



## Carta metodológica

# Pruebas diagnósticas. ¿Cómo describir su validez?

## Diagnostic tests. How to describe its validity?



Juan-José Sancho-Insenser y Ana María González-Castillo \*

Servicio de Cirugía General y Digestiva, Hospital del Mar, Instituto Hospital del Mar de Investigaciones Médicas (IMIM),  
Universitat Autònoma de Barcelona, Barcelona, España

Los cirujanos encaramos rutinariamente el desafío de formular un diagnóstico, y para ello utilizamos toda la información disponible, desde la anamnesis y la exploración física a —cada vez más exclusivamente—, todo tipo de pruebas diagnósticas. En términos generales, entendemos por prueba diagnóstica cualquier fuente de información que puede ser usada para confirmar o descartar la presencia de una enfermedad. Además de confirmar la presencia o ausencia de una enfermedad, las pruebas diagnósticas también aportan información sobre la intensidad de una dolencia y/o su pronóstico<sup>1</sup>. Es esencial conocer la validez de las pruebas diagnósticas, tanto al solicitarlas e interpretarlas para un paciente en concreto, como para incluirlas en un algoritmo diagnóstico y/o terapéutico dentro de un protocolo o guía clínica.

### Necesidad de investigación sobre las pruebas diagnósticas

Dado el enorme beneficio potencial que se deriva de la utilización de las pruebas diagnósticas, ¿no deberíamos utilizarlas masivamente hasta llegar a un diagnóstico exacto? Desgraciadamente, la mayoría de las pruebas diagnósticas no son infalibles y exactas a la vez y, por tanto, raramente confirman o descartan un diagnóstico en la totalidad de la población a la que se aplican. Es decir, tienen una validez limitada. Además de la validez, se deben investigar en cualquier nueva prueba diagnóstica aspectos como sus requerimientos de recursos (tanto económicos como de tiempo profesional, logístico y administrativo), el impacto físico, emocional y económico sobre el paciente en los casos de

falsos positivos o negativos, así como la seguridad derivada de los efectos secundarios inherentes a la prueba diagnóstica en sí, que pueden ir desde una pequeña equimosis en el pliegue del codo a un desenlace fatal por *shock* anafiláctico o perforación de víscera hueca. Finalmente, al investigar una nueva prueba diagnóstica se debe evaluar cuidadosamente su trascendencia clínica. Su utilidad será proporcional solo al cambio de actitud en el manejo del paciente que se siga de su resultado<sup>2</sup>.

La investigación sobre la validez y la seguridad de las pruebas diagnósticas es esencial para el cirujano, que muy frecuentemente fundamenta su indicación y técnica quirúrgicas en los resultados de dichas pruebas. La validez de una prueba diagnóstica es su capacidad para identificar las personas que padecen la enfermedad frente aquellas que no la padecen. El tipo de estudio más utilizado para evaluar la validez de una prueba diagnóstica es el diseño de 2 grupos de individuos, un grupo que padece la enfermedad y otro sin ella. Para determinar la validez de la prueba diagnóstica se comparan los resultados obtenidos con la prueba a evaluar en dichos individuos con un criterio de referencia o patrón de referencia (*gold standard*)<sup>3,4</sup>. Según el tipo variable generado por la prueba diagnóstica distinguiremos 2 situaciones (tablas 1 y 2):

### Validez de las pruebas diagnósticas con resultados dicotómicos

Una prueba dicotómica clasifica a cada paciente como sano o enfermo en función de que su resultado sea positivo o

\* Autor para correspondencia.

Correo electrónico: [amgonzalezcastillo@psmar.cat](mailto:amgonzalezcastillo@psmar.cat) (A.M. González-Castillo).

