

How to Determine the Sample Size Needed to Test Our Research Hypothesis?

Emiliano Rossi 

Cardiologist. Research Department. Hospital Italiano de Buenos Aires.
City of Buenos Aires, Argentina.

Acta Gastroenterol Latinoam 2025;55(3):184-187

Received: 31/08/2025 / Accepted: 22/09/2025 / Published online: 30/09/2025 / <https://doi.org/10.52787/agl.v55i3.540>

In clinical research, the validity of the results depends not only on clearly defining our research question and choosing an appropriate design, but also on having an optimal sample size.

What is the sample size?

It is the number of observational units (e.g., patients) that need to be included in the study in order to answer the research question.

Why must it be optimal?

An insufficient sample size carries the risk of failing to detect a true effect (Type II error). On the other hand, an excessively large sample increases study costs (resources, time) and may even detect statistically significant differences that are not clinically relevant.¹

At what stages of the study should it be considered?

Sample size calculation is a step that should not be overlooked. It must be considered early during the planning stage and described in the study protocol.

Once the study is completed, during the writing of the scientific article, the sample size calculation should be reported in the Methods section. Reporting guidelines for clinical trials (CONSORT) and observational studies (STROBE) recommend how to present this information.

What are its statistical foundations?

Statistical inference seeks to draw conclusions about populations based on the analysis of representative sam-

Correspondence: Emiliano Rossi
Email: emiliano.rossi@hospitalitaliano.org.ar

ples. When comparing samples, the goal is to determine whether they come from the same reference population or not. To do this, it is necessary to define:²

- *Null hypothesis (H_0)*: States that there is no difference between the groups being compared, assuming that the compared samples belong to the same reference population.
- *Alternative hypothesis (H_1)*: States that there is a difference between the groups.
- *Type I error (α)*: The probability of rejecting the null hypothesis when it is actually true. Usually set at 0.05, meaning that 5 out of 100 tests may commit this error.
- *Type II error (β)*: The probability of failing to reject the null hypothesis when it is false.

What information is needed for its calculation?

1. *Significance level (α)*: The maximum acceptable probability of committing a Type I error. If $p > 0.05$, H_0 is not rejected.
2. *Statistical power ($1-\beta$)*: The probability of rejecting the null hypothesis when it is false; in other words, the probability of detecting a difference if it truly exists. It is conventionally set at 80% or 90%.
3. *Clinically relevant effect size*: The minimum magnitude of the effect difference between groups that is intended to be detected.
4. *Outcome variability*: Expressed as standard deviation (SD) for continuous variables or as the expected proportion for categorical variables.³

Where is the necessary information obtained?

Obtaining the accurate information for sample size calculation is essential for the success of the study. Underestimating this step jeopardize the ability to answer the research question. The challenge is that part of the required information is the very data intended to be uncovered through the study protocol.

Available information sources include: first, published scientific evidence (clinical trials, meta-analyses, observational studies, registries). It is important to ensure that the populations studied are similar to those being investigated. Second, expert opinion in the specific research area. Finally, pilot studies (considering their limitations).

What is the role of pilot studies?

Pilot studies are small-scale studies that help estimate unknown parameters such as standard deviation, expected proportion, and effect size. However, their main limitation is that due to their small size, these estimates tend to be imprecise, and often come with considerable wide confidence intervals.² This may lead to either overestimation or underestimation of the required sample size. Therefore, pilot studies should be used mainly to assess the feasibility of studies rather than as the sole source of information for sample size calculation.

How is the calculation performed?

The sample size calculation depends on the type of study (equality, superiority, non-inferiority, or equivalence), the sampling method (e.g., simple random), the number of groups to be compared (e.g., two), the allocation ratio (e.g., 1:1), and the effect measure (e.g., mean, proportion, OR, HR, rate).

Since this is an introductory article, mathematical formulas will not be presented. However, it is important to emphasize that the required sample size increases when: the significance level is reduced (e.g., 0.01 instead of 0.05), the statistical power is increased, the clinically relevant effect size is smaller, and/or the variability of the outcome is greater.

What tools are available?

Several tools are available for sample size calculation. These include: free software such as **G*Power** (<https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower>), and R packages like **pwr** or **TrialSize**; commercial software such as **PASS** or **Stata**; and online calculators such as **OpenEpi** (<https://www.openepi.com/>) or **ClinCalc** (<https://clincalc.com/stats/samplesize.aspx>).

The sample size calculation will be shown using two common scenarios in the clinical context. The first one, involves a study in which hypothesis testing involves the comparison of two means, and the second one, compares two proportions. Given its immediate availability and ease of use, the **ClinCalc** online calculator will be used.

Study 1:

Evaluate the 12-month effect of vitamin E on reducing alanine aminotransferase compared with placebo in patients with nonalcoholic steatohepatitis.

