaboration between clinicians and triallists/statisticians is fruitful. Good clinical research arises from practical experience, and trials need, as much as possible, to be conducted within real practice so that the results of trials are relevant to that practice. Finally, the results of trials then need to be incorporated within practice so that care of future patients is based on clinical experience allied with solid research evidence.

Reference

1. Couchould C (1999) Cytomegalovirus prophylaxis with antiviral agents for solid organ transplantation. In: The Cochrane Library, Issue 2. Update Software, Oxford