

¿Cuántos pacientes selecciono para mi estudio?

J. Muñoz-García, M.I. Santiago-Pérez

Resumen

Introducción. Después de haber trabajado uno de los autores durante años en una unidad de apoyo a la investigación en un hospital grande, llegó al convencimiento de que esta pregunta, y sus variantes, es, probablemente, una de las causas más frecuentes de consulta entre los investigadores clínicos y, en ocasiones, motivo de ansiedad.

Desarrollo. Vamos a dedicar unas breves reflexiones para ayudar a los investigadores clínicos a abordar el cálculo del tamaño muestral para un estudio. La orientación de las mismas intentará reproducir la línea argumental que le funciona al componente no matemático de los autores, e ir a través de las soluciones que le ha dado a los mismos problemas que afrontan otros compañeros para intentar que esto tenga sentido. Así, se huirá de fórmulas matemáticas en detalle, pero se abordarán los diferentes elementos que desempeñan un papel a la hora de calcular un tamaño muestral. La razón de esto es que existen muchos contextos diferentes en los que plantearse un tamaño muestral, en función del tipo de estudio, el interés u objetivo de la investigación, etc. En cada uno de estos contextos diferentes es de aplicación una fórmula específica, lo que hace muy difícil recordar las fórmulas. Por suerte, esto no es necesario, ya que existen múltiples programas disponibles en la red para ayudarnos en el proceso de estimar el tamaño muestral necesario.

Conclusión. En el presente artículo se aborda la cuestión del cálculo del tamaño muestral a través de tres cuestiones: 1) intentar entender por qué tiene sentido y es necesario estimar un tamaño muestral; 2) en qué pensar cuando deseamos hacerlo; y 3) presentación de alguna herramienta disponible para ayudarnos en el proceso.

Palabras clave.

Epidemiología. Estadística. Tamaño muestral

Justificación de la necesidad de considerar el tamaño muestral

Gran parte de lo que podemos contar aquí y en el siguiente apartado es coincidente con el conocimiento intuitivo.

La principal justificación es que merezca la pena lo que hagamos y, al mismo tiempo, no hacer más de lo necesario. ¿Qué queremos decir con esto? En resumen, y sin entrar en consideraciones éticas que, sin duda, matizan esta afirmación, para cualquier estudio, cuanto mayor es el tamaño muestral, mejor.

En igualdad de condiciones (calidad de la información recogida, etc.), ¿me fío más de un estudio grande o de uno pequeño? Al mismo tiempo, en el otro lado de la balanza, ¿por qué repetir 1.000 veces lo mismo si con 200 es suficiente para lo que queremos? Si esto es así, es evidente que hay un margen potencialmente grande para el ahorro de tiempo, recursos, trabajo, personal, etc.

En definitiva, los propios investigadores son los primeros interesados en pensar en el tamaño muestral necesario antes de embarcarse en un estudio.

¿Por qué la pregunta anterior –¿me fío más de un estudio pequeño o de uno grande?– en realidad la vemos como retórica? La explicación es que un estudio grande es más preciso o, en román paladino, es más difícil meter la pata o cometer un error. Para entender de qué tipo de error nos libra (o lo mitiga) aumentar el tamaño de la muestra, es bueno clasificarlos en errores sistemáticos y aleatorios. Un ejemplo tosco de los primeros puede ser, en el caso de un estudio que intente estimar la prevalencia de hipertensión arterial, utilizar esfigmomanómetros mal calibrados que sistemáticamente dan valores altos de presión arterial. Esto resultaría en una sobreestimación de la prevalencia de hipertensión arterial. No importa cuánto aumentemos el tamaño de la muestra: si no arreglamos los aparatos, no nos libramos del problema. De igual manera, pero en el otro sentido, sería utilizar observadores sordos. El otro gran grupo de errores son los aleatorios, los que tienen que ver con el azar, en definitiva, con la suerte. Sí, en un estudio, haciéndolo todo bien, podemos tener mala suerte y meter la pata. El tamaño muestral nos ayuda a planear que la probabilidad de que eso suceda sea pequeña. ¿Cómo se puede equivocar uno por azar en el curso de una investigación? Tal vez se entienda con un ejemplo sencillo. Supongamos una bolsa con un

millón de bolas. En realidad, la mitad de las bolas son rojas y la mitad negras, pero eso es algo que no sé y que quiero estimar.