To calculate the required sample size, the ClinCalc website is used. Start by selecting “number of groups” (two independent groups) and “primary endpoint” (continuous). Then, enter the expected mean of group 1 and group 2 (remember that the difference between these two represents the effect size) and the anticipated standard deviation. Finally, set alpha and power at conventional values.

With significance level $\alpha = 0.05$, power = 90%, clinically relevant effect = 20 U/L (assuming the control group mean = 120 U/L and Vit. E group = 100 U/L) and SD = 15 U/L, 12 patients per group will be needed.

Study 2:

Evaluate the eradication success of *H. pylori* by comparing standard triple therapy vs. bismuth quadruple therapy.

As in the previous example, select two independent groups, but this time the endpoint is dichotomous. Then, enter the anticipated proportions (incidence) in each group and finally, set alpha and power.

With significance level $\alpha = 0.05$, power = 90%, and estimated eradication rate of 80% for standard therapy vs. 95% for quadruple therapy, 100 patients per group will be needed.

It is important to remember that the calculated required number should be increased by the expected loss to follow-up percentage that might occur in the study (e.g., 10%).

Is *post hoc* power an alternative?

Post hoc or observed power is determined once the study has been completed, i.e., after data have been analyzed and results are known.

Power is a monotonic function of the *p*-value. Therefore, it does not provide any new information. It only confirms what the *p*-value already indicates.³ Non-significant *p*-value will always correspond to low observed power.⁴ In short, power is a planning tool, not intended for retrospective analysis.

Many authors and editorial guidelines recommend not using *post hoc* power. Instead, they suggest reporting the confidence intervals of the effect measure, as these better reflect the precision of the estimate.^{4,5}

What is precision-based sample size?

As previously discussed, the power-based sample size approach seeks to detect a difference between groups,

based on predefined significance level (α) and power ($1-\beta$). In contrast, the precision-based approach focuses on the accuracy of the estimate of the parameter of interest (mean, proportion, etc.). First, a maximum acceptable margin of error (half-width of the confidence interval) must be set, and then the number of individuals required is calculated so that the estimate remains within this margin with a confidence level of $1-\alpha$.²

This approach ensures that the estimate is sufficiently precise, although it does not focus on detecting differences (does not compare samples). It is particularly useful in epidemiological studies, where the goal is to estimate a population parameter (e.g., disease prevalence) from a single group.²

Study 3:

Conduct a cross-sectional study in the general population to estimate the prevalence of *H. pylori* infection in Argentina, with a 95% confidence interval and an absolute margin of error of ± 3 percentage points. As a reference, data from a study reporting that *H. pylori* infection affects 36% of the general U.S. population was used.

For the precision-based sample size calculation, OpenEpi website is used. Select “Sample Size”, then “Proportion”, and click on “Enter Data”. Once there, write the anticipated frequency = 36% and the confidence limits (margin of error) = 3%, leaving remaining options as default values. This will indicate that a sample size of 983 individuals should be randomly included to achieve the specified precision. Finally, an adjustment for anticipated losses or non-responses (e.g., 20%) should be added.

Conclusions

Sample size calculation is a critical step in the planning of any research study. It requires a clearly defined hypothesis and information about the magnitude of the effect considered clinically relevant and the expected variability of the outcome.

An adequate sample size calculation ensures that the study is both valid and efficient, providing sufficient statistical power without including more patients than necessary.

Intellectual Property. The author declares that the data presented in the manuscript are original and were carried out at his belonging institution.

Funding. *The author declares that there were no external sources of funding.*

Conflict of interest. *The author declares that he has no conflicts of interest in relation to this article.*

Copyright



© 2025 Acta Gastroenterológica latinoamericana. This is an open-access article released under the terms of the Creative Commons Attribution (CC BY-NC-SA 4.0) license, which allows non-commercial use, distribution, and reproduction, provided the original author and source are acknowledged.

Cite this article as: Rossi E. How to determine the sample size needed to test our research hypothesis?. Acta Gastroenterol Latinoam. 2025;55(3):184-187. <https://doi.org/10.52787/agl.v55i3.540>

References

1. Hickey GL, Grant SW, Dunning J, Siepe M. Statistical primer: sample size and power calculations - why, when and how? Eur J Cardiothorac Surg 2018;54:4-9.
2. Machin D, Campbell MJ, Tan SB, Tan SH. Sample sizes for clinical, laboratory and epidemiology studies. 4th ed. Chichester (UK): John Wiley & Sons; 2018.
3. Russell V Lenth (2001) Some Practical Guidelines for Effective Sample Size Determination, The American Statistician, 55:3, 187-193.
4. Hoenig JM, Heisey DM. *The abuse of power: The pervasive fallacy of power calculations for data analysis.* The American Statistician. 2001;55(1):19-24.
5. International Committee of Medical Journal Editors (ICMJE). *Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals.* Actualización 2019.