



Fisterra

Pruebas diagnósticas: Sensibilidad y especificidad

📅 Fecha de revisión: **07/12/2010**

Índice de contenidos

- [Introducción](#)
- [La validez de una prueba diagnóstica: Sensibilidad y especificidad](#)
- [La seguridad de una prueba diagnóstica. Valores predictivos.](#)
- [La influencia de la prevalencia.](#)
- [Razones de probabilidad](#)
- [Anexo](#)
- [Bibliografía](#)
- [Autores](#)
- [Cálculos Excel](#)

Introducción

La medicina es una ciencia de probabilidades y un arte de manejar la incertidumbre. Dicha incertidumbre se extiende no sólo a las actividades preventivas, terapéuticas y pronósticas sino también a las diagnósticas. En las fases del proceso diagnóstico intervienen la historia clínica, la exploración física y la realización de pruebas complementarias. Cuando existen varias hipótesis diagnósticas, se realizará el diagnóstico diferencial y las pruebas complementarias tratarán de aclarar las dudas existentes. Si solamente hay una sospecha diagnóstica, las pruebas complementarias tratarán de confirmarla. La realización simultánea de varias pruebas complementarias se denomina pruebas complementarias en paralelo y la realización de pruebas complementarias según los resultados de otras previas, se denomina pruebas complementarias en serie. Al realizar pruebas en paralelo aumenta la probabilidad de diagnosticar a un enfermo, pero también aumenta la probabilidad de considerar como enfermo a un sano. El riesgo de la realización de pruebas en serie es no diagnosticar a algunos enfermos. En cambio, pocos sanos serán considerados como enfermos.

Es evidente que una buena prueba diagnóstica es la que ofrece resultados positivos en enfermos y negativos en sanos. Por lo tanto, las condiciones que deben ser exigidas a un test son:

- **Validez:** Es el grado en que un test mide lo que se supone que debe medir. ¿Con que frecuencia el resultado del test es confirmado por procedimientos diagnósticos más complejos y rigurosos? La sensibilidad y la especificidad de un test son medidas de su validez.
- **Reproductividad:** es la capacidad del test para ofrecer los mismos resultados cuando se repite su aplicación en circunstancias similares. La variabilidad biológica del hecho observado, la introducida por el propio observador y la derivada del propio test, determinan su reproductividad.
- **Seguridad:** La seguridad viene determinada por el valor predictivo de un resultado positivo o negativo. ¿Con que seguridad un test predecirá la presencia o ausencia de enfermedad? Ante un resultado positivo de un test ¿qué probabilidad existe de que este resultado indique presencia de la enfermedad? Veremos posteriormente que esta probabilidad está muy influenciada por la prevalencia de la patología.

A su vez, es conveniente que el test sea sencillo de aplicar, aceptado por los pacientes o la población general, que tenga los mínimos efectos adversos y que económicamente sea soportable.

En este trabajo se revisarán fundamentalmente los conceptos que determinan la validez de un test (sensibilidad y especificidad) y su seguridad (valores predictivos positivos y negativos).

La validez de una prueba diagnóstica: Sensibilidad y especificidad

El caso más sencillo que se nos puede plantear es el de una prueba dicotómica, que clasifica a cada paciente como sano o enfermo en función de que el resultado de la prueba sea positivo o negativo. En casos como éste, generalmente un resultado positivo se asocia con la presencia de enfermedad y un resultado negativo con la ausencia de la misma. Cuando se estudia una muestra de pacientes, los datos obtenidos permiten clasificar a los sujetos en cuatro grupos según una tabla 2x2 como la que se muestra en la Tabla 1. En ella, 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 vayamos 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 especificidad:

Sensibilidad

Es la probabilidad de clasificar correctamente a un individuo enfermo, es decir, la probabilidad de que para un sujeto enfermo se obtenga en la prueba un resultado positivo. La sensibilidad es, por lo tanto, la capacidad del test para detectar la enfermedad.

Cuando los datos obtenidos a partir de una muestra de pacientes se clasifican en una tabla como la que se muestra en la Tabla 1, es fácil estimarla partir de ella la sensibilidad como la proporción de pacientes enfermos que obtuvieron un resultado positivo en la prueba diagnóstica. Es decir:



De ahí que también la sensibilidad se conozca como “fracción de verdaderos positivos (FVP)”.

Especificidad

Es la probabilidad de clasificar correctamente a un individuo sano, es decir, la probabilidad de que para un sujeto sano se obtenga un resultado negativo. En otras palabras, se puede definir la especificidad como la capacidad para detectar a los sanos. A partir de una tabla como la Tabla 1, la especificidad se estimaría como:



De ahí que también sea denominada “fracción de verdaderos negativos (FVN)”.

