Determining the sample size in a clinical trial

Adrienne Kirby, Val Gebski and Anthony C Keech

SAMPLE SIZE MUST BE PLANNED carefully to ensure that the research time, patient effort and support costs invested in any clinical trial are not wasted. Item 7 of the CONSORT statement relates to the sample size and stopping rules of studies (see Box 1); it states that the choice of sample size needs to be justified.[1]

Ideally, clinical trials should be large enough to detect reliably the smallest possible differences in the primary outcome with treatment that are considered clinically worthwhile. It is not uncommon for studies to be underpowered, failing to detect even large treatment effects because of inadequate sample size.[2] Also, it may be considered unethical to recruit patients into a study that does not have a large enough sample size for the trial to deliver meaningful information on the tested intervention.

Components of sample size calculation

The minimum information needed to calculate sample size for a randomised controlled trial in which a specific event is being counted includes the *power*, the *level of significance,* the *underlying event rate in the population under investigation* and the *size of the treatment effect sought*. The calculated sample size should then be adjusted for other factors, including expected compliance rates and, less commonly, an unequal allocation ratio.

Power: The power of a study is its ability to detect a true difference in outcome between the standard or control arm and the intervention arm. This is usually chosen to be 80%. By definition, a study power set at 80% accepts a likelihood of one in five (that is, 20%) of missing such a real difference. Thus, the power for large trials is occasionally set at 90% to reduce to 10% the possibility of a so-called "false-negative" result.

Level of significance: The chosen level of significance sets the likelihood of detecting a treatment effect when no effect exists (leading to a so-called "false-positive" result) and defines the threshold "*P* value". Results with a *P* value above the threshold lead to the conclusion that an observed difference may be due to chance alone, while those with a *P* value below the threshold lead to rejecting chance and concluding that the intervention has a real effect. The level

NHMRC Clinical Trials Centre, University of Sydney, Camperdown, NSW.
Adrienne Kirby, MSc, Biostatistician;
Val Gebski, MStat, Principal Research Fellow;
Anthony C Keech, FRACP, MScEpi, Deputy Director.
Reprints will not be available from the authors. Correspondence: Associate Professor A C Keech, NHMRC Clinical Trials Centre, University of Sydney, Locked Bag 77, Camperdown, NSW 1450.
enquiry@ctc.usyd.edu.au

1: CONSORT checklist of items to include when reporting a trial[1]

Selection and topic	Item no.	Description
Methods Sample size	7	How sample size was determined and, when applicable, explanation of any interim analyses and stopping rules

of significance is most commonly set at 5% (that is, $P = 0.05$) or 1% ($P = 0.01$). This means the investigator is prepared to accept a 5% (or 1%) chance of erroneously reporting a significant effect.

Underlying population event rate: Unlike the statistical power and level of significance, which are generally chosen by convention, the underlying expected event rate (in the standard or control group) must be established by other means, usually from previous studies, including observational cohorts. These often provide the best information available, but may overestimate event rates, as they can be from a different time or place, and thus subject to changing and differing background practices. Additionally, trial participants are often "healthy volunteers", or at least people with stable conditions without other comorbidities, which may further erode the study event rate compared with observed rates in the population. Great care is required in specifying the event rate and, even then, during ongoing trials it is wise to have allowed for sample size adjustment, which may become necessary if the overall event rate proves to be unexpectedly low.

Size of treatment effect: The *effect of treatment* in a trial can be expressed as an absolute difference. That is, the difference between the rate of the event in the control group and the rate in the intervention group, or as a relative reduction, that is, the proportional change in the event rate with treatment. If the rate in the control group is 6.3% and the rate in the intervention arm is 4.2%, the absolute difference is 2.1%; the relative reduction with intervention is 2.1%/6.3%, or 33%.

Estimating the plausible effect of treatment to be sought in a randomised controlled trial provides a further challenge, and may be the most common problem for reported trials. Too frequently, studies are designed to identify an implausibly large treatment effect (for example, a 30% to 50% reduction), when most important treatments that have been adopted into clinical practice have shown more modest benefits. When studies are designed to find unrealistically large reductions and fail, smaller real reductions are inevitably rendered statistically non-significant, leading to confusion about the value of the intervention studied. To resolve uncertainty, the study then needs to be repeated elsewhere,

2: Generic expression for calculating sample size

Sample size $\propto \dfrac{\text{(power, inverse function of significance level*)}}{\text{(absolute difference)}^2}$

* As the *P* value becomes smaller, the function of the significance level increases.

but with a larger sample size than before. Wherever possible, the minimum worthwhile difference in response should be determined from phase II or pilot studies and expert opinion from colleagues. Investigators should take into consideration any cost or logistical advantages or disadvantages of the interventional treatment compared with standard care.

