

Sample Size Estimation in Clinical Trials

Divyangkumar N Patel¹

¹Dr N. D. Desai Faculty of Medical Science and Research, Dharmsinh Desai University, Nadiad, India

DOI: 10.55489/njcm.150620243815

ABSTRACT

Sample size estimation remains as a cornerstone in the meticulous planning and execution of clinical trials, pivotal for ensuring studies possess the requisite statistical power to discern meaningful treatment effects. Insufficient sample sizes compromise the robustness of findings, whereas excessively large samples inflate costs and compromise data integrity. This article meticulously explains the multifaceted factors that outline sample size determination, encompassing various factors such as research design, types of hypotheses, error thresholds, effect size considerations, validity and precision. It investigates into the scope of methodologies available for sample size computation, spanning from intricate statistical formulas to pragmatic tabular approaches. Moreover, it underscores the significance of post-hoc power analysis in retrospectively evaluating completed studies, shedding light on their statistical robustness. This literature review furnishes a nuanced understanding of the sample size estimation landscape in clinical trials, delineating their strengths, limitations, and real-world applications. Anticipating participant attrition assumes paramount importance for proactively adjusting sample sizes, ensuring studies remain methodologically sound. Equipped with a profound grasp of these principles, researchers are empowered to conduct scientifically rigorous and impactful clinical trials, furnishing compelling evidence to inform judicious decision-making in healthcare interventions.

Key Words: Sample size, Statistically significant, Clinically significant, Type I error, Type II error, Power

ARTICLE INFO

Financial Support: None declared

Conflict of Interest: None declared

Received: 14-02-2024, **Accepted:** 24-04-2024, **Published:** 01-06-2024

Correspondence: Dr Divyangkumar Patel (Email: drdivpatel@gmail.com)

How to cite this article: Patel DN. Sample Size Estimation in Clinical Trials. Natl J Community Med 2024;15(6):503-508. DOI: 10.55489/njcm.150620243815

Copy Right: The Authors retain the copyrights of this article, with first publication rights granted to Medsci Publications.

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-Share Alike (CC BY-SA) 4.0 License, which allows others to remix, adapt, and build upon the work commercially, as long as appropriate credit is given, and the new creations are licensed under the identical terms.

www.njcmindia.com | pISSN: 0976-3325 | eISSN: 2229-6816 | Published by Medsci Publications

INTRODUCTION

Conducting medical research demands substantial resources, including manpower, time, and materials. A significant challenge arises when, upon completing a study, investigators realize that the research lacks the necessary statistical power to accurately estimate the significance of results, thereby compromising the study's ability to reliably estimate the observed outcomes.^{1,2}

Inadequate sample size raises concerns about the validity of observed findings. Both insufficient and excessively large sample sizes, compared to the required size, present challenges. A small sample size is associated with missed effects and imprecise estimates, while a large sample size leads to increased costs and compromised data quality. Consequently, determining the optimal sample size for each study is essential.^{1,2}

Randomized Controlled Trials (RCTs) are widely regarded as the gold standard for evaluating healthcare interventions. Unlike observational studies, randomization in RCTs is a powerful method for balancing confounding factors between treatment groups, effectively eliminating the impact of potential variables that could skew results. When designing a clinical trial, determining the optimal sample size is a crucial consideration to ensure significant results.¹

In the study by Pierre Charles et al., it was found that 5% of RCTs did not provide any information regarding sample size calculations, and 43% failed to report all the necessary parameters required for calculating sample size. Furthermore, in 30% of the RCTs, there was a deviation exceeding 10% between the reported sample size in the article and the replicated sample size calculation. The reporting of sample size calculations remains inadequate, often containing errors and relying on assumptions that are frequently inaccurate.³

The main goal of determining sample size is to ascertain a suitable number of participants for a specific study design, ensuring a balance between obtaining meaningful results and optimizing resource utilization.

Sample size challenges in a study are addressed in two stages: during the planning phase, researchers calculate the optimal size using relevant methods, and in situations with constraints like finances or rare diseases, post-hoc power analysis is conducted after completing the study with available sample.

This article aims to elucidate the factors that impact sample size estimation in clinical trials and provide guidance for appropriate sample size calculation.

