

El tamaño muestral en la investigación quirúrgica

Sample size in surgical research

M. Elena Peña*

El cálculo del tamaño muestral constituye un componente esencial del diseño metodológico en la investigación médica. Su informe en los trabajos científicos mejora la interpretabilidad de los resultados y refuerza tanto la validez interna como la legitimidad ética del estudio. En cirugía, donde muchos eventos clínicos son relativamente infrecuentes y la variabilidad técnica entre operadores puede ser considerable, esta cuestión adquiere una relevancia particular.

A pesar de su importancia, la ausencia del cálculo del tamaño muestral continúa siendo una debilidad frecuente en publicaciones quirúrgicas^{1,2}. Este artículo propone revisar su definición, relevancia, indicaciones y los principios generales de cálculo.

¿Qué es el cálculo del tamaño muestral?

El cálculo del tamaño muestral consiste en la estimación del número de participantes o unidades de observación que deben incluirse en un estudio para responder de manera adecuada a la pregunta de investigación planteada (véase más adelante)^{3,4}.

En cirugía, las unidades de observación suelen ser pacientes o procedimientos quirúrgicos, aunque también pueden ser centros, cirujanos o episodios clínicos específicos. Por ejemplo, en un estudio que analiza todas las apendicectomías realizadas durante el año 2025 en un hospital comunitario, se incluyen 120 procedimientos. En este caso, el tamaño de la muestra corresponde a 120 apendicectomías (n=120).

¿Cuál es la importancia de calcular el tamaño muestral?

Cuando no se realiza un cálculo previo del tamaño muestral pueden incluirse más o menos unidades de observación de las necesarias. Si se incluyen más participantes de los requeridos, se incrementan innecesariamente los costos, el uso de recursos y el tiempo de investigación. Por el contrario, si el número de casos es insuficiente, el estudio puede no tener suficiente

capacidad para detectar una diferencia real entre los grupos analizados. Esta situación se conoce como error de tipo II o falta de potencia estadística^{3,5}. Este error es equivalente a no ver una diferencia que verdaderamente existe en la realidad. Cuando esto sucede, se acepta la hipótesis nula erróneamente, cuando en realidad debería rechazarse³.

Además, existe una dimensión ética. En los estudios de investigación, los pacientes aportan información sensible con el objetivo de contribuir al conocimiento científico. Diseñar estudios con una probabilidad muy baja de responder la pregunta de investigación puede implicar un uso inadecuado de esos datos. Por ejemplo, intentar analizar retrospectivamente una condición extremadamente infrecuente, como la apendicitis del muñón, contando únicamente con tres casos disponibles, difícilmente permitirá obtener conclusiones robustas.

¿Qué tipos de preguntas de investigación existen?

De forma general, las preguntas de investigación pueden clasificarse en descriptivas y analíticas. A cada tipo de pregunta corresponde un enfoque metodológico específico y diferentes estrategias de cálculo del tamaño muestral^{3,5}.

Preguntas descriptivas

Las preguntas descriptivas buscan caracterizar o describir cómo se manifiesta un fenómeno o cómo se distribuye una variable en una población determinada, sin intentar establecer relaciones causales. Para este tipo de preguntas suelen utilizarse diseños como series de casos o estudios transversales.

Ejemplos:

- ¿Cuál es la duración promedio de la internación luego de una hemicolectomía izquierda laparoscópica?
- ¿Cuál es la mortalidad de los pacientes con diagnóstico de pancreatitis aguda grave?

En estos casos, el tamaño muestral se calcula generalmente para estimar un parámetro poblacional (media o proporción) con un determinado nivel de precisión.

Preguntas analíticas

Las preguntas analíticas buscan evaluar la relación entre dos o más variables con el objetivo de determinar si existe una asociación, diferencia o posible efecto causal entre ellas.

En este tipo de estudios se formula una hipótesis que se expresa como:

- Hipótesis nula (H0): no existe diferencia o asociación entre las variables.
- Hipótesis alternativa (H1): existe una diferencia o asociación.

