

**ORIGINAL**

## Un esfuerzo ímprobo, aunque innecesario. Factores que condicionan el tamaño muestral.

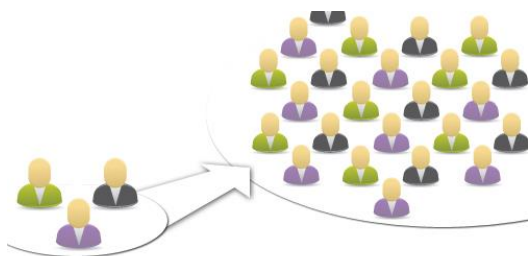
Molina M.

Complejo Hospitalario Universitario de Albacete.

### Resumen

Se revisa la necesidad de realizar un cálculo correcto del tamaño muestral necesario para un estudio, así como los principales factores que influyen en la muestra necesaria para demostrar el efecto considerado importante desde el punto de vista clínico.

### Introducción



Se revisa la necesidad de realizar un cálculo correcto del tamaño muestral necesario para un estudio, así como los principales factores que influyen en la muestra necesaria para demostrar el efecto considerado importante desde el punto de vista clínico.

En nuestros días, la enseñanza de la Medicina y, en general, la enseñanza a nivel universitario está bastante bien definida y estandarizada. Y esto no es solo a nivel nacional, sino también a nivel de nuestro entorno.

Pero esto no ha sido siempre así. En los comienzos, cada uno iba por su camino y había diversidad en los objetivos docentes y en las formas de enseñanza, como veréis en la pequeña historia que os cuento a continuación.

### Un poco de historia

A finales del siglo XIX y comienzos del pasado siglo XX, las escuelas de medicina de Estados Unidos andaban un poco perdidas en cuanto a sus objetivos y formas de enseñanza. Aunque había honrosas excepciones, como las Escuelas de Medicina de Johns Hopkins, Harvard o Michigan, la mayoría tenían una calidad más que deficiente.

El sistema educativo estaba orientado a permitir que el docente tuviese tiempo para sus cosas, con lo que se basaba en las clases magistrales y la escasez de prácticas durante la formación que, en algunas escuelas, se completaba en periodos tan ridículos como de dos semestres.

Así, llegado el año 1906, el Consejo de Educación Médica de la Asociación Médica Americana se empieza a preocupar por el asunto y a recoger información. Ante lo que se encontraron, y por aquello de mantener la objetividad, encargaron a un tercero, la Fundación Carnegie para el Avance de la Educación, el desarrollo de un informe sobre el tema.

Y la Fundación, a su vez, se la encargó a un señor que se llamaba Abraham Flexner, que se había graduado unos 20 años antes en la Johns Hopkins. Este señor no solo no delegó ya en nadie más, sino que se tomó el trabajo con gran empeño: estudió las condiciones de admisión, las instalaciones, la competencia del profesorado y otros aspectos de las escuelas de medicina de Estados Unidos y Canadá.

Hasta aquí nada anormal. Pero lo curioso es que lo estudió en TODAS las escuelas que eran, por aquel entonces, un total de 155. Un gran mérito por un trabajo ímprobo, pero seguramente podría haberse ahorrado esfuerzo (y tiempo y dinero) si hubiese seleccionado un número de escuelas representativas y hubiese reducido, así, el número de establecimientos a investigar.

### **La importancia de seleccionar una muestra adecuada**

Así que ya habéis visto como el señor Flexner fue capaz de incluir en su estudio al 100% de su población diana, cosa de la que pocos pueden presumir. Claro que, además de innecesario, muchas veces esto no es posible e, incluso, puede no ser conveniente.

Una cosa es estudiar escuelas de medicina y otra muy diferente es comparar la eficacia o la seguridad de un nuevo tratamiento respecto al tratamiento estándar o al placebo.

Un principio básico para la investigación biomédica, el principio de indeterminación clínica, nos dice que, para comparar dos tratamientos en un ensayo, el investigador tiene que ignorar realmente cuál de los dos es mejor. Una vez que este principio deja de cumplirse, no es ético continuar el ensayo o realizar otro similar.

La razón es porque, aunque el investigador cree que su nuevo tratamiento va a ser mejor, es posible que sea igual o, incluso, peor que la opción de comparación, con lo que somete a los participantes del ensayo a un riesgo.

Esta es una de las razones que hace tan conveniente el cálculo previo del tamaño muestral necesario: debemos saber cuál es el número mínimo de participantes que necesitamos para poder demostrar estadísticamente el efecto del nuevo tratamiento si este efecto existe, cosa que no sabemos cuando empezamos el estudio.

No sería ético incluir más pacientes de los necesarios solo para obtener la ansiada  $p < 0,05$ . Debemos establecer el efecto clínicamente importante que queremos detectar y calcular el tamaño muestral para que el estudio tenga la potencia necesaria para detectarlo.

