

Sensibilidad y especificidad

El equipo investigador debe valorar la validez de las medidas de las variables y seleccionar la más adecuada para su estudio. La sensibilidad, la especificidad y el valor predictivo son los criterios de validez que cuantifican la capacidad de una prueba para clasificar de manera correcta o errónea a una persona, según la presencia o ausencia de una exposición o una enfermedad. Las pruebas diagnósticas son imperfectas y se cometen errores al clasificar a una persona por su resultado.

SENSIBILIDAD Y ESPECIFICIDAD

La validez de una medida se calcula partiendo de la información contenida en una tabla de 2×2 (tabla A3.1). La presencia o ausencia del resultado (en general una enfermedad) se determina a partir de un criterio de referencia, que idealmente debe ser siempre positivo en los individuos con la enfermedad y negativo en aquellos que no la presentan. Por otro lado, se encuentra el resultado de la prueba, o medida en general, que se quiere evaluar.

TABLA A3.1 Presentación de resultados de un estudio de valoración de una prueba diagnóstica

Resultados de la prueba	Clasificación de los individuos según el criterio de referencia		
	Enfermos	No enfermos	Total
Positivo	a	b	n1
Negativo	c	d	n2
Total	m1	m2	N

Verdaderos positivos (a): número de individuos con la enfermedad, en los que el resultado de la prueba diagnóstica es positivo.

Falsos positivos (b): número de individuos sin la enfermedad, en los que el resultado de la prueba diagnóstica es positivo.

Falsos negativos (c): número de individuos con la enfermedad, en los que el resultado de la prueba diagnóstica es negativo.

Verdaderos negativos (d): número de individuos sin la enfermedad, en los que el resultado de la prueba diagnóstica es negativo.

De esta información se derivan los siguientes índices:

Sensibilidad (a/m_1): probabilidad de obtener un resultado positivo en los individuos que tienen la enfermedad.

Especificidad (d/m_2): probabilidad de obtener un resultado negativo en los individuos que no tienen la enfermedad.

Valor predictivo positivo (a/n_1): probabilidad de que un individuo que presenta un resultado de la prueba positivo tenga la enfermedad.

Valor predictivo negativo (a/n_2): probabilidad de que un individuo que presenta un resultado de la prueba negativo no tenga la enfermedad.

TABLA A3.2 Resultados de un estudio que evalúa un nuevo método diagnóstico de infección urinaria

	Cultivo positivo	Cultivo negativo	Total
Tinción positiva	285	12	297
Tinción negativa	61	706	767
Total	346	718	1.064

Para simplificar la exposición, la mayor parte de las veces se considerará que se trata de una medida dicotómica, por lo que su resultado se clasificará en positivo o negativo.

De los resultados obtenidos en cada una de las casillas de la [tabla A3.1](#) se derivan, entre otros, dos índices: la sensibilidad y la especificidad. La *sensibilidad* responde a la pregunta: si un individuo tiene una enfermedad o factor de riesgo, ¿qué probabilidad existe de que el resultado de la medida que se le aplica sea positivo? En otras palabras, la probabilidad de que una medida clasifique correctamente a un individuo enfermo. La *especificidad* responde a la pregunta: si un individuo no tiene la enfermedad o el factor de riesgo, ¿qué probabilidad existe de que el resultado obtenido sea negativo? Representa la probabilidad de que una medida clasifique correctamente a una persona no enferma. Ambos conceptos son características intrínsecas de la propia medida. Es decir, que si esta se aplica en una población de individuos semejante, y los observadores tienen la misma experiencia, la sensibilidad y la especificidad de una medida no variarán cuando se emplee en distintos estudios.

Ejemplo A3.1. Con el objetivo de evaluar un método simple y económico de tinción directa para el diagnóstico de infección urinaria, en un estudio se analizaron 1.064 muestras de orina ([García Curiel, 1988](#)). Se utilizó como colorante el azul de metileno, que permite teñir las bacterias y el material nuclear de los leucocitos. Esta prueba se comparó con el método tradicional de siembra en placa de agar (criterio de referencia). Las orinas se clasificaron como positivas si existía un crecimiento de $10\text{-}10^5$ UFC/ml (unidades formadoras de colonias). Si

el crecimiento era inferior a 10^5 UFC/ml se consideró que no existía bacteriuria significativa.

