

Sample size calculation: Basic principles

Address for correspondence:

Prof. Sabyasachi Das,
107/4A, North Purbachal,
Kalitala, Kolkata - 700 078,
West Bengal, India.
E-mail: sabyasachi1968@
gmail.com

Sabyasachi Das, Koel Mitra, Mohanchandra Mandal¹

Department of Anaesthesiology and Critical Care, Medical College, Kolkata, ¹Department of Anaesthesiology and Critical Care, North Bengal Medical College, Sushrutanagar, Darjeeling, West Bengal, India

ABSTRACT

Addressing a sample size is a practical issue that has to be solved during planning and designing stage of the study. The aim of any clinical research is to detect the actual difference between two groups (power) and to provide an estimate of the difference with a reasonable accuracy (precision). Hence, researchers should do a *priori* estimate of sample size well ahead, before conducting the study. *Post hoc* sample size computation is not encouraged conventionally. Adequate sample size minimizes the random error or in other words, lessens something happening by chance. Too small a sample may fail to answer the research question and can be of questionable validity or provide an imprecise answer while too large a sample may answer the question but is resource-intensive and also may be unethical. More transparency in the calculation of sample size is required so that it can be justified and replicated while reporting.

Key words: Effect size, medical research, power, randomized control trial, research hypothesis, sample size, study design

Access this article online
Website: www.ijaweb.org
DOI: 10.4103/0019-5049.190621
Quick response code


INTRODUCTION

Besides scientific justification and validity, the calculation of sample size ('just large enough') helps a medical researcher to assess cost, time and feasibility of his project.^[1] Although frequently reported in anaesthesia journals, the details or the elements of calculation of sample size are not consistently provided by the authors. Sample size calculations reported do not match with replication of sample size in many studies.^[2] Most trials with negative results do not have a large enough sample size. Hence, reporting of statistical power and sample size need to be improved.^[3,4] There is a belief that studies with small sample size are unethical if they do not ensure adequate power. However, the truth is that for a study to be ethical in its design, its predicted value must outweigh the projected risks to its participants. In studies, where the risks and inconvenience borne by the participants outweigh the benefits received as a result of participation, it is the projected burden. A study may still be valid if the projected benefit to the society outweighs the burden to society. If there is no burden, then any sample size may be ideal.^[5] Many different approaches of sample size design exist depending on the study design and research question. Moreover, each

study design can have multiple sub-designs resulting in different sample size calculation.^[6] Addressing a sample size is a practical issue that has to be solved during planning and designing stage of the study. It may be an important issue in approval or rejection of clinical trial results irrespective of the efficacy.^[7]

By the end of this article, the reader will be able to enumerate the prerequisite for sample size estimation, to describe the common lapses of sample size calculation and the importance of *a priori* sample size estimation. The readers will be able to define the common terminologies related to sample size calculation.

IMPORTANCE OF PILOT STUDY IN SAMPLE SIZE ESTIMATION

In published literature, relevant data for calculating the sample size can be gleaned from prevalence

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as the author is credited and the new creations are licensed under the identical terms.

For reprints contact: reprints@medknow.com

How to cite this article: Das S, Mitra K, Mandal M. Sample size calculation: Basic principles. Indian J Anaesth 2016;60:652-6.

estimates or event rates, standard deviation (SD) of the continuous outcome, sample size of similar studies with similar outcomes. The idea of approximate ‘effect’ estimates can be obtained by reviewing meta-analysis and clinically meaningful effect. Small pilot study, personal experience, expert opinion, educated guess, hospital registers, unpublished reports support researcher when we have insufficient information in the existing/available literature. A pilot study not only helps in the estimation of sample size but also its primary purpose is to check the feasibility of the study.

The pilot study is a small scale trial run as a pre-test, and it tries out for the proposed major trial. It allows preliminary testing of the hypotheses and may suggest some change, dropping some part or developing new hypotheses so that it can be tested more precisely.^[8] It may address many logistic issues such as checking that instructions are comprehensive, and the investigators are adequately skilled for the trial. The pilot study almost always provides enough data for the researcher to decide whether to go ahead with the main study or to abandon. Many research ideas that seem to show great promise are unproductive when actually carried out. From the findings of pilot study, the researcher may abandon the main study involving large logistic resources, and thus can save a lot of time and money.^[8]

METHODS FOR SAMPLE SIZE CALCULATION

Sample size can be calculated either using confidence interval method or hypothesis testing method. In the former, the main objective is to obtain narrow intervals with high reliability. In the latter, the hypothesis is concerned with testing whether the sample estimate is equal to some specific value.