El diseño del estudio intenta poner a prueba la hipótesis nula mediante análisis estadístico. Si los resultados permiten rechazarla, se acepta la hipótesis alternativa, lo que sugiere la existencia de una diferencia o asociación estadísticamente significativa. Este tipo de preguntas suele abordarse mediante diseños analíticos como estudios de cohorte (prospectivos o retrospectivos), estudios de casos y controles o ensayos clínicos.

Por ejemplo: ¿Cuál es la tasa de infección del sitio quirúrgico en apendicectomías realizadas por vía abierta en comparación con las realizadas por vía laparoscópica? En este caso, la hipótesis nula establecería que no existen diferencias en la tasa de infección entre ambos abordajes. Si el análisis estadístico permite rechazar esta hipótesis, se aceptaría la hipótesis alternativa, que plantea que la tasa de infección difiere entre los grupos.

¿Cómo se calcula el tamaño muestral?

En la práctica, el cálculo del tamaño muestral requiere definir varios elementos clave. Debido a su frecuencia en las publicaciones científicas en cirugía, nos referiremos al cálculo del tamaño muestral en estudios analíticos^{1,3,4,6,7}:

1. Pregunta de investigación

El investigador debe tener clara la pregunta de investigación. Esto le permitirá en etapas previas definir el diseño de su estudio. Por ejemplo, para la pregunta "¿Cuál es la tasa de recidiva en pacientes mujeres operadas por hernia femoral por vía laparoscópica en comparación con la vía convencional?". Se trata en este caso de una pregunta analítica, que se puede responder tanto con un estudio de cohorte como con un ensayo clínico. Su elección dependerá de

la factibilidad, recursos disponibles y preferencia del investigador.

2. Variable principal

Es la variable de interés que permite responder la pregunta de investigación. El tipo de variable (categórica o continua) determina el método estadístico utilizado para el cálculo del tamaño muestral. En el ejemplo anterior, la variable de interés es la tasa de recidiva, que corresponde a una variable categórica (sí/no).

3. Magnitud del efecto

Representa la magnitud mínima de diferencia entre grupos que se consideran clínicamente relevantes. Para obtenerla se necesita evaluar la literatura y conocer cómo se comportan estos tratamientos. Es fundamental la mirada y experiencia de los investigadores vinculados con el ámbito del problema, que se encuentran familiarizados con los efectos de ellos.

Siguiendo el ejemplo, el investigador sabe que la tasa de recidiva de hernia femoral en mujeres informada en la literatura ronda el 5-10% para la vía abierta y el 1-3% para la laparoscópica. En este caso, el investigador podría establecer que una reducción absoluta del 5% en la tasa de recidiva entre la reparación abierta y la laparoscópica constituye una diferencia clínicamente significativa para su estudio.

4. Nivel de significación (α)

El nivel alfa (α) representa la probabilidad de cometer un error de tipo I, es decir, rechazar la hipótesis nula cuando en realidad es verdadera. En términos prácticos, implica concluir que existe una diferencia o asociación entre las variables estudiadas cuando en realidad esta diferencia no existe.

En investigación clínica, el nivel alfa suele fijarse convencionalmente en 0,05 (5%). Esto significa que el investigador acepta una probabilidad del 5% de encontrar una diferencia estadísticamente significativa que, en realidad, sea producto del azar.

5. Poder estadístico

El poder estadístico representa la probabilidad de detectar una diferencia real entre los grupos cuando dicha diferencia realmente existe. Se relaciona directamente con el error de tipo II (β), que ocurre cuando el estudio no detecta una diferencia que verdaderamente existe y, por ende, acepta erróneamente la hipótesis nula (véase anterior). Matemáticamente, el poder se expresa como $1-\beta$. En investigación clínica, el poder estadístico suele fijarse convencionalmente en 80-90%, lo que implica aceptar una probabilidad del 20-10% de cometer un error de tipo II.

6. ¿Prueba de 1 o 2 colas?

Al calcular el tamaño muestral también debe definirse si el contraste de hipótesis será de una o de dos colas. En una prueba de dos colas se evalúa la posibilidad de que exista una diferencia entre los grupos en cualquiera de las dos direcciones. Por este motivo, el nivel alfa se divide entre ambos extremos de la distribución. Este es el enfoque más utilizado en investigación clínica. En una prueba de una cola, el análisis considera solo una dirección específica de la diferencia (por ejemplo, que una técnica sea superior a otra). Aunque requiere un tamaño muestral menor, su uso debe estar bien justificado.

