

Errores en la investigación médica. Una perspectiva epidemiológica

Miguel Delgado Rodríguez

1. Introducción y concepto

Los errores en la investigación médica desde una perspectiva epidemiológica constituyen el eje central de la aportación que los epidemiólogos han realizado en los últimos

veinte años. Un sesgo es todo error sistemático que se comete en un estudio. Los factores relacionados con la selección de los grupos participantes, las tasas de participación y respuesta de la población elegible, la recolección de la información, la metodología empleada y las estrategias analíticas utilizadas introducen a menudo sesgos que, si no son conocidos, pueden conducir a la elaboración de conclusiones incorrectas acerca de los objetivos de una investigación: ya sea simplemente conocer lo que sucede en una población, analizar la relación existente entre una exposición y una enfermedad, investigar los determinantes del pronóstico de un proceso, valorar una prueba diagnóstica o la eficacia de una medida profiláctica o terapéutica, etc.

Mientras que el sesgo puede alterar la validez de los resultados encontrados, el error aleatorio, que por definición sucede por igual en todos los grupos y subgrupos, no afectará en principio a la *validez interna* de un estudio, pero sí reducirá la posibilidad de encontrar una relación verdadera entre una exposición y un efecto. (1) La *validez externa* se interesa por

El presente trabajo es una revisión de los principales errores sistemáticos que se pueden presentar en la investigación clínica-epidemiológica. Tras una definición de los términos de validez y error, se clasifican los sesgos en tres grandes grupos: selección, información y confusión. Dentro de cada uno de ellos se detallan los más frecuentes en su presentación para cada uno de los diseños básicos (estudios de cohortes, de casos y controles, y de corte); exponiéndose el concepto, un ejemplo, así como una introducción de las técnicas dirigidas a su control en cada caso concreto.

la extrapolación de los resultados desde la población de estudio a la población general. Ejemplo, en el estudio de cohortes de los adventistas californianos sobre estilo de vida y enfermedades crónicas (especialmente cáncer),

(2) la *validez interna* se plantea sobre si es verdad que la exposición (el estilo de vida) es un factor de riesgo para el cáncer en los adventistas, en los participantes de la investigación; mientras que la validez externa se cuestiona si ello no es sólo aplicable a los adventistas, sino al resto de la población californiana (otras religiones) o a los habitantes de otros países. Aunque la validez interna es esencial, también es necesario tener en cuenta a la validez externa y para ello la población estudiada ha de ser representativa de la población general. Si se estudia por razones de interés tan sólo una parte de la población (por ejemplo, en el caso anterior un solo grupo ocupacional) la validez externa puede resentirse. Ha de estar claro que sin validez interna no hay validez externa en unos resultados, pero al contrario no: tener validez externa no implica ni mucho menos tener validez interna.

2. Clasificación y generalidades

Los sesgos pueden clasificarse de distintas formas. La manera más aceptada de clasificar los errores sistemáticos es por la etapa de un dise-

ño en que se producen. Esa es una de las grandes aportaciones que KLEINBAUM *et al.*, (1) en su texto pionero, establecieron:

- Sesgo de selección, cuando el error sistemático es introducido en el momento de la incorporación de los sujetos al estudio.
- Sesgo de información, cuando el error se comete en la fase de la recogida de los datos.
- Sesgo de confusión, a nivel de la etapa del análisis, cuando no se ha tenido en cuenta una variable extraña que puede justificar los resultados encontrados en la asociación de exposición-enfermedad estudiada.

Los sesgos pueden ser positivos o negativos. Entendemos por un sesgo positivo cuando la estimación obtenida de un parámetro es superior al valor verdadero del mismo. También recibe los nombres de sesgo hacia afuera o sesgo por exceso. Por el contrario, un sesgo es negativo cuando la estimación alcanzada de un parámetro es inferior a la del valor real del mismo. Se le llama de también sesgo hacia el nulo o sesgo por defecto, ya que disminuye el valor de la asociación. Cuando se estima un parámetro de asociación (como el riesgo relativo, o razón de la frecuencia de un fenómeno en un grupo con respecto a otro) y el valor hallado indica un sentido distinto de la asociación real, esto es, si el valor real del, por ejemplo riesgo relativo, es superior a la unidad (asociación positiva), mientras que se obtiene un valor por debajo de la unidad (asociación negativa), el sesgo recibe el nombre de sesgo de cambio de sentido en la asociación (*switch-over bias*). (1)

3. Sesgo de selección

En sentido estricto, este sesgo surge cuando la variable resultado que se valora (la exposición en los estudios de casos y controles y la enfermedad en los de cohortes) influye diferencialmente la selección de los grupos que se comparan. Sin embargo, aplicando un criterio más laxo se podría incluir en este grupo la elección de grupos poblacionales no estudiados específicamente en una investigación (comparación externa poblacional). Un ejemplo de ello sería el proporcionado por unos autores esta-

dounidenses de Georgia, que compararon la incidencia de cáncer de mama en una cohorte de mujeres que tomaron la píldora, curiosamente identificada (formada en base a las prescripciones anotadas en la farmacia del hospital), con los datos generales de la población (3). En esta población la frecuencia de consumo anticonceptivos orales se aproxima al 50%. Elegir un grupo de comparación que también se encuentra expuesto es inaceptable según los estándares actuales para establecer relaciones de causalidad.

El sesgo de selección se produce con más frecuencia en los estudios en los que se seleccionan los sujetos antes de que la enfermedad ocurra, es decir, es más frecuente en los estudios de casos y controles y en los de corte que en los de cohortes (4). A continuación se estudiarán algunos de los ejemplos más típicos de este tipo de errores.

3.1. Sesgos de selección de estudios de cohortes

3.1.1. Sesgo de respuesta

Puede ocurrir cuando el grupo de estudio no refleja la misma distribución por edad, sexo, etc., de la población diana. Esto puede producirse cuando algunos de los miembros de la cohorte originalmente seleccionados rehúsan participar, o cuando se producen pérdidas o errores en los informes sobre algunos sujetos.

