

## EPIDEMIOLOGIA PARA CLINICOS

### La calidad de la información clínica (I): validez

M. Porta Serra\*, \*\*, C. Alvarez-Dardet\*\*, F. Bolúmar\*\*\*, A. Plasencia\*\*\*\* y E. Velilla\*

\*Unidad de Epidemiología Clínica. Instituto Municipal de Investigación Médica (IMIM).  
 \*\*Facultad de Medicina. Universidad Autónoma de Barcelona. \*\*\*Departamento de Salud Comunitaria, Universidad de Alicante. \*\*\*\*Institut Municipal de la Salut (IMS), Barcelona

Cuando realizamos un estudio de investigación clínica, casi nunca trabajamos con poblaciones completas. Lo que hacemos es, partiendo de observaciones realizadas en un grupo reducido de personas (la llamada muestra), generalizar o extrapolar nuestros resultados a colectivos más amplios<sup>1</sup>. Por ejemplo, si estamos haciendo un ensayo clínico sobre la eficacia de un fármaco antifúngico en las micosis profundas, no lo aplicaremos a todos los afectados de la enfermedad, sino a una muestra de ellos. Sin embargo, implícitamente (y lógicamente), lo que queremos es que nuestros resultados sean aplicables a otros enfermos (quizás, a la mayoría) que desarrollen la enfermedad en otros ámbitos geográficos, temporales o asistenciales.

El hecho de que no trabajemos con poblaciones completas sino con muestras puede introducir en nuestras observaciones errores producidos por el azar, a los que llamaremos errores aleatorios (EA).

Con un ejemplo muy sencillo podremos comprender la naturaleza de este problema (figs. 1 y 2) (tabla 1). Imaginemos que tenemos una población compuesta por 10 chicos y 10 chicas afectados de asma bronquial (fig. 1), y que para establecer la frecuencia de varones que hay en la población, obtenemos diez muestras de cuatro personas cada una. La tabla 1 presenta las personas seleccionadas y las respectivas proporciones de varones en cada una de las 10 muestras de cuatro personas obtenidas al azar. Nótese que en algunas muestras hemos cometido algunos errores (ya que la proporción de varones no es en ellas el 50 %); tales errores pueden atribuirse a la variabilidad aleatoria que conlleva siempre un muestreo. En la figura 2 podemos ver el resultado de llevar a una gráfica las proporciones de varones observadas en cada una de las diez muestras aleatorias (representadas por aspas). Si calculamos la media de tales proporciones (52,5 %) y la comparamos con la frecuencia real de varones (50 %), obtendremos una estimación del EA cometido: EA = 52,5 % - 50 % = 2,5 %.

Es importante darse cuenta de que estos errores se deben al hecho de que estamos muestreando: si utilizáramos a toda la población el EA desaparecería. Asimismo, cuanto mayores sean las muestras empleadas, menores serán los errores que cometamos. El lector interesado puede calcular en el mismo ejemplo el EA que se cometería con muestras de 8, 12, 14 y 20 personas cada una.

Correspondencia: Dr. M. Porta Serra. Unidad de Epidemiología Clínica (IMIM). P. Marítimo, 25-29. 08003 Barcelona

Manuscrito aceptado el 26-8-1987

Med Clin (Barc) 1987; 89: 741-747

Pero el EA no es el único que podemos cometer en un estudio clínico. Existe otro tipo de errores, conocidos como errores sistemáticos (ES), que no están relacionados con el hecho de trabajar con muestras. Además, no tienden a cancelarse cuando aumentamos el número de observaciones. En este artículo desarrollaremos el concepto de ES y el siguiente lo dedicaremos a los EA.

Volvamos a la figura 1. Imaginemos que seguimos teniendo el mismo objetivo, calcular la frecuencia de varones en esta población de 20 asmáticos, y que vamos a seguir obteniendo muestras aleatorias de cuatro personas cada una. Pero ahora resulta que uno de nuestros instrumentos de medida, nuestra definición de lo que es varón, está equivocado, ya que reza: «Varón es todo aquel que lleva pantalones.» Obviamente, con nuestra definición tenderemos a sobrevalorar la importancia de los varones, ya que hay chicas en la figura 1 que llevan pantalones. En la figura 2 hemos representado con puntos negros los resultados obtenidos en las mediciones con la definición errónea, así como el correspondiente ES (tabla 1).

Es importante resaltar que el ES, a diferencia del EA, no tiende a cero cuando aumentamos el número de observaciones. Es más, el ES persiste si hacemos mediciones en toda la población, porque no se debe al hecho de trabajar con muestras al azar, sino al propio diseño del estudio.

#### Tres conceptos fundamentales

Existen tres conceptos que hoy en día se consideran fundamentales para realizar e interpretar investigación clínica: la validez interna, la precisión estadística y la validez externa<sup>2-8</sup> (tabla 2).

TABLA 1

**Error aleatorio y error sistemático: proporción de varones en cada una de las 10 muestras obtenidas al azar y en cada una de las 10 muestras con la definición errónea**

Muestra	Personas seleccionadas	% ♂	% p
1	5, 7, 8, 9	50	50
2	1, 19, 3, 14	50	75
3	16, 11, 8, 1	50	75
4	10, 7, 14, 5	75	75
5	12, 4, 17, 2	50	50
6	20, 13, 15, 5	50	75
7	9, 3, 18, 17	25	50
8	16, 19, 9, 10	75	75
9	4, 9, 13, 20	50	50
10	10, 19, 11, 6	50	75

% ♂ = proporción de varones.

% P = proporción que lleva pantalones.