7. Ajuste por pérdidas

En estudios prospectivos debe considerarse la posibilidad de pérdidas durante el seguimiento, por lo que el tamaño muestral inicial suele incrementarse para compensarlas. Habitualmente se incluye un 10-20% más de unidades de observación.

Una vez definidos estos parámetros, el tamaño muestral puede calcularse mediante fórmulas estadísticas específicas o utilizando software especializado, como Epi Info[®], SPSS[®], R[®], Stata[®] u otras herramientas disponibles en línea^{5,7}.

¿Es siempre necesario calcular el tamaño muestral?

Idealmente, el tamaño de la muestra debería calcularse siempre antes de iniciar cualquier estudio, tanto para preguntas descriptivas como analíticas³.

Este cálculo adquiere especial relevancia en los estudios analíticos y, particularmente, en los ensayos clínicos. Estos estudios experimentales y prospectivos asignan a los participantes a una o más intervenciones con el objetivo de evaluar su eficacia y seguridad. Los ensayos clínicos suelen ser complejos y costosos, ya que requieren infraestructura, seguimiento prolongado, personal capacitado, insumos y, con frecuencia, financiamiento externo. Un cálculo incorrecto del tamaño muestral puede conducir a incluir menos participantes de los necesarios y no detectar un efecto real, desperdiciando tiempo y recursos. Por el contrario, reclutar más participantes de los indispensables aumenta

innecesariamente los costos y la exposición de los pacientes a intervenciones experimentales.

En los estudios retrospectivos, aunque los datos ya existan y los costos sean menores, también es recomendable estimar el tamaño muestral requerido. Esto permite evaluar si la cantidad de casos disponibles posee suficiente potencia estadística para detectar asociaciones o diferencias relevantes. De lo contrario, resultados no significativos pueden reflejar simplemente una falta de tamaño muestral y no la ausencia de un efecto real.

¿Qué hacer cuando el estudio no alcanza al tamaño muestral previsto?

Cuando no es posible alcanzar el tamaño muestral calculado, pueden considerarse diversas estrategias³:

- Reformular el estudio como exploratorio o generador de hipótesis
- Declarar explícitamente la limitación de la potencia estadística
- Realizar colaboraciones multicéntricas para aumentar el tamaño muestral
- Evitar afirmaciones de equivalencia, si el estudio no fue diseñado específicamente para demostrar no inferioridad o equivalencia

Esta transparencia metodológica fortalece la credibilidad científica incluso cuando existen limitaciones inevitables.

Conclusión

El tamaño muestral representa la traducción cuantitativa de la hipótesis clínica y constituye un pilar fundamental del diseño metodológico. En investigación quirúrgica, su cálculo riguroso resulta indispensable para generar evidencia válida, interpretable y éticamente sustentable. La calidad de nuestras publicaciones no depende únicamente de la destreza técnica en el quirófano, sino también del rigor metodológico con el que diseñamos, analizamos y comunicamos nuestra investigación.

■ ENGLISH VERSION

Sample size calculation is an essential component of the methodological design in medical research. In scientific studies, reporting the sample size improves the interpretability of the results, strengthening both the internal validity and ethical legitimacy of the study. This aspect is particularly

important in surgery, given that many clinical events are relatively rare and surgical techniques can vary considerably from surgeon to surgeon.

Despite its significance, failure to calculate the sample size remains a common shortcoming in surgical publications^{1,2}. The purpose of this article is to review

the definition, significance, indications, and general principles of sample size calculation.

What is sample size calculation?

Sample size calculation involves estimating the number of participants or observation units that must be included in a study to adequately answer the research question posed (see below)^{2,4}.

In surgery, the units of observation are usually patients or surgical procedures, as well as hospitals, surgeons, or specific clinical episodes. For example, a study analyzing all appendectomies performed in 2025 at a community-based hospital included 120 procedures. In this case, the sample size consists of 120 appendectomies ($n=120$).

Why is it important to calculate the sample size?

