

medicina, y que para el que comienza puede ser fuente de ansiedad e incertidumbre. Sólo queda recordar que "Aquel que domina la información domina la medicina, o acaso el mundo..."

■ AGRADECIMIENTO

A la Dra. Ana María González, por su colaboración en la redacción de este capítulo. ■

BIBLIOGRAFÍA

1. Hunt DL, Jaeschke R, McKibbin KA. Using Electronic Health Information Resources in Evidence-Based Practice. *JAMA* 2000; 283: 1875-79.
2. Greenhalg T. How to read a paper: the Medline database. *BMJ* 1998; 315: 180-183.
3. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence-based medicine: what it is and what it isn't. *BMJ* 1996; 312: 71-72.
4. ACP-ASIM. Internet site <http://www.acponline.org>.
5. BMJ Clinical Evidence. <http://www.evidence.org>.
6. Bero L, Rennie D. The Cochrane Collaboration: preparing, maintaining, and disseminating systematic reviews of the effects of health care. *JAMA* 1995; 274: 1935-8.
7. Scientific American Medicine website. <http://www.samed.com>.
8. OVID Corporation website. <http://www.ovid.com>.
9. Up To Date website. <http://www.uptodate.com>.
10. MEDLINE Internet sites. <http://www.ncbi.nlm.nih.gov/PubMed>.

CAPÍTULO 4

Conceptos básicos de epidemiología clínica

Gregorio Tomás Obrador

INTRODUCCIÓN

La epidemiología es la ciencia que estudia de forma sistemática la frecuencia, la distribución y los factores que determinan las enfermedades en humanos (B. Mahon y T. F. Pugh). El primer elemento del método epidemiológico consiste en identificar la enfermedad y determinar su frecuencia. El segundo consiste en estudiar su distribución en la población, así como su distribución geográfica y en el tiempo. El conocimiento de la frecuencia y la distribución de la enfermedad lleva a la formulación de hipótesis sobre los posibles factores determinantes o causales de la misma, los cuales deben también ser analizados de forma sistemática. El estudio de la frecuencia y distribución de las enfermedades es el objeto de la epidemiología descriptiva, mientras que el análisis de factores determinantes o causales de las enfermedades es objeto de la epidemiología analítica.

Los estudios del médico británico John Snow sobre las epidemias de cólera de mediados del siglo pasado en Londres, constituyen un buen ejemplo de lo que es el método epidemiológico. Snow observó que las tasas de mortalidad debidas al cólera eran considerablemente mayores en los barrios de Londres que recibían

agua de dos compañías distribuidoras que la obtenían de una zona del río Támesis contaminada por residuos. Entre 1849 y 1854 una de las compañías cambió la fuente de agua a una zona limpia del río. Snow observó que durante esos años las muertes debidas al cólera prácticamente desaparecieron en los barrios que sólo recibían agua de dicha compañía, mientras que las tasas de mortalidad en los barrios que recibían agua de la otra compañía seguían siendo muy elevadas. Estas observaciones de la frecuencia y la distribución de la enfermedad llevaron a Snow a formular la hipótesis de que el cólera se transmite por agua de tomar contaminada, aun cuando el mecanismo preciso (infección por *V. cholerae*) no había sido identificado.

La epidemiología "clínica", como su nombre indica, consiste en la aplicación del método epidemiológico al estudio de problemas clínicos. La epidemiología "clásica" se distingue de la "clínica" en que su objeto fundamental son los problemas de salud pública de las comunidades y no necesariamente los problemas clínicos. El método epidemiológico es como la ciencia básica que permite analizar, de una forma sistemática, los problemas clínicos, con el objeto de obtener conclusiones válidas y ofrecer la mejor evidencia posible que guíe la capacidad de decisión del médico. El espectro de la epidemiología clínica incluye aspectos como la definición de anomalía, la frecuencia, la etiología, el diagnóstico, el pronóstico, el tratamiento y la prevención de las enfermedades. Los eventos de interés de la investigación epidemiológica tradicional, como el desarrollo (pronóstico) o la existencia (diagnóstico) de una enfermedad y la probabilidad de que produzca la muerte (mortalidad), se han expandido, de modo que ahora también se estudia el efecto de la enfermedad en la calidad de vida, el grado de incapacidad o el nivel de insatisfacción del paciente, así como el costo y la óptima utilización de los recursos para el tratamiento y la prevención de las enfermedades.

POBLACIÓN Y MUESTRA

Para llevar a cabo un estudio epidemiológico se requiere de una población o de una muestra. Una población es un grupo grande de individuos con unas características bien definidas. Por ejemplo, todos los habitantes de una región, todos los individuos mayores de 50 años, todos los pacientes que atiende un hospital, etc. Una muestra es un grupo reducido de individuos los cuales han sido seleccionados de una población.

Muchos estudios epidemiológicos se llevan a cabo en muestras debido a su mayor facilidad de ejecución. Para que los resultados sean válidos y se puedan extrapolar, es importante que la selección de la muestra sea representativa de la población que se pretende estudiar.

TIPOS DE ESTUDIOS EPIDEMIOLÓGICOS

En general, los estudios epidemiológicos se clasifican en descriptivos y analíticos (tabla 4-1). Los estudios descriptivos tienen como objeto determinar la frecuencia y la distribución de la enfermedad, mientras que los analíticos se utilizan para examinar la asociación determinante o causal de uno o varios factores con la enfermedad. Los estudios descriptivos proporcionan información de gran importancia para las autoridades sanitarias. Por ejemplo, el conocimiento de la frecuencia y la distribución de las enfermedades en un país o una región determinada, permitieron una distribución eficiente de los recursos para la salud y una planeación adecuada de programas de prevención y de educación. Los estudios descriptivos también frecuentemente proporcionan la primera pista para formular hipótesis sobre los factores determinantes o causales de las enfermedades.

Dentro de los estudios descriptivos se incluyen los estudios correlacionales, los reportes de casos y los estudios transversales. Los estudios correlacionales usan información de poblaciones enteras para comparar las frecuencias de enfermedad entre diferentes grupos durante el mismo periodo de tiempo o en la misma población a diferentes tiempos. Por ejemplo, en un estudio se observó una clara correlación entre el consumo diario *per capita* de carne y cáncer de colon en mujeres de diferentes países. Este tipo de estudios sirve para establecer posibles asociaciones, las cuales deben ser investigadas en más detalle mediante estudios más específicos.

Los reportes de casos consisten en una descripción detallada de un paciente o de un grupo pequeño de pacientes que tienen una enfermedad concreta. Cuando el reporte incluye más de diez pacientes se suele hablar de una "serie" de casos.

Tabla 4-1. Tipos de estudios epidemiológicos

Estudios descriptivos	
• Estudios correlacionales	
• Reporte de un caso	
• Reporte de una serie de casos	
• Estudios transversales o de prevalencia	
Estudios analíticos	
• Observacionales	
Estudios de cohortes	
Estudios de casos y controles	
• Experimentales	
Ensayos clínicos	

Generalmente describen casos de enfermedades nuevas o raras, manifestaciones clínicas inusuales de una enfermedad, o el efecto de terapias nuevas o inusuales.

No es infrecuente que los reportes proporcionen la primera evidencia de que una enfermedad se asocia con la exposición a un factor concreto. Por ejemplo, los primeros reportes de malformaciones congénitas después del uso de talidomida, sugirieron que dicho medicamento tenía efectos teratogénicos, lo cual se comprobó más tarde. Hay que notar, sin embargo, que los reportes de casos tienen limitaciones importantes para que pueda establecerse con certeza que existe una asociación causal o de riesgo. Esto se debe a que la asociación puede ser simplemente debida al azar por el número reducido de casos que típicamente incluyen, y a la falta de un grupo comparativo que no haya sido expuesto al posible factor causal. Por último, los estudios transversales se utilizan para describir la frecuencia (prevalencia) de una enfermedad u otro evento de interés, así como de la exposición a un posible factor determinante o causal en una población y en un momento determinados. Dichos estudios se describen en más detalle en el capítulo 5.

El segundo tipo de estudios epidemiológicos, los estudios analíticos, tienen como objeto examinar de una forma más explícita la asociación determinante o causal de uno o varios factores con la enfermedad. Los estudios analíticos se suelen dividir en observacionales y experimentales. En los estudios observacionales, el investigador simplemente observa la relación entre la ocurrencia de la enfermedad u otro evento de interés y la exposición a uno o varios factores que, se piensa, pueden ser determinantes o causales de la enfermedad.

En este tipo se incluyen los estudios de cohortes y los de casos y controles. Los estudios experimentales se distinguen de los observacionales en que el investigador no simplemente observa, sino que asigna a los individuos al grupo expuesto, o al no expuesto al factor causal o de riesgo y después los sigue respecto a la ocurrencia de la enfermedad o al evento de interés.

El prototipo del estudio experimental es el ensayo clínico. En el capítulo 5 se hace una descripción detallada de los tipos más frecuentes de estudios epidemiológicos, a saber, los estudios transversales o de prevalencia, los de cohortes, los de casos y controles y los ensayos clínicos. También se discuten las ventajas y desventajas de los diferentes tipos de diseños epidemiológicos.

FRECUENCIA Y DISTRIBUCIÓN DE LA ENFERMEDAD

Como se dijo anteriormente, lo primero en epidemiología clínica es determinar la frecuencia y la distribución de una enfermedad u otro evento de interés. Las medidas de frecuencia más comúnmente utilizadas son la prevalencia y la incidencia. La prevalencia es el número de individuos que tienen la enfermedad en

una población, en un momento dado. La prevalencia generalmente se obtiene reanalizando estudios de diseño transversal y expresa la probabilidad de que un individuo esté enfermo en un momento dado. Por ejemplo, en una muestra representativa de individuos de estrato socioeconómico bajo que residían en la ciudad de México y que tenían edades comprendidas entre los 35 y los 64 años, se observó que 304 de los 2,282 individuos encuestados tenían *diabetes mellitus*. Al dividir el número de casos entre el de la población encuestada ($304/2,282 \times 100$) se obtuvo que la prevalencia de *diabetes mellitus* en esa población en el tiempo en que se realizó la encuesta fue del 13.3%. La incidencia, en cambio, se refiere al número de casos nuevos de una enfermedad que ocurren en un periodo de tiempo determinado, en una población que está a riesgo de padecer dicha enfermedad. La incidencia generalmente se obtiene realizando estudios de diseño longitudinal como los de cohortes y expresa la probabilidad de que un individuo desarrolle una enfermedad en un periodo de tiempo determinado (ver capítulo 5).