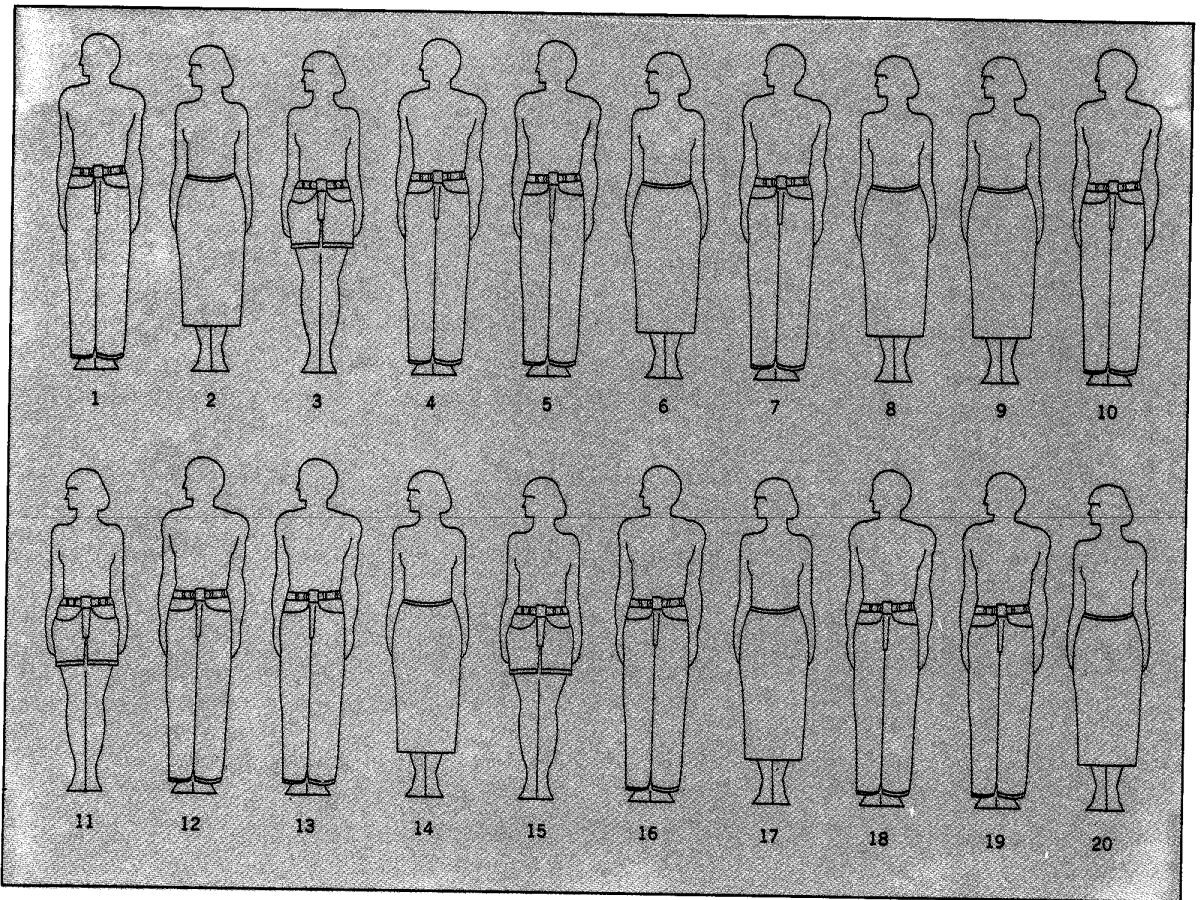


Fig. 1. Error aleatorio y error sistemático. En la población total hay 20 personas asmáticas: 10 chicas y 10 chicos (50 % de varones).

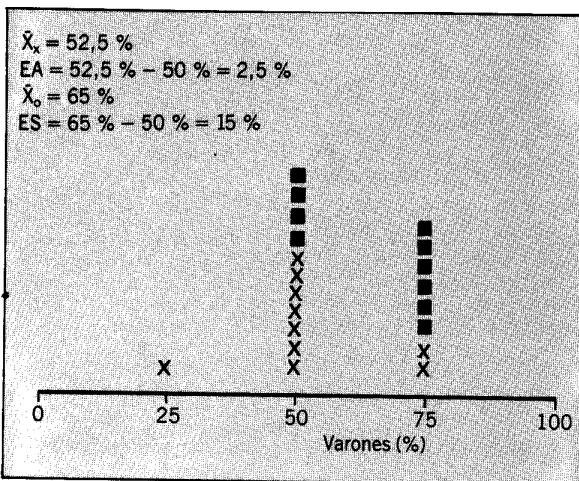


Fig. 2. Error aleatorio y error sistemático en 20 muestras de pacientes asmáticos. X = error aleatorio: proporción de varones en cada una de las 10 muestras obtenidas al azar; ■ = error sistemático: proporción de varones en cada una de las 10 muestras obtenidas con una definición errónea;  $\bar{X}_x$  = media o promedio de varones de las muestras obtenidas al azar; EA = error aleatorio;  $\bar{X}_o$  = promedio de varones de las muestras obtenidas con una definición errónea; ES = error sistemático.

La validez es el grado de ausencia de ES. Su caracterización y la valoración de su importancia es más compleja que la simple aplicación de una prueba estadística. La validez interna es el grado de validez del resultado para los pacientes del estudio.

La precisión estadística es el grado de ausencia de EA. Puede medirse mediante pruebas de significación estadística. Cuanto mayor es la precisión, menor es el papel que el azar juega en el resultado obtenido en un estudio.

El término validez externa plantea la siguiente pregunta: ¿hasta qué punto se pueden aplicar a otros pacientes los resultados de un estudio? Es decir, ¿en qué medida podemos extrapolar o generalizar los resultados obtenidos con unos pacientes a otros pacientes? (tabla 2).

#### Orígenes de los sesgos

Todo estudio clínico es susceptible de cometer dos grandes tipos de errores: los ES y los EA. Los ES o sesgos conllevan siempre que la validez interna del estudio sea baja. Los sesgos o ES también pueden resultar en una validez externa baja. Los EA resultan en la falta de precisión estadística y en una significación estadística baja (tabla 2).

