

Nuevos conceptos sobre la sensibilidad y la especificidad de las pruebas de diagnóstico

Dr. Horacio García Romero*

La sensibilidad y la especificidad de una prueba de diagnóstico son índices que nos señalan la eficacia de esa prueba para establecer o descartar un diagnóstico determinado. El clínico debe conocer con cierta precisión esos valores en las pruebas que maneja cotidianamente, para poder darles el peso adecuado en sus decisiones. Por otra parte, el médico que realiza una investigación sobre una prueba diagnóstica debe determinar su sensibilidad y su especificidad; de otra manera su trabajo será considerado irrelevante.¹

La determinación de estos dos índices se basa en la realización de la prueba en un grupo de personas, algunas de las cuales tienen la enfermedad en estudio y otras no la tienen. Los datos encontrados en estas muestras se extrapolan después a toda la población. La sensibilidad se calcula conociendo el número de personas que tiene la enfermedad, y de éstas, en cuántas la prueba fue positiva. La fórmula clásica para obtener el valor de la sensibilidad es:^{2 3 4}

$$\text{Sensibilidad} = \frac{\text{verdaderos positivos}}{\text{verdaderos positivos} + \text{falsos negativos}}$$

En el numerador aparecen los pacientes con la enfermedad en los que la prueba fue positiva, y en el denominador el número total de pacientes con la enfermedad, independientemente que la prueba haya sido positiva (verdaderos positivos) o negativos (falsos negativos).

La especificidad se calcula conociendo el número de personas sin la enfermedad en los que se realizó la prueba, y de éstas, en cuántas fue negativa la prueba.

Fórmula clásica:^{2 3 4}

$$\text{Especificidad} = \frac{\text{verdaderos negativos}}{\text{verdaderos negativos} + \text{falsos positivos}}$$

En el numerador se encuentran los pacientes sin la enfermedad en los que la prueba fue negativa; y en el denominador las personas estudiadas que no tienen la enfermedad, haya sido la prueba negativa (verdaderos negativos) o positiva (falsos positivos).

Una prueba altamente sensible para un padecimiento tiene pocos falsos negativos. Una prueba altamente específica para un padecimiento tiene pocos falsos positivos.

Nótese que tanto la sensibilidad como la especificidad se refieren a una enfermedad determinada. Una prueba puede tener diferentes valores con relación a diferentes enfermedades. Esto es particularmente cierto en las pruebas de diagnóstico que utilizan imágenes: radiografías, gamagrafías, estudios de ultrasonido, de tomografía axial y otras.

Con relación a la sensibilidad y a la especificidad se presentan algunos problemas a los que se ha puesto poca atención aún en la literatura especializada y a los que en este trabajo se intentará dar una respuesta:

1. Tratándose de pruebas de diagnóstico de imágenes, no es raro que el clínico se enfrente a pruebas dudosas en las que no le es posible realizar un diagnóstico, y sin embargo no puede considerarlas como normales. ¿Que valor debe

* Exsecretario de la División de Estudios de Posgrado, Facultad de Medicina, UNAM.

dársele a estas pruebas? ¿Se deben descartar para calcular los valores de sensibilidad y especificidad?

2. Es frecuente después de practicar un estudio que el diagnóstico no se compruebe. ¿Qué hacer en casos como éste? ¿Deben o no tomarse en cuenta para los cálculos de la sensibilidad y de la especificidad?

3. ¿Qué tan confiables son los resultados? Si los valores fueron obtenidos de una muestra, ¿hasta qué punto pueden aplicarse a la población general?

4. ¿El número de enfermos en los que se practicó el estudio para el cálculo de la sensibilidad y de la especificidad es el apropiado? ¿Cuál debe ser el tamaño de la muestra?

Los casos dudosos. Si el diagnóstico en un determinado estudio es dudoso, esto puede ocurrir por dos razones:

1. Por no poder determinar con precisión si la alteración existe. Por ejemplo: ¿hay una imagen de úlcera en el duodeno deformado de esta serie gastroduodenal? ¿Es anormal esta imagen ultrasonográfica del ovario?

