



# Uso de tests diagnósticos en la práctica clínica (Parte 1). Análisis de las propiedades de un test diagnóstico

Jaime Cerda L. y Lorena Cifuentes A.

## Clinical use of diagnostic tests (Part 1). Analysis of the properties of a diagnostic test

Diagnostic tests are widely used in clinical practice. Qualitative analysis of diagnostic test properties (validity, reproducibility and safety) has been systematized using clinical epidemiology tools that have introduced a series of concepts such as sensitivity, specificity and predictive values. Based on data reported by a clinical investigation, the present article illustrates the theoretical background underlying these concepts, explains the methodology for their calculation, and analyzes their main strengths and limitations. A second article (Part 2: Clinical application and usefulness of a diagnostic test), describes the manner in which information provided by a test should be applied in order to correctly resolve a clinical dilemma.

**Key words:** Diagnostic tests, Clinical Epidemiology, sensitivity, specificity, predictive values.

**Palabras clave:** Tests diagnósticos, epidemiología clínica, sensibilidad, especificidad, valores predictores.

**Pontificia Universidad  
Católica de Chile.**

Facultad de Medicina  
Departamento de Salud  
Pública (JCL).  
Departamento de Pediatría (LCA).  
Programa de Salud Basada en la  
Evidencia (JCL y LCA).

Recibido: 15 de septiembre de  
2009

Aceptado: 28 de abril de 2010

**Correspondencia a:**

Jaime Cerda L.  
jcerda@med.puc.cl

## Introducción

Los tests diagnósticos son herramientas complementarias ampliamente utilizadas en la práctica clínica para sustentar nuestras hipótesis diagnósticas. La diversa gama de exámenes de laboratorio e imagenológicos disponibles obliga al clínico a analizar críticamente sus propiedades (validez, reproducibilidad y seguridad) con el fin de escoger aquel más óptimo para el paciente en términos clínicos, económicos y valóricos<sup>1</sup>. Hoy en día, términos tales como “sensibilidad”, “especificidad” y “valores predictores” son utilizados frecuentemente en el vocabulario médico para referirse al análisis de las propiedades de un test diagnóstico. El presente artículo explica en términos generales los fundamentos teóricos subyacentes al análisis de las propiedades de un test diagnóstico; para tales efectos, se ejemplificarán algunos conceptos en base a los resultados de una investigación clínica, comenzando por la descripción del proceso diagnóstico. Aspectos referentes a la aplicación clínica de los resultados de un test serán abordados en segundo artículo (Parte 2: Aplicación clínica y utilidad de un test diagnóstico).

## El proceso diagnóstico

El llamado “proceso diagnóstico” describe la secuencia de ideas y acciones llevadas a cabo por un médico cuando se enfrenta a un paciente con un diagnóstico desconocido<sup>2</sup>.

En base a la información obtenida vía anamnesis y examen físico, complementada por su nivel de conocimientos y experiencia previa, el médico formula una determinada hipótesis diagnóstica que de explicación al cuadro clínico del paciente. Toda hipótesis diagnóstica emanada de este proceso tiene asociada una probabilidad de ocurrencia, la cual recibe el nombre de probabilidad pre test (el prefijo ‘pre’ hace referencia a que es una probabilidad planteada antes de la realización de un test diagnóstico). Si a juicio del médico esta probabilidad no supera un primer umbral, denominado ‘umbral de estudio adicional’, éste rechazará la hipótesis inicialmente planteada, descartando el diagnóstico. Por el contrario, si su juicio le indica que la probabilidad pre test supera tanto el umbral de estudio adicional como un segundo umbral, denominado ‘umbral terapéutico’, el médico apoyará la hipótesis sin necesidad de realizar tests diagnósticos confirmatorios, dando paso a la prescripción de la terapia pertinente. Existe un tercer escenario, situado entremedio de los umbrales de estudio adicional y terapéutico, en el cual la probabilidad pre test impide apoyar o rechazar la hipótesis diagnóstica, pues no se encuentra por debajo del primer umbral ni logra superar el segundo; en otras palabras, la probabilidad de que el paciente presente el diagnóstico planteado se encuentra en una ‘zona de incertidumbre’ (Figura 1). Este escenario plantea la necesidad de realizar uno o más tests diagnósticos, cuyo fin no es otro que modificar la probabilidad pre test, permitiendo idealmente su desplazamiento por sobre el umbral terapéutico o bien por debajo del umbral de estudio adicional, alejándonos así de la zona de incer-