Así, pues, las preguntas que nos hacemos al valorar el resultado de un estudio son tres. La primera y la más importante es si el resultado de aquel estudio es correcto, es ve-

TABLA 2

**Definiciones de validez, sesgo y algunos conceptos relacionados**

**ERROR ALEATORIO** (*random error; erreur aléatoire*)\*. Diferencia entre la estimación obtenida en el estudio y el parámetro que se pretende estimar debida al azar. Se atribuye esencialmente a la variabilidad inherente al proceso de muestreo, al proceso de medición y a la propia variabilidad aleatoria de las variables (biológicas, sanitarias, etc.).

**ERROR SISTEMÁTICO o SESGO** (*systematic bias; biais systématique*). Cualquier hecho (estructura del estudio, selección y valoración de los pacientes, método de recogida de los datos, análisis de los mismos) que hace que los resultados de un estudio se desvíen o aparten de la realidad. Diferencia entre la estimación del efecto obtenida en el estudio y el parámetro que se pretende estimar.

**VALIDEZ** (*validity; validité*). Grado de ausencia de sesgos o errores sistemáticos. Grado de concordancia entre la estimación del efecto obtenida en el estudio y el parámetro que se pretende estimar.

**VALIDEZ INTERNA** (*internal validity; validité interne*). Validez del resultado para los pacientes del estudio.

**VALIDEZ EXTERNA**, generalizabilidad (*external validity; generalizability; validité externe*). Validez del resultado para los pacientes no incluidos en el estudio.

**PRECISION** (*precision; précision*). Grado de ausencia de errores aleatorios. Grado de dispersión de las observaciones, grado de acercamiento entre la estimación obtenida y el parámetro que se pretende estimar (calcula). Cuanto menor sea la magnitud del error aleatorio, mayor será la precisión.

**GRADO DE SIGNIFICACION ESTADÍSTICA** o valor de la «p» (*statistical significance, p-value; signification statistique*). Un valor asociado con el resultado de una prueba estadística, que indica la probabilidad de que un valor tan o más extremo que el observado pueda producirse únicamente por el azar en sucesivas repeticiones del estudio. Probabilidad de que el resultado observado se deba al azar.

**FACTOR DE CONFUSION** (*confounding factor, confounder; facteur de confusion, variable concomitante*). Un factor de riesgo para la enfermedad que a su vez está asociado con la exposición, de tal modo que el efecto de la exposición sobre la enfermedad aparece mezclado con los efectos del factor sobre ambas.

**AZAR o suerte** (*randomness, chance; chance*). Variabilidad inherente a los fenómenos naturales. Puede tratarse cuantitativamente y describirse mediante modelos de probabilidad que miden, así, el grado de incertidumbre de unas observaciones.

**DESENLACE clínico o EFECTO** (*outcome, effect; dénouement, effet*). Fenómeno resultante o consecuencia de una actividad clínica, de una característica fisiológica, de la exposición a factores ambientales o de ciertos hábitos de vida.

**A DOBLE-CIEGO** (*double-blind; doublement à l'aveugle, en double insu*). Una de las técnicas de enmascaramiento, consistente en que ni el paciente ni el médico conozcan qué tratamiento recibe el primero. Su objetivo es que la percepción, expresión, detección y valoración de los efectos clínicos no se vea influida por el conocimiento del tratamiento o intervención que se recibe o aplica. Es, pues, una técnica para minimizar los sesgos en la información. Otras técnicas de enmascaramiento son el ciego-simple (sólo el paciente desconoce su tratamiento), el triple-ciego (además del paciente y su médico, tampoco el evaluador o analista de los datos conoce qué tratamiento fue asignado a cada paciente) y combinaciones de las anteriores.

\*El primer término es la equivalencia inglesa y el segundo la francesa<sup>14,15,19,20</sup>.

raz, es cierto (validez interna). En caso afirmativo, procede preguntarse (segunda pregunta) cuán precisa o estadísticamente significativa es la respuesta que da al problema planteado; es decir, hasta qué punto los resultados obtenidos pueden deberse al azar. Lógicamente, si el estudio no nos parece válido esta segunda pregunta sobra. La tercera pregunta es si el resultado es aplicable a nuestros propios pacientes (validez externa).

En este momento hay que decir que el número de ES que pueden operar en un estudio es, potencialmente, enorme. En los últimos años han obtenido cierta aceptación algunos catálogos de sesgos<sup>9-12</sup> y la caza del sesgo agazapado ha ganado numerosos adeptos. Ciertamente, es más fácil decir cómo no deben hacerse los estudios que proporcionar orientaciones sobre cómo deben hacerse. Pues a veces parece que investigar consistiera meramente en sortear hábilmente un campo minado con sesgos<sup>13</sup> y que el método epidemio-

lógico no fuera más que un recetario para criticar mordazmente los trabajos del prójimo. Conviene, por el contrario, tener presente que cuando se revisa un estudio, no basta con sugerir la posibilidad de que en él opere un sesgo, sino que es necesario hallar algún indicio razonable de que tal distorsión existe realmente (y valorar adecuadamente su magnitud y sus implicaciones prácticas).

También en los últimos años ha cobrado aceptación la clasificación del mosaico de ES en tres grandes grupos: los sesgos de selección, los sesgos de información y los factores de confusión (tabla 2). De modo que, al leer un estudio clínico y preguntarnos si el resultado obtenido podría deberse a algo que los autores no han tenido en consideración, pensaremos en cuatro posibilidades: 1) los grupos del estudio no son comparables debido a cómo fueron seleccionados los pacientes (diremos entonces que existen errores, distorsiones o sesgos en la selección); 2) los grupos del estudio no son comparables debido a cómo se obtuvieron los datos (existen sesgos en la información); 3) los autores no han recogido información (o la han obtenido pero no la han utilizado) sobre un factor que se relaciona a la vez con la exposición y con el efecto estudiados (en epidemiología a tal factor se le denomina un factor de confusión), y 4) ninguna de las tres situaciones anteriores, con lo que aquel estudio tendrá un grado alto de validez interna (tabla 2).