Las tasas de participación entre los distintos subgrupos invitados a participar pueden diferir. Esto ocurre en los voluntarios, que pueden reflejar un particular estado de salud o exposición, y que por su interés se prestan a la investigación, pudiendo alterar las conclusiones finales del estudio. Son formas de sesgo de auto-selección, producido cuando los sujetos que deciden participar se encuentran en un nivel mayor o menor de riesgo con relación al efecto que se valora. Ejemplo: si se compara el efecto sobre el absentismo de un programa de tratamiento del alcoholismo en voluntarios con respecto a las tasas de absentismo en alcohólicos que rechazaron el programa, las diferencias podrían deberse no al programa, sino a la motivación a abandonar el alcohol en los vo-

luntarios que decidieron participar en el programa (5).

En la literatura no es frecuente la valoración de este sesgo de respuesta, ya que no se suele tener información sobre los sujetos que no deciden incorporarse a una investigación. BENFANTE *et al.* (6) sí pudieron hacerlo, aunque de manera parcial, en un famoso estudio, el Programa de Corazón de Honolulu. Encontraron que la tasa de participación (del 73,4%) en el estudio descendió con la edad de los sujetos y que la mortalidad de los no participantes era mayor para todos los grupos de edad (aunque las diferencias se amortiguaron con el paso del seguimiento y envejecimiento de la cohorte, de un cociente de 1,86 para sujetos de 49-54 años a 1,2 en mayores de 75). Esto fundamentalmente se debió a que los sujetos que participaron tenían menos factores de riesgo que los no participantes. Los autores pudieron afirmarlo porque el 60% (16,8% de la población elegida) de los no participantes respondieron al cuestionario inicial de base que fue enviado por correo. Así, los primeros, aunque estaban más gordos (por peso e índice de QUETELET), fumaban menos (46 vs. 53%), eran más educados (48 vs. 39% al menos completada la escuela secundaria), había menos solteros (7 vs. 13%), etc. Estos datos ya por sí sugieren que en ese estudio las cifras y valores de hábitos de estilo de vida no son las de la población general, son algo mejores que la media. De aquí se deriva que las cifras de incidencia de la enfermedad también han de quedar distorsionadas. Uno puede preguntarse si las alteraciones anteriores pueden afectar al valor del riesgo relativo (RR) y sesgarlo, ya que la muestra está distorsionada. La respuesta es afirmativa. El error es del 15,5% (esto es, casi una sexta parte), como puede apreciarse en la tabla 1 (7).

Tabla 1

Mortalidad total (ajustada por edad).
Programa de Corazón de Honolulu,
según el hábito tabáquico

Fumaban	Tasa/100	RR (95% IC)
---------	----------	-------------

Varones no examinados

Fumadores	19,9	1,36 (1,11-1,67)
No fumadores	14,6	1 (referencia)

Varones examinados

Fumadores	14,3	1,63 (1,45-1,82)
No fumadores	8,8	1 (referencia)

TOTAL

Fumadores	15,2	1,58 (1,43-1,75)
No fumadores	9,7	1 (referencia)

3.1.2. El efecto del trabajador sano

En la valoración de las exposiciones ocupacionales en su repercusión sobre la salud hay que tener siempre en cuenta la posibilidad de este error de selección. Los sujetos, por ellos mismos, o por sus empleadores, pueden trabajar en los puestos en los que normalmente son más idóneos. Los trabajos más duros y peligrosos normalmente están desempeñados por los sujetos con más y mejores posibilidades, observándose en ellos una menor mortalidad que en la población general (8,9). Por ello no se puede concluir que una cierta exposición es beneficiosa o no sin eliminar o paliar los efectos de este error. Este fenómeno parece que fue reconocido por vez primera por OGLE en 1885 (10).

3.1.3. Introducción de casos latentes

Un problema diferente ocurre cuando los sujetos con enfermedad latente se incluyen en el estudio. Deberían de haber sido eliminados al comienzo de la investigación. Ello a veces no es posible, ya que hay ciertas situaciones crónicas con largo período de latencia hasta la aparición de la sintomatología, obstaculizando enormemente la identificación de la enfermedad en esa fase. Al no separar adecuadamente la causa presumible y el efecto (por la selección se valora una exposición que es concomitante con un efecto), no se puede concluir que ésta precede a su consecuencia, porque bien podría ser a la inversa: ser la enfermedad la responsable de que el individuo esté expuesto. Este tipo de problemas se pueden dar en ciertos tipos de cánceres y en la aterosclerosis. Para evitar este tipo de errores, al inicio del período de riesgo es necesario hacer un cribaje de casos latentes de enfermedad. De no hacerse, aparte

del hecho mentado con anterioridad, tampoco se podría diferenciar un factor de riesgo de uno pronóstico (como no se abarque un tiempo suficiente de riesgo). Si una variable altera el curso de la historia natural de la enfermedad (acelerándolo o enlenteciéndolo), puede causar que los casos detectados de la enfermedad diana tengan una distribución diferente a la que sería esperable si ese efecto pronóstico no existiera. Por ello, hay que eliminar los casos latentes de la enfermedad al inicio del seguimiento. Un ejemplo de ello podría existir en el *Nurses' Health Study*, estudio de cohortes con 10 años de seguimiento en el que se valoró la información de base (inicio del estudio) mediante cuestionario a más de 120.000 enfermeras en 11 estados norteamericanos (11). El cáncer de mama fue una de las enfermedades que se estudiaron. Dado que su latencia es prolongada, la mayoría de los tumores que se detectaron a lo largo del seguimiento estarían presentes en una proporción apreciable al inicio del estudio.

3.1.4. Pérdidas selectivas en el seguimiento de cohortes

Las pérdidas son la norma en este tipo de diseños. Son ocasionadas principalmente por la larga duración del seguimiento. Se consideran como pérdidas las salidas de la investigación de los sujetos, ya sea porque el propio sujeto decide no continuar, porque fallezca por un problema no relacionado con la investigación, emigre o simplemente se haya «perdido». Está claro que este problema es tanto mayor cuanto más aumenta el número de las mismas. La primera repercusión que supone la existencia de sujetos perdidos es el debilitamiento del análisis, al rebajar el tamaño muestral del estudio.