Key Factors Influencing Sample Size Determination

Sample size calculation depends on various factors

such as research design, primary outcome measure, hypothesis (one or two-tailed), precision, type I error, type II error, power, effect size, design effect, etc. To grasp the principles of sample size calculation and power analysis, it is essential to comprehend these commonly used terms.

2.1 Null hypothesis and alternate hypothesis⁴: The null and alternative hypotheses are two mutually exclusive statements about a population. The null hypothesis states that no difference exists among treatments or interventions, null hypothesis require to test. Conversely, the alternative hypothesis contradicts the null hypothesis, postulating a difference among treatments or interventions. If the alternative hypothesis cannot be directly tested, it is accepted, if the test of significance rejects the null hypothesis.

Two types of alternative hypotheses exist: the one-tailed hypothesis, specifying a difference in one direction only, and the two-tailed hypothesis, specifying a difference in either direction. For instance, a one-tailed hypothesis might suggest that Drug A is more effective compared to Drug B for the treatment of Diabetes, while a two-tailed hypothesis proposes a difference in effectiveness among Drug A and Drug B in diabetes treatment, either higher or lower effectiveness.

The choice between one- or two-tailed approaches depends on the clinical or biological significance of the research question and prior knowledge about the effect or association. This decision should not be driven by sample size considerations only, as it hinges on the contextual importance of the research question. Value of Z_{α} is lower for one sided hypothesis compared to two-sided hypothesis so sample size lower for one side hypothesis compared to two side hypotheses.

2.2 Type I (α) error, Type II (β) error and Statistical Power ($1-\beta$): In a study involving a sample population, discrepancies between the derived results and reality are termed as "errors". When study carried out among sample population there may be four possibilities exist in identifying differences between two treatments: (1) correctly concluding no difference among two treatments, (2) incorrectly concluding a difference among treatments when none exists, (3) incorrectly concluding no difference among treatment when difference present and (4) correctly concluding a difference among treatments.

Sometimes in reality null hypothesis is true mean treatments are not different but we conclude that treatments are different means we reject the null hypothesis and accept the alternative hypothesis; this the error in experiment and it called as Type I error or α error. The probability of making α error is called as level of significance (P value). Generally, it is considered as 0.05 (5%). What does $P < 0.05$ mean? It means the probability of α error is less than 0.05 (or 1 in 20).

Table 1: Possible outcomes of a randomized trial: type I, type II errors and power

Hypothesis testing outcome		REALITY	
		Treatments are not difference The null hypothesis (H0) is true	Treatments are different The Alternative hypothesis (H1) is true
RESEARCH	Conclude treatments are not different The null hypothesis (H0) is true	Correct decision (Cell a)	Type II error β error (Cell b)
	Conclude treatments are different The Alternative hypothesis (H0) is true	Type I error α error (Cell c)	Correct decision Power (1- β) (Cell d)

If you see the right half of Table- 1, there may be two types of possibilities of result. When in reality treatments are different, one possibility is result conclude that treatments are different (Cell d) and other possibility is result conclude that treatments are not different, this is an error, it is called as type II error or β error. It is the situation when in reality null hypothesis is rejected but results conclude that null hypothesis is true.

Power is the correctly identifying that treatments are different mean in reality alternative hypothesis is true and result also conclude that alternative hypothesis is true. It is denoted by 1- β . In simple meaning, power of study is the number of participants require to avoid type II error.

Table 2 provides values for Z_α and Z_β for selected α and β . Sample size is inversely proportion to Type I and Type II error and directly proportional to Power (1- β).

Table 2: Unit normal deviates Z_α and Z_β for select-ed values of α or β

α or β	One sided Z_α and Z_β	Two sided Z_α
0.05	1.64	1.96
0.10	1.28	1.64
0.20	0.84	1.28

Z_β is the same for one sided and two-sided tests

2.3 Error: Two types of errors can be introduced into a study: random error and systematic error. Random error is a result of chance, while systematic error is not solely due to chance. Systematic error is challenging to measure accurately. Random error can be minimized by employing strategies such as averaging, increasing sample size, and repeating the experiment to the acceptable level.