ASOCIACIÓN

Aunque el espectro de la epidemiología clínica es muy amplio, muchos estudios tratan de identificar los factores que determinan o causan las enfermedades. Los estudios descriptivos frecuentemente proporcionan la primera pista sobre dichos factores, mientras que los estudios analíticos evalúan de una forma más precisa si efectivamente existe una asociación entre un factor y la enfermedad. Asociación es un término estadístico que expresa la dependencia entre dos variables (el factor y la enfermedad). La asociación expresa si la tasa de la enfermedad ($D+$) es mayor o menor en individuos que están expuestos a un factor ($E+$), que en los que no están expuestos ($E-$). Si la tasa de enfermedad es mayor en los expuestos ($D+/E+ > D+/E-$), el factor se considera de riesgo o causal de la enfermedad (más adelante las condiciones para establecer causalidad). Si la tasa de enfermedad es menor en los expuestos, el factor se considera protector contra el desarrollo de la enfermedad ($D+/E+ < D+/E-$). Obviamente, el concepto de asociación no se limita al estudio de factores etiológicos, sino que también se utiliza en estudios que tienen por objeto comparar tratamientos (un medicamento será más efectivo que otro si se asocia con una tasa mayor de curación), métodos diagnósticos (un estudio diagnóstico será más efectivo que otro si detecta la enfermedad con mayor precisión, ocasiona menos molestias al paciente y cuesta menos), etc.

VALIDEZ INTERNA DE LOS ESTUDIOS EPIDEMIOLÓGICOS

Un aspecto esencial en la evaluación de los estudios epidemiológicos es determinar su validez interna; es decir, la probabilidad de que los resultados obtenidos en

la muestra estudiada sean correctos. Se le llama validez "interna" porque se aplica únicamente al grupo de individuos analizado y no necesariamente a otros. La validez interna evalúa si la relación entre la exposición a un factor y la enfermedad u otro evento de interés es real o no. Para concluir que efectivamente la asociación es real se deben excluir explicaciones alternativas, entre las que se incluyen el azar y los sesgos.

■ AZAR

Como se dijo anteriormente, la mayoría de los estudios epidemiológicos se realizan en muestras de la población. Si el tamaño de la muestra es grande, es altamente probable que el resultado sea similar al que se obtendría si se realizara el estudio en toda la población. No obstante, cabe la posibilidad de que la muestra estudiada no sea representativa de la población, simplemente debido al azar. Por ejemplo, si uno lanza una moneda al aire 100 veces, hay teóricamente un 50% de probabilidades de que caiga de un lado. No obstante, tenemos experiencia de que esto no es necesariamente cierto, pues puede variar el número de veces que la moneda caiga de un lado si repetiéramos el lanzamiento de 100 veces en varias ocasiones. Dicha variación es debida al azar.

Pruebas de significancia estadística, como la T de Student o la chi-cuadrada, evalúan la probabilidad de que las diferencias en la frecuencia de la enfermedad o de otro evento de interés entre individuos expuestos y no expuestos a un factor no sean simplemente debidas al azar. Para ello, arbitrariamente se dice que si la probabilidad de observar una diferencia tan extrema, como la que se obtuvo en la muestra estudiada, es menor o igual a 5% (valor de $p < 0.05$), o de 1 en cada 20, es improbable que sea debida al azar. Por el contrario, un valor de $p > 0.05$ indica que la diferencia entre los expuestos y los no expuestos al factor no es estadísticamente significativa; es decir, que no se puede descartar que sea debida al azar. Otra forma de evaluar la posibilidad de que las diferencias observadas sean debidas al azar son los intervalos de confianza, los cuales se explican en detalle en el capítulo 15.

■ SESGOS

Sesgos son errores sistemáticos que se introducen durante el diseño o el desarrollo de un estudio epidemiológico y que invalidan sus resultados. Por ejemplo, si en un estudio se observa que el medicamento A es más efectivo que el medicamento B, el resultado podría ser real, o bien podría deberse a diferencias en las características de los pacientes (sesgo de selección) o a la forma en que se colectó o interpretó la información (sesgo de información). El sesgo de selección ocurre cuando los gru-

pos de pacientes que se están comparando son diferentes en algún aspecto distinto del factor que se está estudiando y que puede influenciar el resultado.

Por ejemplo, la observación de que el medicamento A es más eficaz que el medicamento B podría deberse a que los pacientes que tomaron el medicamento A estaban menos enfermos que los que tomaron el medicamento B, y no a que el medicamento A sea más eficaz que el B.

El sesgo de información ocurre cuando la información que se obtiene de los grupos que se están estudiando no es comparable. Este sesgo puede deberse a que el investigador o los pacientes de los grupos que se están comparando interpretan o reportan la información de manera diferente. Por ejemplo, puede ocurrir que el medicamento A sea nuevo y que estudios previos en grupos pequeños de pacientes mostraron que es bastante efectivo, mientras que el medicamento B es viejo y no del todo bien estudiado. Si los investigadores y los pacientes conocen esta información y saben que medicamento están tomando, podrían tender a interpretar que el medicamento A es más eficaz que el B.

■ FACTORES DE CONFUSIÓN

Otro problema que puede invalidar los resultados de un estudio epidemiológico son los factores de confusión, que consisten en que la asociación entre la exposición a un factor y la enfermedad o evento de interés se debe a un tercer factor (factor confusor) que acompaña a la exposición y, que independientemente de ella, se asocia con el desarrollo de la enfermedad o el evento de interés. Por ejemplo, supongamos que los pacientes que tomaron el medicamento A aumentaron su ingesta de fibra debido a que produce estreñimiento, mientras que los pacientes que tomaron el medicamento B no lo hicieron por no producir este efecto secundario.

Podría ocurrir que el medicamento A es más eficaz en tratar la enfermedad de base (por ejemplo, colon irritable) que el medicamento B, porque propició que los pacientes aumentaran su ingesta de fibra en la dieta, y no porque el medicamento A sea más eficaz *per se*. En este caso, el aumento de la ingesta de fibra acompaña a la toma del medicamento A ("viajan juntos") y mejora los síntomas de colon irritable independientemente del medicamento A.

CRITERIOS DE CAUSALIDAD

Una vez que se ha establecido la validez interna de un estudio, es decir, que se han excluido alternativas como el azar, un sesgo, o un factor confusor, que podrían explicar la asociación entre el factor de riesgo y la enfermedad, se debe evaluar si la naturaleza de la asociación es causal.

Para que un factor se considere causal, es decir, que se asocie con una alta probabilidad de que el sujeto que lo tiene desarrolle la enfermedad, se deben cumplir las siguientes condiciones:

1. La magnitud de la asociación entre el factor de riesgo y la enfermedad debe ser alta; es decir, el riesgo relativo de desarrollar la enfermedad debe ser considerablemente mayor en los individuos que han sido expuestos al factor que en los no expuestos ($D+/E+ \gg D+/E-$).
2. La asociación debe ser factible desde el punto de vista de los mecanismos biológicos que hacen que se produzca.
3. Los resultados obtenidos por varios investigadores respecto de la asociación deben ser consistentes.
4. Debe haber una relación dosis-efecto; es decir, un aumento de la exposición al factor debe acompañarse de un aumento en la probabilidad de desarrollar la enfermedad, y viceversa.
5. La secuencia temporal debe ser tal que la exposición al factor preceda al desarrollo de la enfermedad.

VALIDEZ EXTERNA DE LOS ESTUDIOS EPIDEMIOLÓGICOS

La validez externa se refiere a la probabilidad de que los resultados de un estudio epidemiológico se puedan aplicar a otros grupos de pacientes; es decir, de que se puedan generalizar. Por ejemplo, si el estudio que mostró que el medicamento A es más eficaz que el medicamento B únicamente incluyó mujeres de raza negra, dicho resultado no es necesariamente aplicable a hombres de raza blanca. En cambio, si el estudio incluyó una muestra grande de pacientes de ambos sexos y razas, los resultados del estudio serían aplicables a hombres y mujeres de ambas razas. En la tabla 4-2 se resumen los aspectos esenciales del análisis de estudios epidemiológicos.

CONCLUSIONES

La epidemiología tradicional es la ciencia que estudia de forma sistemática la frecuencia, la distribución y los factores que determinan las enfermedades en humanos. La epidemiología clínica consiste en la aplicación del método epidemiológico al estudio de problemas clínicos.

Tabla 4-2. Análisis de estudios epidemiológicos

Determinar la validez interna del estudio:

- ¿Es la asociación entre la exposición al factor y la enfermedad o el evento de interés válida?
- ¿Podría la asociación ser explicada por otros factores como el azar, sesgos (por ejemplo, de selección o de información) o un factor confusor?

Si la asociación es válida, ¿es la relación de causa-efecto?

- ¿Es la magnitud de la asociación importante?
- ¿Es la asociación factible desde el punto de vista de los mecanismos biológicos que hacen que se produzca?
- ¿Son los resultados del estudio similares a los obtenidos por otros investigadores?
- ¿Hay una relación de dosis-efecto?
- ¿Es la secuencia temporal adecuada?

Determinar la validez externa del estudio:

- ¿Pueden aplicarse los resultados del estudio a otros grupos?

Los estudios epidemiológicos se realizan en una muestra de individuos o en toda la población y se suelen clasificar en descriptivos o analíticos. Estos últimos, a su vez, se clasifican en observacionales o experimentales.

