

¿Cómo determinamos el tamaño muestral necesario para testear nuestra hipótesis de investigación?

Emiliano Rossi 

Médico cardiólogo. Departamento de Investigación, Hospital Italiano de Buenos Aires.
Ciudad Autónoma de Buenos Aires, Argentina.

Acta Gastroenterol Latinoam 2025;55(3):180-183

Recibido: 31/08/2025 / Aceptado: 22/09/2025 / Publicado online: 30/09/2025 / <https://doi.org/10.52787/agl.v55i3.540>

En investigación clínica la validez de los resultados no solo depende de que hayamos definido claramente cuál es nuestra pregunta de investigación y elegido un diseño adecuado, sino también de que contemos con un tamaño muestral óptimo.

¿Qué es el tamaño muestral?

Es el número de unidades de observación (ej.: pacientes) que debemos incluir en el estudio para responder a la pregunta de investigación.

¿Por qué debe ser óptimo?

Un tamaño muestral insuficiente nos expone al riesgo de no detectar una diferencia de efecto presente (error de tipo II), mientras que uno excesivamente grande nos lleva a aumentar los costos del estudio (recursos materiales, tiempo) e, incluso, a detectar diferencias estadísticamente significativas, aunque clínicamente no relevantes.¹

¿En qué etapas del estudio debemos considerarlo?

El cálculo de tamaño muestral es un paso que no se debe eludir. Tenemos que considerarlo tempranamente en la etapa de planificación y describirlo en el protocolo del estudio.

Una vez finalizado el estudio, en la etapa de escritura del artículo científico, debemos mencionarlo en la sección Métodos. Las guías de reporte de ensayos clínicos (CONSORT) y de estudios observacionales (STROBE) establecen recomendaciones sobre cómo debemos presentar el cálculo de tamaño muestral realizado.

¿Cuáles son sus fundamentos estadísticos?

La inferencia estadística busca sacar conclusiones sobre poblaciones a partir del estudio de muestras representativas. Al comparar muestras buscamos determinar si provienen de la misma población de referencia o no.

Correspondencia: Emiliano Rossi
Correo electrónico: emiliano.rossi@hospitalitaliano.org.ar

Para ello debemos definir:²

- *Hipótesis nula* (H_0): establece que no hay diferencia entre los grupos comparados (asumiendo que las muestras comparadas pertenecen a la misma población de referencia).

- *Hipótesis alternativa* (H_1): establece que hay diferencia entre los grupos.

- *Error de tipo I* (α): es la probabilidad de rechazar la hipótesis nula cuando esta es verdadera. Habitualmente se establece en 0,05. Es decir, se considera aceptable que 5 de cada 100 pruebas presenten este error.

- *Error de tipo II* (β): es la probabilidad de no rechazar la hipótesis nula cuando esta es falsa.

¿Qué información necesitamos para su cálculo?

1. *Nivel de significación* (α): es la máxima probabilidad aceptada de cometer un error de tipo I. Si el valor de $p > 0,05$, no se rechaza la H_0 .

2. *Poder estadístico*: es $1 - \beta$, por ende, es la probabilidad de rechazar la hipótesis nula cuando esta es falsa. Esto equivale a decir que es la probabilidad de detectar una diferencia si realmente existe. Convencionalmente se establece en 80 o 90%.

3. *Tamaño del efecto clínicamente relevante*: es la magnitud mínima de la diferencia de efecto entre los grupos que deseamos detectar.

4. *Variabilidad del resultado*: expresada como el desvío estándar (DS) (en caso de variables continuas) o la proporción esperada (en caso variables categóricas).³

¿De dónde obtenemos la información necesaria?

Obtener la información correcta para el cálculo de tamaño muestral es imprescindible para el éxito del estudio. Subestimar la importancia de este paso pone en riesgo la factibilidad de responder a la pregunta de investigación. El desafío que se nos presenta en este momento es que parte de la información necesaria es a la vez la que queremos averiguar con la realización del protocolo.

