

Cómo estudiar un estudio y probar una prueba: lectura crítica de la literatura médica¹

Segunda edición

Richard K. Riegelman y Robert P. Hirsch

PARTE VI:

Capítulo 16. Definición de enfermedad: la prueba de oro

Capítulo 17. Discriminación diagnóstica de las pruebas

Capítulo 18. Resumen: la prueba de una prueba

¹Título original: *Studying a Study and Testing a Test. How to Read the Medical Literature* Second edition. © Richard K. Riegelman, Robert P. Hirsch. Publicado por Little, Brown and Company, Boston, Massachusetts 02108, Estados Unidos de América. Los pedidos del libro en inglés deben dirigirse a esta dirección.

Versión en español autorizada por Little, Brown and Company; se publica simultáneamente en forma de libro (Publicación Científica 531) y como serie en el *Boletín de la Oficina Sanitaria Panamericana*. Traducción de José María Borrás, revisada por el Servicio Editorial de la Organización Panamericana de la Salud.

© Little, Brown and Company, 1989. Todos los derechos reservados. Ninguna parte de esta publicación puede ser reproducida ni transmitida en ninguna forma ni por ningún medio de carácter mecánico o electrónico, incluidos fotocopia y grabación, ni tampoco mediante sistemas de almacenamiento y recuperación de información, a menos que se cuente con la autorización por escrito de Little, Brown and Company.

DEFINICIÓN DE ENFERMEDAD: LA PRUEBA DE ORO

Cuando se aplica cualquier prueba diagnóstica, ya sea a personas que padecen una enfermedad o a las que no la padecen, los resultados representan un recorrido de valores. En los enfermos, la variabilidad de los resultados puede reflejar diferencias en la gravedad de la enfermedad o una respuesta individual a la misma. A pesar de esta variabilidad, es esencial definir un grupo de pacientes que, sin lugar a dudas, padecen la enfermedad.

LA PRUEBA DE ORO

La prueba o criterio utilizado para definir inequívocamente una enfermedad se conoce como *prueba de oro*¹ (*gold standard*). La prueba de oro puede ser una biopsia, un angiograma, una necropsia posterior o cualquier otra prueba reconocida. El uso de un criterio de oro con el fin de identificar definitivamente a los que tienen la enfermedad es un requisito para examinar la utilidad diagnóstica de cualquier prueba nueva o no evaluada. En otras palabras, la utilidad de la nueva prueba se basa en su comparación con la de oro. De este modo, una prueba nueva se compara con una prueba (o pruebas) antigua y más aceptada para determinar si la nueva ofrece el mismo rendimiento que la de referencia. Observe que se parte del supuesto de que, utilizando la mejor de las pruebas antiguas, es posible tener un 100% de posibilidades de realizar un diagnóstico correcto; la suposición de partida es la imposibilidad de “inventar una mejor trampa para ratones”, dado que no se puede superar el 100%. Puede existir una trampa para ratones más barata o más práctica pero, por definición, ninguna con la que se atrapen más ratones.

Puede parecer equívoco afirmar que la única forma de evaluar la capacidad diagnóstica de una prueba nueva es suponer que ya es posible realizar diagnósticos perfectos. Lamentablemente, esa es la posición en que nos encontramos al evaluar una prueba nueva. Solo podemos preguntarnos si la prueba está a la altura de la mejor de las pruebas antiguas, esto es, la prueba de oro.

A pesar de la limitación intrínseca de nuestra capacidad para evaluar inicialmente una prueba nueva, el tiempo y las aplicaciones repetidas están del lado de la mejor trampa para ratones. Una vez que se aplica a la práctica clínica, puede hacerse evidente que, en realidad, la prueba nueva predice mejor el curso clínico subsiguiente que la de referencia. Incluso es posible que con el tiempo la prueba nueva sea aceptada como prueba de oro. No obstante, el problema que surge con frecuencia es que, si bien se puede realizar el diagnóstico definitivo, la prueba acarrea un riesgo excesivo o se realiza demasiado tarde para rendir sus máximos beneficios clínicos. Es decir, existe una prueba de oro adecuada que no es práctica en el sentido clínico. En estos casos, es útil comprobar que la prueba nueva está a la altura de la de oro. Debe entenderse, repetimos, que el objetivo de evaluar una prueba se limita a compararla con la mejor prueba disponible. Por esta razón, es preciso estar seguro de que se está utili-

¹ N del E Se traducirá como prueba de oro, criterio de oro, y prueba o criterio de referencia, según el contexto

zando la mejor prueba de oro disponible. Examinemos, a modo de ejemplo, lo que puede suceder cuando la prueba de oro utilizada no es la más adecuada.

Se practicó la autopsia a 100 individuos que fueron ingresados en un hospital con "ondas Q diagnósticas" en su electrocardiograma (ECG) y que fallecieron en la hora siguiente al ingreso, con objeto de determinar si habían sufrido infarto de miocardio (IM). La necropsia, que se utilizó como criterio de oro del IM, reveló pruebas de IM en solo 10 sujetos. Los autores concluyeron que el ECG no era un método útil para realizar el diagnóstico de IM e insistieron en aceptar el diagnóstico anatomopatológico como la prueba de oro.

La utilidad de todas las pruebas diagnósticas se determina comparándolas con pruebas de referencia cuya aptitud para medir las características estudiadas ya se ha establecido con la práctica. Los diagnósticos por necropsia se utilizan frecuentemente como criterio de oro contra el cual se juzgan las otras pruebas. Sin embargo, la necropsia no siempre constituye una forma ideal de medir la enfermedad, como muestra este ejemplo, dado que a veces debe pasar bastante tiempo antes de que se manifiesten los signos patológicos del IM. Es posible que las ondas Q diagnósticas de un ECG reflejen mejor el IM que los cambios patológicos observables en una necropsia. El investigador debe cerciorarse de que el criterio de oro utilizado ha sido realmente establecido como la mejor referencia posible, antes de usarlo como base de comparación.

Por desgracia, incluso las mejores pruebas de referencia disponibles muchas veces no distinguen inequívocamente a los enfermos de los sanos. Puede que los casos de enfermedades leves o en sus fases iniciales no satisfagan los criterios de la prueba de oro. A menudo, los investigadores están tentados de incluir solamente a aquellos individuos que presentan pruebas claras de la enfermedad, tal como se miden con la prueba de referencia. A pesar de la certeza intelectual que parece proporcionar, esto puede redundar en una investigación que se limita a los individuos que tienen una enfermedad grave o en fase muy avanzada. Este peligro se ilustra con el siguiente ejemplo.

Un investigador comparó la capacidad de la citología de la orina para diagnosticar el cáncer de vejiga urinaria con la del diagnóstico inequívoco por biopsia de casos de cáncer invasor de vejiga que cumplían los criterios diagnósticos de la prueba de oro. Mediante el examen citológico se identificó a 95% de las personas que tenían cáncer. Sin embargo, cuando se aplicó en la práctica clínica, la citología de orina solo detectó 10% de los casos.

Al considerar solo los casos avanzados de cáncer invasor de vejiga urinaria, los investigadores habían eliminado los casos dudosos o en etapas iniciales de la enfermedad. Por lo tanto, no debe sorprender que, al aplicar la prueba en la práctica clínica, su rendimiento no fuera tan bueno como el obtenido cuando se comparó con una prueba de oro definitiva.