tidumbre. La nueva probabilidad, estimada a la luz de los resultados de los exámenes practicados, recibe el nombre de probabilidad post test<sup>2</sup>. Esta probabilidad es la que finalmente nos ayudará a tomar alguna decisión en relación a nuestro paciente, ya sea iniciar tratamiento, descartar el diagnóstico o bien considerar nuevos exámenes en caso de permanecer en la 'zona de incertidumbre'. La utilidad de un test diagnóstico se sustenta, por lo tanto, en la magnitud con que éste logre modificar la probabilidad pre test, apoyando la toma de decisiones. Esta característica se dimensiona mediante el cálculo de *likelihood ratios* (i.e. razones de verosimilitud) de un test, concepto que será abordado en la segunda parte del presente artículo.

Las propiedades inherentes a un test diagnóstico se caracterizan a través de su sensibilidad y especificidad. Para calcular el valor de estos estimadores, el test diagnóstico se contrasta con un estándar de oro, definido como aquel método, procedimiento o medición ampliamente aceptado como el mejor disponible para el diagnóstico de dicha enfermedad<sup>3</sup>. (e.g. biopsia, cirugía, autopsia y/o seguimiento a largo plazo). Dos requisitos básicos (entre otros) para realizar esta comparación son que: 1) el test y el estándar de oro deben ser evaluados en forma ciega entre sí (con el objetivo de que el conocimiento del resultado de uno no influya en la interpretación del resultado del otro) y que; 2) tanto los resultados del test como los del estándar de oro sean categóricos, exhaustivos (i.e. que abarquen

todos los resultados posibles) y mutuamente excluyentes (e.g. positivo vs negativo)<sup>4</sup>. En caso de tratarse de mediciones continuas (e.g. proteína C reactiva, recuento de leucocitos), es necesario categorizarlas en un cierto número de niveles mediante el establecimiento de puntos de corte, cuya determinación se realiza arbitrariamente o bien construyendo curvas ROC (*Receiver Operating Characteristic Curves*)<sup>5</sup>, excediendo su descripción el objetivo del presente artículo. En la práctica, no siempre se cuenta con un estándar de oro contra el cual contrastar el test diagnóstico, presentando este escenario una dificultad en el cálculo de la verdadera sensibilidad y especificidad del test. Ante tal escenario, la presencia de la enfermedad se determina bajo criterios clínicos, con el riesgo de que los valores de sensibilidad y especificidad calculados no sean los correctos. Los resultados de la siguiente investigación clínica permitirán explicar los fundamentos teóricos y cálculos matemáticos utilizados para evaluar las propiedades de un test diagnóstico.

### Ejemplo

En 1993, Steed y cols<sup>6</sup>, evaluaron la sensibilidad y especificidad del test directo de *Streptococcus pyogenes* (*Group A Streptococcus Direct Test, GASD*) para el diagnóstico de faringoamigdalitis estreptocócica. Para ello, se tomó un cultivo faríngeo a 277 individuos consultantes pediátricos, siendo sus resultados contrastados con aquellos aportados por el GASD (detección de cARN estreptocócico en secreción faríngea). La presencia de un cultivo positivo para *S. pyogenes* fue considerado como el estándar de oro para el diagnóstico de faringoamigdalitis estreptocócica. Por su parte, el GASD podía arrojar un resultado positivo o negativo, denotando la presencia o ausencia de material genético de la bacteria. La Tabla 1 resume los resultados de este estudio.

