

# Conceptos avanzados en la interpretación de pruebas diagnósticas

Francisco López Jiménez  
Luis Eduardo P. Rohde  
Max Alberto Luna-Jiménez

## INTRODUCCIÓN

En el capítulo previo se describieron los requerimientos básicos en el análisis crítico de una prueba diagnóstica. En este capítulo se describen maneras adicionales de interpretar y aplicar la información de estudios que evalúan pruebas diagnósticas.

Este capítulo introduce un abordaje alternativo en el proceso diagnóstico que se basa en la probabilidad de tener la enfermedad antes de la prueba (probabilidad *a priori* o *preprueba*) y ahonda en el concepto de razón de verosimilitud. Además, hace particular énfasis en las limitantes de dividir los resultados de una prueba diagnóstica en “positivos” y “negativos”. También se abordan otros problemas en la interpretación de pruebas diagnósticas en circunstancias especiales, como

en el caso de nuevos métodos diagnósticos que son mejores que el estándar de referencia (estándar de "oro") o cuando no hay un buen estándar de referencia. Finalmente, se discuten diferentes interpretaciones de resultados falsos negativos.

## PROBLEMAS AL DIVIDIR EL RESULTADO DE UNA PRUEBA DIAGNÓSTICA EN NORMAL Y ANORMAL

### ■ JUSTIFICACIÓN PARA DICOTOMIZAR LOS RESULTADOS EN NORMALES Y ANORMALES

Las pruebas diagnósticas son los instrumentos que utiliza el clínico para confirmar o descartar un diagnóstico. En otras palabras, el clínico espera definir, después de conocer el resultado de la prueba, si la enfermedad está presente o ausente. Este tipo de dicotomización (enfermedad presente-ausente) es facilitada si el método diagnóstico proporciona también un resultado dicotomizado (prueba positiva-negativa). En condiciones ideales, una prueba positiva significaría presencia de enfermedad, y una prueba negativa significaría ausencia de enfermedad.<sup>1-3</sup>

Por desgracia, pocas situaciones en medicina cumplen este requisito. Para evaluar una prueba cuyo resultado se ha dicotomizado en positivo-negativo se utilizan las tablas de 2 x 2 y los valores probabilísticos resultantes que han sido explicados en el capítulo anterior. Este abordaje, aunque útil, presenta algunas restricciones importantes. Algunas de ellas se señalan a continuación.

### ■ PÉRDIDA DE INFORMACIÓN ÚTIL AL CONVERTIR DATOS CUANTITATIVOS EN DOS ÚNICAS CATEGORÍAS

Una desventaja importante es la pérdida de información útil al convertir valores cualitativos (mg/L, mmHg, mmol/dl, +/++++, etc.) en dos únicas opciones: normal o anormal. De esta manera, todos los resultados que caen en una categoría intermedia, son forzados a incluirse en una de las dos posibilidades. Por ejemplo, para el diagnóstico de infarto agudo del miocardio con creatin-fosfoquinasa fracción MB (CK-MB) se establece un nivel de normalidad de 8 U/L. Un paciente con un valor de 6.9 U/L sería considerado normal, mientras que un paciente con valores de 8.2 U/L sería considerado anormal aunque la diferencia de 6.9 a 8.2 sea mínima. El problema de pérdida de información útil puede ser aún más relevante si el diagnóstico de cierta condición se fundamenta en una o dos pruebas diagnósticas únicamente.<sup>4</sup>

Información	Población general	Dolor precordial de esfuerzo	Factores de riesgo para cardiopatía isquémica	Prueba de esfuerzo positiva
Probabilidad	+	++	+++	++++
Información	Población general	Dolor precordial puntiforme, no asociado a esfuerzo	Sin factores de riesgo para cardiopatía isquémica	Prueba de esfuerzo positiva
Probabilidad	+	+	+	+

Figura 7-1. Evolución de la probabilidad de tener cardiopatía isquémica, según la información existente.