Null hypothesis

This hypothesis states that there is no difference between the control and the study group in relation to randomized controlled trial (RCT). Rejecting or disproving the null hypothesis – and thus concluding that there are grounds for believing that there is a difference between the two groups, is a central task in the modern practice of science, and gives a precise criterion for rejecting a hypothesis.^[9,10]

Alternative hypothesis

This hypothesis is contradictory to null hypothesis, i.e., it assumes that there is a difference among the groups, or there is some association between the predictor and the outcome [Figure 1].^[9,10] Sometimes, it is accepted by exclusion if the test of significance rejects

the null hypothesis. It may be one-sided (specifies the difference in one direction only) or two-sided (specifies the difference in both directions).

Type I error (α error) occur if the null hypothesis is rejected when it is true. It represents the chance that the researcher detects a difference between two groups when in reality no difference exists. In other words, it is the chance of false-positive conclusion. A value of 0.05 is most commonly used.

Type II error (β error) is the chance of a false-negative result. The researcher does not detect the difference between the two groups when in reality the difference exists. Conventionally, it is set at a level of 0.20, which translates into <20% chance of a false-negative conclusion. Power is the complement of beta, i.e., (1-beta). In other words, power is 0.80 or 80% when beta is set at 0.20. The power represents the chance of avoiding a false-negative conclusion, or the chance of detecting an effect if it really exists.^[11]

TYPES OF TRIALS

Parallel arm RCTs are most commonly used, that means all participants are randomized in two or more arms of different interventions treated concurrently. Various types of parallel RCTs are used in accordance with the need: Superiority trials which verify whether a new approach is more effective than the conventional from statistically or clinical point of view. Here, the concurrent null hypothesis is that the new approach is not more effective than the conventional approach. Equivalence trials which are designed to ascertain that the new approach and the standard approach are equally effective. Corresponding null hypothesis states that the difference between both approaches is clinically relevant. Non-inferiority trials which are designed to ascertain that the new approach is

		H ₀ : Not out			
Umpire verdict			Hypothesis testing		
Verdict	Actual situation		Decision	Actual situation	
	Not out	Out		H ₀ true	H ₁ true
Not out	Correct	Error	Accept H ₀	1 - α	Type II error (β)
Out	Error	Correct	Reject H ₀	Type I error (α)	Power (1 - β)
				False positive	False negative

Figure 1: Result possibilities during hypothesis tasting. H₀ – Null hypothesis; H₁ – The alternative hypothesis

equal if not superior to the conventional approach. Corresponding null hypothesis is that the new approach is inferior to the conventional one.

PREREQUISITES FOR SAMPLE SIZE ESTIMATION

At the outset, primary objectives (descriptive/analytical) and primary outcome measure (mean/proportion/rates) should be defined. Often there is a primary research question that the researcher wants to investigate. It is important to choose a primary outcome and lock that for the study. The minimum difference that investigator wants to detect between the groups makes the effect size for the sample size calculation.^[7] Hence, if the researcher changes the planned outcome after the start of the study, the reported *P* value and inference becomes invalid.^[11] The level of acceptable Type I error (α) and Type II error (β) should also be determined. The error rate of Type I error (alpha) is customarily set lower than Type II error (beta). The philosophy behind this is the impact of a false positive error (Type I) is more detrimental than that of false negative (Type II) error. So they are protected against more rigidly.

Besides, the researcher needs to know the control arm mean/event rates/proportion, and the smallest clinically important effect that one is trying to detect.

THE RELATION BETWEEN PRIMARY OBJECTIVE AND THE SAMPLE SIZE

The type of primary outcome measure with its clear definition help computing correct sample size as there are definite ways to reach sample size for each outcome measure. It needs special attention as it principally influences how impressively the research question is answered. The type of primary outcome measure also is the basis for the mode of estimation regarding population variance. For continuous variable (e.g., mean arterial pressure [MAP]), population SD is incorporated in the formula whereas the SD needs to be worked out from the proportion of outcomes for binomial variables (e.g., hypotension - yes/no). In literature, there can be several outcomes for each study design. It is the responsibility of the researcher to find out the primary outcome of the study. Mostly sample size is estimated based on the primary outcome. It is possible to estimate sample size taking into consideration all outcome measures, both primary and secondary at the cost of much larger sample size.

ESSENTIAL COMPONENTS OF SAMPLE SIZE ESTIMATION

The sample size for any study depends on certain factors such as the acceptable level of significance (*P* value), power ($1 - \beta$) of the study, expected 'clinically relevant' effect size, underlying event rate in the population, etc.^[7] Primarily, three factors *P* value (depends on α), power (related with β) and the effect size (clinically relevant assumption) govern an appropriate sample size.^[12-14] The 'effect size' means the magnitude of clinically relevant effect under the alternative hypothesis. It quantifies the difference in the outcomes between the study and control groups. It refers to the smallest difference that would be of clinical importance. Ideally, the basis of effect size selection should be on clinical judgement. It varies with different clinical trials. The researcher has to determine this effect size with scientific knowledge and wisdom. Available previous publications on related topic might be helpful in this regard. 'Minimal clinically important difference' is the smallest difference that would be worth testing. Sample size varies inversely with effect size.