2. Por no poder interpretar una alteración existente. El estudio nos permite reconocer una alteración, pero no nos orienta al diagnóstico. ¿La úlcera gástrica que se observa en esta serie gastroduodenal es benigna o maligna? ¿El crecimiento pancreático que se observa en este estudio tomográfico representa una pancreatitis crónica o un carcinoma?

Estos casos dudosos se presentan, y no es posible soslayarlos. Constituyen un porcentaje importante del número total de estudios de gabinete.

¿Cómo manejar estos datos? Si la imagen es dudosa no nos orienta al diagnóstico de un padecimiento ni lo descarta. Es un factor en contra de la sensibilidad y de la especificidad de la prueba, y como tal debe cuantificarse. Si una prueba fuera dudosa en el 90 por ciento de los

pacientes, positiva en 5 por ciento y negativa en 5 por ciento, aunque en estos dos últimos casos coincidiera con la presencia o ausencia de la enfermedad, jamás diríamos que la sensibilidad y la especificidad de la prueba son de 100 por ciento. Esto ocurriría si descartáramos los casos dudosos. Lo adecuado sería establecer el diagnóstico correcto de los pacientes en los que la prueba resultó dudosa; y aplicar las fórmulas siguientes, que se proponen en sustitución de las fórmulas clásicas:

$$\text{Sensibilidad} = \frac{\text{verdaderos positivos}}{\text{verdaderos} + \text{falsos negativos} + \text{dudosos positivos}}$$

$$\text{Especificidad} = \frac{\text{verdaderos negativos}}{\text{verdaderos} + \text{falsos positivos} + \text{dudosos negativos}}$$

Los dudosos positivos son aquellos estudios que no muestra datos que permitan el diagnóstico, realizados en pacientes que padecen la enfermedad. Los dudosos negativos son aquellos estudios que no permiten descartar un padecimiento, realizados en pacientes que no padecen la enfermedad.

La aplicación de estas fórmulas que toman en cuenta estos casos dudosos, nos acerca más a la verdadera eficacia diagnóstica de la prueba.

Los casos no comprobados. El diagnóstico de un padecimiento debe comprobarse mediante alguna “prueba de fuego” que no deje dudas sobre la certeza del diagnóstico. Esta prueba puede ser una biopsia, una laparoscopia, una intervención quirúrgica o algún otro padecimiento. Con cierta frecuencia, en un número importante de estudios de imágenes el diagnóstico no se comprueba, sobre todo en aquellos casos en los que el resultado es negativo. Los pacientes con colecistografías negativas sólo ocasionalmente son operados. Por otra parte, en estudios de ultrasonido con imágenes que sugieren metástasis de una neoplasia, los pacientes raramente son sujetos a laparoscopia, laparotomía o autopsia.

Estos casos no comprobados ¿cómo deben manejarse? En rigor, se desconoce si en estos estudios se trata de un verdadero o falso positivo o de un verdadero o falso negativo. Se sabe que no es posible aclarar esta información. Sin embargo, no es raro que muchos de estos pacientes se incluyan erróneamente en las investigaciones sobre sensibilidad y especificidad de una prueba. Habitualmente se eliminan los casos positivos no comprobados (ecocardiografías en padecimientos congénitos) y se incluyen los casos negativos no comprobados (las colecistografías negativas se consideran como verdaderas negativas *a priori*).

No es correcto incluir los datos de los casos no comprobados en la determinación de los valores de diagnóstico de una prueba, pues esta información no tiene peso en la determinación de su eficacia. Por otra parte, es muy conveniente señalar en un trabajo el criterio de comprobación del diagnóstico. Puede ser también útil que se indique el número de estudios cuyo diagnóstico no se comprobó.

Confiabilidad de los resultados. La confiabilidad de los datos expresados por la sensibilidad y la especificidad de una prueba se basa en el grado de representatividad de la muestra objeto del estudio y en el establecimiento de los límites de confianza de los valores obtenidos.