Las medidas más frecuentemente utilizadas de la frecuencia de una enfermedad u otro evento de interés son la prevalencia y la incidencia. Los factores que determinan las enfermedades se estudian comparando dos grupos, uno expuesto y otro no expuesto al posible factor determinante o causal de la enfermedad.

La medida de la asociación o riesgo relativo expresa la dependencia estadística que existe entre el factor y la enfermedad o el evento de interés. Para que la asociación sea causal se deben cumplir una serie de requisitos de causalidad. Para determinar la validez interna de un estudio epidemiológico se deben excluir explicaciones alternativas de la asociación, como el azar, los sesgos y los factores confusores. También se debe analizar si los resultados del estudio se pueden generalizar a otros grupos de pacientes (validez externa). ■

BIBLIOGRAFÍA

1. MacMahon B, Pugh TF. *Epidemiology: Principles and Methods*. Boston, Little, Brown, 1970.
2. Snow J. *On the Mode of Communication of Cholera*. 2a ed., London, Churchill, 1855. Reproducido en *Snow on Cholera*, New York, Hafner, 1965.

3. Hennekens CH, Buring JE. *Epidemiology in Medicine*. Boston, Little, Brown, 1987.
4. Fletcher RH, Fletcher SW, Wagner EH. *Clinical Epidemiology. The Essentials*. 3a ed., Baltimore, Williams & Wilkins, 1996.
5. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. *Clinical Epidemiology: A Basic Science for Clinical Medicine*. 2a ed., Boston, Little, Brown, 1991.
6. Rothman K, Greenland S. *Modern epidemiology*. 2nd ed., Philadelphia, Lippincott-Raven, 1988.
7. González C, Stern MP, Villalpando E, Hazuda H, Haffner S, Lisc E. Prevalencia de diabetes mellitus e intolerancia a la glucosa en una comunidad urbana de nivel económico bajo: Estudio en población abierta. *Rev Invest Clin* 1992; 44: 321.

CAPÍTULO 5

Introducción a los diseños de estudios clínicos

Miguel A. Hernán

INTRODUCCIÓN

La medicina basada en la evidencia requiere información adecuada para la toma de decisiones. Los resultados de estudios epidemiológicos correctamente diseñados constituyen la principal fuente de dicha información. El diseño de cada estudio depende del tipo de datos que se pretende obtener: si el investigador necesita determinar la prevalencia de cáncer de estómago en la población, realizará un estudio de corte transversal; si necesita determinar la eficacia de un nuevo fármaco, intentará llevar a cabo un ensayo clínico. A veces, distintos diseños pueden proporcionar un mismo tipo de información. Por ejemplo, el investigador también podría emplear un estudio de cohortes observacional para evaluar la eficacia del nuevo fármaco.

La elección entre ensayo clínico y estudio de cohortes dependerá de una combinación de factores tales como la validez de los resultados, la ética de la intervención y el costo del estudio.

En este capítulo se revisarán brevemente los diseños más frecuentes que se emplean en epidemiología clínica, divididos aproximadamente en dos grandes grupos: *a*, estudios transversales, cuyo principal objetivo es determinar la frecuencia (prevalencia) de la enfermedad y *b*, estudios longitudinales (ensayo clínico, estudio de cohortes, estudio de casos y controles), cuyo principal objetivo es determinar la frecuencia (riesgo, incidencia) de la enfermedad o su asociación con algún factor.

Se pondrá especial énfasis en describir, a grandes rasgos, las ventajas y desventajas de cada diseño en este segundo grupo.

ESTUDIO TRANSVERSAL

El diseño de un estudio transversal es conceptualmente simple. En primer lugar, se define una enfermedad o un evento de interés (por ejemplo, estar resfriado), una población (por ejemplo, habitantes de Madrid) y una fecha (por ejemplo, 1 de enero del 2000). Después se cuenta el número de personas en la población en esa fecha y el número de éstas que presentan el evento de interés. Un cálculo simple —número de personas resfriadas dividido por habitantes de Madrid, multiplicado por 100— permite obtener la proporción de personas de la población que presentan la enfermedad en la fecha determinada (por ejemplo, 5% de madrileños estaban resfriados el 1 de enero del 2000). Esta proporción es conocida como prevalencia. La prevalencia es la medida de frecuencia de la enfermedad que proporcionan los estudios transversales, a los que a veces se denomina estudios de prevalencia. Por supuesto, además de medir la frecuencia de una enfermedad o un evento, los estudios transversales pueden proporcionar la prevalencia de características (grupo sanguíneo, polimorfismos genéticos) o conductas individuales (uso de tabaco, drogas, fármacos, alcohol). Fuera de la epidemiología, los estudios transversales se usan en muchas otras áreas: encuestas de intención de voto, censos de población, etc.

La principal dificultad de muchos estudios transversales estriba en la imposibilidad, técnica o económica, de incluir a toda la población de interés. Por ejemplo, sería extremadamente costoso, y probablemente inviable, estudiar a la totalidad de los madrileños para obtener la prevalencia del resfriado en la ciudad, en un día determinado. Por esta razón, en la mayor parte de los casos es necesario seleccionar un subgrupo, o muestra, de la población. Esta muestra debe cumplir dos condiciones fundamentales: debe tener la misma prevalencia de la enfermedad que la población a la que representa (es decir, debe ser una muestra representativa) y debe ser lo suficientemente grande como para garantizar una precisión aceptable del resultado (es decir, para que el intervalo de confianza de la prevalencia estimada sea estrecho). Asegurar la obtención de una muestra representativa es el ca-

ballo de batalla de gran parte de los estudios transversales. Un método ampliamente utilizado para ello, es seleccionar la muestra de forma que mantenga las proporciones existentes en la población con respecto a factores asociados con la enfermedad (muestreo estratificado). Por ejemplo, si el resfriado es más común entre fumadores que entre no fumadores, la muestra debería tener la misma proporción de fumadores que la población general (es importante recalcar que la representatividad muestral, así definida, no será un requerimiento para la correcta realización de estudios longitudinales cuyo único interés radique en la identificación de asociaciones entre la enfermedad y algún factor).

La prevalencia es una medida de frecuencia que informa sobre la carga de la enfermedad en un momento determinado. Es, por tanto, esencial para la planificación de servicios sanitarios. Además, una serie temporal de prevalencias en una población (por ejemplo, a intervalos de 5 años) se puede usar para evaluar la evolución de la carga de la enfermedad a lo largo del tiempo e, indirectamente, para evaluar el impacto de una intervención a nivel poblacional. Sin embargo, a pesar de la indiscutible utilidad descriptiva de la prevalencia, la comparación de prevalencias en dos o más grupos presenta serias limitaciones para la identificación de factores de riesgo. Por ejemplo, si un estudio transversal encontrara mayor prevalencia de cáncer de ovario entre mujeres con un polimorfismo genético A que entre mujeres con un polimorfismo B, no podría deducirse de este hallazgo que el riesgo de cáncer de ovario es mayor entre las primeras. De hecho, el riesgo incluso podría ser mayor entre las mujeres con el polimorfismo B. Este sería el caso si una alta proporción de mujeres B desarrollaran el cáncer y éste fuera rápidamente fatal, mientras que las mujeres A, con una baja proporción de casos, desarrollarían un cáncer menos agresivo que no condujera necesariamente a la muerte. Un estudio transversal no detectaría una gran parte de los casos B, porque muchas de las mujeres afectadas habrían fallecido, pero detectaría la mayor parte de los casos A. Por consiguiente, la prevalencia de cáncer sería mayor entre las mujeres A que entre las B, no debido a un mayor riesgo sino a una mejor supervivencia.

Los problemas del uso de la prevalencia para la investigación etiológica derivan de su dependencia de dos factores: la incidencia y la duración media de la enfermedad. Una prevalencia alta no permite distinguir si existe un alto riesgo de desarrollar la enfermedad o si los pocos que la desarrollan permanecen enfermos un largo tiempo, o ambas. Para obviar este problema, es necesario utilizar diseños que proporcionen la incidencia relativa de la enfermedad. A ellos nos dedicamos ahora.

ESTUDIOS LONGITUDINALES

En contraste con el estudio transversal, los estudios longitudinales son ampliamente utilizados en el proceso de inferencia causal. Los tres diseños longitudinales más fre-

cuentes son el ensayo clínico, el estudio de cohortes y el estudio de casos y controles. Antes de describirlos por separado, presentaremos su estructura básica y el tipo de información que pueden proporcionar.

Las versiones más simples de estos diseños consisten en seguir a un grupo de individuos durante un cierto periodo y contar los eventos (por ejemplo, defunciones) que se producen en ese tiempo. Al final del periodo se puede calcular la incidencia del evento en el grupo estudiado. Por ejemplo, en un estudio que incluía a 200 pacientes con cáncer de estómago, se podría haber encontrado que 160 fallecieron tras cinco años de seguimiento. Este 80%—número de fallecimientos dividido por el total de sujetos estudiados, multiplicado por 100— se denomina riesgo, incidencia acumulada o proporción de incidencia. En ocasiones es preferible calcular la tasa de incidencia o número de eventos dividido por el total de persona-tiempo estudiado (el persona-tiempo total es la suma del tiempo que cada sujeto fue seguido). La proporción y la tasa de incidencia son medidas de frecuencia, como la prevalencia descrita más arriba; pero, a diferencia de ésta, no se ven afectadas por la duración de la enfermedad. La elección entre proporción y tasa de incidencia no es siempre opcional y el uso de una u otra dependerá de las características del problema estudiado y de los datos disponibles.

Por ejemplo, algunos estudios longitudinales incluyen individuos que son seguidos durante periodos variables de tiempo (esto puede ocurrir por el propio diseño del estudio o por otras razones tales como el abandono o, en general, las pérdidas de seguimiento de algunos sujetos). En estas situaciones, la proporción de incidencia no puede ser calculada adecuadamente para periodos largos.