Por muy tentador que sea estudiar tan solo a los individuos con enfermedades claramente definidas, es engañoso sacar conclusiones sobre la utilidad de una prueba que se ha aplicado exclusivamente a individuos con una enfermedad avanzada o grave. Cuando se valora la discriminación diagnóstica de una prueba, es importante preguntarse si se utilizó el mejor criterio de referencia para definir a las personas con la enfermedad. También es importante preguntarse si con los enfermos estudiados se abarcó todo el espectro de la enfermedad. Debemos reconocer que a veces es imposible lograr ambos objetivos simultáneamente.

Aunque se cumplan estas condiciones, es preciso apreciar que el propósito de probar una prueba se limita a determinar si la prueba estudiada es tan buena como la prueba de referencia establecida. Los métodos empleados no contemplan la posibilidad de que la prueba nueva sea mejor que la de oro.

DISCRIMINACIÓN DIAGNÓSTICA DE LAS PRUEBAS

Hoy día es posible medir la capacidad de una prueba para discriminar entre los enfermos y los sanos. Al hacer esa valoración, es importante considerar los tres puntos siguientes:

1. Variabilidad de la prueba: medición de la reproducibilidad del resultado de la prueba. El intervalo de variabilidad debe ser relativamente menor que el intervalo de la normalidad.
2. Variabilidad de la población sana: determinación de los valores del intervalo de la normalidad para la prueba.
3. Definición de la prueba de oro: identificación de los grupos de individuos que definitivamente tienen la enfermedad y de los que no la tienen según la prueba de oro.

SENSIBILIDAD Y ESPECIFICIDAD

Las medidas tradicionales del valor diagnóstico de una prueba son la *sensibilidad* y la *especificidad*. Estas miden la discriminación diagnóstica de la prueba comparada con la del criterio de referencia, que, por definición, tiene una sensibilidad y una especificidad de 100%. La sensibilidad y la especificidad se han seleccionado como medidas, porque son características intrínsecas de una prueba que deben ser idénticas, ya sea que se aplique a un grupo de pacientes en los cuales la enfermedad es rara o a un grupo de pacientes en los que es frecuente.¹ Por esta razón, proporcionan medidas de la discriminación diagnóstica de una prueba, que deben ser las mismas sea cual fuere la probabilidad de enfermedad antes de realizar la prueba. La estabilidad de la sensibilidad y la especificidad permite a los investigadores de Los Ángeles, París o Tokio aplicar la misma prueba diagnóstica y esperar resultados similares a pesar de las diferencias importantes que existen entre las poblaciones. Estas medidas también permiten a los investigadores y a los clínicos comparar directamente el rendimiento de una prueba con el de otras.

La sensibilidad mide la proporción de los individuos con la enfermedad que son identificados correctamente por la prueba. En otras palabras, mide lo sensible que es la prueba para detectar la enfermedad. Puede ser útil recordar la sensibilidad como *positiva en los enfermos* (PEE). La especificidad mide la proporción de los individuos sanos que son correctamente identificados como tales por la prueba. La especificidad se puede recordar como *negativa en los sanos* (NES).

Observe que la sensibilidad y la especificidad solamente indican la proporción o porcentaje de los que han sido correctamente clasificados como sanos o como enfermos. Estas medidas no predicen el número real de individuos que serán clasificados correctamente, cifra que dependerá de la frecuencia de la enfermedad en el grupo al que se aplique la prueba.

¹ Es posible que esto no sea estrictamente cierto, si la proporción de enfermos en estadios iniciales de la enfermedad cambia junto con la frecuencia de la enfermedad. Una prueba puede tener sensibilidad y especificidad distintas para una fase inicial de la enfermedad y para una avanzada.

La sensibilidad y la especificidad son medidas útiles, porque permiten a los lectores y a los investigadores obtener los mismos resultados cuando evalúan una prueba en grupos de pacientes que difieren en la frecuencia de la enfermedad. Sin embargo, los valores numéricos pueden ser diferentes según que se obtengan de un grupo de pacientes en los estadios iniciales de la enfermedad o de otros en estadios avanzados.

Primero mostraremos la forma de calcular la sensibilidad y la especificidad y luego sus implicaciones y limitaciones. Para calcular la sensibilidad y la especificidad de una prueba en comparación con la de oro, se siguen los siguientes pasos:

1. Los investigadores seleccionan una prueba de oro que se usará para identificar los individuos enfermos.
2. Seguidamente, escogen a un grupo de pacientes que según el criterio de referencia padecen la enfermedad y a otro grupo de individuos que según el mismo criterio están sanos. Al aplicar este criterio, es importante saber si los investigadores incluyeron a grupos representativos de individuos con y sin la enfermedad. En otras palabras, ¿representan los individuos seleccionados el espectro completo de los que tienen la enfermedad y de los que no la tienen o representan únicamente los dos extremos del espectro? Una práctica habitual en la selección de estos individuos es la de escoger tantos sujetos sanos como enfermos, definidos según el criterio de referencia. Sin embargo, esta división a medias no es necesaria.²
3. Los investigadores deben usar la prueba investigada para clasificar a todos los individuos como positivos o negativos. Para las pruebas cuyos resultados se presentan en valores numéricos, es preciso disponer de un intervalo de la normalidad. Por ejemplo, si la mayoría de los individuos con la enfermedad presentan valores por encima del intervalo de la normalidad, los investigadores usan el límite superior del intervalo de la normalidad como límite de demarcación. A continuación, aplican la nueva prueba a todos los individuos y los clasifican como positivos o negativos.
4. Los investigadores ya han clasificado a cada paciente como sano o enfermo, de acuerdo con la prueba de oro, y como positivo o negativo, según el resultado de la prueba. Ahora, ya pueden calcular el número de individuos en los que la prueba estudiada y la de oro concuerdan y en los que discrepan, y presentar los resultados de la siguiente manera:

PRUEBA EN ESTUDIO	PRUEBA DE ORO ENFERMOS	PRUEBA DE ORO SANOS
Positivos	a = Número de individuos enfermos y positivos	b = Número de individuos sanos y positivos
Negativos	c = Número de individuos enfermos y negativos	d = Número de individuos sanos y negativos
	a + c = Total de individuos enfermos	b + d = Total de individuos sanos

² La división a medias proporciona el mayor poder estadístico para un tamaño muestral determinado. Sin embargo, difícilmente veremos aplicar pruebas de significación estadística para valorar pruebas diagnósticas, dado que el tamaño de la muestra generalmente es pequeño y, por esa razón, el poder estadístico suele ser bajo.