2.4 Effect Size: Effect size (ES) denotes the minimum difference that an investigator aims to detect between study groups, often referred to as the minimal clinically relevant difference. It signifies the magnitude of an effect in the alternative hypothesis, encapsulating the smallest difference that holds clinical or biological significance.^{5,6} Its relevance varies between studies; for instance, a 1% reduction in mortality may be significant, while a 20% reduction in transient asthma might not be clinically relevant. Increasing the effect size reduces type II error, and

power is contingent on both effect size and sample size. Larger sample sizes are needed for 'small effects' to achieve a given power compared to 'large effects'.⁵ The estimation of effect size can be accomplished through three methods: pilot studies, data from previous reports, or informed guesses derived from clinical experiences.⁶

2.5 Validity and Precision: Validity means the degree to which the variable actually measure what it is supposed to measure.⁸ It is a function of systematic error or bias. Precision, or reliability, indicates how consistently a variable maintains the same value across multiple measurements.² It is a measure of consistency and can be related to the width of the confidence interval. Precision is influenced by factors such as random error, sample size, confidence interval width, and variance of the outcome variable. A larger sample size contributes to more precise estimates^{2,8}.

2.6 Design effect: Cluster sampling is a cost-effective and practical approach compared to simple random sampling, simplifying studies by grouping units into clusters. When opting for cluster sampling instead of simple random sampling, it becomes necessary to adjust the sample size using a correction factor known as the "design effect." This design effect is the ratio of variance observed in a cluster survey to that in a simple random sample (SRS) of the same size.

The design effect can be determined from prior studies and is often expressed as a constant, typically ranging between 1 and 2. In practical applications, a design effect of 2 is commonly assumed. This assumption implies that, to achieve the same level of precision as with SRS, twice as many individuals would need to be included in the study when employing cluster sampling. Ideally, the design effect should be derived from previous data or pilot studies for more accurate estimation.

2.7 Type of clinical types¹: There are various clinical trial designs, with the parallel randomized controlled trial (RCT) being the most common. In this design, participants are randomly assigned to different interventions simultaneously, typically comparing a new treatment to a standard one.

- Superiority trials aim to establish that a new treatment is statistically or clinically more effective than a standard treatment. The null hypothe-

sis in this case states that the new treatment is not more efficacious than the control treatment.

- Equivalence trials seek to demonstrate that the new treatment and the standard treatment are equally effective. The null hypothesis for equivalence trials posits that the two treatments differ by a clinically relevant amount.
- Non-inferiority trials are conducted to show that the new treatment is as effective as, but not necessarily superior to, the standard treatment. The corresponding null hypothesis is that the new treatment is inferior to the control treatment by a clinically relevant amount.

Superiority and non-inferiority trials typically employ a one-sided test, while equivalence trials use a two-sided test. The sample size may be lower for superiority and non-inferiority trials compared to equivalence trials. Selection of one sided or two-sided hypothesis or test should not be based on the sample size consideration, it must be based on clinical

significance. It is always better to use two sided hypothesis and test.⁹

Sample size estimation methods

The determination of sample size can be achieved through the using mathematical formulas, ready-made reference tables, or specialized software packages. Essential information necessary for the calculation of sample size includes various parameters, which are the study design, anticipated effect size, desired statistical power, the requisite level of significance, and the design effect.

Statistical Formula:

The formula for calculating the sample size in a clinical trial varies based on the primary outcome, whether it is a mean or a proportion. The specific calculation formula is outlined below.

Example:

Sample size for comparative studies: means

To calculate the sample size, a previous study was taken in to account in which the Drug A was reduced a 50 mg/dl blood pressure after 3 weeks of treatment with SD 20 md/dl and Drug B was reduced a 40 mg/dl blood pressure after 3 weeks of treatment with SD of 10 mg/dl. If we desire an equal sample size in both groups, what would be the minimal sample size in each group to detect a difference with a power of 80% at 95% confidence level?

$$n = \frac{\left[\left(Z_{\alpha/2} + Z_{\beta} \right)^2 \times \{2(\sigma)^2\} \right]}{(\mu_1 - \mu_2)^2}$$