### Sensibilidad y especificidad

El análisis de la validez de un test diagnóstico, es decir, el grado en que el test mide lo que supuestamente debe medir, contempla el cálculo de su sensibilidad y especificidad<sup>1</sup>. La sensibilidad de un test diagnóstico corresponde a la proporción de individuos enfermos, identificados como tal por el test<sup>7-9</sup>. En el ejemplo, de un total de 59 individuos enfermos (definidos como tal por el cultivo faríngeo, el estándar de oro), el GASD identificó correctamente a 51 de ellos. De este modo, la sensibilidad del GASD para el diagnóstico de faringoamigdalitis estreptocócica es  $51/59 = 0,86$  (86%). Por su parte, la especificidad de un test diagnóstico corresponde a la proporción de individuos sin la enfermedad, identificados como tal por el test<sup>7-9</sup>. En el ejemplo, de un total de 218 individuos sin la enfermedad (definidos como tal por el cultivo faríngeo,

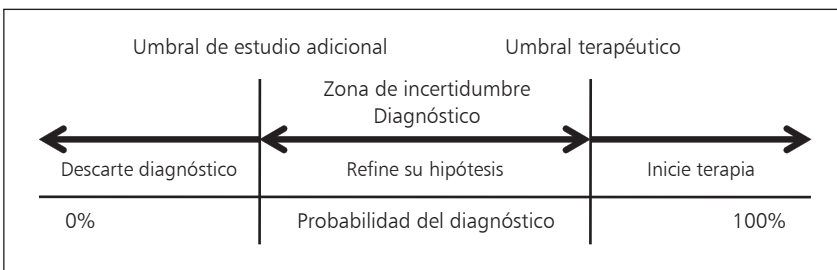


Figura 1. Umbrales de decisión presentes en el proceso diagnóstico.

Tabla 1. Resultados del cultivo faríngeo y test directo de <i>Streptococcus pyogenes</i> (GASD) en 277 individuos consultantes <sup>6</sup>			
GASD	Cultivo faríngeo		Total
	Positivo	Negativo	
Positivo	51	10	61
Negativo	8	208	216
Total	59	218	277

Prevalencia de la enfermedad =  $(59/277) = 21\%$ ; Sensibilidad =  $(51/59) = 86\%$ ; Especificidad =  $(208/218) = 95\%$ ; VPP =  $(51/61) = 84\%$ ; VPN =  $(208/216) = 96\%$ .



**Tabla 2. Comparación de los resultados de un test diagnóstico con el estándar de oro<sup>8,9</sup>**

GASD	Estándar de oro (+)	Estándar de oro (-)
Test (+)	a Verdaderos positivos	b Falsos positivos
Test (-)	c Falsos negativos	d Verdaderos negativos

Sensibilidad =  $(a/a+c)$ ; Especificidad =  $(d/b+d)$ ; Valor predictor positivo =  $(a/a+b)$ ; Valor predictor negativo =  $(d/c+d)$

el estándar de oro), el GASD identificó correctamente a 208 de ellos. De este modo, la especificidad del GASD para el diagnóstico de faringoamigdalitis estreptocócica es  $208/218 = 0,95$  (95%).

Si se interpretan estos resultados desde una óptica diferente, 51 pacientes fueron correctamente diagnosticados como enfermos por el GASD (pacientes verdaderamente positivos), mientras que 8 pacientes fueron incorrectamente diagnosticados como sanos por el test (pacientes falsamente negativos). La probabilidad de obtener un resultado verdaderamente positivo (sinónimo de 'sensibilidad') y la probabilidad de obtener un resultado falsamente negativo son valores mutuamente excluyentes, debiendo sumar 1,00 (en el ejemplo,  $51/59 + 8/59 = 1,00$ ). En consecuencia, es posible afirmar que a mayor sensibilidad de un test diagnóstico, menor es la probabilidad de que éste arroje un resultado falsamente negativo. Por otro lado, 208 individuos fueron correctamente diagnosticados como sanos por el GASD (individuos verdaderamente negativos), mientras que 10 individuos fueron incorrectamente diagnosticados como enfermos por éste (individuos falsamente positivos). La probabilidad de obtener un resultado verdaderamente negativo (sinónimo de 'especificidad') y la probabilidad de obtener un resultado falsamente positivo también son mutuamente excluyentes, debiendo sumar 1,00 (en el ejemplo,  $208/218 + 10/218 = 1,00$ ). En consecuencia, es posible afirmar que a mayor especificidad de un test diagnóstico, menor la probabilidad de que éste arroje un resultado falsamente positivo. La Tabla 2 resume la forma en que se realizan estos cálculos.