5. Finalmente, los investigadores aplican las definiciones de sensibilidad y de especificidad, y calculan directamente sus valores.

$$\text{Sensibilidad} = \frac{a}{a + c} = \begin{array}{l} \text{Proporción de individuos con la enfermedad} \\ \text{según la prueba de oro e identificados} \\ \text{como positivos por la prueba en estudio.} \end{array}$$

$$\text{Especificidad} = \frac{d}{b + d} = \begin{array}{l} \text{Proporción de individuos sanos según la prueba} \\ \text{de oro e identificados como negativos} \\ \text{por la prueba en estudio.} \end{array}$$

Para ilustrar este método numéricamente, imaginemos que se aplica una nueva prueba a 500 individuos que tienen la enfermedad de acuerdo con el criterio de referencia y a 500 individuos que están sanos según el mismo criterio. Podemos construir la tabla de 2×2 como sigue:

PRUEBA EN ESTUDIO	PRUEBA DE ORO ENFERMOS	PRUEBA DE ORO SANOS
	Positivos	a
Negativos	c	d
	500	500

Vamos a suponer que con la nueva prueba 400 de los 500 individuos con la enfermedad son identificados como positivos y que 450 de los 500 individuos sanos son identificados como negativos. Ya estamos en condiciones de rellenar la tabla de 2×2 :

PRUEBA EN ESTUDIO	PRUEBA DE ORO ENFERMOS	PRUEBA DE ORO SANOS
	Positivos	400
Negativos	100	450
	500	500

Ahora se pueden calcular la sensibilidad y la especificidad.

$$\text{Sensibilidad} = \frac{a}{a + c} = \frac{400}{500} = 0,80 = 80\%$$

$$\text{Especificidad} = \frac{d}{b + d} = \frac{450}{500} = 0,90 = 90\%$$

Una sensibilidad de 80% y una especificidad de 90% describen una prueba diagnóstica que, aunque no es ideal, tiene la misma calidad que muchas pruebas que se usan en la medicina clínica para diagnosticar enfermedades.

Observe que la prueba se ha aplicado a un grupo de pacientes, de los cuales 500 tienen la enfermedad y 500 están sanos, según el criterio de referencia. Esta división a medias entre sanos y enfermos es la que se emplea habitualmente al realizar estudios de este tipo. Sin embargo, la sensibilidad y la especificidad habrían sido las mismas independientemente del número de pacientes enfermos y sanos escogidos. Una forma de convencerse de la autenticidad de este importante principio es observar cómo se calculan la sensibilidad y la especificidad, esto es,

$$\text{Sensibilidad} = \frac{a}{a + c} \text{ y especificidad} = \frac{d}{b + d}$$

Observe que a y c —que son necesarios para calcular la sensibilidad— se encuentran en la columna de la izquierda de la tabla. De la misma manera, b y d —que son necesarios para calcular la especificidad— se encuentran en la columna de la derecha de la tabla. De esta forma, el número total de individuos en cada columna no es importante realmente, dado que la sensibilidad y la especificidad se relacionan, respectivamente, solo con la división de los pacientes que se encuentran en una simple columna.

Una vez que se han calculado la sensibilidad y la especificidad, es posible volver atrás y completar la tabla cuando se trabaja con distintos números de individuos enfermos y de sanos definidos según la prueba de oro. Esta vez vamos a suponer que hay 900 individuos sanos y 100 enfermos. En otras palabras, nos encontramos en una situación en la cual 10% de los individuos a los que se aplica la prueba tienen la enfermedad. Por lo tanto, el individuo promedio tiene una probabilidad de 10% de padecer la enfermedad antes de que se realice la prueba.

PRUEBA EN ESTUDIO	PRUEBA DE ORO ENFERMOS	PRUEBA DE ORO SANOS
Positivos	a	b
Negativos	c	d
	100	900

Apliquemos ahora las medidas de la sensibilidad y la especificidad, tal y como hemos hecho previamente.

La sensibilidad es igual a 80%; por lo tanto, 80% de los que tienen la enfermedad serán correctamente identificados como positivos (80% de 100 = 80), y 20% de los que tienen la enfermedad serán incorrectamente identificados como negativos (20% de 100 = 20).

La especificidad es igual a 90%; por consiguiente, 90% de los que no tienen la enfermedad serán correctamente identificados como negativos (90% de 900 = 810), y 10% de los que no tienen la enfermedad serán incorrectamente identificados como positivos (10% de 900 = 90).

Ahora podemos construir la siguiente tabla de 2×2 .

PRUEBA EN ESTUDIO	PREVALENCIA DE 10%	
	PRUEBA DE ORO ENFERMOS	PRUEBA DE ORO SANOS
Positivos	80	90 positivos falsos
Negativos	20 negativos falsos	810
	100	900

En esta situación, 10% de los pacientes estudiados tienen la enfermedad, según la prueba de oro; por lo tanto, podemos afirmar que, en este grupo de pacientes, la verdadera probabilidad de tener la enfermedad es de 10%.

Comparemos esta tabla con la que construimos al calcular por primera vez la sensibilidad y la especificidad. En realidad, utilizamos un grupo de pacientes cuya probabilidad de tener la enfermedad era de 50%, dado que trabajábamos con 500 individuos enfermos y 500 sanos.

PRUEBA EN ESTUDIO	PREVALENCIA DE 50%	
	PRUEBA DE ORO ENFERMOS	PRUEBA DE ORO SANOS
Positivos	400	50 positivos falsos
Negativos	100 negativos falsos	450
	500	500

Observe que con nuestra división inicial a medias (esto es, con una prevalencia de 50%) se identificaron erróneamente 100 individuos como negativos y 50 como positivos. Sin embargo, en el grupo de pacientes en los que la prevalencia de la enfermedad era de 10%, se identificaron incorrectamente 20 individuos como negativos y 90, también erróneamente, como positivos. El cambio en las cifras se debe únicamente a la diferencia de la frecuencia relativa de la enfermedad o prevalencia en los dos grupos de pacientes estudiados (50% *versus* 10%). Observe que en el ejemplo en que se utilizó una prevalencia de 10% había realmente más positivos que estaban sanos (90) que positivos enfermos (80).

Esto puede sorprender, habida cuenta de que la sensibilidad y la especificidad son relativamente altas. Sin embargo, ilustra un principio que debe conocerse para aplicar los conceptos de sensibilidad y especificidad. A pesar de que la sensibilidad y la especificidad no están influidas directamente por la frecuencia relativa o prevalencia de la enfermedad, el número real de individuos que se identifican erróneamente como positivos o como negativos depende de la frecuencia relativa de la enfermedad.

Ahora analizaremos una situación más extrema, en la cual solo 1% de los integrantes del grupo estudiado tienen la enfermedad. Esta situación es la que

aparece típicamente cuando se realiza el tamizaje de un grupo de individuos que están expuestos a factores de riesgo de una enfermedad común, pero que no tienen signos clínicos. La tabla correspondiente podría parecerse a la siguiente:

PRUEBA EN ESTUDIO	PREVALENCIA DE 1%	
	PRUEBA DE ORO ENFERMOS	PRUEBA DE ORO SANOS
Positivos	8	99 positivos falsos
Negativos	2 negativos falsos	891
	10	990

En esta situación hemos utilizado de nuevo la misma prueba que tiene una sensibilidad de 80% y una especificidad de 90%. Esta vez encontramos 8 positivos verdaderos y 99 positivos falsos o, dicho de otra forma, 12 positivos falsos por cada positivo verdadero. Por esta razón, la sensibilidad y la especificidad por sí solas no proporcionan indicación suficiente de la utilidad de un resultado para el diagnóstico de una enfermedad en un individuo concreto. Como clínicos y usuarios de una prueba diagnóstica, necesitamos saber algo más que la sensibilidad y la especificidad de la prueba. Hemos de ser capaces de formular preguntas clínicas tales como ¿cuál es la probabilidad de que haya enfermedad si el resultado de la prueba es positivo?; ¿cuál es la probabilidad de que no haya enfermedad si el resultado es negativo? Antes de que podamos responder a estas preguntas, hemos de preguntarnos ¿cuál es la probabilidad de que el paciente tenga la enfermedad antes de realizar la prueba? Esta *probabilidad anterior a la prueba*, junto con la sensibilidad y la especificidad, nos permite calcular la medida denominada *valor predictivo de la prueba*.