Si las pérdidas ocurren por igual en los grupos expuesto y no expuesto no afectarán a la validez interna, asumiendo que la tasa de enfermedad de los sujetos perdidos es la misma que la del grupo del que procedían. Pero si la tasa de enfermedad en las pérdidas es diferente de la que presenta la cohorte donde se encontraban, sí habrá una alteración de la validez in-

terna. Este problema resulta similar en sus consecuencias a un sesgo de selección, cuando los sujetos que no deciden participar tienen un riesgo diferente de enfermar del de los participantes. El auténtico problema radica en determinar la tasa de enfermedad en un grupo de sujetos que ya no se encuentran bajo vigilancia. Si las pérdidas son desiguales entre los expuestos y los no expuestos se afectará el resultado final si las pérdidas tienen relación con el efecto a observar, mostrando una probabilidad diferente de la investigación para desarrollar la enfermedad. En este último caso es cuando se afirma que las pérdidas en el estudio se encuentran relacionadas con la exposición y con el efecto y sesgan el resultado final (5, 7, 12).

3.1.5. Sesgo por riesgos competitivos

En esencia, este sesgo se produce cuando se omite en una investigación un factor de riesgo para la enfermedad que se estudia (13). Los métodos de tabla de vida o actuariales son los que se emplean para calcular la probabilidad de desarrollar un fenómeno determinado durante un cierto intervalo, probabilidad condicionada a que el sujeto haya sobrevivido a intervalos de tiempo precedentes. El método actuarial se basa en el aserto de que las restantes causas de pérdida durante el seguimiento (riesgos competitivos) operan independientemente del factor de riesgo que se está estudiando: el sujeto se afecta por la enfermedad diana porque no ha muerto por otras causas. Esto es una visión poco realista en ciertos casos, ya que algunos factores de riesgo son comunes con varias causas de muerte. SCHATZKIN y SLUD (13) proponen dos ejemplos con dos factores de riesgo (uno de los cuales es desconocido u omitido, X) y dos enfermedades. El fenómeno se presenta cuando una de las enfermedades (causa de muerte) sucede en mayor grado por la presencia conjunta de los dos factores de riesgo (que por otra parte no es necesario que estén asociados entre sí). Esto puede motivar que cuando relacionamos el factor conocido con la otra enfermedad se observe una asociación cuando en realidad no exista.

3.2. Sesgos de selección en estudios de casos y controles

3.2.1. Sesgo de Berkson

Se produce en estudios con un asentamiento exclusivamente hospitalario. Surge cuando la probabilidad de hospitalización en los casos es diferente de la de los controles, además de existir una probabilidad diferencial con arreglo a la exposición. Por ejemplo, supóngase que se decide estudiar la relación entre cáncer de pulmón y tabaco mediante un diseño de casos y controles. Los cánceres de pulmón tienen una probabilidad de hospitalización del 100%, con independencia de que hayan fumado o no. Si se elige como grupo de referencia una muestra de patología no cancerosa del hospital, ¿tienen los fumadores la misma probabilidad de ser hospitalizados que los no fumadores? Dado que los fumadores suelen acudir al hospital con más frecuencia, una muestra hospitalaria estará formada por una sobrerrepresentación de fuma-

dores, sesgando el resultado final (en el presente caso, hacia el nulo) de la asociación entre el tabaco y el cáncer de pulmón. Este sesgo es mínimo cuando la probabilidad de hospitalización es la misma para los casos y los controles o cuando la probabilidad de hospitalización de los expuestos con la enfermedad índice o de comparación es pequeña, menor del 5%. No existirá, es obvio, cuando las poblaciones de donde se extraen los casos y controles son mutuamente excluyentes (14, 15).

3.2.2. Falacia de Neyman

Es el sesgo producido por la inclusión de casos prevalentes o supervivientes de la enfermedad, cuando la exposición es un factor pronóstico de la misma (14). Por ello se suele exigir en todo estudio de casos y controles como criterio de calidad el que utilice casos incidentes. Se puede entender fácilmente con un ejemplo. Imagínese un estudio de cohortes como el siguiente:

	E +	E—	Total	
FR +	100	9900	10000	$RR = (100/10000) / (50/10000) = 2$
FR—	50	9950	10000	

FR + : expuestos; FR—: no expuestos; E + : enfermos; E—: no enfermos.

Supóngase que la letalidad de la enfermedad es del 50% en los expuestos y del 20% en los no expuestos. Al cabo de un tiempo morirán 50 casos expuestos y 10 de los no expuestos.

Si se eligen casos prevalentes y una muestra de controles (1% de la muestra total), la distribución será la siguiente:

	E +	E—	
FR +	50	99	$RR = \text{razón de ventaja} = (50*100) / (40*99) = 1,2$
FR—	40	100	

Nótese como en el ejemplo anterior, la razón de ventaja (el estimador RR que se utiliza en los estudios de casos y controles, ya que no se pueden calcular incidencias) produce un valor claramente distinto al de RR.

3.2.3. Sesgo por inclusión/exclusión

Se produce cuando se seleccionan en el grupo control enfermedades relacionadas con la exposición. Este es un sesgo que esencialmente

se produce cuando el grupo de referencia es hospitalario. Para su correcta identificación hay que proporcionar con el mayor detalle posible la patología elegida en el grupo de referencia.

El no proporcionarla está en contra de la metodología normalmente aceptada. Un ejemplo bastante probable de existencia de este sesgo es el estudio de HARRIS *et al.* (16) en el que se estudió la relación, entre otras, entre los anti-

conceptivos orales y el riesgo de cáncer de cuello uterino. Los autores decidieron elegir dentro del grupo de referencia a mujeres con otros tumores (algunos de los cuales se relacionan con los anticonceptivos) y enfermedad benigna, también relacionada con la exposición.

3.2.4. Sesgo de detección

Cuando por alguna razón la gente es sometida a una mayor vigilancia, lo que facilita el diagnóstico de la enfermedad, se puede producir este error. Si se pretende relacionar una enfermedad con la razón que motiva su mayor vigilancia es cuando se produce el sesgo (esto en un estudio de cohortes sería mala clasificación). Es el equivalente del sesgo de casos latentes en las cohortes. Este sesgo se produce, por ejemplo, con gran frecuencia en el estudio de la asociación anticoncepción oral y riesgo de cáncer de cervix uterino (17). Las mujeres que los toman asisten con más frecuencia a los servicios de detección precoz de cáncer genital. Este sesgo no hace referencia a una mayor profundidad diagnóstica en el grupo de expuestos.

