

Sample size determination for clinical research

Duolao Wang¹
 Ameet Bakhai¹
 Angelo Del Buono²
 Nicola Maffulli^{3,4}

¹ Department of Medical Statistics, London, UK

² Department of Orthopaedic and Trauma Surgery, Campus Biomedico University of Rome, Rome, Italy

³ Centre for Sports and Exercise Medicine, Queen Mary University of London, Barts and The London School of Medicine and Dentistry, Mile End Hospital, London, UK

⁴ Department of Physical and Rehabilitation Medicine, University of Salerno, Salerno, Italy

Corresponding author:

Nicola Maffulli
 Centre for Sports and Exercise Medicine,
 Queen Mary University of London,
 Barts and The London School of Medicine and Dentistry, Mile End Hospital, London, UK
 Department of Physical and Rehabilitation
 Medicine, University of Salerno, Salerno, Italy
 E-mail: n.maffulli@qmul.ac.uk

Summary

Calculating the sample size is a most important determinant of statistical power of a study. A study with inadequate power, unless being conducted as a safety and feasibility study, is unethical. However, sample size calculation is not an exact science, and therefore it is important to make realistic and well researched assumptions before choosing an appropriate sample size accounting for dropouts and also including a plan for interim analyses during the study to amend the final sample size.

KEY WORDS: randomized control trials, statistics, sample size.

Introduction

A well-designed clinical trial asks an important question regarding the effectiveness or safety about a treatment, and provides a reliable answer using statistical analysis. The major determinant of the reliability of the answer is the sample size of the trial. Therefore sample

size and trial results go hand in hand. In this article, we will review the issues determining the appropriate sample size for a randomised control trial.

What is sample size for a randomised control trial?

The sample size of a randomised controlled trial (RCT) is the number of patients to be included in the study. This number determines the likelihood of being able to detect a certain magnitude of difference (usually an anticipated benefit is assumed) between treatment groups. Either that likelihood or that magnitude can be varied. The likelihood is called the “power” of the study, and the magnitude of benefit is called the “expected treatment effect”. The higher the sample size of patients, the better the power with which to detect a treatment effect, or the smaller the treatment effect that can be detected as significant. Conversely, the lower the sample size the less power with which to detect a treatment effect or the greater the effect must be to be detected as significant. The sample size calculation is also based on the design of a trial, so how the primary outcome is to be determined must also be clarified in advance of determining sample size.

Why do we have to choose a sample size?

When resources are limited we must decide how best to invest them to achieve maximal benefit. Should we use treatment X or treatment Y? To answer this, we cannot wait indefinitely as patients will continue to be given or refused a treatment without evidence. We need to decide how hard we will look for the answer. We may say it is only worth looking at the question, if we are fairly likely to capture a 10% improvement with the new treatment. To create this “fair chance” of capturing such a difference (if it exists) we have to choose a sample size wisely based on realistic initial assumptions. More importantly it is unethical to a study which has no chance of capturing a real difference, as we will have spent precious resources performing the study at no additional gain. From this we can appreciate that choosing an appropriate sample size for a study is not an exact science and needs good judgement.

What factors determine sample size?

Several factors are critical in determining an appropriate sample size for a RCT:

- the event rate of patients in the control or standard treatment arm;
- the smallest treatment effect or benefit to be detected;
- the significance level at which we reject the null hypothesis of no difference in treatment effects;
- the power with which we want to detect an effect;
- the design of the study;
- the rate of dropouts during the study.

Are negative trials due to small sample sizes?

A negative clinical trial is a trial in which the observed differences between the proposed and control treatments was not large enough to satisfy a specified "significance" level (the risk of false positive error or Type I error), and the results were declared to be "not statistically significant". With the benefit of hindsight, analyses of many "negative" clinical trials have shown that the assumptions chosen by investigators often lead them to choose a sample size which is too small to offer a reasonable chance of avoiding a false negative error (Type II error). A branch of medical statistics known as meta-analysis often combines results from many such small studies to try to closer estimate a true mean effect. If this shows a favourable benefit of the new treatment, then ideally a larger definitive RCT should be performed to verify this, but always one has to balance the outlay of resources to realise the potential benefit, and even then there are always surprises found in large RCT's of unexpected effects.

So is that it - just apply the formula?

Statisticians have developed various methodologies for determining sample sizes, depending on the number of treatments, the type of primary endpoint, statistical analysis methods for outcomes and study design. When determining sample size it is wise to have finalised the main objective of the protocol and then work with a medical statistician to examine a range of assumptions. This will provide a range of sample sizes and a balance can then be struck between ideal

statistical power and available resources and time before finalising the sample size. Even then in an RCT interim analyses of overall event rates (i.e. still unaware of event rates by treatment group) will guide as to whether the sample size needs to be increased or decreased as the RCT proceeds. When event rates by treatment arms are measured during the course of a study, the data and safety committee may recommend early termination of the study, if a large benefit or element of harm is seen with the new treatment early in the trial.

Are there websites available for calculating sample size?

The following user-friendly calculators are available on the Internet for sample size calculations:
<http://ebook.stat.ucla.edu/calculators/powercalc>;
<http://www.dssresearch.com/SampleSize/default.asp>;
http://hedwig.mgh.harvard.edu/quant_measur/defs.html#cross.

Conclusion

The sample size is the most important determinant of statistical power of a study, and a study with inadequate power, unless being conducted as a safety and feasibility study, is unethical. However, sample size calculation is not an exact science and therefore it is important to make realistic and well researched assumptions before choosing an appropriate sample size accounting for dropouts and also including a plan for interim analyses during the study to amend the final sample size.

Further reading

Pocock SJ. Clinical Trials: A Practical Approach. Wiley, New York, 1983.
Moher D, Dulberg CS, Wells GA. Statistical power, sample size, and their reporting in randomized controlled trials. JAMA 1994; 272:122-124.