TECHNICAL NOTES

SAMPLE SIZE IN CLINICAL RESEARCH, THE NUMBER WE NEED

Pratap Patra
Department of Pediatrics, Govt. Medical College, Vadodara, Gujarat, India
Correspondence to: Pratap Patra (pratap_patra3@yahoo.co.in)

ABSTRACT

In the era of evidence based medicine where we should base our clinical practice on the available evidence of growing medical literatures, large number of available published literatures do not have enough sample size to detect their desirable primary or secondary outcome. It is surprised that even in journal with high impact factors sample size calculation is still inadequate, often erroneous and based on assumption which are inaccurate. Sample size calculation is always a difficult issue, which should be address diligently before commencement of the study. The objective of the current narrative review is to illustrate the methodology of sample size calculation for comparative and descriptive research, which will be most useful for the beginners.

Key Words: Sample Size; Type 1 Error; Type 2 Error; Comparative and Descriptive Research

INTRODUCTION

The importance of determining sample size before any study can never be under-estimated. It is essential to have a priori sample size calculation before any study because, they minimize the chance of type 1 (probability of rejecting null hypothesis when it is true) and type 2 (failure to reject null hypothesis when it is false) statistical errors, which is directly linked to the power of the study to detect any meaningful intergroup difference that exists. Secondly, it is equally important to understand the sample size to interpret the relevance of the findings.[1,2] Though sample size calculation is always a burning issue as far as any clinical research is concerned, surprisingly, it is often neglected in clinical practice. It is important for the readers to know that many underpowered studies continue to get published.[3] If the sample size is too small, it is impossible to generalize the situation in the parent population and at the same time it is wastage of resources, if it is too big. Therefore, it should be just appropriate.[3] Though it is a very important issue, only few good literatures are available so far.

Components of Sample Size Calculation

The basic components of sample size calculation for comparative studies are selection of acceptable level of type 1 and type 2 errors, appropriate statistical power, effect size, significance and probability, and estimated measurement of variability.

Type 1 and type 2 errors

In comparative trial the investigator tries to reject the null hypothesis and accept the alternate hypothesis. The probability of rejecting the null hypothesis when it is true is called type 1 error or alpha (α) or level of statistical significance. The probability of committing type 2 error (fail to reject null hypothesis when it is false is) called beta (β). Ideally, both α and β should be set at zero eliminating the possibility of false positive and false negative results but practically not possible and they are made as small as possible.[4]

Power

The power of a study is its ability to detect the difference of a given size represented by

(1-β).[4] Higher the power better is the chance to detect a difference between groups if it exists. The power of the study is depends upon various factor, but as a general rule higher power is achieved by increasing the sample size.[5] Traditionally the power should be at least 80% which means there is 80% chance of detecting a particular state of difference that exists between groups.

Effect size

Effect size is the difference between the experimental group and control group which investigator tries to detect. Selecting an appropriate effect size is the most difficult aspect of sample size planning. Smaller the effect size, bigger the sample size required to detect the statistical significant difference if present in between two groups. Estimation of effect size is based on the clinical judgment and purely subjective.[6]

Significance and probability

Siginficance is arbitrary cut off point chosen for type 1 error usually set at α (0.05). *P* value is the boundary between significance and non-significance which is based on probability where 'p' stands for the probability that finding of interest was reached by chance.[7] When 'p' value is less than (α) threshold (0.05) said to be statistically significant and when it is above the threshold, not statistically significant.

One and two sided test

When we compare two groups we usually starts with null hypothesis that there is no difference between the population from which the data comes but if this is not true than alternate must be true, that there is a difference. Since null hypothesis dose not specifies any direction and so does the alternative hypothesis we have a two sided test. A new drug intended to relive nausea might exacerbate it. In the language of statisticians, this means one needs a two tailed test unless very much convinced with the evidence that the difference be only in one direction.

Estimated measurement of variability

It is represented by the expected standard deviation in the measurement made within each comparison group. If statistical variability increases, the sample size needed to detect the minimum difference increases. A separate estimate of measurement of variability is not required when the measurement is a proportion in contrast to mean because standard deviation is mathematically derived from the proportion.[8]

With the backdrop of some essential prerequisite regarding sample size calculation, we can proceed to discuss the methodology for the sample size calculation. But before that, one need to choose the acceptable level of type 1 error (α) and type 2 error (β) expressed mathematically this dependence as Z_α and Z_β. Table 1 provides values of Zα and Z_β corresponding to commonly used levels of significance and power.