Generalmente el grado de representatividad de las muestras es diferente según se estudie la sensibilidad o la especificidad. En la investigación de la sensibilidad, es probable que se tome un grupo más o menos homogéneo de los pacientes que tienen el padecimiento y que representen adecuadamente al grupo de enfermos en general. Cuando se investiga una prueba de diagnóstico en estos enfermos no es común que se haga una selección al azar, lo que originaría una muestra verdaderamente representativa. Sin embargo, salvo excepciones que siempre deben ser consideradas, la muestra de pacientes con la enfermedad, si es suficientemente grande, representa satisfactoriamente la población de enfermos con ese padecimiento. Es relativamente simple determinar la sensibilidad de una prueba, siempre y cuando los pacientes sean suficientemente numerosos para incluir las diversas moda-

lidades del padecimiento.

No ocurre lo mismo con la especificidad. Esta se refiere a las personas que no padecen la enfermedad, las que forman un conjunto muy heterogéneo, en el que pueden distinguirse tres grupos:

1. Personas "normales"
2. Pacientes con padecimientos, comunes o no, que no tienen ninguna relación con el padecimiento en estudio.
3. Pacientes con problemas de alguna manera relacionados con el padecimiento en estudio.

En todos estos casos, el clínico debe conocer la probabilidad de enfrentarse a un estudio falso positivo con referencia al padecimiento por el que se realizó la prueba. Para estudiar su especificidad, debe investigarse si la prueba es negativa en los tres grupos de personas mencionadas, lo que raramente se hace. En el cálculo de la especificidad, una muestra representativa debe incluir un número importante de estos grupos de elementos.

En lo que se refiere a los límites de confianza de nuestros valores, sabemos que éstos han sido calculados en una muestra y que solamente se aproximan a los valores reales de la sensibilidad y la especificidad de la prueba en estudio. Para conocer a ciencia cierta estos valores se tendría que realizar la prueba en toda la población de pacientes.

Cuando se señala que la prueba de Papanicolaou de citología gástrica tiene una sensibilidad de 90 por ciento y una especificidad de 95 por ciento, estamos afirmando que en una serie de estudios en un número grande de pacientes la prueba resultó positiva en el 90 por ciento de los que tenían cáncer (sensibilidad) y fue negativa en el 95 por ciento de los que no tenían cáncer (especificidad). Si la muestra fue representativa, sabemos que los datos obtenidos es poco probable que se alejen mucho de los valores reales de sensibilidad y especificidad de la prueba, valores que se obtendrían si ésta se aplicase a toda la población de pacientes con y sin cáncer gástrico. Pero ¿qué significa en estas frases las expresiones "poco probable" y "que se alejen mucho"? Esto nos lo contesta la Esta-

dística. Si al valor obtenido en nuestra investigación le sumamos y restamos dos veces su desviación estándar, tendremos con cerca de 95 por ciento de probabilidades los límites entre los cuales se haya el valor real. Si la sensibilidad de una prueba es de .80 (80 por ciento) y su desviación estándar S es de .05 ($2S = .10$), podemos asegurar que con 95.4 por ciento de probabilidad la prueba será positiva en $.80 \pm .10$, o sea, .70 a .90 (70 al 90 por ciento) de los casos. En este ejemplo el valor real de la sensibilidad de la prueba aplicada a la población general se encuentra entre 70 y 90 por ciento, con 95.4 por ciento de probabilidad. En este caso el rango de valores probables es muy amplio y puede dejarnos poco satisfechos de nuestros resultados. Si la desviación estándar fuera más chica, por ejemplo: .01, estos límites se encontrarían entre 78 y 82 por ciento; aquí estamos seguros que el valor de nuestra muestra se acerca mucho al valor de la población general.

Con lo anterior, se observa claramente la conveniencia de señalar la desviación estándar de los valores de sensibilidad y especificidad obtenidos en una investigación. Esto se hace sólo excepcionalmente, lo que no deja de ser una deficiencia en el análisis de los resultados de una investigación de este tipo. El cálculo de la desviación estándar de las proporciones (como son la sensibilidad, la especificidad y los valores de predicción de una prueba), se realiza mediante la fórmula:

$$S = \frac{p(1-p)}{n}$$

en la que S es la desviación estándar, p es la proporción en la que una característica se presenta en la muestra (por ejemplo: valor de la sensibilidad), y n es el número de elementos de la muestra.

