

Metodología de investigación y lectura crítica de estudios

Generalidades de las pruebas diagnósticas, y su utilidad en la toma de decisiones médicas

María Camila Medina¹

Resumen

Introducción: La utilidad de las pruebas diagnósticas se basa en la capacidad de separar adecuadamente las personas sanas de las enfermas, y de no producir falsos positivos o negativos. Las propiedades de las pruebas diagnósticas se describen tradicionalmente como sensibilidad, especificidad y valores predictivos. Una forma adicional de considerar la probabilidad de la enfermedad es mediante el cálculo de la razón de probabilidades o *likelihood ratio*; este permite conocer cuánto más probable es un resultado (positivo o negativo) según la presencia o ausencia de la enfermedad. **Objetivo:** Describir en términos generales las propiedades de las pruebas diagnósticas. Se explican los conceptos de sensibilidad, especificidad, valores predictivos y *likelihood ratio*, así como su utilidad y aplicabilidad en la práctica médica. **Metodología y desarrollo:** Se realizó una revisión de artículos y textos sobre las propiedades de las pruebas diagnósticas, para definir y analizar sus características operativas; posteriormente, se exponen ejemplos para facilitar su entendimiento e interpretación.

Palabras clave: sensibilidad, especificidad, valor predictivo positivo, valor predictivo negativo, *likelihood ratio*, pruebas diagnósticas.

Title: Overview of Diagnostic Tests and Their Usefulness in Medical Decisions Making

Abstract

Introduction: The usefulness of diagnostic tests consists of the ability to detect and separate a person with a disease from a healthy person, and not to produce false positives or negatives. The properties of diagnostic tests have traditionally been described using sensitivity, specificity, and positive and negative predictive values. An additional way to consider the probability of a disease is by calculating the likelihood ratio which helps to estimate how many times more (or less) likely patients with a disease are to have a particular result than patients without the disease. **Objective:** To describe the main properties of diagnostic tests. Important concepts, such as sensitivity, specificity, predictive value and likelihood ratio are explained, as well as their usefulness and applicability in clinical practice. **Method:** Articles and texts have been reviewed to define and analyze the operative characteristics of diagnostic tests. Similarly, examples are given to facilitate understanding and interpretation.

Key words: Sensitivity, specificity, positive predictive value, negative predictive value, likelihood ratio, diagnostic tests.

¹ Médica interna. Asistente de investigación, Departamento de Epidemiología Clínica y Bioestadística, Facultad de Medicina, Pontificia Universidad Javeriana. Bogotá, Colombia.

En caso de encontrarse en un umbral intermedio, en el cual ni se rechaza ni se apoya la hipótesis diagnóstica, la probabilidad de que el paciente presente un diagnóstico planteado se encuentra en una zona de incertidumbre, y este es el caso donde se justifica la realización de pruebas adicionales.

Ejemplo 1

Hombre de 30 años de edad, quien es llevado por sus familiares al servicio de urgencias por presentar temblor, taquicardia, diaforesis y cifras tensionales elevadas. Un familiar refiere que el paciente suspendió las bebidas alcohólicas hace 2 semanas. Con esta información el médico considera que el paciente presenta un síndrome de abstinencia, por lo cual inicia tratamiento.

Ejemplo 2

Hombre de 65 años de edad, quien sale de una cirugía de reemplazo de cadera, y posteriormente presenta alteración de la conciencia, desorientación, alucinaciones e hiperactividad motora. Dada la probabilidad alta de *delirium*, se debe establecer cuál es la condición médica de base, por lo que se justifica la realización de pruebas diagnósticas.

Ejemplo 3

Estudiante de medicina, quien se encuentra realizando sus prácticas en Medicina Legal. Durante la necropsia sufre un accidente: se cortó con un bisturí con el que

estaba realizando el procedimiento; no se conoce si tiene VIH. Dados la gravedad de esta condición y los costos que genera, es necesario que se realice una prueba diagnóstica.

Ejemplo 4

Paciente de 25 años de edad, quien súbitamente presenta dificultad para respirar, sudoración, palpitaciones y dolor torácico; la probabilidad de que sufra un infarto de miocardio es nula. El médico conjetura una crisis de pánico, por lo cual no considera la toma de exámenes.