Where,

n = sample size required in each group,

μ_1 = mean change in pain score from baseline to week 24 in Drug A = 50,

μ_2 = mean change in pain score from baseline to week 24 in Active drug = 40,

$\mu_1 - \mu_2$ = clinically significant difference = 10

σ = combined standard deviation = 16.549

$Z_{\alpha/2}$: This depends on level of significance, for 5% this is 1.96

Z_{β} : This depends on power, for 80% this is 0.84

$$n = \frac{[(1.96+0.84)^2 \times 2 \times (16.54)^2]}{(50-40)^2} = \frac{7.84 \times 547}{100} = 43 \text{ in each group}$$

Sample size for comparative studies: proportion

A placebo-controlled randomized trial proposes to assess the effectiveness of Drug A in curing children suffering from otitis media. A previous study showed that proportion of subjects cured by Drug A is 40% and a clinically important difference of 10% as compared to placebo is acceptable.

$$n = \frac{\left(Z_{\alpha/2} + Z_{\beta} \right)^2 [P_1(1 - P_1) + P_2(1 - P_2)]}{(P_1 - P_2)^2}$$

n = sample size required in each group,

p1 = proportion of subject cured by Drug A = 0.40,

p2 = proportion of subject cured by Placebo = 0.30,

p1-p2 = clinically significant difference = 0.10

$Z_{\alpha/2}$: This depends on level of significance, for 5% this is 1.96

Z_{β} : This depends on power, for 80% this is 0.84

$$n = \frac{(1.96 + 0.84)^2 [0.40(1 - 0.40) + 0.30(1 - 0.30)]}{(0.10)^2} = 352 \text{ in each group}$$

Tabular method:

Readymade table are available to calculate the sample size required for each group.

This table is designed for studies where the main outcome is a proportion or binary outcome. It helps to calculate sample size based on the effectiveness of

two interventions derived from previous studies or pilot studies. To determine the sample size, you only need the lower of the two cure rates and the difference in cure rates between the two treatment groups. The term "cure rate" can be substituted with "effectiveness" or the "desired outcome".

Lower of the Two Cure Rates	DIFFERENCES IN CURE RATES BETWEEN THE TWO TREATMENT GROUPS													
	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70
0.05	420	130	69	44	36	31	23	20	17	14	13	11	10	8
0.10	680	195	96	59	41	35	29	23	19	17	13	12	11	8
0.15	910	250	120	71	48	39	31	25	20	17	15	12	11	9
0.20	1,090	290	135	80	53	42	33	26	22	18	16	12	11	9
0.25	1,250	330	150	88	57	44	35	28	22	18	16	12	11	—
0.30	1,380	360	160	93	60	44	36	29	22	18	15	12	—	—
0.35	1,470	370	170	96	61	44	36	28	22	17	13	—	—	—
0.40	1,530	390	175	97	61	44	35	26	20	17	—	—	—	—
0.45	1,560	390	175	96	60	42	33	25	19	—	—	—	—	—
0.50	1,560	390	170	93	57	40	31	23	—	—	—	—	—	—

Source: Gordis Epidemiology – Six edition

Figure 1: Sample size needed in clinical trial ($\alpha=0.05$) and Power=0.80 (Two-sided test)

Lower of the Two Cure Rates	DIFFERENCES IN CURE RATES BETWEEN THE TWO TREATMENT GROUPS													
	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70
0.05	330	105	55	40	33	24	20	17	13	12	10	9	9	8
0.10	540	155	76	47	37	30	23	19	16	13	11	11	9	8
0.15	710	200	94	56	43	32	26	22	17	15	11	10	9	8
0.20	860	230	110	63	42	36	27	23	17	15	12	10	9	8
0.25	980	260	120	69	45	37	31	23	17	15	12	10	9	—
0.30	1,080	280	130	73	47	37	31	23	17	15	11	10	—	—
0.35	1,160	300	135	75	48	37	31	23	17	15	11	—	—	—
0.40	1,210	310	135	76	48	37	30	23	17	13	—	—	—	—
0.45	1,230	310	135	75	47	36	26	22	16	—	—	—	—	—
0.50	1,230	310	135	73	45	36	26	19	—	—	—	—	—	—

Source: Gordis Epidemiology – Six edition

Figure 2: Sample size needed in clinical trial ($\alpha=0.05$) and Power=0.80 (one sided test)

Post hoc power analysis

Researchers have the option to calculate power after completing a study using various software packages. An example of a straightforward and open-source tool is OpenEpi, which can be accessed through www.openepi.com. By navigating to the menu bar, users can open the power analysis section and input parameters such as sample size for two groups, effect size, or proportion of effect in both groups.