The ideal study to make a researcher happy is one where power of the study is high, or in other words, the study has high chance of making a conclusion with reasonable confidence, be it accepting or rejecting null hypothesis.^[9] Sample size matrix, includes different values of sample sizes using varying dimensions of alpha, power ($1-\beta$), and effect size. It is often more useful for the research team to choose the sample size number that fits conveniently to the need of the researcher [Table 1].

FORMULAE AND SOFTWARE

Once these three factors are fixed, there are many ways (formulae, nomogram, tables and software) for estimating the optimum sample size. At present, there are a good number of softwares, available in the internet. It is prudent to be familiar with the instructions of any software to get sample size of one arm of the study. Perhaps the most important step is to check with the most appropriate formula to get a correct sample size. Websites of some of the commonly used softwares are provided in Table 2.^[2,6]

The number of formulae for calculating the sample size and power, to answer precisely different study designs and research questions are no less than 100. It is wise to check appropriate formula even while

using software. Although there are more than 100 formulae, for RCTs numbers of formulae are limited. It essentially depends on the primary outcome measure such as mean ± SD, rate and proportion.^[6] Interested readers may access all relevant sample size estimation formulae using the relevant links.

Calculating the sample size by comparing two means

A study to see the effect of phenylephrine on MAP as continuous variable after spinal anaesthesia to counteract hypotension.

MAP as continuous variable:

- n = Sample size in each of the groups
- μ_1 = Population mean in treatment Group 1, μ_2 = Population mean in treatment Group 2
- $\mu_1 - \mu_2$ = The difference the investigator wishes to detect
- σ = Population variance (SD)
- a = Conventional multiplier for alpha = 0.05, b = Conventional multiplier for power = 0.80.

$$n = \frac{2([a+b]^2 \sigma^2)}{(\mu_1 - \mu_2)^2}$$

Value of $a = 1.96$, $b = 0.842$ [Table 3]. If a difference of 15 mmHg in MAP is considered between the phenylephrine and the placebo group as clinically significant ($\mu_1 - \mu_2$) and be detected with 80% power and a significance level alpha of 0.05.^[7] $n = 2 \times ((1.96 + 0.842)^2 \times 20^2) / 15^2 = 27.9$. That means 28 subjects per group is the sample size.

Calculating the sample size by comparing two proportions

A study to see the effect of phenylephrine on MAP as a binary variable after spinal anaesthesia to counteract hypotension.

MAP as a binary outcome, below or above 60 mmHg (hypotension – yes/no):

- n = The sample size in each of the groups
- p_1 = Proportion of subjects with hypotension in treatment Group 1
- q_1 = Proportion of subjects without hypotension in treatment Group 1 ($1 - p_1$)
- p_2 = Proportion of subjects with hypotension in treatment Group 2
- q_2 = Proportion of subjects without hypotension in treatment Group 2 ($1 - p_2$)

Table 1: The matrix showing changes of sample size with varying dimensions of alpha, power (1-β), and effect size

Sample size matrix using nMaster			
Effect size	Power (1-β)	Error (α)	Sample size
0.2	80	0.5	586
0.2	80	0.1	773
0.2	90	0.5	746
0.4	80	0.5	146
0.4	80	0.1	193
0.4	90	0.5	186
0.6	80	0.5	65
0.6	80	0.1	86
0.6	90	0.5	83

Table 2: Websites for some useful statistical software

Software for analysis	Website
SPSS	www.spss.com
STATA	www.stata.com
Power and precision	www.poweranalysis.com
Epi-Info	www.cdc.gov/epiinfo
nMaster	www.cmc-biostatistics.ac.in/nmaster
nQuery	www.statsols.com/nquery

Table 3: The constant Z values for conventional α and β values

Z values for conventional α values		
Direction of testing: Two-sided		
Level of significance (%)	α	Z _α value (two-sided)
5	0.05	1.96
1	0.01	2.58
Direction of testing: One-sided		
Level of significance (%)	α	Z _α value (one-sided)
5	0.05	1.65
1	0.01	2.33
Z values for conventional β values		
Power (1-β) (%)	β	Z (1-β) value
80	0.02	0.84
90	0.10	1.28
95	0.05	1.64

- x = The difference the investigator wishes to detect
- a = Conventional multiplier for alpha = 0.05
- b = Conventional multiplier for power = 0.8.

$$n = \frac{([a + b]^2 [p_1q_1 + p_2q_2])}{x^2}$$

Considering a difference of 10% as clinically relevant and from the recent publication the proportion of subjects with hypotension in the treated group will be 20% ($p_1 = 0.2$) and in the control group will be 30% ($p_2 = 0.3$), and thus q_1 and q_2 are 0.80 and 0.70, respectively.^[7] Assuming a power of 80%, and an alpha of 0.05, i.e. 1.96 for a and 0.842 for b [Table 3] we get:

$[(1.96 + 0.842)^2 \times (0.20 \times 0.80 + 0.30 \times 0.70)] / 0.10^2 = 290.5$. Thus, 291 is the sample size.