VALOR PREDICTIVO DE LAS PRUEBAS POSITIVAS Y NEGATIVAS

Como hemos comentado anteriormente, la principal ventaja que ofrecen la sensibilidad y la especificidad en la valoración de una prueba es que no dependen directamente de la prevalencia o de la probabilidad de la enfermedad anterior a la prueba. Esta ventaja es especialmente útil para los artículos de la literatura médica. Sin embargo, también tienen limitaciones para responder a dos preguntas importantes desde el punto de vista clínico: si la prueba es positiva, ¿cuál es la probabilidad de que el individuo tenga la enfermedad?; si la prueba es negativa, ¿cuál es la probabilidad de que no la padezca? Estas preguntas tienen una importancia práctica para los clínicos.

Las medidas que responden a estos interrogantes se conocen como *valor predictivo*.

Valor predictivo de una prueba positiva = Proporción de los individuos con una prueba positiva que tienen la enfermedad.

Valor predictivo de una prueba negativa = Proporción de los individuos con una prueba negativa que no tienen la enfermedad.

Los términos *prevalencia* y *valor predictivo* aparecen en los artículos de investigación en relación con grupos de individuos. Por fortuna, en la práctica clínica se utilizan términos equivalentes, aunque el médico trata a un solo paciente a la vez. Desde la perspectiva de la actividad clínica, la *prevalencia* de una enfermedad corresponde a la mejor estimación de la probabilidad de enfermedad antes de realizar la prueba. En términos clínicos, la prevalencia se conoce como probabilidad *anterior a la prueba*. El *valor predictivo* significa lo mismo que la probabilidad de que la enfermedad esté presente (o ausente) después de obtener los resultados de la prueba. Por esta razón, los valores predictivos pueden considerarse clínicamente como la *probabilidad posterior a la prueba*. Si los términos *prevalencia* y *valor predictivo* le parecen confusos, puede sustituirlos por los de *probabilidad de la enfermedad* antes y después de realizar la prueba.

Como orientación práctica y sencilla para interpretar los valores de esas medidas, puede ser útil usar las siguientes aproximaciones de las probabilidades anteriores a la prueba:

- 1% = La probabilidad anterior a la prueba de los que están expuestos a factores de riesgo de una enfermedad común, pero asintomáticos.
- 10% = La probabilidad anterior a la prueba cuando la enfermedad es improbable, pero clínicamente posible y el clínico desea descartarla.
- 50% = La probabilidad anterior a la prueba cuando la incertidumbre es considerable, pero la presentación clínica es compatible con la enfermedad.
- 90% = La probabilidad anterior a la prueba cuando la enfermedad es muy probable clínicamente, pero el clínico desea confirmarla por medio de una prueba diagnóstica.

Mediante las tablas de 2×2 mostraremos cómo se calcula el valor predictivo. Recuerde que lo hacemos para una determinada prevalencia o probabilidad anterior a la prueba.

PRUEBA EN ESTUDIO	PRUEBA DE ORO ENFERMOS	PRUEBA DE ORO SANOS
Positivos	a = Número de individuos enfermos y positivos	b = Número de individuos sanos y positivos
Negativos	c = Número de individuos enfermos y negativos	d = Número de individuos sanos y negativos

a + b = Total de positivos c + d = Total de negativos

Para calcular el valor predictivo de las pruebas negativas y positivas se emplean las siguientes fórmulas.

$$\text{Valor predictivo de una prueba positiva} = \frac{a}{a + b} = \text{Proporción de individuos con una prueba positiva que realmente tienen la enfermedad (medida con la prueba de oro).}$$

$$\text{Valor predictivo de una prueba negativa} = \frac{d}{c + d} = \text{Proporción de individuos con un resultado negativo que realmente no tienen la enfermedad (medida con la prueba de oro).}$$

Ahora, calculemos estos valores empezando con probabilidades anteriores a la prueba de 90%, 50%, 10% y 1%. Recuerde que el número de positivos y de negativos será diferente para cada prevalencia de la enfermedad.

PRUEBA EN ESTUDIO	PROBABILIDAD ANTERIOR DE 90%	
	PRUEBA DE ORO ENFERMOS	PRUEBA DE ORO SANOS
Positivos	720	10
Negativos	180	90
	900	100

Probabilidad anterior a la prueba de 90%

$$\text{Valor predictivo de una prueba positiva} = \frac{a}{a + b} = \frac{720}{730} = 98,6\%$$

$$\text{Valor predictivo de una prueba negativa} = \frac{d}{c + d} = \frac{90}{270} = 33,3\%$$

Empleando el mismo método, los otros valores predictivos son:

Probabilidad anterior a la prueba de 50%

$$\text{Valor predictivo de una prueba positiva} = \frac{a}{a + b} = \frac{400}{450} = 88,9\%$$

$$\text{Valor predictivo de una prueba negativa} = \frac{d}{c + d} = \frac{450}{550} = 81,8\%$$

Probabilidad anterior a la prueba de 10%

$$\text{Valor predictivo de una prueba positiva} = \frac{a}{a + b} = \frac{80}{170} = 47,1\%$$

$$\text{Valor predictivo de una prueba negativa} = \frac{d}{c + d} = \frac{810}{830} = 97,6\%$$

Probabilidad anterior a la prueba de 1%

$$\text{Valor predictivo de una prueba positiva} = \frac{a}{a + b} = \frac{8}{107} = 7,5\%$$

$$\text{Valor predictivo de una prueba negativa} = \frac{d}{c + d} = \frac{2}{891} = 99,8\%$$

Para una prueba con una sensibilidad de 80% y una especificidad de 90%, los datos pueden resumirse de la siguiente forma:

Probabilidad anterior a la prueba	1%	10%	50%	90%
Valor predictivo de una prueba positiva	7,5%	47,1%	88,9%	98,6%
Valor predictivo de una prueba negativa	99,8%	97,6%	81,8%	33,3%

Los cálculos de los valores predictivos tienen importantes implicaciones clínicas. Indican que la probabilidad de que la enfermedad esté presente o ausente después de obtener los resultados de una prueba depende de la mejor estimación posible de la probabilidad de la enfermedad antes de realizar la prueba. Cuando la probabilidad de una enfermedad es moderadamente alta antes de realizar la prueba, por ejemplo de 50%, incluso una prueba negativa, como en el ejemplo utilizado, conduce a una probabilidad de que la enfermedad esté presente de 18,2% ($100\% - 81,8\%$). Cuando la probabilidad de la enfermedad es relativamente baja antes de realizar la prueba, por ejemplo, 10%, incluso una prueba positiva conduce a una probabilidad de que la enfermedad no esté presente de 52,9% ($100\% - 47,1\%$).

