

Diseño y aplicación de los estudios de casos y controles

ARMANDO JUÁREZ-BENGOA,^{a,b} RICARDO FIGUEROA-DAMIÁN^{a,c}

RESUMEN

Los estudios de casos y controles constituyen un tipo de enfoque metodológico adecuado para la investigación clínica, si el resultado que se investiga es poco frecuente o se requiere un tiempo de observación muy prolongado. Se debe seleccionar una muestra de la población e identificar claramente los casos, que son los sujetos que padecen el evento o la enfermedad, así como a los controles, que son los sujetos que no lo padecen. Una característica muy importante de estos estudios, es que parten de la identificación de los casos y los controles en el presente, pero la información sobre la exposición al factor en estudio, se recolectó en el pasado, generalmente, a partir de entrevistas, cuestionarios o registros hospitalarios. Esto genera un grado de incertidumbre sobre la veracidad de la información, porque ésta puede ser insuficiente o inadecuada, lo que representa la principal debilidad de los estudios con este diseño. En cambio, tienen la ventaja de que pueden utilizarse para investigar problemas de salud poco frecuentes o con un tiempo de evolución prolongado entre la exposición al factor de riesgo y el inicio de la enfermedad, consumen poco tiempo, su costo es bajo y se obtiene una gran cantidad de información a partir de un número relativamente pequeño de individuos. En esta revisión, se presenta una forma práctica para abordar los estudios de casos y controles, así como las bases estadísticas para decidir la validez de los resultados.

PALABRAS GUÍA: Metodología, casos y controles, investigación clínica.

INTRODUCCIÓN

En la actualidad los eventos clínicos relacionados con el diagnóstico, la causalidad o el tratamiento son sometidos, en un número cada vez mayor, a proyectos

de investigación. Dada la diversidad de situaciones, el abordaje de cada uno de los problemas planteados enfrenta diferentes obstáculos en su investigación, siendo relativamente simples en algunos casos, pero muy difíciles en otros. El método de estudio depende de las características del evento, del tiempo necesario para llevarlo a cabo, de su factibilidad y del costo del mismo, entre otros factores. Uno de los métodos de investigación clínica ampliamente utilizado en la actualidad, se conoce con el nombre de: estudio de casos y controles. Esta revisión tiene por objetivo presentar los conceptos básicos necesarios para comprender este tipo de estudios y está dirigido a los médicos que pretenden iniciarse en la investigación y esperan encontrar los aspectos prácticos de su aplicación.

^a Instituto Nacional de Perinatología Isidro Espinosa de los Reyes.

^b HGZ 53, IMSS.

^c Hospital de Infectología, Centro Médico La Raza, IMSS.

Correspondencia:
Dr. Armando Juárez Bengoa
Instituto Nacional de Perinatología
Montes Urales 800 Col. Lomas de Virreyes
C.P. 11000, México, D.F.
Tercer piso Torre de Investigación.

Recibido: 23 de noviembre de 2007.

Aceptado: 29 de noviembre de 2007.

¿PARA QUÉ SE UTILIZAN?

Los estudios de casos y controles se utilizan en aquellas situaciones en las que se desea conocer la probabilidad de que ocurra un evento clínico cuando existe el antecedente de exposición a un factor de riesgo. Por ejemplo, conocer la probabilidad de aparición de cáncer cervicouterino dado el antecedente de infección por el virus del papiloma humano. Una estrategia para encontrar la respuesta a este problema podría ser esperar la aparición del evento después de la exposición a la condición, probablemente causal, y buscar la asociación entre ellos, pero esta estrategia trae consigo dificultades que en ocasiones son insalvables, en primer término, los impedimentos éticos. También puede ocurrir que el tiempo de seguimiento de los pacientes sea muy prolongado, que el estudio resulte muy costoso o que los sujetos de la investigación abandonen el estudio.

Los estudios de casos y controles son una alternativa adecuada, especialmente cuando la ocurrencia del fenómeno de desenlace es poco frecuente o se requiere de un seguimiento prolongado para observar el resultado.

¿CUÁLES SON LOS CASOS?