Researcher may follow some measures like using continuous variables as the primary outcome, measuring the outcome precisely or choose outcomes that can be measured properly. Use of a more common outcome, making one-sided hypothesis may help achieving this target. Published literature and pilot studies are the basis of sample size calculation. At times, expert opinions, personal experience with event rates and educated guess becomes helpful. Variance, effect size or event rates may be underestimated during calculation of the sample size at the designing phase. If the investigator realizes that this underestimation has led to 'too small a sample size' recalculation can be tried based on interim data.^[15]

SUMMARY

Sample size calculation can be guided by previous literature, pilot studies and past clinical experiences. The collaborative effort of the researcher and the statistician is required in this stage. Estimated sample size is not an absolute truth, but our best guess. Issues such as anticipated loss to follow-up, large subgroup analysis and complicated study designs, demands a larger sample size to ensure adequate power throughout the trial. A change in sample size is proportional to variance (square of SD) and inversely proportional to the detected difference.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

REFERENCES

1. Moher D, Dulberg CS, Wells GA. Statistical power, sample size, and their reporting in randomized controlled trials. *JAMA* 1994;272:122-4.
2. Chan YH. Randomised controlled trials (RCTs) – Sample size: The magic number? *Singapore Med J* 2003;44:172-4.
3. Abdulatif M, Mukhtar A, Obayah G. Pitfalls in reporting sample size calculation in randomized controlled trials published in leading anaesthesia journals: A systematic review. *Br J Anaesth* 2015;115:699-707.
4. Noordzij M, Tripepi G, Dekker FW, Zoccali C, Tanck MW, Jager KJ. Sample size calculations: Basic principles and common pitfalls. *Nephrol Dial Transplant* 2010;25:1388-93. Erratum in: *Nephrol Dial Transplant* 2010;25:3461-2.
5. Bacchetti P, Wolf LE, Segal MR, McCulloch CE. Ethics and sample size. *Am J Epidemiol* 2005;161:105-10.
6. Zodpey SP. Sample size and power analysis in medical research. *Indian J Dermatol Venereol Leprol* 2004;70:123-8.
7. Gupta KK, Attri JP, Singh A, Kaur H, Kaur G. Basic concepts for sample size calculation: Critical step for any clinical trials! *Saudi J Anaesth* 2016;10:328-31.
8. Woken MD. The Center for Teaching and Learning, University of Illinois. Available from: <http://www.uis.edu/ctl/wp-content/uploads/sites/76/2013/03/ctlths7.pdf>. [Last accessed on 2016 Aug 24].
9. Whitley E, Ball J. Statistics review 4: Sample size calculations. *Crit Care* 2002;6:335-41.
10. Farrugia P, Petrisor BA, Farrokhyar F, Bhandari M. Practical tips for surgical research: Research questions, hypotheses and objectives. *Can J Surg* 2010;53:278-81.
11. Noordzij M, Dekker FW, Zoccali C, Jager KJ. Sample size calculations. *Nephron Clin Pract* 2011;118:c319-23.
12. Zhong B. How to calculate sample size in randomized controlled trial? *J Thorac Dis* 2009;1:51-4.
13. Farrokhyar F, Reddy D, Poolman RW, Bhandari M. Why perform *a priori* sample size calculation? *Can J Surg* 2013;56:207-13.
14. Clark T, Berger U, Mansmann U. Sample size determinations in original research protocols for randomised clinical trials submitted to UK research ethics committees: Review. *BMJ* 2013;346:f1135.
15. Wittes J. Sample size calculations for randomized controlled trials. *Epidemiol Rev* 2002;24:39-53.

Announcement

Members of Indian Society of Anaesthesiologists (ISA): Registration for IJA

Those members of ISA not registered for online / hard copy for 2017 issues can register for either or both at www.ijaweb.in or send request with details, by E mail to the Editor in Chief, IJA (Name, Online / Hard Copy / Both, Date of Birth, ISA No., e mail i.d., Mobile no., Address with PIN code (mandatory). Last Date is Sep. 30th 2016.

Those registered in 2016 and desirous of receiving hard copies in 2017 (Jan to Dec issues) should renew their registration between August 1st and September 30th; click on Renewal Button in the website (www.ijaweb.in). Renewals after Sep. 2016 are not allowed.

Editor In Chief

E mail: editorija@yahoo.in/sbalabhaskar@gmail.com