

Cálculo del tamaño muestral en estudios de casos y controles

Pértegas Díaz S., Pita Fernández S.

Unidad de Epidemiología Clínica y Bioestadística. Complejo Hospitalario Juan Canalejo. A Coruña. Cad Aten Primaria 2002; 9: 148-150. Actualización 17/09/2002.

Estimación del riesgo en los estudios de casos y controles.

Los estudios de casos y controles son un tipo de diseño habitualmente empleado en investigaciones clínicas que tienen por objeto la identificación de factores de riesgo. Si bien los estudios de cohortes reúnen las características idóneas para llevar a cabo este tipo de análisis, los estudios de casos y controles cuentan con la ventaja de que suelen exigir menos tiempo y ser menos costosos que aquellos^{1,2}.

Una de las características esenciales de los estudios de casos y controles, y su principal diferencia con los estudios de cohortes, es la forma de selección de los sujetos de estudio. En un estudio de cohortes se selecciona a los sujetos en base a su exposición. Por el contrario, en un estudio de casos y controles, se identifica a un grupo de personas con una enfermedad (casos) y se les compara con un grupo apropiado que no tenga la enfermedad (controles). Ya que los individuos son seleccionados en función de la presencia o ausencia del evento de estudio y no por su exposición, no podrá estimarse directamente el riesgo de enfermar entre los sujetos expuestos y los no expuestos. La relación entre uno o varios factores y la presencia de la enfermedad deberá estudiarse comparando la frecuencia de exposición entre los casos y los controles. Si la frecuencia de exposición es mayor en el grupo de casos que en los controles, podremos decir que hay una asociación entre la causa y el efecto.

En estudios de este tipo, la distribución de n sujetos estudiados según presenten o no la enfermedad y según su exposición a cada factor se puede mostrar en una tabla 2 x 2, similar a la [Tabla 1](#): así, existirán en la muestra casos expuestos (a), casos no expuestos (c), controles expuestos (b) y controles no expuestos (d). Como medida de la frecuencia de exposición entre los casos se puede utilizar el cociente:

$$\Omega_1 = \frac{p_1}{1 - p_1}$$

donde p_1 es la probabilidad de exposición entre los casos. A partir de una muestra como la de la [Tabla 1](#), Ω_1 se puede estimar dividiendo los casos expuestos entre los casos no expuestos:

$$\hat{\Omega}_1 = \frac{a/c}{c/(a+c)} = \frac{a}{c}$$

De modo similar, se valora la frecuencia de exposición entre los controles mediante el cociente:

$$\Omega_2 = \frac{p_2}{1 - p_2} \Rightarrow \hat{\Omega}_2 = \frac{b/d}{d/(b+d)} = \frac{b}{d}$$

donde p_2 es la probabilidad de exposición entre los controles.

La medida más utilizada para cuantificar la asociación entre la exposición y la presencia de enfermedad es el "odds ratio" (OR) y su cálculo se estima mediante el cociente de las dos cantidades anteriores:

$$OR = \frac{\Omega_1}{\Omega_2} = \frac{p_1(1 - p_2)}{p_2(1 - p_1)} \Rightarrow OR \hat{=} \frac{a/c}{b/d} = \frac{a \times d}{b \times c}$$

La interpretación del OR es la siguiente: si el OR es igual a 1, la exposición no se asocia con la enfermedad, mientras que si el OR es menor de 1 la exposición tiene un efecto protector (es decir, la exposición disminuye la probabilidad de desarrollar la enfermedad). Por último, si el valor del OR es mayor de 1, la exposición aumenta las posibilidades de desarrollar la enfermedad. De cualquier modo, las estimaciones del OR se deben realizar con su 95% intervalo de confianza para poder confirmar o rechazar la asociación de la exposición con la enfermedad.

Como se puede observar, el valor del OR puede obtenerse de la [Tabla 1](#) multiplicando “en cruz” sus cuatro valores. De ahí que también reciba el nombre de “razón de productos cruzados”, o también “razón de ventajas”. Bajo suposiciones adecuadas, el OR puede ser un estimador adecuado de la razón de tasas de incidencia o del riesgo relativo, medidas habitualmente utilizadas para valorar la asociación entre una exposición y un evento. Cuando la frecuencia de exposición es reducida, el valor del OR y del riesgo relativo son muy similares.

Cálculo del tamaño muestral en estudios de casos y controles.

