



## FORMACIÓN CONTINUADA

# Las pruebas diagnósticas. Su aplicación en los estudios epidemiológicos

O. Ávalos

Médico Epidemiólogo y Bioestadístico. Unidad de Investigación. Hospital de Gran Canaria Dr. Negrín. Las Palmas de Gran Canaria.

Las pruebas diagnósticas son de uso frecuente en los estudios epidemiológicos, bien sea a partir de observaciones clínicas o de técnicas de laboratorio, mediante las cuales los sujetos se clasifican como sanos o como pertenecientes a un grupo de una determinada enfermedad. De ahí su importancia en medicina y, en particular, para el diagnóstico precoz de diferentes entidades nosológicas<sup>1</sup>.

La utilización de diferentes pruebas diagnósticas en el ámbito sanitario constituye una práctica cada vez más extendida, dada su amplia gama de aplicación en todas las especialidades médicas. La necesidad de conocer —entre sus múltiples usos— la eficacia de un determinado tratamiento para una afección dada (ya sea a partir de medicación habitual, radioterapia, quimioterapia, tipo de intervención quirúrgica, etc.), qué dosis a indicar a los fines de obtener una respuesta satisfactoria en el tratamiento de distintas enfermedades y la validación de diferentes «puntos de corte» para determinaciones analíticas.

Otros casos en los que resulta conveniente el empleo de estos tests son: para la detección precoz de enfermedades a partir de diferentes pruebas, en la medición de los valores predictivos de varios tests para una determinada afección dada la prevalencia de la enfermedad en la población objeto de estudio, en la evaluación de la precisión diagnóstica de varias pruebas para la detección de algunas entidades patológicas (v.g., estudios electrocardiográficos antes y después de esfuerzos físicos); así como otros múltiples usos.

Este trabajo tiene como *objetivos*: ofrecer las definiciones y formas de cálculo de los distintos indi-

cadoreos relacionados con las pruebas diagnósticas y brindar algunos ejemplos de diferentes tipos de investigaciones en las cuales se han utilizado este tipo de pruebas.

Al fin de lograr los objetivos propuestos se realizó una amplia revisión bibliográfica de las publicaciones que han aparecido en los últimos años acerca del tema y de sus diferentes aplicaciones, aunque sólo se seleccionaron unos pocos artículos para ofrecerlos como ejemplos dadas las limitaciones obvias de espacio.

### PRUEBAS DIAGNÓSTICAS: CONCEPTOS BÁSICOS. CUÁNDO Y CÓMO UTILIZARLAS. ANÁLISIS DE CADA UNA DE ELLAS

Esta tabla —una tabla clásica de  $2 \times 2$ — en la cual aparecen en las columnas la presencia o ausencia de enfermedad y en las filas el resultado de la prueba diagnóstica (positiva o negativa), permitirá comprender con mayor facilidad los conceptos básicos de los tests diagnósticos.

	ENFERMEDAD		
	+	-	
PRUEBA DIAGNÓSTICA	+	<i>a</i>	<i>b</i>
	-	<i>c</i>	<i>d</i>

Como se puede apreciar, *a* identifica el número de individuos en presencia de enfermedad y con prueba diagnóstica positiva; *b* indica a aquellos sanos y con prueba positiva; la celda *c* se corresponde con los enfermos y resultado negativo y la casilla *d* con los sanos y prueba negativa.

### Pruebas diagnósticas simples

1. *Sensibilidad*: Es la proporción de individuos con la enfermedad que tienen una prueba positiva:

**Correspondencia:** Dr. José Carlos Rodríguez Pérez  
Unidad de Investigación  
Hospital de Gran Canaria Dr. Negrín  
Bco. la Ballena, s/n  
35020 Las Palmas de Gran Canaria

$a/(a + c)$ . Donde  $a$  = número de *verdaderos positivos* y  $c$  = número de *falsos negativos*.

De esta manera, la sensibilidad nos indica cuán buena es una prueba diagnóstica para identificar una enfermedad; por lo que se le llama también *tasa o proporción de verdaderos positivos*.

2. *Especificidad*: Es la proporción de sanos que tienen una prueba negativa:  $d/(b + d)$ . Es oportuno recordar que  $d$  = número de *verdaderos negativos* y  $b$  = número de *falsos positivos*.

Es decir, la especificidad valora la utilidad de una prueba a los fines de identificar a los no enfermos (*tasa o proporción de verdaderos negativos*).