3.2.5. Sesgo diagnóstico

Se produce si un proceso se diagnostica más cuando existe la exposición, esto es, la exposición se comporta como un criterio diagnóstico más. Es muy similar al anterior. La diferencia estriba en que aquí es la técnica sobre los pacientes la que cambia en función de la exposición. Por ejemplo, DELGADO *et al.* (18) han mostrado cómo el diagnóstico de la displasia congénita de cadera se realiza con más intensidad en los partos de nalgas. Cualquier estudio epidemiológico realizado en la población tenderá a encontrar como factor de riesgo el parto de nalgas, simplemente porque la detección de la enfermedad es más exhaustiva en ese grupo de niños. Este tipo de error también ha sido considerado en la asociación entre aspirina y síndrome de REYE (19) o al inicio de la epidemia del SIDA, cuando se identificó a los haitianos como grupo de riesgo.

3.2.6. Sesgo de declaración

Este tipo de error es muy similar al caso anterior. Su presencia queda originada cuando una

enfermedad se declara más cuando se conoce la exposición. Sólo tiene repercusión para estudios que se asientan sobre registros de enfermedades. Por ejemplo, piénsese en la cuestionada entre el tampón y el shock tóxico. Los CDC norteamericanos solicitaron que para facilitar la investigación epidemiológica se declararan los casos. Se planteó como razonable el que se declararan a los servicios centrales aquellos casos en los que se conocía la exposición (20). Los médicos en su deseo de ayudar estropeaban el futuro estudio.

3.2.7. Sesgo protopático

Este es un sesgo infrecuente, pero que hay que tenerlo en cuenta casi siempre, especialmente en la epidemiología de las enfermedades crónicas. Se produce cuando la exposición es condicionada por una etapa precoz de la enfermedad. Ejemplo de ello es la asociación entre estrógenos y cáncer de endometrio. El cáncer de endometrio, de especial incidencia en la transición a la menopausia, produce en su inicio metrorragias, que son controladas mediante la administración de estrógenos (21). Un estudio posterior encontraría un aumento de la exposición en los casos, concluyendo falsamente que podría ser un factor de riesgo.

4. Confusión

La importancia de la confusión es capital para la inferencia causal. Cuando se establecen comparaciones con respecto a la exposición se asume siempre que la incidencia de la enfermedad en los expuestos sería igual a la de los no expuestos si no padecieran la exposición. (Esto es, el tabaco favorece el cáncer de pulmón porque se asume que si no fumaran tendrían el riesgo de los no fumadores). Si eso es verdad es que no existe confusión en nuestros datos. Pero si no es verdad, la causa puede residir en la existencia de variables extrañas no tenidas en cuenta en el análisis de resultados. No existe ningún método empírico para confirmar la afirmación anterior. Lo único que se puede realizar es tan sólo buscar diferencias entre los expuestos y los no expuestos por otras variables extrañas que sean conocidas como factores de riesgo para la enfermedad. Si cuando se reali-

za eso se observa que el efecto inicialmente medido está distorsionado en los distintos estratos, se produce una violación clara en el aserto con el que se inició esta exposición. Los factores que producen esa distorsión son los llamados factores de confusión. Por decirlo de otra manera, el efecto producido por un factor de confusión es mezclar su influencia en la enfermedad con la que podría tener la exposición bajo estudio.

Es obligado citar a OLLI S. MIETTINEN, finés afincado en Norteamérica, como uno de los autores (quizá el más importante) que ha contribuido a la detección y tratamiento de este tipo de error, en la década de los 70. (22-25) La confusión también ha recibido otros nombres, aunque en la actualidad han caído en desuso: paradoja de SIMPSON, sesgo de susceptibilidad, sesgo de especificación, o asociación espuria (7). Es el sesgo que con más frecuencia se presenta en la literatura. Hoy en día, con las herramientas analíticas estadísticas de que se dispone, no debiera presentarse este tipo de error.

Es importante diferenciar el error introducido por los factores de confusión —llamado sesgo de confusión— del sesgo de selección. La regla para hacerlo es bien sencilla. El sesgo de confusión es controlable en el análisis, mientras que el sesgo de selección no. Para controlarlo en el análisis de los datos, la única condición existente es que la variable considerada como factor de confusión haya sido recogida en el protocolo.

En general, la condición para que se produzca confusión epidemiológica es que haya una asociación del factor (F) con la exposición (Ex) bajo investigación y con la enfermedad diana (E), que puede producirse por razones diferentes: — Por una asociación causal a nivel de la población diana (no en los resultados del estudio) entre el F con Ex y/o E.

— Por la forma en que los sujetos han sido seleccionados, que condiciona los criterios anteriores. Esto se produce sobre todo cuando se utilizan criterios de restricción en la selección de la población de estudio (se incorporan limi-

taciones para que cierto tipo de sujetos no sean incorporados al estudio).

— Por el uso de ciertos métodos (erróneos) en la recolección de los datos, en la clasificación de los sujetos y en el análisis.

Por todo ello, para una adecuada valoración de la confusión será necesario un conocimiento previo de las causas de la enfermedad en estudio, consideraciones estadísticas y los problemas particulares del diseño empleado en cada situación. Véase con un ejemplo cómo se podría detectar un posible factor de confusión. Considérese el estudio de FRASER *et al.* (26) sobre consumo de frutas y riesgo de cáncer de pulmón en adventistas del séptimo día californianos (tabla 2). El análisis global reveló una asociación protectora entre el consumo de fruta y el riesgo de cáncer de pulmón. Los autores pensaron que podría existir una confusión debida al tabaco, porque así lo habían encontrado otros autores. El análisis estratificado reveló un primer detalle interesante en lo referente a los consumidores de fruta de 3 o más veces a la semana. Resulta que el efecto protector del consumo de fruta es menor que en el análisis global, tanto en los fumadores como en los no fumadores. Si se estudia con detenimiento la tabla, se apreciará que el tabaco es un factor de riesgo independiente para el cáncer de pulmón. Para ser independiente el tabaco tiene que aumentar el riesgo de cáncer en los consumidores infrecuentes de fruta. De hecho, el RR (o más propiamente la razón de densidades de incidencia) fue de 3,9 en este grupo, y de 4,7 en los de consumo alto. Otro de los factores a tener en cuenta para considerar que hay confusión en los datos es que el tabaco se asocie con el consumo de fruta. Si se analizan los datos de la tabla, se apreciará que los consumidores infrecuentes de fruta tienen una mayor proporción de fumadores que los consumidores frecuentes, 44% versus consumidores 24%. Otro detalle importante, aunque no se deriva de los datos numéricos (sino del conocimiento sobre los mecanismos de producción de la enfermedad), es que el consumo de fruta parece que no actúa protectoramente frente al cáncer de pulmón del menor consumo de tabaco. El

sesgo que produce el tabaco es de tipo positivo: se asocia negativamente con la exposición, siendo él mismo un factor positivo, por lo que su ausencia es la que motiva el aumento de la protección.