Ejemplo:

Como ejemplo de lo visto hasta ahora, consideremos los datos de un estudio en el que se incluyó a 2.641 pacientes con sospecha de cáncer prostático que acudieron a una consulta de Urología durante un periodo de tiempo determinado. Durante su exploración, se recogió el resultado del tacto rectal realizado a cada uno de estos pacientes, según fuese éste normal o anormal, y se contrastó con el posterior diagnóstico obtenido de la biopsia prostática. Los datos del estudio y los resultados obtenidos se muestran en la Tabla 2. Se encontraron en total 1.121 casos de cáncer, lo cual representa un 42,45% del total de sujetos estudiados. La sensibilidad del tacto rectal para detectar cáncer fue de 56,56% (634/1121) y la especificidad de 82,3% (1251/1520). Así, el tacto fue anormal en un 56,56% de los casos de cáncer prostático y normal en un 82,3% de los casos que presentaron finalmente otras patologías. Esto significa que un $100 - 56,56 = 43,44\%$ de los pacientes que efectivamente tenían cáncer presentaban tactos normales. Claramente ello indica la necesidad de utilizar otros marcadores más sensibles, como el PSA o sus derivados, para poder establecer el diagnóstico de forma más precisa.

Resulta obvio que lo ideal sería trabajar con pruebas diagnósticas de alta sensibilidad y especificidad, pero esto no siempre es posible. En general, las pruebas de screening deben ser de alta sensibilidad para poder captar a todos los enfermos. Una prueba muy sensible será especialmente adecuada en aquellos casos en los que el no diagnosticar la enfermedad puede resultar fatal para los enfermos,

como ocurre con enfermedades peligrosas pero tratables, como los linfomas o la tuberculosis, o en enfermedades en las que un falso positivo no produzca serios trastornos psicológicos o económicos para el paciente (por ejemplo, la realización de mamografía en el cáncer de mama).

Por otra parte, la especificidad se refiere, como se señaló previamente, a la probabilidad de que un sujeto sano sea clasificado adecuadamente. En general, las pruebas confirmatorias del diagnóstico deben ser de alta especificidad, para evitar falsos positivos. Los tests de alta especificidad son necesarios en enfermedades graves pero sin tratamiento disponible que las haga curables, cuando exista gran interés por conocer la ausencia de enfermedad o cuando diagnosticar a un paciente de un mal que realmente no padece pueda acarrear graves consecuencias, ya sean físicas, psicológicas o económicas (por ejemplo, en el caso del SIDA).

La seguridad de una prueba diagnóstica. Valores predictivos.

Los conceptos de sensibilidad y especificidad permiten, por lo tanto, valorar la validez de una prueba diagnóstica. Sin embargo, carecen de utilidad en la práctica clínica. Tanto la sensibilidad como la especificidad proporcionan información acerca de la probabilidad de obtener un resultado concreto (positivo o negativo) en función de la verdadera condición del enfermo con respecto a la enfermedad. Sin embargo, cuando a un paciente se le realiza alguna prueba, el médico carece de información a priori acerca de su verdadero diagnóstico, y más bien la pregunta se plantea en sentido contrario: ante un resultado positivo (negativo) en la prueba, ¿cuál es la probabilidad de que el paciente esté realmente enfermo (sano)? Así pues, resulta obvio que hasta el momento sólo hemos abordado el problema en una dirección. Por medio de los valores predictivos completaremos esta información:

Valor predictivo positivo:

Es la probabilidad de padecer la enfermedad si se obtiene un resultado positivo en el www. El valor predictivo positivo puede estimarse, por tanto, a partir de la proporción de pacientes con un resultado positivo en la prueba que finalmente resultaron estar enfermos:



Valor predictivo negativo:

Es la probabilidad de que un sujeto con un resultado negativo en la prueba esté realmente sano. Se estima dividiendo el número de verdaderos negativos entre el total de pacientes con un resultado negativo en la prueba:



Retomando el ejemplo anterior sobre cáncer prostático, el valor predictivo positivo es en este caso del 70,21% (634/903) y el valor predictivo negativo del 71,98% (1251/1738). Ello significa que en un 70,21% de los pacientes con un tacto anormal finalmente se confirmó la presencia de cáncer, mientras que de los que no se detectaron anomalías en el tacto un 71,98% estaban efectivamente sanos.