*Para que un indicador sea útil debe presentar una alta sensibilidad y especificidad.*

A partir de estos dos conceptos podemos hablar de una *proporción de falsos negativos*: probabilidad de que una persona enferma obtenga un resultado negativo en la prueba (enfermos que han tenido un resultado negativo de entre el total de enfermos) y de forma similar encontraremos una *proporción de falsos positivos*: probabilidad de que un individuo sano tenga un resultado positivo (sanos con prueba positiva de entre todos los sujetos que no presentan la enfermedad).

Un ejemplo de la utilización de estos dos conceptos es la investigación que desarrollaron Carol Parkes y cols.<sup>2</sup> con el propósito de evaluar mediciones séricas del antígeno específico de la próstata como una prueba diagnóstica potencial en el diagnóstico del cáncer de próstata. Se trató de un estudio prospectivo en 265 hombres asintomáticos que posteriormente desarrollaron un cáncer prostático y en 1.055 controles pareados por edad, centro hospitalario; y duración y almacenamiento de las diferentes muestras que les fueron tomadas. Calcularon la sensibilidad (proporción de casos en el estudio con un nivel del antígeno específico por encima de un determinado punto de corte) y la tasa de falsos positivos (1-especificidad), definida como la proporción de controles con un valor superior a ese mismo punto de corte.

Los autores encontraron, entre otros resultados del trabajo, que el hecho de realizar las pruebas a los hombres asintomáticos y efectuar un seguimiento hasta el fallecimiento o hasta la presencia clínica de cáncer, proporciona una evaluación no sesgada de la medición del antígeno específico de la próstata como test diagnóstico.

3. *Valor predictivo positivo (VPP)*: Es la probabilidad de que un individuo con resultado positivo en la prueba, tenga la enfermedad:  $a/(a + b)$ . Es decir, se trata de los enfermos con prueba positiva de entre todos los tests positivos. Este concepto se denomina también probabilidad «a posteriori» o probabilidad «post-test».

4. *Valor predictivo negativo (VPN)*: Es la probabilidad de que un individuo con resultado negativo en la prueba no tenga la enfermedad:  $d/(c + d)$ . Se trata de los sujetos libres de enfermedad y con test negativo dentro de todos aquéllos con prueba negativa.

Los valores predictivos, tanto el positivo como el negativo dependen de forma muy importante de la prevalencia de la enfermedad (*medida de la frecuencia de la enfermedad que aportan los estudios transversales, es decir, para un momento dado. Indica la proporción de personas afectadas en ese momento*). Al aumentar la prevalencia crece el valor predictivo positivo para una misma sensibilidad y especificidad; lo cual se debe, fundamentalmente, a que disminuye el número de falsos positivos. Por otra parte, cuando disminuye la prevalencia, se reduce también el valor predictivo positivo y aumenta el negativo, dado que para una misma sensibilidad y especificidad, disminuyen los falsos negativos<sup>3</sup>.

Como ejemplo de la aplicación de estos dos últimos conceptos está el estudio prospectivo llevado a cabo por Stanley Usen y otros investigadores<sup>4</sup>, con el objetivo de determinar las correlaciones clínicas con la hipoxemia en niños internados en un hospital de Gambia por presentar infecciones agudas respiratorias. A estos efectos, construyeron varios modelos de regresión logística para identificar cuáles eran las variables predictoras. Posteriormente calcularon el valor diagnóstico de aquéllas identificadas como tales, a partir de la sensibilidad, la especificidad y los valores predictivos positivo y negativo.

5. *Valor global de la prueba o precisión de la prueba*: Es la probabilidad de que un individuo sea clasificado correctamente por la prueba:  $(a + d)/(a + b + c + d)$ . Así, en el numerador aparecen los enfermos con resultado positivo ( $a$  = verdaderos positivos) y los sanos con test negativo ( $d$  = verdaderos negativos); mientras que en el denominador se encuentran todos los sujetos.

6. *Razón de verosimilitud positiva (RVP)*: Es la probabilidad de una prueba positiva en presencia de la enfermedad (*sensibilidad*), dividida por la probabilidad de una prueba positiva en ausencia de la enfermedad (*1-especificidad*):  $(a/a + c)/(b/b + d)$ . Entonces, esta razón no es más que *sensibilidad/(1-especificidad)*.

7. *Razón de verosimilitud negativa (RVN)*: Es la probabilidad de una prueba negativa en presencia de la enfermedad (*1-sensibilidad*), dividida por la probabilidad de una prueba negativa en ausencia de la enfermedad (*especificidad*):  $(c/a + c)/(d/b + d)$ . Por lo que esta razón sería:  $(1-sensibilidad)/especificidad$ .