Los principios de valoración de una determinada variable como factor de confusión: (27, 28).

1. Un factor de confusión (Fc) debe ser (o asumirse que es) un factor de riesgo para la enfermedad que se estudia, o una variable fuertemente correlacionada con un factor no medido, que es el factor de riesgo para la enfermedad. Además, debe ser independiente en ese efecto como factor de riesgo de la exposición (Ex). Esto último quiere decir que debe ser capaz de ejercer su efecto en ausencia de la exposición problema. Por ello, se utiliza como criterio secundario de un factor de confusión el que la *odds ratio* (OR o razón de ventaja, un estimador del riesgo relativo) del factor de confusión sobre el efecto (E) sea distinta de 1 en los no expuestos ($OR_{Fc:E,Ex} < > 1$).

Tabla 2
Dieta, Tabaco y Cáncer de Pulmón (modificado de FRASER *et al.* «26»).

Consumo de fruta	Casos	Personas-año	RR ¹ (IC 95%)
TOTAL			
> 3/sem.	44	166.013	1 (referencia)
< 3/sem.	8	9.128	3.2 (1.5-7.1)
NO FUMADORES			
> 3/sem.	18	124.474	1 (referencia)
< 3/sem.	2	5.144	2.7 (0,6-12.1)
FUMADORES			
> 3/sem.	26	38.539	1 (referencia)
< 3/sem.	6	3.984	2.2 (0,9-5,5)

¹ Estimación calculada por el autor en base a los datos crudos.

2. Debe asociarse con la exposición problema en la población de estudio. Este hecho es fácil comprobarlo en los estudios de cohortes. Por ello se utiliza como criterio secundario en la valoración de un Fc el que la OR entre la

exposición y el factor extraño sea distinta de 1 en los estudios de cohortes ($OR_{Fc:Ex} < > 1$), o que la OR entre el factor y la exposición en los controles (grupo que se aproxima a la población general) sea distinta de 1 en los estudios de casos y controles ($OR_{Fc:Ex:E} < > 1$). Este y el criterio anterior deben cumplirse cuando se controlan los demás factores. Es posible que varios factores que de manera individual se comportan como factores de confusión, cuando se analizan en conjunto se cancelen entre sí sus efectos deletéreos, esto es, la función de riesgo con todas las convariables puede ser idénticas en los expuestos y en los no expuestos.

3. No debe ser una variable intermediaria en la producción del efecto de la exposición en estudio, ya que al estratificar se reduce parte del propio efecto producido por la exposición problema, aunque sea ocasionado a través de otra variable: el presunto factor de confusión. (28, 29) Este criterio no depende de los datos, sino del conocimiento *a priori* existente sobre el efecto en estudio. Es un criterio discutido desde el punto de vista estadístico.

En algunas ocasiones se ha afirmado que uno de los criterios de valoración de confusión en los datos es el hecho de que una estimación ajustada del parámetro que valora el efecto difiera del valor crudo (1). Esto ha sido rotundamente negado y demostrado por otros autores (25).

Además de lo anterior, es necesario tener en cuenta ciertos criterios en el manejo de los factores de confusión:

1. Cuando una exposición cambia a lo largo del tiempo, y ese cambio puede ser función de los niveles alcanzados a inicio del estudio, es aconsejable en estos casos tratar como Fc el nivel inicial de la exposición problema.

2. Cuando se intenta controlar varios Fc en un estudio, no se puede hacer independientemente para cada uno de ellos. Es necesario controlarlos todos al unísono.

3. La identificación de una variable como Fc no debería depender de la investigación estadística de los propios datos obtenidos, ya que quizá no reflejen la verdadera situación de la población general. Recuérdese lo que se afir-

maba al principio sobre el origen de un Fc en los datos.

Un valor elevado del RR es difícil que sea explicado por completo por un factor de confusión (30).

5. Sesgo de información

Existen tres fuentes básicas de sesgo de información en los estudios observacionales:

1. El error en la medición, especialmente la mala clasificación de la enfermedad y/o la exposición, que puede producir el sesgo de mala clasificación en la estimación del efecto.
2. La tendencia que presenta el valor de una variable continua que es extremo en su primera medición a ser más cercano a la media de la población en mediciones posteriores, es decir, la regresión a la media.
3. La falacia ecológica o sesgo de agregación, que se puede producir cuando la unidad de análisis es el grupo, no el individuo. No será estudiado aquí, por tener menor repercusión sobre la clínica.

5.1. Sesgo de mala clasificación

Este tipo de errores se da en la medición de la exposición, el efecto o ambos. Los errores en la valoración de la exposición son considerablemente más frecuentes que los de la enfermedad (31). A pesar del reconocimiento de la importancia que tienen este tipo de errores y de la existencia de procedimientos para tratar de corregirlos, no se utilizan en la práctica. Varias razones pueden ser las responsables. Primero, algunos de los procedimientos existentes son para análisis bivariado, cuando los datos epidemiológicos en el momento presente son multivariados. Segundo, se requieren intervalos de confianza en el momento presente y la mayoría de los procedimientos propuestos no incorporan un estimador de la varianza del parámetro corregido. Tercero, algunos de los procedimientos de corrección más complejos se han desarrollado para modelos que no se usan en el análisis epidemiológico. Cuarto, la mayoría de los artículos publicados sobre el tema están escritos en lenguaje muy sofisticado, por lo que permanece críptico (y sin aplicación, por

ende) para la mayoría de sus potenciales usuarios. Y quinto, los datos de validez y fiabilidad necesarios para aplicar los procedimientos de corrección no están disponibles (31).

