Sample size calculation in clinical research

Priya Ranganathan, Vishal Deo[1], C. S. Pramesh[2]

Department of Anaesthesiology, Tata Memorial Centre, Homi Bhabha National Institute, [2]Department of Surgical Oncology and Administration, Tata Memorial Centre, Homi Bhabha National Institute, Mumbai, Maharashtra, [1]National Institute for Research in Digital Health and Data Science, Indian Council of Medical Research, New Delhi, India

Abstract Calculation of sample size is an essential part of research study design since it affects the reliability and feasibility of the research study. In this article, we look at the principles of sample size calculation for different types of research studies.

Keywords: Epidemiologic methods, research design, sample size

Address for correspondence: Dr. Priya Ranganathan, Department of Anaesthesiology, Tata Memorial Centre, Homi Bhabha National Institute, Mumbai - 400 012, Maharashtra, India.
E-mail: drpriyaranganathan@gmail.com
Received: 30-05-24, **Accepted:** 04-06-24, **Published:** 04-07-24.

INTRODUCTION

In previous articles, we have looked at a variety of research study designs.[1-5] Once the research design, objectives and endpoints have been finalized, the researcher needs to calculate the sample size for the study. The sample size is the number of participants or other units required in a study to be able to answer the research question reliably. The sample size drives the budget of the study, allows the researcher to determine the feasibility of the study, and could lead to changes in the proposed study design, methodology, or outcome.

WHY IS SAMPLE SIZE DETERMINATION IMPORTANT?

The sample size of a study determines the reliability of the study results. The internal validity of a study, especially in terms of precision or power, depends largely on the sample size.[6] The sample size also impacts the cost and duration of the study. A study with an adequate sample size is sufficiently powered to detect a treatment effect. As discussed in the next section, a study that is too small may not be able to detect an effect reliably, as the smaller sample size leads to a wider confidence interval of the estimated effect. This, in turn, increases the chance of an observed effect to be statistically insignificant, and the study may fail to detect an existing important treatment effect. On the other hand, a study that is too large is overpowered and may detect effects that are statistically significant but are too small to be of clinical relevance. Even if the real effect size is considerably large and clinically relevant, opting for larger than the required sample size may be deemed unethical for unnecessarily subjecting more participants to an inferior treatment in the control arm. In addition, this may lead to substantial wastage of resources and affect the feasibility of the study. Thus, it is important that the sample size is calculated properly using relevant methods to power the study adequately.

Irrespective of the type of study, the determination of sample size requires information on certain parameters associated with the outcome of interest. Since many

Access this article online	
Quick Response Code:	Website: www.picronline.org
	DOI: 10.4103/picr.picr_100_24

How to cite this article: Ranganathan P, Deo V, Pramesh CS. Sample size calculation in clinical research. Perspect Clin Res 2024;15:155-9.

of these parameters may be estimates (and not true measurements), the researcher should note that sample size calculations are approximations and not absolute measures. In the subsequent sections, we will discuss sample size calculation for some commonly used research study designs, such as descriptive and comparative or analytical studies (clinical trials, cross-sectional studies, case–control studies, and cohort studies).

KEY CONCEPTS IN SAMPLE SIZE CALCULATION

Population versus sample

Studies are based on samples that are considered to be representative of the population. Since the sample is a subset of the population, the sample-based estimates are likely to differ from the true population values. Moreover, since samples are selected randomly from the population, if the same study is conducted repeatedly, we will get different estimates of the true but unknown population measure.[7] Thus, we may practically visualize a range of estimates being produced if the same study is repeated under the same setting. With a valid study design and an adequate sample size, the width of this range is minimized at a given level of confidence (discussed later in this section). Therefore, the sample result is always reported as a summary statistic with a range of values that represent the possible interval containing the true population values, and the width of the interval depends on the level of confidence determined by how certain the researcher wants to be about the estimate.

Size of the population

Sample size estimates require knowledge of the size of the population. Most often, this is unknown and is taken as infinite (for very large populations). Occasionally, the researcher may have a more definite idea of the size of the population and may be able to provide an estimate. For example, Alghamdi *et al* looked at work-related stress and burnout levels among Saudi commercial pilots.[8] Since this was a defined and small population, they were able to estimate the total number of pilots (the population size) to be approximately 2000.

