

**[5]**

**Sesgos en las  
observaciones clínicas**

**Francisco Javier Carrera Hueso  
Piedad López Sánchez**

Uno de los mayores problemas en investigación es la ocurrencia de errores, que pueden darse de forma aleatoria o sistemática. Existe sesgo cuando se comete un error de forma sistemática, por causas relativas al propio observador, al fenómeno observado o a aquello con lo que se observa. Este capítulo trata de cómo minimizar o evitar los sesgos.



# [5]

## Introducción

Como titula el clásico libro sobre seguridad sanitaria, errar es humano, y la medición es una actividad humana sometida a error. En efecto, ya sabemos que se debe tener muy presente la precisión con una simple pesada en cualquier laboratorio, es decir, que la medida real estará dentro de un intervalo, con una confianza del 95% (normativa admitida en el ámbito sanitario), y es debido al potencial error que cometamos. Por tanto, una medida siempre contendrá dos partes, la medida real y el error que se haya cometido<sup>1</sup>. Se debe recordar que no es lo mismo utilizar un instrumento de medida u otro, como por ejemplo una balanza de precisión.

Cuando medimos sustancias fisiológicas o fármacos en una persona (orina, suero, sangre, etc.), estamos cometiendo un posible error, tanto en el procedimiento analítico utilizado como en la propia medición<sup>2</sup>. Cuestión que se complica, y mucho, al medir condiciones patológicas, síntomas o síndromes de los pacientes de manera directa, ya que esta medición se ve influida además por muchos otros factores del propio paciente y del observador que mide. Este sería el caso del registro de la presión arterial de un paciente, en el que no solo influye el instrumento utilizado (tensiómetro calibrado y su uso correcto), sino también las condiciones en las que se realiza (posición relajada del paciente al menos 5 minutos antes de medir) y del propio observador que la registra (cuán avezado sea para esta labor y cómo la lleva a cabo: sin redondear medidas, etc.). Todo ello se puede incluso complicar si se desea conocer la calidad de vida relacionada con la salud de los pacientes, ya que se debe definir claramente cómo se realiza la medida, a quiénes, cuándo y por qué, quién lo realiza y cómo se realiza, entre otras cuestiones.

Queda claro, pues, que la medición de las observaciones clínicas está sujeta a muchos y múltiples errores, que además son de naturaleza muy dispar. En este capítulo pretendemos orientar a los farmacéuticos de hospital en formación sobre cómo realizar medidas exentas de sesgos y permitirles ser los agentes sanitarios que avalen, con su contribución profesional, las medidas que se registran tanto en investigación como en la asistencia sanitaria de los equipos

multidisciplinares que atienden a los pacientes. En definitiva, que sean garantías de la calidad del registro de los datos y su gestión posterior, para llegar a resultados fiables y válidos que nos acerquen más a la verdad.

## Fiabilidad y validez

Los errores en una medición pueden ser aleatorios o sistemáticos (llamados sesgos). Un error aleatorio es un error que no es constante en cada medición y no ocurre siempre en la misma dirección<sup>2</sup>. Sin embargo, el error sistemático siempre se produce en la misma dirección; por ejemplo, una balanza que produce un peso de +5 gramos. En la figura 1 se muestra el clásico ejemplo del tiro a una diana como símil de la precisión y exactitud de una medición. Como puede comprobarse, las dianas 1 y 2 presentan un error sistemático en la misma dirección mientras que en las dianas 3 y 4 el error es aleatorio, no podemos saber en qué dirección se produce.

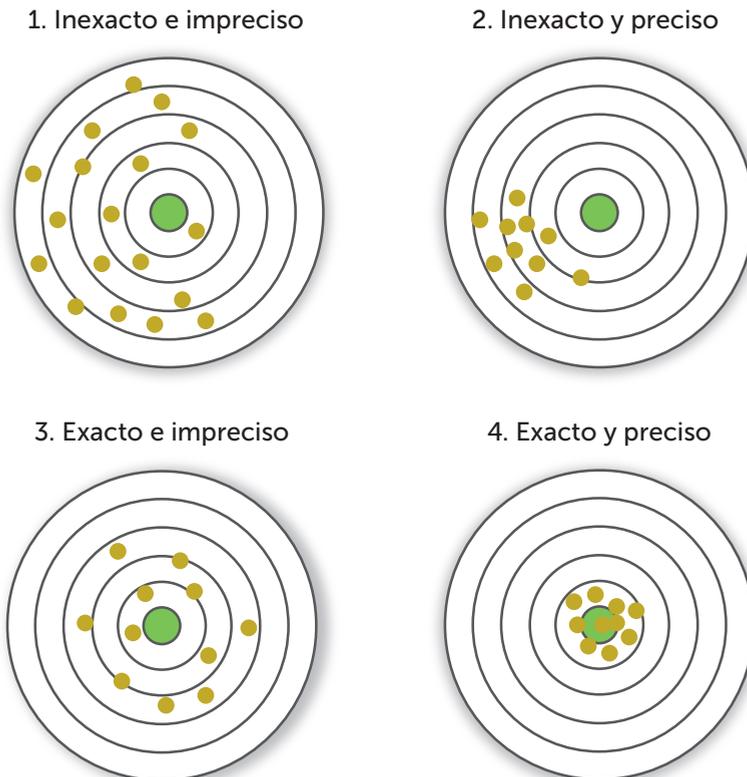


Figura 1. Precisión y exactitud.

La precisión o *fiabilidad* de una medición valora la reproducibilidad de esa medición, es decir, la capacidad de poder obtener un mismo valor cuando una medición se realiza sobre la misma persona en más de una ocasión y en condiciones similares. Se habla de *reproducibilidad* tanto si utilizamos un instrumento de medida varias veces en una misma persona como si valoramos medidas realizadas en una misma persona pero por investigadores diferentes. Se analiza esta fiabilidad con el estadístico conocido como índice kappa (recomendamos su estudio en textos especializados).