When the sample size is not preliminarily calculated, the number of observation units included may be greater or fewer than necessary. When more participants than necessary are included, costs, resource usage, and research time are unnecessarily increased. Conversely, if the sample size is too small, the study may not have sufficient power to detect a real difference between the groups being analyzed. This situation is known as a type II error or lack of statistical power³⁻⁵. This error is equivalent to failing to perceive a difference that actually exists. When this happens, the null hypothesis is erroneously accepted, when in fact it should be rejected³.

Moreover, an ethical dimension is also implicated. In research studies, patients provide sensitive information with the aim of contributing to scientific knowledge. The design of studies in which the probability of answering the research question is very low may result in an inappropriate use of the data. For instance, a retrospective analysis of an extremely rare condition, such as stump appendicitis, based on only three available cases, would not likely yield robust conclusions.

What types of research questions exist?

Generally speaking, research questions can be classified as descriptive and analytical. Each type of question requires a specific methodological approach and different strategies for calculating the sample size³⁻⁵.

Descriptive questions

The aim of descriptive questions is to characterize or describe how a phenomenon manifests

itself or how a variable is distributed within a given population, without attempting to establish causal relationships. Case series or cross-sectional studies are commonly used for this type of question.

Examples:

- What is the average length of hospital stay following a laparoscopic left hemicolectomy?
- What is the mortality rate among patients diagnosed with severe acute pancreatitis?

In these cases, the sample size is typically calculated to estimate a population parameter (mean or proportion) with a specified level of precision.

Analytical questions

Analytical questions are designed to evaluate the relationship between two or more variables to determine whether there is an association, a difference, or a possible causal effect between them.

In this type of study, a hypothesis is formulated and expressed as follows:

- Null hypothesis (H0): There is no difference or association between the variables.
- Alternative hypothesis (H1): There is a difference or association.

The study design aims to test the null hypothesis using statistical analysis. If the results reject the null hypothesis, the alternative hypothesis is accepted, suggesting the existence of a statistically significant difference or association. These types of questions are typically addressed using analytical designs such as prospective or retrospective cohort studies, case-control studies, or clinical trials.

For example, what is the surgical site infection rate for open appendectomies compared to laparoscopic appendectomies? In this case, the null hypothesis would state that there are no differences in the infection rate between the two approaches. If the statistical analysis rejects this hypothesis, the alternative hypothesis—which posits that the infection rate differs between the groups—would be accepted.

How is the sample size calculated?

The calculation of sample size requires defining several key elements. We will refer to the calculation of sample size in analytical studies given that it is frequently mentioned in scientific publications in the field of surgery^{1,3,4,6,7}:

1. Research question

The researcher must have a clear understanding of the research question in order to define the study design in previous stages. For example, for the question “What is the recurrence rate in female patients undergoing laparoscopic femoral hernia repair compared to conventional surgery?” This is an analytical question that can be answered using either a cohort study or a clinical trial. The type of study chosen will depend on the feasibility, available resources, and the researcher’s preferences.

2. Primary variable

The primary variable is the variable of interest that allows answering the research question. The type of variable (categorical or continuous) determines the statistical method used to calculate the sample size. In the previous example, the variable of interest is the recurrence rate, which is a categorical variable (yes/no).

3. Effect size

The effect size represents the smallest difference between groups that is considered clinically significant. Determining it requires reviewing the literature to understand how these treatments work. It also relies on the insight and experience of researchers in the field who are familiar with these effects.

Continuing with the example, the researcher knows that the reported recurrence rate for femoral hernias in women is approximately 5–10% for open surgery and 1–3% for laparoscopic surgery. In this case, the researcher could determine that a 5% absolute reduction in the recurrence rate between open and laparoscopic repair constitutes a clinically significant difference in this study.

4. Significance level (α)

The alpha (α) level is defined as the probability of committing a type I error, that is, of rejecting the null hypothesis when it is in fact true. In practice, this can lead to the conclusion of difference or association between the variables under study when, in fact, no such difference exists.

In clinical research, the alpha level is conventionally set at 0.05 (5%). This means that the researcher acknowledges a 5% probability of finding a statistically significant difference that is actually due to chance.