Hypothesis

Every research study starts with a hypothesis, which is a statement of belief that the researcher wishes to prove. Since this statement forms the basis of the research, it is often called a research hypothesis.[9] While constructing a statistical test within a study to scientifically prove or disprove this belief, a null hypothesis is defined, suggesting that the effect being claimed does not exist. This is complemented by an alternate hypothesis which is in agreement with the research hypothesis. The null and alternate hypotheses are statistical hypotheses. Based on the empirical evidence produced by the study, we either accept or reject the null hypothesis. Rejection of the null hypothesis results in acceptance of the alternate hypothesis, suggesting the presence of a statistically significant effect. For example, for comparative studies, the null hypothesis may be stated as "there is no difference between groups" and an alternate hypothesis as "there is a difference between groups." For single-arm studies such as prevalence studies, where we want to detect or estimate the prevalence of a disease or condition, the prevalence is taken as zero in the null hypothesis and is taken as the estimate, say "x" percent ($x > 0$), in the alternate hypothesis based on the sample estimate. To be precise, the research hypothesis needs to be laid out clearly at the start of the study as it determines the primary objective and the basis for sample size calculation.

Summary statistic

The study outcome is reported as a single statistic that serves as an estimate of the population measure of interest and is calculated as a function of sample observations. The choice of statistic depends on the population measure to be estimated, which further depends on the objectives of the study. This could be a mean, a proportion, an odds ratio, a risk ratio, a hazard ratio, etc. The calculation of sample size depends on the type of statistic that will be used.

Expected event rates

Both single-arm and comparative studies require knowledge of the baseline event rate (prevalence in single-arm studies and control group event rate in comparative studies). In addition, for comparative studies, the researcher needs to know the event rate in the comparator group (the effect size). This can be estimated by scoping the published literature, reviewing clinical data, by carrying out pilot studies, or by "guesstimation" but only in case none of the other options are feasible.

Variance

It is a concept used for numerical data and refers to the degree of spread of data within the population. The larger the spread (the more heterogenous the data), the higher the sample size that will be needed to estimate outcomes. Since it is not possible to directly measure the variance in the population, we often assume that the sample variance represents the population variance.

Type 1 error/Level of confidence

This represents the certainty with which the estimated intervals of the summary statistic contain the true

value of the population measure being estimated. It is conventionally set at 5% (which means that at least 95 out of the 100 times we conduct the study under the same setting, the resultant intervals of the estimate will contain the true population value).[7,9]

SAMPLE SIZE FOR SINGLE GROUP STUDIES (DESCRIPTIVE OR CROSS-SECTIONAL SURVEYS OR PREVALENCE STUDIES)

These studies aim to measure the prevalence of a particular factor in the population. The inputs needed for sample size calculation are:

a. The predicted or expected prevalence in the population - Sample sizes are very high for extremely low (<10%) and extremely high (>90%) prevalence rates. Between prevalence rates of 10 and 90%, the sample size is maximum for prevalence rates of 50%. Sometimes, it is suggested that an arbitrary value of 50% may be used if the prevalence is completely unknown. However, this approach has its drawbacks
b. The degree of precision - This represents the margin of error with which we want to estimate the prevalence and determines the width of the confidence intervals of the estimated prevalence. For example, if we want to estimate a prevalence of 25% with a 2% margin of error, it means that we want to estimate the prevalence to be between 23% and 27%
c. The level of confidence
d. Population size – It can be either known (a finite number) or be assumed to be unknown (infinite). The mathematical formula for calculating sample size is different for the two scenarios. In general, it is preferred to assume the population to be infinite, unless the target population is small
e. Degree of attrition or nonresponse.