### ■ POCA UTILIDAD DIRECTA DE LOS VALORES DE SENSIBILIDAD Y ESPECIFICIDAD

Una vez que se obtiene el resultado de una prueba diagnóstica, los valores de sensibilidad y especificidad no son de utilidad directa para predecir la presencia o ausencia de enfermedad en un paciente en particular. Por ejemplo, la especificidad del electrocardiograma de esfuerzo es de 90%. Sin embargo, una prueba positiva en un paciente de 35 años sin factores de riesgo coronario no incrementa mucho la probabilidad de tener cardiopatía isquémica, como lo haría una prueba positiva en un sujeto de 65 años.<sup>5</sup> Para un clínico, las cifras realmente útiles son los valores predictivos, pues éstos dan la probabilidad que el paciente tenga la enfermedad en caso que la prueba haya sido positiva, o que no tenga la enfermedad dado el caso que la prueba haya sido negativa. Sin embargo, los resultados de estudios que evalúan pruebas diagnósticas en muchas ocasiones sólo expresan los valores de sensibilidad y especificidad.

### ■ LOS VALORES PREDICTIVOS INFORMADOS EN LOS ESTUDIOS SE BASAN EN PREVALENCIAS DIFERENTES A LA ENCONTRADA EN NUESTROS PACIENTES

Bajo una definición clínica, prevalencia es la probabilidad de que nuestros pacientes tengan la enfermedad, antes de conocer el resultado de una prueba diagnóstica (probabilidad *a priori*).

Aun cuando los estudios informen los valores predictivos, éstos están intrínsecamente relacionados a la prevalencia de la enfermedad en la población estudiada, que puede ser muy diferente de la población en la que utilizamos la prueba.<sup>6</sup>

Como se mencionó anteriormente, la prevalencia de la enfermedad cambia los valores predictivos de una prueba. Sin embargo en muchas ocasiones tampoco sabemos la prevalencia de la enfermedad en la población en la que utilizamos la prueba; únicamente intuimos la probabilidad de cada individuo (antes de la prueba) de tener la enfermedad, con base en la historia clínica y hallazgos en la exploración física. En resumen, los clínicos necesitamos saber los valores predictivos para cada paciente en particular. Conociendo de antemano que los valores predictivos se modifican con la prevalencia de la enfermedad, entonces necesitamos saber la manera de convertir los valores de sensibilidad y especificidad en valores predictivos de acuerdo a la probabilidad *a priori* de cada individuo de tener la enfermedad. Esto es posible utilizando la razón de verosimilitud, que más adelante explicaremos con detalle.

### ■ DIFICULTAD DE COMPARAR DIFERENTES PRUEBAS DIAGNÓSTICAS

En ocasiones el clínico desea saber cuál prueba es mejor para diagnosticar cierta enfermedad. La decisión es sencilla si una prueba es más sensible y más específica que otra. El problema se presenta cuando una prueba es ligeramente más sensible pero menos específica que otra, o viceversa.

## EL PROCESO DIAGNÓSTICO

El escenario común en la práctica clínica es un paciente con ciertas características (v. gr., fumador, obeso y con historia familiar de cardiopatía isquémica) que acude al consultorio por la aparición de cierto síntoma (v. gr., dolor precordial al subir escaleras), que se somete a uno o varios estudios de diagnóstico (v. gr., electrocardiograma de reposo y en esfuerzo) en los que se detecta cierta anomalía (v. gr., infradesniveles del segmento ST de 2 mm después de 4 minutos de ejercicio en una banda sinfin) y que finalmente recibe un diagnóstico definitivo o presuntivo (v. gr., angina de pecho, probable obstrucción de la arteria coronaria descendente anterior). La probabilidad de que un paciente tenga la enfermedad va cambiando durante el proceso diagnóstico (ver figura 7-1). Las pruebas diagnósticas pueden proporcionar información adicional, que al igual que la historia clínica y los hallazgos físicos, pueden incrementar o disminuir nuestro apoyo a cierto diagnóstico (ver figura 7-2).<sup>7</sup>