La precisión se puede valorar con los intervalos de confianza de una estimación, por lo que, a mayor precisión, menor rango entre los límites del intervalo de confianza. Habrá precisión o mayor fiabilidad de una medición cuando los errores aleatorios sean pequeños. Volviendo al ejemplo de la figura 1, comprobamos que las mediciones de las dianas 1 y 4 son precisas por su bajo grado de dispersión.

El error aleatorio no afecta en sí a la validez interna, que veremos a continuación, pero sí dificulta la posibilidad de encontrar una asociación entre dos variables por la imprecisión de las medidas.

Los *errores sistemáticos* se caracterizan por:

- Afectar más a la validez que a la precisión del estudio.
- Desviar las estimaciones de la asociación en una sola dirección, pero varios errores se pueden potenciar o contrarrestar.
- Ser generalmente predecibles, dando una subestimación o sobreestimación del efecto.
- Ser evitables mejorando el diseño y/o análisis de los datos.

La *validez* de una medición consiste en su capacidad de medir lo que realmente se quiere medir. En la figura 1 la media de las mediciones de las dianas 1 y 2 no se aleja del verdadero valor, que es el centro de la diana, y, por tanto, las mediciones no tendrán validez. Sin embargo, el valor promedio de las dianas 3 y 4 está muy próximo al verdadero valor y, por tanto, tales mediciones serán válidas. La validez puede ser interna y externa.

La *validez interna* de una investigación valora el grado en que los resultados obtenidos son correctos para las personas estudiadas. Se obtiene evitando errores o sesgos en el diseño del estudio pero también en el análisis de los datos, como puede suceder si se utiliza un método estadístico erróneo que da lugar a un sesgo de mala especificación. Los instrumentos que se usan, la ma-

nera en que se mide y el tiempo de seguimiento correcto de los participantes son claves para evaluar la validez interna de un estudio científico<sup>2</sup>.

La *validez externa* de un estudio es la capacidad de poder generalizar los resultados de nuestra muestra a la población de referencia o a aquella a la que se pretenden aplicar las conclusiones del estudio. La muestra, por tanto, debe ser representativa, y la medición del efecto ha de realizarse durante un seguimiento suficiente de los pacientes. Aquí entran en juego las pérdidas del seguimiento realizado, además de la comparación basal de nuestros pacientes, edad, sexo, gravedad, etc.<sup>2</sup>

Un estudio sin validez interna nunca puede tener validez externa. Los errores sistemáticos (dianas 1 y 2 de la figura 1) afectan a la validez interna del estudio y, por tanto, indirectamente también a su validez externa<sup>2</sup>.

## Sesgos

### Definición

Se define sesgo, o error sistemático, como cualquier error diferencial, en relación con los grupos que se comparan, en que se puede incurrir durante el diseño, conducción o análisis del estudio y que invariablemente resulta en una conclusión errónea, ya sea por una estimación más alta o más baja del valor real<sup>3</sup>.

### Clasificación

Una forma simple de comprender las distintas posibilidades de cometer un sesgo durante la investigación es pensar en los tres ejes que dominan una investigación: *lo que* se observará o medirá, es decir, la variable en estudio; *el que* observará o medirá, es decir, el observador; y *con lo que* se observará o medirá.

En la tabla 1 se recogen los tipos de sesgos más comunes en los distintos tipos de estudios observacionales<sup>4</sup>.

Los sesgos se pueden clasificar según la frecuencia en que se presentan y la etapa del estudio en que se originan, es decir, el sesgo en las observaciones clínicas, que se puede clasificar en tres categorías<sup>3</sup>: sesgo de selección, sesgo de información y sesgo de confusión.

**Tabla 1. Tipos de sesgos más comunes en estudios observacionales**

Posibilidad de	Cohortes	Casos y controles	Corte transversal	Estudios ecológicos
Sesgo de selección	Baja	Alta	Media	No aplica
Sesgo de memoria	Baja	Alta	Alta	No aplica
Sesgo de confusión	Baja	Media	Media	Alta
Pérdidas de seguimiento	Alta	Baja	No aplica	No aplica
Tiempo necesario	Elevada	Media	Media	Baja
Coste	Alta	Media	Media	Baja

Adaptada de Manterola y Otzen T, 2015<sup>4</sup>.

### **Sesgos de selección**

Este tipo de sesgo se genera durante la selección o el seguimiento de la población de estudio. Ocurre cuando hay un error sistemático en los procedimientos utilizados para seleccionar a los sujetos del estudio. Por tanto, conduce a una estimación del efecto diferente del obtenible para la población diana. Los grupos de pacientes a menudo difieren en muchos aspectos: edad, sexo, gravedad de la enfermedad, etc. Si comparamos la experiencia de los dos grupos que difieren respecto a una característica de especial interés (por ejemplo, un tratamiento o una causa posible de enfermedad), pero también son diferentes en otros aspectos y estas diferencias guardan relación con el resultado, la comparación es sesgada y apenas pueden extraerse conclusiones sobre los efectos independientes de la característica de interés. En el ejemplo utilizado, se habría producido un sesgo de selección si los pacientes que recibieron el tratamiento A gozaran de mejor salud que los pacientes que recibieron el tratamiento B.

Los sesgos de selección pueden ocurrir en cualquier tipo de diseño de estudio, sin embargo se dan con mayor frecuencia en series de casos retrospectivas, estudios de casos y controles, de corte transversal y de aplicación de encuestas. Este tipo de sesgo impide extrapolar las conclusiones en estudios realizados con voluntarios extraídos de una población sin el evento de interés.

Un ejemplo de esta situación es el denominado *sesgo de Berkson*, llamado también *falacia o paradoja de Berkson*, o *sesgo de admisión o de diagnóstico*. Se define como el conjunto de factores selectivos que conducen a diferencias sistemáticas que se pueden generar en un estudio de casos y controles con

casos hospitalarios<sup>5</sup>. Ocurre en aquellas situaciones en las que la combinación entre una exposición y el evento de interés en estudio incrementa el riesgo de ingreso a un centro hospitalario, lo que conduce a una tasa de exposición sistemáticamente más elevada entre los casos hospitalarios respecto de los controles. Un ejemplo es el caso de la asociación negativa entre cáncer y tuberculosis pulmonar, en el que la tuberculosis actuaba como factor de protección para el desarrollo de cáncer, lo que se explicó por la baja frecuencia de tuberculosis en los hospitalizados por cáncer, hecho que no significa que entre estos sujetos la frecuencia de la enfermedad sea menor.