<https://doi.org/10.1016/j.ciresp.2022.02.006>

0009-739X/© 2022 AEC. Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

**Tabla 1 – Relación entre el resultado de una prueba diagnóstica y el diagnóstico verdadero de una enfermedad**

|                                 |          | Diagnóstico verdadero            |                               |
|---------------------------------|----------|----------------------------------|-------------------------------|
|                                 |          | Enfermos                         | Sanos                         |
| Resultado de la prueba          | Positiva | Enfermos con prueba +<br>o<br>VP | Sanos con prueba +<br>o<br>FP |
|                                 | Negativa | Enfermos con prueba –<br>o<br>FN | Sanos con prueba –<br>o<br>VN |
| Sensibilidad:                   |          | VP/VP + FN                       | VP/enfermos                   |
| Especificidad:                  |          | VN/VN + FP                       | VN/sanos                      |
| Valor predictivo positivo:      |          | VP/VP + FP                       | VP/positivos                  |
| Valor predictivo negativo:      |          | VN/VN + FN                       | VN/negativos                  |
| Cociente probabilidad positivo: |          | Sensibilidad/100-especificidad   |                               |
| Cociente probabilidad negativo: |          | 100-sensibilidad/especificidad   |                               |
| Índice de Youden:               |          | Sensibilidad + especificidad-1   |                               |
| Valor predictivo global:        |          | VN + VP/total                    | Aciertos/todos                |

FN: falsos negativos; VP: verdaderos positivos.

**Tabla 2 – Principales características y limitaciones de las pruebas utilizadas para describir la validez de una prueba diagnóstica con resultados dicotómicos**

|   | Característica   | Limitación   |
|---|--|--|
| Sensibilidad y especificidad                    | Describen bien la validez de una prueba diagnóstica.<br>Ideal para el epidemiólogo.  | Precisan conocer el diagnóstico real antes de calcularse.  |
| Valor predictivo (positivo y negativo)          | No precisan conocer el diagnóstico real antes de calcularse.<br>Ideal para el clínico.   | Son muy sensibles a valores extremos de la prevalencia.<br>No sirven para comparar pruebas en poblaciones distintas. |
| Cocientes de probabilidad (positivo y negativo) | No precisan conocer el diagnóstico real antes de calcularse.<br>Son independientes de la prevalencia.<br>Ideales para el investigador. | Deben interpretarse correctamente.   |
| Valor predictivo global                         | Resume en una cifra la validez de una prueba.  | Da el mismo peso a especificidad y sensibilidad. Valor muy limitado.   |

negativo. El resultado positivo indica la presencia de enfermedad y el resultado negativo su ausencia.

La validez diagnóstica se cuantifica con las proporciones clásicas: sensibilidad y especificidad. Suelen expresarse como porcentajes, y objetivan la capacidad de la prueba para clasificar correcta o erróneamente a una persona, según la presencia o ausencia de una enfermedad.

Los datos obtenidos distribuyen los individuos en 4 grupos, usualmente representados en una tabla 2x2 en la que se cruza el resultado de la prueba diagnóstica (en las filas) con el estado real de los individuos (en las columnas). El resultado de la prueba puede ser correcto (verdadero positivo y verdadero negativo) o incorrecto (falso positivo y falso negativo).

En muchas ocasiones el diagnóstico real y absoluto de todos los individuos es desconocido, y se toma como comparador el resultado de la prueba de referencia (*gold standard*) utilizada para el diagnóstico hasta la aparición de la nueva prueba diagnóstica a evaluar.

La **sensibilidad** es la capacidad de la prueba para detectar la enfermedad. Matemáticamente es la proporción de individuos con la enfermedad que obtienen un resultado positivo, es decir

la probabilidad de clasificar correctamente a un individuo enfermo.

$$\text{Sensibilidad} = \frac{\text{Verdaderos Positivos}}{\text{Verdaderos positivos} + \text{Falsos Negativos}}$$

$$= \frac{\text{Verdaderos Positivos}}{\text{Todos los Enfermos}}$$

Podemos pues referirnos a ella como la fracción de verdaderos positivos.

Las pruebas diagnósticas orientadas al cribado deben tener una sensibilidad alta a fin de poder detectar a todos los enfermos, aún a costa de algunos falsos positivos. Una prueba muy sensible resulta especialmente apropiada para las situaciones en las que no diagnosticar la enfermedad resulta peligroso para el paciente, como en el caso de enfermedades peligrosas pero tratables, o en enfermedades en las que un falso positivo no origina serios trastornos al paciente. Antes de iniciar un tratamiento, suelen requerir una prueba confirmatoria, que idealmente tendrá una alta especificidad.

La **especificidad** es la capacidad de una prueba para detectar a los sanos. Matemáticamente es la proporción de individuos sin la enfermedad que presentan un resultado

negativo e indica la utilidad de la prueba para identificar a los individuos que no tienen la enfermedad.

$$\text{Especificidad} = \frac{\text{Verdaderos Negativos}}{\text{Verdaderos Negativos} + \text{Falsos Positivos}} = \frac{\text{Verdaderos Negativos}}{\text{Todos los Sanos}}$$

Podemos también referirnos a ella como la fracción de verdaderos negativos.

