

# Investigación factores de confusión

## Metodología y lectura crítica de estudios Factores de confusión

**Carlos Gómez-Restrepo<sup>1</sup>**  
**Mayumi Okuda Benavides<sup>2</sup>**

### Resumen

*Objetivo:* describir la definición de los factores de confusión, dar algunos ejemplos y herramientas para su determinación y sugerir estrategias para reducir la confusión dentro del diseño de estudios. *Metodología:* se hace una descripción corta de las expresiones *confusión* y *factores de confusión*, y se describen algunos ejemplos. Se explican algunos métodos para su consideración dentro de la realización de diferentes estudios y las formas de mejorar su diseño. *Conclusión:* la confusión debe ser evaluada para describir la verdadera relación entre las variables.

**Palabras clave:** confusión, sesgo, factor de riesgo.

**Title:** Medical Studies` Methodology and Critical Reading. Factors leading to Confusion.

### Abstract

*Objective:* To describe the definition of confounding factors, give some examples and tools for their determination, and to suggest strategies to reduce confusion in study designs. *Methods:* The terms *confusion* and *confounding factors* are briefly explained and some examples are given. Some of the methods used to assess them in the course of different studies and the ways to improve the design of these studies are also addressed. *Conclusion:* Confusion should be evaluated in order to describe the true relationship between variables.

**Key words:** Confusion, bias (epidemiology), risk factor.

### Introducción

Al hacer la lectura crítica de un estudio o durante el planteamiento del diseño de un protocolo, se debe

pensar en el papel que desempeña el azar y el sesgo en la valoración de la presencia de una asociación entre diferentes variables (1). Sin embargo, existe otra importante conside-

<sup>1</sup> Médico psiquiatra, psicoanalista, MSc. en Epidemiología Clínica, director del Departamento de Psiquiatría y Salud Mental, profesor de la Unidad de Epidemiología Clínica y Bioestadística, Facultad de Medicina, Pontificia Universidad Javeriana, Bogotá, Colombia.

<sup>2</sup> Médica, asistente de investigación, Facultad de Medicina, Unidad de Epidemiología Clínica y Bioestadística, Pontificia Universidad Javeriana Bogotá, Colombia.

ración en el diseño de un estudio y en el análisis de los resultados, la cual ha sido denominada *confusión*. A diferencia de las otras, ésta no es introducida por el investigador ni los participantes y se debe a la interrelación tan compleja que existe entre diferentes exposiciones y la enfermedad (1).

### **Definición**

La confusión puede ser vista como la mezcla del efecto de un factor de riesgo o protección (variable de exposición) con un tercer factor asociado a dicho factor y que a la vez es una variable de riesgo o protección para la enfermedad. Este factor puede llevar a sobreestimar o subestimar el efecto de los factores de exposición sobre el desenlace (enfermedad) (1).

El término *confusión* se refiere “al efecto de una variable externa que de manera completa o parcial es responsable del efecto aparente de la exposición analizada en un estudio, o la causante del enmascaramiento de una asociación subyacente verdadera” (2).

Cuando se habla de confusión en un experimento clínico, podemos decir que el cálculo del efecto de una intervención está sesgado, debido a la presencia de un tercer factor que induce una diferencia entre los grupos de comparación, más allá de las diferencias esperadas y planeadas dentro del estudio. De esta manera, para que un factor sea

causante de confusión, es necesario que difiera en los grupos de comparación y predecir el desenlace de interés (3-4). Los criterios que debe cumplir una característica para ser considerada una variable de confusión incluyen:

1. El factor de confusión debe estar asociado al factor de exposición.
2. Este no debe estar en la cadena de causalidad.
3. El factor de confusión debe estar relacionado independientemente con el desenlace en términos de riesgo o protección. (1)

De esta manera, podemos representar el factor de confusión así:

FC: factor de confusión.

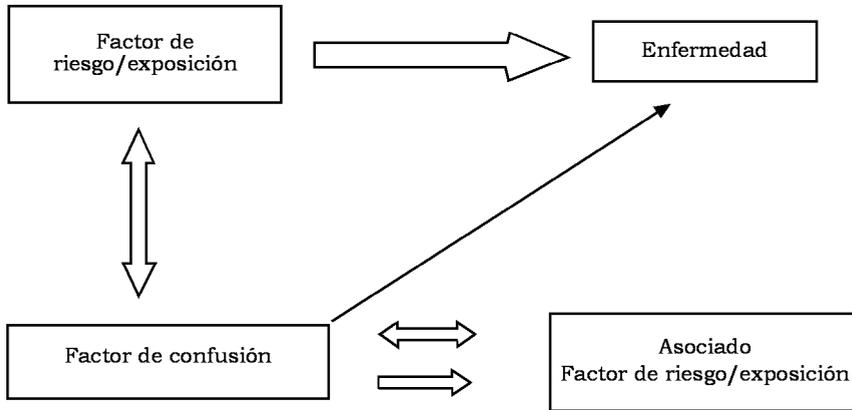
FR: factor de riesgo o protección.

