



Artículo Especial

Validación de test diagnósticos (primera parte). Test categóricos

Validation of diagnostic tests (part one). Categorical tests

Antonio Martín Conejero

Hospital Clínico San Carlos. Madrid

Resumen

La validez de una prueba diagnóstica se refiere a la capacidad de un test de identificar correctamente como enfermo al paciente que es enfermo y como sano al que es sano. Hay que tener en cuenta que el punto de partida es la comparación entre el resultado obtenido en nuestro test diagnóstico y los resultados obtenidos por lo que consideramos prueba de referencia. En esta primera parte analizaremos conceptos como sensibilidad, especificidad y valores predictivos (positivo y negativo), así como los cocientes de probabilidad o verosimilitud.

Palabras clave:

Sensibilidad.
Especificidad. Valor
predictivo. Razón
de verosimilitud.

Abstract

The validity of a diagnostic test refers to the ability of a test to correctly identify the patient who is sick as sick and the one who is healthy as healthy. It must be considered that the starting point is the comparison of the result obtained in our diagnostic test with the results obtained by what we consider to be the "gold standard" or reference test. In this first part we will analyze concepts such as sensitivity, specificity, predictive values (positive and negative), as well as probability or likelihood ratios.

Keywords:

Sensitivity. Specificity.
Predictive value.
Likelihood ratio.

INTRODUCCIÓN

Como médicos que somos, seguro que en muchas ocasiones nos hemos encontrado en nuestra práctica clínica en situaciones similares a las que exponemos a continuación:

- Un eco Doppler carotídeo se informa como normal. Sin embargo, nos preguntamos si podría tratarse de un error y que este paciente tuviese una estenosis severa. ¿Habría alguna

forma de saber si el eco Doppler es por tanto un buen test diagnóstico?

- Un representante de un laboratorio ha ofrecido un test que en Japón se emplea porque tiene muy alta sensibilidad y especificidad para la enfermedad de Takayasu. Es un muy buen test, pero para una enfermedad muy poco prevalente en nuestro medio. ¿Es aconsejable realizar este test en nuestro medio, dado lo bueno que es?

Recibido: 18/10/2022 • Aceptado: 18/10/2022

Conflicto de interés: el autor declara no tener ningún conflicto de interés.

Martín Conejero A. Validación de test diagnósticos (primera parte). Test categóricos. Angiología 2023;75(2):101-105

DOI: <http://dx.doi.org/10.20960/angiologia.00473>

Correspondencia:

Antonio Martín Conejero. Hospital Clínico San Carlos.
C/ del Prof. Martín Lagos, s/n. 28040 Madrid
e-mail: amartinconejero.cto@gmail.com

- En la camilla de la urgencia tenemos un paciente con un eco fast sugerente de AAA roto. ¿Hay alguna forma de saber las probabilidades de que el paciente esté enfermo o sano si se informa de que el eco es positivo para hematoma?
- Sabemos que el diagnóstico de TVP se basa en dos test: dímero D y eco Doppler. ¿Por qué se hace primero uno y luego el otro y no al revés?

A continuación, vamos a intentar responder a cada una de estas situaciones que se nos plantean. Creo que no va a ser difícil dar respuesta a cada una de estas cuestiones.

CONCEPTO DE VALIDEZ

La validez de un test diagnóstico se refiere a la capacidad del test de identificar correctamente como enfermo al enfermo y como sano al sano. Antes de empezar a ver diferentes parámetros, hay que tener en cuenta que el punto de partida es la comparación del resultado obtenido en nuestro test diagnóstico con los resultados obtenidos por la prueba de referencia.

En la tabla I podemos ver las cuatro situaciones en las que puede encontrarse un sujeto en función del resultado de los dos test cualitativos: test a validar y prueba de referencia.