Otro subtipo de sesgo de selección es el denominado *sesgo de Neymann (de prevalencia o de incidencia)*, que se produce cuando la condición en estudio determina la pérdida prematura por fallecimiento de los sujetos afectados por ella. Por ejemplo, si en un grupo de 1.000 sujetos con hipertensión arterial (factor de riesgo de infarto de miocardio) y 1.000 no hipertensos, seguidos durante 10 años, se observa una intensa asociación entre hipertensión arterial e infarto de miocardio, y, sin embargo, puede ocurrir que no se llegue a obtener esa asociación debido a la no incorporación en el análisis de los sujetos que fallecen por infarto de miocardio durante el seguimiento.

### **Sesgo de información o de medición**

Este tipo de sesgo ocurre cuando se produce un defecto al medir la exposición o la evolución que genera información diferente entre los grupos en estudio que se comparan (precisión). Se produce cuando los métodos de determinación son consistentemente distintos en grupos diferentes de pacientes. En la práctica puede presentarse como la clasificación incorrecta de sujetos, variables o atributos, dentro de una categoría distinta de aquella a la que deberían haberse asignado.

Las probabilidades de clasificación incorrecta pueden ser las mismas en todos los grupos en estudio, y entonces se denomina *clasificación incorrecta no diferencial* (el grado de clasificación incorrecta es similar en los grupos en estudio), o puede variar entre estos, lo que se denomina *clasificación incorrecta diferencial* (el grado de clasificación incorrecta es mayor en un grupo que en otro). Un subtipo de esta última opción es el *sesgo de recuerdo o de memoria (recall bias)*, que ocurre fundamentalmente en estudios de carácter retrospectivo en los que se estudia como posible exposición antecedentes de circunstancias acaecidas en etapas previas de la vida, en las que existe la posibilidad de olvido, en especial si las mediciones son de alta variabilidad. Un ejemplo de un potencial sesgo de medición sería el de un estudio que comparase la frecuen-

cia de la utilización de los anticonceptivos orales entre mujeres ingresadas en un hospital debido a una tromboflebitis y un grupo de mujeres ingresadas por otras razones. Es perfectamente posible que las mujeres con tromboflebitis conocieran la asociación documentada entre estrógenos y acontecimientos trombóticos, y pudieran indicar la utilización de anticonceptivos orales de forma más completa que las mujeres sin tromboflebitis porque ya habrían oído hablar de la asociación. Por las mismas razones, los clínicos podrían obtener y registrar información sobre la utilización de anticonceptivos orales de forma más completa para las mujeres con tromboflebitis que para las mujeres que no la experimentan. En este caso, podría observarse una asociación entre los anticonceptivos orales y la tromboflebitis debido a la forma en la cual se mencionó la historia de exposición y no porque exista realmente una asociación.

### ***Sesgo de confusión o mezcla de efectos***

Se produce cuando dos factores están asociados entre sí o “viajan juntos” y el efecto de uno está confundido o distorsionado por el efecto del otro. Esto podría producirse debido al sesgo de selección, por azar o porque en realidad ambos se asocian por naturaleza. El sesgo de selección y el de confusión no son mutuamente excluyentes, pero se describen por separado porque presentan problemas en diferentes momentos de una observación o estudio clínico. El sesgo de selección debe considerarse principalmente cuando los pacientes son escogidos para una observación y, por lo tanto, es importante en el diseño de un estudio. El sesgo de confusión debe abordarse durante el análisis de los datos, una vez que se han hecho las observaciones. Más adelante se describe detenidamente este sesgo.

Durante el curso de una investigación los sesgos pueden clasificarse también en función de la etapa del estudio en la que ocurren: sesgos en la *etapa de planificación*, sesgos en la *etapa de muestreo y seguimiento*, o sesgos en la *etapa final del estudio*, que se encuentran relacionados con la *medición*, el *análisis*, la *interpretación* y su *comunicación*<sup>4</sup> (Tabla 2).

### ***Sesgos durante la planificación del estudio***

Estos pueden relacionarse con la fuente de financiación, la información previa disponible y la planificación del estudio propiamente:

- *Sesgos presupuestarios*: debidos a la influencia de la financiación sobre el proyecto; supone que en algún caso puede concederse excesiva importancia a ciertos aspectos y obviarse otros que pueden ser relevantes.

**Tabla 2. Momentos de ocurrencia de sesgos en las distintas etapas de la investigación y estrategias para evitarlos**

Etapa del estudio	Estrategias para evitarlos
Planificación	<ul style="list-style-type: none"> <li>• Utilización de protocolos rigurosos</li> <li>• Definición clara, precisa y concisa de objetivos</li> <li>• Empleo de información relacionada con la investigación obtenida a partir de clínicos, epidemiólogos, estadísticos</li> </ul>
Comienzo	<ul style="list-style-type: none"> <li>• Determinar sesgos de muestreo y evitar los evitables</li> <li>• Previsión de abandonos y minimización de pérdidas de seguimiento</li> <li>• Asignación aleatoria de los grupos en estudio</li> <li>• Enmascaramiento de las mediciones</li> <li>• Utilización de instrumentos de medición válidos y confiables</li> <li>• Entrenamiento de los investigadores</li> </ul>
Desarrollo	<ul style="list-style-type: none"> <li>• Detección de incoherencias</li> </ul>
Finalización	<ul style="list-style-type: none"> <li>• Corrección estadística de los sesgos</li> <li>• Interpretación de resultados en función de los sesgos cometidos</li> </ul>

Adaptada de Manterola y Otzen T, 2015<sup>4</sup>.

- *Sesgos de evaluación inicial del proyecto:* se deben a la existencia de prejuicios o datos erróneos que condicionan el planteamiento de la investigación.
- *Sesgos de concepto:* se cometen cuando no se tienen en consideración ciertas variables que pueden actuar como factores de confusión, o cuando la duración del estudio es inadecuada.