La situación empeora cuando la prueba se emplea como instrumento de tamizaje. Por ejemplo, podríamos aplicar la prueba a un grupo de individuos expuestos a un factor de riesgo cuya probabilidad de tener la enfermedad activa es de 1%. Nuestro ejemplo de 1% de prevalencia o probabilidad anterior a la prueba nos enseña que cuando se aplica la prueba que tiene una sensibilidad de 80% y una especificidad de 90% a este grupo de individuos, los que dan resultados positivos tienen una probabilidad de enfermedad de 7,5%. Esto es lo que significa un valor predictivo de 7,5% de una prueba positiva. Si no se entiende el efecto de la probabilidad anterior a la prueba sobre el valor predictivo, se puede cometer el error que describimos a continuación.

Se evaluó una prueba nueva y barata para diagnosticar el cáncer de pulmón aplicándola a un grupo de 100 individuos con cáncer de pulmón y a 100 sin la enfermedad. El valor predictivo de la prueba positiva fue de 85%; es decir, que 85% de los que tuvieron pruebas positivas padecían cáncer de pulmón. Los autores concluyeron que la prueba era adecuada para el tamizaje de ese cáncer en la población general, dado que 85% de los que tuvieran resultados positivos padecerían cáncer de pulmón.

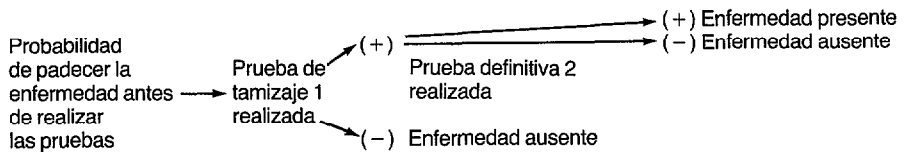
El valor predictivo de una prueba positiva es el porcentaje de personas con un resultado positivo que tienen la enfermedad. Ese valor predictivo depende de la prevalencia de la enfermedad en el grupo de individuos a los que se haya aplicado la prueba. A menudo una prueba se evalúa aplicándola a un grupo de individuos de los que se sabe que la mitad tienen la enfermedad. En este ejemplo, la prueba se aplicó a un grupo en el cual 50% de los individuos padecían la enfermedad (100 con cáncer de pulmón y 100 sin cáncer). Así, la prevalencia o probabilidad anterior de la enfermedad en este grupo era de 50%. Cuanto menor sea la probabilidad anterior de la prueba, menor será el valor predictivo de una prueba positiva.

En la comunidad, la prevalencia de cáncer de pulmón es mucho menor de 50%, incluso entre los fumadores. Por lo tanto, el valor predictivo de una prueba positiva en un individuo promedio de la comunidad —aunque esté expuesto a los factores de riesgo de cáncer de pulmón— será mucho menor de 85%. La capacidad de una prueba positiva para predecir la presencia de enfermedad cambia sustancialmente, según se aplique a grupos de individuos con probabilidades diferentes de presentar la enfermedad. Una prueba positiva puede tener un valor predictivo muy elevado en un grupo de pacientes; no obstante, en otro grupo con una prevalencia o probabilidad anterior distinta de la enfermedad, la misma prueba puede tener un valor predictivo mucho más bajo. La prueba puede ser útil para el diagnóstico en un grupo de pacientes que se sospecha pueden padecer la enfermedad, pero ser inútil para el tamizaje de la población general en la cual la sospecha de enfermedad es baja.

COMBINACIÓN DE PRUEBAS

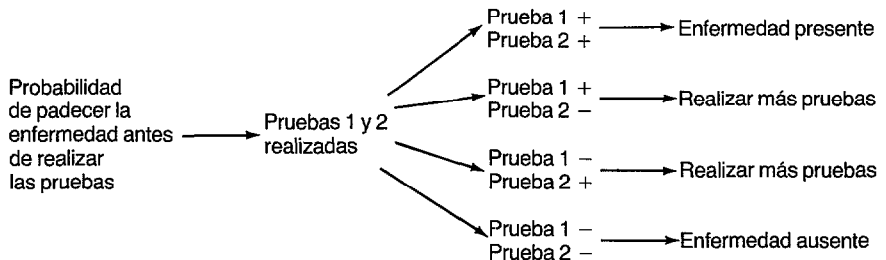
En la práctica clínica y en los artículos de investigación sobre análisis de decisión, que se publican cada vez con mayor frecuencia, los investigadores examinan los efectos de la combinación de pruebas. Hay dos formas básicas de combinar dos pruebas: en serie o en paralelo.

El uso de dos pruebas en serie puede conducir al diagnóstico mediante la siguiente estrategia:³



Al usar las dos pruebas en serie, la prueba número 2 solo se realiza en los individuos que son positivos a la prueba 1. Cuando los resultados de ambas pruebas son positivos, la probabilidad de tener la enfermedad se calcula considerando que el valor predictivo de una prueba positiva adicional es igual a la probabilidad de la enfermedad antes de realizar la prueba 2. Las pruebas en serie, aunque suelen tomar más tiempo, permiten a los médicos descartar la enfermedad empleando menos pruebas.⁴ El valor predictivo de una prueba positiva o negativa no está influido por la prueba que se realiza en primer lugar.⁵ Habitualmente, son razones de seguridad y de costo las que determinan cuál es la prueba que se efectuará primero.

La estrategia de utilizar las pruebas en paralelo exige realizar ambas pruebas al mismo tiempo, y la probabilidad de enfermedad se calcula después de realizadas, tal como se presenta en el siguiente gráfico.⁶



³ Tomado de Riegelman RK y Povar GJ, eds. *Putting prevention into practice*. Boston: Little, Brown; 1988

⁴ También puede ser de ayuda realizar pruebas en serie cuando ambas pruebas tienen una especificidad baja.

⁵ Se presume que las pruebas son independientes. En otras palabras, se parte del supuesto de que la segunda prueba tendrá la misma sensibilidad y especificidad, sea cual fuere el resultado de la primera.

⁶ Véase la nota 3.

Esta estrategia en paralelo funciona bien cuando ninguna de las pruebas tiene una sensibilidad especialmente elevada, pero cada una es capaz de detectar un tipo o estadio distinto de la enfermedad. A veces una prueba puede detectar un estadio temprano, mientras que la otra puede detectar uno más avanzado. En otros casos, una prueba puede detectar una enfermedad de evolución rápida o agresiva mientras que la segunda detecta una de progresión gradual o lenta.

Las estrategias en serie y en paralelo suponen que la segunda prueba proporciona una información adicional superior a la que proporciona la primera. En caso contrario, el rendimiento de ambas estrategias es menor que el esperado. Por ejemplo, imagine el siguiente uso de la estrategia en paralelo.

Se estudiaron la exploración mamaria y la termografía como pruebas en paralelo para detectar el cáncer de mama. Se observó que la exploración tenía una sensibilidad de 40% y la termografía, de 50%. Utilizando las dos conjuntamente, se comprobó que con una u otra prueba solo se detectaban 50% de los cánceres. Los investigadores estaban sorprendidos, dado que habían previsto ser capaces de detectar la mayor parte de los cánceres de mama.