Supongamos que se quiere llevar a cabo un estudio de casos y controles con el fin de determinar si existe una relación significativa entre la exposición a un factor y la presencia de una determinada enfermedad. A continuación se explica cómo calcular el tamaño de muestra necesario para contrastar la hipótesis de que el OR sea igual a 1.

Si se conoce la probabilidad de exposición entre los controles p_2 , y se prevé que el OR asociado al factor de estudio es w , el valor de p_1 , la frecuencia de exposición entre los casos, puede obtenerse fácilmente:

$$OR = w = \frac{p_1(1-p_2)}{p_2(1-p_1)} \Rightarrow wp_2(1-p_1) = p_1(1-p_2) \Rightarrow p_1(1-p_2+wp_2) = wp_2 \Rightarrow$$

$$\Rightarrow p_1 = \frac{wp_2}{(1-p_2)+wp_2}$$

Así, el problema del cálculo del tamaño muestral podrá abordarse mediante las fórmulas habituales empleadas en la comparación de dos proporciones, asumiendo aquí que las proporciones esperadas son p_1 y p_2 [3](#).

Recurriendo a las fórmulas habituales para determinar el tamaño muestral mínimo necesario para la comparación de dos proporciones, se precisará conocer [3](#):

- a) La magnitud de la diferencia a detectar, que tenga interés clínicamente relevante. En este caso, como ya vimos, bastaría con conocer dos de los siguientes tres parámetros:
 - Una idea del valor aproximado del odds ratio que se desea estimar (w)
 - La frecuencia de la exposición entre los casos (p_1)
 - La frecuencia de la exposición entre los controles (p_2)
- b) La seguridad con la que se desea trabajar (α), o riesgo de cometer un error de tipo I. Generalmente se trabaja con una seguridad del 95% ($\alpha = 0,05$).
- c) El poder estadístico ($1-\beta$) que se quiere para el estudio, o riesgo de cometer un error de tipo II. Es habitual tomar $\beta = 0,2$, es decir, un poder del 80%.

Con estos datos, y para un planteamiento bilateral, para el cálculo del tamaño muestral se utilizará la expresión [4.5](#):

$$n = \frac{\left[z_{1-\alpha/2} \sqrt{2p(1-p)} + z_{1-\beta} \sqrt{p_1(1-p_1) + p_2(1-p_2)} \right]^2}{(p_1 - p_2)^2} \quad \text{Ecuación 1}$$

donde

$$p = \frac{p_1 + p_2}{2}$$

y los valores $z_{1-\alpha/2}$ y $z_{1-\beta}$ son valores que se obtienen de la distribución normal estándar en función de la seguridad y el poder elegidos para el estudio⁶. En particular, para una seguridad de un 95% y un poder estadístico del 80% se tiene que $z_{1-\alpha/2} = 1,96$ y $z_{1-\beta} = 0,84$.

Hasta ahora se ha asumido un tamaño muestral igual para casos y controles. En caso de que el número de casos y controles no esté balanceado, la expresión anterior deberá ser ligeramente modificada. Denotando ahora por n el número de casos y por m el número de controles la fórmula a aplicar sería⁴:

$$n = \frac{\left[z_{1-\alpha/2} \sqrt{(c+1)p(1-p)} + z_{1-\beta} \sqrt{cp_1(1-p_1) + p_2(1-p_2)} \right]^2}{c(p_2 - p_1)^2} \quad \text{Ecuación 2}$$

donde $c = m/n$ es el número de controles por cada caso. Así, el número de controles vendría dado por $m = c \times n$.

Debe precisarse que en el presente trabajo se ha tratado de exponer de modo lo más sencillo posible el procedimiento a seguir en el cálculo del tamaño de la muestra en un estudio de casos y controles. No obstante, en ocasiones se utilizan para este cálculo expresiones más complejas basadas en una corrección de la fórmula del cálculo del tamaño muestral para la comparación de dos proporciones⁴. Así mismo, existen fórmulas específicas para el cálculo del tamaño de la muestra en el caso de que el diseño corresponda a un estudio de casos y controles pareados⁷⁻⁸.

Ejemplo del cálculo del tamaño muestral en un estudio de casos y controles.