En la medida en que los valores de las razones de verosimilitud se alejen de 1 hacia  $\infty$  (en el caso de la positiva), o hacia 0 (en la negativa); mejor

será el cociente y la información que aporte a la prueba.

Para una misma prevalencia, una prueba diagnóstica con una razón de verosimilitud positiva alta tiende a aumentar la probabilidad «post test» de un resultado. En sentido contrario: para una misma prevalencia, una prueba diagnóstica con un valor de la razón de verosimilitud negativa alto, tiende a disminuir la probabilidad «post test» de un resultado.

Dicho de otra manera, estas dos medidas indican la razón entre la probabilidad de un resultado en presencia de enfermedad y la probabilidad de un resultado en ausencia de la enfermedad. Estos cocientes resumen el mismo tipo de información que la sensibilidad y la especificidad expresando, además, cuántas veces es más probable que se encuentre un resultado en personas enfermas en comparación con las sanas<sup>3</sup>.

Estas razones de verosimilitud (tanto positiva como negativa), tienen la ventaja de que, al igual que la sensibilidad y la especificidad, no varían con la prevalencia de la enfermedad.

## COMBINACIÓN DE PRUEBAS DIAGNÓSTICAS

1. *Pruebas en serie*: Consisten en realizar otra prueba diagnóstica simple (o la misma) a los sujetos clasificados como *positivos* en una prueba anterior. Los sujetos con resultado negativo en una prueba anterior ya se consideran «negativos» y, por tanto, no se someten a la siguiente prueba. De esta manera ( $a + b$ ) de la tabla mostrada antes, sería ahora ( $a + b + c + d$ ) en esta segunda tabla; la celda  $a$  pasaría a ser  $a + c$  y, en igual sentido,  $b$  se convertiría en  $b + d$ .

Una característica importante de estas *pruebas en serie* es que *disminuyen la sensibilidad y aumentan la especificidad*, lo cual se explica por el hecho de que no se someten a esta segunda prueba los individuos que ya habían sido clasificados como negativos en el primer test.

2. *Pruebas en paralelo*: Consisten en realizar una nueva prueba o la misma, a las personas que habían sido identificadas previamente como *negativas*. Así, los sujetos con resultado positivo son excluidos de esta próxima prueba, quedando sólo aquellos que habían sido clasificados como negativos. Entonces,  $c + d$  de la tabla anterior pasaría a ser  $a + b + c + d$  en esta nueva prueba,  $c$  se convertiría en  $a + c$ ; y  $d$  en  $b + d$ . Es oportuno señalar que estas *pruebas en paralelo aumentan la sensibilidad y disminuyen la especificidad*, lo cual es obvio dada la exclusión de los que habían resultado positivos<sup>5</sup>.

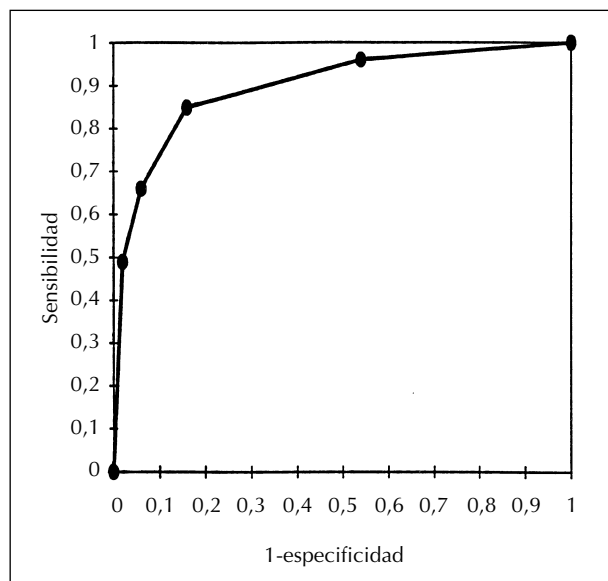


Fig. 1.—Ejemplo teórico de una Curva ROC para determinar el mejor punto de corte, de entre cuatro diferentes, de una nueva prueba diagnóstica para una enfermedad dada. En este ejemplo se han incluido a 90 personas con antecedentes de padecer la enfermedad y a 130 supuestamente sanos. Los puntos de corte que hemos escogido fueron: 70, 75, 80 y 85 unidades. Una vez realizadas las pruebas diagnósticas simples de sensibilidad y especificidad para cada uno de ellos, se calculó la Curva ROC (utilizando el programa EPIDAT<sup>3</sup>) cuyos resultados mostraron una sensibilidad de 49% y una especificidad de 98%, que coinciden con el punto de corte de 75. El área obtenida debajo de la curva fue de 0,897.