### Valores predictores positivo y negativo

Los conceptos de sensibilidad y especificidad permiten dimensionar la validez de un test diagnóstico; sin embargo, carecen de utilidad práctica en clínica, por cuanto proporcionan información sobre la probabilidad de que un test arroje un resultado concreto (positivo o negativo)

en función de la verdadera condición del individuo con respecto a la enfermedad (enfermo vs sano)<sup>1</sup>. Frente a un paciente, el clínico ignora *a priori* cuál es el verdadero diagnóstico del paciente (precisamente por eso realiza el test diagnóstico), planteándose una interrogante en sentido opuesto al de la sensibilidad y especificidad. Esta interrogante adopta dos formas: 1) ¿cuál es la probabilidad de que el individuo consultante esté enfermo, dado que el resultado del test diagnóstico fue positivo? y 2) ¿cuál es la probabilidad de que el individuo consultante esté sano, dado que el resultado del test diagnóstico fue negativo?. Estas dos probabilidades son conocidas como valores predictores positivo (VPP) y negativo (VPN), respectivamente, y son calculadas en base a los datos de la misma tabla de contingencia a partir de la cual se calculó la sensibilidad y especificidad<sup>8-10</sup> (Tabla 1). En el ejemplo, el VPP se calcula como  $51/61 = 0,84$  (84%) y el VPN se calcula como  $208/216 = 0,96$  (96%). En otras palabras, la probabilidad de que un individuo tenga faringoamigdalitis estreptocócica dado que el resultado del GASD fue positivo es 84%, mientras que la probabilidad de que un individuo no tenga faringoamigdalitis estreptocócica dado que el resultado del GASD fue negativo es 96%. La Tabla 2 resume la forma en que se realizan estos cálculos.

### Fortalezas y limitaciones de la sensibilidad, especificidad y valores predictores

La sensibilidad y especificidad de un test diagnóstico tienen la fortaleza de ser propiedades inherentes al test, independientemente de la prevalencia de la enfermedad en la población a la cual se aplica. Sin embargo, no proporcionan información relevante al momento de tomar una decisión clínica, debido a que responden a interrogantes formuladas en sentido inverso al proceso diagnóstico. Por el contrario, el sentido de las preguntas respondidas por los valores predictores les otorgan gran utilidad al momento de tomar decisiones clínicas; sin embargo, tienen el defecto de ser dependientes de la prevalencia de la enfermedad en la población de personas de similares características al individuo consultante. Al respecto, mientras mayor sea la prevalencia de la enfermedad, mayor es el VPP y menor es el VPN del test diagnóstico; por el contrario, mientras menor sea la prevalencia de la enfermedad, mayor es el VPN y menor es el VPP del mismo. En el ejemplo, la prevalencia de la enfermedad (definida como el total de individuos con faringoamigdalitis estreptocócica según el estándar de oro/total de individuos evaluados) es  $59/277 = 0,21$  (21%). La Tabla 3 ilustra los valores que adoptaría la tabla de contingencia en un escenario hipotético de mayor prevalencia de faringoamigdalitis estreptocócica (e.g.  $100/200 = 0,50 = 50\%$ ). En tal escenario, la sensibilidad



**Tabla 3. Resultados del cultivo faríngeo y test directo de *Streptococcus pyogenes* (GASD) en un escenario ficticio de mayor prevalencia de la enfermedad en la población de personas similares al individuo consultante**

GASD	Cultivo faríngeo		Total
	Positivo	Negativo	
Positivo	86	5	91
Negativo	14	95	109
Total	100	100	200

Prevalencia de la enfermedad =  $(100/200) = 50\%$ ; Sensibilidad =  $(86/100) = 86\%$ ; Especificidad =  $(95/100) = 95\%$ ; VPP =  $(86/91) = 95\%$ ; VPN =  $(95/109) = 87\%$ .