Tabla I. Clasificación de los sujetos de estudio

Enfermos	Sanos	Total
VP	FP	Total de positivos
FN	VN	Total de negativos
Total de enfermos	Total de sanos	Total de la muestra

Los parámetros que miden esta validez son la sensibilidad (S) y la especificidad (E). La S hace referencia al porcentaje de enfermos que tiene el test positivo dentro del grupo de enfermos. Es decir, que la condición o punto de partida es el total de enfermos (denominador), y de ellos, los que tienen nuestro test

o test a validar positivo. Por tanto, en el numerador se sitúan los verdaderos positivos (VP):

$$S = \frac{VP}{(\text{Total de enfermos})}$$

La E hace referencia al porcentaje de pacientes que resultan negativos entre el total de sujetos sanos, es decir, el cociente de verdaderos negativos (VN) entre el total de sanos:

$$E = \frac{VN}{(\text{Total de sanos})}$$

Es decir, un test sensible será aquel que considerará como enfermos a todos los sujetos (incluso a los que no lo son), con el riesgo de tener muchos falsos positivos (FP). Un test específico, sin embargo, considerará como sanos a la mayoría de los sujetos, con el riesgo que implica de falsos negativos. Ahora que conocemos los parámetros de S y E, la pregunta que debemos hacernos es cuándo usamos cada uno de estos parámetros.

Nos ayudarán a entender esto las tres siguientes situaciones diferentes en el diagnóstico de una trombosis venosa profunda (TVP).

- *Paciente con dímero D (test sensible) negativo.* Descartamos enfermedad. El riesgo son los FP; sin embargo, aparecido un negativo, apostamos a que viene de un sano, descartamos enfermedad y, por lo tanto, no tenemos que hacer más pruebas diagnósticas. Utilidad de un test sensible negativo: descartar enfermedad.
- *Paciente con dímero D positivo.* Como es muy sensible, el resultado puede venir de un sujeto sano y, por lo tanto, ser un FP. Con el fin de excluir estos FP, hacemos en un segundo tiempo un eco Doppler (test específico), que resulta negativo. Concluimos que el paciente no presenta TVP. Un test específico negativo elimina los falsos positivos de la prueba sensible.
- *Paciente con dímero D positivo y eco Doppler positivo.* Confirmamos la enfermedad, ya que los test específicos casi no tienen FP. Un test específico positivo confirma la enfermedad.

Resulta obvio que lo ideal sería trabajar con pruebas diagnósticas de alta sensibilidad y especificidad, pero esto no siempre es posible. En general, las prue-

bas de cribado deben ser de alta sensibilidad para que no se “escapen” enfermos. Una prueba muy sensible será especialmente adecuada en aquellos casos en los que no diagnosticar la enfermedad puede resultar fatal para los enfermos, como ocurre con enfermedades graves o tratables.

Por otra parte, las pruebas específicas tienen una mayor utilidad como confirmatorias del diagnóstico. Los test de alta especificidad son necesarios en enfermedades graves, pero sin tratamiento disponible y cuando los FP supongan un grave problema psicológico para el paciente.

Ya hemos visto que la S y la E miden la validez de una prueba diagnóstica. Hay que tener en cuenta que el punto de partida de ambos parámetros es el resultado de la prueba de referencia, del total de enfermos para la sensibilidad y del total de sanos para la especificidad.

Sin embargo, cuando a un paciente se le realiza una eco de abdomen para el diagnóstico de un AAA, el resultado puede ser positivo o negativo. Es decir, que a este paciente, de forma individual, no le importa tanto el valor de la S y la E, sino el valor de parámetros que tengan como condición o punto de partida ser positivos o negativos, es decir, los valores predictivos.

Los valores predictivos miden la probabilidad de estar enfermo o sano según el resultado del test que vamos a validar. En este caso, por tanto, el punto de partida es el resultado del test a validar, de forma que un valor predictivo positivo (VPP) alto significa que, si un sujeto tiene un test positivo, lo más probable es que tenga la enfermedad. De igual modo, un valor predictivo negativo (VPN) alto significa que, si un test es negativo, casi seguro estará sano.

