

Selección del tamaño de muestra en estudios clínicos

Sample size for a clinic trial

Javier Hernández B. MD,¹

RESUMEN

En el diseño de una investigación contamos con varias herramientas: la utilización de programas computacionales, ecuaciones de estadística y fórmulas matemáticas sencillas para poder determinar el tamaño de muestra adecuado, que permita inferir el valor de un parámetro en la población. Idealmente, deberíamos evaluar a toda la población elegible pero por razones de costo y tiempo los estudios deben hacerse a partir de una muestra de esa población.

La selección de la muestra tiene implicaciones éticas e influye en el poder del estudio para detectar diferencias cuando las hay. Con algo de práctica, se puede llegar a calcular de manera adecuada el tamaño de muestra necesario. Esta revisión aportará datos útiles a tener en cuenta al momento de calcular la muestra.

PALABRAS CLAVE

Tamaño de muestra, error alfa, poder del estudio.

ABSTRACT

In the design of an investigation we possess several tools like computer programs, statistical equations and simple mathematical formulae to be able to determine the adequate sample size, which in turn will allow us to infer the value of a parameter in the population. Ideally, we should evaluate the whole eligible population but for reasons of cost and time the studies must be done from a sample of this population.

The selection of the sample has ethical implications and influences the power of the study to detect differences when they exist. With some practice, it is possible to come to calculate in an adequate way the necessary sample size. This review will contribute useful information to bear in mind when calculating the sample size.

KEY WORDS

Sample size, alpha error, power of the study.

Cuando planeamos una investigación, por dificultades de costo y tiempo, es imposible abarcar toda la población elegible, y debemos limitarnos a una muestra de ella. Esta muestra debe ser lo suficientemente representativa, para que los datos obtenidos permitan inferir el valor del parámetro en la población. Los resultados obtenidos en la muestra reciben el nombre de estimadores, y representan el valor poblacional (parámetro) siempre y cuando la selección de la muestra haya sido adecuada. La realidad, entendida como el parámetro poblacional, no la conocemos y es precisamente lo que deseamos medir.

Adicionalmente, cuando leemos críticamente un estudio, necesitamos saber si la muestra fue suficiente. Esto es de particular importancia para los estudios con resultados negativos, en donde el no haber encontrado diferencias puede deberse a un tamaño de muestra insuficiente, y si en estos casos aceptamos el resultado negativo, estaríamos frente a un error tipo II.

No es ético realizar estudios con un tamaño de muestra insuficiente para obtener conclusiones, ya sean éstas positivas o negativas. Igualmente, no es ético realizar estudios con un tamaño de muestra superior al requerido, porque innecesariamente se expone a un número adicional de sujetos (estudios experimentales) o se incrementan los costos.

Desde la fase de diseño, debemos preocuparnos por hacer un cálculo del tamaño de muestra correcto, y éste es un requisito no sólo para los estudios experimentales, sino también para los estudios observacionales incluidos los estudios de prevalencia que con alguna frecuencia vemos publicados en la literatura local. De esos estudios observacionales, las series de casos o reportes de caso por razones obvias no requerirán cálculo de tamaño de muestra.

La declaración CONSORT* y cada vez más las revistas biomédicas, exigen al momento de enviar

¹ Médico Internista, Residente de Gastroenterología, Candidato a Maestría en Epidemiología Clínica.

Pontificia Universidad Javeriana-Bogotá D.C, Colombia
Fecha recibido: 26-04-06 / Fecha aceptado: 30-05-06

un artículo a publicación, que se especifique cómo fue calculado el tamaño de muestra.

***CONSORT:** Es una herramienta que mediante un enfoque basado en evidencia busca mejorar y homogenizar la forma como se reportan los experimentos clínicos controlados.

El razonamiento para la determinación del tamaño de muestra es sencillo, y existen varios programas para su cálculo, por ejemplo STATA, TAMAMU, Epi-Info, que permiten calcular el tamaño de muestra para cada uno de los diseños de investigación, siempre y cuando podamos alimentar el programa con los datos requeridos, por lo que si lo hacemos con alguna frecuencia, adquiriremos la habilidad para hacerlo correctamente, sin el apoyo de un bioestadístico, quien permanecerá jugando un importante papel al momento del análisis estadístico de los resultados.

Este artículo no busca profundizar en la matemática del cálculo del tamaño de muestra, que se describe en los libros de estadística, y cuya fórmula de cálculo concreto dependerá del parámetro que se va a estimar (una proporción, una media, un coeficiente de correlación, etc.) y del modelo probabilístico con el que se supone se distribuye ese parámetro, obteniéndose finalmente una fórmula que depende del error prefijado, de la mínima diferencia D entre los parámetros que se considera de importancia clínica, de la probabilidad de no detectar esa diferencia, o su complementaria (1 —la probabilidad de detectarla) o poder de la prueba, y de la variabilidad de los datos expresada en función de la desviación estándar.