### **Sesgos durante el proceso de muestreo**

Se deben a la falta de representatividad de la muestra, es decir, a que la muestra no es una reproducción correcta de la población. Pueden deberse a que la población diana es distinta de la población a la que se pretenden inferir o extrapolar los resultados:

- *Sesgos debidos a falta de representatividad de la población:* a menudo existen diferencias entre la población que se quiere analizar y la población diana; por ejemplo, para estudiar la prevalencia de colestasis en la población general se ha de realizar una técnica de muestreo que permita obtener una muestra representativa de la población general y no estudiar una muestra de un hospital.
- *Sesgos debidos a falta de representatividad de la muestra:* una vez se tiene definida la población diana, se debe realizar el muestreo, para lo que existen estrategias probabilísticas y no probabilísticas, y su elección dependerá de una serie de hechos. Sin embargo, es obvio que un muestreo probabilístico

puede ayudar en la eliminación de posibles sesgos. En los muestreos no probabilísticos se debe considerar que prácticamente siempre suponen el establecimiento de sesgos, pues en la mayoría de los casos la muestra analizada se elige en función de aspectos de accesibilidad y posibilidades de colaboración, por lo que puede no ser representativa de la población.

### **Sesgos durante la recolección de datos**

Se producen durante el proceso de la recogida de la información, ya sea por la obtención de información incompleta o errónea o por la modificación de la muestra durante la ejecución de la investigación:

- *Sesgos debidos al encuestado*: la información que este proporciona puede ser incorrecta bien por olvido, subjetividad, confusión, desconfianza, ignorancia, incomprensión o bien por modificación de la respuesta por la propia encuesta o medición incorrecta de parámetros.
- *Sesgo por olvido*: el factor tiempo es un aspecto importante y que afecta de manera distinta a los distintos acontecimientos: los problemas agudos (por ejemplo, una gripe) en general se recuerdan mejor que los problemas subagudos.
- *Sesgo por subjetividad*: independiente del sesgo de olvido, podemos obtener respuestas que no se ajusten a la realidad cuando una pregunta se acota a un periodo de tiempo: "¿Cuántos resfriados ha tenido en el último año?". En caso que no existan registros, o que estos sean de mala calidad, obtendremos una respuesta aproximada que puede reflejar "más o menos" lo acontecido en el periodo en estudio.
- *Sesgo por confusión e ignorancia*: ocurre cuando se confunde el rol de ciertas variables, exposiciones o eventos de interés. Este fenómeno puede producirse por ignorancia o por falta de previsión por parte del investigador.
- *Sesgo de medición*: se genera por la elección incorrecta del instrumento de medición o por estimaciones subjetivas de la medición.
- *Sesgo de abandono*: puede ocurrir en el curso de estudios longitudinales, ya sea por abandono del estudio o por desaparición del individuo (muerte, cambio de ciudad, etc.).

### **Sesgos durante la etapa de análisis e interpretación**

Una vez que se dispone de los datos recopilados, se procede a su análisis. En esta fase pueden ocurrir errores sistemáticos por transcripción incorrecta de la información a la base de datos. También se pueden cometer errores en los métodos estadísticos empleados, que pueden ser inadecuados para los datos

analizados. Finalmente, hay que tener en cuenta las interpretaciones erróneas de los resultados, por ejemplo, por la presencia de variables de confusión no contempladas.

### **Sesgo de publicación**

Se puede considerar un tipo de sesgo de selección que ocurre cuando el investigador piensa que los estudios publicados son todos los realmente realizados. Es sabido que muchos estudios nunca llegan a ser publicados por diversas razones (no se concluyen, el autor considera que los resultados son irrelevantes, no son aceptados para publicación, etc.).

### **Criterios de causalidad**

Todos sabemos que los niños no vienen de París ni los traen las cigüeñas, pero este dicho sobre la causalidad de los nacimientos producidos en una población viene definido por una asociación espuria de las variables que intervienen, como se muestra en la figura 2<sup>2</sup>. Se comprueba que el número de nacimientos está correlacionado tanto con el número de cigüeñas como con el número de campanarios, y este a su vez con el número de habitantes de la población que se estudia. El grafo también nos indica que el número de cigüeñas y el número de nacimientos son variables condicionalmente independientes dado el número de habitantes o el número de iglesias<sup>6</sup>.

Como hemos mostrado en este ejemplo, la causalidad implica correlación entre variables, pero la correlación no implica causalidad, conceptos muy distintos y que no debemos olvidar a la hora de analizar un estudio epidemiológico.

Los principales criterios para conocer la causalidad de un efecto fueron definidos por Hill en 1965. Con el tiempo se han ido añadiendo más, e incluso se siguen discutiendo. Se admiten en la actualidad los representados en las páginas 98 y 99.

Ninguno de los criterios relacionados, excepto el de precedencia temporal, es esencial para definir una relación causal o para definir un factor de riesgo. Tan solo dos hechos son fundamentales: la existencia de *asociación* (no debida a errores) y la *ausencia de ambigüedad temporal*.

Finalmente, conviene no confundir un factor de riesgo con un factor pronóstico. Un *factor de riesgo* de una enfermedad es toda variable que aumenta



**Figura 2. Grafo dirigido acíclico de causalidad del número de nacimientos.**

Adaptada de Díez Vegas, 2005<sup>2</sup>.

la probabilidad de que se presente la enfermedad. Sin embargo, un *factor pronóstico* de una enfermedad es un factor cuya presencia o ausencia indica peor o mejor evolución de la enfermedad. Los factores pronósticos pueden ser *explicativos*: cuando su presencia produce un cambio en el pronóstico (relación causal); y *predictivos*: cuando la presencia del factor solo indica un cambio en el pronóstico (relación no causal)<sup>6</sup>.