La mala clasificación puede ser de dos tipos: diferencial y no diferencial (1, 28).

1. *Sesgo de mala clasificación no diferencial:* Cuando la cantidad de clasificación correcta de una variable (la enfermedad o la exposición) es la misma por los distintos niveles de la otra variable (la exposición o la enfermedad, respectivamente). Por ejemplo el sesgo de mala clasificación no diferencial de la enfermedad significa que la sensibilidad y especificidad diagnóstica es la misma en los expuestos y en los no expuestos. Produce de manera constante una infraestimación del valor del RR (32) lo que ha llevado a KENNETH ROTHMAN a afirmar que esa es la razón por la que los epidemiólogos no han concedido mucha importancia a este tipo de errores. Esto se traduce en que la mala clasificación no diferencial necesita de un tamaño muestral aumentado para tener la misma potencia estadística que si no lo tuviera (33,34), y que conforme la fuerza de la asociación es mayor, mayor es la pérdida de potencia (35). En general, se ha sugerido que en los estudios de cohortes, dado que la enfermedad suele ser un fenómeno raro, es más importante una prueba más específica que sensible que *viceversa*, porque el número de sujetos clasificados incorrectamente es menor.

2. *Sesgo de mala clasificación diferencial:* Cuando la cantidad de clasificación correcta de una variable (la enfermedad o la exposición) es diferente por los distintos niveles de la otra variable (la exposición o la enfermedad, respectivamente). En un estudio de cohortes, por ejemplo, la clasificación de la enfermedad puede ser mala y diferencial si hay sesgos derivados de la observación desigual de las cohortes durante el seguimiento: a) cuando por alguna razón la gente es sometida a una mayor vigilancia, lo que facilita el diagnóstico de la enfermedad, se puede producir *el sesgo de detección*. Si se pretende relacionar esta enfermedad con la razón que motiva la mayor vigilancia es cuando se produce el sesgo.

b) Cuando el médico diagnostica más un proceso en presencia de la exposición, esto es, la exposición es un criterio diagnóstico más, es cuando se produce el *sesgo diagnóstico*. Ejemplo: el tampón y el síndrome del shock tóxico. Es fácil evitarlo en un estudio de cohortes prospectivo: la persona que examina al sujeto participante en busca del efecto diana debe ignorar el estado de exposición del sujeto. En un estudio retrospectivo, por razones obvias, su control es mucho más difícil.

c) Por la misma razón que en el caso anterior, el *sesgo de declaración* se produce cuando una enfermedad se declara más cuando se conoce la exposición. Posiblemente el lector piense que este error es infrecuente en un estudio de cohortes. Recuértese que los estudios de cohortes en los que se recogen unas características de base, que sirven para tipificar a los sujetos su estado de exposición, y el seguimiento se hace en base a contacto con registros de base poblacional, son frecuentes. En ellos es necesario tener en cuenta este tipo de error.

En los estudios de casos y controles destaca el sesgo anamnésico por encima de todos. Los estudios que valoran retrospectivamente una exposición, especialmente los estudios de casos y controles que no validan la información, son más proclives a presentar este error (14, 36, 37). Este tipo de error puede ser considerado como una forma de sesgo diferencial de mala clasificación (38) en el que la capacidad de recuerdo de los grupos que se comparan es distinta. En los estudios de casos y controles, sobre todo en los que valoran resultados reproductivos, se asume que surge de la mayor voluntad de participar de los casos, lo que condiciona un mayor esfuerzo para bucear en su memoria de los mismos (39). Este tipo de error no sólo se puede producir porque la exposición sea recordada de manera diferencial por los grupos que se comparan, sino porque ésta conduzca más fácilmente a un diagnóstico, lo que provoca una forma de sesgo de detección (39). Los estudios de cohortes no están inmunes a este tipo de error. No existe la influencia de la enfermedad para recordar la exposición, pero sí la contraria, especialmente cuando no existen

datos objetivos en la valoración de la enfermedad y el diseño es retrospectivo. Piénsese en los peligros que arrastra la divulgación de los riesgos de ciertas exposiciones (40), especialmente en epidemiología ocupacional, en la que el sujeto al conocer las consecuencias de cierta substancia con la que ha estado en contacto intenta encontrarse la sintomatología que ella determina. Es una consecuencia más del aserto clásico de que el estado emocional influencia la capacidad de la memoria (41). Esta influencia se manifiesta también en la entrevista, donde se ha comprobado que el tono general hedónico de la experiencia del presente (por ejemplo, a través de olores agradables) sesga la memoria (42).

El sentido del sesgo, cuando los casos se recuerdan mejor y más que los controles, es alejarse del nulo, siempre que el factor que se investiga se relacione positivamente con la enfermedad (razón de ventaja superior a la unidad). Lo contrario sucede cuando la variable que se estudia es protectora (38). Esto es, el sesgo es positivo para factores de riesgo y negativo para factores protectores.

En la literatura cada vez son más frecuentes los intentos de valoración de este tipo de error. No en vano las respuestas directas de los participantes en una investigación constituyen con frecuencia la única fuente de valoración de una exposición (o de una condición o estado fisiológico). El campo en el que se ha estudiado más el sesgo de declaración o anamnésico ha sido en epidemiología perinatal (43).

La corrección de los errores de mala clasificación puede ser complicada, sesgada o impráctica porque:

1. Esta aproximación ignora tipos de sesgo, con los que puede coincidir: el sesgo de selección y el de confusión. Por regla general, no tenemos información para estimar la magnitud del sesgo de mala clasificación en la presencia de otros errores. Además, puede haber mala clasificación en las variables por las que se empareja (y en general por factores de confusión), lo que complica la estimación del error. De hecho, cuando existe mala clasificación de una variable que es además un factor de confusión en

una relación exposición-efecto, puede producir, una incapacidad en el control de los factores confusión (32), esto es, producir confusión residual. Esto además fomenta la presencia de heterogeneidad de los parámetros de efecto en las distintas categorías del factor de confusión.

2. Con frecuencia carecemos de información suficiente para afirmar si los errores de clasificación son independientes o no, o si son diferenciales o no diferenciales.