El tamaño de la muestra. Como se aprecia en la fórmula, la desviación estándar puede variar según sea la proporción encontrada (p) y depende también del tamaño de la muestra (n). En una prueba realizada en 100 pacientes en la que se

encuentra una especificidad de .80, la desviación estándar será de .04, lo que significa que los límites de confianza de esta especificidad serán de $p \pm 2S = .88$ y $.72$ con un 95.4 por ciento de probabilidad. Si la misma prueba se realiza a sólo 50 pacientes, y se encuentra la misma especificidad, la desviación estándar será mayor, en este caso de .056, y los límites de confianza estarán comprendidos, con un 95.4 por ciento de probabilidad, entre .69 y .91. Si uno desea que los límites de confianza permanezcan dentro de ciertos márgenes (por ejemplo, no más de 2.5 o 5 por ciento del valor encontrado), debe establecerse el tamaño de la muestra apropiada.

Para calcular el número de elementos de una muestra en la que se establece una proporción, se ha empleado el coeficiente de variación, que expresa la desviación estándar como porcentaje de la proporción. El tamaño de la muestra se busca en tablas en las que se incluye este coeficiente de variación y la proporción estimada⁵.

Es probable que en la práctica el médico prefiera que los límites de confianza no excedan de valores fijos, como pueden ser el 2.5, el 5 o el 10 por ciento, por encima y por debajo de los valores esperados. El tamaño de la muestra en estos casos se puede encontrar en la siguiente tabla (página 490) elaborada por el autor de este trabajo.

En la tabla, los valores de p pueden ser sustituidos por los de sensibilidad y especificidad esperados. Si no se tiene ninguna información previa que permita suponer algunos valores aproximados, éstos deben determinarse en un estudio piloto. Si se desea que al final del estudio los límites de confianza se encuentran entre $p \pm .025$ (2.5 por ciento arriba y abajo del valor encontrado), el tamaño de la muestra debe buscarse en la columna correspondiente, en la línea horizontal del valor de p . Lo mismo debe hacerse si se encuentran 5 o 10 por ciento por encima o debajo de los valores encontrados. Si la sensibilidad de una prueba es aproximadamente de .90, para que al final del estudio los resultados se encuentran entre $.90 \pm .05$ (85 y 95 por ciento), en la tabla puede encontrarse que la muestra debe ser de un mínimo de 144 elementos.

TABLA 1
Tamaño de la muestra en los estudios de sensibilidad y especificidad de una prueba de diagnóstico

Proporción esperada	Tamaño de la muestra		
	$p \pm .025$	$p \pm .05$	$p \pm .10$ %
.95	291	76	18
.90	552	144	35
.85	782	204	49
.80	981	256	62
.75	1150	300	72
.70	1288	336	81
.65	1422	364	88
.60	1472	384	92
.55	1518	396	95
.50	1562	400	96
.45	1518	396	95
.40	1472	384	92
.35	1422	364	88
.30	1288	336	81
.25	1150	300	72
.20	981	256	62
.15	782	204	49
.10	552	144	35
.05	291	76	18

REFERENCIAS

1. Haynes R. B., "How to read clinical journals: to learn about a diagnostic test", *CMA journal*, 124: 703-710, 1981.
2. Ingelfinger J.A., Mosteller F., Thibodeau L.A., Ware J.H., *Bioestatistic in clinical Medicine*, Nueva York, Toronto, Londres, E.D. Macmillan Publishing Co. pp. 1-23, 1983.
3. Griner P.F., Mayewsk. R.J., Muchlin A. y Greenland P., "Selection and interpretation of diagnostic tests and procedure", *Annals of Internal Medicine*, 94: 555-600, 1981.
4. Galen R.S., Gambino S.R., *Beyond normality: The predictive value and efficiency of medical diagnosis*: Nueva York-Londres. Ed. John Wiley and sons. pp. 9-15, 1975.
5. Cañedo D.L., García Romero H. Méndez R.I., *Principios de Investigación Médica*, México, DIF, p. 395, 1980.