**Sesgos en la selección de los pacientes**

Empecemos por la primera situación: el resultado obtenido puede explicarse, total o parcialmente, por cómo se seleccionaron los pacientes del estudio. Este tipo de error ocurre en fases de un estudio tales como la entrada de los pacientes en el estudio (muestreo, criterios de inclusión y de exclusión), la asignación de los tratamientos (cómo se selecciona quien tomará cada tratamiento)<sup>14</sup>, la elección del grupo testigo o de control (por ejemplo, en estudios sobre la etiología de una enfermedad) y el seguimiento de los pacientes (cómo se forman los grupos que finalmente, tras el período de seguimiento, son analizados). Los ES cometidos en la valoración del estado y los antecedentes clínicos de los pacientes (al inicio, durante y al término del estudio) se clasifican como sesgos en la información.

Veamos un primer ejemplo. Siempre que una serie de casos clínicos compara los resultados de un tratamiento con los resultados de otras series, cabe preguntarse si las diferencias que se observan pueden deberse a otros factores que no sean los tratamientos<sup>14</sup>. Puede que algo haya cambiado a lo largo del tiempo. Por ejemplo, la historia gestacional de las mujeres de una determinada serie puede ser distinta de la de las mujeres seleccionadas en otra serie anterior, incluso si ambas proceden del mismo hospital (ya que en la actualidad es superior el porcentaje de mujeres cuyo primer parto sucede a una edad mayor). La historia gestacional puede alterar factores hormonales que modulan la respuesta al tratamiento. Por ello, para que en nuestro ejemplo la comparación esté libre de sesgos es importante que los grupos se seleccionen de forma que sus historias gestacionales sean parecidas.

Tal y como indicamos en el artículo sobre las medidas de frecuencia<sup>15</sup>, el uso de la prevalencia o el uso de casos prevalentes en estudios clínicos y epidemiológicos puede ocasionar algunos problemas, ya que los casos prevalentes representan lo que existe o queda en un momento determinado, pero no todo lo que ocurrió anteriormente. Recordemos que para cualquier enfermedad con una cierta letalidad, los casos prevalentes en un momento dado equivaldrán a los casos incidentes menos los casos que han muerto hasta ese momento.

TABLA 3

**Un ejemplo de sesgo de selección: sesgo por supervivencia selectiva**

La exposición prolongada al benceno a dosis altas (10 ppm/año) aumenta el riesgo de morir de leucemia y mieloma múltiple. Probablemente, la exposición a dosis más bajas o de menor duración no aumenta sensiblemente tales riesgos<sup>18</sup>. Supongamos que, en un momento determinado, realizamos un estudio de casos y controles con el que pretendemos estudiar el riesgo de sufrir una leucemia que tienen las personas expuestas al benceno. Nótese que, aunque tal es el objetivo, por tratarse de un estudio caso-control la secuencia causal<sup>19</sup> es la inversa: partimos del desenlace (afecto de leucemia, no afecto de leucemia) y observamos la causa (expuesto o no a benceno); por supuesto, la elección de tal secuencia causal está justificada, ya que la leucemia es una enfermedad relativamente poco frecuente y la asociación que estudiamos tiene un período de latencia largo<sup>19</sup>.

Supongamos asimismo que los casos de leucemia que seleccionamos son aquellos que están vivos en el momento en que realizamos el estudio. Es decir, no incluimos todos los casos de leucemia que se produjeron a lo largo de un período de tiempo entre todos los expuestos a benceno en un lugar determinado, sino que incluimos en el estudio sólo a los que sobrevivieron y excluimos a los que murieron. Así, pues, el estudio incluye sólo casos de leucemia ya existentes (casos prevalentes), pero no incluye casos nuevos (casos incidentes)<sup>15</sup>. Una vez seleccionados los casos de leucemia y los controles (sin leucemia), comparamos las respectivas proporciones de pacientes que habían estado expuestos a benceno. Hallamos que la proporción de casos de leucemia expuestos a benceno es similar a la proporción de controles expuestos a benceno. Siguiendo la lógica del diseño caso-control, y tras calcular la correspondiente *odds ratio* y sus intervalos de confianza al 95 % (que incluyen al 1)<sup>19</sup>, concluimos que el benceno no parece aumentar el riesgo de leucemia.

En realidad, lo que puede haber ocurrido es que los expuestos a dosis altas de benceno hayan tenido una mayor mortalidad (lo que ha hecho menos probable que sean incluidos en el estudio, ya que éste selecciona únicamente a los casos existentes en un momento determinado), mientras que los expuestos a dosis bajas han sobrevivido lo suficiente como para sí ser incluidos.

Ejemplo. En el año 1972 Rogentine et al<sup>16</sup> publicaron un trabajo en el que comunicaban haber hallado una asociación entre leucemias agudas linfocíticas y el antígeno HLA-A2. Pero el año siguiente, en otro artículo<sup>17</sup>, los mismos autores reconocieron que el HLA-A2 no sólo no era un factor de riesgo para las leucemias agudas linfocíticas, sino que en realidad parecía ser un marcador de buen pronóstico para las mismas. Esta aparente paradoja se produjo a causa de un sesgo en la selección de los pacientes: en el primer estudio habían seleccionado casos prevalentes (existentes en un momento dado) de leucemia aguda, y no casos incidentes (todos los casos nuevos ocurridos durante un período determinado). Como el HLA-A2 parece ser un factor de buen pronóstico, los casos con HLA-A2 están sobrerrepresentados en su muestra, debido a su mayor supervivencia. La tabla 3 presenta otro ejemplo de sesgo por supervivencia selectiva.