Los resultados principales se exponen en la [tabla A3.2](#). Se encontraron 346 bacteriurias, lo que representa el 32,5% del total de las muestras estudiadas. La sensibilidad de la tinción para detectar una infección urinaria fue del 82,4% ($284/346$), y la especificidad fue del 92,3% ($706/718$). Así pues, en el grupo de orinas con bacteriuria significativa se identificó con la nueva medida un 82,4%. De igual modo, el resultado fue negativo en un 98,3% de las muestras consideradas negativas. La tinción fue falsamente negativa en un 17,6% de los casos ($100 - \text{sensibilidad}$) y solo fue falsamente positiva en un 1,7% de las ocasiones ($100 - \text{especificidad}$).

Variando el criterio de normalidad se modifican los valores de estos parámetros. Si en el ejemplo A3.1 se considerara que un paciente presenta bacteriuria cuando los valores fueran superiores a 10^4 UFC/ml, se aumentaría la sensibilidad de la medida a costa de reducir su especificidad. Del mismo modo, si se fuese más exigente en el criterio de normalidad, aumentaría la especificidad y disminuiría la sensibilidad.

VALORES PREDICTIVOS

En la práctica clínica, cuando un médico solicita una prueba diagnóstica desconoce si el paciente tiene la enfermedad. Los médicos deben efectuar inferencias sobre la presencia o ausencia de la enfermedad a partir de los resultados de la prueba. Existen dos modos de cuantificar esta inferencia: los valores predictivos y los cocientes de probabilidad.

El *valor predictivo positivo (VPP)* es la probabilidad de que un individuo con un resultado positivo tenga la enfermedad, y el *valor predictivo negativo (VPN)* es la probabilidad de que si el resultado es negativo el paciente no tenga la enfermedad. Los valores predictivos dependen no solo de la sensibilidad y la especificidad, sino también de la prevalencia de la enfermedad.

Ejemplo A3.2. En el ejemplo A3.1 (v. [tabla A3.2](#)), el VPP es del 96% (285/297) y el VPN es del 92% (706/767). El trabajo se llevó a cabo en pacientes ingresados en un hospital o que acudían a las consultas externas, y la prevalencia fue del 32,5%. Si se aplicara la misma prueba a la población que consulta en un centro de salud, la prevalencia de infección urinaria sería, muy probablemente, menor, y por consiguiente el VPP disminuiría. Supongamos que la prevalencia es del 5% y se aplica la prueba a 1.000 personas: la sensibilidad y la especificidad serían las mismas que las calculadas en la [tabla A3.2](#). Los resultados de este estudio hipotético se muestran en la [tabla A3.3](#). La predictividad de la prueba positiva ha disminuido de un 96 a un 71,9%.

La predictividad de una medida no se puede evaluar sin considerar la prevalencia de la enfermedad; si es alta, un resultado positivo tiende a confirmar su presencia, mientras que si es negativo, no ayudará a excluirla. Contrariamente, cuando la prevalencia es baja, un resultado negativo permitirá descartar la enfermedad con un elevado margen

de confianza, pero si es positivo no permitirá afirmar su existencia.

La prevalencia es el factor más determinante de los valores predictivos. La sensibilidad y la especificidad, al ser características intrínsecas de una medida, no sufrirán grandes variaciones según el lugar donde se apliquen, siempre y cuando se realicen en condiciones similares. Sin embargo, esta asunción no siempre se cumple. El espectro de pacientes también varía según el lugar donde se aplica la prueba. Por ejemplo, una misma prueba, cuando se emplea en un programa de detección precoz se aplica a sujetos asintomáticos, mientras que cuando se usa con fines diagnósticos en un hospital de alta tecnología muchos de los pacientes que la reciben tienen una enfermedad avanzada.