Estándar de referencia, o *gold standard*

La primera condición para evaluar una nueva prueba es que exista un método confiable con el fin de hacer el diagnóstico; es decir, cuál es, hasta el momento del estudio, la mejor alternativa diagnóstica existente para estudiar una determinada enfermedad o un evento de interés. Habitualmente este es un examen demasiado difícil o costoso para ser utilizado de manera rutinaria. La prueba de oro puede ser desde una imagen radiológica o un electroencefalograma hasta una necropsia.

El uso de una prueba de oro con el fin de identificar definitivamente a quienes tienen la enfermedad es un requisito para evaluar la utilidad diagnóstica de cualquier prueba nueva o no evaluada. De este modo una prueba nueva se compara con una prueba an-

tigua y más aceptada, buscando determinar si la nueva ofrece el mismo rendimiento que la de referencia.

Medición de la sensibilidad y la especificidad

El método tradicional para la enseñanza de estos conceptos se basa en una tabla de 2x2, conocida también como *tabla de contingencia*. En esta tabla si se compara una prueba diagnóstica con el estándar de referencia se pueden obtener cuatro combinaciones si los resultados de la prueba se expresan de forma dicótoma (positivo o negativo, ausente o presente).

En la tabla de contingencia se enfrenta el resultado de la prueba diagnóstica (en filas) con el estado real de los pacientes (en columnas), o, en su defecto, el resultado de la prueba de referencia, o *gold standard*, que se vaya a utilizar. El resultado de la prueba puede ser correcto (verdadero positivo y verdadero negativo) o incorrecto (falso positivo y falso negativo). El análisis

de su validez puede obtenerse calculando los valores de sensibilidad y de especificidad (tabla 1).

Sensibilidad

La sensibilidad puede definirse como la capacidad de la prueba para clasificar correctamente al enfermo como enfermo, o como la probabilidad de tener un resultado positivo si se tiene la enfermedad.

Para calcular la sensibilidad se debe dividir el número de enfermos con prueba positiva por la sumatoria de los enfermos con prueba positiva y los enfermos con prueba negativa:

$$a / (a + c); \text{ o } VP / VP + FN$$

Se puede deducir que una prueba diagnóstica de alta sensibilidad tiene pocos falsos negativos, y, al contrario, una prueba con baja sensibilidad (poca capacidad para detectar al enfermo como enfermo) tendrá una alta tasa de resultados falsos negativos. La utilización de una prueba muy sensible asegura

Tabla 1. Tabla de 2x2, en la que se explica la generación de los conceptos VP, VN, FP, FN

Estado respecto a la enfermedad según el estándar de oro		
Resultado de la prueba de estudio	Enfermo	Sano
Positivo	Verdadero positivo (VP)	Falso positivo (FP)
Negativo	Falso negativo (FN)	Verdadero negativo (VN)

VP: el paciente tiene la enfermedad y la prueba es positiva. FP: el paciente no tiene la enfermedad y la prueba es positiva. VN: el paciente no tiene la enfermedad y la prueba es negativa. FN: el paciente tiene la enfermedad y la prueba es negativa.

que un resultado negativo probablemente será un resultado verdadero, pues tiene una gran capacidad para detectar a los enfermos como enfermos.

Especificidad

La especificidad es la capacidad de la prueba para clasificar adecuadamente a los sanos como sanos; es el porcentaje de personas que no tienen la condición de estudio y dan resultados “negativos” o “normales”.

Para calcular la especificidad se debe dividir el número de sujetos “no enfermos” con prueba positiva por la sumatoria de los sujetos “no enfermos” con prueba positiva y los sujetos “no enfermos” con prueba negativa:

$$b/(b + d); \text{ o } FP/FP + VN$$

El aspecto más importante es que el examen puede clasificar co-

rectamente al paciente sano como sano; es decir, los verdaderos negativos. Un examen con una alta especificidad es muy útil cuando el resultado es positivo, pues la tasa de falsos positivos es muy baja (tabla 2).

Ejemplo

En 1997 Whooley *et al.* evaluaron la sensibilidad y la especificidad de 2 preguntas para el tamizaje de depresión comparado con 6 instrumentos ya validados (SDDS-PC, CES-D, CES-D-SHORT, MOS, BDI, BDI-SHORT). Las preguntas evaluadas fueron: ¿Durante el último mes se ha sentido triste, deprimido o sin esperanza? ¿Durante el último mes ha sentido poco interés o placer en la realización de sus actividades?