La relativa complejidad de los ejemplos anteriores (nada inusual, por lo demás, en las revistas médicas) nos recuerda también que, aunque ciertamente imprescindible, el sentido común no suele bastar para investigar en Medicina<sup>20</sup>. En cambio, un cierto conocimiento de los métodos de investigación, sin necesidad de grandes sofisticaciones matemáticas, sí permite juzgar por uno mismo la validez de lo que otros publican.

Además, suele ser factible superar ciertas limitaciones metodológicas. En el caso de los dos ejemplos anteriores, podemos mencionar dos posibles soluciones<sup>8</sup>. Primera, seleccionar únicamente casos nuevos (incidentes) de la enfermedad; si ello no resulta viable, podemos incluir casos prevalentes pero que fueron diagnosticados lo más cerca posible (en el tiempo) del momento en que se inicia el estudio. Segunda, estratificar a los casos según la fecha de diagnóstico, analizando la frecuencia de la exposición en cada

estrato; si tal frecuencia es distinta en los diferentes estratos, existe el riesgo de que la duración de la enfermedad o la existencia de una supervivencia selectiva estén provocando un sesgo en el estudio.

Cuando los estudios se realizan en hospitales debe tenerse en cuenta la posibilidad de que ocurra el llamado sesgo de Berkson (tabla 4). Allander<sup>23</sup> ha señalado que la influencia de este sesgo no es debidamente apreciada por los clínicos, y Feinstein<sup>24</sup> ha sugerido diversas alternativas para evitar los sesgos derivados de las diferentes formas de acceso de los pacientes a los hospitales.

Otras formas de sesgos en la selección de los pacientes pueden ocurrir cuando comparamos enfermos con voluntarios (sanos o enfermos), pacientes ingresados con pacientes atendidos en consultas externas, enfermos diagnosticados mediante diferentes técnicas, pacientes con diferentes antecedentes familiares, o pacientes que aceptan y enfermos que rechazan el tratamiento propuesto<sup>14</sup>.

En un estudio sobre la utilidad de una prueba diagnóstica, se produce un sesgo cuando sólo se seleccionan casos que claramente tienen la enfermedad y casos que claramente no la tienen. La utilidad de la prueba estará sobrestimada en el estudio, y resultará inferior cuando la prueba se utilice en pacientes en los que simplemente se sospecha la enfermedad. Para hacer un estudio válido hay que incluir un amplio espectro de pacientes y que se parezcan a aquellos en los que nos plantearemos solicitar la prueba.

**Sesgos en la información**

Examinemos ahora la posibilidad de que los grupos del estudio no sean comparables debido a cómo se obtuvieron los datos, es decir, los sesgos en la información.

Un simple caso de este tipo de errores es el siguiente: el investigador detecta o valora de forma sistemáticamente distinta el estado clínico de los pacientes según el tratamiento que reciben; por ejemplo, se realizan exploraciones que permiten detectar o atribuir un efecto adverso con mayor intensidad o frecuencia en los pacientes que toman un fármaco ya conocido que en los pacientes que toman un medicamento en fase de estudio.

También son frecuentes los errores en la información en aquellos estudios cuyos datos proceden exclusivamente de las historias clínicas, sin verificación alguna.

Por otra parte, sucede a menudo que las personas que sufren una enfermedad recuerdan sus antecedentes clínicos, familiares y laborales de forma diferente que las personas que no sufren la enfermedad. Por ejemplo, las mujeres afectas de cáncer de endometrio suelen tener una probabilidad mayor de recordar si han sido tratadas con estrógenos que las mujeres que sufren una infección vaginal, ya que su esfuerzo memorístico se halla condicionado por los comentarios del personal sanitario que las atiende y por el hecho de padecer una enfermedad grave.

La gravedad de una enfermedad, su duración y su popularidad en un momento determinado pueden influir en que la anamnesis farmacológica sea más completa. Así, parece plausible que se produzca un sesgo si nos conformamos con comparar la proporción de mujeres afectas de cáncer de endometrio que refieren (o en cuyas historias clínicas consta) la ingesta de estrógenos con la proporción de mujeres afectas de una enfermedad benigna y aguda que refieren (o en cuyas historias se menciona) dicha ingesta. Para minimizar el sesgo podemos tomar varias medidas: comprobar en las historias clínicas los tratamientos prescritos, seleccionar como mujeres-control a las que acuden a los mismos servicios que las mujeres-caso por síntomas que requieren exploraciones similares, entrenar a quienes realizan las entre-

TABLA 4

Un ejemplo del sesgo de Berkson<sup>21,22</sup>

El llamado sesgo de Berkson<sup>22</sup> fue descrito en el año 1946, pero curiosamente nadie hasta una época muy reciente había comprobado empíricamente su importancia. Ello ocurrió en 1978 en un estudio de base comunitaria<sup>21</sup>. De una muestra aleatoria de 2.784 individuos de una comunidad, 257 (9,2 %) habían estado hospitalizados en los seis meses anteriores. Además de los datos sobre las hospitalizaciones, los investigadores recogieron información sobre la presencia o ausencia de enfermedades respiratorias y del aparato locomotor. Se realizó un análisis de los datos referidos únicamente a los enfermos que habían sido hospitalizados (tabla 4a), y otro análisis referido al total de la muestra comunitaria (tabla 4b).

TABLA 4a

## Análisis de los pacientes hospitalizados

		Enfermedades del aparato locomotor		
		sí	no	
Enfermedades respiratorias	sí	5	15	257
	no	18	219	
OR = 3,57		IC95 % = 11,5 y 1,41		

TABLA 4b

## Análisis de todos los pacientes

		Enfermedades del aparato locomotor		
		sí	no	
Enfermedades respiratorias	sí	17	207	2.784
	no	184	2.376	
OR = 1		IC95 % = 1,77 y 0,67		

En la tabla 4a puede observarse como, entre los enfermos hospitalizados, existe una asociación estadísticamente significativa entre las enfermedades respiratorias y las del aparato locomotor (OR = 3,57, LSIC = 11,5, LIIC = 1,41). Por el contrario, en la tabla 4b no se observa la asociación (OR = 1, LSIC = 1,77, LIIC = 0,67).