Dado que muchos de los pacientes que son vistos en atención primaria con un resultado positivo en una prueba diagnóstica (es decir, la suma de verdaderos y falsos positivos) son derivados a otro nivel de atención, es de esperar que la especificidad se reduzca.

Ejemplo A3.3. En una muestra de 2.000 pacientes visitados en atención primaria con un diagnóstico de sospecha de apendicitis aguda, los que tenían una elevada probabilidad de padecerla eran derivados al hospital de referencia para confirmación y tratamiento. Una comparación entre los resultados observados en ambos niveles de atención mostró que la prevalencia de la enfermedad fue del 14% en las consultas de atención primaria y del 63% entre los pacientes derivados a los servicios de

TABLA A3.3 Resultados de un estudio que evalúa un nuevo método diagnóstico de infección urinaria

	Infección urinaria	No infección urinaria	Total
Tinción positiva	41	16	57
Tinción negativa	9	934	943
Total	50	950	1000

Los valores de sensibilidad (82,4%) y especificidad (98,3%) son los mismos que los calculados en la [tabla A3.2](#).

Valor predictivo positivo: $41/57 = 71,9\%$.

Valor predictivo negativo: $934/943 = 99,0\%$.

urgencia de los hospitales. Este aumento de prevalencia se acompañó también de una distinta prevalencia de los síntomas y signos diagnósticos. Así, el dolor en el cuadrante inferior derecho se observó en el 21% de los pacientes visitados en atención primaria, mientras que en los derivados al hospital esta prevalencia fue del 82%. La derivación de pacientes con resultados falsos positivos supuso que la especificidad del signo dolor en el cuadrante inferior derecho disminuyera desde el 89 hasta el 16%. Como consecuencia, un signo diagnóstico útil en atención primaria (cociente de probabilidad de una prueba positiva de 8 y cociente de probabilidad de una prueba negativa de 0,2) carece de utilidad en los hospitales de referencia (cociente de probabilidad de una prueba positiva y negativa de 1) (Sackett y Haynes, 2002).

La especificidad no siempre se reduce cuando los pacientes son derivados al nivel secundario o terciario de atención, por lo que no existe un «factor» que sirva para ajustar los resultados en función del nivel de atención. La única forma de evitar este problema es repitiendo el mismo estudio en distintas poblaciones y lugares.

RAZONES DE PROBABILIDAD

La razón o cociente de probabilidad compara la probabilidad de obtener un determinado resultado en un individuo que presente la enfermedad con la de obtenerlo en un sujeto en el que se ha descartado su presencia.

La razón de probabilidad de una prueba positiva (RPP) se calcula dividiendo la proporción de casos que tiene un resultado positivo de la prueba (sensibilidad), entre la proporción de personas que no tienen la enfermedad, pero en las que la prueba también ha dado un resultado positivo (1 – especificidad):

$$RPP = \frac{\text{sensibilidad}}{1 - \text{especificidad}}$$

Ejemplo A3.4. En el ejemplo A3.1 (v. tabla A3.2), la RPP sería 82,4% / (100% – 98,3%) = 49, lo cual permitiría interpretar que en el grupo de pacientes diagnosticados de infección urinaria la probabilidad de encontrar un resultado

positivo a la tinción con azul de metileno es 49 veces mayor que en los individuos en los que se ha descartado la enfermedad.

Análogamente, la razón de probabilidad de una prueba negativa (RPN) se calcula dividiendo los casos que tienen un resultado negativo (1 – sensibilidad) entre la proporción de sujetos que no tienen la enfermedad, en los que el resultado de la prueba es negativo (especificidad):

$$RPN = \frac{1 - \text{sensibilidad}}{\text{especificidad}}$$

Ejemplo A3.5. Siguiendo con el mismo ejemplo anterior, la RPN sería (100% – 82,4%) / 98,3 = 0,18, lo que indica que un resultado negativo se encontró 5,5 veces (1 / 0,18 = 5,5) más frecuentemente en los individuos sin infección urinaria que entre aquellos que sí la padecieron (v. tabla A3.2).