### **Condicionantes del tamaño muestral**

El tamaño de la muestra es diferente en cada situación y depende de numerosos factores. No vamos a ver en esta entrada la forma de calcular el tamaño muestral en cada una de las situaciones, sino que nos vamos a limitar a reflexionar sobre los condicionantes que nos pueden influir en la forma de calcularlo y en el tamaño necesario obtenido.

Veamos algunos de estos factores que deberemos tener en cuenta cuando planifiquemos el tamaño muestral necesario para nuestro estudio.

### **Cuál es la diferencia clínicamente importante que queremos detectar**

Un vicio muy extendido es buscar de forma ansiosa obtener una  $p$  que sea estadísticamente significativa. Cuando vemos la  $p < 0,05$  se nos ilumina la cara y ya no pensamos en nada más.

Craso error: la significación de la  $p$  depende, entre otras cosas, del tamaño de la muestra. Y, como ya hemos comentado, no se trata de obtener una  $p$  significativa, sino de estudiar una magnitud de efecto que consideremos clínicamente importante.

Esta diferencia la marca el investigador, habitualmente en base a sus conocimientos sobre el tema que esté estudiando o según lo que se haya publicado o se conozca por estudios previos.

Cuando comparamos dos intervenciones en un ensayo clínico partimos siempre de la hipótesis nula de que ambas intervenciones son igual de eficaces. Sabemos que, simplemente por azar, aunque la hipótesis nula sea cierta, el valor de la variable de resultado que obtengamos será diferente en los dos grupos.

Por ejemplo, supongamos que estudiamos dos hipotensores A y B y medimos la diferencia de presión arterial media entre el final y el inicio de la intervención. De entrada, la hipótesis nula supone que las diferencias serán iguales en los dos grupos.

Sin embargo, como ya sabemos, los valores que obtendremos serán diferentes, así que nos preguntaremos cuál es la probabilidad de que esa diferencia se deba al azar. Si la probabilidad es menor del 5% ( $p < 0,05$ ), nos sentiremos con la confianza necesaria para rechazar la hipótesis nula y concluiremos que uno de los tratamientos es más eficaz que el otro.

El problema es que, por muy pequeña que sea la diferencia entre los dos grupos, se puede conseguir la significación estadística (la  $p < 0,05$ ) si se aumenta el tamaño muestral lo suficiente.

Imaginad que el tratamiento A disminuye la presión arterial 20 mmHg y el B, 18 mmHg. Si metemos el número suficiente de participantes podremos obtener la  $p < 0,05$  pero ¿realmente podemos concluir que A es mejor que B solo con esta diferencia? Evidentemente, no. Desde un punto de vista clínico, yo diría que tienen una eficacia similar.

Nosotros deberíamos marcar qué diferencia nos parece importante. Por ejemplo, podemos decidir que queremos detectar una diferencia entre los dos fármacos de 20 mmHg o más. Con esta diferencia, calcularemos el número de participantes necesarios para, si existe esta diferencia, que la  $p$  sea significativa. No haría falta ni uno más ni uno menos.

Si nos quedamos por debajo de este tamaño necesario, aunque detectemos una diferencia de 20 mmHg, la  $p$  puede no ser significativa. El estudio no tiene la potencia necesaria para detectar el efecto por falta de tamaño muestral.

Si la diferencia detectada es menor de 20 mmHg, la  $p$  tampoco será significativa. No pasa nada, no existe una diferencia clínicamente importante entre los dos tratamientos. Lo que no tendría sentido es aumentar el tamaño de la muestra para demostrar la significación estadística de un efecto menor que el considerado clínicamente importante.

Una advertencia antes de dejar este punto: todo lo que hemos dicho se desarrolla en el reino de las probabilidades, por lo que siempre tenemos una cierta probabilidad de cometer un error al realizar el contraste de hipótesis (error de tipo I y error de tipo II).

**La variabilidad en la población del parámetro en estudio**

Este es otro factor importante. Cuanto mayor sea la variabilidad de la variable de resultado de nuestro estudio en la población diana, mayor será el tamaño de muestra necesario para detectar una misma magnitud de efecto.

La variabilidad en la población se refleja en la desviación estándar, que influye en los cálculos del error estándar y de los intervalos de confianza. Cuanto mayor sea el error estándar de la variable, mayor será el tamaño muestral necesario, ya que las estimaciones sobre la población son menos precisas.

Lo mismo ocurre con la precisión de la estimación que queremos realizar. Mientras más precisa queramos que sea nuestra estimación, mayor será el tamaño de la muestra necesaria, y viceversa.

### **La confiabilidad que esperamos del estudio**

La confiabilidad del estudio depende de dos parámetros cuyo valor debemos fijar para realizar el cálculo del tamaño de la muestra: el nivel de confianza y la potencia del estudio.

El nivel de confianza refleja el grado de seguridad que tenemos de que, si repetimos el estudio en las mismas condiciones, volveremos a obtener un resultado similar. Habitualmente se elige un nivel de confianza del 95%, aunque podemos subirlo o bajarlo en función de lo estrictos que queramos ser con el grado de seguridad necesario.