La combinación de la exploración mamaria y la termografía añade poco al uso de cada prueba por separado, ya que ninguna de ellas detecta la enfermedad en una fase temprana. Los resultados de las pruebas nos proporcionan básicamente la misma información y, por este motivo, con una tenemos suficiente. Las pruebas de este tipo son poco útiles cuando se aplican conjuntamente, sea en paralelo o en serie.

Por eso, al diseñar una estrategia de diagnóstico para aplicar las pruebas, necesitamos saber algo más que la sensibilidad y la especificidad; necesitamos saber si las pruebas miden fenómenos diferentes o independientes. También es preciso saber los tipos de enfermedad que pasan por alto las pruebas, por ejemplo, si son incapaces de detectar enfermedades en sus fases iniciales o las de evolución lenta.

RESUMEN: LA PRUEBA DE UNA PRUEBA

El marco utilizado para probar una prueba nos exige evaluarla mediante los conceptos de reproducibilidad y exactitud, y determinar la variabilidad de los resultados en los que no tienen la enfermedad, mediante el concepto del intervalo de la normalidad, y de los que tienen la enfermedad, mediante su medición con la prueba de oro o de referencia. Seguidamente, esta información se combina para valorar la discriminación diagnóstica de la prueba, que se mide en función de su sensibilidad, especificidad y valor predictivo.

La variabilidad de una prueba se mide mediante su reproducibilidad, o sea, repitiendo la prueba en condiciones idénticas. Las repeticiones de la prueba se interpretan sin conocer los resultados originales. La reproducibilidad no garantiza por sí sola la exactitud de la prueba. Una prueba reproducible puede reproducir resultados inexactos cuando existe un sesgo en una dirección. Aunque se puede prever alguna variabilidad en los resultados, esta debe ser bastante menor que la variabilidad biológica medida con el intervalo de la normalidad.

La variabilidad entre individuos sin la enfermedad se mide mediante el intervalo de la normalidad. Este intervalo muchas veces comprende solo 95% de los valores de los individuos considerados sanos. Además, depende del grupo de referencia seleccionado. Recuerde que ese intervalo es meramente una descripción de cómo son las cosas entre los individuos presuntamente sanos. No es diagnóstico: estar fuera de los límites de la normalidad no equivale a estar enfermo, y estar dentro de los límites de la normalidad no equivale a estar sano; un cambio dentro del intervalo de la normalidad puede ser patológico, y estar dentro de los límites de la normalidad no es necesariamente igual a tener los valores más convenientes. Se pueden ajustar los valores que demarcan los límites de la normalidad, alterando de ese modo la especificidad, pero debe tenerse en cuenta el precio que se paga con el número de positivos y negativos falsos. Finalmente, al emplear el concepto del intervalo de la normalidad para definir a un grupo de individuos sanos, es preciso establecer límites inequívocos, con objeto de determinar cuáles son los pacientes positivos y cuáles los negativos a la prueba.

El grupo de sujetos con la enfermedad se define mediante la prueba de oro, que es el mejor método disponible y generalmente aceptado para diagnosticar la enfermedad. Al definir a los enfermos, es conveniente incluir a individuos que tengan el mismo tipo de estado patológico que encontraremos al aplicar la prueba en el medio clínico.

Después de haber comprobado que la prueba es reproducible, que se ha definido lo que es una prueba positiva y una negativa mediante el intervalo de la normalidad y que, mediante el criterio de referencia, se ha identificado a un grupo de individuos enfermos y a uno de sanos, ya se puede valorar la discriminación diagnóstica de una prueba.

Cuando se aplica la prueba en estudio a los individuos identificados como enfermos o como sanos mediante el criterio de referencia, se calculan su sensibilidad y especificidad comparando los resultados de la prueba estudiada con los de la prueba de referencia. Dado que se supone que la prueba de referencia tiene una dis-

criminación diagnóstica perfecta, o de 100%, la prueba en estudio generalmente no estará a la altura de la de referencia. Esto será cierto aunque la prueba estudiada sea intrínsecamente mejor que la de la referencia, ya que cualquier discrepancia se resolverá a favor de la de oro.

La sensibilidad mide la proporción de los individuos que tienen la enfermedad, diagnosticada según el criterio de referencia, que son identificados correctamente como enfermos por la prueba estudiada. La especificidad mide la proporción de los que no tienen la enfermedad, según el criterio de referencia, que son identificados correctamente como sanos por la prueba en estudio. La sensibilidad y la especificidad son importantes porque teóricamente son independientes de la prevalencia o probabilidad anterior a la prueba de que la enfermedad exista en el grupo de individuos estudiados. Esto permite comparar los resultados que obtienen en la prueba distintos grupos de pacientes. Las pruebas se pueden comparar directamente en función de su sensibilidad y especificidad. Sin embargo, es importante reconocer que estas medidas pueden ser distintas en los estadios tempranos de una enfermedad, comparados con los avanzados.

La sensibilidad y la especificidad no responden a la siguiente pregunta clínica: ¿cuál es la probabilidad de tener la enfermedad si la prueba es negativa o si es positiva? El valor predictivo de una prueba positiva nos informa de la corrección con que la prueba confirma la enfermedad en una determinada situación clínica. El valor predictivo de una prueba negativa nos indica la corrección con que la prueba descarta la enfermedad en una determinada situación clínica. Los valores predictivos, a diferencia de la sensibilidad y la especificidad, dependen de la prevalencia o probabilidad de la enfermedad anterior a la prueba.

Desde el punto de vista clínico, esto significa que se debe realizar la mejor estimación posible de la probabilidad de la enfermedad antes de ejecutar la prueba. El valor predictivo informa al clínico sobre la probabilidad de que la enfermedad esté presente después de realizar la prueba. El clínico ha de tener cuidado de no extrapolar el valor predictivo de un contexto clínico a otro. Una prueba muy útil para diagnosticar la enfermedad en presencia de síntomas puede ser prácticamente inútil para el tamizaje de individuos asintomáticos.

Las pruebas pueden combinarse en serie o en paralelo. Cuando se utilizan dos pruebas en serie, la segunda prueba se realiza únicamente si la primera da un resultado positivo. Las pruebas en serie a veces permiten establecer un diagnóstico realizando un menor número de pruebas. Las pruebas en paralelo pueden emplearse cuando ninguna de ellas tiene una sensibilidad suficientemente alta. Esta estrategia funciona bien cuando las dos pruebas detectan diferentes estadios o tipos de la enfermedad, diferenciando, por ejemplo, entre un estadio incipiente y uno avanzado o entre una enfermedad fulminante y una de desarrollo lento. En esta situación, el empleo de las pruebas en paralelo garantiza que se detectará un porcentaje más alto de individuos con la enfermedad. La obtención del máximo beneficio de cualquiera de estas estrategias requiere que los resultados de las dos pruebas midan fenómenos distintos o detecten la enfermedad en etapas diferentes de su evolución.

PREGUNTAS ÚTILES PARA PROBAR UNA PRUEBA

La siguiente lista de preguntas de comprobación ayudará a reforzar los principios necesarios para probar una prueba.