Suponiendo que los datos del estudio incluyeron a 500 pacientes con sospecha de depresión, los resultados se contrastaron con los del SDDS-PC. Los datos obtenidos se muestran en la siguiente tabla.

Tabla 2. Tabla de 2x2 en la que se explica la generación de las celdas con las que se realizan los cálculos para las medidas de S, E y VP

Estado respecto a la enfermedad según el estándar de referencia		
Resultado de la prueba de estudio	Presente	Ausente
Positivo	a (enfermos con prueba +)	b (no enfermos con prueba +)
Negativo	c (enfermos con prueba -)	d (no enfermos con prueba -)

Sensibilidad: proporción de personas con la enfermedad, que tendrán un resultado positivo $\{a/(a+c)\}$.

Especificidad: proporción de personas sin la enfermedad que tendrán un resultado negativo $\{b/(b+d)\}$.

De un total de 500 pacientes se identificó a 100 de ellos con la condición de interés (definido por el SDDS-PC), y se encontró que 86 fueron correctamente identificados por el tamizaje de las 2 preguntas (verdaderos positivos); la sensibilidad de la prueba fue del 86%. Los resultados negativos (falsos negativos) incluyeron a 14 pacientes que fueron incorrectamente diagnosticados como sanos.

Por otro lado, de un total de 400 pacientes, 380 fueron correctamente diagnosticados como sanos por el tamizaje de las 2 preguntas (verdaderos negativos), mientras que 20 fueron incorrectamente diagnosticados como enfermos (falsos positivos); la especificidad de la prueba fue, pues, del 95% (tabla 3).

¿Qué pruebas utilizar?

Según todo lo anterior, si se pretende “detectar” una enfermedad que sea muy riesgosa o hay un precio importante que pagar si no se hace el diagnóstico, se utilizarán

pruebas de alta sensibilidad: estas son más útiles cuando los resultados son negativos, por lo cual si los falsos negativos tienen alto costo se prefiere alta sensibilidad.

En el caso de estudios de tamizaje, cuando la probabilidad de la condición sea baja y el propósito sea descubrirlo se debe utilizar una prueba de alta sensibilidad.

Si un falso positivo tiene alto riesgo o costo se prefieren pruebas con una alta especificidad, pues son más útiles cuando los resultados son positivos; por tener una baja tasa de falsos positivos, se usan para la confirmación del diagnóstico.

Valores predictivos

En la clínica el médico se enfrenta a situaciones donde debe considerar el resultado del examen, sabiendo que la prueba la mayoría de las veces no es perfecta.

Ante un resultado positivo o negativo en la prueba se preguntará cuál es la probabilidad de que el paciente esté enfermo, dado que el

Tabla 3. Evaluación de la sensibilidad y especificidad Symptom Driven Diagnostic System for Primary Care (SDDS-PC)

Dos preguntas	Positivo	Negativo	Total
Positivo	86	20	106
Negativo	14	380	394
Total	100	400	500

Prevalencia de la enfermedad=(100/500)=20%; Sensibilidad=(86/100)= 86%; Especificidad=(380/400)=95%; VPP= (86/106)= 81%; VPB=(380/394)= 96%.

resultado del test es positivo; o cuál es la probabilidad de que el paciente esté sano, dado un resultado negativo. Estas probabilidades se determinan mediante los valores predictivos.

Valor predictivo positivo (VPP)

Es la probabilidad de presentar la enfermedad si se obtiene un resultado positivo en el test.

Para calcular el VPP de una prueba diagnóstica se ha de dividir el número enfermos con prueba positiva por la sumatoria de los enfermos con prueba positiva y los sujetos “no enfermos” con prueba positiva.

$$a / (a + b); \text{ o } VP / VP + FP$$

Valor predictivo negativo (VPN)

Es la probabilidad de que un paciente con un resultado negativo en la prueba esté realmente sano.

Para calcular, entonces, el VPN se debe dividir el número de enfermos

con prueba negativa por la sumatoria de los enfermos con prueba negativa y los sujetos “no enfermos” con prueba negativa.

$$d / (c + d); \text{ o } VN / FN + VN$$

Usando los mismos datos de la tabla anterior, se calcula el VPP como $86 / (86 + 20) = 81\%$. La interpretación de este resultado es: la probabilidad de que un paciente tenga depresión, dado que el resultado de las dos preguntas para tamizaje fue positivo, es del 81%.