Example: After conducting a study with a sample size of 352 in each group, we obtained the following effect. What is the statistical power of this study?

- n = sample size in each group=352
- p1 = proportion of subject cured by Drug A = 0.40
- p2 = proportion of subject cured by Placebo = 0.30

Power for Randomized Clinical Trials		
Confidence Interval (%) {two-sided}	95	Enter between 0 and 100, usually 95%
	Treatment Group 1	Treatment Group 2
Sample Size	352	352
Percent with outcome (%)	40	30

Figure 3: Calculation of power of RCT

Power for Randomized Clinical Trials	
	Input Data
Two sided-confidence interval (%)	95
Sample size of Treatment Group 1	352
Percent with outcome in Treatment Group 1 (%)	40
Sample size of Treatment Group 2	352
Percent with outcome in Treatment Group 2 (%)	30
Risk ratio detected	1.3
Power based on:	
Normal approximation	79.56%
Normal approximation with continuity correction	77.2%

Results from OpenEpi, Version 3, open source calculator--PowerRCT
Print from the browser with ctrl-P
or select text to copy and paste to other programs.

Figure: 4 Result of calculation of power of RCT

The determined statistical power is 80%. We previously computed a sample size of 352, considering a similar effect size, with the aim of achieving 80% power and a 95% confidence interval.

In prospective studies, including clinical trials, the anticipated loss to follow-up or participant dropout (d) must be considered when determining the final sample size. The adjusted sample size (N1) to account for potential dropouts is calculated using the formula $N1 = n / (1 - d)$. For instance, if the calculated sample size (n) is 350 and the dropout rate is 10%, the final adjusted sample size (N1) would be $350 / (1 - 0.10) = 389.7$, rounded up to 390 participants.⁷ Dropout rate can be based on historical data from similar studies or pilot studies. A common rule of thumb is to assume a 10-20% dropout rate.

CONCLUSION

Sample size estimation is a crucial aspect of designing clinical trials, as it plays a pivotal role in ensuring the validity and reliability of study outcomes. The re-

sults obtained from such investigations serve as compelling evidence for making informed decisions regarding interventions.

Consequently, it is imperative for investigators to employ rigorous methods for calculating sample size. The utilization of various established approaches allows researchers to determine the optimal number of participants required for a study. A comprehensive understanding of the key factors influencing sample size calculations is essential for researchers to conduct scientifically sound and impactful clinical trials. By comprehensively reviewing the various sample size estimation methods, this research article aims to equip researchers, statisticians, and clinicians with a comprehensive understanding of the principles and practices involved in determining sample sizes for clinical trials.

REFERENCES

1. Zhong B. How to Calculate Sample Size in Randomized Controlled Trial? J Thorac Dis. 2009 Dec;1(1):51-4.
2. Zodpey SP. Sample size and power analysis in medical research. Indian J Dermatol Venereol Leprol. 2004;70(2):123-8.
3. Charles P, Giraudeau B, Dechartres A, Baron G, Ravaud P. Reporting of sample size calculation in randomised controlled trials: review. BMJ. 2009 May 12;338:b1732.
4. Wang X, Ji X. Sample Size Estimation in Clinical Research. Chest. 2020 Jul;158(1):S12-20.
5. 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 Sep;10(3):328.
6. Dahiru T, Aliyu A, Kene TS. Statistics in Medical Research: Misuse of Sampling and Sample Size Determination.
7. Sakpal TV. Sample Size Estimation in Clinical Trial. Perspect Clin Res. 2010;1(2):67-9.
8. Jones SR, Carley S, Harrison M. An introduction to power and sample size estimation. Emerg Med J. 2003 Sep;20(5):453-8.
9. Chadha V. Sample size determination in health studies. NTI Bulletin. 2006;42(3 & 4):55-62.