Lo más importante al iniciar este tipo de estudios es especificar claramente la definición de *caso*. El término se refiere a la identificación precisa de un individuo que presenta el evento en estudio.^{1,2} Los criterios con los que se define deben minimizar las probabilidades de que se omita la identificación de una persona afectada, es decir, la definición debe ser suficientemente sensible para identificar todos los casos. A la vez, se debe minimizar la clasificación falsa de un individuo no afectado como caso; dicho de otro modo, los criterios también deben ser específicos. En la práctica esto se realiza a través de la definición clara de los *criterios de inclusión*, que se establecen antes de comenzar un estudio. Dado que la identificación de todos los casos en una población determinada puede consumir demasiado tiempo (o puede ser incluso irrealizable, si ésta es muy grande) el estudio suele realizarse solamente con una *muestra* de la población.³

¿CUÁLES SON LOS CONTROLES?

Los *controles* son individuos que no presentan el evento en estudio, aunque pudieron haberse expuesto a la variable potencialmente causal que se está investigando.

Las características que los identifican también se definen a través de los *criterios de inclusión* que se establecen antes de iniciar el estudio. Con el fin de evitar sesgos de selección que pueden afectar el resultado, es conveniente que los controles sean obtenidos de la misma población de la que provienen los casos.^{4,5}

Una estrategia para disminuir el efecto confusor de otras variables es no incluir individuos, tanto en el grupo de casos como en el grupo de controles, que presenten características que puedan influir en el resultado. Otra alternativa para reducir el efecto de las variables de confusión es el *pareamiento*, que consiste en la selección de controles con características similares (sexo, edad, etc.) a los casos, excepto por la condición de estar o no afectados. El pareamiento disminuye el sesgo asociado a los factores de confusión, pero impide analizar el efecto del riesgo de las variables de pareamiento que por definición son iguales en los casos y en los controles.⁶

¿CÓMO ES EL DISEÑO?

Los estudios de investigación clínica se dividen en dos grupos principales: los experimentales y los observacionales. En los primeros el investigador aplica una maniobra que provoca un efecto en el resultado; por ejemplo, la administración de un medicamento nuevo para el tratamiento de una enfermedad. En los estudios observacionales la investigación se realiza sin una maniobra de intervención y el investigador sólo observa a los sujetos y al fenómeno motivo del estudio. Los estudios observacionales, a su vez, se dividen en descriptivos y analíticos; los primeros sólo señalan las características de los sujetos y del fenómeno en estudio, efectuándose algunas mediciones descriptivas como frecuencia, proporción, tasas de prevalencia o tasas de incidencia. Los estudios observacionales analíticos realizan además comparaciones, asociaciones o establecen relaciones que permiten calcular un riesgo determinado dentro del fenómeno analizado y proporcionan resultados y conclusiones más sólidos.

Los estudios de casos y controles pertenecen al grupo de estudios observacionales analíticos. Una característica importante en este diseño es la perspectiva de tiempo para asociar la causa con el efecto y, por lo tanto, son clasificados como longitudinales. Además, dado que se parte de un resultado identificado en el presente y a partir de éste se trata de dilucidar la causa en el pasado, este tipo de



estudios son retrospectivos. El diseño de un estudio de casos y controles se muestra gráficamente en la figura 1.

La primera fase en el diseño de cualquier estudio de investigación consiste en la definición de la pregunta que se pretende resolver, con base en ella, se establece la necesidad de un estudio cuya perspectiva de evaluación es hacia el futuro (prospectivo) o hacia el pasado (retrospectivo). Dentro de un estudio de casos y controles, después del planteamiento de la pregunta de investigación, la siguiente etapa es la selección de los casos de acuerdo con los criterios establecidos de antemano (criterios de inclusión) y junto con ellos se selecciona a los controles. Posteriormente, se procede a obtener la información precisa de cada individuo respecto al antecedente de exposición al factor que se investiga y, finalmente, se compara el resultado entre los dos grupos establecidos.^{7,8}

VARIANTES DEL DISEÑO

En ocasiones es necesario hacer modificaciones a los diseños básicos para efectuar la investigación. Por ejemplo, cuando se estudian enfermedades poco frecuentes y la medición de la variable predictora es costosa, se hace uso de una variante del diseño de casos y controles, conocida como *casos y controles anidado en una cohorte*. En esta situación se parte de un grupo de sujetos a los que se les hace una medición basal quedando expuestos al factor de riesgo que se investiga y se sigue a lo largo del tiempo, constituyendo así una cohorte. En el seguimiento de esta cohorte se

identifica a los individuos que desarrollan la enfermedad, siendo éstos los que conformarán el grupo de casos. Entre los individuos de la cohorte que no desarrollan la enfermedad se selecciona a los integrantes del grupo de controles;¹ sólo en los individuos seleccionados como casos y controles de la cohorte, se efectuarán las evaluaciones, partiendo de la medición basal que se les realizó a todos.