### ■ RAZÓN DE VEROSIMILITUD: SOLUCIÓN A VARIOS PROBLEMAS EN LA INTERPRETACIÓN DE RESULTADOS DE PRUEBAS DIAGNÓSTICAS

En el capítulo previo se introduce al concepto de razón de verosimilitud, que es una alternativa para resolver en parte las limitantes de utilizar los valores de sensibilidad y especificidad de manera separada. La razón de verosimilitud es un valor probabilístico que combina los valores de sensibilidad y especificidad en uno solo, y se obtiene dividiendo la probabilidad de que un paciente tenga la prueba positiva dado que tiene la enfermedad (sensibilidad), con la probabilidad de que tenga la prueba positiva dado que no tiene la enfermedad (1 - especificidad).<sup>8,9</sup>

Para utilizar la razón de verosimilitud se requiere convertir en momios la probabilidad preprueba (*a priori*). Momios son los números que expresan la probabilidad de que suceda un evento dividido por la probabilidad de que no suceda  $\{p/(1-p)\}$ . V. gr., si la probabilidad clínica (preprueba) de que cierto paciente tenga cáncer de páncreas es alrededor de 75%, los momios serán 75:25, o 3:1, que equivale a 3. Al multiplicar los momios preprueba por la razón de verosimilitud, se obtienen los momios posprueba (ver figura 7-2). Dado que la razón de verosimilitud del ultrasonido para diagnóstico de cáncer de páncreas es de 4, entonces los momios

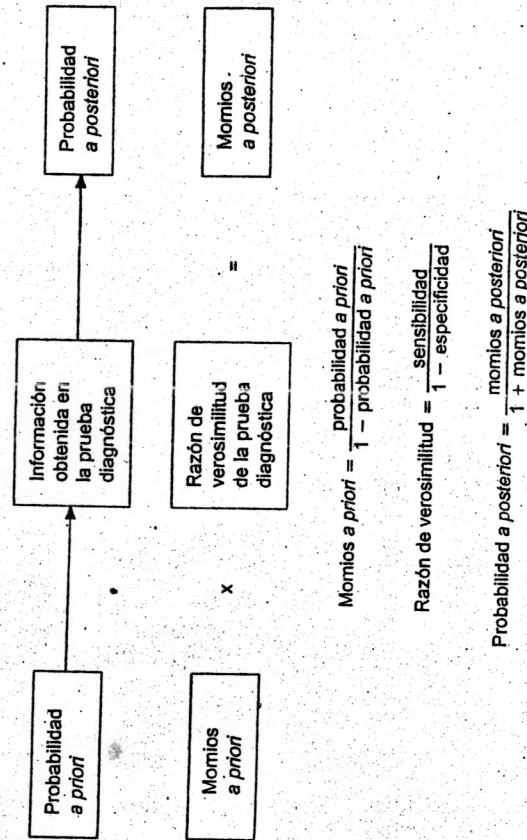


Figura 7-2. El uso de la razón de verosimilitud es conceptualmente similar al proceso diagnóstico usual.