Las pruebas confirmatorias de un diagnóstico deben tener una elevada especificidad, a fin de eludir los falsos positivos. Las pruebas de elevada especificidad son esenciales cuando existe gran interés por conocer la ausencia de enfermedad, en enfermedades graves, pero sin tratamiento que las haga curables, o cuando diagnosticar a un paciente de una enfermedad que realmente no padece pueda acarrear graves consecuencias, ya sean físicas, psicológicas o económicas<sup>5</sup>.

Sensibilidad y especificidad son propiedades inherentes de la prueba diagnóstica y, por tanto, son independientes de la prevalencia de la enfermedad en la población a estudio<sup>2</sup>. La sensibilidad y especificidad valoran la validez de una prueba diagnóstica, pero son conceptos no aplicables directamente en la práctica clínica diaria. Estos parámetros podrán servir para recomendar una prueba en un protocolo o guía clínica, pero no son útiles para valorar el resultado de una prueba en un paciente concreto. La razón es que tanto sensibilidad como especificidad valoran la probabilidad de obtener un resultado concreto (positivo o negativo) en función del verdadero diagnóstico. Sin embargo, cuando a un paciente se le realiza una prueba diagnóstica, desconocemos su verdadero diagnóstico (¡por eso le pedimos la prueba!). El clínico se plantea por tanto la pregunta en sentido contrario: si la prueba sale positiva, ¿cuál es la probabilidad de que el paciente esté realmente enfermo? Debemos calcular los valores predictivos para obtener esta información<sup>3</sup>.

## Valores predictivos positivo y negativo

El **valor predictivo positivo** es la probabilidad de que un individuo con un resultado positivo tenga la enfermedad. Matemáticamente es la proporción de pacientes con un resultado positivo en la prueba que están realmente enfermos<sup>4</sup>.

### Valor Predictivo Positivo

$$= \frac{\text{Verdaderos Positivos}}{\text{Verdaderos Positivos} + \text{Falsos Positivos}} = \frac{\text{Verdaderos Positivos}}{\text{Todos los Positivos}}$$

El **valor predictivo negativo** es la probabilidad de que un individuo con resultado negativo no tenga en realidad la enfermedad. Matemáticamente es la proporción de pacientes con un resultado negativo que están realmente sanos<sup>4</sup>.

### Valor Predictivo Negativo

$$= \frac{\text{Verdaderos Negativos}}{\text{Verdaderos Negativos} + \text{Falsos Negativos}} = \frac{\text{Verdaderos Negativos}}{\text{Todos los Negativos}}$$

## El efecto de la prevalencia sobre los valores predictivos

Los valores predictivos varían según la prevalencia de la enfermedad y, por tanto, valoran tanto la validez de la prueba diagnóstica como la prevalencia de la enfermedad en la población estudiada y, por tanto, el clínico debe conocer al valorarlos la prevalencia de la enfermedad que pretende confirmar o descartar en la población a la que pertenece el individuo estudiado, y corregir consecuentemente su valoración a la luz de la prevalencia. Prevalencias muy bajas hacen que un resultado negativo pueda descartar la enfermedad con mayor seguridad y, por tanto, tendrán un valor predictivo negativo artificialmente elevado. En el escenario opuesto, prevalencias muy altas hacen que una prueba positiva tenga muchas probabilidades de ser cierta y, por tanto, arroje un valor predictivo positivo engañosamente alto si se aplica a otra población con una prevalencia menor.

Todos estos porcentajes tienen un intervalo de confianza que puede calcularse fácilmente, y que debería acompañar siempre tanto a los valores de especificidad, sensibilidad como a los valores predictivos positivo y negativo. Todos estos porcentajes deben siempre valorarse conjuntamente: sensibilidad con especificidad y valor predictivo positivo con valor predictivo negativo, pues cada par es como las 2 caras de una moneda. Sin embargo, para obviar la dependencia de la prevalencia se han propugnado una serie de parámetros que combinando los anteriores, reflejen la validez de una prueba diagnóstica en una sola cifra y sean independientes de la prevalencia de la enfermedad a diagnosticar: las razones de probabilidad,

## Razones de probabilidad

Como sea que la prevalencia es un factor muy influyente en los valores predictivos de un test, estos no pueden emplearse como medida para comparar 2 métodos diagnósticos diferentes. Por ello, resulta necesario utilizar otros parámetros de valoración que sean a la vez clínicamente útiles y no dependan de la prevalencia de la enfermedad en la población a estudiar: Son la razón de verosimilitudes, razón de probabilidad o cociente de probabilidades positivo y negativo, que comparan la probabilidad de obtener un determinado resultado (positivo o negativo) en un individuo que presente la enfermedad, con la de obtenerlo en un sujeto en el que se ha descartado la presencia de la misma<sup>6</sup>.