## Confusión

La confusión es la intromisión de una o más variables entre una exposición o causa y su desenlace o efecto (Figura 3), lo que provoca que el efecto de la causa se confunda con el agente que se quiere estudiar. Por tanto, una variable o factor de confusión se puede definir como la que distorsiona la medida de asociación entre otras dos variables. Su resultado puede ser la observación de un efecto que no existe en realidad o la exageración de una asociación que sí existe (confusión positiva), o la atenuación de la asociación real o incluso un cambio de sentido de la misma (confusión negativa)<sup>6</sup>.

## CRITERIOS DE CAUSALIDAD

1

*Fuerza de asociación medida por la magnitud del riesgo relativo ( $RR \neq 1$ ). Cuanto mayor es la fuerza de la asociación, más probable es que esta refleje una verdadera relación de causalidad, de modo que los posibles sesgos pueden haber influido sobre la cuantía de la asociación, pero es menos probable que hayan influido sobre su dirección.*

2

*Consistencia de la asociación en distintas personas, lugares, circunstancias, etc. Aquí desempeña un papel esencial el análisis estratificado de nuestro estudio y de otros anteriores. Cuando se desea confirmar una hipótesis, se debería dar prioridad a los resultados similares obtenidos en estudios de diseño diferente.*

3

*Especificidad entre una causa y un efecto: relación biunívoca exclusiva. Aunque no siempre es así, ya que, por ejemplo, el hábito tabáquico puede ser causante no solo del cáncer de pulmón sino de otros tumores, o de bronconeumonía obstructiva crónica y cardiopatía isquémica.*

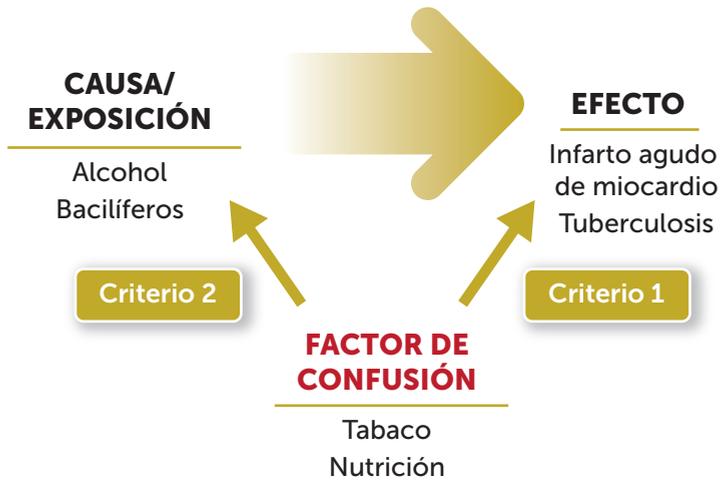
4

*Precedencia temporal de la causa, es decir, que la causa preceda al efecto. Aunque este criterio parezca obvio, no siempre lo es, sobre todo cuando se estudian enfermedades crónicas.*

5

*Gradiente biológico o gradiente dosis-respuesta. Aquí es donde intervendrían los factores de confusión que lo suelen incumplir. La frecuencia o la gravedad de la enfermedad en estudio debe guardar relación con la dosis o con la duración de la exposición al factor de riesgo considerado. Aunque para la mayoría de las relaciones de causalidad esta consideración puede ser válida, hay que recordar que la frecuencia o la intensidad de algunos efectos (por ejemplo, reacciones de hipersensibilidad alérgica) no guardan relación con la dosis administrada.*





**Figura 3. Factor de confusión.**

Adaptada de De Irala, 2008<sup>6</sup> y Rothman, 2002<sup>7</sup>.

El fenómeno o sesgo de la confusión, a diferencia del resto de sesgos, es el resultado de relaciones específicas entre variables consideradas en un determinado estudio que no tiene por qué ser el resultado de errores cometidos por parte del investigador o de problemas en la valoración de los participantes o pacientes (revisense los apartados de sesgos de selección y de información).

Veámoslo mejor con el siguiente ejemplo sobre la asociación del consumo de alcohol y los infartos de miocardio, cuyos resultados se exponen en la tabla 3. La razón de tasas obtenida en toda la población es de 2,2, por lo que la probabilidad de sufrir un infarto de miocardio es más del doble si se consume alcohol moderadamente. Sin embargo, es sabido que las personas que consumen alcohol fuman más que las que no lo hacen, es decir, ambas variables están asociadas, y esta asociación podría llegar a confundir o distorsionar. Si estratificamos por la variable, en fumadores y no fumadores, comprobamos que la asociación cambia incluso de sentido, pasando a ser 0,8 en ambos estratos. Por tanto, el consumo moderado de alcohol protege de sufrir un infarto de miocardio, tanto en los fumadores como en los no fumadores de este estudio. Esto muestra que el tabaco es un factor de confusión que debemos tener en cuenta para dar una correcta conclusión en nuestro trabajo.

Podemos comprobar, en la tabla 3, que el tabaco está asociado a una mayor tasa de infarto de miocardio (criterio 1 de la figura 3). En efecto, las razones de tasas (IR) de infarto de miocardio entre fumadores y no fumadores son ma-

**Tabla 3. Infarto de miocardio asociado al consumo moderado de alcohol por grupos de tabaquismo**

Infarto agudo de miocardio			Casos	Personas año	Tasa (por 1.000)	IR
Global	Consumo moderado de alcohol	Sí	187	39.600	4,7	2,2
		No	42	20.400	2,1	
Fumadores	Consumo moderado de alcohol	Sí	173	21.600	8	0,8
		No	24	2.400	10	
No fumadores	Consumo moderado de alcohol	Sí	14	18.000	0,8	0,8
		No	18	18.000	1	
Comprobación del <b>criterio 1</b> (factor de confusión con efecto)						
Beben alcohol	Fumadores	Sí	173	21.600	8	10
		No	14	18.000	0,8	
No beben alcohol	Fumadores	Sí	24	2.400	10	10
		No	18	18.000	1	
Comprobación del <b>criterio 2</b> (factor de confusión con causa o exposición)						
			Fumadores		No fumadores	
Consumo moderado de alcohol			21.600 (90%)		18.000 (50%)	
No beben			2.400 (10%)		18.000 (50%)	
Total			24.000 (100%)		36.000 (100%)	

IR: razón de tasas.

