

Índices de rendimiento de las pruebas diagnósticas

V. Abraira

Unidad de Bioestadística Clínica. Hospital Ramón y Cajal. Madrid.

Ya se vio en otra nota¹ un índice para cuantificar la reproducibilidad o concordancia de un instrumento de medida cuando sus resultados se expresan mediante una variable binaria (tipo “sí” o “no”), que es la forma más simple en la que se pueden expresar los resultados de las pruebas diagnósticas cuando se usan para determinar si un individuo tiene, o no, una cierta enfermedad, un síntoma o un signo. En esta nota se van a repasar los índices empleados para evaluar el otro tipo de error de medición, el debido a la falta de validez². Por ejemplo, en la tabla 1 se muestran los resultados de la medición, en ancianos, de la concentración plasmática de péptido natriurético tipo B para diagnosticar la disfunción ventricular izquierda, comparados con el diagnóstico basado en el ecocardiograma³ ¿es suficientemente buena la medición de la concentración plasmática de este péptido para diagnosticar la disfunción ventricular? Obviamente para contestar a esta pregunta hemos de asumir que existe otro procedimiento que permite hacer el diagnóstico correcto con el que se compara la prueba a evaluar. La elección de este procedimiento, habitualmente denominado patrón de oro, es el principal problema del diseño de un estudio de evaluación de una prueba diagnóstica⁴. En el ejemplo se eligió el ecocardiograma como patrón de oro.

Entendiendo por “positivo” el resultado que predice la presencia de la enfermedad o condición de interés, se definen como índices de validez: la sensibilidad (*Sen*) o proporción de verdaderos positivos, es decir, la proporción de positivos en los enfermos y especificidad (*Esp*) o proporción de verdaderos negativos. En el ejemplo $Sen = 11/12 = 0,92$ y $Esp = 93/143 = 0,65$. Ambos son fácilmente interpretables: toman valores entre 0, que indica que la prueba es totalmente inválida y 1 si la prueba fuera perfectamente válida. Aunque estos índices corresponden al punto de vista “natural” del investigador que evalúa la prueba y son por ello los más usados en las publicaciones científicas, en

Tabla 1. Resultados de la evaluación del péptido tipo B

Ecocardiograma	Péptido natriurético tipo B (pmol/l)			Total		
	Normal (< 18,7)	Elevado (≥18,7)				
Normal	93	a	b	50	$r=a+b$	143
Disfunción	1	c	d	11	$s=c+d$	12
Total	94	$t=a+c$	$u=b+d$	61	$N=a+b+c+d$	155

la práctica clínica lo que suele ser conocido es el resultado de la prueba y por lo tanto, lo que tiene interés es la proporción de individuos con resultado “positivo” que verdaderamente tienen la enfermedad. A esta proporción se le denomina *valor predictivo positivo (VP+)* y, simétricamente, se denomina *valor predictivo negativo (VP-)* a la proporción de individuos con resultado “negativo” que verdaderamente no tienen la enfermedad. En el ejemplo $VP+ = 11/61 = 0,18$ y $VP- = 93/94 = 0,99$, que nos indican que la inmensa mayoría (el 99%) de los individuos en que la prueba es negativa no tienen la enfermedad, pero que sólo una minoría (el 18%) de aquéllos en que el resultado es positivo la padecen.

A pesar de que estos últimos índices representan las proporciones de interés en la práctica, su valor es muy limitado porque dependen de la proporción de enfermos en el grupo en el que la prueba se ha evaluado². Es decir, los valores predictivos anteriores sólo son válidos en aquellos ambientes en los que la prevalencia de la disfunción ventricular izquierda sea la misma que la del estudio, es decir $12/155 = 0,08$. Hay que tener en cuenta además que lo que importa para interpretar el resultado de una prueba diagnóstica no es realmente la prevalencia (proporción de enfermos en la población), sino la denominada probabilidad preprueba, es decir la probabilidad de que el individuo tenga la enfermedad antes de aplicar la prueba, probabilidad que será estimada a partir de la anamnesis y de las pruebas previas (en general, la probabilidad preprueba coincide con la prevalencia sólo en el caso de una prueba aplicada como cribado poblacional).

Hay otros índices que, por ser más recientes, son menos usados pero que sin embargo son los más útiles para interpretar el resultado de una prueba diagnóstica⁵, los denominados *cociente de probabilidad positivo (CP+)* y *cociente*

Correspondencia:
Dr. V. Abraira.
Unidad de Bioestadística Clínica.
Hospital Ramón y Cajal. Crta. Colmenar km 9,100.
28034 Madrid.
Correo electrónico: victor.abraira@hrc.es

de probabilidad negativo (CP^-) que, usando la notación de la tabla 1, se definen como:

$$CP+ = \frac{Sen}{1 - Esp} = \frac{\frac{d}{b}}{\frac{b}{b \times s}} = \frac{d \times r}{b \times s} \quad CP- = \frac{1 - Sen}{Esp} = \frac{\frac{c}{a}}{\frac{a}{a \times s}} = \frac{c \times r}{a \times s}$$

Es decir, $CP+$ representa cuánto más frecuente es el resultado positivo entre los enfermos que entre los no enfermos. Si el resultado positivo se obtuviera con la misma frecuencia en los enfermos y en los no enfermos, el hecho de que la prueba sea positiva no contendría ninguna información y $CP+ = 1$; cuanto más frecuente sea el resultado positivo en los enfermos con respecto a los no enfermos más información contiene este resultado y mayor sería $CP+$. Simétricamente CP^- representa cuánto más frecuente es el resultado negativo entre los enfermos que entre los no enfermos. Si el resultado negativo se obtuviera con la misma frecuencia en los enfermos y en los no enfermos, este resultado no contendría ninguna información y $CP^- = 1$; cuanto menos frecuente sea el resultado negativo en los enfermos con respecto a los no enfermos más información contiene este resultado y menor sería CP^- .