Las fuentes de información que tenemos a disposición son, en primer lugar, la evidencia científica ya publicada (ensayos clínicos, metaanálisis, estudios observacionales, registros). Al consultarla es importante buscar que se haya incluido a una población de características semejantes a la nuestra. En segundo lugar, la opinión de expertos en el área temática a investigar. Y, finalmente, estudios piloto (considerando sus limitaciones).

¿Cuál es el rol de los estudios piloto?

Son estudios en pequeña escala que permiten la estimación de parámetros desconocidos, como el desvío estándar o a la proporción esperada y el tamaño del efecto. Sin embargo, su principal limitación radica en que, al tener un tamaño pequeño, estas estimaciones suelen ser imprecisas y tienen intervalos de confianza muy amplios.² Esto puede llevar a sobreestimar o subestimar el tamaño muestral requerido. Por lo tanto, la recomendación es usarlos principalmente como estudios de factibilidad, más que como fuente única de información para el cálculo de tamaño muestral.

¿Cómo realizamos el cálculo?

Esto va a depender del tipo de estudio (igualdad, superioridad, no inferioridad o equivalencia), la técnica de muestreo (ej. aleatorio simple), el número de grupos a comparar (ej. dos), la asignación de individuos a los grupos (ej. 1:1) y de cuál sea nuestra medida de efecto (media, proporción, OR, HR, tasa).

Dado que este es un artículo con un enfoque introductorio, no vamos a presentar fórmulas matemáticas. Sólo recalcaremos que el tamaño muestral será mayor a medida que reduzcamos el nivel de significación (ej. 0,01 en vez de 0,05), aumentemos el poder, disminuyamos el tamaño del efecto clínicamente relevante que queremos detectar y/o aumentemos la variabilidad del resultado.

¿Qué herramientas tenemos disponibles?

Las herramientas que tenemos para realizar el cálculo de tamaño muestral son diversas. Tenemos a disposición software libre como *G*Power* (<https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower>), paquetes de R como *pwr* o *TrialSize*; software comercial como *PASS* o *Stata*, y calculadores online de sitios web como *OpenEpi* (<https://www.openepi.com/>) o *ClinCalc* (<https://clincalc.com/stats/samplesize.aspx>).

Vamos a ejemplificar el cálculo del tamaño muestral planteando dos situaciones frecuentes en el contexto clínico. La primera es un estudio en el que el testeo de hipótesis plantea comparar dos medias y, la segunda, dos proporciones. Dada su inmediata disponibilidad y uso intuitivo utilizaremos el calculador online de *ClinCalc*.

Estudio 1:

Evaluar el efecto a 12 meses de la vitamina E en

la reducción de la alanina aminotransferasa comparada con placebo en pacientes con esteatohepatitis no alcohólica.

Ingresaremos en el sitio web de ClinCalc, seleccionaremos la cantidad de grupos a comparar (dos grupos independientes) y el tipo de punto final (continuo). Posteriormente, ingresaremos la media esperada del grupo 1 y 2 (recordemos que su diferencia permite establecer el tamaño del efecto) y el desvío estándar anticipado. Finalmente, estableceremos el valor de alfa y el poder deseado en valores convencionales.

Estableciendo un nivel de significación (α) = 0,05, un poder = 90%, un efecto clínicamente relevante = 20 U/L (estimamos que la media del grupo control es 120 U/L y la del tratado con Vit. E 100 U/L) y considerando un DS = 15 U/L, necesitaríamos incluir 12 pacientes por grupo.

Estudio 2:

Evaluar el éxito en la erradicación de *H. pylori* comparando un esquema de terapia triple estándar vs. cuadriterapia con bismuto.

Al igual que en el ejemplo anterior seleccionamos dos grupos independientes, aunque esta vez el punto final es dicotómico. Completamos la proporción anticipada (incidencia) del punto final en cada grupo y, finalmente, estableceremos el valor de alfa y el poder.

Si establecemos un nivel de significación (α) = 0,05, un poder = 90%, una incidencia estimada de erradicación del 80% con el tratamiento estándar vs. 95% con cuadriterapia, necesitaríamos incluir 100 pacientes por grupo.

En este punto es importante recordar que al número calculado como necesario debemos sumarle el porcentaje de pérdida de seguimiento que podríamos tener en nuestro estudio (ej.: 10%).