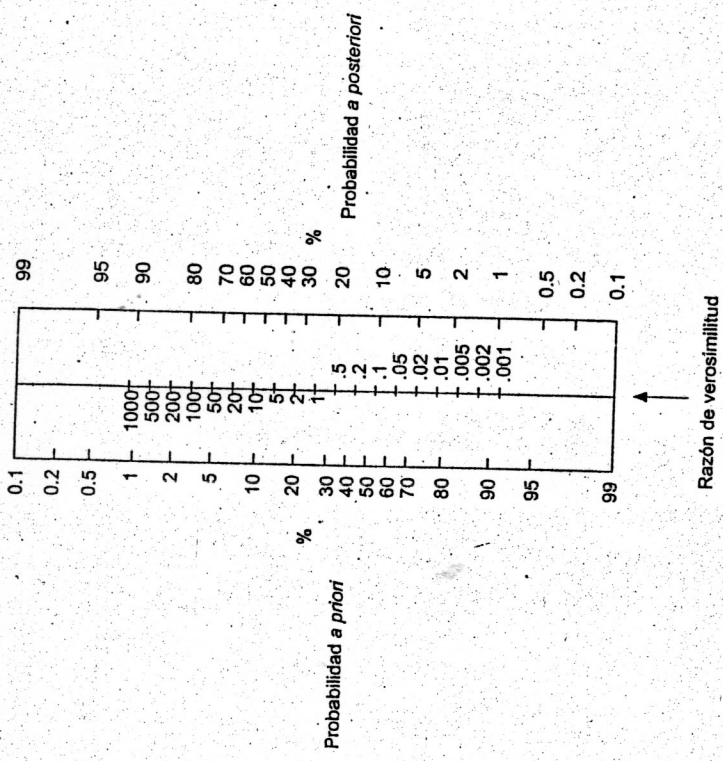


Figura 7-3. Nomograma para interpretar pruebas diagnósticas. Para obtener la probabilidad a posteriori, primero se selecciona la probabilidad a priori en la columna de la izquierda, luego se cruza una línea recta desde el punto seleccionado y que pase por la razón de verosimilitud obtenida. Finalmente, se prolonga la línea hasta la columna de la derecha, que corresponde a la probabilidad a posteriori.

posprueba son  $4 \times 3 = 12$ , que convertidos a probabilidad,  $\{m/(1 + m)\}$ , dan una probabilidad posprueba (a posteriori) de 92% de tener cáncer pancreático.

Para hacer más fácil la conversión de probabilidad preprueba a probabilidad posprueba se diseñó un nomograma<sup>10</sup> (figura 7-3). El clínico únicamente tiene que escoger en la columna de la izquierda la probabilidad clínica de que el paciente tenga la enfermedad, marcar una línea recta que cruce en la columna del centro el valor de la razón de verosimilitud de la prueba, y prolongar la línea recta para identificar en la columna de la derecha la probabilidad posprueba de tener cáncer pancreático.

Las ventajas de utilizar la razón de verosimilitud son varias. Primero, combina dos valores en un solo número, lo que facilita comparar diferentes pruebas diag-

nósticas o diferentes puntos de corte en una misma prueba diagnóstica. Este valor único puede ser empleado en el proceso diagnóstico como el equivalente a la información adicional proporcionada por la prueba diagnóstica (figura 7-2). El conocer el valor de la razón de verosimilitud de diferentes puntos de corte, nos permite aprovechar la información diagnóstica que proporcionan pruebas diagnósticas con valores cuantitativos, especialmente los valores cercanos al escogido como "normal". Además que proporciona un valor probabilístico individualizado a cada paciente de acuerdo a sus características clínicas particulares, y no de acuerdo a la prevalencia global de la población estudiada o de la población heterogénea donde se aplica la prueba.

### SELECCIÓN DEL MEJOR PUNTO DE CORTE

Una vez que se utiliza la razón de verosimilitud para diferentes puntos de corte en una prueba diagnóstica con resultados cuantitativos, no se requiere establecer una línea entre normalidad y anormalidad. Sin embargo, reconociendo que el proceso de cambio puede tomar tiempo, y que muy probablemente se continúa utilizando el método de dicotomizar los valores en normales y anormales, es importante reconocer las alternativas para escoger el mejor punto de corte.

El método más empleado quizá sea la curva de las características operativas del receptor (COR). Este método consiste en graficar los valores de sensibilidad y especificidad obtenidos en diferentes puntos de corte (ver figura 6-1); después, se selecciona el punto más cercano a la esquina superior izquierda, que corresponde al punto donde la sensibilidad y especificidad combinadas tienen el máximo valor.

La principal desventaja de seleccionar el punto A como el punto de corte ideal, es que corresponde a un razonamiento más matemático que clínico, pues presupone un mismo valor clínico para los resultados falsos positivos que para los falsos negativos. Por ejemplo, supongamos que cierta prueba X en su valor crítico óptimo tiene una sensibilidad y especificidad de 90%, mientras que en otro valor subóptimo la sensibilidad es del 98% pero la especificidad desciende a 70%. Es evidente que el valor crítico óptimo representa menos errores totales (10 pacientes con la prueba negativa, pero con la enfermedad —falsos negativos— y 10 pacientes con la prueba positiva pero sin enfermedad —falsos positivos—), comparado con el nivel "subóptimo", donde existirían dos falsos negativos y 30 falsos positivos. Sin embargo, supongamos que la enfermedad a diagnosticar tiene una mortalidad de 90% a corto plazo si no recibe tratamiento y sólo de 10% en caso de que sí reciba el tratamiento, cuyos efectos colaterales son mínimos. Clínicamente, es preferible el nivel crítico "subóptimo", pues solamente dejaríamos de tratar dos