Datos recopilados de De Irala, 2008<sup>6</sup>.

yores que el valor nulo (IR=10) tanto en consumidores moderados de alcohol como en los no bebedores. Por otra parte, el hábito tabáquico no se distribuye de manera homogénea en ambos grupos de consumo de alcohol (criterio 2 de la figura 3), ya que, como podemos comprobar en la tabla, el 90% de los bebedores son además fumadores, mientras que en el grupo de no bebedores las personas que fuman descienden al 50%<sup>6</sup>.

En conclusión, al estimar la razón de tasas global del efecto protector del consumo moderado de alcohol sobre la incidencia del infarto de miocardio, este es contrarrestado por el mayor riesgo asociado al tabaco, ya que existe una mayor prevalencia del hábito tabáquico en las personas que consumen alcohol.

El resultado global mostrado en la tabla 3 (IR=2,2) hace referencia a la razón de tasas de incidencias *cruda o bruta* porque se obtiene sin tener en cuenta

la variable "fumar" de las personas estudiadas. Por el contrario, las estimaciones de las razones de tasas obtenidas en cada estrato de la tabla 3 (IR=0,8) se denominan estimaciones *ajustadas o controladas* por el hábito tabáquico. El ajuste por una variable de confusión consiste en la eliminación del efecto de distorsión que produce<sup>6</sup>.

Con este ejemplo facilitamos la comprensión de la definición de una variable de confusión, aspecto clave en epidemiología e investigación clínica. En concreto, para que una variable pueda considerarse un factor de confusión de la asociación entre una exposición o factor de riesgo y una respuesta o desenlace (Figura 3) tiene que cumplir tres condiciones o criterios<sup>7</sup>:

- **Criterio 1:** el factor de confusión (F) debe estar asociado con el desenlace o efecto (E), independientemente con la exposición de interés, es decir, con las personas no expuestas (nExp), y en los estudios de casos y controles con los controles. Además, debe producir efecto aunque la exposición no esté presente. Por tanto, con variables binarias, la medida que valora la asociación (razón de *odds*: OR, riesgo relativo: RR, o razón de tasas: IR) deberá ser distinta a 1:

$$OR_{F-E/nExp} \neq 1$$

La experiencia indica que los factores con OR (o IR) entre 0,67 y 1,5 se pueden descartar porque la asociación es muy débil. El criterio de  $p < 0,20$  tiene el inconveniente de estar afectado por el tamaño muestral, ya que si la muestra es muy grande, valores de  $p < 0,20$  pueden reflejar asociaciones muy débiles.

- **Criterio 2:** un factor de confusión (F) debe asociarse con la causa o exposición (C) pero no ser un resultado, un efecto de la misma.
  - En los estudios de cohortes, el criterio 2 con variables binarias es que la medida de asociación (OR) entre el factor de confusión (F) y la exposición sea distinta a 1:

$$OR_{F-Exp} \neq 1$$

- En los estudios de casos y controles, el criterio 2 con variables binarias es que la medida de asociación (OR) entre el factor de confusión (F) y la exposición en los controles (Co), ya que representan a la población, sea distinta a 1:

$$OR_{F-Ex/Co} \neq 1$$

- **Criterio 3:** el factor de confusión no debe ser un eslabón causal intermedio entre la exposición y el desenlace, es decir, no debe ser una variable intermedia. Este último criterio no suele depender de los datos sino del conocimiento *a priori* existente sobre el efecto en estudio.

En nuestro ejemplo de la tabla 3 se cumple este tercer criterio. Ciertamente no es plausible que el consumo moderado de alcohol ejerza su papel protector del infarto de miocardio a través de la modificación del hábito tabáquico de los participantes del estudio.

Ahondando en este concepto debido a su importancia, veremos otro ejemplo, mostrado en la tabla 4, sobre la asociación entre la adquisición de tuberculosis y el convivir con personas bacilíferas<sup>1</sup>. Este ejemplo viene representado en la figura 3. El resultado muestra un riesgo cuatro veces superior si se convive con bacilíferos positivos. Se sospechó que estos resultados podrían estar influidos por el nivel de nutrición de los participantes en el estudio. Al estratificar por esta posible variable confundidora comprobamos que cambia el RR en ambos estratos respecto al análisis global o crudo: 8,4 en los malnutridos y 3,16 en los bien nutridos. Nótese que los riesgos son diferentes en ambos estratos, cuestión que no sucedía en el anterior ejemplo. Pero veámoslo paso a paso<sup>6</sup>.

*Comprobación del criterio 1.* La malnutrición debe ser un factor de riesgo de tuberculosis en ausencia de convivientes bacilíferos. Para ello vemos en la tabla 4 que el riesgo de tener tuberculosis en los que representan a la población de bacilíferos negativos, de estar malnutridos o no, es RR=1; por tanto,

**Tabla 4. Tuberculosis y antecedentes de convivientes bacilíferos según el estadio de nutrición**

Tuberculosis			Casos	Personas-año	Tasa (por 100)	RR
Global	Bacilíferos	+	500	10.000	5,0	4,0
		-	125	10.000	1,25	
Con malnutrición	Bacilíferos	+	168	1.600	10,5	8,40
		-	5	400	1,25	
Sin malnutrición	Bacilíferos	+	332	8.400	3,95	3,16
		-	120	9.600	1,25	
Comprobación de <b>criterio 1</b> (factor de confusión con efecto)						
Bacilíferos -	Malnutrición	Sí	5	400	1,25	1
		No	120	9.600	1,25	

RR: riesgo relativo.

Datos recopilados de Delgado *et al.*, 2012<sup>1</sup>.

no cumple este criterio, y ya podemos asegurar que la malnutrición no es un factor de confusión.

Si la malnutrición no es un factor de confusión, y sabemos que puede variar y ser igual entre estratos o no variar respecto a la medida de asociación estudiada, ¿por qué entonces varía el riesgo entre los distintos estratos estudiados? Es debido a que la malnutrición es un modificador del efecto, que en el siguiente apartado estudiaremos.