En cualquier caso, diseños longitudinales como el ensayo clínico y el estudio de cohortes proporcionan una medida de frecuencia: la (proporción o tasa de) incidencia del evento. Esto puede constituir, en sí mismo, el objetivo del estudio. Sin embargo, la mayor parte de los estudios longitudinales van más allá y no se limitan a describir la incidencia del evento, sino que suelen comparar las incidencias entre dos (o más) grupos de individuos. El hallazgo de incidencias distintas en dos grupos es un primer paso hacia la inferencia causal. Por ejemplo, de un ensayo clínico o un estudio de cohortes puede obtenerse la incidencia de muerte en dos grupos de pacientes con cáncer de estómago: uno tratado con un nuevo fármaco A y el otro con el tratamiento establecido B. A continuación, los investigadores comparan las dos incidencias, bien dividiéndolas (riesgo relativo o tasa relativa) o bien restándolas (diferencia de riesgos o diferencia de tasas). Estas razones o diferencias son denominadas medidas de asociación, y los estudios longitudinales que las producen se conocen a veces con el nombre de estudios de asociación.

La adición de un intervalo de confianza a estas medidas, mediante un simple cálculo estadístico, completa el mínimo análisis requerido para caracterizar una asociación entre el tratamiento (en general, la exposición) y el evento estudiado.

El intervalo de confianza permite evaluar el papel que el azar ha podido jugar en la obtención del resultado.

La obtención de medidas de asociación como razones o diferencia de frecuencias es un proceso sencillo en teoría, pero que puede implicar serias dificultades prácticas. Imaginemos un estudio cuyo objetivo sea determinar si la incidencia de cáncer de ovario es mayor entre mujeres con un polimorfismo genético A que entre personas con un polimorfismo B. Puesto que el cáncer de ovario es relativamente infrecuente, el cálculo de la incidencia entre personas con cada polimorfismo exige el seguimiento de un gran número de personas durante largos periodos, lo que eleva el costo del estudio y puede incluso comprometer su viabilidad. Por este motivo, cuando el objetivo principal del estudio es determinar asociaciónes, sería ideal contar con un diseño que pudiera proporcionar la medida de asociación deseada (por ejemplo, la tasa relativa) sin requerir previamente el cálculo de las medidas de frecuencia en cada grupo (por ejemplo, tasas de incidencia). Afortunadamente, como veremos con más detenimiento, ese diseño existe: el estudio de casos y controles. Este diseño produce una medida de asociación (odds relativa u *odds ratio**) que es un estimador de la tasa relativa, sin necesidad de calcular tasas. Veamos ahora cada uno de los diseños por separado.

■ ENSAYO CLÍNICO

El ensayo clínico es un experimento, es decir, un estudio en el que el investigador interviene en la asignación del tratamiento (en general, exposición) cuyo efecto se desea estudiar. La naturaleza experimental del ensayo clínico le diferencia radicalmente de otros diseños longitudinales, en los que el investigador se limita a observar (de ahí su denominación de estudios observacionales) a individuos que pueden estar o no expuestos por motivos no siempre conocidos. La descripción que sigue se refiere a ensayos clínicos cuyo objetivo es la comparación de dos o más tratamientos (o exposiciones). Aquéllos con otros objetivos (por ejemplo, ensayos clínicos de fase I) no serán considerados.

Otra característica fundamental de un buen ensayo clínico comparativo es que la asignación de la exposición por parte del investigador ocurre al azar: el investiga-

* La traducción del término *odds ratio* es aún motivo de discusión entre los epidemiólogos hispanos. Puesto que el concepto anglosajón de *odds* no tiene equivalente directo en las lenguas latinas, cualquier intento de traducción de *odds ratio* será arbitrario (por ejemplo, oportunidad relativa), si no lingüísticamente cuestionable (por ejemplo, razón de momios). Hasta que se alcance un consenso, hemos preferido mantener, provisionalmente, la palabra *odds* en su lengua original.

dor sortea el tratamiento que cada participante recibirá. A este sorteo se le llama también aleatorización (o, directamente del inglés, randomización). Cuando es posible, el investigador utiliza un sistema de sorteo que impide a los participantes saber a qué grupo de tratamiento han sido asignados (diseño ciego) y que, además, impide a los médicos saber qué tratamiento administran a cada paciente (diseño doble ciego). En breve describiremos las ventajas del sorteo y del diseño ciego.

El prototipo de ensayo clínico al que nos referimos es, pues, un estudio longitudinal en el que el investigador selecciona a un grupo de individuos (mediante estrictos criterios de inclusión/exclusión) y sorteas al azar el grupo de tratamiento al que cada individuo pertenecerá. Idealmente, este sorteo se realiza de forma doblemente ciega. Tras administrar a cada individuo el tratamiento que le ha tocado en la lotería, se le sigue durante un tiempo predeterminado, y se registra si el tratamiento ha tenido o no éxito. A continuación se calcula la incidencia de éxito o fracaso (medida de frecuencia) en cada grupo de tratamiento. Finalmente, se comparan las incidencias en cada grupo de tratamiento mediante una medida de asociación, que puede ser una razón de medidas de frecuencia (por ejemplo, riesgo relativo) o una diferencia (por ejemplo, diferencia de riesgos), y su intervalo de confianza. Una variación común es que el efecto del tratamiento se mide mediante una variable continua (por ejemplo, presión arterial), en lugar de mediante una variable dicotómica (éxito: sí, no). En este caso se comparan, en vez de las incidencias, las medias o medianas de esta variable en cada grupo. El diseño de ensayo clínico que hemos descrito se llama paralelo, en contraposición a los ensayos clínicos cruzados, que se caracterizan por estar divididos en dos periodos de forma que cada participante recibe, secuencialmente pero en orden aleatorio, los dos tratamientos estudiados.

Consideremos ahora el siguiente escenario: un investigador quiere estimar el efecto causal de un nuevo tratamiento en la curación de una enfermedad. Para ello, diseña un ensayo clínico (paralelo) en el que divide al azar a mil pacientes en dos grupos de igual tamaño. Administra el tratamiento nuevo a los individuos en el primer subgrupo y el tratamiento tradicional a los otros. Si la incidencia de curación es claramente mayor entre los pacientes tomando el nuevo tratamiento que entre los otros, mucha gente aceptará que es probable la existencia de una relación causal entre el tratamiento y la curación. ¿Por qué se considera al ensayo clínico el diseño más apropiado para establecer relaciones causales? El motivo más importante es, precisamente, el sorteo del tratamiento. Puesto que cualquier individuo ha tenido la misma probabilidad de recibir el nuevo tratamiento, no puede alegarse que los pacientes tratados con él eran distintos de los no tratados. Es decir, los pacientes que recibieron el nuevo tratamiento no tenían un mejor pronóstico al entrar en el estudio, no fumaban menos o eran más jóvenes que los del otro grupo. La misma proporción de pacientes con mejor pronóstico, no fumadores o jóvenes debería existir en ambos grupos, porque el tratamiento ha sido asignado al azar. Los pacientes de uno y otro grupo son comparables en todo menos

en el tratamiento que recibieron y, por tanto, cualquier diferencia respecto a la incidencia de curación tiene que ser debida a aquello en lo que difieren. Evidentemente, la asignación aleatoria del tratamiento no garantiza que ambos grupos tendrán exactamente las mismas proporciones de fumadores o de jóvenes. Por ejemplo, por puro azar puede ocurrir que un grupo tenga un 42% de fumadores frente a un 39% en el otro, pero estas diferencias tenderán a ser pequeñas si se sorteas a un número alto de pacientes. En caso de que existan desequilibrios sustanciales entre grupos de tratamiento con respecto a factores pre-tratamiento que afectan a la probabilidad de curación (pronóstico, tabaco, edad, etc.), nuestra medida de asociación estará confundida; es decir, no se podrán atribuir las diferencias en incidencia al efecto del tratamiento. Por esta razón, estos factores reciben el nombre de confusores. La asignación aleatoria del tratamiento es el mejor arma de que dispone el investigador para luchar contra la confusión del efecto del tratamiento.

A pesar del sorteo, que minimiza la posibilidad de confusión, aún cabe la posibilidad de que el efecto del tratamiento no pueda ser correctamente estimado por otras causas. Por ejemplo, si el paciente sabe qué tratamiento está recibiendo, puede modificar su conducta: podría comenzar un estilo de vida más saludable o iniciar tratamientos concomitantes para compensar lo que el percibe como un tratamiento experimental de dudosa eficacia o, al contrario, puede estar tan seguro de la utilidad del tratamiento que descuida otros aspectos de su estilo de vida. En cualquier caso, las diferencias en incidencia no serían únicamente debidas al tratamiento en sí y, por tanto, la medida de asociación será un estimador sesgado del efecto del tratamiento. Algo similar podría ocurrir si el médico conociera el tratamiento que está administrando y decidiera, inconscientemente, prestar más atención a aquellos pacientes que reciben el tratamiento nuevo o, por el contrario, a los que reciben un peligroso tratamiento tradicional. La forma más eficaz de luchar contra este sesgo es el diseño doble ciego.