pacientes (1.8 muertes) a cambio de administrar el tratamiento a 30 individuos que no lo necesitan. Utilizando el punto de corte "óptimo", serían casi 8 muertes a cambio de evitar los inconvenientes del tratamiento en 20 pacientes que no lo necesitan. Otra desventaja de utilizar la curva COR es que no se incluye la prevalencia de la enfermedad dentro de la decisión. Como mencionamos más arriba, la prevalencia de la enfermedad cambia la distribución de los falsos positivos y negativos dentro del número total de errores. Si utilizamos la prueba Z, que tiene una sensibilidad de 95% pero una especificidad de 70% en su valor crítico óptimo, y la aplicamos a una población cuya prevalencia sea de 50%, por cada 2,000 pacientes estudiados (es decir 1,000 pacientes con la enfermedad y 1,000 pacientes sin ella) tendríamos 300 falsos positivos y 60 falsos negativos (5 falsos positivos por cada falso negativo). Si la desventaja de diagnosticar como sano a un paciente que tiene la enfermedad es 5 veces mayor que la desventaja de diagnosticar como enfermo a un paciente sano, entonces la optimización de la prueba se aplica solamente a una prevalencia de la enfermedad de 50%. Si la desventaja clínica de tener un resultado falso negativo es equivalente a la desventaja de tener un falso positivo (1:1) (v. gr., tratamientos con alto riesgo de complicaciones, cambios emocionales importantes por etiquetar al paciente de enfermo, etc.), entonces necesitaríamos tener una prevalencia de 85.7% para así obtener el mismo número de falsos negativos que de falsos positivos, que en este ejemplo sería de 86.

Se debe reconocer que, la asignación de un valor numérico a la relación entre los inconvenientes entre los falsos positivos y los falsos negativos es un procedimiento cualitativo e inexacto, por lo que se deben utilizar sólo como aproximaciones a lo deseable clínicamente, incluyendo diferentes aspectos del paciente como beneficio real, efectos colaterales y costo del tratamiento; calidad de vida, etc.

El procedimiento para obtener la prevalencia necesaria de acuerdo al valor relativo de los falsos negativos, comparado con los positivos, se da con más detalle en el Apéndice.

## PROBLEMAS CON LA DEFINICIÓN DEL ESTÁNDAR DE REFERENCIA

### ■ CUANDO LA PRUEBA A EVALUAR ES MEJOR QUE EL ESTÁNDAR DE REFERENCIA ESTABLECIDO

El estándar de "oro" o estándar de referencia es la prueba diagnóstica con mayor sensibilidad y especificidad de entre todas las disponibles. Idealmente sería una prueba cuyo resultado positivo confirme la enfermedad, y en caso de ser negativo la descarte. Sin embargo, la mayoría de estándares de referencia son imperfectos,

por lo que son susceptibles a ser comparados a una prueba con mejor desempeño diagnóstico. Si se compara un método diagnóstico novedoso que tiene mayor sensibilidad que el estándar, el nuevo método aparenta ser menos específico, pues detecta casos reales que el estándar de referencia no los identifica como tales, por lo que los cataloga como falsos positivos. Por el contrario, si la prueba novedosa es más específica que el estándar, aparecerá como si fuera menos sensible, pues algunos casos que la prueba nueva cataloga correctamente sin enfermedad, para el estándar son falsos negativos.

Múltiples métodos diagnósticos se han comparado a estándares de referencia inferiores. La enzima creatin-kinasa se comparó con el electrocardiograma en el diagnóstico de infarto agudo del miocardio, y posteriormente la fracción MB de la misma enzima se enfrentó a la utilidad diagnóstica del valor total de creatin-kinasa. Actualmente otros marcadores serológicos como la troponina y la miosina son comparadas con la creatin-kinasa, fracción MB. En todos estos casos, el proceso de aceptación del nuevo método diagnóstico como el nuevo estándar de referencia se ha dificultado por el hecho de ser comparados con estándares menos óptimos que la prueba misma.