E: enfermedad.

### **Evaluación de la presencia de factores de confusión**

Dentro del análisis del diseño de un estudio debemos preguntar, en primera instancia, si se ha hecho un esfuerzo sistemático para identificar y medir los posibles factores de confusión. Éstos deben ser identificados y medidos durante la fase de protocolo del estudio. En algunas ocasiones, debido a la falta de conocimiento del papel de algunas variables o factores sobre el curso de una patología, esta identificación puede ser difícil (5); sin embargo, en caso de duda, es preferible tomar las variables como si fueran de confusión, con el fin de

Gráfico 1. Factor de confusión



tenerlas en cuenta en el momento de la recolección. Así mismo, aun cuando la distribución de un factor de confusión desconocido puede darse de manera uniforme dentro de los grupos de comparación, los investigadores deben tratar de determinar qué tan susceptibles fueron los resultados a estos factores de confusión desconocidos (5).

### ¿Cómo reducir la confusión?

Debido a lo expuesto con anterioridad, se han diseñado diferentes formas para controlar los posibles factores de confusión. La primera opción consiste en la exclusión o restricción de las posibles variables de confusión dentro del diseño del estudio (5). Por ejemplo, si se están analizando concentraciones de acetilcolina (ACh) y su relación con el inicio de esquizofrenia, es importante conocer que las cantidades de ACh podrían estar asociadas con la

edad para aumentar o disminuir el riesgo. En este caso se puede determinar no incluir a personas mayores de cincuenta años de edad, esto es restringirlas, si se conoce que el comportamiento de las personas de este grupo de edad será muy diferente a los que se incluyen. Otra opción consiste en incluirlos dentro del análisis del estudio haciendo un “control matemático”, ya sea mediante análisis multivariados de regresión logística o la estratificación.

La regresión logística consiste en el cálculo de la forma como se relacionan los confusores y el desenlace. Esto nos permite hacer el cálculo ajustado del efecto de la intervención (6). Esta última se hace mediante el cálculo de la asociación de cada variable independiente con la variable dependiente, luego de ajustar por los efectos de todas las otras variables.

Por otro lado, la estratificación comprende el proceso donde la muestra se divide en subgrupos, a

partir de la presencia de las características que se consideran posibles confusores. De esta manera, cada subgrupo se “balancea”, en términos de los factores de confusión y cada subgrupo se analiza de manera independiente, apreciando el efecto de cada uno de éstos (6).

Por último, tenemos la opción de hacer un apareamiento (*matching*) con los posibles efectos indeseados que pueden resultar de su uso, ya que éste puede aumentar la dificultad en el análisis. El sobreapareamiento por variables como la edad y el sexo puede llevar a la disminución o pérdida en la capacidad de detección de diferencias cuando éstas existen o, por otro lado, a la magnificación de los sesgos y la reducción de la eficiencia del estudio. Debido a esto, la recomendación actual es la de hacer apareamiento cuando el factor de confusión es un factor de riesgo muy grande para la enfermedad (OR mayores a 7), cuando la exposición es muy rara o en casos de tamaños de muestra pequeños (7).

En consecuencia, el buen diseño de un estudio no necesariamente implica la completa anulación de los posibles factores de confusión dentro de su realización ni mucho menos dentro del análisis de sus resultados, sino la apreciación de su existencia y la toma de decisiones que lleve a un análisis del impacto o influencia en los resultados.

### Algunos ejemplos del impacto de la confusión

Un investigador piensa que las concentraciones de colesterol son un factor de riesgo para la presencia de depresión. Para ello plantea un estudio en que desea evidenciar la relación de las concentraciones de colesterol y el trastorno depresivo mayor (DM), y realiza un estudio de casos y controles en el cual se encuentran los siguientes datos:

Tabla 1. Estudio hipotético de asociación de colesterol con DM

| Colesterol | DM |    |
|------------|----|----|
|            | +  | -  |
| Alto       | 71 | 52 |
| Bajo       | 29 | 48 |

OR: 2,26

En este ejemplo, el riesgo relativo indirecto (OR crudo) muestra una asociación de colesterol alto con depresión mayor (2,26 más riesgo). No obstante, puede pensarse que esta relación esté confundida por un tercer factor, entre otros, la edad, puesto que conocemos que está asociada con las concentraciones de colesterol y, a la vez, es un factor de riesgo para depresión. De esta manera, si estratificamos con el fin de controlar la confusión, de acuerdo con la edad dividiéndolos en mayores de 45 años y menores de esta edad, tendríamos la siguiente tabla:

Tabla 2. Estudio hipotético de asociación de colesterol con DM estratificado por edad

| Edad               | Concentración de colesterol | DM (+) | DM (-) | OR    |
|--------------------|-----------------------------|--------|--------|-------|
| Mayores 45 años    | Alto                        | 63     | 36     | OR: 1 |
|                    | Bajo                        | 7      | 4      |       |
| Menores de 45 años | Alto                        | 8      | 16     | OR: 1 |
|                    | Bajo                        | 22     | 44     |       |

Con la estratificación por edad, observamos que la asociación se pierde y pasa de 2,26 a 1 (no diferencia entre grupos). De lo anterior podemos concluir que la relación de colesterol y DM se confunde por la edad y que realmente no existe la relación entre el colesterol alto o bajo y la presencia o no de depresión, que se evidenciaba con el OR crudo. Esta situación, igualmente, nos permite ilustrar la paradoja de Simpson, donde la presencia de diferencias

inexistentes en el OR crudo se ven diluidas a 1 al realizarse la estratificación.