Esta aparente paradoja se produce porque la muestra hospitalaria no es realmente una muestra al azar de la comunidad, sino que está condicionada por la tasa de admisión hospitalaria que tiene cada uno de los elementos que hay en las celdillas de las tablas de 2 x 2. Así, el porcentaje de personas hospitalizadas entre las que no tienen ninguna de las enfermedades es bajo (219/2.376 = 9 %). También lo es entre las que presentan o sólo enfermedad respiratoria (15/207 = 7 %) o sólo enfermedad del aparato locomotor (18/184 = 9,8 %). Sin embargo, la frecuencia de ingresos entre las que presentan las dos enfermedades es más del doble (5/17 = 29 %), lo que explica que las personas con las dos condiciones estén sobrerrepresentadas y que obtengamos una medida de la asociación (la OR)<sup>19</sup> que sobrealora la magnitud de la asociación entre ambas enfermedades.

OR = odds ratio; IC95 % = intervalos de confianza al 95 %; LSIC = límite superior del intervalo de confianza; LIIC = límite inferior del intervalo de confianza.

vistas y no informarles sobre quiénes son los casos y quiénes los controles, y preguntar a todas las pacientes por los nombres comerciales de los preparados, enseñándoles las cajas de tales preparados para ayudarles a recordar.

Conviene señalar al respecto que lo importante (para realizar una comparación válida de un grupo con otro) es, en primer lugar, que la detección de los antecedentes sea igual de completa en ambos grupos; secundariamente, intentaremos que la detección sea totalmente completa en ambos grupos. Este mismo principio vale también para los estudios prospectivos de cohortes y para los ensayos clínicos con asignación aleatoria<sup>19</sup>. En ellos, nos interesa que la probabilidad de detectar un efecto o desenlace clínico sea igual en el grupo de expuestos y en el de no expuestos; secundariamente, interesa detectar la totalidad de los efectos. Pero puede que no exista un sesgo si se nos escapan algunos efectos, de la misma naturaleza, en ambos grupos por igual. La igualdad o similitud en la detección y valoración no es fácil de alcanzar. Uno de los sesgos en la información más frecuente ocurre cuando, influidos por lo que sabemos, deseamos o esperamos de un tratamiento, buscamos con más atención determinados efectos en los pacientes que toman aquel que en los pacientes expuestos a otro tratamiento. El llamado doble-ciego y las otras técnicas de enmascaramiento<sup>25</sup> (tabla 2) ayudan precisamente a minimizar este tipo de errores<sup>26</sup>.

Un ejemplo clásico de la influencia de la enfermedad sobre la información que los pacientes proporcionan se refiere a la posible agregación familiar de la artritis reumatoidea (AR). En un primer análisis, Schull y Cobb<sup>27</sup> hallaron que, en comparación con un grupo control formado por personas sanas, una mayor proporción de los pacientes con AR mencionaba una historia familiar de esta enfermedad. Sin embargo, la asociación entre AR e historia familiar de AR desaparecía cuando se comparaban los antecedentes fami-

liares del grupo de control con los antecedentes familiares referidos por los hermanos sin AR de los pacientes con AR. Así, mientras que un 8 % de los controles decía que ambos padres habían sufrido AR, esta proporción era el 15 % en los casos de AR, pero solamente el 8 % (igual que en los controles) en los hermanos sanos de los casos.

Por otra parte, se ha definido como sesgo de sospecha diagnóstica<sup>9</sup> al hecho de que el conocimiento de que el paciente ha estado expuesto a un factor de riesgo condicione la intensidad con la que se investiga un determinado diagnóstico. Por ejemplo, si sabemos que un niño de 8 años con una erupción mucocutánea generalizada de seis semanas de duración ha estado tomando sulfamidas, es más probable que pensemos en un síndrome de Stevens-Johnson que si desconocemos tal exposición.

También el diagnóstico de tromboembolismo venoso puede ser más probable en una mujer que ha estado tomando contraceptivos orales que en una que no los tomaba (o que desconocíamos los tomaba), a pesar de que la sintomatología de ambas sea idéntica. Incluso si realizamos una verificación del diagnóstico independiente de la anamnesis (para minimizar un sesgo de información), el envío y la admisión al hospital de la paciente pueden haber estado ya condicionados por el conocimiento de su historia farmacológica, lo que supone un sesgo de selección<sup>28</sup>.

Es vital recordar que la inmensa mayoría de sesgos en la selección y en la información no podrán ser corregidos después de recoger los datos. No existe estadístico, informático o epidemiólogo que pueda salvarnos de un estudio mal diseñado o ejecutado. Por ello, antes de empezar a recoger los datos hay que pensar en qué sesgos pueden ocurrir y en cómo diseñaremos el estudio para poderlos controlar. Un buen investigador clínico busca ayuda metodológica mucho antes de empezar el estudio y sólo analiza datos de cuya validez interna está convencido.

TABLA 5

**Consideraciones sobre los factores de confusión<sup>2-4</sup>**

Para crear confusión (en el sentido epidemiológico) un factor externo debe cumplir tres condiciones:

- Ser un factor de riesgo para la enfermedad
- Estar asociado con la exposición (en el caso concreto de la población de donde provienen los enfermos)
- No ser un simple factor intermedio en la vía etiopatogénica entre la exposición y la enfermedad

Cuando se cumplan las tres condiciones anteriores, el riesgo relativo (o la medida de la asociación entre la exposición y el efecto que en cada caso se utilice) bruto (es decir, el que no tiene en cuenta al factor externo) será distinto del riesgo relativo ajustado (el que sí tiene en cuenta al factor)

La comparación del riesgo relativo bruto y del riesgo relativo ajustado permite conocer la dirección y la magnitud del sesgo provocado por el factor de confusión

La valoración de la magnitud del sesgo provocado por el factor de confusión es una cuestión de juicio subjetivo; las pruebas de significación estadística no son adecuadas para ello