La razón de probabilidad relaciona la sensibilidad y la especificidad en un solo índice, por lo que no varía con la prevalencia. Pueden obtenerse razones de probabilidad según varios valores de una nueva medida y no es necesario expresar la información de forma dicotómica, como resultado normal o anormal, o positivo y negativo.

A partir de la razón de probabilidad se pueden calcular los valores predictivos (o probabilidad *a posteriori*) de una prueba. El primer paso consiste en expresar la prevalencia en *odds* de enfermedad. Si, por ejemplo, la prevalencia es del 25%, la *odds* será de 1:3 (25% / 75%). A continuación se multiplica la *odds* de la enfermedad por la RPP, obteniendo así la *odds posprueba*:

$$\text{Odds posprueba} = \text{odds preprueba} \times RPP$$

Esta *odds posprueba* se puede convertir en *probabilidad posprueba*, o lo que es lo mismo, en los valores predictivos, de la siguiente forma:

$$\begin{aligned} &\text{probabilidad posprueba} \\ &= \frac{\text{odds posprueba}}{(\text{odds posprueba} + 1)} \end{aligned}$$

Ejemplo A3.6. En el ejemplo anterior, la prevalencia era del 32,5%, o lo que es lo

mismo, una *odds* preprueba de 1:2, y la RPP era de 49. Por lo tanto, la *odds* posprueba = $1:2 \times 49 = 49:2$ y la probabilidad posprueba = $24,5 / (24,5 + 1) = 0,96$. Es decir, el VPP es del 96%, el mismo que el obtenido anteriormente.

Una prueba útil desde el punto de vista clínico es aquella que, una vez aplicada, genera cambios desde la estimación diagnóstica de la probabilidad preprueba hasta la estimación de la probabilidad posprueba. Desde el punto de vista clínico, las pruebas proporcionan una ganancia de información cuando la prevalencia de la enfermedad es intermedia, superior al 10%. Las pruebas con una RPP superior a 10 o una RPN inferior a 0,1 tienen una gran utilidad clínica; si los valores de las razones de probabilidad están comprendidos entre 5 y 10 o 0,1 y 0,2, tienen una utilidad moderada.

CURVAS ROC

Cuando los valores de la prueba diagnóstica siguen una escala cuantitativa, la sensibilidad y la especificidad varían según el punto de corte elegido para clasificar a la población como enferma o no enferma. En esta situación, una forma más global de conocer la exactitud de una prueba en el conjunto de puntos de corte es mediante el uso de curvas ROC (*receiver operating characteristics* o curvas de características operativas para el receptor).

La curva ROC es un gráfico en el que en el eje de ordenadas se sitúa la sensibilidad (proporción de verdaderos positivos) y en el eje de abscisas el complementario de la especificidad (1 – especificidad, o proporción de falsos positivos) (fig. A3.1). Cada punto de la curva representa el valor de la RPP correspondiente a un punto de corte determinado.

El área bajo la curva (*area under the curve*, *AUC*) se define como la probabilidad de clasificar correctamente a un par de individuos, uno sano y otro enfermo, seleccionados al azar, al aplicarles la prueba, y es independiente de la prevalencia de la enfermedad. Por ejemplo, un área bajo la curva de 0,75 significa que un individuo seleccionado aleatoriamente del grupo de enfermos tendrá el

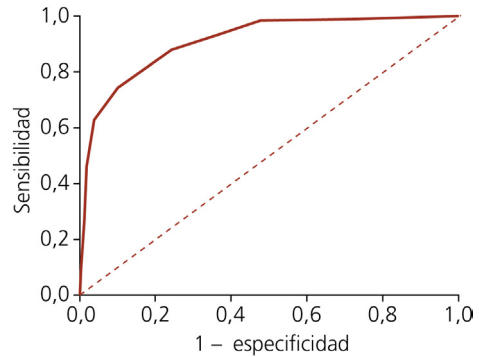


FIGURA A3.1 Ejemplo de curva ROC.