From these components, sample size can be calculated as shown in Box 2. It can be seen that the required sample size increases as the chosen significance level becomes smaller and as the chosen power increases. Also, even a small change in the expected absolute difference with treatment has a major effect on the estimated sample size, as the sample size is inversely proportional to the square of the difference. Thus, if 1000 participants per treatment group are required to detect an absolute difference of 4.8%, 4000 per treatment group would be required to detect a 2.4% difference. Precise calculation of sample size for different types of outcomes (continuous, binary and time-to-event) is discussed in standard texts.[3-5] A checklist for determining sample size is given in Box 3.

Effect of compliance

A major limitation of many sample size calculations is the failure to account for patients' predictable lack of compliance with their allocated treatments. As compliance losses directly affect the size of the achievable treatment difference, they also affect the estimated sample size in a non-linear fashion. For example, a placebo-controlled study needing 100 patients per treatment arm, with 100% compliance, would require about 280 patients per arm if compliance is only 80% in each group (that is, 20% of patients allocated the investigational treatment fail to take it, and 20% of patients allocated to the placebo-control arm cross over to the investigational treatment). The compliance adjustment formula is adjusted *n* per arm equals $N/([c1+c2-1]^2)$, where $c1$ and $c2$ are the average compliance rates per arm (so, in

the above example, adjusted $n = 100/([0.8+0.8-1]^2) = 280$).

Allocation ratio

A one-to-one allocation to intervention and control treatment arms is the most common form of random allocation and results in the smallest sample size requirement. Sometimes different allocation ratios are chosen, resulting in a larger total sample size needed to achieve the same power. This may be justified where the investigational treatment is unusually expensive or complicated to administer.

Reporting the sample size section of the protocol

The sample size calculation should be described in sufficient detail to allow its use in other protocols. The power, level of significance and the control and intervention event rates should be clearly documented. Information on the scheduled duration of the study, any adjustment for non-compliance and any other issues that formed the basis of the sample size calculation should be included. For continuous outcomes, in particular (eg, blood pressure), assumptions made about the distribution or variability of the outcome should be explicitly stated.

Conclusion

Estimating sample size is important in the design of clinical trials, and the quality of the estimate ultimately depends on the quality of the information used to derive it. Care should be taken to avoid overestimating the likely event rate and the feasible effects of treatment. The objectives and outcome measures of the study must be clearly stated,[6] and the information used in calculating the sample size should reflect as closely as possible the type of data that will be gathered from the trial in question. Professional advice should be sought before embarking on any major trial project.

References

1. Altman DG, Schulz KF, Moher D, et al, for the CONSORT group. The revised CONSORT statement for reporting randomised trials: explanation and elaboration. *Ann Intern Med* 2001; 134: 663-694.
2. Frieman JA, Chalmers TC, Smith H Jr, Kuebler RR. The importance of beta, the type II error and sample size in the design and interpretation of the randomised control trial: survey of 71 "negative" trials. *N Engl J Med* 1978; 299: 690-694.
3. Altman DG. Statistics and ethics in medical research: III. How large a sample? *BMJ* 1980; 281: 1336-1338.
4. Gore SM. Assessing clinical trials — trial size. *BMJ* 1981; 282: 1687-1689.
5. Friedman L, Furberg C, DeMets D. Fundamentals of clinical trials. 3rd ed. New York: Springer-Verlag; 1998.
6. Gebski V, Marschner I, Keech AC. Specifying objectives and outcomes for clinical trials. *Med J Aust* 2002; 176: 491-492.

3: Checklist for determining sample size for clinical trials

■ Estimate the event rate in the control group by extrapolating from a population similar to the population expected in the trial.
■ Determine, for the primary outcome, the smallest difference that will be of clinical importance.
■ Determine the clinically justifiable power for the particular trial.
■ Determine the significance level or probability of a "false positive" result that is scientifically acceptable.
■ Adjust the calculated sample size for the expected level of non-compliance with treatment.