3. Incluso si se conoce el tipo de mala clasificación, se puede no conocer la validez apropiada necesarios para hacer la corrección. Esta limitación es especialmente problemática cuando los errores de clasificación se correlacionan o son diferenciales porque se requiere más parámetros de validez. Además, los parámetros de validez de ciertas variables pueden cambiar substancialmente a lo largo de poblaciones diferentes y con diferentes procedimientos de recogida y clasificación de la información.

5.2. *Regresión a la media*

El sesgo de información puede surgir también de la selección de sujetos cuyas valores iniciales de una variable continua son extremos (muy altos o bajos en relación a otros sujetos). Hay una tendencia para que las mediciones siguientes en estos sujetos den un valor más cercano a la media poblacional, porque algunos de ellos se deben a error en la medición. Este fenómeno se conoce como regresión a la media (44). Al menos tres tipos de estudios se pueden ver afectados por este problema. Uno de ellos es el estudio de corte que servirá como punto de partida a un seguimiento. Un ejemplo lo constituye el Lipid Research Clinics Prevalence Study, en el que se analizó de manera más intensa una muestra aleatoria de los sujetos inicialmente contactados, aparte de estudiar a todos los sujetos con el colesterol y/o triglicéridos elevados. En este segundo grupo es donde se espera que los valores sean a continuación más bajos. En el primero no, porque los sujetos no están elegidos en base a valores extremos, sino en base a todo el espectro de los valores del colesterol/triglicéridos.

Un segundo tipo de estudio que debería tener

en cuenta la regresión a la media es aquél en el que no haya un grupo de referencia. Un ejemplo sería el grupo de sujetos con tensión arterial elevada en el que se prueba de nuevo fármaco, sin que haya un grupo que reciba un placebo.

El tercero de los diseños que se puede ver afectado por la regresión a la media son los estudios experimentales, especialmente el ensayo clínico. Este problema se presenta sobre todo en los ensayos con selección de sujetos con valores extremos en dos etapas. Se produce un problema de selección del punto de corte, por debajo del cual los sujetos serán eliminados. El segundo problema es la elección de la cifra basal o de referencia que indique el éxito en el tratamiento. El descenso en una variable puede ser mezcla del tratamiento y de la regresión a la media. La elección de una cifra basal adecuada puede disminuir mucho este sesgo (44, 45).

Es recomendable que en un estudio toda la escala de variabilidad de la exposición esté adecuadamente representada (27). Esto puede conducir a muestrear en exceso los sujetos con valores extremos (normalmente los más infrecuentes en una población), con lo que el error de regresión a la media puede hacerse más aparente.

Hay varios procedimientos para reducir o evitar este tipo de error:

1. Usar un grupo de comparación que no recibe el tratamiento experimental. Si el diseño se realiza de manera correcta, la valoración del efecto no se verá distorsionada por la regresión a la media. Es más, si el grupo de referencia no recibe ningún tratamiento o sólo un placebo, se puede valorar la cantidad de regresión a la media.
2. La realización de varias mediciones y su promedio reduce la variabilidad introducida por el azar. Cuantas más mediciones se hagan, menor regresión a la media se observará.
3. Podemos usar varias mediciones iniciales independientes para: a) determinar la elegibilidad; y b) proporcionar un valor de base al que se compararán las mediciones

hechas durante el seguimiento. Estas mediciones se hacen en días distintos.

4. Podemos estimar la regresión a la media de manera separada al efecto mediante el uso de técnicas analíticas, que suponen asumir varios hechos acerca de la variabilidad del efecto en la población de origen (46).

Miguel Delgado Rodríguez, *Departamento de Medicina Preventiva y Salud Pública, Facultad de Medicina, Unidad de Epidemiología Clínica, Ciudad Sanitaria «Virgen de las Nieves», Granada.*

Referencias bibliográficas

1. KLEINBAUM, DG.; KUPPER, LL.; MORGENSTERN, H.: *Epidemiologic Research*. Belmont (CA): Lifetime Learning Publications, 1992.
2. MILLS, PK.; NEWELL, GR.; BEESON, WL.; FRASER, GE.; PHILLIPS, RL.: «History of cigarette smoking and risk of leukemia and myeloma: results from the Adventist health study», *J. Natl Cancer Inst.*, 1990; 82: 1832-6.
3. GAMBRELL, RD. Jr.; MASSEY, FM.; CASTANEDA, TA.; BODDIE, AW.: «Breast cancer and oral contraceptive therapy in premenopausal women». *J. Reprod Med.* 1979; 23: 265-71.
4. KLEINBAUM, DG.; MORGENSTERN, H.; KUPPER, LL.: «Selection bias in epidemiologic studies». *Am. J. Epidemiol* 1981; 113: 452-63.
5. GÁLVEZ, R.; DELGADO, M.; BUENO, A. Estudios de cohortes en: «Piédrola G. et al». *Medicina Preventiva y Salud Pública*. 9.ª ed. Barcelona: Salvat, 1991.
6. BENFANTE, R.; REED, D.; MACLEAN, C.; KAGAN, A.: «Response bias in the Honolulu Heart Program». *Am. J. Epidemiol* 1989; 130: 1088-100.
7. DELGADO, M.: *Estudios de Cohortes*. «Curso Superior de Salud Pública», Universidad de Oviedo, 1991, pág. 254 (mecanografiado).
8. CHECKOWAY, H.; PEARCE, N.; CRAWFORD-BROWN, DJ.: *Research methods in occupational epidemiology*. Nueva York: Oxford University Press, 1989.
9. HEMON, D.: «Epidemiologie des risques d'origine professionnelle: quelques problemes methodologiques». *Rev. Epidemiol Santé Publ.* 1986; 34: 230-6.
10. FOX, AJ.; COLLIER, PF.: «Low mortality rates in industrial cohort studies due to selection for work and survival in the industry». *Br. J. Prev. Soc. Med.* 1976; 30: 225-30.
11. ROMJEU, I.; WILLETT, W.; COLDITZ, GA.; STAMPFER, MJ.; ROSNER, B.; HENNEKENS, CH.; SPEIZER, FE.: «Prospective study of oral contraceptives and breast cancer in women». *J. Natl. Cancer Inst.* 1989; 81: 1313-21.
12. GREENLAND, S.: «Response and follow-up bias in cohort studies». *Am J. Epidemiol* 1977; 106: 184-7.
13. SCHATZKIN, A.; SLUD, E.: «Competing risks bias arising from an omitted risk factor». *Am. J. Epidemiol* 1989; 129: 850-6.
14. SACKETT, DL.: «Bias in analytic research». *J. Chron Dis* 1979; 32: 51-63.
15. WALTER, SD.: «Berkson' bias and its control in epidemiologic studies». *J. Chron Dis* 1980; 33: 721-5.
16. HARRIS, RWC.; BRINTON, LA.; COWDELL, RH.; SKEGG, DCG.; SMITH, PG.; VESSEY, MP.; DOLL, R.: «Characteristics of women with dysplasia or carcinoma in situ of the cervix uteri». *Br. J. Cancer.* 1980; 42: 359-69.
17. DELGADO-RODRÍGUEZ, M.; SILLERO-ARENAS, M.; MARTÍN-MORENO, J. M.; GÁLVEZ-VARGAS, R.: «Oral contraceptives and cancer of the cervix uteri. A meta-analysis». *Acta Obstet Gynecol Scand.* 1992; 71: 368-76.
18. DELGADO-RODRÍGUEZ, M.; JIMÉNEZ, C.; LÓPEZ-MORATALLA, M.; SILLERO, M.; GÁLVEZ, R.: «Validity and diagnostic bias in the clinical screening for congenital dysplasia of the hip». *Acta Orthop Belg.* 1993 (en prensa).