La influencia de la prevalencia.

Hemos visto cómo los valores de sensibilidad y especificidad, a pesar de definir completamente la validez de la prueba diagnóstica, presentan la desventaja de que no proporcionan información relevante a la hora de tomar una decisión clínica ante un determinado resultado de la prueba. Sin embargo, tienen la ventaja adicional de que son propiedades intrínsecas a la prueba diagnóstica, y definen su validez independientemente de cuál sea la prevalencia de la enfermedad en la población a la cual se aplica.

Por el contrario, el concepto de valores predictivos, a pesar de ser de enorme utilidad a la hora de tomar decisiones clínicas y transmitir a los pacientes información sobre su diagnóstico, presenta la limitación de que dependen en gran medida de lo frecuente que sea la enfermedad a diagnosticar en la

población objeto de estudio. Cuando la prevalencia de la enfermedad es baja, un resultado negativo permitirá descartar la enfermedad con mayor seguridad, siendo así el valor predictivo negativo mayor. Por el contrario, un resultado positivo no permitirá confirmar el diagnóstico, resultando en un bajo valor predictivo positivo.

Ilustraremos lo anterior con un sencillo ejemplo. Para el diagnóstico del VIH se emplean tests que han confirmado tener una alta validez, con valores aproximados de sensibilidad y especificidad de un 99,5%. Supongamos que se aplicase esta prueba a la totalidad de la población gallega, que se cifra en 2.800.000 habitantes. Si asumimos que en Galicia existen 6.000 pacientes VIH positivos (lo cual implicaría una prevalencia de $6000/2.800.000=0,21\%$), el test resultaría positivo en un total de 19.940 sujetos, obteniéndose un valor predictivo positivo del 29,9% (Tabla 3). Así pues, sólo un 29,9% de los sujetos con un resultado positivo en el test resultarían estar realmente afectados, mientras que un 70,1% de los mismos no presentarían la enfermedad. Resulta obvio que en una comunidad como la gallega la utilización de esta prueba no resultaría útil, debido a la alta proporción de falsos positivos que conllevaría.

Veamos ahora que ocurriría si se aplicase la misma prueba a una población en la que el número de enfermos VIH+ fuese de 800.000 (resultando en una prevalencia mucho mayor de un $800.000/2.800.000=28,6\%$). En este caso, la predictividad de una prueba positiva aumenta de un 29,9% a un 98,7%, disminuyendo la proporción de falsos positivos a tan sólo un 1,3% (Tabla 4). Por lo tanto, si la prevalencia es alta, un resultado positivo tiende a confirmar la presencia de la enfermedad, mientras que si la prevalencia es baja, un resultado positivo no permitirá afirmar su existencia.

Razones de probabilidad

Queda claro pues cómo la prevalencia es un factor determinante en los valores predictivos de un www. Por lo tanto, éstos, no pueden ser utilizados como índices a la hora de comparar dos métodos diagnósticos diferentes, ni tampoco a la hora de extrapolar los resultados de otros estudios a datos propios. Por ello, resulta necesario determinar otros índices de valoración que sean a la vez clínicamente útiles y no dependan de la prevalencia de la enfermedad en la población a estudiar. Así, además de los conceptos de sensibilidad, especificidad y valores predictivos, se suele hablar del concepto de razón de verosimilitudes, razón de probabilidad, o cociente de probabilidades. Estos miden cuánto más probable es un resultado concreto (positivo o negativo) según la presencia o ausencia de enfermedad:

- **Razón de verosimilitudes positiva o cociente de probabilidades positivo:** se calcula dividiendo la probabilidad de un resultado positivo en los pacientes enfermos entre la probabilidad de un resultado positivo entre los sanos. Es, en definitiva, el cociente entre la fracción de verdaderos positivos (sensibilidad) y la fracción de falsos positivos (1-especificidad):



- **Razón de verosimilitudes negativa o cociente de probabilidades negativo:** se calcula dividiendo la probabilidad de un resultado negativo en presencia de enfermedad entre la probabilidad de un resultado negativo en ausencia de la misma. Se calcula por lo tanto, como el cociente entre la fracción de falsos negativos (1-sensibilidad) y la fracción de verdaderos negativos (especificidad):



Volvamos de nuevo al ejemplo planteado en la Tabla 2 sobre el diagnóstico de cáncer prostático a partir del tacto rectal. En este caso, se obtiene un cociente de probabilidades positivo de 3,20. Ello viene a indicarnos que un tacto anormal es, por lo tanto, 3 veces más probable en un paciente con cáncer prostático que en otro sujeto sin cáncer.