Ésta es la situación típica de un estudio de prevalencia en el que queremos estimar qué pasa en la población a partir de una muestra extraída de la misma. En el problema de la bolsa es obvio que si saco dos bolas puedo acertar con lo que pasa en el conjunto de la bolsa (una roja y una negra, prevalencia de bolas rojas = 50%), pero también puedo fallar de manera catastrófica en muchos casos (dos rojas, prevalencia = 100%, o dos negras, prevalencia = 0%). El problema es que no sé cuál de las muestras es la que he extraído.

Si en lugar de dos bolas, extraigo 10, pueden ser aún todas rojas o todas negras, pero, intuitivamente, es más difícil que antes. Y cada vez lo es más si extraigo 100, 500 o 1.000.

En las diferentes muestras posibles de tamaño 1.000, habrá muchas que tengan 495, 497, 500, 503, 504, etc. bolas rojas. ¿Es posible elegir por azar una muestra extrema? (por ejemplo, 1.000 o 990 bolas rojas o negras). Sí, si hay 500.000 bolas rojas y el azar nos puede jugar una mala pasada y hacer que de todas las posibles muestras de tamaño 1.000 nos toque una muy extrema. Pero ¡malo será! Así es, cuanto mayor es el tamaño muestral, más difícil es que tengamos por azar una muestra extrema. De hecho, la distribución de todas las posibles muestras diferentes de 1.000 bolas sacadas de esa gran bolsa sigue una distribución aproximadamente normal: las muestras cercanas a la realidad (prevalencia de bolas rojas = 50%) son muy frecuentes, mientras que cuanto más se alejan de esta realidad menos frecuentes son. Al aumentar la muestra, lo que hacemos es aumentar la probabilidad de que el azar seleccione una muestra que se acerca a la realidad. El tamaño muestral nos ayuda, por tanto, a controlar los errores aleatorios, no los sistemáticos.

¿Cuál es la enseñanza de todo esto? Un concepto básico, pero que como clínicos difícilmente se tiene en cuenta, es que de una población de tamaño N hay muchas posibles muestras diferentes de tamaño n . Cuando hacemos nuestro estudio, sólo tenemos una muestra (la que seleccionamos al azar), ¡y no sabemos cuál es! Mejor, así, que sea grande, por lo ya dicho. ¿De verdad hay muchas muestras posibles? Si pensamos en la lotería primitiva, hay casi 14.000.000 muestras distintas de tamaño seis de la población de 49 números (¡por eso es tan difícil que nos toque!). Si en lugar de una población de tamaño 49 y una muestra de tamaño 6, tenemos la situación de una población de tamaño 50.000 (cualquier ciudad pequeña) y muestras posibles de tamaño 1.000, en lugar de millones de posibles muestras, estaríamos hablando de billones, trillones o más. (Quien tenga curiosidad, puede calcularlo como las posibles combinaciones de 50.000 tomadas de 1.000 en 1.000). En resumen, muchas posibles muestras, sólo disponemos de una y no sabemos cuál es.

¿En qué debemos pensar para calcular el tamaño de la muestra?

Para responder a esta pregunta, debe tenerse presente algo que ya enunciamos con anterioridad: ¿para qué queremos seleccionar participantes? Responder a esta pregunta nos permitirá acotar qué aspectos hay que tener presentes para estimar qué tamaño muestral necesitamos.