Readers can refer to other papers for further information on this topic.[10,11]

EXAMPLES IN PUBLISHED LITERATURE

Al-Ramahi carried out a cross-sectional study of adherence to medications among Palestinian hypertensive patients.[12] Since there was no available literature showing the prevalence of adherence among the Palestinian community, a 50% expected prevalence was used. The calculated sample size was 384, and to account for attrition, 500 patients were included in the study. In the study by Alghamdi *et al*, the sample size of 311 was determined considering a 5% margin of error, 95% confidence level, and a 60.3% burnout prevalence rate.[8] Zaidi *et al*. surveyed the awareness and practice of self-medication among the general public in Jeddah and Makkah.[13] At a 95% confidence level with an estimated 50% response distribution, a margin of error ± 5%, and accounting for a 5% nonresponse rate, the estimated sample size was 404.

SAMPLE SIZE FOR CLINICAL TRIALS

A typical two-arm phase 3 clinical trial with a superiority hypothesis aims to prove that a treatment is superior to a control. The inputs needed for sample size calculation are:[14]

a. Type 1 error: This refers to the probability of falsely rejecting the null hypothesis. For a superiority study, it refers to the probability of finding a difference by chance where a true difference does not exist
b. Type 2 error: This refers to the probability of falsely accepting the null hypothesis. For a superiority study, it refers to the probability of missing a true treatment effect. The power of the study is the converse of the type 2 error and refers to the ability to find the existing difference. Conventionally, the type 2 error is set as 10% or 20%, which means that the power of the study is 90% or 80%
c. Expected effect size: Refers to the minimum difference in effect between the study arms that would be considered significant. The effect size is inversely proportional to the sample size. Very small effect sizes require large studies and may not be feasible or clinically meaningful. Very large effect sizes require small studies but may not be scientifically plausible and will result in negative studies
d. Variance of the outcome measure
e. Allocation ratio between control and experimental arms: Most often, participants are randomized in a 1:1 ratio between experimental and control arms. However, sometimes, investigators may choose to allot additional patients to the experimental arm or to the control arm, which will impact sample size.

EXAMPLES IN PUBLISHED LITERATURE

Fernandez-Lopez compared flexible and conventional treatments for gestational diabetes mellitus.[15] They found that the variance in the weight of newborn babies of mothers with gestational diabetes mellitus was 1.8 kg. To detect a difference in mean weight of 700 g between newborns in the two groups, at 95% confidence and 80% power, it would be necessary to have 82 patients per group. Assuming 15% attrition, 96 patients were needed per group.

The DREAMS study hypothesized that preoperative dexamethasone would reduce postoperative vomiting in

patients undergoing elective bowel surgery.[16] The initial sample size was calculated at 80% power and 5% type I error, based on a 24% proportional reduction in the number of participants experiencing vomiting in the first 24 h after surgery (corresponding to a reduction from 37% to 28% based on an earlier large trial). After accounting for a 10% loss to follow-up, the sample size was determined to be 950 patients.

SAMPLE SIZE FOR COMPARATIVE COHORT STUDIES

Cohort studies include groups of exposed and nonexposed individuals who are followed up to determine outcomes. The inputs needed for sample size calculation are similar to that of a clinical trial, and include the power of the study, the type 1 error, the probability of outcomes in the exposed and unexposed groups, and the ratio of exposed to unexposed participants.[17]

SAMPLE SIZE FOR COMPARATIVE CASE–CONTROL STUDIES

Case–control studies include individuals with the outcome (cases) and without the outcome (controls) who are then assessed for exposure status. The sample size depends on the power of the study, the type 1 error, the ratio of cases to controls, and the expected odds of exposure in the cases versus the controls.[18]

ADDITIONAL ASPECTS

- The calculated sample size for comparative studies usually refers to the sample required per arm of the study.
- Guidelines for reporting research studies mandate that all aspects of sample size calculation should be reported in the research manuscript. This allows readers and reviewers to assess whether there has been any bias in the conduct of the study.
- In some cases, sample size calculations may be more complex. The sample size may need to be adjusted for interim analyses if the overall type 1 error is to be maintained at the desired level. Factorial designs, cluster randomized designs, equivalence and noninferiority trials, and early phase trials, all require additional inputs for sample size calculation.[19] Similarly, sample size calculation may be based on study outcomes such as correlation, association, agreement, sensitivity, or specificity.[20,21]
- A variety of online tools are available for the calculation of sample size. However, readers need to be aware that there are many nuances to sample size calculation, and it is best to take expert input from a statistician rather than rely completely on these tools.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