Con el advenimiento de nuevas técnicas de diagnóstico molecular y genético, se incrementa el dilema entre la importancia de detectar la presencia biológica de la enfermedad y la importancia de lo que sea o llegue a ser presente clínicamente. Por ejemplo, las mutaciones BRCA1 y BRCA2 están relacionados a una predisposición de padecer cáncer de mama. Sin embargo, más de la mitad de las mujeres portadoras de dichos genes nunca desarrollan la enfermedad, mientras que un gran número de pacientes con cáncer de mama no tienen tales genes.<sup>11</sup> Lo mismo se aplica a algunas pruebas con PCR que dada su alta sensibilidad son capaces de detectar ADN, aun cuando el germen nunca se manifestó clínicamente.

### ■ ¿CÓMO DETERMINAR SI UNA PRUEBA NUEVA ES MEJOR QUE EL ESTÁNDAR DE REFERENCIA?

Para averiguar si una prueba nueva es potencialmente mejor que el estándar de referencia, se deben utilizar diferentes abordajes. Uno, juzgar si la posibilidad es biológicamente plausible. En el caso de la troponina T, se sabe que su expresión es única en el miocardio, mientras que la CK-MB tiene expresión en diferentes tejidos.<sup>12</sup> Después se puede realizar experimentación básica. Al tratar de detectar troponina T en diferentes tejidos triturados, incluyendo músculo estriado, miocardio y tejido cerebral, sólo se ha detectado troponina en el tejido cardíaco triturado, mientras que la CK-MB se ha detectado en diferentes. Finalmente, se pueden utilizar subrogados diagnósticos como complicaciones, nuevos eventos clínicos, etc. Los pacientes que durante cirugía presentan elevaciones de troponina T pero que no cumplen los criterios diagnósticos de infarto agudo del miocar-

dio tienen una probabilidad de casi 5 veces de tener un evento cardiaco grave en los próximos 6 meses, mientras que los pacientes con elevaciones transoperatorias de CK-MB tienen un pronóstico similar al de aquéllos en quienes la CK-MB se mantuvo normal.<sup>13</sup>

### ■ CUANDO NO HAY ESTÁNDAR DE REFERENCIA

La selección del estándar de referencia depende de la perspectiva desde la que se está evaluando al paciente. En algunos casos no importa tanto la presencia biológica de la enfermedad, sino su repercusión clínica. En el caso del carcinoma foliular de tiroides o del carcinoma de próstata, la mayor parte de los pacientes con microtumores probablemente morirán de otra causa, y también es probable que el tumor nunca se haga evidente.<sup>14</sup> Por tal motivo, se ha propuesto que la definición de estándar de referencia incluya también variables clínicas de interés como probabilidad de progresión clínica, síntomas y mortalidad específica. Volviendo al ejemplo del infarto agudo del miocardio, llama la atención ver que en unos estudios se utiliza la curva de CK-MB más cambios electrocardiográficos como estándar de referencia al evaluar valores aislados de CK-MB,<sup>15</sup> en otros se utiliza el ecocardiograma para evaluar la troponina I<sup>16</sup>, y en otros se utiliza la troponina para evaluar subformas de la CK-MB. Resulta claro que no hay un estándar de referencia para infarto agudo del miocardio. La discrepancia puede radicar en la definición de infarto agudo del miocardio. Si lo que se requiere es identificar pacientes con infarto que se pueden beneficiar con trombolíticos, nada supera al electrocardiograma. Si se requiere evaluar infartos con repercusión en la contractilidad cardíaca, la mejor alternativa es el ecocardiograma. Si se trata de evaluar la muerte de algunas miofibrillas, probablemente la troponina es el mejor método. Cada definición de infarto miocárdico tiene una utilidad clínica diferente: la terapéutica, la funcional y la celular.