No tener en cuenta posibles factores de confusión puede disminuir la validez interna del estudio. Ajustar por factores externos que no son realmente factores de confusión puede disminuir la precisión estadística

Al contrario que los sesgos de selección y los sesgos de información, los sesgos debidos a factores de confusión pueden ser corregidos después de recoger los datos, durante su análisis. Pero para ello habrá que decidir previamente (durante el diseño del estudio) qué factores de confusión será necesario controlar en la fase de análisis

La decisión de qué factores externos se tendrán en cuenta durante el análisis de los datos como posibles factores de confusión debe hacerse:

- Antes de recoger los datos
- Basándose en conocimientos sólidos sobre la etiopatogenia de la enfermedad
- En función de la hipótesis del estudio

**Efectos mezclados**

Veamos ahora la tercera y última situación que, como mencionábamos al inicio del artículo, puede poner en peligro la validez interna de un estudio clínico: los autores no han tenido en consideración un factor que se relaciona a la vez con la exposición y con el efecto estudiados (un factor de confusión). Como siempre, la exposición puede ser un fármaco, un hábito de vida, una variable biológica (el grupo sanguíneo, una hormona, un determinado título de anticuerpos), o cualquier otro factor de riesgo, intervencional preventiva o tratamiento. El efecto puede ser una enfermedad, una curación, un diagnóstico temprano, o cualquier desenlace de interés clínico<sup>15</sup> (tabla 2).

Para definir lo que es un factor de confusión podemos considerar la situación más común en investigación etiológica: aquella en la que estudiamos la asociación entre una exposición (E) y un efecto (Y), teniendo en cuenta la influencia de otro factor (F). En tal caso, un factor de confusión es un factor de riesgo para la enfermedad, el cual, a su vez, se asocia con la exposición. Así, la asociación entre E e Y aparece en nuestros datos mezclada con F. Para obtener un cálculo no sesgado de la magnitud de la asociación entre la exposición y el efecto debemos tener en cuenta al posible factor de confusión.

En otras palabras, un factor de riesgo para la enfermedad será un factor de confusión al estudiar la asociación entre la exposición y el efecto sólo en el caso que el parámetro con el que medimos la asociación entre E e Y (por ejemplo, el riesgo relativo)<sup>19</sup> sea distinto cuando no tenemos en consideración a F y cuando sí tenemos en cuenta a F (tabla 5). Supongamos que un equipo de clínicos estudia la asociación entre cáncer de la cavidad bucal (efecto, Y) y consumo elevado de bebidas alcohólicas (exposición, E)<sup>3</sup>. Un factor de riesgo para este tipo de neoplasias es el tabaco (F). Además, diversos estudios han hallado una asociación entre fu-

mar y beber alcohol: no sólo existen personas abstemias (ni fuman ni beben) y personas que fuman mucho y beben mucho, sino que, en muchas poblaciones, la proporción de fumadores es mayor entre los bebedores que en los no bebedores, y viceversa, la proporción de bebedores es mayor entre los fumadores que en los no fumadores.

Puesto que el tabaco (F) es un factor de riesgo para el cáncer de boca (Y), y puesto que el tabaco también puede asociarse a la ingesta elevada de alcohol (E), los investigadores consideran, acertadamente, que el tabaco puede ser un factor de confusión; en consecuencia, deciden tenerlo en consideración al analizar la magnitud de la asociación entre E e Y.

En su primer análisis, en el que no tienen en cuenta a F, el riesgo de neoplasia es 6,7 veces mayor en los bebedores que en los no bebedores (riesgo relativo o RR = 6,7). En un segundo análisis de los mismos datos, al tener en cuenta el hábito tabáquico de los expuestos y de los no expuestos al alcohol, el RR disminuye a 3,2. Puesto que el RR bruto (6,7) difiere del RR ajustado (2,3), F sí era un factor de confusión en este estudio. El análisis bruto (que no ajusta por F) estaba sesgado y, para ser más concretos, sobrestimaba el riesgo de cáncer. La dirección del sesgo es, en este caso, sobrestimar el riesgo (en otras ocasiones la dirección será la contraria: subestimar el riesgo real). La magnitud del sesgo viene dada por la diferencia entre 6,7 y 3,2 (tabla 5).

**Validez interna y validez externa**

El mensaje central de este artículo se resume en dos tesis: 1) la validez interna de todo estudio es siempre un requisito necesario previo a la valoración de la validez externa<sup>2-8</sup> y 2) generalizar o extrapolar no es en modo alguno una simple cuestión estadística o mecánica, sino un proceso valorativo que exige juzgar qué es relevante y qué es irrelevante en cada caso (desde un punto de vista clínico, biológico, social, etc.)<sup>3</sup>.

Digamos para terminar que actualmente se observa una tendencia a dar más importancia a la validez interna que a la validez externa de los estudios<sup>29</sup>. En los estudios etiológicos, esta tendencia parece plenamente justificada. Por ejemplo, el hallazgo de que fumar aumenta significativamente el riesgo de fallecer de numerosas patologías, realizado por un estudio con un grado alto de validez interna, aunque la población estudiada sean los médicos británicos, es en gran medida generalizable a los médicos españoles, al conjunto de adultos españoles y a otros grupos de población. El hallazgo de que el primer embarazo se asocia a una disminución de larga duración de los niveles plasmáticos de prolactina<sup>30</sup>, aunque fue obtenido en una muestra de mujeres del sureste de Estados Unidos es, en principio, generalizable a muchos otros grupos de mujeres.

Sin embargo, en los estudios sobre la eficacia de las intervenciones médicas (tecnología, cirugía, medicamentos, etc.) la generalización de los resultados es más delicada, ya que el ambiente concreto en el que aquéllas se practican altera notablemente su rendimiento<sup>31,32</sup>.