Las fórmulas de ambos parámetros son:

$$VPP = \frac{VP}{(\text{Total de positivos})}$$

$$VPN = \frac{VN}{(\text{Total de negativos})}$$

Supongamos que la eco abdominal para el diagnóstico de aneurisma de aorta (AAA) tiene, en una determinada comunidad, un VPP del 90 %. Esto significará que, si un paciente tiene una ecografía abdominal positiva, la probabilidad de que realmente sea portador de un AAA es del 90 %. De la misma manera,

si se nos informa de que el VPN es del 95 %, podremos afirmar que un paciente con ecografía negativa tendrá un 95 % de probabilidades de no tener AAA.

Acabamos de ver cuatro parámetros esenciales en la validación de pruebas diagnósticas: S, E, VPP y VPN. La pregunta ahora que nos hacemos es si existe alguna relación entre ellos. Un test muy sensible implica que, aparecido un negativo, se descarta la enfermedad. Por ejemplo, si la eco de abdomen no muestra AAA, se descarta que el paciente sea portador de esta patología. Esto es exactamente lo mismo que mide el VPN: que, siendo negativa la prueba diagnóstica, el sujeto sea sano. De igual forma hemos dicho que un test específico sirve para quitar falsos positivos, de forma que los positivos restantes son VP. Es decir, mejoraríamos el VPP. En resumen: si la sensibilidad es alta, mejoramos el VPN del test diagnóstico; si la especificidad es alta, mejoramos el VPP.

¿CÓMO INFLUYE LA PREVALENCIA EN LOS RESULTADOS DE UN TEST DIAGNÓSTICO?

Supongamos que dos pacientes son remitidos a su consulta porque en Atención Primaria se les ha hecho un índice tobillo-brazo (ITB) dentro de un programa de cribado en pacientes asintomáticos que ha resultado de 0,6. Un paciente tiene 85 años (supongamos una prevalencia de enfermedad arterial periférica [EAP] del 40 %), el otro es su hijo de 60 años (supongamos una prevalencia de enfermedad del 20 % en este rango de edad). ¿Debemos pensar que como en ambas el resultado del ITB es patológico la probabilidad de enfermedad arterial es la misma en ambos pacientes?

Evidentemente, la probabilidad de EAP será mayor para el paciente de 85 años, aunque el resultado de la prueba haya sido similar en ambos casos. Es decir, aparecido un positivo en el ITB, la probabilidad de EAP es diferente en función de la edad del paciente. Esto se debe a que los valores predictivos son parámetros dependientes de la prevalencia.

Cuando la prevalencia de una determinada enfermedad es muy elevada, pensamos que un test positivo muy probablemente provenga de un enfermo, o lo que es lo mismo, aparecido un positivo, asumimos que muy probablemente se tratará de un VP.

Al contrario, un test negativo para una enfermedad muy prevalente no nos merecerá credibilidad y pensaremos que un test negativo casi seguro será un FN. Se cumple, por tanto:

$$\uparrow \text{prevalencia} = \uparrow \text{VPP} \text{ y } \downarrow \text{VPN}$$

De igual forma, si realizamos un test diagnóstico en una comunidad de muy baja prevalencia, un test positivo probablemente será un FP. Sin embargo, un test negativo casi seguro será un VN, cumpliéndose:

$$\downarrow \text{prevalencia} = \downarrow \text{VPP} \text{ y } \uparrow \text{VPN}$$

Sirva como ejemplo lo que mostramos en las tablas II y III. Modificando la prevalencia de enfermedad, se modifican los valores predictivos, tal y como hemos comentado anteriormente. Nótese que S y E no se han modificado en ambas tablas, aunque sí se han modificado los valores predictivos por ser dependientes de prevalencia.