## CURVAS ROC (RECEIVER-OPERATING CHARACTERISTICS)

Son la representación gráfica de la capacidad discriminativa de una determinada prueba a partir de todos los posibles «puntos de corte». En el eje de las ordenadas se sitúan los distintos valores de la *sensibilidad* y en el de las abscisas los valores de *1-especificidad* (que indican las proporciones de los falsos positivos)<sup>5</sup>. La curva que se encuentra en esta representación gráfica situada más por encima de la línea diagonal (fig. 1) nos mostrará cuál es la prueba (o el indicador) con mayor precisión. Los valores de estas áreas debajo de las curvas oscilan entre 0,5 y 1. Una cifra de 0,5 nos indicará que esa prueba no tiene valor discriminante, mientras que un valor del área de 1, nos hace ver que se trata de un indicador o prueba diagnóstica «perfecto»<sup>6</sup>.

Ventajas de las curvas ROC:

a) Pueden evaluar el rendimiento de un test, de entre una amplia gama de pruebas diagnósticas.

b) Puede compararse el resultado de una prueba con el de otras.

c) Permiten seleccionar el mejor punto de corte de entre todos los elegidos<sup>7</sup>. El punto de mayor *eficiencia* podría, de esta manera, ser identificado a partir del mejor punto de corte de estas curvas.

Problemas que podrían presentarse con las curvas ROC:

a) El mejor punto de corte no necesariamente tiene que coincidir con la curva que tenga el vértice superior izquierdo más alto.

b) El área debajo de la curva evalúa el conjunto de los resultados de los tests, pero puede incluir algunos puntos de corte irrelevantes o no útiles<sup>7</sup>.

Otra gran aplicación de estas curvas es en la Regresión Logística, con la finalidad de evaluar la precisión en la predicción del modelo seleccionado.

Dos ejemplos de investigaciones en las cuales se utilizaron las curvas ROC: Para verificar la hipótesis de que los niveles plasmáticos elevados de los péptidos natriuréticos pueden servir para diagnosticar a los pacientes con disfunción ventricular izquierda, Frank Muders y cols.<sup>8</sup>, realizaron un estudio en el cual construyeron las curvas ROC para esos fines y, además, para identificar a los pacientes con hemodinámica normal y patológica y parámetros angiográficos. También obtuvieron la sensibilidad, especificidad y el valor global de la prueba (precisión), utilizando la prevalencia de disfunción ventricular de una muestra de los pacientes. Las diferencias resultantes entre los diversos tests diagnósticos elaborados, fueron contrastadas con la prueba de Chi-cuadrado.

El otro ejemplo es el trabajo elaborado por Johan Hedstrom y cols.<sup>9</sup>, que tenía como objetivo estimar la utilidad de las concentraciones séricas del complejo de tripsina 2 y (alfa) 1 antitripsina en el diagnóstico y determinación de la severidad de la pancreatitis aguda; en comparación con la proteína sérica C reactiva, la amilasa y las concentraciones del tripsinógeno 2 (como marcadores de referencia). Los marcadores fueron medidos a los pacientes que iban ingresando consecutivamente por presentar dolor abdominal agudo, bien debido a pancreatitis o a otras enfermedades no relacionadas con el páncreas (controles). Se calcularon la sensibilidad y la especificidad para estimar la capacidad de las diferentes pruebas para diferenciar entre una pancreatitis aguda moderada o severa y las enfermedades no pancreáticas, a partir de diferentes puntos de corte. Posteriormente se validaron estos tests a través de las curvas ROC, para describir la precisión diagnóstica de cada uno de ellos.

Las pruebas diagnósticas, al igual que otros tipos de investigaciones, son susceptibles de errores debidos al azar y a los sesgos.

El error aleatorio puede ocurrir gracias al azar, ya que algunos pacientes con una determinada enfermedad pueden tener resultados negativos en una prueba. Aunque este tipo de error es inevitable es, sin embargo, cuantificable. La estrategia que brinda más información para cuantificar el error aleatorio es la construcción de *intervalos de confianza* para la sensibilidad y para la especificidad. Esto permite comparar los rangos de un test con los de otros.