Una de las ventajas de este diseño es que la totalidad de los casos son incluidos en la evaluación (aún los que hayan fallecido) puesto que la medición de la exposición se efectuó al inicio, al momento de inclusión en la cohorte. Por otro lado, tanto los casos como los controles se obtienen de la misma población, lo que elimina el sesgo de selección de los individuos participantes. Además, como la medición se hace solamente en una muestra de los controles y no en todos los que conforman la cohorte original, el costo disminuye.^{1,9} Este tipo de estudio tiene el inconveniente de que el seguimiento es más difícil de realizar, porque requiere de cohortes, en ocasiones hay muchas pérdidas y no siempre existe material almacenado o conservado adecuadamente para efectuar los estudios *a posteriori*.¹

¿CÓMO SE OBTIENE LA INFORMACIÓN?

En el diseño de casos y controles se trata de identificar el factor causal en el pasado y esto puede generar incertidumbre del tiempo preciso en que inició la enfermedad. Por lo tanto, es crítico conocer el momento exacto de la exposición al factor

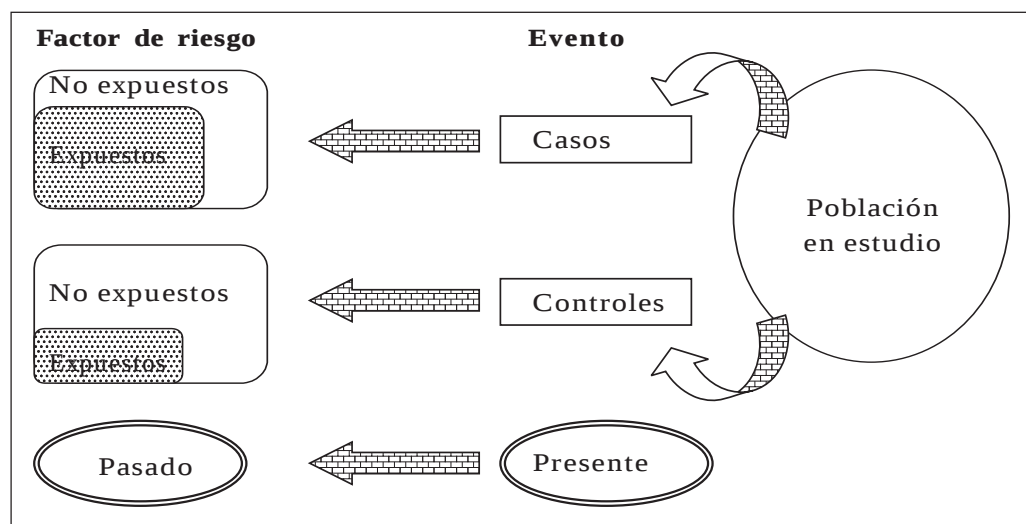


Figura 1
Diseño de los estudios de casos y controles
En un estudio de casos y controles se parte de la presencia (casos) o ausencia (controles) de un evento y se compara con la exposición previa al probable factor de riesgo

de riesgo. Tratándose de enfermedades de lenta evolución, puede ser difícil determinar el tiempo que pasa entre la exposición y el inicio de la enfermedad.³

Debido a que el factor de riesgo se busca de manera retrospectiva, las entrevistas y los cuestionarios son los medios más comunes para determinar la historia de exposición del sujeto. Para asegurar que la información de los casos y los controles se obtiene de la misma manera, las entrevistas deberán realizarse de manera uniforme por personas entrenadas. La utilidad de las entrevistas radica en que se puede explorar un campo extenso de factores de riesgo, son de bajo costo y, a través de ellas, se puede obtener información de exposiciones ocurridas varios años antes del inicio de la enfermedad.³

La información también puede ser obtenida de registros, ya sea médicos o de otro tipo, evitando el sesgo potencial que puede surgir de las entrevistas, aunque la información que se obtiene de informes elaborados en el pasado, tiene la desventaja de que suele ser incompleta.³

La utilidad principal de los estudios de casos y controles se encuentra en la investigación de problemas de salud poco frecuentes o de aquellas enfermedades que tienen un tiempo de latencia muy largo entre la exposición y la enfermedad;^{8,9} son más eficientes que otros diseños para este tipo de eventos y en muchas ocasiones son la única opción posible.¹⁰