Hemos visto que la aleatorización y el ciego colocan al ensayo clínico, en comparación con los estudios observacionales, en una posición privilegiada para la inferencia causal. Dadas las ventajas de este diseño, uno esperaría encontrar a la mayor parte de los investigadores concentrados en organizar ensayos clínicos únicamente. Sin embargo, los ensayos clínicos constituyen un porcentaje minoritario de los estudios epidemiológicos que se llevan a cabo. ¿Por qué? Hay cuatro razones fundamentales por las que el ensayo clínico no es utilizado siempre:

1. **Viabilidad.** Muchas de las preguntas que los investigadores formulan no pueden ser traducidas a un experimento convincente. Por ejemplo, consideremos a un investigador interesado en estimar el efecto protector del uso de preservativos sobre la transmisión del VIH en hombres. Nuestro investigador podría reclutar a un grupo de hombres VIH-negativos que no hubieran usado nunca preservativo en sus relaciones sexuales. Podría incluso repartir

a estos hombres al azar en dos grupos (uso de preservativo: sí, no), estudiar los dos años y calcular la incidencia de infección por VIH en cada grupo. Pero, ¿cómo podría asegurarse que las diferencias, si es que hay alguna, se deben exclusivamente al uso del preservativo? La pregunta causal que el investigador estaba formulando no puede ser adecuadamente tratada por medio de un ensayo clínico tradicional. La probable falta de cumplimiento con el tratamiento asignado y la imposibilidad de un diseño ciego limitan los objetivos del estudio a la estimación del efecto de un programa de educación sobre el uso del preservativo, algo muy distinto de lo que el investigador pretendía originalmente. Ejemplos aún más radicales de no viabilidad (al menos actualmente) de experimentos aleatorizados, son aquellos estudios en los que se pretende elucidar el efecto causal de ciertos polimorfismos genéticos o del sexo.

2. **Ética.** Incluso si pudieran cristalizarse en un ensayo clínico viable, muchas, si no la mayoría, de las preguntas que los investigadores formulan no pueden ser traducidas a un experimento éticamente aceptable. Por ejemplo, todos los estudios cuyo objetivo es estimar el efecto causal de algún factor de riesgo (tabaco, alcohol, residuos tóxicos, etc.) en humanos son impensables desde el punto de vista experimental. Los experimentos nazis condujeron a la adopción de normas internacionales que restringen la experimentación con seres humanos y enfatizan la necesidad de que la participación de los sujetos experimentales sea voluntaria. En la mayor parte de los países, cualquier ensayo clínico precisa obtener la aprobación de un comité ético que vela por el cumplimiento de estas normas. Además, los participantes deben recibir información adecuada sobre los riesgos y beneficios que pueden derivar de su participación en el estudio y pueden abandonarlo cuando así lo deseen (esto puede crear problemas metodológicos adicionales).

3. **Validez externa.** Muchos ensayos clínicos, a pesar de ser viables y éticamente irreprochables, proporcionan respuestas incompletas a las preguntas formuladas. Esto es generalmente debido a los estrictos criterios de inclusión/exclusión que limitan la selección de sujetos en estos estudios. Por ejemplo, muchos ensayos clínicos restringen la participación de mujeres en edad fértil debido a los posibles efectos teratogénicos de los fármacos estudiados. Otros limitan la participación a pacientes con alta probabilidad de permanecer vivos durante la duración del estudio. Estas restricciones hacen difícil la extrapolación de los resultados del estudio a la población que podría beneficiarse de ellos.

4. **Costo.** En general, los ensayos clínicos son más caros que estudios observacionales comparables en tamaño, ya que es necesario añadir la asignación y administración del tratamiento, el mantenimiento del ciego, etc.

Por estas razones, los ensayos clínicos han quedado relegados al estudio de exposiciones supuestamente beneficiosas, pero no claramente mejores o peores que aquéllas con las que se comparan. Los estudios observacionales son necesarios para el resto de la investigación biomédica y para completar los resultados de los ensayos clínicos. En un capítulo posterior se dan mayores detalles de este diseño. Veamos ahora los diseños observacionales más frecuentemente usados.

■ ESTUDIO DE COHORTES

El estudio de cohortes es el primo no experimental (observacional) del ensayo clínico comparativo. Es decir, el investigador identifica dos grupos de individuos (expuestos y no expuestos), los sigue durante un periodo de tiempo, calcula la incidencia del evento de interés por separado en cada grupo y, finalmente, compara las incidencias mediante el cálculo de una medida de asociación. Por supuesto, puede haber más de dos grupos definidos por distintos niveles de exposición. La variable objetivo de algunos estudios de cohortes puede ser cuantitativa (por ejemplo, presión arterial) y, en ese caso, se comparan las medias o medianas de esa variable.

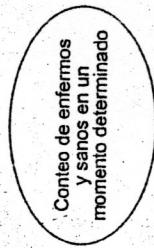
La mayor diferencia con el ensayo clínico radica en la asignación del tratamiento: en el estudio de cohortes el investigador no interviene en este proceso. Esto abre la puerta a la confusión del efecto del tratamiento por factores que determinen, parcialmente al menos, tanto el nivel de exposición que los sujetos reciben como su riesgo de enfermedad. Por ejemplo, si un estudio de cohortes halla una asociación inversa entre consumo de café y el riesgo de desarrollar la enfermedad de Parkinson, una pregunta inmediata será si esa asociación no estará confundida por el uso de tabaco (los fumadores tienen una menor incidencia de enfermedad de Parkinson y suelen beber más café que los no fumadores).

La lucha contra la confusión se basa generalmente en recoger información sobre posibles confusores para calcular la medida de asociación dentro de cada nivel de éstos. Por ejemplo, en el estudio descrito en el párrafo anterior, podría calcularse la tasa relativa de enfermedad de Parkinson en bebedores de café respecto a no bebedores de café, pero separadamente para fumadores y no fumadores. Si se desea una mejor eliminación del posible efecto confusor del tabaco, se puede dividir a los fumadores en subgrupos (de acuerdo a la cantidad de tabaco que fuman, por ejemplo) y calcular la tasa de incidencia dentro de cada grupo. A este proceso se le llama estratificación y es un método habitualmente utilizado para luchar contra la confusión del efecto del tratamiento. Un paso final consiste en calcular una media ponderada de las tasas relativas en cada estrato del confusor para obtener un único número: la tasa relativa ajustada. Los modelos de regresión son una mera extensión del análisis estratificado. En estudios más complejos, se

necesitan métodos alternativos (por ejemplo, los modelos causales desarrollados por Robins) para conseguir un control adecuado de la confusión.

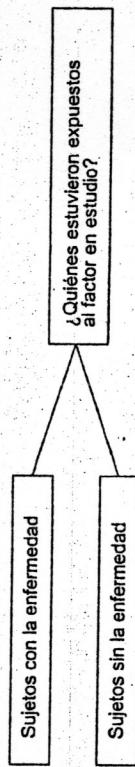
Por tanto, los estudio de cohortes (y en general, todos los observacionales) imponen una condición ineludible para asegurar la interpretación causal de la medida de asociación: el investigador debe conocer los factores de confusión, recoger información sobre ellos y ajustar los análisis de forma apropiada. Este requerimiento adicional coloca en posición de desventaja al estudio de cohortes respecto al ensayo clínico, ya que es imposible demostrar que la condición es cierta. Sin embargo, este es el precio que debe pagar el investigador para llevar a cabo estudios que no pueden diseñarse en forma de experimentos. En la figura 5-1 se esquematizan los diferentes diseños observacionales.

Estudio transversal



Población muestra

Estudio de casos y controles



Estudio de cohortes

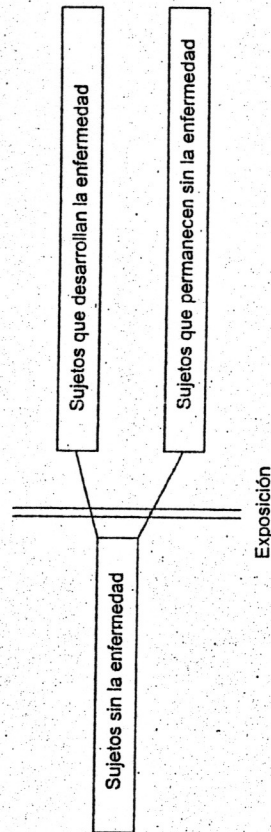


Figura 5-1. Representación esquemática de los diferentes diseños observacionales.

Algunos estudios utilizan cohortes definidas en el pasado. Por ejemplo, supongamos que se quiere estimar el riesgo relativo de cáncer entre trabajadores de una central nuclear que manejan sustancias radiactivas, comparados con los que no tienen contacto con esas sustancias. El investigador podría utilizar los archivos de la central para identificar la cohorte de trabajadores contratados entre 1970-80 y su puesto de trabajo. Después podría acudir al registro de cáncer para identificar quiénes desarrollaron la enfermedad. Este estudio de cohortes retrospectivo es, en principio, tan válido como el estudio de cohortes prospectivo descrito más arriba. Sin embargo, suelen existir grandes dificultades prácticas para llevarlo a cabo (información incompleta sobre la exposición, la enfermedad, los confusores o, incluso, sobre quiénes pertenecían a la cohorte de interés).

Hasta ahora nos hemos referido a estudios de cohortes cerradas, es decir, cohortes cuya composición queda determinada al principio del seguimiento y a las que sus miembros no pueden dejar de pertenecer a menos que se consideren pérdidas de seguimiento. Existen también estudios de cohortes abiertas, de las que los sujetos pueden entrar y salir, pero su descripción excede los límites de este capítulo.

En el siguiente apartado veremos que los estudios de casos y controles presentan dificultades adicionales comparados con los estudios de cohortes. Sin embargo, los investigadores prefieren muchas veces diseñar estudios de casos y controles. La razón es simple: muchas enfermedades de interés (infarto, cáncer, esclerosis múltiple, etc.) suelen ser relativamente raras y su estudio exigiría el seguimiento de cohortes enormes, lo que plantea serios problemas logísticos y económicos. Los estudios de casos y controles, de diseño más complejo, disminuyen estos problemas, siempre y cuando el investigador esté dispuesto a pagar un precio.