Table-1: Values of Z_α and Z_β corresponding to specified value of significance level and power.

Signifi-cance level	Z crit Value *Two sided test **One sided test	Statistical Power	Z power value
0.01 (99)	2.576 2.326	0.80	0.842
0.02 (98)	2.326 1.645	0.85	1.036
0.05 (95)	1.960 1.282	0.90	1.282
0.10 (90)	1.645	0.95	1.645

* Investigator expecting difference only one direction.
** Investigator expecting the difference in either direction.

Sample Size for Comparative Studies (Dichotomous Outcome)

Sample size requirements in terms of risk difference, is the most commonly used in design of clinical trial and a straightforward application of traditional sample size formula for comparison of two proportions.[9] If we have event rate of proportion in the experiment group if we called it P_E, event rate in the control group called P_C and $\delta = P_E - P_C$ is the difference in the event rate which is scientifically and clinically to detect then; Equation No.1:

$$N = \frac{\left[Z\alpha\sqrt{P^*(1-P^*)} + Z\beta\sqrt{P_E(1-P_E) + P_C(1-P_C)}\right]^2}{\delta^2}$$

Where $P^* = (P_E+P_C)/2$.

Example: Let us consider that, an investigator hypothesizes that caffine is better than aminophylline interms of reducing apnea of prematurity. Previous studies have reported an efficacy of 40% for aminophylline. To detect a 5 % difference between them with power of 80% and two tailed test of 5% significance level, than what sample size would be needed?

$N= \{1.960\sqrt{[0.375(1-0.375)]} + 0.840\sqrt{[0.35(1-0.35) + 0.4(1-0.4)]}\}^2/0.05^2$

Sample size required per group is 876. For correction of continuity and high degree of accuracy one need to increase the sample size by $2/ (P_E - P_C)$, than final sample size would be 896 per group.

Sample Size Calculation for Comparative Study (Continuous Outcome)

The approach to sample size for continuous outcome (Infinite no. of values) is different from the sample size calculation for dichotomous outcome, where standard deviation of the measured variable is always taken in to account. The sample size for the above measures can be obtained by using Equation No 2.

$N = 4\,\sigma^2 (Z_\alpha + Z_\beta)^2 / D^2$

Where N is the total sample size require, σ is the standard deviation of the outcome variable same for both the group, Z_α is the confidence level and Z_β is the power, D is the effect size which the investigator wants to detect.

Example: let us consider an investigator plans a randomized control trial of the effect of salbutamol and ipratropium bromide on FEV_1 after 2 weeks of treatment. Previous study has reported mean FEV_1 in persons treated with asthma was 2 liters with a standard deviation of 1 liter. If the investigator tries to detect a difference of 10% between them, how many individual will be required for the study?

$N= 4*1^2(1.960+0.842)^2 / 0.2^2 = 785$ person required.

Sample Size for Descriptive Study

The principle of calculating sample size is different from comparative studies which is also known as observational study, where an investigator observes the subjects without otherwise intervening. The principal descriptive quantity is either mean or proportion and descriptive study commonly reports confidence intervals around the mean or proportion which measures the precision of a sample estimate. In calculating the required sample size for descriptive study, one needs to know the characteristics of the variable of interest, which is either continuous (mean) or dichotomous (proportion).

Sample Size for Continuous Variable

Here the investigator should have a priori information about the standard deviation of the interested attribute or variate and chosen confidence interval as well as confidence level. The sample size can be calculated form Equation No.3

$N = 4*Z_\alpha^2 S^2 / W^2$

Where Zα = Confidence level, S = Standard deviation, W = Width of the confidence interval.

Example : Suppose an investigator wants to detect the mean weight of newborns between 30-34 week of gestation with 95% confidence interval not more than not more than ±0 .1 kg. From the previous study the standard deviation has been reported of 1 kg, then the sample size required would be,

$N = 4*1.96^2*1^2/0.2^2=384$ newborns required.

Sample Size for Dichotomous Variable

While calculating sample size for the dichotomous variable the investigator needs to select the required confidence level and width of the confidence interval. Then the required sample size can be calculated from Equation No.4

$N = 4*Z\alpha^2*p\,(1-p) / w^2$

where Zα is the confidence level, w is the width of the confidence interval and P is the pre-study estimate of the proportion to be measured.