Comparado con otro tipo de estudios, el de casos y controles requiere menos tiempo y en general es menos costoso. Pueden identificarse simultáneamente los efectos de diversos factores de riesgo posibles y requiere de un menor número de sujetos.¹⁰ Además, se obtiene una gran cantidad de información a partir de un número relativamente pequeño de individuos.⁷

Por otro lado, hay una serie de desventajas que no pueden dejar de considerarse. En principio, dado que es un estudio retrospectivo, puede haber errores o sesgos en su interpretación.⁹ Los sesgos provienen principalmente de dos fuentes:

- 1) El hecho de que los casos y los controles se obtienen de muestras diferentes.
- 2) La medición de las variables predictoras es retrospectiva.⁷

Dado que los casos se obtienen de una reserva disponible de la población, es difícil asegurar una

comparación adecuada de los factores de riesgo, debido sobre todo a que la exposición se mide después de que la enfermedad se ha desarrollado.¹¹ Otra desventaja, es que puede haber un sesgo al seleccionar a los individuos, tanto los casos como los controles, ya que esta selección depende del investigador.¹² Por otro lado, no hay una manera directa de estimar la incidencia o la prevalencia porque la información disponible es limitada.⁷

¿QUÉ PRUEBAS ESTADÍSTICAS SE UTILIZAN?

La prueba estadística dependerá del diseño del estudio y de las variables a estudiar. En general, si se trata de variables continuas con distribución normal y se trata de comparar promedios entre el grupo de casos y el grupo de controles, se utilizan pruebas paramétricas como la prueba *t de Student*. Si son variables discretas o no tienen una distribución normal, se deberán utilizar pruebas no paramétricas, como la prueba de *chi* cuadrada.

Al tratar de investigar factores de riesgo, el parámetro de medición es la razón de momios.^{12,13} Para esto se elabora una tabla de 2 por 2, en la que se coloca en cada una de las celdas el valor correspondiente a los casos y los controles, tanto de los expuestos como de los no expuestos, tal y como se indica en la tabla 1.

En la tabla 1, las celdas *a* y *b*, representan el número de los casos expuestos y los no expuestos, en tanto que las celdas *c* y *d*, representan los controles expuestos y los no expuestos, respectivamente. Para el cálculo de la razón de momios, en primer lugar se obtiene la razón entre los casos expuestos y los no expuestos (*a/c*); después la razón de los controles expuestos y los no expuestos (*b/d*); finalmente, la razón de momios se obtiene dividiendo estas dos razones.¹⁴

$$\text{Razón de momios} = \frac{a/c}{b/d}$$

La razón de momios indica el número de veces que tiene la probabilidad de exposición al factor de riesgo entre los casos con respecto a la probabilidad de falta de exposición. Dicho de otra manera, si la razón de momios es mayor de uno, la exposición está asociada positivamente a la enfermedad, tanto



Tabla 1
Cálculo de la razón de momios

	Expuestos	No expuestos	Total
Casos	a	b	a + b
Controles	c	d	c + d
Total	a + c	b + d	N

En una tabla de 2 por 2 las letras *a* y *b* representan los casos expuestos y los no expuestos respectivamente; *c* y *d* representan los controles expuestos y los no expuestos respectivamente. Los totales se muestran en las celdas marginales.

más cuanto mayor sea el valor de la razón de momios. Si el valor es de uno o menor, entonces no hay un efecto de la exposición sobre la aparición de la enfermedad.

El valor de la razón de momios, por sí solo, no es suficiente para decir con certeza si la exposición está asociada a la enfermedad. Se requiere establecer un intervalo de valores que garantice que ese valor es significativo y el fenómeno ocurrirá nuevamente si el estudio se repite muchas veces. Este rango se conoce como intervalo de confianza y se obtiene mediante la siguiente fórmula:¹⁵

$$e^{[\ln(RM) \pm 1.96(DE)]}$$

En la fórmula *ln* es el logaritmo natural y *RM*, es la razón de momios. *DE* es la desviación estándar y se calcula de la siguiente manera:

$$DE = \sqrt{(1/a) + (1/b) + (1/c) + (1/d)}$$

donde:

a, *b*, *c*, y *d*, representan los valores contenidos en las celdas de la tabla 1.

Mientras más amplio es el intervalo de confianza, el resultado es menos preciso. Si el intervalo comprende valores mayores de 1, la asociación entre el factor de riesgo y la enfermedad es real. Si el intervalo de confianza incluye la unidad, entonces la asociación es cuestionable.