1. Propiedades intrínsecas de una prueba.
 - a. Reproducibilidad: ¿producen resultados prácticamente idénticos las

- repeticiones múltiples de una prueba realizada en las mismas condiciones?
- b. Exactitud: ¿corresponden los resultados de la prueba a los verdaderos valores del fenómeno anatómico, bioquímico o fisiológico?
 - c. Exactitud clínica: ¿proporciona la prueba mediciones similares a las experimentales cuando se realiza en las condiciones reales de la práctica clínica?
2. Variación biológica: el concepto del intervalo de la normalidad y de la variabilidad en los individuos sanos.
 - a. ¿Se ha establecido el intervalo de la normalidad de forma apropiada para que incluya a un porcentaje definido, a menudo 95%, de los individuos considerados sanos?
 - b. ¿Se ha distinguido entre estar fuera del intervalo de la normalidad y estar enfermo?
 - c. ¿Se ha distinguido entre estar dentro del intervalo de la normalidad y estar sano?
 - d. ¿Se puede aplicar de forma generalizada el grupo de referencia utilizado o existen grupos identificables con diferentes intervalos de la normalidad?
 - e. ¿Reconocieron quienes aplicaron la prueba que el intervalo de la normalidad es una descripción del grupo presuntamente sano y que los cambios dentro del mismo para un individuo pueden ser patológicos?
 - f. ¿Se ha distinguido el intervalo de la normalidad del deseable?
 - g. ¿Han justificado los investigadores la modificación del intervalo para cumplir con objetivos diagnósticos específicos?
 3. Variabilidad de los individuos enfermos.
 - a. ¿Han seleccionado los investigadores la mejor prueba de oro disponible para definir a los pacientes que tienen la enfermedad en estudio?
 - b. ¿Han incluido los investigadores a individuos que representen todo el espectro de la enfermedad, para establecer un intervalo realista de los resultados posibles?
 4. Discriminación diagnóstica: distinción entre los enfermos y los sanos.
 - a. ¿Cuán correctamente identifica la prueba a los enfermos? ¿Cuán alta es su sensibilidad? ¿Con qué frecuencia es positiva en los enfermos?
 - b. ¿Cuán correctamente identifica a los que no tienen la enfermedad? ¿Cuán alta es su especificidad? ¿Con qué frecuencia es negativa en los que están sanos?
 - c. ¿Se ha reconocido que, si bien en teoría la sensibilidad y la especificidad no son influidas por la probabilidad de la enfermedad posterior a la prueba, estas medidas pueden ser distintas en las fases tempranas y en las avanzadas de la enfermedad?
 - d. ¿Se ha distinguido entre la sensibilidad y la especificidad de la prueba y su valor predictivo cuando es positiva y cuando es negativa?

PRUEBAS DE LABORATORIO

Examinemos algunas de las pruebas básicas de laboratorio que se utilizan en la medicina clínica para valorar su exactitud, reproducibilidad, intervalo de la normalidad y discriminación diagnóstica.

Hematócrito

Exactitud y reproducibilidad

El hematócrito es una medida del porcentaje de la sangre total compuesta por glóbulos rojos aglomerados. Los hematócritos de rutina se miden mediante la punción en el dedo, con el fin de valorar la sangre en los capilares, o mediante punción venosa. Ambos métodos permiten medir exactamente la cantidad relativa de eritrocitos en la sangre, pero su reproducibilidad depende de que se preste atención a los detalles técnicos. Se puede esperar que la sangre de los capilares tenga un hematócrito entre 1 y 3% más bajo que la venosa. El hecho de apretar excesivamente el dedo puede extraer más plasma y disminuir erróneamente el hematócrito. En presencia de anemia grave, la punción del dedo ofrece valores menos exactos.

Para valorar la exactitud con que el hematócrito mide el estado fisiológico, cabe recordar que se está midiendo la masa relativa, no absoluta, de glóbulos rojos. Los resultados pueden ser erróneos si el volumen plasmático se ha reducido debido a deshidratación o diuresis. Los individuos con un volumen de plasma reducido pueden tener hematócritos por encima del intervalo de la normalidad. Estas son variaciones normales que pueden confundirse con la policitemia (hematócrito patológicamente elevado).

Intervalo de la normalidad

El intervalo de la normalidad de las concentraciones del hematócrito es diferente en los hombres y en las mujeres. Esto es de conocimiento general en los laboratorios y en sus informes los intervalos normales para hombres y mujeres se presentan por separado. Menos a menudo se reconoce que los intervalos de la normalidad son distintos en las diferentes fases del embarazo, en diversas edades y en las personas que viven a distintas altitudes. El hematócrito suele descender durante el embarazo, empezando en algún momento entre el tercer y el quinto mes. Entre el quinto y el octavo no es inusual que se observe una reducción de 20% respecto a los valores anteriores. Sin embargo, por lo general aumenta ligeramente cerca del término del embarazo y vuelve a sus valores normales seis semanas después del parto.

La edad tiene un efecto marcado sobre el hematócrito, especialmente en los niños. El intervalo de la normalidad del primer día de vida es 54 ± 10 (esto es, oscila entre 44 y 64). Al 14o. día, el intervalo es 42 ± 7 , y a los seis meses, $35,5 \pm 5$. El hematócrito promedio aumenta gradualmente hasta la adolescencia, y alcanza un promedio de 39 entre los 11 y los 15 años de edad. El intervalo de la normalidad de los hombres adultos es 47 ± 5 y el de las mujeres, 42 ± 5 .

La presión barométrica baja también tiene un efecto pronunciado sobre el intervalo de la normalidad del hematócrito. Las personas que nacen y viven a grandes alturas tienen en general hematócritos más altos. Por ejemplo, el intervalo de la normalidad a los 1 200 metros es $49,2 \pm 4,5$ en los hombres adultos y $44,5 \pm 4,5$ en las mujeres adultas.

El intervalo de valores de los hematócritos normales es bastante amplio. Por esta razón, si un individuo tiene un valor cercano al límite superior de la normalidad, puede llegar a perder hasta una quinta parte del volumen de eritrocitos antes de que se pueda demostrar la existencia de anemia mediante un hematócrito bajo. Las comparaciones con los hematócritos anteriores son importantes para valorar el desarrollo de la anemia. Individualmente, el hematócrito se mantiene dentro de límites

fisiológicos bastante estrechos, por eso, sus cambios constituyen una medida diagnóstica mejor que la de una sola observación.

Discriminación diagnóstica

Al evaluar el hematócrito, también es necesario conocer la forma en que este responde a la pérdida aguda de sangre. Durante una hemorragia aguda se pierde sangre entera y en un primer momento, la restante se aproximará al hematócrito original. Pueden pasar hasta 24 horas o más antes de que la pérdida se compense mediante un aumento del volumen plasmático. Solamente después de esta compensación puede el hematócrito reflejar totalmente la magnitud de la pérdida de sangre. Si no se reconoce este fenómeno, se pueden obtener resultados negativos falsos en la identificación de los sangrados agudos.