El VPN se calcula como $380 / (380 + 14) = 96\%$. La interpretación de este resultado es: la probabilidad de que un individuo no tenga depresión, dado que el resultado de las dos preguntas fue negativo, es del 96% (figura 2).

La influencia de la prevalencia

La sensibilidad y la especificidad de un test diagnóstico tienen la fortaleza de ser propiedades inherentes

Figura 2. Evaluación de los valores predictivos

	Presente	Ausente	
Resultado de la prueba (positivo)	86 A	20 B	$a / (a + b)$ $86 / (86 + 20) = 81\%$
Resultado de la prueba (Negativo)	14 c	380 D	
	100	400	$d / (c + d)$ $(380 + 14) = 96\%$

al test, independientemente de la prevalencia de la enfermedad; sin embargo, no proporcionan información relevante en el momento de tomar decisiones clínicas.

A diferencia de la sensibilidad y la especificidad, los valores predictivos permiten tomar decisiones adecuadas con respecto a la interpretación de los resultados, pero presentan limitaciones que dependen de la prevalencia o de la probabilidad preexamen, lo cual no las hace extrapolables a todas las situaciones. Por eso, para cada valor de prevalencia o probabilidad preexamen se deben calcular los valores predictivos.

Mientras mayor sea la prevalencia de la enfermedad, mayor es el VPP y menor el VPN; por el contrario, mientras menor sea la prevalencia de la enfermedad, mayor es el VPN, y menor el VPP (tabla 4).

Razón de probabilidades (Likelihood ratio, o RP)

El verdadero valor de una prueba diagnóstica en la práctica clínica no

solo depende de su sensibilidad, su especificidad y sus valores predictivos, sino de la información que proporcionan los resultados de dicha prueba, para lograr un cambio sustancial en la probabilidad pretest y postest con respecto a una determinada enfermedad.

Para ello resulta necesario determinar otras propiedades que sean, a la vez, clínicamente útiles, una forma adicional de considerar la probabilidad de presencia de la enfermedad es mediante el cálculo del *likelihood ratio* o razón de probabilidades. Estos miden cuánto más probable es un resultado (positivo o negativo) según la presencia o ausencia de enfermedad.

La ventaja de la razón de probabilidades radica en que, a diferencia de los valores predictivos, no varía con los cambios en la prevalencia de una enfermedad. La otra ventaja importante es que las RP suelen venir calculadas para múltiples puntos de corte, y así facilitan el cálculo de probabilidad postexamen, de acuerdo con el valor del resultado obtenido.

Tabla 4. Variación de los valores predictivos con respecto a la prevalencia

Prevalencia	Sensibilidad	Especificidad	VPP	VPN
10%	86%	95%	64%	98%
20%	86%	95%	81%	96%
50%	86%	95%	94%	87%

La razón de probabilidades para un resultado positivo

Razón entre la probabilidad de que un examen tenga resultado positivo en los pacientes que tienen la enfermedad *vs.* los que no la tienen.

$$RP (+) = [a/(a+c)] / [b/(b+d)]$$

La razón de probabilidades para un resultado negativo

Razón entre la probabilidad de que un examen tenga resultado negativo en los pacientes que tienen la enfermedad *vs.* los que no la tienen.

$$RP (-) = [c/(a+c)] / [d/(b+d)]$$

Usando los mismos datos de los ejemplos anteriores, se calcula la razón de probabilidades. Una razón de probabilidades para un resultado positivo de 17 se interpreta como que

es 17 veces más posible que resulte positiva una prueba en presencia de su enfermedad que en su ausencia.

También se puede afirmar que los resultados negativos ocurren 0,14 veces en personas enfermas con respecto a cada persona sana (figura 3).

Guía para la interpretación de razones de probabilidad

- Razones de probabilidad mayores que 10 y menores que 0,1 generan amplios y frecuentes cambios concluyentes de la probabilidad preprueba a la posprueba.
- Razones de verosimilitud de 5-10 y de 0,1-0,2 generan cambios moderados de la probabilidad preprueba a la posprueba.
- Razones de verosimilitud de 2-5 y de 0,5-0,2 generan pequeños (pero algunas veces importantes) cambios en probabilidad.