¿El poder *post hoc* es una alternativa?

El poder *post hoc* u observado es aquel que se determina una vez finalizado el estudio. Es decir, que como investigadores ya hemos realizado el análisis y conocemos los resultados.

Debemos tener en cuenta que el poder es una función monótona del *p* valor y, por tanto, no aporta información adicional. Sólo confirma lo que el *p* valor ya nos indica.³ Valores de *p* no significativos siempre se corresponderán con valores de poder observado bajos.⁴ En síntesis, el poder es una herramienta de planificación, no de análisis retrospectivo.

Distintos autores y guías editoriales recomiendan no

utilizar el poder *post hoc* y, en su lugar, presentar los intervalos de confianza de nuestra medida de efecto. El intervalo de confianza nos permitirá reflejar la precisión de nuestra estimación.⁴⁻⁵

¿Qué es tamaño muestral basado en la precisión?

Como hemos visto, el enfoque de cálculo de tamaño muestral basado en el poder busca detectar una diferencia de efecto entre grupos, dados un nivel de significación (α) y poder preestablecidos ($1 - \beta$). En cambio, el enfoque basado en la precisión se centra en la exactitud con que se estima el parámetro de interés (media, proporción, etc.). Para ello debemos fijar primero un margen de error máximo aceptable (hemiamplitud del intervalo de confianza) y, luego, calcular el número necesario de individuos a incluir en el estudio para que nuestra estimación no lo exceda con un nivel de confianza de $1 - \alpha$.²

Esta aproximación permite que la estimación sea lo suficientemente precisa, aunque no se basa en detectar diferencias (no compara muestras). Es útil en estudios epidemiológicos en donde hay una única muestra y se pretende estimar un parámetro poblacional (ej. prevalencia de enfermedad).²

Estudio 3:

Realizar un estudio de corte transversal en la población general para estimar la prevalencia de infección por *H. pylori* en la Argentina, con un intervalo de confianza del 95% y un margen de error absoluto de ± 3 puntos porcentuales. Para establecer la prevalencia esperada utilizaremos como referencia a un estudio que reportó que la infección por *H. pylori* afecta al 36% de la población general de los Estados Unidos.

Para el cálculo de tamaño muestral por precisión, ingresaremos en el sitio web de OpenEpi, seleccionaremos «Tamaño de la muestra», «Proporción», «Introducir datos» y allí completaremos la frecuencia anticipada = 36% y los límites de confianza (margen de error) = 3%. No modificaremos los otros datos establecidos por defecto. Esto nos indicará que necesitamos incluir en forma aleatoria a 983 individuos para alcanzar la precisión especificada. Finalmente, debemos agregar el ajuste por pérdidas o no respuestas anticipadas (ej. 20%).

Conclusiones

El cálculo del tamaño muestral es un paso decisivo en la planificación de un estudio de investigación. Requiere que identifiquemos claramente cuál es nuestra

hipótesis y tengamos información sobre la magnitud del efecto que consideramos clínicamente relevante y la variabilidad esperada del resultado.

Un cálculo del tamaño muestral adecuado garantiza que el estudio sea válido y eficiente, permite asegurar el poder suficiente y evita incluir a más pacientes de los necesarios.

Propiedad intelectual. El autor declara que los datos presentes en el manuscrito son originales y se realizaron en su institución perteneciente.

Financiamiento. El autor declara que no hubo fuentes de financiación externas.

Conflicto de interés. El autor declara no tener conflictos de interés en relación con este artículo.

Aviso de derechos de autor



© 2025 Acta Gastroenterológica Latinoamericana. Este es un artículo de acceso abierto publicado bajo los términos de la Licencia Creative Commons Attribution (CC BY-NC-SA 4.0), la cual permite el uso, la distribución y la reproducción de forma no comercial, siempre que se cite al autor y la fuente original.

Cite este artículo como: Rossi E. ¿Cómo determinamos el tamaño muestral necesario para testear nuestra hipótesis de investigación?. *Acta Gastroenterol Latinoam*. 2025;55(3):180-183. <https://doi.org/10.52787/agl.v55i3.540>

Referencias

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.