Otro ejemplo hipotético sería uno en el cual se pretende evidenciar la influencia del patrón de relaciones tempranas madre-hijo (RT-MH), calificado como buenas o malas relaciones, y el padecer durante la adultez de un trastorno hipocondríaco (TH). En este estudio hipotéticamente encontramos estas asociaciones:

Tabla 3. Asociación hipotética de relaciones tempranas madre-hijo (RT-MH) y trastorno hipocondríaco (TH)

| RT-MH \ TH | TH |     |
|------------|----|-----|
|            | +  | -   |
| Buenas     | 34 | 110 |
| Malas      | 48 | 91  |

OR: 0,59

Con este resultado crudo podríamos concluir que las buenas relaciones protegen contra el TH. No obstante, podemos pensar que la edad puede ser un factor de confusión, asociado con el tipo de relación que se tuvo, y que sea, además, factor de riesgo o protección para padecer un TH. De hecho, se conoce que en algunos rangos de edad se presenta con mayor frecuencia el TH.

Por este motivo el investigador debería estratificar por edad, de alguna manera. Con el ánimo de explicar lo podríamos dividir en: menores de 40 años y mayores de 40 años. Con la hipótesis de que los mayores presentaron crianzas diferentes a las que recibieron las personas de menor edad y que probablemente generaron vínculos distintos.

Tabla 4. Asociación hipotética de RT-MH y TH estratificado por edad

| Edad               | RT-MH  | TH (+) | TH (-) | OR       |
|--------------------|--------|--------|--------|----------|
| Menores de 40 años | Buenas | 30     | 12     | OR: 1,7  |
|                    | Malas  | 25     | 17     |          |
| Mayores de 40 años | Buenas | 4      | 98     | OR: 0,73 |
|                    | Malas  | 23     | 74     |          |

En este ejemplo se observa que el OR crudo se mantiene como un factor protector para los mayores de 40 años, y se convierte en factor de riesgo en los menores de tal edad. Esta relación no habría sido identificada si no se hubiera pensado en que la edad pudiera ser un factor de confusión y la necesidad de estratificarlo.

### Conclusión

Al leer o al obtener los resultados de un estudio, se debe considerar la posibilidad de que éstos sean explicados por la confusión. Debido a esto, es importante tener claro el concepto de *factor de confusión* y es necesario usar los métodos disponibles para un buen diseño y análisis de resultados. La valoración de la existencia de factores de confusión y los métodos utilizados en un estudio cuando nos encontramos ante la presencia de estas variables nos guiará en la descripción y comprobación de una verdadera relación estadística válida entre las variables que se van a estudiar.

### Referencias

1. Hennekens CH, Buring JE. Analysis of epidemiologic studies: evaluating the role of confounding. In: Epidemiology in medicine. Boston: Little Brown and Company; 1987.
2. Schlesselman JJ. Case-control studies: design, conduct, analysis. Oxford: Oxford University Press; 1982.
3. Rochon PA, Gurwitz JH, Sykora K, Mamdani M, Streiner DL, Garfinkel S, et al. Reader's guide to critical appraisal of cohort studies: 1. Role and design. *BMJ*. 2005;330:895-7.
4. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al. The revised CONSORT statement for reporting randomized controlled trials: explanation and collaboration. *Ann Intern Med*. 2001;134:663-94.
5. Mamdani M, Sykora K, Li P, Normand SL, Streiner DL, Austin PC, et al. Reader's guide to critical appraisal of cohort studies: 2. Assessing potential for confounding. *BMJ*. 2005;330:960-2.
6. Normand SL, Sykora K, Li P, Mamdani M, Rochon PA, Anderson GM. Reader's guide to critical appraisal of cohort studies: 3. Analytical strategies to reduce confounding. *BMJ*. 2005;330:1021-3.
7. Ruiz A, Morillo LE, editores. Epidemiología clínica: investigación clínica aplicada. México: Editorial Médica Internacional; 2004. Capítulo 15: estudios de casos y controles.

*Recibido para evaluación:* 10 de marzo de 2006  
*Aprobado para publicación:* 17 de mayo de 2006

*Correspondencia*  
*Carlos Gómez-Restrepo*  
*Departamento de Psiquiatría y Salud Mental*  
*Hospital Universitario de San Ignacio*  
*Carrera 7 N° 40-62, piso 8*  
*Bogotá, Colombia*  
*Correo electrónico: cgomez@javeriana.edu.co*