REFERENCES

1. Aggarwal R, Ranganathan P. Study designs: Part 2 – Descriptive studies. Perspect Clin Res 2019;10:34-6.
2. Ranganathan P, Aggarwal R. Study designs: Part 3 – Analytical observational studies. Perspect Clin Res 2019;10:91-4.
3. Aggarwal R, Ranganathan P. Study designs: Part 4 – Interventional studies. Perspect Clin Res 2019;10:137-9.
4. Ranganathan P, Pramesh CS, Aggarwal R. Equivalence trials. Perspect Clin Res 2022;13:114-7.
5. Ranganathan P, Pramesh CS, Aggarwal R. Non-inferiority trials. Perspect Clin Res 2022;13:54-7.
6. Shih WJ, Aisner J. Statistical Design, Monitoring, and Analysis of Clinical Trials: Principles and Methods. Boca Raton: Chapman and Hall/CRC; 2021.
7. Ranganathan P, Pramesh CS, Buyse M. Common pitfalls in statistical analysis: "P" values, statistical significance and confidence intervals. Perspect Clin Res 2015;6:116-7.
8. Alghamdi AA, Alghamdi AH. Determining the best approach: Comparing and contrasting the impact of different coping strategies on work-related stress and burnout among Saudi commercial pilots. Cureus 2023;15:e41948.
9. Ranganathan P, Cs P. An introduction to statistics: Understanding hypothesis testing and statistical errors. Indian J Crit Care Med 2019;23:S230-1.
10. Naing L, Nordin RB, Abdul Rahman H, Naing YT. Sample size calculation for prevalence studies using Scalex and ScalaR calculators. BMC Med Res Methodol 2022;22:209.
11. Khaled Fahim N, Negida A. Sample size calculation guide – Part 1: How to calculate the sample size based on the prevalence rate. Adv J Emerg Med 2018;2:e50.
12. Al-Ramahi R. Adherence to medications and associated factors: A cross-sectional study among Palestinian hypertensive patients. J Epidemiol Glob Health 2015;5:125-32.
13. Zaidi SF, Hakami AY, Khan MA, Khalid AA, Haneef AK, Natto SS, *et al*. The awareness and practice of self-medication among the general public in Jeddah and Makkah. Cureus 2023;15:e39706.
14. Negida A, Fahim NK, Negida Y, Ahmed H. Sample size calculation guide – Part 5: How to calculate the sample size for a superiority clinical trial. Adv J Emerg Med 2019;3:e49.
15. Fernández-López M, Blanco-Carnero JE, Guardia-Baena JM, de Paco-Matallana C, Aragón-Alonso A, Hernández-Martínez AM. Flexible treatment of gestational diabetes mellitus adjusted according to intrauterine fetal growth versus treatment according to strict maternal glycemic parameters: A randomized clinical trial. BMJ Open Diabetes Res Care 2022;10:e002915.
16. DREAMS Trial Collaborators and West Midlands Research Collaborative. Dexamethasone versus standard treatment for postoperative nausea and vomiting in gastrointestinal surgery: Randomised controlled trial (DREAMS trial). BMJ 2017;357:j1455.
17. Khaled Fahim N, Negida A. Sample size calculation guide – Part 2: How to calculate the sample size for an independent cohort study. Adv J Emerg Med 2019;3:e12.
18. Fahim NK, Negida A, Fahim AK. Sample size calculation guide – Part 3: How to calculate the sample size for an independent case-control study. Adv J Emerg Med 2019;3:e20.

19. Negida A. Sample size calculation guide – Part 6: How to calculate the sample size for a non-inferiority or an equivalence clinical trial. Adv J Emerg Med 2020;4:e15.
20. Negida A, Fahim NK, Negida Y. Sample size calculation guide – Part 4: How to calculate the sample size for a diagnostic test accuracy study based on sensitivity, specificity, and the area under the ROC curve. Adv J Emerg Med 2019;3:e33.
21. Negida A. Sample size calculation guide – Part 7: How to calculate the sample size based on a correlation. Adv J Emerg Med 2020;4:e34.