ACTA SCIENTIFIC NUTRITIONAL HEALTH

Volume 3 Issue 2 February 2019

Review Article

The Concept Behind Sample Size Calculation for Randomised Control Trials. A Review

Vishwanath P Hiremath[1]* and Shivalingappa Javali[2]

[1]*Professor, Department of Community Dentistry, Maratha Mandal Dental College, Belgaum, India*

[2]*Professor in Statistics, S.D.M. Dental College, Dharwad, India*

***Corresponding Author:** Vishwanath P Hiremath, Professor, Department of Community Dentistry, Maratha Mandal Dental College, Belgaum, India.

Received: December 29, 2018; **Published:** January 22, 2019

Abstract

To design clinical trials, research duration and sample size calculations are the key for the success of a clinical trial. In the beginning the paper discusses the statistical theory behind sample size calculations. The paper highlights that researchers should also be aware that the study design helps one to choose an appropriate sample size calculation method. An emphasis on the design of randomised controlled trial is considered. Further the factors like effect size, type of primary outcomes, methods and steps involved in sample size calculation for randomized controlled trials is dealt with.

Keywords: Statistical Power; Sample Size; Randomized Controlled Trial

Introduction

Even a well-executed study may fail to answer its research question if the sample size is too small. On the other hand, study with large samples will be more difficult to carry out and it will not be cost effective [2]. Also, the confidence intervals and values, both of these have been shown to depend strongly on the size of the study sample in question, with larger samples generally resulting in narrower confidence intervals and smaller P values [1].

In medical research, the sample size has to be "just large enough" to provide a reliable answer to the research question. How to arrive at this magic number?

Statistical theory behind sample size calculation

The Null Hypothesis is set up to be rejected.

The argument is that it is easier to prove that a statement is false than to prove it's true. For example, we want to prove that "all pens are black", and one could notice only black pens everywhere, there's still doubt that a white pen could be lying somewhere. But once we get a white pen, the hypothesis of 'all pens are black' is rejected [6]. In statistical hypothesis testing, the null hypothesis set out for a particular significance test and it always occurs in conjunction with an alternative hypothesis. The null hypothesis is set up to be rejected, thus if we want to compare two interventions, the null hypothesis will be "there is no difference" versus the alternative hypothesis of "there is a difference". However, not being able to reject the null hypothesis does not mean that that it is true, it just means that we do not have enough evidence to reject the null hypothesis.

Type I error Alpha error

In any Research there is no such thing that the results of my findings are correct. The only way this can be expressed that there is a small possibility of me being wrong. This is rejecting null hypothesis when it is actually correct [4]. In classical statistical terms, type I error is always associated with the null hypothesis and is called as alpha error. For example, in a two-arm trial with Null hypothesis there is no difference in two therapies, considering a statistical significance level of $\alpha = 0.05$, a positive P value of 0.03 was found at the end. Assuming that all the bias is controlled we can have two possibilities, one possibility is that a real difference exists between the two interventions and the other is that this difference is by chance, but there is only 3% chance that this difference is just by chance. Hence, if the p-value is closer to 0 then the chances of difference occurring due to "chance" are very low. This also explains why a two-sided test is usually preferred compared to one-sided test, which requires smaller sample size [5]. The type I error is usually set at two sided 0.05, not all, but some study design is exceptive.

Type II error Beta error

As null hypothesis is associated with type I error, the alternative hypothesis is associated with type II error, when we are not able to reject the null hypothesis when it is actually false. This is given by the power of the research (1- type II error/ β): the probability of rejecting the null hypothesis when it is false.). As mentioned, the main aim of a clinical research is to reject the null hypothesis and we could achieve this by controlling the type II error [3]. Usually, the power is set at 0.80.(1- β), as higher the power, more the size of the sample. The study for the researcher in which the power is high means that the study has a high chance of detecting a difference between groups if one exists. Eventually, if the study demonstrates no difference between groups the researcher can be reasonably confident in concluding that none exists in reality.

To calculate Sample size for an RCT to ethically answer the research question the following factors are considered.

1. Type of comparison
2. Type of configuration
3. Type of the primary outcome.

Type of comparison (Considering the design of the study)

Parallel RCT design is most commonly used, which means all participants are randomized to two (the most common) or more arms of different interventions treated concurrently [7-9].

Superiority trials

To Study verify that a new treatment is more effective than a standard treatment from a statistical point of view or from a clinical point of view. The null hypothesis is that: The new treatment/ test therapy is not more efficacious/better than the control treatment by a statistically/clinically relevant amount. Based on the nature of relevant amount, superiority design contains statistical superiority trials and clinical superiority trials.