Con respecto a los sesgos, los más frecuentes son los debidos al tamaño de la muestra (en la que influye la prevalencia de la enfermedad) y a la medición del error del resultado de una prueba, que puede ser evaluado por la realización de otros tests<sup>10</sup>.

### EJEMPLO DE PRUEBAS DIAGNÓSTICAS SIMPLES

Se quieren conocer los valores predictivos de una nueva prueba para diagnosticar precozmente la nefropatía diabética. Para ello son sometidos a examen 25.000 pacientes con diabetes mellitus tipo I en la población objeto de estudio. Se estima (por investigaciones realizadas en otros países) que la sensibilidad de la prueba es del 85% y la especificidad del 95%. La prevalencia estimada es del 16%.

Los resultados obtenidos (se utilizó el programa EPIDAT<sup>5</sup>) fueron los siguientes:

Test diagnóstico	Valor	Intervalos de Confianza	
		Límite Inferior	Límite Superior
Sensibilidad	83,3	82,5	84,1
Especificidad	93,8	93,4	94,1
Valor predictivo positivo	88,2	87,5	88,9
Valor predictivo negativo	90,9	90,5	91,3
Valor global de la prueba	90	89,6	90,4
Razón de verosimilitud positiva	13,3	12,5	14,2
Razón de verosimilitud negativa	0,18	0,17	0,19

### CONSIDERACIONES FINALES

En este trabajo se han intentado ofrecer las definiciones, los métodos de cálculo y las aplicaciones de los diferentes tipos de pruebas diagnósticas, destacando de forma especial a las curvas ROC por el hecho de que resumen —y a la vez contrastan— los resultados de la sensibilidad y la especificidad.

Por otra parte, se ha podido conocer su gran utilidad en las investigaciones epidemiológicas, a partir de los ejemplos citados extraídos de la literatura médica. De esta manera, se ha mostrado su aplicación en estudios de: disfunción sistólica del ventrículo izquierdo, hipoxemia infantil, cáncer de próstata, y pancreatitis aguda.

## BIBLIOGRAFÍA

1. Armitage P, Berry G: Estadística para la Investigación Biomédica. Barcelona: Doyma, S. A. 554-555, 1992.
2. Parkes C, Wald NJ, Murphy P, George L, Watt HC, Kirby R, Knekt P, Helzlsouer KJ, Tuomilehto J: Prospective observational study to assess value of prostate specific antigen as screening test for prostate cancer. *Br Med J* 311: 1340-1343, 1995.
3. Gómez de la Cámara A: Caracterización de Pruebas Diagnósticas. *Epidemiología Clínica y Bioestadística*. Madrid: IDEPSA, 43-46, 1998.
4. Usen S, Weber M, Mulholland K, Jaffar S, Oparaugo A, Omisigho C, Adegbola R, Greenwood B: Clinical predictors of hypoxaemia in Gambian children with acute lower respiratory tract infection: Prospective Cohort Study. *Br Med J* 318: 86-91, 1999.
5. Organización Panamericana de la Salud/Xunta de Galicia: EPIDAT. Análisis Epidemiológico de Datos Tabulados. Versión 2.0 para Windows. Galicia: Cograf, 89-97, 1997.
6. Vermeylen, F: StatNews #07: Assessing Logistic Regression. Statistical Consulting News. Cornell University, USA, 1998.
7. Boston University School of Public Health: Clinical Epidemiology. EB725. Session V: Diagnostics Test Performance II: ROC Curves and Likelihood Ratios. Boston, USA, 1996.
8. Muders F, Kromer EP, Griesse DP, Pfeifer M, Hense HW, Riegger GAJ, Elsner D: Evaluation of plasma natriuretic peptides as markers of left ventricular dysfunction. *Am Heart J* 134(3): 442-450, 1997.
9. Hedstrom J, Sainio V, Kempainen E, Haapiainen R, Kivilaasko E, Schroder T, Leinonen J, Setnman UH: Serum complex of trypsin 2 and (alpha)(sub 1) antitrypsin as diagnostic and prognostic marker of acute pancreatitis: clinical study in consecutive patients. *Br Med J* 313: 333-337, 1996.
10. Browner WS, Newman TB, Cummings SR: Designing Clinical Research. An Epidemiologic Approach. Chapter 9. Designing a New Study: III. Diagnostic Tests. Stephen B. Hulley & Steven R. Cummings. Williams & Wilkins, Baltimore, USA, 92-96, 1988.