El **cociente de probabilidad de una prueba positiva (CPP) o razón de verosimilitud positiva (Positive Likelihood Ratio)** se calcula dividiendo la proporción de casos que tienen resultado positivo (sensibilidad) entre la proporción de personas que no tienen la enfermedad, pero en los que la prueba ha dado un resultado positivo (100-especificidad, falsos positivos).

$$\text{CPP} = \frac{\text{Sensibilidad}}{100 - \text{Especificidad}} = \frac{\text{Fracción de Verdaderos Positivos}}{\text{Fracción de Falsos Positivos}}$$

El **cociente de probabilidad de una prueba negativa (CPN) o razón de verosimilitud negativa (Negative Likelihood Ratio)** se

calcula dividiendo la proporción de casos que tienen resultado negativo en presencia de la enfermedad (falsos negativos) entre la probabilidad de un resultado negativo en ausencia de la misma.

$$CPN = \frac{100 - \text{Sensibilidad}}{\text{Especificidad}} = \frac{\text{Fracción de Falsos Negativos}}{\text{Fracción de Verdaderos Negativos}}$$

Los cocientes de probabilidad ofrecen la ventaja de que relacionan sensibilidad y especificidad de una prueba en un solo índice y no dependen de la prevalencia. Esto permite utilizarlos para comparar diferentes pruebas para un mismo diagnóstico. Pueden calcularse cocientes de probabilidad según varios niveles de una nueva medida y no es necesario expresar la información de forma dicotómica (positivo o negativo)<sup>7</sup>.

### Otros valores sintéticos para valorar la validez de una prueba diagnóstica

#### Valor predictivo global

Otro de los parámetros que fueron utilizados para describir en una sola cifra la validez de una prueba es el valor predictivo global (*Overall Accuracy*). A pesar de su imponente denominación, resulta el parámetro menos objetivo para describir la validez de una prueba diagnóstica con resultado dicotómico.

#### Valor Predictivo Total

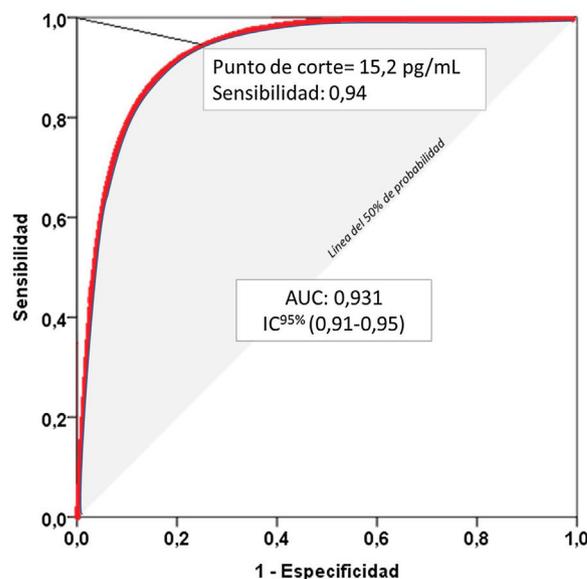
$$= \frac{\text{Verdaderos Positivos} + \text{Verdaderos Negativos}}{\text{Total}}$$

Como medidas sintéticas, —en una sola cifra—, de la validez de una prueba diagnóstica, asumen que tienen la misma importancia la sensibilidad y la especificidad para el caso que nos ocupa y, por tanto, son de utilidad limitada a casos concretos en los que sólo disponemos de una prueba para sospechar el diagnóstico y para confirmarlo. Son claramente inferiores a las razones de verosimilitud y deberían actualmente estar en desuso<sup>8</sup>.