Como ejemplo, supongamos que se desea estudiar la existencia de una asociación entre el consumo de tabaco y el hecho de sufrir un infarto de miocardio. Para poner en evidencia dicha asociación y cuantificar su magnitud se diseña un estudio de casos y controles en el que se investigará el consumo de tabaco de una serie de pacientes que han padecido un infarto de miocardio (casos) y una serie de pacientes sanos (controles). Se cree que alrededor de un 40% de los controles son fumadores y se considera como diferencia importante entre ambos grupos un odds ratio de 4. Con estos datos, podemos calcular el tamaño de muestra necesario en cada grupo para detectar un odds ratio de 4 como significativamente diferente de 1 con una seguridad del 95% y un poder del 80%. De acuerdo con lo expuesto con anterioridad, conocemos los siguientes parámetros:

- Frecuencia de exposición entre los controles: 40%
- Odds ratio previsto: 4
- Nivel de seguridad: 95%
- Poder estadístico: 80%

De acuerdo con estos datos, se estima que la frecuencia de exposición entre los casos vendrá dada por:

$$p_1 = \frac{wp_2}{(1-p_2) + wp_2} = \frac{4 \times 0,40}{(1-0,40) + 4 \times 0,40} = \frac{1,6}{0,60 + 1,6} = 0,73$$

Esto es, se estima que aproximadamente un 73% de los casos son fumadores. Aplicando la [Ecuación 1](#), se obtiene:

$$n = \frac{\left[1,96\sqrt{2 \times 0,565 \times (1-0,565)} + 0,84\sqrt{0,73 \times (1-0,73)} + 0,4 \times (1-0,4)\right]^2}{(0,73-0,4)^2} \approx 35$$

Es decir, se necesitaría estudiar a 35 sujetos por grupo (35 pacientes con infarto de miocardio y 35 controles) para detectar como significativo un valor del odds ratio de 4.

Si se reduce el tamaño del efecto a detectar, asumiendo que el odds ratio es aproximadamente igual a 3, se obtiene:

$$p_1 = \frac{wp_2}{(1-p_2)+wp_2} = \frac{3 \times 0,40}{(1-0,40)+3 \times 0,40} = 0,67$$

y, de acuerdo con la [Ecuación 1](#), serían necesarios n=54 pacientes por grupo para llevar a cabo el estudio.

En algunos estudios, el investigador reúne un número mayor de controles que de casos con el objeto de incrementar el poder estadístico. Supongamos que en el presente ejemplo se planea obtener dos controles por caso, y se asume que el odds ratio a detectar es aproximadamente igual a 3. Aplicando la [Ecuación 2](#):

$$n = \frac{\left[1,96\sqrt{(2+1) \times 0,565 \times (1-0,565)} + 0,84\sqrt{2 \times 0,73 \times (1-0,73)} + 0,4 \times (1-0,4)\right]^2}{2 \times (0,73-0,4)^2} \approx 40$$

Por tanto, se necesitaría un grupo de n=40 casos (pacientes con infarto de miocardio) y m=2x40=80 controles para llevar a cabo la investigación.

El cálculo del tamaño de la muestra en los estudios de casos y controles debe formar parte del diseño metodológico del mismo, ya que la ejecución de este tipo de estudios es costosa. El iniciar un estudio sin conocer el poder estadístico y la seguridad para detectar diferencias, si es que existen, podría ser motivo de cometer un error de tipo II en el sentido de no detectar diferencias cuando realmente las hay.

TABLA 1. Disposición de los sujetos incluidos en un estudio de casos y controles. Tabla de 2 x 2.

	Casos	Controles	
Expuestos	a	b	a + b
No expuestos	c	d	c + d
	a + c	b + d	n

Bibliografía

1. Kelsey JL, Thompson WD, Evans AS. Methods in Observational Epidemiology. New York: Oxford University Press; 1986.
2. Hennekens CH, Buring JE. Epidemiology in Medicine Boston: Litle, Brown and Company; 1987.
3. Pita Fernández S. Determinación del tamaño muestral. Cad Aten Primaria 1996; 3: 138-141. [[Texto completo](#)]
4. Fleiss JL. Statistical methods for rates and proportions. 2nd edition. New York: John Wiley & Sons; 1981.
5. Lwanga SK, Lemeshow S. Determinación del tamaño de las muestras en los estudios sanitarios. Manual práctico. Ginebra: O.M.S.; 1991

6. Pértega Díaz S, Pita Fernández S. La distribución normal. Cad Aten Primaria 2001; 8: 268-274. [[Texto completo](#)]
7. Fleiss JL, Levin B. Sample size determination in studies with matched pairs. J Clin Epidemiol 1988; 41(8): 727-730. [[Medline](#)]
8. Schlesselman JJ. Case-Control Studies. New York: Oxford University Press; 1982.