La razón de probabilidades ofrece la ventaja de que relaciona la sensibilidad y la especificidad de una prueba en un solo índice. Además, pueden obtenerse razones de probabilidad según varios niveles de

una nueva medida y no es necesario expresar la información de forma dicotómica, como resultado de normal o anormal o bien positivo y negativo. Por último, al igual que sucede con la sensibilidad y la especificidad, no varía con la prevalencia. Esto permite utilizarlo como índice de comparación entre diferentes pruebas para un mismo diagnóstico.

Hasta ahora hemos abordado el caso de una prueba con un resultado dicotómico (positivo o negativo), pero en muchas situaciones la confirmación de un diagnóstico debe hacerse a partir de un parámetro numérico, sobre todo cuando éste se realiza a partir de determinaciones analíticas. La generalización a estas situaciones se consigue mediante la elección de distintos valores de corte que permitan una clasificación dicotómica de los valores de la prueba según sean superiores o inferiores al valor elegido. La diferencia esencial con el caso más simple es que ahora contaremos no con un único par de valores de sensibilidad y especificidad que definan la exactitud de la prueba, sino más bien con un conjunto de pares correspondientes cada uno a un distinto nivel de decisión. La estrategia de análisis adecuada consistiría en representar gráficamente los pares (1-especificidad, sensibilidad) obtenidos al considerar todos los posibles valores de corte de la prueba, obteniéndose así una curva llamada curva ROC. El área bajo dicha curva se convierte así en el mejor indicador de la capacidad predictiva del test, independiente de la prevalencia de la enfermedad en la población de referencia y en base al cual se podrán establecer comparaciones entre diferentes pruebas diagnósticas.

En definitiva, es sumamente importante el saber valorar la validez y seguridad de las diferentes pruebas diagnósticas con el fin de seleccionar la más adecuada en cada momento. La sensibilidad, la especificidad y los valores predictivos son los criterios tradicionalmente utilizados para valorar la capacidad predictiva de un test. Los estudios de evaluación de tests diagnósticos son el instrumento adecuado para obtener esta información. No obstante, no debemos olvidar que existen determinados aspectos en el diseño de este tipo de investigaciones que pueden afectar a la precisión y a la validez de las estimaciones realizadas. Una vez más, el cálculo de intervalos de confianza puede ayudarnos a conocer la precisión de los índices calculados. La población de estudio, la estrategia de muestreo, la selección del criterio de referencia y la forma de aplicación de las pruebas diagnósticas serán algunos de los elementos a cuidar para evitar la presencia de sesgos.

Anexo

Resultado de la prueba	Verdadero diagnóstico	
	Enfermo	Sano
Positivo	Verdaderos Positivos (VP)	Falsos Positivos (FP)
Negativo	Falsos Negativos (FN)	Verdaderos Negativos (VN)

Resultado del tacto rectal	Resultado de la biopsia prostática		
	Cáncer	Patología benigna	Total
Anormal	634	269	903
Normal	487	1251	1738
Total	1121	1520	2641

Tabla 2. Resultados de la exploración y biopsia prostática de una muestra de pacientes con sospecha de cáncer de próstata.

$$\begin{aligned} \text{Sensibilidad} &= \frac{634}{634+487} = \frac{634}{1121} = 0,5656 \Rightarrow 56,56\% \\ \text{Especificidad} &= \frac{1251}{269+1251} = \frac{1251}{1520} = 0,8230 \Rightarrow 82,30\% \\ \text{Valor predictivo positivo} &= \frac{634}{634+269} = \frac{634}{903} = 0,7021 \Rightarrow 70,21\% \\ \text{Valor predictivo negativo} &= \frac{1251}{487+1251} = \frac{1251}{1738} = 0,7198 \Rightarrow 71,98\% \\ \text{Razón de verosimilitud positiva} &= \frac{\text{Sensibilidad}}{1-\text{Especificidad}} = \frac{0,5656}{1-0,8230} = 3,19 \\ \text{Razón de verosimilitud negativa} &= \frac{1-\text{Sensibilidad}}{\text{Especificidad}} = \frac{1-0,5656}{0,8230} = 0,53 \end{aligned}$$

Tabla 3. Resultados de la aplicación del test de VIH en una población de baja prevalencia.