Si, por ejemplo, estamos tratando de diseñar una encuesta transversal para estimar la prevalencia de diabetes en la población, el tamaño de muestra necesario deberá garantizar que esa prevalencia se estime con una determinada precisión, la que nosotros queramos para el estudio. Fijar la precisión es equivalente a controlar el error aleatorio que cometeremos al calcular la prevalencia poblacional de diabetes con base en el resultado proporcionado por una muestra. A estas alturas, es bueno incorporar el concepto de 'intervalo de confianza', sobre todo de cómo interpretarlo. Aunque muchos lectores lo saben, vamos a recordar lo mencionado con anterioridad de error aleatorio y el ejemplo de la bolsa de bolas (si se prefiere, cámbiese 'bolas negras' por 'diabéticos', 'diagnóstico de arteriopatía periférica' o lo que sea). Hemos sacado 1.000 bolas y 490 son negras, ¿cuál es la prevalencia real (dentro de la bolsa) de bolas negras? Nuestra mejor estimación a partir de la única muestra estudiada (¡sólo tenemos una y hay billones!) es 49%. Pero ¿está seguro? Aquí entra el intervalo de confianza (precisión). Supongamos que calculamos el intervalo de confianza del 95% en este ejemplo y da $49 \pm 2\%$ (no hemos hecho los cálculos, ¡es sólo un ejemplo!). ¿Cómo se interpreta esto? La interpretación chusca, pero muy útil a efectos de entender los intervalos de confianza, es la que sigue: No, no estoy seguro de que en la bolsa de bolas (población) la prevalencia de bolas negras sea exactamente de 49%, pero malo será que sea más del 51% o menos del 47%. ¿Cómo de malo? Estoy el 95% seguro de que es así. En términos un poco más técnicos (no mucho), el intervalo de confianza del 95% (en este caso 47-51%) se obtiene por un procedimiento que garantiza que el 95% de los intervalos así contruidos (con las infinitas muestras posibles) contiene el verdadero valor del parámetro. ¿Podemos equivocarnos? Por supuesto: tenemos una probabilidad del 5% de equivocarnos, es decir, de que ese intervalo sea uno de los que no contiene la prevalencia real. ¿Es mucho? Se podría fijar un valor más pequeño, por ejemplo, el 1%, con lo cual, tendríamos más 'confianza' en nuestro intervalo (el 99%), pero eso haría que fuese más amplio y, por tanto, menos

informativo (por ejemplo, en lugar de 47-51%, 46-52% o 45-53%). Sin embargo, aún podría aumentarse el nivel de confianza sin perder precisión y seguro que se imagina cómo. ¡Bien, acertó, aumentando el tamaño muestral! Si uno quiere reducir los errores aleatorios, tiene que aumentar el tamaño muestral. Es una decisión que depende del investigador.

En otro contexto están los estudios analíticos y los de intervención. Pensemos, por ejemplo, en un ensayo clínico de un nuevo fármaco comparado con un placebo; aquí el objetivo podría ser detectar una reducción del riesgo de mortalidad después de un año de seguimiento. Entramos en el mundo de los tests de hipótesis y, de manera rápida, son dos los tipos de errores aleatorios que podemos sufrir al comparar dos grupos:

- Tipo I (o α): no existen diferencias entre los grupos pero en mi estudio aparecen.
- Tipo II (o β): existen diferencias de determinada magnitud, pero mi estudio no es capaz de detectarlas ($1 - \text{error } \beta$ es la famosa potencia del estudio).

Ahora, el cálculo del tamaño de muestra se realiza para garantizar una determinada potencia.

Ya tenemos claro el primer elemento que interviene en el cálculo del tamaño de muestra: precisión o potencia, según el objetivo del estudio (¡queremos tener errores aleatorios pequeños!).