19. FORSYTH, BW.; HORWITZ, RL.; ACAMPORA, D. y 10 más: «New epidemiologic evidence confirming that bias does not explain the aspirin/Reye's syndrome association». *Jama* 1989; 261: 2517-24.
20. SCHUCHAT, A.; BROOME, CV.: «Toxic syndrome and tampons». *Epidemiol Rev.* 1991; 13: 199-12.
21. FEINSTEIN, AR.; HORWITZ, RL.: «An algebraic analysis of biases due to exclusion, susceptibility, and protopathic prescription in case-control research». *J. Chron Dis.* 1981; 34: 393-403.
22. MIETTINEN, OS.: «Standardization of risk ratios». *Am. J. Epidemiol.* 1972; 96: 383-8.
23. MIETTINEN, OS.: «Proportion of disease caused or prevented by a given exposure, trait or intervention». *Am. J. Epidemiol* 1974; 99: 325-32.
24. MIETTINEN, OS.: «Stratification by a multivariate confounder score». *Am. J. Epidemiol.* 1976; 104: 609-20.
25. MIETTINEN, OS.; COOK, EF.: «Confounding: essence and detection». *Am. J. Epidemiol.* 1981; 114: 593-603.
26. FRASER, GE.; BEESON, L.; PHILLIPS, RL.: «Diet and lung cancer in California Seventh-Day Adventists». *Am. J. Epidemiol.* 1991; 133: 683-93.
27. MIETTINEN, OS.: *Theoretical epidemiology. Principles of occurrence research in medicine.* Nueva York: Wiley, 1985.
28. ROTHMAN, KJ.: *Modern Epidemiology.* Boston: Little, Brown & Co., 1986.
29. GREENLAND, S., NEUTRA, R.: «Control of confounding in the assessment of medical technology». *Int. J. Epidemiol.* 1980; 9: 61-7.
30. SCHLESSELMAN, JJ.: «Assessing effects of confounding variables». *Am. J. Epidemiol.* 1978; 108: 3-8.
31. WILLETT, W.: «An overview of issues related to the correction of nondifferential exposure measurement error in epidemiologic studies». *Stat Med.* 1989; 8: 1031-40.
32. GREENLAND, S.: «The effect of misclassification in the presence of covariates». *Am. J. Epidemiol* 1980; 112: 564-9.
33. WHALEY, FS.; QUADE, D.; HALEY, RW.: «Effects of method error on the power of a statistical test: implications of imperfect sensitivity and specificity in retrospective chart review». *Am. J. Epidemiol* 1980; 111: 534-42, 648-53.
34. WHITE, E.: «Effect of misclassification of disease status in follow-up studies: implications for selecting disease classification criteria». *Am. J. Epidemiol* 1986; 124: 816-25.
35. GLADEN, BC.; ROGAN, WJ.: «Misclassification and the design of environmental studies». *Am. J. Epidemiol* 1979; 109: 607-16.
36. SCHLESSELMAN, JJ.: *Case-control studies. Design, conduct, analysis.* Nueva York: Oxford University Press, 1982.
37. COUGHLIN, SS.: «Recall bias in epidemiologic studies». *J. Clin Epidemiol* 1990; 43: 87-91.
38. RAPHAEL, K.: «Recall bias: a proposal for assessment and control». *Int. J. Epidemiol* 1987; 16: 167-70.
39. MACKENZIE, SG.; LIPPMAN, A.: «An investigation of report bias in a case-control study of pregnancy outcome». *Am. J. Epidemiol* 1989; 129: 65-75.
40. NEUGEBAUER, R. NG. S.: «Differential recall as a source of bias in epidemiologic research». *J. Clin. Epidemiol.* 1990; 43: 1337-41.
41. RHODES, WS.; RISKIND, JH.; LANE, JW.: «Emotional states and memory biases: effects of cognitive priming and mood». *J. Pers. Soc. Psychol.* 1987; 52: 91-9.
42. EHRLICHMAN, H.; HALPERN, JN.: «Affect and memory: effects of pleasant and unpleasant odors on retrieval of happy and unhappy memories». *J. Pers. Soc. Psychol.* 1988; 55: 769-79.
43. KELSEY, JL.; THOMPSON, WD.; EVANS, AS.: *Methods in Observational Epidemiology.* Nueva York: Oxford University Press, 1986.
44. DAVIS, CE.: «The effect of regression to the mean in epidemiologic and clinical studies». *Am. J. Epidemiol* 1976; 104: 493-8.
45. MCMAHAN, CA.: *Regression to the mean in a two-stage selection program,* 1982; 116: 394-401.
46. CHIM, S.; HELLER, RF.: «Some further results concerning regression to the mean». *Am. J. Epidemiol.* 1981; 114: 902-5.