Equivalence trials

The objective of this design is to ascertain that the new treatment and standard treatment are equally effective. The null hypothesis of that is: Both the treatments differ clinically by relevant amount.

Non-inferiority trials

Non-inferiority trials are conducted to show that the new treatment is as effective but need not be superior when compared to the standard treatment. The corresponding null hypothesis is: The new treatment is inferior to the control treatment by a clinically relevant amount.

One-sided test is performed in both superiority and non-inferiority trials, and two-sided test is used in equivalence trials.

Factor to be considered	Effect in Magnitude	Influence on appreciation of effect	Sample size needed
P value	Small	'significance' is difficult to achieve	Large
	Large	'significance' is easier to attain	Small
Power	Low	Identification is difficult	Small
	High	Identification possible	Large
Effect	Small	Difficult to appreciate it.	Large
	Large	Easy to appreciate.	Small

Table 1: Factors that affect sample size in randomised controlled trials.

Type of configuration [4]

Effect size of therapies

The effect size specifies the accepted clinical difference between two therapies that a researcher wants to observe in a study.

The difference between two groups in a study will usually be explored in terms of an estimate of effect, appropriate confidence interval and P value. The confidence interval indicates the likely range of values for the true effect in the population; while the P value determines how likely it is that the observed effect in the sample is due to chance. A related quantity is the statistical power of the study. The probability of rejecting the null hypothesis when it is false Simply put, this is the probability of correctly identifying a difference between the two groups in the study sample when one genuinely exists in the populations from which the samples were drawn [1].

There are three usual ways to get the effect size [6]:

a) From past literature.
b) If no past literature is available, one can do a small pilot study to determine the estimated effect sizes.
c) Clinical expectations.

Type of Primary outcome.

Parameter definitions Dichotomous variable

N = size per group; p = the response rate of standard treatment group; p0 = the response rate of new drug treatment group; z_x =

the standard normal deviate for a one or two sided x; d = the real difference between two treatment effect; δ_0 = a clinically acceptable margin ; S2 = Polled standard deviation of both comparison groups.

Dichotomous variable

For non-inferiority design, the formula is:

$$N = 2 \times \left(\frac{z_{1-\alpha} + z_{1-\beta}}{\delta_0} \right)^2 \times p \times (1-p)$$

For equivalence design, the formula is:

$$N = 2 \times \left(\frac{z_{1-\frac{\alpha}{2}} + z_{1-\beta}}{\delta_0} \right)^2 \times p \times (1-p)$$

For statistical superiority design, the formula is:

$$N = \frac{1}{2} \times \left(\frac{Z_{\frac{\alpha}{2}} + Z_{\beta}}{\arcsin\sqrt{p} - \arcsin\sqrt{P_0}} \right)^2$$

For clinical superiority design, the formula is:

$$N = 2 \times \left(\frac{z_{1-\alpha} + z_{1-\beta}}{\mathrm{d} - \delta_0} \right)^2 \times p \times (1-p)$$

Let us consider a problem: The research question is whether there is a difference in the efficacy of Vitamin D (new drug) and Vitamin C (standard drug) for the treatment bleeding in gums (Gingivitis) in 6-week treatment duration. All parameters were assumed as follows: p = 0.40; p_0 = 0.58; α = 0.05; β = 0.20; δ = 0.18; δ_0 = 0.10.

Then:

$$N_{\text{non-inferiority}} = 2 \times \left(\frac{1.645 + 0.845}{0.10} \right)^2 \times 0.40 \times (1-0.40) = 298$$

$$N_{\text{equivalence}} = 2 \times \left(\frac{1.96 + 0.845}{0.10} \right)^2 \times 0.40 \times (1-0.40) = 378$$

$$N_{\text{statistical superiority}} = \frac{1}{2} \times \left(\frac{1.96 + 0.845}{\arcsin\sqrt{0.40} - \arcsin\sqrt{0.58}} \right)^2 = 121$$

$$N_{\text{clinical superiority}} = 2 \times \left(\frac{1.645 + 0.845}{0.18 - 0.10} \right)^2 \times 0.40 \times (1-0.40) = 466$$

Continuous variable

N=size per group; δ = π_1 = mean change in standard treatment group; π_2 = mean change in the new drug treatment group; z_x= the standard normal deviate for a one or two sided x; d= the real difference between two treatment effect; δ_0= a clinically acceptable margin ; S^2= Polled standard deviation of both comparison groups. δ = π_2- π_1