ESTUDIO DE CASOS Y CONTROLES

El estudio de casos y controles se basa en el hecho de que todos los casos provienen de alguna cohorte. En lugar de invertir recursos en estudiar detenidamente a esa cohorte, bastaría con diagnosticar a los nuevos casos que se producen y, después, identificar a una muestra aleatoria de miembros de la cohorte (los "controles" que no tengan la enfermedad en el momento en que los casos son diagnosticados. Una vez seleccionados casos y controles, se obtiene información de cada individuo (exposición y confusores), y se calcula una medida de asociación: la odds relativa (razón de momios) que es un estimador insesgado de la tasa relativa. La comparación de casos y controles debería producir, prácticamente, los mismos resultados que el estudio de la cohorte entera, ya que los casos son comunes al estudio de casos y controles y a la cohorte, y los controles en el estudio de casos son representativos (en lo que a la distribución de la exposición se refiere) del persona-tiempo de la cohorte.

enfermedad o síndrome y se describe su evolución clínica desde el diagnóstico hasta la cura, muerte o estado crónico. En general, las series de casos son un método rudimentario sin un grupo control, en el que no se pueden hacer análisis de inferencia para extraer conclusiones de causalidad o asociación. Sin embargo, en contadas ocasiones las series de casos pueden generar hipótesis importantes que conduzcan a un agente etiológico. Por ejemplo, los primeros casos de síndrome de inmunodeficiencia adquirida representaron para los médicos una entidad muy rara, pero con características muy similares y con ciertas peculiaridades. Tales características hicieron pensar desde un principio que se trataba de un agente infeccioso, muy probablemente transmitido de manera sexual. Las primeras recomendaciones preventivas se dieron aun antes de identificarse el virus causal.

Finalmente, es necesario aclarar que la presentación de muchos conceptos de este capítulo ha sido simplificada por razones didácticas. La realidad puede ser algo más complicada. Por ejemplo, algunos estudios de casos y controles están libres de sospecha de sesgo de selección ya que los controles se seleccionan de una cohorte perfectamente identificada (estudios de casos y controles anidados). Por otro lado, las pérdidas de seguimiento, o la falta de cumplimiento con el tratamiento asignado, pueden inducir sesgo de selección en ensayos clínicos y estudios de cohortes. Este capítulo de introducción es un esqueleto compuesto por conceptos simplificados, sobre los que el lector puede organizar los contenidos de los siguientes capítulos. ■

BIBLIOGRAFÍA

1. Rothman KJ, Greenland S. *Modern Epidemiology*. Philadelphia, Lippincott-Raven, 1998.
2. Wacholder S, McLaughlin JK, Silverman DT, Mandel JS. Selections of controls in case-control studies. I. Principles. *Am J Epidemiol* 1992; 135: 1019-28.
3. Wacholder S, Silverman DT, McLaughlin JK, Mandel JS. Selections of controls in case-control studies. II. Types of controls. *Am J Epidemiol* 1992; 135: 1029-41.
4. Tsuang MT, Hsieh C-C, Fleming JA. Odds ratio and its interpretation in case-control studies. In: Hsu LKG, Hensen M (eds.). *Research in Psychiatry: Issues, Strategies, and Methods*. New York, Plenum Press, 1992: 125-32.
5. Robins JM. Causal inference from complex longitudinal data. In: Berkane M (ed.). *Latent Variable Modeling and Applications to Causality*. Lecture Notes in Statistics 120. New York, Springer-Verlag, 1997: 69-117.

CAPÍTULO 6

Interpretación de pruebas diagnósticas

Mario Enrique Rendón Macías

INTRODUCCIÓN

El diagnóstico clínico consiste en la integración de la sintomatología (padecimiento) del paciente en una entidad fisiopatológica conocida. La precisión del diagnóstico en la práctica médica permite la toma de decisiones terapéuticas adecuadas, que pueden ir desde la indicación de un(os) medicamento(s) o una(s) maniobra(s) terapéutica(s), hasta la simple observación del comportamiento de la enfermedad.

Los datos clínicos con los que el médico realiza el proceso diagnóstico están sujetos a inconsistencias derivadas de las técnicas subjetivas que utiliza al recopilarlos. La información suele llevar consigo un grado variable de incertidumbre diagnóstica, a consecuencia de las variaciones biológicas en los propios pacientes, así como de la interpretación de los síntomas y signos obtenidos por el médico.

Para aumentar la precisión diagnóstica se realizan estudios paracrínicos de laboratorio y gabinete, también conocidos como pruebas diagnósticas. No obstante su utilidad, su uso indiscriminado puede resultar en una alta frecuencia de

resultados falsos positivos. La presencia de datos anormales en un paciente sano o viceversa puede tener consecuencias igual o más graves que la decisión de no realizarlos. Además, el auge de la tecnología así como la presión social ejercida sobre los médicos por las demandas de la mala práctica, ha hecho que en los últimos años se hayan desarrollado nuevas pruebas, las cuales no siempre se asocian con una mayor precisión diagnóstica y sí con un mayor costo. Es por ello, que el médico debe de mantener una actitud crítica hacia la aceptación de nuevas pruebas diagnósticas y conocer la metodología que le permita evaluar los beneficios en cuanto a validez, riesgos y costo de las mismas.

DISEÑO DE LOS ESTUDIOS DIAGNÓSTICOS

El objetivo primario del diseño es determinar la validez de una prueba diagnóstica, es decir su capacidad de discriminar entre pacientes enfermos y sanos. Esto se logra con estudios de diseños transversales comparativos, donde la prueba es aplicada en dos grupos de individuos, los enfermos y los sanos o sujetos con otras enfermedades diferenciales. Si la prueba reconoce sin equivocación el estado de enfermedad o de salud del individuo se dice que es perfecta o exacta.

Para determinar la validez de una prueba diagnóstica es necesario compararla con otra que se considere el estándar-ideal o patrón de oro. El estándar de oro es el conjunto de criterios que establecen de forma definitiva la presencia o no de una enfermedad. La comparación de las pruebas diagnósticas debe realizarse de forma cegada en cuanto al estado de salud o enfermedad por quienes realizan o interpretan los diferentes estudios.

INTERPRETACIÓN DE LOS ESTUDIOS DIAGNÓSTICOS CON RESULTADOS DICOTÓMICOS (POSITIVO O NEGATIVO)

El siguiente ejemplo nos ayudará a exponer varios conceptos. Una mujer de 45 años acude a un consultorio ginecológico por notar un tumor en su seno derecho. Después de una evaluación clínica cuidadosa, el médico sospecha una neoplasia, pero la información clínica es insuficiente para determinar si es maligna o benigna. El recurso diagnóstico ideal es la biopsia con estudio histopatológico del tejido. Como el procedimiento conlleva ciertos riesgos, el médico se pregunta si otras pruebas no invasivas como la mastografía y la centelleografía mamarias serían útiles para hacer el diagnóstico, cual de las dos tiene mayor validez, confia-

bilidad, tiene menos efectos secundarios y cuesta menos. En otras palabras, ¿qué estudio sería la mejor opción?

Para su respuesta, se llevó cabo un estudio que comparó el desempeño de ambas pruebas en pacientes con y sin tumores malignos de mama. Para determinar, fue necesario reunir un grupo de mujeres en quienes la condición de malignidad o benignidad se estableció con un criterio inequívoco como el estudio histopatológico (estándar de oro). El sesgo de sospecha diagnóstica y evolución se controlaron con la realización de ambos procedimientos en forma simultánea y paralela antes de la realización de la biopsia. Los médicos que realizaron e interpretaron la mastografía y la centelleografía mamaria estaban cegados a la condición de salud o enfermedad de las pacientes al momento de las pruebas. La evaluación de las imágenes se hizo con criterios previamente establecidos para cada técnica y por personal experto.

Los resultados se resumen en una tabla de 2 x 2, designándose a cada cuadro una letra: "a", "b", "c" o "d" (tabla 6-1). En el cuadro "a" se anota el número de mujeres con neoplasia mamaria maligna en quienes la prueba fue reportada positiva para malignidad (verdaderas positivas). En el cuadro "b" se anota el número de mujeres diagnosticadas falsamente como portadoras de la enfermedad (falsas positivas). En el cuadro "c" se anota el número de mujeres erróneamente calificadas como sanas, dado que tienen un tumor maligno (falsas negativas). En el cuadro "d" se anota el número de mujeres en quienes se descartó correctamente la ausencia de enfermedad maligna (verdaderas negativas).

Los criterios de validez de una prueba diagnóstica incluyen la exactitud, la sensibilidad (sen), la especificidad (esp) y los valores predictivos positivo (VP+) y negativo (VP-). Los tres primeros son propiedades inherentes de la prueba diagnóstica, y reflejan su eficacia. Como se ve en la tabla 6-1, hay dos casillas en que las observaciones fueron veraces (a y d) y dos erróneas (b y c). La exactitud de la prueba es el porcentaje de aciertos con el estándar de oro $[(a+d)/(a+b+c+d) \times 100]$. En este ejemplo, la centelleografía fue más exacta que la mastografía (96 vs 69%). La sensibilidad evalúa el funcionamiento de la prueba diagnóstica en individuos enfermos y se define como el porcentaje de pacientes con la enfermedad que salen positivos a la prueba (a/a+c). En el ejemplo, todas las mujeres con una neoplasia mamaria maligna fueron detectadas por ambos métodos, dando así una sen del 100%; es decir, tienen gran valor para detectar a las mujeres con cáncer de mama. Este criterio tiene aplicación sobre todo para epidemiólogos y profesionales de la salud pública porque permite el rastreo de casos en las poblaciones. La especificidad (esp) evalúa el desempeño de la prueba en la población sana y se define como el porcentaje de sujetos sin la enfermedad que salen negativos a la prueba (d/b+d). En el ejemplo, la mastografía realizada en mujeres con tumores benignos reportó benignidad en 54% de los casos, mientras que la centelleo-

Tabla 6-1. Desempeño diagnóstico de la mastografía y la centelleografía mamaria para el diagnóstico de neoplasia mamaria maligna al compararse con el reporte histopatológico obtenido por biopsia