Figura 3. Evaluación de la razón de probabilidades

RP	Depresión	Ausente	
Prueba positiva	86 A	20 B	106
Prueba negativa	14 c	380 D	394
	100	400	
$RP (+) = [86/(86+14)] / [20/(20+380)] = 17$			
$RP (-) = [14/(14+86)] / [380/(380+20)] = 0,14$			

- Razones de verosimilitud de 1-2 y de 0,5-1 alteran la probabilidad en un pequeño (y rara vez importante) grado.

Conclusiones

Gracias a la evolución de la medicina y a los avances tecnológicos, la disponibilidad de pruebas diagnósticas ha sido tan amplia, y sus costos

tan altos, que se debe determinar la validez de dichos estudios, saber interpretar correctamente los resultados y, finalmente, aplicar los resultados a los pacientes.

El presente artículo ilustró la forma como se analizan las propiedades de un test diagnóstico. La sensibilidad, la especificidad y los valores predictivos son propiedades importantes de las pruebas diag-

Tabla 5. Resumen de las propiedades de las pruebas diagnósticas

Propiedades del test	Pregunta por responder	Fórmula	Comentarios
Sensibilidad	Qué tan bueno es el test detectando a posibles enfermos	$a/(a + c)$	Si es un test muy sensible, un resultado negativo descarta la enfermedad
Especificidad	¿Qué tan bueno es el test detectando a pacientes sanos?	$b/(b + d)$	Si un test es muy específico, un resultado positivo confirma la enfermedad
VPP	Si un paciente tiene una prueba positiva, ¿qué tan probable es que tenga la condición?	$a/(a + b)$	Los valores predictivos varían de acuerdo con la prevalencia
VPN	Si un paciente tiene una prueba negativa, ¿qué tan probable es que no tenga la condición?	$d/(c + d)$	
RP	¿Qué tanto más probable es encontrar determinado valor del test en alguien enfermo comparado con alguien sano?	—	Pueden calcularse tantos RP como valores posibles tiene el test. En el caso de un test dicotómico tendrá un valor positivo y negativo
RP +	¿Qué tanto más probable es encontrar el test positivo en alguien enfermo que en alguien sano?	$[a/(a+c)] / [b/(b+d)]$	También se puede calcular: sensibilidad/1-especificidad
RP -	¿Qué tanto más probable es encontrar el test negativo en alguien enfermo que en alguien sano?	$[c/(a+c)] / [d/(b+d)]$	También se puede calcular: 1-sensibilidad/especificidad

nósticas, aunque cada una de ellas tiene sus limitaciones.

La razón de probabilidades o el *likelihood ratio* son útiles para cambiar la probabilidad pretest a postest, por lo cual son de gran utilidad en la práctica clínica, y su cálculo y su aplicación en la medicina resultan fundamentales (tabla 5).

Agradecimientos

Al Departamento de Epidemiología Clínica y Bioestadística, y a sus profesores Carlos Gómez-Restrepo, Álvaro Ruiz Morales y Juan Gabriel Ruiz Peláez.

Bibliografía sugerida

- Callahan CM, Unverzagt FW, Hui SL, et al. Six item screener to identify impairment among potential subjects for clinical research. *Med Care*. 2002;40:771-81.
- Cerda J, Cifuentes L. Uso de test diagnósticos en la práctica clínica (Parte 1). Análisis de las propiedades de un test diagnóstico. *Rev Chil Infect*. 2010;27:205-8.
- Riegelman RK, Hirsch RP. Studying a study and testing a test: how to read the medical literature. Boston: Little Brown; 1989.
- Ruiz A, Cuervo L. Pruebas diagnósticas, generalidades de su interpretación. *Rev Col Neumología*. 2008;20:98-105.
- Ruiz A. Uso de pruebas diagnósticas en medicina clínica. En: Ruiz A, Gómez C, Londoño D. editores. *Epidemiología clínica: investigación clínica aplicada*. Bogotá: CEJA; 2001. p. 263-80.
- Whooley M. Finding instruments for depression, two questions are as good as many. *J Gen Intern Med*. 1997;12:439-45.

Conflictos de interés: La autora manifiesta que no tiene conflictos de interés en este artículo.

*Recibido para evaluación: 28 de septiembre de 2011
Aceptado para publicación: 10 de noviembre de 2011*

*Correspondencia
María Camila Medina
Departamento de Epidemiología Clínica y Bioestadística
Hospital Universitario de San Ignacio.
Carrera 7ª No. 40-62, piso 2
Bogotá, Colombia
medina.camila01@gmail.com*