En cada uno de estos dos contextos, estimación de un parámetro o realización de pruebas de hipótesis, las fórmulas para determinar tamaños de muestra dependen, a su vez, del parámetro concreto que se trata de estimar o de la prueba estadística que se aplicará, respectivamente. En cualquier caso, es importante matizar que el resultado de todas estas ‘abigarradas’ fórmulas debe constituir más una orientación que un valor definitivo e incuestionable. Un elemento que inevitablemente forma parte del proceso de decisión es el de los recursos disponibles: ¿de qué me sirve calcular que necesito una muestra de 550 sujetos si sólo puedo pagar, por ejemplo, determinaciones analíticas de 400? En una situación como ésta, habrá que valorar lo ‘bueno’ que sería el estudio con esos 400, y si la precisión que nos garantiza esa muestra no es suficiente, quizá lo mejor es no hacer un estudio que, inevitablemente, arrojará resultados poco fiables. Dicho de otra manera, frente a posicionamientos maniqueos muy frecuentes, el proceso de estimar el tamaño de la muestra para disponer de una determinada precisión (como en los ejemplos de estudios para estimar la prevalencia de una determinada enfermedad, o, para el caso, la media de una variable cuantitativa) o de una determinada capacidad para detectar diferencias entre grupos (como en los casos de estudios de factores de riesgo o en los ensayos clínicos), es igual de razonable y lícito como el proceso inverso: en función del tamaño muestral asequible (por razones logísticas, presupuestarias, de viabilidad del estudio u otras), ¿qué precisión o qué capacidad para detectar diferencias nos ofrece? Y ¿merece la pena, entonces, realizar el estudio?

Respecto al tamaño de los errores aleatorios, cuanto más pequeños, mejor. Pero como eso afecta de manera importante al tamaño muestral, por convenio se utilizan frecuentemente errores α de 0,05 (la $p < 0,05$ que aparece después en el análisis) y errores β de 0,2 (equivale este último a una potencia para detectar diferencias del 80%). Estos valores, en función de los diseños y objetivos de los estudios, pueden ser otros, pero habitualmente más exigentes, no menos.

Sobre el tamaño de la población objeto de estudio, aquella de la que se seleccionará la muestra, pueden hacerse algunas consideraciones. En muchas ocasiones no se conoce exactamente el número de individuos que la componen, o incluso se desconoce por completo, pero, por suerte, este elemento tiene muy poca influencia en el cálculo del tamaño de muestra la mayoría de las veces. Supongamos, por ejemplo, que se planifican dos estudios de prevalencia de obesidad, uno en una ciudad de un millón de habitantes mayores de 18 años, y otro en uno de sus municipios que tiene 10.000. Se quiere estimar el porcentaje de obesos con un error no mayor del 3% y se espera que dicho porcentaje esté en torno a un 30%. Para un nivel de confianza del 95%, los tamaños de muestra que se obtienen con los datos anteriores y muestreo aleatorio simple (más adelante veremos alguna herramienta para hacer este cálculo) son de 896 en la ciudad y 823 en el municipio ¡casi iguales! En el peor de los casos, cuando el tamaño de la población es totalmente desconocido (veremos que realmente no es tan ‘malo’), simplemente puede obviarse a la hora de hacer el cálculo del tamaño muestral, asumiendo que la muestra se obtendrá de una hipotética población infinita, es decir, de tamaño mucho mayor. Para aclarar esto, volvamos al ejemplo de la obesidad, ¿qué ocurre si desconocemos el tamaño de la población en la ciudad? Pues resulta que necesitaríamos seleccionar 897 sujetos y, ante este dato, podemos preguntarnos, ¿es apreciable la diferencia con los números anteriores? La respuesta queda para el análisis del lector.

Por último, la cuestión más controvertida en el cálculo del tamaño de muestra es que es necesario conocer de antemano el valor de uno o varios parámetros que sólo podrán estimarse una vez realizado el estudio. Aquí se plantea una clara contradicción para la que no hay ninguna ‘receta’ mágica. Una posibilidad es realizar un estudio piloto para obtener un valor aproximado del parámetro desconocido; la alternativa más frecuente consiste en utilizar valores obtenidos de estudios previos similares.