5. Statistical power

Statistical power represents the probability of detecting a real difference between groups when such

a difference actually exists. It is directly related to a type II (β) error, which occurs when the study fails to detect a difference that actually exists and therefore erroneously accepts the null hypothesis (see above). Mathematically, statistical power is expressed as $1 - \beta$. In clinical research, statistical power is typically set at 80%–90%, which implies accepting a 20%–10% probability of committing a type II error.

6. One-tailed test or two-tailed test?

When calculating the sample size, it is also necessary to determine whether a one-tailed or two-tailed hypothesis test should be used. A two-tailed test assesses the possibility of a difference in either direction. For this reason, it splits the alpha level across both ends of the probability distribution. This approach is most commonly used in clinical research. A one-tailed test specifies a single direction of effect or difference (for example, that one technique is superior to another). Although smaller sample sizes are required, its implementation must be well justified.

7. Adjusting for losses

In the context of prospective studies, the potential for losses during follow-up must be considered. To address this, the initial sample size is typically increased to compensate for these potential losses. Usually, an additional 10–20% of observation units are included.

Once these parameters have been defined, the sample size can be calculated using specific statistical formulas or specialized software, such as Epi Info, SPSS, R, Stata, or other online tools^{5,7}.

Is it always necessary to calculate the sample size?

Ideally, the sample size should always be calculated before initiating any study, either for descriptive or analytical questions³.

Calculating the sample size is particularly important in analytical studies, especially in clinical trials. These experimental and prospective studies assign participants to one or more interventions in order to evaluate their efficacy and safety. Clinical trials are often complex and expensive, as they require infrastructure, long-term follow-up, trained staff, supplies, and, frequently, external funding. Errors in calculating the sample size can result in the inclusion of an insufficient number of participants, thus failing to detect a genuine effect. In turn, this can result in a waste of time and resources. Conversely, recruiting more participants than strictly necessary unnecessarily increases costs and patients’ exposure to experimental interventions.

In retrospective studies, even though the data

already exist and the costs are lower, it is also advisable to estimate the required sample size. This enables the determination of whether there are enough cases to detect relevant associations or differences with sufficient statistical power. Otherwise, non-significant results may simply reflect a lack of sample size rather than the absence of a real effect.

- Avoid claims of equivalence if the study was not specifically designed to demonstrate non-inferiority or equivalence

This methodological transparency strengthens scientific credibility, even in the presence of unavoidable limitations.

How to proceed when the study does not reach the planned sample size?

When it is not possible to achieve the calculated sample size, various strategies may be considered³:

- Redefine the study as exploratory or hypothesis-generating
- Explicitly state the limitation in statistical power
- Conduct multicenter collaborations to increase the sample size

Conclusion

The sample size represents the quantitative translation of the clinical hypothesis and is a fundamental pillar of the methodological design. In surgical research, rigorous calculation is essential for generating valid, interpretable, and ethically sound evidence. The quality of our publications depends not only on our technical skill in the operating room, but also on the methodological rigor with which we design, analyze, and communicate our research.

Bibliografía / References

1. Maggard MA, O'Connell JB, Liu JH, Etzioni DA, Ko CY. Sample size calculations in surgery: Are they done correctly? *Surgery*. 2003;134(2):275-9. doi:10.1067/msy.2003.235.
2. Speich B. Adequate reporting of the sample size calculation in surgical randomized controlled trials. *Surgery*. 2020;167(5):812-4. doi:10.1016/j.surg.2019.10.011.
3. Giunta DH. *Bioestadística Handbook Fundamentos Básicos*. Buenos Aires: Merck S.A; 2018.
4. García-García JA, Reding-Bernal A, López-Alvarenga JC. Cálculo del tamaño de la muestra en investigación en educación médica. *Investigación en Educación Médica*. 2013;2(8):217-24. doi:10.1016/S2007-5057(13)72715-7.
5. Althubaiti A. Sample size determination: A practical guide for health researchers. *J Gen Fam Med*. 2023;24(2):72-8. doi:10.1002/jgf2.600.
6. Sheps S. Sample Size and Power. *J Invest Surg*. 1993;6:469-75.
7. Staffa SJ, Zurakowski D. Statistical power and sample size calculations: A primer for pediatric surgeons. *J Pediatr Surg*. 2020;55(7):1173-9. doi:10.1016/j.jpedsurg.2019.05.007