La potencia, por su parte, refleja la probabilidad de que los resultados que obtenemos en el estudio representan la realidad. Como ya hemos dicho, es la probabilidad de que el estudio detecte el efecto, si este existe. Habitualmente se suele marcar en un 80%, aunque también puede aumentarse hasta el 90% en algunos estudios.

Como es fácil de intuir, a mayor nivel de confianza y mayor potencia del estudio, el tamaño muestral necesario será mayor, y viceversa.

### **El tipo de estudio que vamos a realizar**

Estamos hablando todo el tiempo de ensayos clínicos, pero el cálculo del tamaño muestral se aplica también a otros diseños metodológicos.

Así, podemos calcular el tamaño muestral necesario para hacer estimaciones de prevalencia en estudios transversales con una precisión determinada, comparar la asociación y las medidas de riesgo en estudios observacionales, establecer la correlación entre dos variables, etc.

Lógicamente, el tipo de diseño influirá en la forma de calcular el tamaño de la muestra y en el número de participantes necesario.

### **Otra vez los grupos apareados**

Es importante establecer la relación que existe entre los dos grupos que queremos comparar que, como ya sabemos, pueden ser independientes o apareados.

Como ya es sabido, la variabilidad es mayor entre grupos independientes que entre grupos apareados, lo que influirá en el tamaño muestral necesario, que será mayor siempre cuando manejemos grupos independientes.

### **La dirección del contraste**

El contraste de hipótesis puede ser unilateral o bilateral (con una o dos colas).

El contraste bilateral asume en su hipótesis alternativa que hay una diferencia entre las dos intervenciones

comparadas, pero no entra en cuál de las dos es más eficaz. Por su parte, el contraste unilateral sí establece en la hipótesis alternativa cuál de las dos intervenciones es superior.

Lo más habitual es elegir el contraste bilateral, ya que cuando realizamos un experimento no sabemos la dirección que puede tomar el resultado. Sin embargo, si estamos seguros de cuál va a ser la dirección del efecto, podemos adoptar un contraste unilateral.

El contraste bilateral es más conservador, por lo que es más difícil alcanzar la significación estadística que con el contraste unilateral y, además, requiere un tamaño de la muestra superior.

De todas formas, no nos confundamos: lo elegante es realizar un contraste bilateral y, si optamos por uno unilateral, nunca debería ser para alcanzar la  $p$  significativa con más facilidad o con menos participantes.

### Las características de la variable

Lógicamente, el tamaño de la muestra será diferente si queremos medir una o más variables y dependerá también del tipo de variables. Este aspecto está también ligado a algo de los que hemos hablado ya, la precisión con la que queremos estimar cada variable.

### Resumiendo

Ya hemos visto los factores que nos pueden influir en el número de participantes que deberá tener nuestro estudio si queremos que sea capaz de detectar un efecto que nosotros consideramos clínicamente importante.

Para resumir, podemos decir que el tamaño de la muestra necesaria será mayor cuanto menor sea la probabilidad de error de tipo I y de tipo II que

aceptemos, cuanto mayor sea la dispersión de la variable en la población de estudio y cuanto menor sea el tamaño de efecto.

También aumentará el tamaño muestral cuando comparemos grupos independientes, cuando queramos comparar más de una variable y cuando optemos por un contraste de hipótesis bilateral.

### Nos vamos...

Y aquí vamos a dejar el tema por hoy.

Por si tenéis curiosidad sobre qué pasó con el señor Flexner, puedo decir que su informe fue demoledor. Concluyó que con 31 escuelas se podría formar mejor a los médicos que cómo se hacía con las 155 que estudió. Por tanto, recomendó reducir el número de escuelas y, como consecuencia, el número de estudiantes.

Según Flexner, se formaban demasiados médicos para las necesidades del mercado. No sé, creo que esto me suena de algo...

Y ahora sí que no vamos. Hemos hablado mucho de la importancia de que la muestra tenga el tamaño adecuado y de los factores que pueden influir en el mismo. Sin embargo, no es suficiente con que el tamaño esté bien calculado.

De nada servirá una muestra de tamaño adecuado si la técnica de muestreo nos proporciona una muestra que no sea representativa de la población de estudio. Pero esa es otra historia...

### Bibliografía

— Lwanga SK, Lemeshow S. Determinación del tamaño de las muestras en los estudios sanitarios. Manual práctico. Organización Mundial de la Salud. Ginebra, 1991. ([HTML](#))

– Velasco Rodríguez VM, Martínez Ordaz VA, Roiz Hernández J, Huazano García F, Nieves Rentería A. Muestreo y tamaño de muestra. Una guía práctica para personal de salud que realiza investigación. e-libro.net. Coahuila (México), 2002. ([HTML](#))

---

**Correspondencia al autor**

*Manuel Molina*  
[mma1961@gmail.com](mailto:mma1961@gmail.com)  
*Servicio de Gastroenterología.*  
*Hospital Infantil Universitario La Paz.*  
*Madrid. España.*

---

Aceptado para el blog en abril de  
2021