ESTIMACIÓN DEL TAMAÑO DE LA MUESTRA

Un aspecto importante para el investigador es conocer el número de sujetos que se requiere para

que el estudio tenga mayor solidez en sus resultados. En general, mientras más grande es la muestra, es mayor la probabilidad de que el análisis de las comparaciones efectuadas en el estudio tenga más sustento; pero también, con un número mayor de sujetos en el estudio, es mayor el gasto de la investigación; de aquí la importancia de establecer un número adecuado de individuos.

Entre la serie de factores que deben ser considerados para definir el número de sujetos a incluir en el estudio, el primero a considerar es la manera como ha sido planteada la pregunta de investigación, porque de ésta se deriva la hipótesis del estudio. A su vez, se requiere de la hipótesis para definir las bases del cálculo muestral.¹⁶ El contenido de la hipótesis implica las variables a estudiar, la asociación que se establece entre ellas, su escala de medición y la fuerza de evaluación. Al mismo tiempo en la hipótesis se señala si la asociación entre las variables predictor y de desenlace, es en una dirección o en dos, lo que se nombra en estadística de una o de dos colas. Una hipótesis de una cola requiere una muestra de menor tamaño.

Las pruebas estadísticas están encaminadas a evaluar una hipótesis con una cierta probabilidad de certeza y, por lo tanto, también tienen una probabilidad de error. El error de tipo I, también llamado alfa, consiste en aceptar como verdadera una hipótesis que en realidad es falsa. El error tipo II, también conocido como beta, consiste en rechazar una hipótesis que es verdadera. El investigador fija la probabilidad máxima aceptable de cometer los errores de tipo I y II, antes de efectuar el estudio. Habitualmente se fijan los valores de alfa y de beta en 0.05 y 0.20, respectivamente¹⁷ y son incluidos en las fórmulas del cálculo de la muestra.

Otro aspecto que influye en el tamaño de la muestra es la magnitud de la asociación entre la variable predictor y la variable de desenlace. El investigador debe elegir el grado de la asociación, llamado también *tamaño del efecto*, que le gustaría ser capaz de detectar en la muestra.

En el cálculo de la muestra también es importante considerar la variabilidad de los datos. Mientras mayor es la variabilidad en la medida de desenlace, es más probable que los datos de los grupos se superpongan y, por lo tanto, será más difícil demostrar una diferencia entre ellos.¹⁶ Esto hace necesario incluir un número mayor de sujetos en el estudio.

En resumen, para el cálculo del tamaño de la muestra en primera instancia se debe plantear de manera adecuada una pregunta de investigación, a partir de la cual, se desprenderá la hipótesis de estudio donde se señala el tamaño del efecto elegido, así como la variabilidad de los datos; finalmente se fijan los valores de alfa y de beta.

En el caso concreto de los estudios de casos y controles se debe calcular el número requerido para contrastar la hipótesis de que la razón de momios es igual a 1. Para ello, es necesario conocer la probabilidad de exposición al factor de riesgo de los controles p_2 y asignar un valor r a la razón de momios que se pretende probar en el estudio. Con estos datos se puede calcular la frecuencia de exposición entre los casos p_1 :

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

Conociendo este valor se calcula el tamaño de la muestra con la siguiente fórmula:¹⁷

$$n = \frac{[Z_{1-\alpha/2} \sqrt{2p(1-p)} + Z_{1-\beta} \sqrt{p_1(1-p_1) + p_2(1-p_2)}]^2}{(P_1 - P_2)^2}$$

Donde:

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

Dado que los valores de $z_{1-\alpha/2}$ y $z_{1-\beta}$ se establecen de antemano y corresponden a 1.96 y 0.84, respectivamente, el cálculo de la muestra queda como:

$$n = \frac{[1.96 \sqrt{2p(1-p)} + 0.84 \sqrt{p_1(1-p_1) + p_2(1-p_2)}]^2}{(p_1 - p_2)^2}$$

Cuando se utiliza un número diferente de casos y controles, se designa n al número de casos, m al número de controles, $c = m/n$ es el número de controles por cada caso. Por último, el cálculo de la muestra se hace de la siguiente manera:

$$n = \frac{[1.96 \sqrt{(c+1)p(1-p)} + 0.84 \sqrt{cp_1(1-p_1) + p_2(1-p_2)}]^2}{c(p_1 - p_2)^2}$$

CONCLUSIONES

Dentro de los métodos observacionales utilizados para responder a preguntas clínicas, los *estudios de casos y controles* son muy eficientes porque requieren de pocos recursos, consumen poco tiempo y su costo es relativamente bajo. Están plenamente indicados cuando la variable de desenlace es poco frecuente. Una vez hechas estas consideraciones, el investigador decidirá si un estudio de casos y controles es el método adecuado para responder a su pregunta de investigación.