El lector avezado habrá notado que no hemos utilizado pruebas de significación estadística para comprobar los criterios que debe cumplir un factor de confusión; es decir, para valorar la confusión hemos dado más importancia al efecto que tiene el ajuste sobre la magnitud de los cambios de las medidas de asociación. Por tanto, queda un elemento de subjetividad en el investigador sobre si debe o no ajustar por una variable de confusión porque esa relación sea lo suficientemente importante. Para despejar estas dudas Maldonado y Greeland<sup>8</sup> proponen ajustar la variable confundidora si produce un cambio en la magnitud del efecto superior al 10% independientemente de su dirección; de tal manera que:

$$| OR_{Cruda} - OR_{Ajustada} | / OR_{Ajustada} > 0,10 \text{ (10\%)}$$

Aplicando este criterio al ejemplo de la tabla 3, vemos que la diferencia es  $2,2 - 0,8 = 1,4$ ; que, dividido por la  $OR_{Ajustada}$   $1,4 / 0,8 = 1,75$ , es decir, que varía la magnitud un 175%; por tanto, efectivamente el tabaquismo confunde la medida de asociación inicial y deberá ajustarse por el mismo para dar una medida de asociación correcta.

A continuación veremos cómo debemos proceder si el factor de confusión es variable continua. Una posible solución sería categorizar las variables cuantitativas y utilizar el OR o RR, pero solo deben categorizarse cuando la relación exposición-respuesta no es lineal. Lo que se suele hacer es utilizar el *coeficiente rho* de Spearman como medida de asociación y considerar valores de  $r \leq 0,10$  indicadores de asociaciones débiles. Asimismo, también podemos utilizar el criterio de la  $p$  hasta 0,20 ( $p \leq 0,20$ ) cuando usemos el *estadístico r* de correlación entre las variables (*correlations* con Stata y *correlaciones bivariadas* con SPSS) para muestras no muy grandes, que suele ser en la mayoría de las investigaciones. Para estudiar la correlación si el efecto o la causa son variables nominales y el factor de confusión, podríamos utilizar el *estadístico eta*<sup>1</sup>.

Cuando queremos valorar varios factores de confusión a la vez, llegamos fácilmente a la situación de que nos faltan participantes en los estratos para

poder llegar a una estimación válida en cada uno de ellos. Recuérdese, además, el factor de la pérdida de muestra cuando realizamos estudios de seguimientos, normalmente los de cohortes, por lo que incluso partiremos con menos personas a estudiar. Una alternativa más eficiente para poder considerar el valor de la confusión de diversas variables simultáneamente es el *análisis multivariante*, con el que obtendremos, con la regresión logística, las OR ajustadas de todas las variables y también del efecto que deseamos estudiar. Las variables que se incluyen en el modelo normalmente son aquellas que tienen una  $p < 0,20$ . En los paquetes estadísticos es posible hacerlo directamente indicándoles que extraigan las variables que no lo cumplan con un  $Pout = 0,20$ .

Recomendamos estudiar cada variable de confusión de manera individualizada antes de iniciar el análisis multivariante de las mismas. Deben tenerse en cuenta varios aspectos, ya que no podemos comprobar a la vez todas las variables recogidas o que se nos ocurran porque las variables sexo y edad siempre se contemplan en el modelo (se fuerza su inclusión), y por cada variable incluida deben existir al menos 10 sucesos o eventos. En el ejemplo de la tabla 3 se produjeron  $187 + 42 = 229$  infartos de miocardio en una población total de 60.000 personas-año, por tanto, podremos comprobar como máximo 22 variables en este estudio. Normalmente los estudios clínicos de investigación son más pequeños que el epidemiológico descrito<sup>6</sup>.

## Modificación del efecto (interacción)

Existe interacción biológica entre dos o más factores causales cuando existe una interdependencia entre sus mecanismos de acción para causar, prevenir o controlar una enfermedad, de manera que la incidencia de enfermedad atribuible a este conjunto de factores causales es diferente a la incidencia esperable de sus efectos individuales. Como puede comprobarse en el ejemplo de la tabla 4, es diferente la razón de tasas de incidencia si se tiene en cuenta o no la variable que modifica el efecto.

Por tanto, una variable es modificadora del efecto de una exposición sobre una respuesta cuando, tras estratificar por la variable modificadora, se observa que la magnitud de la asociación (OR, RR o IR) cambia en cada estrato. Ya comprobamos que la malnutrición no era una variable confundidora de la relación entre convivientes bacilíferos y casos de tuberculosis, pero sí es una variable que modifica este riesgo ya que al estratificar pasamos de un  $RR = 4,0$  a un  $RR = 8,4$  con malnutrición y a un  $RR = 3,16$  sin malnutrición.

El concepto de modificación del efecto hace referencia, dentro de un modelo basado en probabilidades, al criterio de interacción causal, y puede ser de sinergismo o antagonismo. El *sinergismo* se refiere a la necesidad de que concurren más de dos variables para que se produzca el efecto (por ejemplo, tabaco y alcohol en el cáncer); y el *antagonismo* hace referencia a la neutralización o cancelación del efecto entre variables (por ejemplo, vitamina A y fumar en el cáncer de pulmón)<sup>1</sup>.

La respuesta depende del modelo biológico que consideremos para entender y explicar el efecto de la exposición a varios factores. Si asumimos que, en presencia de las dos exposiciones, los efectos de cada factor se suman, hablaríamos de un modelo *aditivo*, pero si consideramos que se multiplican, entonces sería un modelo *multiplicativo*<sup>9</sup>.

Existen dos características importantes del concepto de modificación del efecto y son:

- A partir de una misma base de datos se puede llegar a concluir que hay modificación del efecto o que no la hay según la escala escogida, aditiva o multiplicativa, para valorar el fenómeno.
- Cuando se describe una modificación del efecto es preciso especificar la escala en la cual se mide: se hablaría de una desviación de la aditividad cuando se valoran efectos absolutos como diferencias de riesgos, o de multiplicidad cuando se valoran efectos relativos como razones de tasas o riesgos relativos.