Resultado del test	Verdadero diagnóstico		
	VIH+	VIH-	Total
Positivo	5.970	13.970	19.940
Negativo	30	2.780.030	2.780.060
Total	6.000	2.794.000	2.800.000

Prevalencia = $6.000/2.800.000 = 0,21\%$
 Sensibilidad = $\frac{VP}{6.000} \Rightarrow VP = 6.000 \times 0,995 = 5.970$
 Especificidad = $\frac{VN}{2.794.000} \Rightarrow VN = 2.794.000 \times VN = 2.780.030$
 $VP+ = \frac{5.970}{5.970+13.970} = 0,299 \Rightarrow 29,9\%$
 $VP- = \frac{2.780.030}{30+2.780.030} = 0,999 \Rightarrow 99,9\%$

Tabla 4. Resultados de la aplicación del test de VIH en una población de alta prevalencia.

Resultado del test	Verdadero diagnóstico		
	VIH+	VIH-	Total
Positivo	796.000	10.000	806.000
Negativo	4.000	1.990.000	1.994.000
Total	800.000	2.000.000	2.800.000

Tabla 4. Resultados de la aplicación del test de VIH en una población de alta prevalencia.

$$\text{Prevalencia} = 800.000/2.800.000 = 28,6\%$$

$$\text{Sensibilidad} = \frac{VP}{800.000} \Rightarrow VP = 800.000 \times 0,995 = 796.000$$

$$\text{Especificidad} = \frac{VN}{2.000.000} \Rightarrow VN = 2.000.000 \times 0,995 = 1.990.000$$

$$VP+= \frac{796.000}{796.000+10.000} = 0,987 \Rightarrow 98,7\%$$

$$VP-= \frac{1.990.000}{4.000+1.990.000} = 0,998 \Rightarrow 99,8\%$$

Bibliografía

- Sackett DL, Haynes RB, Guyatt GH, Tugwell P. Epidemiología clínica. Ciencia básica para la medicina clínica. 2ª ed. Madrid: Editorial médica panamericana; 1994.
- Sandler G. The importance of the history in the medical clinic and the cost of unnecessary www. Am Heart J 1980; 100: 928. [[Medline](#)]
- Morrison AS. Screening in Chronic disease. Second edition. New York: Oxford University Press; 1992.
- Altman D.G., Bland J.M. Statistics Notes: Diagnostic tests 1: sensitivity and specificity. BMJ 1994; 308: 1552. [[Medline](#)]
- Altman D.G., Bland J.M. Statistics Notes: Diagnostic tests 2: predictive values. BMJ 1994; 309: 102. [[Medline](#)]
- Dujardin B, Van der Ende J, Van Gompel A, Unger JP, Van der Stuyft P. Likelihood ratios: a real improvement for clinical decision making? Eur J Epidemiol 1994; 10: 29-36. [[Medline](#)]
- Burgueño MJ, García Bastos JL, González Buitrago JM. Las curvas ROC en la evaluación de las pruebas diagnósticas. Med Clin (Barc) 1995; 104: 661-670. [[Medline](#)]
- Zweig MH, Campbell G. Receiver-operating characteristics (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin Chem 1993; 39: 561-577. [[Medline](#)]
- Altman D.G., Bland J.M. Statistics Notes: Diagnostic tests 3: receiver operating characteristic plots. BMJ 1994; 309: 188. [[Medline](#)]
- Lopez de Ullibarri Galparsoro I, Pita Fernández S. Curvas ROC. Cad Aten Primaria 1998; 5(4): 229-235.
- Argimon Pallás JM, Jiménez Villa J. Métodos de investigación clínica y epidemiológica. 2ª ed Barcelona: Harcourt; 2000.
- Fletcher RH, Fletcher SW, Wagner EH. Clinical epidemiology: the essentials. 3ª ed. Baltimore: Williams and Wilkins; 1996.
- Cabello López JB, Pozo Rodríguez F. Estudios de evaluación de las pruebas diagnósticas en cardiología. Rev Esp Cardiol 1997; 50: 507-519. [[Medline](#)]
- Greenhalgh T. How to read a paper: papers that report diagnostic or screening tests. BMJ 1997; 315: 540-543. [[Medline](#)] [[Texto completo](#)]

Autores

Pita Fernández, S., Pértegas Díaz, S.

Unidad de Epidemiología Clínica y Bioestadística. Complejo Hospitalario Universitario de A Coruña (España)

Cad Aten Primaria 2003; 10: 120-124.

Cálculos Excel

- Estudio de la capacidad predictiva de una prueba diagnóstica

© Descargado el 03/03/2025 13:22:45 Para uso personal exclusivamente. No se permiten otros usos sin autorización.
Copyright © . Elsevier Inc. Todos los derechos reservados.