#### Validez de pruebas diagnósticas con resultados de valores continuos

Los resultados de muchas pruebas diagnósticas son valores numéricos continuos (p. ej., lipasa sérica). En ellas debemos decidir a partir de qué valor numérico o punto de corte, los resultados se considerarán positivos o negativos. Se debe buscar el compromiso entre una mayor sensibilidad y menor especificidad o a la inversa, dependiendo de la intención concreta de la prueba diagnóstica. Para calibrar su validez pueden dicotomizarse a partir de ese punto de corte, pero es mucho mejor utilizar las curvas de características funcionales<sup>5</sup>.

Las curvas de características funcionales (Receiver-operating characteristic [ROC]) representan gráficamente los pares de valores de sensibilidad y especificidad resultantes de la variación continua de los puntos de corte en todo rango de resultados observados. El área bajo la curva ROC (Area Under



**Figura 1 – Representación de una curva de características funcionales o curva Receiver-Operating Characteristic (ROC). Area Under the Curve (AUC): Área bajo la curva, representa la validez diagnóstica positiva de la prueba. Cuanto más cercana a la unidad, mejor capacidad de distinguir entre enfermos y sanos, sin conocimiento previo del verdadero diagnóstico y sin influencia de la prevalencia de la enfermedad a diagnosticar. Intervalo de confianza del 95% (IC 95%) para el AUC.**

the Curve [AUC]) oscila entre 0 y 1, y es un excelente indicador de la validez diagnóstica de la prueba, de modo que a mayor valor de esta área mayor capacidad de la prueba de discriminar correctamente entre enfermos y no enfermos<sup>9</sup>.

La representación gráfica de varias curvas ROC permite una apreciación directa de las diferencias entre varias pruebas diagnósticas o entre varios puntos de corte de una misma prueba diagnóstica. Se suele incorporar a su representación gráfica una línea diagonal que representa un AUC de 0,5, o lo que es lo mismo, la probabilidad de dar un resultado positivo lanzando una moneda al aire. Por tanto, para interpretar correctamente el AUC de una curva ROC debemos restar 0,5 del AUC. Así, una AUC de 0,750 puede parecer convincente, pero significa que la prueba diagnóstica que estamos evaluando mejora solo en un 25% el resultado que obtendríamos lanzando una moneda al aire (área sombreada en la figura 1).

En conclusión, es primordial valorar la validez y seguridad de las pruebas diagnósticas con el fin de seleccionar la más adecuada en cada situación clínica.

#### BIBLIOGRAFÍA

1. Sackett D, Haynes R, Tugwell P. *Epidemiología clínica*. Madrid: Díaz de Santos. 1989.
2. Escrig-Sos J, Martínez-Ramos D, Miralles-Tena J. Pruebas diagnósticas: nociones básicas para su correcta

- interpretación y uso. *Cir Esp.* 2006;79:267-73. [http://dx.doi.org/10.1016/s0009-739x\(06\)70904-6](http://dx.doi.org/10.1016/s0009-739x(06)70904-6).
3. Altman D, Bland J. Diagnostic tests. 1 Sensitivity and specificity. *BMJ.* 1994;308:1552. <http://dx.doi.org/10.1136/bmj.308.6943.1552>.
  4. Altman D, Bland J. Diagnostic test - Predictive values. *BMJ.* 1994;309:102. <http://dx.doi.org/10.1016/j.ebobjyn.2005.10.001>.
  5. Pita-Fernández S, Pértegas-Díaz S. Pruebas diagnósticas. *Cad Aten Primaria.* 2003;10:120-4.
  6. Moskowitz CS, Pepe MS. Comparing the predictive values of diagnostic tests: Sample size and analysis for paired study designs. *Clin Trials.* 2006;3:272-9. <http://dx.doi.org/10.1191/1740774506cn147oa>.
  7. María Eugenia Burgos D, Carlos Manterola D. Cómo interpretar un artículo sobre pruebas diagnósticas. *Rev Chil Cir.* 2010;62:301-8. <http://dx.doi.org/10.4067/s0718-40262010000300018>.
  8. Alberg AJ, Park JW, Hager BW, Brock MV, Diener-West M. The use of «overall accuracy» to evaluate the validity of screening or diagnostic tests. *J Gen Intern Med.* 2004;19:460-5. <http://dx.doi.org/10.1111/j.1525-1497.2004.30091.x>.
  9. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. *Clin Chem.* 1993;39:561-77. <http://dx.doi.org/10.1093/clinchem/39.4.561>.