Example: Let us consider that an investigator wish to determine the incidence of nosocomial pneumonia (NP) in neonatal intensive care with 95% confidence level. He selected a confidence interval of ± 10 and the mean incidence NP has been reported earlier is 20%. Then the required sample size would be

$N = 4*1.96^2*0.20(1-0.20)/ 0.20^2 = 62$

DISCUSSION

Sample size calculation is a vexed issue and an important aspect of clinical research which should be clearly documented in sufficient detail in the protocol section of each study.[10,11] What all we have discussed is all about the calculation of sample size in few of the commonest clinical scenarios. However not all the sample size problems are the same and an investigator might have difficulty in obtaining sample size in various other circumstances. In comparative research if an investigator is keen to demonstrate that there is no difference between two treatments which is also known as equivalence study, or wants to illustrate that an intervention is twice or thrice better then placebo, than the approach become different. Obtaining sample size based on selection of appropriate power for comparative studies is a tool that exists since decades has been subjected to debate in recent years. Although lot of emphasis has been given to confidence interval rather the power for sample size calculation, power calculation still remained an instrumental in increasing sample sizes to levels where studies can provide much more useful information.[12]

In descriptive research one can imitate the sample size of similar studies, using published tables, or else applying formulas to calculate a sample size. Effortless approach is to use the same sample size as those of studies similar to the one you plan. But doing so, without reviewing the procedures employed in previous studies one may run the risk of repeating errors that were made in determining the sample size for another study. Although tables can provide a useful guide for determining the sample size, one may need to calculate the necessary sample size for a different combination of levels of precision, confidence, and variability. The best approach to determining sample size is the application of one of several formulas.[13]

Once in a while it becomes extremely difficult for the researcher to get the desired number especially, when no prior information is available regarding the subject he wished to carry out the research. In such situation the strategy should be extensive search for previous and related findings on the topic and research questions but, still no information is available, it is prudent to go for a pilot study which is useful for estimating the standard deviation of a measurement or the proportion of subjects with particular characteristic. If all this fails, the investigator might make an educated guess about the likely value of missing ingredients but should be always a last resort.

CONCLUSION

To conclude, the number of individual needed for a clinical research is always a complicated issue, which should be determine before initiating the study and the approach depends upon the type of research one planned for. At last one shouldn't forget to utilize the various statistical package software that are available for sample size calculation, and help from a good statisticians indeed needed in situation which is cumbersome.

REFERENCES

1. Todd M Michael. Clinical Research Manuscript in anesthesiology. Anesthesiology 2001;95; 1054 – 67.
2. Sarmukadam SB, Garad SG. On validity of assumption while determining sample size. Indian J Community Med 2004;2:87-91.
3. Browner SW, Newman BT, Hully BS. Estimating sample size and Power application and example In Hully BS, Cummings RS, Browner SW, Grady GD, Newman BT editors. Designing clinical research. Lippincott Williams and Wilkins; 2006. P. 65-94.
4. Florey C DV. Sample size for beginners. BMJ 1993;36:1181-1184.
5. Whitely E, Ball J. Statistics Review 4: Sample size calculations Critical Care 2002;6:335-341.
6. Freedman BK, Bernstein J. Sample size and Statistical Power in Clinical Orthopedic Research. J Bone Joint Surg Am 1999;81:1454-60.
7. Kain NZ, Maclaren J. P less than .05 what does it really mean? Pediatrics 2007;119:608-610.
8. H Russel. Sample size estimation: How many individuals should be studied. Radiology 2003;227:309-313.

9. Donner A. Approaches to sample size estimation in the design of clinical trials -a review. Stat med 1984;3:199-214.
10. Kirby A, Gaskin V, Keech CA. Determining the sample size in a clinical trial. MJA 2002;177:256-257.
11. Schulz FK, Grimes AD. Sample size calculation in randomized trials: mandatory and mystical. Lancet 2005; 365:1348-53.
12. Bland MJ. The tyranny of power: is there any better way to calculate sample size? BMJ 2009;339:1134-1135.
13. Kasiulevicius V, Sapoka V, Filipaviciute R.Sample size calculation in epidemiological studies. Gerontologija 2006; 7: 225–231.

Source of Support: Nil
Conflict of interest: None declared