Por ello, la traducción al propio contexto asistencial de los estudios clínicos con un grado alto de validez interna depende, en última instancia, de cada uno de nosotros<sup>33,34</sup>.

**BIBLIOGRAFIA**

1. Whiting-O'Keefe QE, Henke C, Simborg DW. Choosing the correct unit of analysis in medical care experiments. *Med Care* 1984; 22: 1.101-1.114.
2. Sussner M. Causal thinking in the health sciences. Nueva York: Oxford University Press, 1973.

3. Rothman K. Modern epidemiology. Boston: Little, Brown and Co, 1986.
4. Kleinbaum DG, Kupper LL, Morgenstern H. Epidemiologic research. Principles and quantitative methods. Belmont, CA: Lifetime Learning Publications, 1982.
5. Miettinen OS. Theoretical epidemiology. Nueva York: John Wiley and Sons, 1985.
6. Breslow NE, Day NE. Statistical methods in cancer research. Vol I: the analysis of case-control studies. Lyon: IARC Scientific Publications, 1981.
7. Kelsey JL, Thompson WD, Evans AS. Methods in observational epidemiology. Nueva York: Oxford University Press, 1986.
8. Schlesselman JJ, Stolley PD. Case-control studies. Design, conduct, analysis. Nueva York: Oxford University Press, 1982.
9. Sackett DL. Bias in analytic research. En: Ibrahim MA, Spitzer WO, ed. The case-control study: consensus and controversy. Nueva York, Pergamon Press 1979; 51-63.
10. Jenicek M, Cl eroux R. Epid miologie. Principes, techniques, applications. Paris, Maloine 1984; 236-241.
11. Michael M, Boyce WT, Wilcox AJ. Biomedical bestiary: an epidemiologic guide to flaws and fallacies in the medical literature. Boston: Little, Brown and Co, 1984.
12. Feinstein AR. Clinical epidemiology. The architecture of clinical research. Filadelfia: WB Saunders, 1985; 282-287, 303-307, 317-318, 458-475, 504-513, 618-622.
13. Koepsell TD. Recensi n bibliogr fica de Clinical epidemiology: the architecture of clinical research (Feinstein AR; Filadelfia: WB Saunders, 1985). J Gen Intern Med 1986; 1: 272.
14. Bonfill X, Porta Serra M.  Es necesario un grupo de control? Med Clin (Barc) 1987; 89: 429-435.
15. Alvarez-Dardet C, Bol mar F, Porta Serra M. La medici n de la frecuencia de la enfermedad. Med Clin (Barc) 1987; 88: 287-291.
16. Rogentine GN, Yankee RA, Gart JJ, Nam J, Trapani RJ. HLA antigens and disease: acute lymphocytic leukemia. J Clin Invest 1972; 51: 2.420-2.428.
17. Rogentine GN, Trapani RJ, Yankee RA, Henderson ES. HLA antigens and acute lymphocytic leukemia: the nature of the association. Tissue Antigens 1973; 3: 470-476.
18. Rinsky RA, Smith AB, Hornung R et al. Benzene and leukemia. An epidemiologic risk assessment. N Engl J Med 1987; 316: 1.044-1.050.
19. Alvarez-Dardet C, Bol mar F, Porta Serra M. Tipos de estudios. Med Clin (Barc) 1987; 88: 296-301.
20. Porta Serra M. La observaci n cl nica y el razonamiento epidemiol gico. Med Clin (Barc) 1986; 87: 816-819.
21. Roberts RS, Spitzer WO, Delmore T, Sackett DL. An empirical demonstration of Berkson's bias. J Chronic Dis 1978; 31: 119-128.
22. Berkson J. Limitations of the application of fourfold table analysis to hospital data. Biom Bull 1946; 2: 47-53.
23. Allander E. The prevalence of rheumatoid arthritis: present knowledge and some future directions of research. En: Lawrence RC, Shulman LE, ed. Current topics in rheumatology. Epidemiology of the rheumatic diseases. Proceedings of the Fourth International Conference, National Institutes of Health. Nueva York, Gower Medical Publishing 1984; 127-131.
24. Feinstein AR. Clinical Epidemiology. The architecture of clinical research. Filadelfia, WB Saunders 1985; 469-490.
25. Meinert C. Clinical trials. Design, conduct, analysis. Nueva York: Oxford University Press, 1986; 68-70, 97-100, 200.
26. Porta Serra M. El control de sesgos en el dise o de los estudios epidemiol gicos experimentales. Atenci n Primaria (Barc) 1985; 2: 100-106.
27. Schull WJ, Cobb S. The intrafamiliar transmission of rheumatoid arthritis. J Chronic Dis 1969; 22: 217-222.
28. Miettinen O, Slone D, Shapiro D. Current problems in drug-related epidemiologic research. En: Colombo F, Shapiro S, Slone D, Tognoni G, ed. Epidemiological evaluation of drugs. Littleton, PSG Publishing 1977; 295-303.
29. Alvarez-Dardet C, Mur P, Gasc n E, Nolasco A, Bol mar F. La investigaci n cl nica en Espa a: tipos de dise os utilizados. Med Clin (Barc) 1987; 89: 221-223.
30. Musey VC, Collins DC, Musey PI, Martino-Saltzman D, Preedy JRK. Long-term effect of a first pregnancy on the secretion of prolactin. N Engl J Med 1987; 316: 229-234.
31. Tognoni G, Bonati M. Second-generation clinical pharmacology. Lancet 1986; 2: 1.028-1.029.
32. Porta M. What is the question? The relevance of the explanatory/pragmatic distinction. En: Hartzema AG, Porta M, Tilson HH, ed. Introduction to pharmaco-epidemiology. Cincinnati: Harvey Whitney, 1987 (en prensa).
33. An nimo. The natural laboratory. Lancet 1986; 2: 1.019.
34. Sackett DL, Haynes RB, Tugwell P. Clinical epidemiology: a basic science for clinical medicine. Boston: Little, Brown and Co, 1985; IX-XII.