For non-inferiority design, the formula is:

$$N = 2 \times \left(\frac{z_{1-\alpha} + z_{1-\beta}}{\delta_0} \right)^2 \times s^2$$

For equivalence design, the formula is:

$$N = 2 \times \left(\frac{z_{1-\frac{\alpha}{2}} + z_{1-\beta}}{\delta_0} \right)^2 \times s^2$$

For statistical superiority design, the formula is:

$$N = 2 \times \left(\frac{z_{1-\frac{\alpha}{2}} + Z_{1-\beta}}{\delta} \right)^2 \times s^2$$

For clinical superiority design, the formula is:

$$N = 2 \times \left(\frac{z_{1-\alpha} + z_{1-\beta}}{\delta - \delta_0} \right)^2 \times s^2$$

Continuous variable

Problem: The research question is whether there is a difference in the efficacy of A CE II antagonist (new drug) and A CE inhibitor (standard drug) for the treatment of primary hypertension. The primary measurement is change of diastolic blood pressure (DBP, mmHg) compared to baseline. The parameters assumed were as

follows: mean change in DBP for the new drug treatment group = 18 mm Hg; mean change of DBP in standard treatment group = 14 mm Hg; $\alpha = 0.05$; $\beta = 0.20$; $\delta = 4$ mm Hg; $\delta 0 = 3$ mm Hg; $s = 6$ mm Hg.

Then:

$$N_{\text{non-inferiority}} = 2 \times \left(\frac{1.645 + 0.845}{3} \right)^2 \times 6^2 = 50$$

$$N_{\text{equivalence}} = 2 \times \left(\frac{1.96 + 0.845}{3} \right)^2 \times 6^2 = 63$$

$$N_{\text{statistical superiority}} = 2 \times \left(\frac{1.96 + 0.845}{4} \right)^2 \times 6^2 = 36$$

$$N_{\text{clinical superiority}} = 2 \times \left(\frac{1.645 + 0.845}{4 - 3} \right)^2 \times 6^2 = 112$$

Discussion

1. It is very clear the steps involved in calculating sample size go hand in hand with the steps involved in designing a RCT. Most importantly initially the researcher should specify the null and alternative hypotheses, along with the type I error rate and the power (1- type II error rate).

2. Secondly, the appropriate parameters for data is collected by the researcher but sometimes a pilot study may be required.

3. Thirdly, we have to keep in mind the objective of the study when the null hypothesis and the alternative hypothesis is to be established. Also the sample size can be determined on basis of certain reasonable parameters of clinically significant. This needs a lot of clinical expertise not just the statistician calculations.

4. If δ is too large, several drugs which are inefficacious are considered for they can be judged as non-inferiority/equivalence; On the contrary, if δ is too small, we could well reject some effective drugs coming into the market [10].

Conclusion

This review highlights the importance of design and measures of outcome to calculate the sample size. It also provides some knowledge on what information will be needed when coming to consult a biostatistician for sample size determination. Research Scholars should be able to help statistician with some knowledge of sample size determination in RCT especially in giving the value of δ the clinical difference which should be reasonable.

Bibliography

1. Elise Whitley and Jonathan Ball. "Statistics review for Sample size calculations". *Critical Care* 6.4 (2002): 335-341.

2. Zhong BL. "Randomized Controlled Trial Sample Size Calculation". *Journal of Thoracic Disease* 1 (2009): 51-54.

3. Muller KE and Benignus VA. "Increasing scientific power with statistical power". *Neurotoxicology and Teratology* 14 (1992): 211-219.

4. Chan YH. "Study design considerations study configurations, introduction to clinical trials". editors Karlberg J and Tsang K. (1998): 249-57.

5. Wittes J. "Sample size calculations for randomized controlled trials". *Epidemiology Review* 24 (2002): 39-53.

6. Chan YH. "Randomised controlled trials (RCTs) – Sample Size: The Magic Number?" *Singapore Medical Journal* 44.4 (2003): 172-174.

7. Lesaffre E. "Superiority, equivalence, and non-inferiority trials". *Bulletin of the NYU Hospital for Joint Diseases* 66 (2008): 150-154.

8. Schall R., *et al*. "Introduction to clinical trials". editors Karlberg J and Tsang K, (1998): 258-266.

9. ICH Harmonised Tripartite Guideline. Statistical principles for clinical trials.

10. Liu YX., *et al*. "Statistical Methods in Clinical Noninferiority/ Equivalence Evaluation". *Chinese Journal of Clinical Pharmacology and Therapeutics* 5 (2000): 344-349.

Volume 3 Issue 2 February 2019