Reporte de histopatología		Reporte de histopatología	
	Positivo para neoplasia maligna	Negativo para neoplasia maligna	Total
Mastografía sugestiva de neoplasia maligna	19	17	36
Mastografía sugestiva de neoplasia benigna	0	20	20
Total	19	37	56
Sensibilidad = $a / a + c = 19 / 0 + 19 = 1$ o 100% Especificidad = $d / b + d = 20 / 17 + 20 = 0.54$ o 54% Exactitud = $a + d / a + b + c + d = 19 + 20 / 19 + 17 + 0 + 20 = 0.69$ o 69% VP(+) = $a / a + b = 19 / 19 + 17 = 0.52$ o 52% VP(-) = $d / c + d = 20 / 0 + 20 = 1$ o 100% Prevalencia = $a + c / a + b + c + d = 19 + 0 / 19 + 17 + 0 + 20 = 0.33$ o 33%			
Reporte de histopatología		Reporte de histopatología	
	Positivo para neoplasia maligna	Negativo para neoplasia maligna	Total
Centelleografía sugestiva de neoplasia maligna	19	2	21
Centelleografía sugestiva de neoplasia benigna	0	35	35
Total	19	37	56
Sensibilidad = $a / a + c = 19 / 0 + 19 = 1$ o 100% Especificidad = $d / b + d = 35 / 2 + 35 = 0.94$ o 94% Exactitud = $a + d / a + b + c + d = 19 + 35 / 19 + 2 + 0 + 35 = 0.96$ o 96% VP(+) = $a / a + b = 19 / 19 + 2 = 0.90$ o 90% VP(-) = $d / c + d = 35 / 0 + 35 = 1$ o 100% Prevalencia = $a + c / a + b + c + d = 19 + 0 / 19 + 17 + 0 + 20 = 0.33$ o 33%			

grafía lo hizo en 94%. Si el objetivo fuera detectar las mujeres con tumores mamarios malignos en una población, las dos pruebas serían útiles; pero, la centelleografía tendría más utilidad para descartar mujeres con tumores benignos. La sensibilidad y la especificidad de una prueba diagnóstica se calculan conociendo *a priori* quien tiene la enfermedad y quien no, y por tanto son de poca utilidad clínica. En la práctica médica, sin embargo, los pacientes acuden con datos clínicos y el médico pide estudios paraclínicos para confirmar un diagnóstico. Los VP+ y VP- de una prueba diagnóstica evalúan su capacidad para determinar que la enfermedad existe si el resultado es positivo, o que no existe si es negativo.

El valor predictivo positivo (VP+) es la probabilidad de que un sujeto tenga la enfermedad si la prueba diagnóstica es positiva ($VP+ = a/a+b$). En el ejemplo, la mastografía y la centelleografía tuvieron un VP+ de 52 y 90%, respectivamente. Esto significa que el médico obtendrá un mayor beneficio diagnóstico con una centelleografía positiva que con una mastografía positiva. Con la centelleografía, la probabilidad de un resultado falso positivo, es decir, de que la mujer tenga un tumor benigno a pesar de que la prueba es positiva para malignidad, es de un 10%, mientras que con la mastografía es de 48%. El valor predictivo negativo (VP-) es la probabilidad de que un sujeto no tenga la enfermedad si la prueba es negativa. En el ejemplo, ambas pruebas mostraron un VP- de 100%, por lo que mujeres con resultados negativos dan la seguridad de que no tienen una neoplasia maligna.

INTERPRETACIÓN INDIVIDUAL DE UNA PRUEBA DIAGNÓSTICA

La probabilidad de que una prueba diagnóstica detecte la enfermedad está relacionada con la prevalencia de la misma o de los datos clínicos manifestados por el paciente en dicha enfermedad. Esto se conoce como probabilidad preprueba o *a priori*. En general, cuando la prevalencia de la enfermedad (probabilidad preprueba, ppop) es alta, el VP+ es mayor, mientras que si es baja el VP+ (probabilidad posprueba, ppop) es menor.

En el ejemplo se evaluó el desempeño de ambas pruebas en un grupo de 56 mujeres, 19 con cáncer de mama y 36 con tumores benignos. La prevalencia de cáncer de mama en este estudio fue del 33% ($19/56 = 0.33$). Con esta prevalencia (ppop), el VP+ (ppop) de la mastografía y de la centelleografía fueron de 52 y 90%, respectivamente.

Si la ppop de una enfermedad en la población no se conoce, es útil emitir una prevalencia basada en la información disponible en la literatura médica. En nuestro caso, se estimó que la prevalencia del cáncer de mama en la población es de un 10%.

Con este dato podemos recalcular la tabla de 2 x 2, cambiando el número de pacientes en cada uno de los cuadros, basados en los datos de la prevalencia (10%), la sensibilidad y la especificidad obtenidos previamente (no olvidar que la sensibilidad y la especificidad son propiedades inherentes a las pruebas diagnósticas, y por tanto no cambian con la prevalencia).

En la tabla 6-2 se muestra cómo el VP+ de ambas pruebas disminuye al disminuir la prevalencia.

Tabla 6-2. Desempeño diagnóstico de la mastografía y la centelleografía mamaria para el diagnóstico de neoplasia mamaria maligna al disminuir la prevalencia de enfermos

Reporte de histopatología			
	Positivo para neoplasia maligna	Negativo para neoplasia maligna	Total
Mastografía sugestiva de neoplasia maligna	19	79	98
Mastografía sugestiva de neoplasia benigna	0	92	92
Total	19	171	190

$Sensibilidad = a / a + c = 19 / 0 + 19 = 1$ o 100%
 $Especificidad = d / b + d = 92 / 79 + 92 = 0.54$ o 54%
 $Exactitud = a + d / a + b + c + d = 19 + 92 / 19 + 79 + 0 + 92 = 0.58$ o 58%
 $VP(+)= a / a + b = 19 / 19 + 79 = 0.19$ o 19%
 $VP(-)= d / c + d = 92 / 0 + 92 = 1$ o 100%
 $Prevalencia = a + c / a + b + c + d = 19 + 0 / 19 + 79 + 0 + 92 = 0.10$ o 10%

Reporte de histopatología			
	Positivo para neoplasia maligna	Negativo para neoplasia maligna	Total
Centelleografía sugestiva de neoplasia maligna	19	10	29
Centelleografía sugestiva de neoplasia benigna	0	161	161
Total	19	171	190

$Sensibilidad = a / a + c = 19 / 0 + 19 = 1$ o 100%
 $Especificidad = d / b + d = 161 / 10 + 161 = 0.94$ o 94%
 $Exactitud = a + d / a + b + c + d = 19 + 161 / 19 + 10 + 0 + 161 = 0.94$ o 94%
 $VP(+)= a / a + b = 19 / 19 + 10 = 0.65$ o 65%
 $VP(-)= d / c + d = 35 / 0 + 35 = 1$ o 100%
 $Prevalencia = a + c / a + b + c + d = 19 + 0 / 19 + 10 + 0 + 171 = 0.10$ o 10%

torio reporta el resultado como anormal (cifras que representan valores superiores o inferiores a los encontrados en 95% de la población sana). Una segunda opción es calcular la probabilidad diagnóstica de que el sujeto esté enfermo dada la cifra obtenida en la prueba.

PRIMERA OPCIÓN: ESTABLECER UN NIVEL DE CORTE

Para explicar este concepto, revisaremos un estudio que evaluó la validez del calcio sérico total y calcio calculado para el diagnóstico de hipocalcemia en unidades de terapia intensiva neonatal. El calcio sérico iónico se utilizó como estándar de oro, puesto que representa el calcio metabólicamente activo y directamente relacionado con las manifestaciones clínicas de la hipocalcemia. El calcio sérico iónico no siempre está disponible pero sí el calcio sérico total. En la tabla 6-3A se observa que las medianas del calcio sérico total e iónico son diferentes en neonatos con y sin hipocalcemia ($p < 0.05$). En estas pruebas, los autores dividieron los resultados como positivos y negativos con relación a una cifra mínima de acuerdo con lo reportado en la literatura. Con esta estrategia se calculó la sen, esp y los VP+ de cada prueba, como se indicó en el apartado anterior (tabla 6-3B). Otra manera de determinar con exactitud el nivel óptimo de corte es mediante la

Tabla 6-3 A. Cifras obtenidas en los diferentes métodos de valoración de calcemia en recién nacidos con y sin hipocalcemia iónica

Variable	Con hipocalcemia. N = 21 neonatos		Sin hipocalcemia. N = 32 neonatos		p
	Mediana	Intervalo	Mediana	Intervalo	
Calcio iónico (mM/L)	0.93	0.5 - 1.03	1.23	1.13 - 1.84	0.001
Calcio total (mg/dL)	8.1	5.8 - 11.1	9.2	6.1 - 15.5	0.003

* Calcio iónico considerado normal 1.10 mM/L.

B. Evaluación de las pruebas diagnósticas de hipocalcemia en neonatos usando los puntos de corte de la literatura (LIT) y los calculados con curvas ROC

Punto de corte	Parámetro de evaluación			
	Sen.	Esp.	VP+	VP-
Calcio total (mg/dL)	LIT < 8.0 ROC < 8.2	47% 61%	87% 78%	71% 65%
Calcio calculado (mg/dL)	LIT < 9.2 ROC < 9.3	55% 55%	75% 75%	73% 73%

* sen = sensibilidad, esp = especificidad, VP+ = valor predictivo positivo, VP- = valor predictivo negativo. Con autorización de los autores.