Cuando no hay un estándar de referencia clínicamente válido, se pueden utilizar subgrupos clínicos como mortalidad específica o progresión de la enfermedad a una etapa sintomática. En otras ocasiones se pueden usar evaluaciones hechas por paneles de expertos o índices de concordancia interobservadores. En psiquiatría se ha explorado el uso de análisis de estructura latente, que integran información múltiple que es analizada mediante regresión logística.<sup>17</sup>

### INTERPRETACIÓN DE LOS FALSOS NEGATIVOS: FALSOS NEGATIVOS POR AZAR O POR LA PRUEBA PER SE

Una prueba diagnóstica puede tener resultados erróneamente negativos (falsos negativos), esencialmente por dos diferentes mecanismos: por la variabilidad

atribuida al azar o por que la prueba no puede detectar la anomalía en estudio. Los errores atribuidos a la variabilidad por azar incluyen la variabilidad intraobservador, cambios biológicos, y otros factores que estrictamente no dependen del azar, pero que si no son sistemáticos, se puede esperar que ocurran de manera aleatoria.

Los falsos negativos que son el resultado de la imposibilidad de la prueba para detectar la enfermedad pueden depender tanto de la etapa en la que se encuentra la enfermedad a diagnosticar, como del buen funcionamiento de la prueba diagnóstica. Un ejemplo es el diagnóstico de infección por el virus de la inmunodeficiencia humana (VIH). El resultado de la prueba de ELISA para VIH que detecta anticuerpos virales, puede ser erróneamente negativo a pesar de que el paciente tenga anticuerpos. En este caso sería un error atribuido al azar como parte de la variabilidad esperable en la detección del anticuerpo por medio de la inmunoadsorción. Por otro lado, la prueba puede resultar erróneamente negativa porque el paciente se encuentra durante la ventana inmunológica, periodo en el que el paciente está infectado pero no tiene suficientes anticuerpos para ser detectados por la prueba.

La importancia en distinguir entre ambos tipos de falsos negativos radica en que un resultado falso negativo secundario a variabilidad aleatoria es muy probable que desaparezca si uno repite la prueba. En una prueba en la que se espera una tasa de falsos negativos por azar de 3%, al repetirse la prueba después de un resultado erróneamente negativo, la probabilidad de que vuelva a ser un falso negativo se reduce a menos de 1/1000. Sin embargo, si el falso negativo fue debido a que la prueba no puede detectar la anomalía, la probabilidad de que obtengamos otro resultado falso negativo al repetir la prueba es casi de 100%.

## CONCLUSIONES

Los valores de sensibilidad y especificidad tienen utilidad limitada para predecir enfermedad, a menos que se involucre la prevalencia dentro del cálculo. La razón de verosimilitud nos facilita este proceso.

Cuando un método diagnóstico sea potencialmente mejor que el estándar de referencia, se deben considerar los posibles errores en la interpretación al compararse el uno con el otro. Estándares de comparación subóptimos, pueden hacer que una prueba diagnóstica excelente parezca poco sensible o poco específica.

Existen dos clases de resultados falsos negativos y el clínico debe estar consciente de ambos ante un resultado negativo que probablemente sea erróneo.

Para una utilización óptima de las pruebas diagnósticas, el clínico debe estar familiarizado con los problemas y posibles soluciones en su interpretación. Un

diagnóstico correcto es la piedra angular en la que se basa el manejo y el pronóstico de los pacientes.

## APÉNDICE

Fórmula para obtener la prevalencia necesaria para que la relación entre resultados falsos positivos y falsos negativos sea similar a la relación entre los valores clínicos de ambos resultados erróneos, en caso de seleccionar el nivel de normalidad con la curva COR.

$$P = \frac{Rr}{Rr + Rc}$$

donde P = prevalencia de la enfermedad; Rr = relación real = tasa de falsos positivos/tasa de falsos negativos multiplicado por (1 - especificidad)/(1 - sensibilidad); Rc = relación clínica = valor clínico de un falso positivo<sup>a</sup>/valor clínico de un falso negativo<sup>b</sup>.