Desde el punto de vista del análisis estadístico, las desviaciones del efecto aditivo o multiplicativo se pueden valorar utilizando modelos que se basan en escalas multiplicativas como la regresión logística, que estima OR, o la regresión de Cox, que estima razones de tasas (HR: *hazard ratios*). La interacción se evalúa incorporando al modelo una variable que representa el producto entre los factores, por ejemplo: *Tabaco* × *Alcohol* (siendo de primer orden), y su significación estadística nos indica la existencia de interacción.

A diferencia de la confusión, que es un error o sesgo que hay que corregir, el fenómeno de interacción no debe controlarse sino que ha de ser descrito del modo más exhaustivo y comprensible que sea posible, porque constituye una riqueza de los datos<sup>8</sup>. Por tanto, deben presentarse los resultados desagregados por estratos, como hemos visto en el ejemplo de la tuberculosis (un resultado para malnutridos y otros para normonutridos), no siendo correcto presentar el resultado global. Así es como las guías de práctica clínica recomiendan actualmente que se haga para evitar mal interpretaciones o que se limiten los sesgos por género<sup>10</sup>.

## Aplicaciones prácticas

Sin pretender ser exhaustivos ni crear dogma, a continuación proponemos unos consejos prácticos a la hora de estudiar o analizar los datos de una investigación clínica de cara a minimizar el impacto de los sesgos:

- Establecer una hipótesis del estudio y la relación entre variables y definir claramente el objetivo del estudio, utilizando una pregunta que dé respuesta al formato PICO (*patients, intervention, comparator y outcomes*). Para ello es obligada una buena revisión y estudio de la bibliografía científica. Estableceremos un objetivo principal, el más importante, y unos secundarios que nos ayudarán a entender mejor nuestros resultados, como el análisis estratificado, etc.
- Establecer el tipo de diseño que utilizaremos para contestar el objetivo de nuestro estudio (estudios de casos y controles, de cohortes, transversales, etc.). Debe quedar clara su ubicación tanto temporal como espacial así como las características de los pacientes incluidos y los excluidos.
- Realizar un protocolo para la correcta recogida de datos, definiendo cada una de las variables. Para facilitar el conocimiento de todos los que intervienen en la investigación, desde el investigador principal al analista o personas que recogen los datos, recomendamos siempre realizar previamente a la recogida de datos un diccionario de variables. Este diccionario contiene la información mínima de cada variable: cómo se llama en la base de datos (por ejemplo, *sex*), tipo de variable (*nominal*); su descripción (*sexo*); valores (*0=masculino; 1=femenino*).
- Siempre es importante definir de manera operativa la recogida de datos, es decir, quién la realiza, en qué momento, por qué, etc. Para ello construiremos nuestro formulario o cuestionario de recogida de datos. Y no menos importante es definir correctamente cómo se registrarán los datos digitalmente y su custodia para su posterior análisis. Es muy recomendable consultar al comité ético de investigación clínica (CEIC) de referencia para conocer en profundidad estos términos y realizar una correcta investigación bioética. La aprobación por el CEIC es indispensable para iniciar nuestra investigación.
- La base de datos utilizada para almacenar los datos recogidos puede ser de diferentes características pero siempre debemos realizarla de la manera más segura posible, exenta de errores. Normalmente se utilizan bases de datos como MS Access o SQL, que minimizan los errores de introducción de datos; incluso, actualmente, se escanean directamente los formularios o cuestionarios de recogida de datos para eliminar este paso. Aun cuando realicemos estudios locales y/o pequeños en los que utilicemos MS Excel, debemos recoger los datos de manera segura. Se puede consul-

tar en Youtube la manera de hacerlo (<https://www.youtube.com/channel/UCXAaPw-1M4sUnlrTN8uSyxQ/videos>).

- Por último, se deben describir con antelación los estadísticos que se utilizarán para cada uno de los objetivos propuestos inicialmente. También es bueno conocer si se realizará el análisis de manera ciega o no.
- Antes de proceder al análisis de datos, una vez disponemos de toda la información en nuestra base de datos, es obligado realizar una depuración de los mismos que nos ayude a cumplimentar los faltantes y los que se introdujeron de manera errónea. Hay que utilizar pruebas lógicas y de rango para comprobar cada variable. Este proceso es reiterativo, y para ello nos ayudamos de paquetes estadísticos. En los vídeos de Youtube mencionados anteriormente se explica cómo realizar esto de manera directa para estudios pequeños.

## Bibliografía

1. Delgado M, Llorca J, Doménech JM. Investigación científica: Fundamentos metodológicos y estadísticos. 5ª ed. Barcelona: Signo; 2012.
2. Díez Vegas FJ. Probabilidad y teoría de la decisión en medicina. Programa modular Herramientas de gestión e investigación sanitaria. Madrid: UNED; 2005.
3. Hernández-Ávila M, Garrido F, Salazar-Martínez E. Sesgos en estudios epidemiológicos. Salud Pública de México. 2000;42(5).
4. Manterola C, Otzen T. Los sesgos en investigación clínica. Int J Morphol. 2015;33(3):1156-64.
5. Fletcher RH, Fletcher SW, Wagner EH. Epidemiología clínica. Aspectos fundamentales. 2ª ed. Barcelona: Elsevier Masson; 1998.
6. De Irala J. Unidad 12: Validez y fiabilidad de las mediciones: sesgos, confusión, interacción. Diploma Superior en Metodología de la Investigación 2ª edición. Madrid: OMC-ISCIII; 2008.
7. Rothman KJ. Epidemiology: An introduction. London: Oxford University Press; 2002.
8. Maldonado G, Greenland S. Simulation study of confounder-selection strategies. Am J Epidemiology. 1993;138(11):923-36.
9. De Irala J, Martínez-González MA, Guillén Grima F. ¿Qué es una variable modificadora del efecto? Med Clin (Barc). 2001;117:297-302.
10. Rico R, Aldasoro E, Salcedo F, Calvo M, Barandiaran M, López de Argumedo, et al. Guías de Práctica Clínica libres de sesgos de género. Vitoria-Gasteiz: Departamento de Salud, Gobierno Vasco; 2013.