75% de las veces un valor de la prueba mayor que un individuo sano elegido al azar. No significa que un resultado positivo se produzca en los individuos enfermos con una probabilidad de 0,75 ni que esté asociado con la enfermedad el 75% de las veces.

Para una prueba con una sensibilidad y una especificidad del 100%, la curva ROC estaría representada por los lados izquierdo y superior de la figura (área bajo la curva igual a 1). Cuando la prueba no tiene ningún poder de discriminación, es decir, se observan los mismos resultados en los individuos enfermos que en los sanos, la curva ROC está representada por la diagonal principal del gráfico (área bajo la curva igual a 0,5). Como norma general, si el área bajo la curva es mayor de 0,9, la prueba es muy exacta, mientras que valores comprendidos entre 0,7 y 0,9 indican una exactitud moderada. Los valores comprendidos entre 0,5 y 0,7 se corresponden con una exactitud baja. En definitiva, cuanto más próxima es una curva ROC a la esquina superior izquierda, más alta es la exactitud global de la prueba.

La curva ROC facilita la elección del punto de corte. La mayoría de las curvas ROC presentan un segmento de gran pendiente, en el cual la sensibilidad aumenta mucho, mientras la proporción de falsos positivos prácticamente no varía. No tiene mucho sentido elegir el punto de corte en esta zona, porque al desplazarse hacia arriba por la curva aumentará la sensibilidad sin reducirse sustancialmente la especificidad. Igualmente, tampoco es aconsejable elegir el punto de corte en la zona

plana de la curva, ya que la sensibilidad se mantiene prácticamente inalterada mientras que la proporción de falsos positivos aumenta. Por lo general, si el coste que supone cometer un falso positivo es similar al de cometer un falso negativo, el mejor punto de corte es el más próximo al ángulo superior izquierdo del gráfico. Si el coste de un falso positivo y uno negativo difieren, se deberá tener en cuenta este distinto coste para el cálculo del punto óptimo de corte.

Los intervalos de confianza del área bajo la curva permiten efectuar comparaciones estadísticas entre distintas pruebas diagnósticas, siempre que se apliquen en la misma población y para responder a una misma duda diagnóstica. Cuando los resultados de las pruebas se expresan de forma dicotómica, se pueden comparar mediante la *odds ratio* diagnóstica, que se calcula del siguiente modo:

$$OR \text{ diagnóstica} = \frac{\text{sensibilidad} \times \text{especificidad}}{(1 - \text{sensibilidad}) \times (1 - \text{especificidad})}$$

Se puede demostrar matemáticamente que la *odds ratio* diagnóstica es indepen-

diente de los cambios que ocurren en la prevalencia y del espectro de enfermedad que presentan los pacientes. Los investigadores pueden usar técnicas multivariantes, como la regresión logística, para identificar el punto de corte que optimiza el resultado de la *odds ratio*. Las pruebas útiles suelen tener *odds ratio* superiores a 20 (es decir, una RPP de 7 y una RPN de 0,3).

BIBLIOGRAFÍA DE LOS EJEMPLOS

- García Curiel A. Diagnóstico precoz de las infecciones del tracto urinario: examen microscópico y cuantitativo de orina total teñida con azul de metileno. *Enferm Infecc Microbiol Clin*. 1988;6:303-7.
- Sackett DL, Haynes RB. The architecture of diagnostic research. *BMJ*. 2002;324:539-41.

BIBLIOGRAFÍA

- Fletcher RH, Fletcher SW, Fletcher GS. *Epidemiology: the essentials*. 5th ed. Philadelphia: Wolters Kluwer – Lippincott Williams & Wilkins; 2014.
- Gordis L. *Epidemiology*. 5th ed. Philadelphia: Saunders Elsevier; 2014.
- Rothman KJ, Greenland S, Lash TL. *Modern epidemiology*. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2008.