INTERPRETACIÓN DE LOS ESTUDIOS DIAGNÓSTICOS CON RESULTADOS CUANTITATIVOS

Cuando los resultados de las pruebas diagnósticas no son dicotómicos (positivo o negativo) sino cuantitativos (p. ej., 1-3, 3-5, > 5), el análisis e interpretación de los resultados es más complicado. Una primera opción es considerar un valor crítico o de corte, es decir, una cifra límite por arriba o por debajo de la cual el labora-

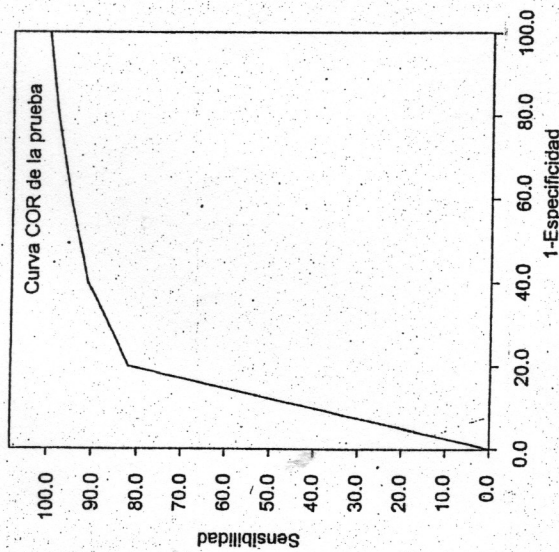


Figura 6-1. Curva operante del receptor. Es el nivel de corte donde se obtiene la mayor sensibilidad y especificidad.

construcción de curvas operante del receptor (COR). Estas curvas se construyen calculando la sen y la esp de la prueba a diferentes niveles de corte y escogiendo la cifra que proporciona una mayor exactitud (tabla 6-3B). Dicha cifra corresponde gráficamente al punto más cercano al ángulo superior izquierdo de la curva (figura 6-1).

Los resultados muestran (tabla 6-3) que el calcio calculado ofrece una mayor precisión diagnóstica que el calcio sérico total (exactitud 75 vs 72%). Por tanto, si uno no dispone de calcio iónico para hacer el diagnóstico de hipocalcemia, se puede utilizar el calcio calculado, aunque su VP+ es relativamente bajo.

SEGUNDA OPCIÓN: CALCULAR LAS RAZONES DE VEROSIMILITUD

Consiste en interpretar las pruebas diagnósticas con resultados cuantitativos a través de calcular la probabilidad diagnóstica de que el sujeto esté enfermo dada la cifra obtenida en la prueba. Para ello, primero se deben calcular las razones de

Tabla 6-4. Valores de calcio calculado a diferentes niveles de corte para calcular la sensibilidad, especificidad y razones de verosimilitud en 53 neonatos gravemente enfermos de una terapia intensiva neonatal

Valor Ca calculado mg/dl	Hipocalcemia iónica < 1mmol/L	Normocalcemia iónica	Sensibilidad	Especificidad	Razón de verosimilitud
8.21-8.40	0	2	10	100	10
8.41-8.60	4	1	20	95	4
8.61-8.80	3	3	35	90	3.5
8.81-9.00	3	2	50	80	2.5
9.01-9.20	7	11	80	75	3.2
9.21-9.40	2	13	90	40	1.5
> 9.41	2	0	100	0	1
Total	21	32			

* RV o RV estrato = tasa de positivos a una cifra/ tasa de negativos a una cifra. Para el estrato 8.41-8.60 = $(4/21)/(1/32)$, para el estrato 8.61-8.80 = $(3/21) / (3/32)$.

verosimilitud (en inglés, *likelihood ratios*), que resultan de dividir la tasa de enfermos con una cifra determinada entre la tasa de sanos con esta misma cifra. En la tabla 6-4 se reportan las razones de verosimilitud a diferentes niveles de corte del calcio calculado para diagnosticar hipocalcemia en neonatos; el valor más alto fue de 3.5.

Segundo, se calculan los momios a favor (*odds*) que tiene el sujeto de estar enfermo antes de realizarle la prueba; para esto se debe conocer la ppep o prevalencia de la enfermedad o de los síntomas. Si un neonato presenta temblor fino de extremidades, que es interpretado con un 0.40 o 40% de probabilidad de tener hipocalcemia, se convierte esta probabilidad en *odds* con la siguiente fórmula:

$$\text{Odds ppep} = \text{probabilidad}/1 - \text{probabilidad} = 0.40/1 - 0.40 = 0.66.$$

Finalmente, se calcula la probabilidad de estar enfermo, dado el resultado de la prueba, multiplicando el *odd* ppep por la RV. Si el nivel de calcio calculado fue de 8.6 mg/dL, la RV a este nivel es de 3.5. Los *odds* ppop de hipocalcemia se calculan:

$$\text{Odds ppop} = \text{Odds ppep} \times \text{RV} = 0.66 \times 3.5 = 2.31.$$

Para convertir el resultado de *odds* a probabilidades se usa la siguiente fórmula:

$$\text{Odds ppop} = \text{Odds ppop} / 1 + \text{Odds ppop} = 2.31/1 + 2.31 = 0.69 \text{ o } 69\%.$$

Es decir, la probabilidad del neonato de tener hipocalcemia es de 69%.

DECISIÓN DE USO DE UNA PRUEBA DIAGNÓSTICA

La tabla 6-5 lista los criterios más importantes a evaluar en una prueba diagnóstica. Aunque el principal criterio para determinar la utilidad de una prueba diagnóstica es su validez, en la práctica clínica hay otros criterios que deben ser considerados, tales como:

1. **La reproducibilidad:** la calidad de algunas pruebas está influenciada por la forma en que se obtienen y procesan las muestras y por la habilidad de quien las interpreta. Si la prueba no es reproducible quizá no vale la pena solicitarla.
2. **El beneficio:** si se dispone de dos pruebas igualmente válidas se seleccionará la que sea más disponible, sencilla en la preparación, menos molesta, menos compleja y más económica. En ocasiones ésta puede ser usada como de primera fase y dejar la más exacta, con mayor costo y riesgo para una segunda fase. Éste es el principio de los estudios seriados.
3. **La inocuidad:** se debe escoger la prueba que sea más inocua. El médico debe conocer las contraindicaciones, los cuidados y las complicaciones, e informar al paciente sobre la preparación, los riesgos y beneficios que se derivarán de la prueba.
4. **El espectro de la enfermedad:** cuando se está decidiendo sobre el uso de una prueba diagnóstica, hay que tomar en cuenta si la prueba fue validada no sólo en los casos de enfermedad clásica y avanzada, sino también en los casos de presentación en sus diferentes grados de afección y compromiso de otros órganos y sistemas.

Tabla 6-5. Criterios a evaluar en el análisis de un artículo que evalúa una prueba diagnóstica

• ¿Hubo un estándar de referencia (de oro) al cual se comparó la prueba en estudio?
• ¿Fue la comparación con el estándar de referencia cegada e independiente?
• ¿Se describió adecuadamente la población en estudio, así como el tamizaje por el que los pacientes pasaron, antes de ser incluidos en el estudio?
• ¿Se incluyeron pacientes con diferentes grados de severidad de la enfermedad (espectro adecuado) y no sólo pacientes con enfermedad avanzada, o clínicamente evidente?
• ¿Se describió la manera de realizar la prueba diagnóstica con claridad de modo que se pueda reproducir fácilmente?
• ¿Se expresaron con claridad los valores de sensibilidad, especificidad y valores predictivos?
• ¿Se definió la manera en que se delimitó el nivel de "normalidad"?
• ¿Se propone la prueba diagnóstica como una prueba adicional o como una prueba sustituto de la utilizada más comúnmente en la práctica clínica?
• ¿Se informa de las complicaciones o de los efectos adversos potenciales de la prueba?
• ¿Se proporcionó información relacionada al costo monetario de la prueba?

5. **La normalidad:** se debe conocer cual es el criterio de normalidad establecido en una prueba diagnóstica, el cual puede ser basado en lo más frecuentemente encontrado en la población. Aunque la prevalencia de las enfermedades no es siempre la misma y la distribución de algunos datos no siempre es normal, en este criterio se utilizan el promedio y desviaciones estándares o las percentilas según el caso. Otros criterios son: considerar anormal a la persona que tiene un riesgo extra de presentar algún desenlace no deseado, o los sujetos que se benefician de una decisión médica, y/o aquellos que tienen datos clínicos.

CONCLUSIÓN

La decisión de realizar un estudio debe estar fundada en el cumplimiento de los diferentes criterios, de los cuales el principal es la exactitud o validez del estudio, seguido de la evaluación de las condiciones necesarias para su adecuada realización evitando complicaciones o daño al paciente. ■

BIBLIOGRAFÍA

1. Feinstein AR. *Clinimetrics*. New Haven, CT, Yale University Press, 1987.
2. Fletcher RH, Fletcher SW, Wagner EH. *Clinical Epidemiology: The Essentials*. 3ª ed., Williams & Wilkins, USA, 1996.
3. García-Fernández R, Maravilla A, Pichardo-Romero P, Rubio V, Iwasaki-Otake L, Arévila-Ceballos N. Diagnóstico de tumores mamaros por centelleografía versus mastografía. *Rev Invest Clin* 1997; 50: 53-6.
4. Jenicek M. Epidemiology. The Logic of Modern Medicine. EPIMED International Montreal, Canada 1996. En: Kassirer JP, Kopelman RI. *Learning Clinical Reasoning*. Williams & Wilkins, USA, 1991.
5. López-Jiménez F, Rohde LE, Luna-Jiménez MA. Problemas y soluciones en la interpretación de pruebas diagnósticas. *Rev Invest Clin* 1998; 50: 65-72.
6. Rendón-Macias ME, Lazcano J. La importancia del interrogatorio y la exploración física para el diagnóstico clínico pediátricos. *Gac Med Mex* 1996; 53: 374-8.
7. Rendón-Macias ME, Morales-Castillo ME, Escobar-Padilla B, Fajardo-Gutiérrez A. Improved medial efficiency during clinical reasoning by numeric specification of diagnostic probabilities. *Bol Med Hosp Infant Mex* 1998; 55: 576-86.
8. Sackett DL, Haynes RB, Tugwell P. *Clinical Epidemiology. A basic science for clinical medicine*. Little Brown, USA, 1985: 3-138.
9. Serret Montoya J, Villegas-Silva R, Muño-Flores R, Peña L, Villasis-Keever MA. Evaluación de pruebas diagnósticas de hipocalcemia en neonatos gravemente enfermos. *Rev Invest Clin* 1998; 50: 471-6.