<sup>a</sup> incluye las desventajas clínicas de diagnosticar con la enfermedad a un paciente que no la tiene, v. gr., efectos colaterales del tratamiento, efectos psicológicos por estar etiquetado como enfermo, riesgos putativos a la hospitalización como infecciones nosocomiales, efectos adversos secundarios a errores médicos o de enfermería, etc.

<sup>b</sup> incluye las desventajas clínicas de diagnosticar como libre de enfermedad a un paciente que realmente está enfermo, v. gr., no administrar el tratamiento indicado, no poner al paciente en vigilancia médica, etc. ■

## BIBLIOGRAFÍA

1. Departamento de Epidemiología Clínica y Bioestadística, Universidad McMaster. Cómo leer revistas médicas. II Para aprender sobre una prueba diagnóstica. *Rev Invest Clin* 1988; 40: 73-83.
2. Jaeschke R, Guyatt GH, Sackett DL. User's guide to medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? *JAMA* 1994; 271: 703-7.
3. Glasziou, PP. Probability revision: medical decision making. *Prim Care* 1995; 22: 235-45.
4. Radack KL, Roijan G, Hedges J. The likelihood ratio. An improved measure for reporting and evaluating diagnostic tests results. *Arch Pathol Lab Med* 1986; 110: 689-93.
5. Diamond GA, Forrester JS. Analysis of probabilities as an aid in the clinical diagnosis of coronary artery disease. *N Engl J Med* 1979; 300: 1350-8.

6. Black ER, Panzer RJ, Mayewski RJ, Griner PF. Characteristics of diagnostic tests and principles for their use in quantitative decision making. In: Panzer RJ, Black ER, Griner PF (eds.). *Diagnostic strategies for common medical problems*. Philadelphia American College of Physicians Press, 1991: 1-16.
7. Weinstein MC, Fineberg HV (eds.). The use of diagnostic information to revise probabilities. In: *Clinical decision analysis*. Philadelphia, PA, W. B. Saunders Co., 1980: 75-130.
8. Fletcher RH, Fletcher SH, Waner EH (eds.). *Clinical Epidemiology. The essentials*. 3ª ed., Baltimore, Williams and Wilkins, 1996: 53-74.
9. Sackett DL, Haynes RB, Guyatt GH, Tugwell P (eds.). *Clinical Epidemiology. A basic science for clinical medicine*. Boston, Little Brown, 1991: 69-152.
10. Fagan TI. Nomogram for the Bayes theorem. *N Engl J Med* 1975; 293: 257.
11. Struwing JP, Hartge P, Wacholder S et al. The risk of cancer associated with specific mutations of BRCA1 and BRCA2 among Ashkenazi Jews. *N Engl J Med* 1997; 336: 1401-8.
12. Katus HA, Remppis A, Scheffold T, Dieterich KW, Kuebler W. Intracellular compartmentation of cardiac troponin T and its release kinetics in patients with reperfused and nonreperfused myocardial infarction. *Am J Cardiol* 1991; 67: 1360-7.
13. López-Jiménez F, Goldman L, Sacks D et al. Prognostic value of cardiac troponin T after noncardiac surgery: six month follow-up data. *J Am Coll Cardiol* 1997; 29: 1241-5.
14. López-Jiménez F, Luna-Jiménez MA. Cuándo y en quién hacer diagnóstico temprano de enfermedad. *Rev Invest Clin* 1997; 49: 67-74.
15. López-Jiménez F, Goldman L, Thomas EJ, Kuntz KM, Sacks DB, Lee TH. Predictive value of creatine kinase MB for the diagnosis of acute myocardial infarction after major noncardiac surgery. *J Invest Med* 1995; 43(Suppl): 345A.
16. Adams JE III, Sicard GA, Allen BT, Bridwell KH, Lenke LG, Dávila-Roman VG et al. Diagnosis of perioperative myocardial infarction with measurement of cardiac troponin I. *N Engl J Med* 1994; 330: 676-4.
17. Faraone SV, Tsuang MT. Measuring diagnostic accuracy in the absence of a 'gold standard'. *Am J Psychiatry* 1994; 151: 650-7.