Esta revisión aportará algunos datos útiles a tener en cuenta al momento de calcular la muestra.

DETERMINANTES DEL TAMAÑO DE MUESTRA

Dependiendo de si el diseño es descriptivo o analítico, los factores que a continuación se detallan serán determinantes para el cálculo del tamaño de muestra:

1. El diseño del estudio: Para cada diseño existe una fórmula particular de cálculo de tamaño de muestra, requiriendo cada uno datos diferentes para alimentar la fórmula. Es importante considerar no sólo el tipo de diseño sino la naturaleza (nominales, ordinales, de intervalo o de razón) de las variables a analizar, y si se planea realizar mediciones repetidas. Cuando el estudio pretenda estimar varios parámetros, se debe hacer un cálculo para

cada uno de esos estimativos y seleccionar como tamaño de muestra el mayor calculado.

2. La variabilidad: Se refiere a la dispersión de los datos que esperamos encontrar y se representa por la varianza y su raíz cuadrada (desviación estándar). En el caso de variables ordinales o nominales, no se necesita este valor en el cálculo del tamaño de muestra, pero para variables continuas (intervalo o razón) es necesario conocer ese dato. La varianza puede obtenerse de estudios previos o en su ausencia, requerir la realización de una prueba piloto para estimarlo.
3. Frecuencia basal del fenómeno: Si la frecuencia reportada del evento que deseamos estudiar es baja, el tamaño de muestra será mucho mayor. La frecuencia basal (prevalencia o incidencia) generalmente se encuentra en la literatura y puede obtenerse a partir de estudios más pequeños.
4. Hipótesis: Las hipótesis a una cola, requieren tamaño de muestra menor que las hipótesis a dos colas, sin embargo, los trabajos en general, deben manejar hipótesis a 2 colas, a menos que exista evidencia previa suficiente que sugiera la posible direccionalidad de los hallazgos.
5. Nivel de error tipo I o nivel de significación: El error tipo I ocurre cuando se acepta que una intervención es mejor que la otra cuando en realidad no lo es. Se mide mediante el valor de p , que se define como la probabilidad de obtener un resultado tanto o más extremo que el encontrado, simplemente por azar. El nivel de error tipo I (error alfa) normalmente incluido en la fórmula de tamaño de muestra debe ser inferior a $0,05$. Si el estudio incluye comparaciones múltiples, el nivel del error tipo I considerado en el cálculo de la muestra debe ser menor (usualmente $0,01$), esto debido a que pruebas estadísticas repetidas, pueden llevar a un resultado positivo solamente por azar. A menor nivel de error tipo I, mayor será el tamaño de muestra requerido.
6. Es importante recordar que el valor de p nos habla de la diferencia estadísticamente significativa, mas no dice si la diferencia hallada es clínicamente significativa o no.

7. Nivel de error tipo II (error β) y poder del estudio: El poder del estudio para detectar una diferencia cuando realmente existe, está estrechamente relacionado con el error tipo II, siendo aceptado normalmente un poder de 0,8 a 0,9. El poder del estudio es igual a $1-\beta$. A mayor poder del estudio, mayor tamaño de muestra.
8. Magnitud de la diferencia que se considerará clínicamente significativa: De antemano, el investigador deberá determinar la diferencia que espera encontrar como clínicamente significativa y que le permitirá probar la hipótesis del estudio. Entre más grande sea esa diferencia, menor será el tamaño de muestra.
9. Ajustes por pérdidas en el seguimiento: Se deben prever las pérdidas de seguimiento durante el estudio para garantizar que el tamaño al final del estudio no sea menor al inicialmente requerido. Si n representa el tamaño de muestra estimado sin ajuste, $n(a)$ representa el tamaño de muestra ajustado por pérdidas en el seguimiento y X es la proporción de pérdidas a través del tiempo del estudio, tendremos:

$$n(a) = \frac{n}{(1-X)^2}$$

MÉTODOS SECUENCIALES

En los análisis secuenciales se efectúa el cálculo del tamaño de muestra necesario a medida que se van recogiendo los casos y se compara el resultado obtenido hasta ese momento con dos umbrales, superados los cuales, se detiene el experimento. En este tipo de diseño, para mantener la probabilidad global de error tipo I en un valor inferior al prefijado (0,05), es necesario utilizar un nivel de significación menor para cada paso. Se debe tener cuidado al analizar este tipo de diseños, pues un resultado positivo puede ser un resultado sesgado, producto de los análisis repetidos.

Algunas fórmulas para el cálculo de tamaño de muestra

- Estudios que buscan estimar un promedio:

$$n = f(1-\alpha)S^2/\Delta^2$$

Donde:

$f(1-\alpha)$ es un valor constante para cada nivel de confianza. El nivel de confianza lo determina el investigador.

Tabla 1. Valores de $f(1-\alpha)$

Nivel de confianza	$f(1-\alpha)$
90%	2,706
95%	3,842
99%	6,635

Δ : Amplitud del intervalo de confianza.