Tabla II. Parámetros de validez para una prevalencia de enfermedad del 20 %

	Enfermos	Sanos	Total		
+	16	8	24	Sensibilidad	80,00
-	4	72	76	Especificidad	90,00
Total	20	80	100	VPP	66,67
				VPN	94,714

Tabla III. Parámetros de validez para una prevalencia de enfermedad del 40 %

	Enfermos	Sanos	Total		
+	32	6	38	Sensibilidad	80,00
-	8	54	62	Especificidad	90,00
Total	40	60	100	VPP	84,21
				VPN	87,10

RAZONES DE VEROSIMILITUD Y COCIENTES DE PROBABILIDAD (O LIKELIHOOD RATIO)

Como se ha comentado anteriormente, los valores predictivos son dependientes de prevalencia,

por lo que un alto VPP puede deberse más a la aplicación del test en una comunidad con elevada prevalencia que a la bondad del test diagnóstico. Las razones de verosimilitud tienen la ventaja de informar sobre la bondad del test diagnóstico sin que el resultado esté influido por la prevalencia. Estos parámetros miden cuánto más probable es un resultado concreto (positivo o negativo) según la presencia o ausencia de enfermedad.

La razón de verosimilitud positiva (RV+) nos informa de cuánto es más probable que un test positivo venga de un enfermo con respecto a la probabilidad de que venga de un sano. Por ello, nos interesa una RV+ alta, es decir, que cuanto más alto sea el valor, mayor es la apuesta que hacemos de que un positivo provenga de un enfermo. Su fórmula es la siguiente:

$$RV+ = \frac{S}{(1 - E)}$$

Es decir, aproximadamente verdaderos positivos / falsos positivos (VP/FP). Imaginemos que la RV+ del eco Doppler carotídeo es 7. Esto significa que, si un test es patológico, el paciente tendrá 7 veces más probabilidades de ser enfermo que de ser sano, de tener una estenosis carotídea que de no tenerla.

La razón de verosimilitud negativa (RV-) es un cociente en el que los falsos negativos se ubican en el numerador y los verdaderos negativos se sitúan en el denominador. Sería aproximadamente un cociente falsos negativos / verdaderos negativos (FN/VN).

$$RV- = \frac{(1 - S)}{E}$$

Supongamos que una RV- es 0,2. Recordemos que el cociente es FN/VN. Para interpretarlo de forma sencilla, lo haremos a partir de su inverso, por lo que en el numerador tendremos los VN.

Es decir, FN/VN = 0,2. Si hacemos el inverso en ambos lados de la ecuación no la alteramos, pero nos facilita mucho su interpretación, de forma que nos quedaría VN/FN = 1/0,2 = 5.

Diremos en consecuencia que si una RV- es 0,2, aparecido un test negativo, este es 5 veces más probable que venga de un sano a que venga de un enfermo.

Como hemos comentado, las razones de verosimilitud sirven para determinar la probabilidad

de enfermo-sano, dependiendo del resultado de un test diagnóstico. Por ello, se emplean para cuantificar la probabilidad de cambio en nuestra decisión clínica. Por ejemplo, si un dímero D en el diagnóstico de trombosis venosa profunda tiene una RV positiva de 1,5, la probabilidad de que hagamos un eco Doppler de confirmación será baja. Sin embargo, si la RV es de 8, casi con seguridad sí realizaremos el eco Doppler porque la apuesta de TVP es muy alta (Tabla IV).

En esta ocasión hemos visto cómo calcular e interpretar la bondad de una prueba diagnóstica categórica o cualitativa. Es decir, las pruebas en las que la respuesta es "positivo/negativo".

En el siguiente artículo conoceremos si es mejor la velocidad sistólica que la diastólica en el diagnóstico de estenosis carotídea, si el valor de ITB es un buen predictor de mortalidad o qué valor de diámetro de aorta en un AAA es predictor de rotura. Es decir, analizaremos la interpretación de las pruebas diagnósticas cuantitativas.

Tabla IV. Cambios en la decisión clínica en función de las razones de verosimilitud

Valor de CP+	1-2	2-5	5-10	> 10
Valor de CP-	1-0,5	0,5-0,2	0,2-0,1	< 0,1
Interpretación	Variación insignificante	Pequeños cambios	Cambios moderados	Grandes cambios concluyentes

CP: cociente de probabilidad.