ABSTRACT

Case-control studies are suitable methods in clinical research if the outcome in study is uncommon or a long time is demanded. A sample must be selected from the target population and to identify clearly the cases, which are the subjects that suffer the clinical event or sickness, as well as the controls, who are the subjects who do not suffer the resulting event. A very important characteristic in these studies is that they start from the identification of cases and controls in present time but the information about exposure to the risk factor is obtained from the past time, usually from interviews, questionnaires or hospital records. The main weakness of this method of research is the way for obtaining information because it can be either not enough or inadequate, which produces uncertainty about its reliability. On the other hand, they can be used to investigate uncommon health problems or diseases with a long period between exposure to the risk factor and the start of the disease, the required time is short, the cost is low and a great amount of information is obtained from a few number of subjects. In this review, a practical way to perform control-case studies is presented as well as statistical basis to decide the truthfulness of the results.

KEYWORDS: *Methodology, case controls, clinical research.*

REFERENCIAS

1. Gómez C, Ardila J. Estudio de casos y controles. En: Ruiz H, Morillo L (eds.). *Epidemiología Clínica: investigación clínica aplicada*. México: Editorial Médica Panamericana; 2004, p. 265-78.
2. Cañedo L. Estudios longitudinales. En: *Investigación Clínica*. México: McGraw-Hill Interamericana; 1987, p. 51-70.
3. Greenberg R, Flanders D, Eley W, Daniels S, Boring J. Estudios de casos y controles. En: Greenberg R, Flanders D. (eds.). *Epidemiología Médica*. 3a Ed. México: El Manual Moderno; 2002, p. 139-53.
4. Grimes D, Schulz K. Compared to what? Finding controls for case-control studies. *Lancet* 2005; 365: 1429-33.
5. Mant J, Vessey M. Using case-control studies for prescribing research. *J Clin Phar Therap* 2002; 27: 67-74.
6. Lewallen S, Courtright P. Epidemiology in practice: case-control studies. *Community Eye Health* 1998; 11: 57-8.
7. Newman T, Browner W, Cummings S, Hulley S. Designing a new study: II. Cross-sectional and case-control studies. En: Hulley S, Cummings S. (eds.). *Designing clinical research* by Stephen B. Hulley, et al. 3rd Ed. Lippincott WW; 2007, p. 109-23.
8. Schulz K, Grimes D. Case-control studies: research in reverse. *Lancet* 2002; 359: 431-4.
9. Argimon J, Jiménez J. Estudios de casos y controles. En: Argimón J, Jiménez J (eds.). *Métodos de investigación clínica y epidemiológica*. 2a Ed. Madrid: Ediciones Harcourt; 2000, p. 51-64.
10. Mann C. Observational research methods. Research design II: cohort, cross sectional, and case-control studies. *Emerg Med J* 2003; 20: 54-60.
11. Grimes D, Schulz K. Bias and causal associations in observational research. *Lancet* 2002; 359: 248-52.
12. Shapiro E. Case-control studies. *Pediatr Infect Dis J* 2003; 22: 85-7.
13. Peipert J, Grimes D. The case-control study: A primer for the obstetrician-gynecologist. *Obstet Gynecol* 1994; 84: 140-5.
14. Fletcher R, Fletcher S, Wagner E. Estudio de casos. En: Fletcher R, Fletcher S (eds.).

- Epidemiología Clínica. 2a Ed. Masson; 2003, p. 217-36.
15. Posada M. La elección del tipo de diseño en los estudios de investigación clínica. Estudios de casos y controles. Neurol 2004; 19(Supl 1): 13-22.
 16. Browner W, Newman T, Hulley S. Getting ready to estimate sample size: Hypotheses and underlying principles. En: Hulley S. Designing clinical research. 3rd Ed. Lippincott WW; 2007, p. 51-63.
 17. Díaz P, Fernández P. Cálculo del tamaño muestral en estudios de casos y controles. Cad Aten Prim 2002; 9: 148-50.