En resumen, hay un componente subjetivo en la determinación de cualquier tamaño de muestra: ¿qué nivel de confianza?, ¿por qué el 95% y no otro?, ¿cuál es la precisión adecuada para mi estudio?, ¿qué valor asumo de antemano para la proporción?... Así, dos investigadores que, por separado, estén diseñando el mismo estudio pueden obtener tamaños muestrales muy diferentes, aun usando la misma fórmula. Por supuesto, esto no significa que haya que pasarse al extremo opuesto y ‘renegar’ por completo de los métodos formales. Simplemente, debemos tener presente que los cálculos nos orientan, pero en la decisión final intervendrán otros elementos como los ya mencionados recursos materiales, la ‘historicidad’ del problema, los tamaños usados en trabajos similares y, ¿por qué no?, el sentido común. Una completa reflexión sobre el cálculo del tamaño de muestra puede encontrarse en libros específicos [1].

Herramientas disponibles para determinar tamaños de muestra

Actualmente existe una amplia variedad de programas que nos facilitan el cálculo del tamaño muestral, unos comerciales y otros de libre distribución. Entre estos últimos, existen dos interesantes programas disponibles en español: GRANMO, desarrollado específicamente para determinar tamaños de muestra, y Epidat, que incluye opciones para efectuar este cálculo. Los dos se caracterizan por la sencillez del manejo, tanto en la entrada de datos como en la exportación de los resultados. Y en ambos casos no hay que saberse las fórmulas, pero sí qué nos va a preguntar el programa para hacer los cálculos: prevalencia estimada, tipo de estudio, precisión deseada, diferencias que se desean detectar, errores α y β , etc., según el caso.

GRANMO (versión 5.2, enero 2000) es un programa en Windows desarrollado por el Instituto Municipal de Investigación Médica (IMIM) y se puede descargar desde su sitio web (<http://www.imim.es>). Los algoritmos usados por el programa para el cálculo del tamaño de muestra se encuentran detallados en un artículo publicado por los autores del mismo [2].

Epidat (versión 3.1, enero 2006) es un programa de libre distribución dirigido a epidemiólogos y otros profesionales de la salud para el manejo de datos tabulados, y que ha sido desarrollado por la Dirección Xeral de Saúde Pública de Galicia en colaboración con la Organización Panamericana de la salud (OPS-OMS). La versión 3 opera en entorno Windows e incluye 12 módulos con diferentes métodos estadísticos y epidemiológicos [3]. Específicamente, tiene un módulo de muestreo para cálculo de tamaños de muestra y selección muestral. El Epidat 3.1, que está traducido a tres idiomas (gallego, inglés y portugués), está disponible en la web de la Dirección Xeral de Saúde Pública (<http://dxsp.sergas.es>).

En resumen, el cálculo del tamaño de la muestra debe abordarse muy al principio de la fase de diseño del estudio, y el investigador, que tendrá que tomar decisiones al respecto, debe considerarlo como una ayuda a la hora de diseñar su estudio y valorar su viabilidad. En esta fase, debido al número elevado de asunciones que necesita, algunas de ellas con poca información disponible, considerarlo de manera rígida es un error y es más razonable hacerlo como un muy valioso elemento orientativo para los objetivos descritos.

Bibliografía

1. Silva LC. Diseño razonado de muestras y captación de datos para la investigación sanitaria. Madrid: Díaz de Santos; 1997.
2. Marrugat J, Vila J, Pavesa J, Sanz F. Estimación del tamaño de muestra en la investigación clínica y epidemiológica. Med Clin (Barc) 1998; 111: 267-76.
3. Hervada-Vidal X, Santiago-Pérez MI, Vázquez-Fernández E, Castillo-Salgado C, Loyola-Elizondo E, Silva-Ayçaguer LC. Epidat 3.0: Programa para el análisis epidemiológico de datos tabulados. Versión 3.0. Rev Esp Salud Pública 2004; 78:277-80.