Los cocientes de probabilidad permiten transformar la probabilidad preprueba en probabilidad postprueba mediante la relación²:

$$odds_{post} = CP_R \times odds_{pre}$$

Recuérdese que se denomina $odds$ ⁶ al cociente entre la probabilidad de que un evento (en este caso "estar enfermo") ocurra y la probabilidad de que no ocurra, por lo tanto

$$odds = \frac{p}{1-p} \quad \Rightarrow \quad p = \frac{odds}{1+odds}$$

los subíndices $post$ y pre indican postprueba y preprueba respectivamente y el subíndice R representa el resultado obtenido.

Los CP de la prueba del ejemplo son

$$CP+ = \frac{d \times r}{b \times s} = \frac{11 \times 143}{50 \times 12} = 2,62 \quad CP- = \frac{c \times r}{a \times s} = \frac{1 \times 143}{93 \times 12} = 0,13$$

supóngase que se le aplica la prueba a un anciano al que por su sintomatología e historia se le estima una probabilidad alta (por ejemplo 0,7) de padecer disfunción ventricular izquierda ¿cómo se modificaría esta probabilidad si se le midiera la concentración plasmática del péptido y resultara "positiva" ($\geq 18,7$)? A partir de la probabilidad preprueba, primero se calcula el $odds_{pre} = 0,7 / (1-0,7) = 2,33$; como el resultado es positivo $CP_R = CP+ = 2,62$ por tanto $odds_{post} = 2,62 \times 2,33 = 6,10$ y $p_{post} = 6,10 / 7,10 = 0,86$; si el resultado de la prueba hubiera sido negativo $CP_R = CP- = 0,13$ y $odds_{post} = 0,13 \times 2,33 = 0,30$ y por lo tanto $p_{post} = 0,30 / 1,30 = 0,23$. Existen calculadoras tanto gráfi-

Tabla 2. Guía aproximada del valor informativo de los CP

$CP > 10$ ó $CP < 0,1$	Cambios amplios
$5 < CP < 10$ ó $0,1 < CP < 0,2$	Cambios moderados
$2 < CP < 5$ ó $0,2 < CP < 0,5$	Cambios pequeños
$1 < CP < 2$ ó $0,5 < CP < 1$	Cambios insignificantes

Puntos clave

- La *sensibilidad y especificidad*, que son los índices más establecidos como índices de validez de las pruebas diagnósticas, corresponden al punto de vista "natural" del investigador que evalúa la prueba, pero requieren de ciertos cálculos para interpretar el resultado de una prueba en la práctica clínica.
- Los *valores predictivos* representan las proporciones de interés en la práctica, pero su valor es muy limitado porque dependen de la proporción de enfermos en el grupo en el que la prueba se ha evaluado.
- Los *cocientes de probabilidad* son los índices más útiles porque permiten transformar la probabilidad preprueba en probabilidad postprueba mediante una sencilla relación y no dependen de la proporción de enfermos en el grupo en el que se ha evaluado la prueba.

cas⁵ como electrónicas que facilitan estos cálculos. La utilidad de una prueba diagnóstica reside en la modificación que su resultado produce en la probabilidad de estar enfermo, obsérvese que en este ejemplo el cambio es mayor si la prueba tiene un resultado negativo, debido a que el CP^- es más distinto de 1 que el $CP+$. Como guía aproximada de la utilidad en función del cambio que produce en la probabilidad se puede usar la que figura en la tabla 2 debida a Jaeschke et al⁵.

Nota: en el apartado de software de la página web de la Unidad de Bioestadística Clínica del Hospital Ramón y Cajal (<http://www.hrc.es/bioest.html>) está disponible una calculadora que, a partir de los datos "crudos" de un estudio sobre la validez de una prueba diagnóstica, calcula todos estos índices y sus intervalos de confianza, así como la probabilidad postprueba a partir de la probabilidad preprueba.

BIBLIOGRAFÍA

1. Abraira V. El índice kappa. SEMERGEN 2001;27:247-9.
2. Abraira V. Errores en las mediciones y clasificaciones clínicas: Precisión y validez. [consultado: 11/01/2002]. Disponible en: http://www.hrc.es/bioest/Intro_errores.html
3. Smith H, Pickering RM, Struthers A, Simpson I, Mant D. Biochemical diagnosis of ventricular dysfunction in elderly patients in general practice: observational study. Br Med J 2000;320:906-8.
4. Cabello López JB, Pozo Rodríguez F. Métodos de investigación en cardiología clínica (X). Estudios de evaluación de las pruebas diagnósticas en cardiología. Rev Esp Cardiol 1997;50:507-19.
5. Jaeschke R, Guyatt G, Sackett DL for the Evidence-Based Medicine Working Group. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What were the results and will they help me in caring for my patients? JAMA 1994;271:703-7.
6. Abraira V. Medidas del efecto de un tratamiento (II): *odds ratio* y número necesario para tratar. SEMERGEN 2001;27:418-20.