$(86/100 = 0,86 = 86\%)$  y especificidad  $(95/100 = 0,95 = 95\%)$  no cambian, por cuanto son valores inherentes al test diagnóstico. Sin embargo, el VPP  $(86/91 = 0,95 = 95\%)$  aumenta en comparación al original (84%); por el contrario, el VPN  $(95/109 = 0,87 = 87\%)$  disminuye en comparación al original (96%). En otras palabras, frente a un GASD positivo, la probabilidad de que un individuo consultante efectivamente tenga faringoamigdalitis estreptocócica es mayor en aquellas poblaciones en las cuales la prevalencia de esta enfermedad es más alta (VPP 95% vs VPP 84%); por el contrario, frente a un GASD negativo, la probabilidad de que un individuo consultante efectivamente no tenga faringoamigdalitis estreptocócica es mayor en aquellas poblaciones en las cuales la prevalencia de esta enfermedad es menor (VPN 96% vs VPN 87%). Este hecho genera un importante dilema clínico, el cual es posible resumir en la siguiente pregunta: ¿existe alguna herramienta útil para la toma de decisiones clínicas que contemple únicamente las fortalezas descritas tanto para la sensibilidad y especificidad como para los valores predictores, excluyendo sus limitaciones?. La respuesta a esta interrogante la aportan los *likelihood ratios* (i.e. razones de verosimilitud), cuyo cálculo y aplicación serán abordados en la segunda parte del presente artículo.

## Conclusión

El presente artículo ilustró la forma en que se analizan las propiedades de un test diagnóstico, utilizando herramientas teóricas aportadas por la Epidemiología Clínica. Dos estimadores que caracterizan estas propiedades, sensibilidad y especificidad, son descriptores del test diagnóstico *per se*, presentando escasa utilidad en la resolución de un dilema clínico. Para el médico clínico resulta más relevante conocer la magnitud en la cual un test diagnóstico modifica sus probabilidades pre test y contribuye a la toma de decisiones. Al respecto, los valores predictores positivo y negativo responden interrogantes en sentido concordante al proceso diagnóstico; sin embargo, tienen la limitante de cambiar según la prevalencia de la enfermedad en la población de personas de similares características al individuo al cual se aplica el test. Para dimensionar la utilidad clínica de un test diagnóstico resulta fundamental el cálculo y aplicación de estimadores conocidos como *likelihood ratios* (i.e. razones de verosimilitud).

## Resumen

Los tests diagnósticos constituyen herramientas complementarias ampliamente utilizadas en la práctica clínica. La Epidemiología Clínica ha sistematizado el análisis cualitativo de sus propiedades (validez, reproducibilidad y seguridad) mediante la introducción de una serie de conceptos tales como sensibilidad, especificidad y valores predictores. En base a los datos aportados por una investigación clínica, el presente artículo ilustra el sustrato teórico subyacente a los conceptos anteriormente mencionados, detalla su forma de cálculo y analiza sus principales fortalezas y limitaciones. En un segundo artículo (Parte 2: Aplicación clínica y utilidad de un test diagnóstico) se describe la forma en que la información aportada por una prueba debe ser aplicada correctamente en la resolución de un dilema clínico.

## Referencias

- 1.- Pita S, Pértegas S. Pruebas diagnósticas. Cad Aten Primaria 2003; 10: 120-4.
- 2.- Capurro D, Rada G. El proceso diagnóstico. Rev Med Chile 2007; 135: 534-8.
- 3.- Last J. A Dictionary of Epidemiology, New York: Oxford University Press, 2001.
- 4.- Valenzuela L, Cifuentes L. Validez de estudios de tests diagnósticos. Rev Med Chile 2008; 136: 401-4.
- 5.- Altman D G, Bland J M. Statistics Notes: Diagnostic tests 3: receiver operating characteristic plots. Br Med J 1994; 309: 188.
- 6.- Steed L L, Korgenski E K, Daly J A. Rapid detection of *Streptococcus pyogenes* in pediatric patient specimens by DNA probe. J Clin Microbiol 1993; 31: 2996-3000.
- 7.- Altman D G, Bland J M. Statistics Notes: Diagnostic tests 1: sensitivity and specificity. Br Med J 1994; 308: 1552.
- 8.- Greenhalgh T. How to read a paper: papers that report diagnostic or screening tests. Br Med J 1997; 315: 540-3.
- 9.- Salech F, Mery V, Larrondo F, Rada G. Estudios que evalúan un test diagnóstico: interpretando sus resultados. Rev Med Chile 2008; 136: 1203-8.
- 10.- Altman D G, Bland J M. Statistics Notes: Diagnostic tests 2: predictive values. Br Med J 1994; 309: 102.