S : Desviación estándar de las observaciones. Se obtiene de la literatura o estudio piloto.

- Estudios que buscan estimar una proporción:

Donde:

$$n = f(1-\alpha)P(1-P)/\Delta^2$$

P : Proporción que se espera encontrar en la población de estudio. Se obtiene de reportes previos en la literatura o en la experiencia local.

Δ : Amplitud del intervalo de confianza

$f(1-\alpha)$ es un valor constante (ver tabla 1)

- Estudios para evaluación de diferencias de promedios:

$$n = 2f(\alpha, \beta)S^2/\Delta^2$$

Donde:

Δ : La magnitud de la diferencia que se considera clínicamente relevante.

S : Desviación estándar de las observaciones. Se obtiene de estudios previos o estudio piloto.

$f(\alpha, \beta)$: Valor determinado en función del error α y β seleccionados por el investigador y de si la evaluación de la hipótesis es a una o dos colas (ver tabla 2).

- Estudios para evaluación de diferencias de proporciones:

Se puede usar la siguiente fórmula:

$$n = \frac{P_1(1-P_1) + P_2(1-P_2)}{(P_1 - P_2)^2} f(\alpha, \beta)$$

Donde:

P_1 : Proporción en el grupo control

P_2 : Proporción en el grupo intervenido

$f(\alpha, \beta)$: Varía en función de α, β (ver tabla 2)

Tabla 2. Valores para $f(\alpha, \beta)$

Poder	Nivel de error tipo I			
	Una cola		Dos colas	
	0,05	0,01	0,05	0,01
0,8	6,18	10,04	7,85	11,68
0,9	8,56	13,02	10,51	14,88

El problema aquí puede radicar en que las proporciones P_1 o P_2 , sean desconocidas, en cuyo caso debemos basarnos en estudios observacionales previos o un estudio piloto para determinar P_1 y considerar la diferencia (D) de proporciones que clínicamente consideraremos significativa, entonces tendremos que $P_2 = P_1 + D$.

Por último, debemos recordar que un gran tamaño de muestra no es garantía de un buen resultado. Ahora que están de moda los mega estudios, debemos tener bien claro, que los grandes estudios sólo son necesarios para demostrar pequeños efectos o diferencias pequeñas en los efectos y que aunque los grandes tamaños de muestra de esos trabajos nos impresionen, el tamaño no lo es todo, ya que en la calidad de los resultados también influye la calidad y características del diseño y ésta siempre es mucho más difícil de garantizar en los grandes estudios. Si la magnitud de la diferencia que esperamos encontrar es grande, es probable que necesitemos tamaños de muestra muy pequeños, lo cual de ninguna manera invalida los hallazgos.

REFERENCIAS

1. Beck RW. Sample Size for a Clinical Trial: Why Do Some Trials Need Only 100 Patients? *Ophthalmology* 113; (5): 721-2.
2. Molinero L. Cálculo del tamaño de muestra. Métodos secuenciales. En www.seh-lilha.org/stat1.htm. Noviembre 2002.
3. Dennis R, Pérez A. "Cálculo del tamaño de muestra". En *Epidemiología clínica. Investigación clínica aplicada*, Bogotá D.C. Editorial médica Panamericana, 2004: 141-62.
4. Pérez A, Rodríguez N, Gil JFA, Ramírez G. Tamaño de la muestra. A computer program to estimate the required sample size and power in clinical research. *Journal of clinical epidemiology* 1999; 52 (suppl 1): 38s.
5. Pagano M, Kimberlee G. Cálculo del tamaño de muestreo". En *Fundamentos de Bioestadística*, México D.F, International Thomson Editores S.A, 2001: 330-41
6. Paesmans M, Bleiberg H. Are we cautious enough when we interpret results of randomized but underpowered comparisons? *J Clin Oncol* 2006; 24(12): 1964-5.
7. Sinha S, Mkhherjee B. A score test for determining sample size in matched case-control studies with categorical exposure. *Biom J* 2006; 48(1): 35-53.
8. Spilker B. "Sample size and number of parts of a clinical trial" En *Guide to clinical trials*. Philadelphia, Pennsylvania. Lippincott Raven Publishers 1996: 65-8.
9. Daniel W. "Estimación". En *Bioestadística. Base para el análisis de las ciencias de la salud*. México D.F. Editorial Limusa SA 2002; 180-5.
10. CONSORT Checklist of items to include when reporting a randomized trial. En <http://www.consortstatement.org/Downloads/checklist.pdf>. Fecha de acceso mayo de 2006.
11. Epi Info™ Versión 3.3.2. En <http://www.cdc.gov/epiinfo/>. Fecha de acceso mayo de 2006.
12. Dupont WD and Plummer WD: PS power and sample size program available for free on the Internet. *Controlled Clin Trials* 1997; 18: 274.
13. STATA. Statistical software for professionals. <http://www.stata.com/>. Acceso mayo 2006.