El cuerpo suele ser capaz de compensar la desintegración o pérdida de sangre que ocurre lentamente. Por lo tanto, es posible que en los pacientes con sangrado lento o hemólisis no se detecte anemia, aunque el número de reticulocitos esté elevado. Si mediante la prueba del hematócrito se espera diagnosticar las enfermedades que predisponen a las pérdidas de sangre o a las hemólisis, encontraremos un número relativamente alto de negativos falsos. Enfermedades como la beta-talasemia o los déficit de glucosa-6-fosfato deshidrogenasa (G-6FD) se manifiestan muchas veces como anemias compensadas que cursan con un hematócrito normal o en el límite inferior de la normalidad. Los resultados positivos falsos se pueden producir entre los individuos que tienen el volumen plasmático aumentado como una variante de la normalidad. Desde otro punto de vista, estos individuos reflejan el hecho de que para cualquier intervalo de la normalidad habrá individuos sanos con valores fuera de los límites del intervalo de la normalidad.

Nitrógeno ureico en la sangre y creatinina sérica

Exactitud y reproducibilidad

Para medir las concentraciones de nitrógeno ureico en la sangre (NUS) y la creatinina sérica contamos con pruebas automatizadas bastante reproducibles. Dado que esas concentraciones reflejan la acumulación de nitrógeno ureico y creatinina no excretados, constituyen una indicación de la incapacidad del riñón para eliminar esas sustancias. Cuando se utilizan como medidas de la función renal, sirven para valorar la tasa de filtración glomerular, pero reflejan más exactamente esta tasa cuando hay un deterioro importante de la función renal. Como los riñones suelen tener una capacidad de reserva funcional considerable, puede producirse una pérdida funcional importante, por ejemplo, de un riñón, sin que aumente la acumulación de NUS y creatinina. Estas sustancias reflejarán la tasa de filtración glomerular solo después de una pérdida de 50% o superior de la filtración glomerular. Cuanto más desciende la tasa de filtración, más rápidamente aumentan el NUS y la creatinina. Cuando el NUS y la creatinina están claramente elevados, su incremento porcentual refleja mejor el porcentaje de pérdida de la tasa de filtración glomerular que su cambio numérico.

Cuando la concentración de creatinina se encuentra dentro de los límites normales y se precisa disponer de una valoración exacta de la tasa de filtración glomerular, es posible medir la concentración de creatinina en orina de 24 horas y a la vez la sérica, y calcular entonces el aclaramiento de creatinina. Aunque esta prueba presenta problemas relacionados con su reproducibilidad y con el intervalo de la normali-

dad, es capaz de detectar cambios mucho menores de la tasa de filtración glomerular cuando la concentración sérica de creatinina no está elevada.

Intervalo de la normalidad

Los intervalos de la normalidad de uso estándar en laboratorios varían aproximadamente entre 10 y 20 mg/dl para el NUS y entre 0,6 y 1,4 mg/dl para la creatinina sérica. Para utilizar esos intervalos, es necesario comprender los factores que los afectan en ausencia de enfermedad. El NUS refleja el estado proteico y de hidratación de un paciente. Por eso, el NUS puede cambiar marcadamente sin indicar la presencia de una enfermedad específica. La creatinina es un producto muscular y, como tal, varía ampliamente entre individuos sanos, de acuerdo con su masa muscular. La concentración de creatinina suele ser más baja en las mujeres que en los hombres. En los ancianos, considerados como grupo, la masa muscular y la concentración de creatinina son relativamente menores. A pesar de estas importantes diferencias, los laboratorios habitualmente presentan la concentración de creatinina en referencia a un solo intervalo de la normalidad.

La concentración de creatinina, al contrario que la del NUS, varía menos de día a día y de mes a mes en respuesta a factores no renales. Muchas veces, las comparaciones de las concentraciones de creatinina medidas en distintos momentos pueden proporcionar información muy importante sobre el estado renal. Una concentración de creatinina sérica de 1,3 mg/dl en una mujer anciana puede reflejar una pérdida considerable de la tasa de filtración glomerular, especialmente si ha aumentado en comparación con valores anteriores. La misma concentración en un hombre joven y musculoso puede ser estable y no indicar disminución de la tasa de filtración glomerular.

Discriminación diagnóstica

Como hemos comentado anteriormente, a menudo se obtienen resultados negativos falsos para la enfermedad renal utilizando el NUS y la creatinina, dado que la concentración de estos metabolitos no empieza a aumentar hasta que se ha producido una pérdida sustancial de la filtración glomerular. Además, también se producen resultados positivos falsos tanto con la creatinina como con el NUS. La concentración de NUS puede ser baja en enfermedades que producen malnutrición. Por ejemplo, los alcohólicos con frecuencia tienen concentraciones bajas de NUS. El NUS también refleja la degradación hemática que se produce durante una pérdida rápida de sangre en el tubo digestivo. Estas elevaciones del NUS no son resultados positivos falsos en sentido estricto, pues indican la presencia de otras enfermedades distintas de las renales. Sin embargo, son positivos falsos cuando se está tratando de valorar la filtración glomerular. En presencia de enfermedades musculares se puede detectar una elevación falsa de la concentración de creatinina. En este caso, los resultados tampoco son positivos falsos, pero sugieren que la enfermedad no es renal.

Es posible utilizar las pruebas del NUS y la creatinina en paralelo para entender y localizar anomalías. Normalmente, la razón entre la concentración de NUS y la de creatinina es de 15:1. La elevación desproporcionada del NUS en relación con la creatinina sugiere la presencia de una enfermedad pre o posrenal, más que una enfermedad propiamente renal. La deshidratación puede presentarse como un patrón prerrenal, con elevación desproporcionada del NUS respecto a la creatinina. A veces, alguna enfermedad posrenal, como la obstrucción prostática, también produce una

elevación desproporcionada del NUS respecto a la creatinina. Por este motivo, el uso paralelo o simultáneo de las pruebas del NUS y la creatinina ejemplifica una situación en la cual el uso de dos pruebas ofrecerá más información diagnóstica que el de una sola, dado que miden fenómenos distintos.

Acido úrico

Exactitud y reproducibilidad

La concentración de ácido úrico en la sangre se puede medir de forma reproducible mediante técnicas automatizadas. Para ello existen diversos métodos, cada uno de los cuales proporciona valores ligeramente distintos. Es importante comparar las concentraciones obtenidas con el mismo método. Con las técnicas automatizadas generalmente se obtienen valores algo más altos. La concentración de ácido úrico puede variar en breve plazo; por ejemplo, puede aumentar rápida y significativamente debido a deshidratación.

La concentración de ácido úrico en la sangre mide exactamente la concentración del ácido úrico en el suero. Sin embargo, no valora exactamente todos los parámetros fisiológicos importantes del ácido úrico. No es un indicador exacto del ácido úrico corporal total. Por ejemplo, el ácido úrico cristalizado y depositado no se refleja en la concentración sérica. En relación con el desarrollo de gota, el ácido úrico cristalizado y depositado es frecuentemente el criterio diagnóstico decisivo. Además, la concentración sérica de ácido úrico es solo uno de los factores que influyen en su excreción. Algunos individuos que no tienen una concentración elevada de ácido úrico en el suero pueden excretarlo en grandes cantidades, lo cual los predispone a formar piedras de ácido úrico.

Intervalo de la normalidad y discriminación diagnóstica

El intervalo de valores normales del ácido úrico es bastante amplio y en la mayor parte de los laboratorios varía aproximadamente entre 2 y 8 mg/dl, en función del método empleado. Muchos individuos tienen valores ligeramente por encima de este intervalo y muy pocos por debajo. Son muy pocas las personas con valores ligeramente elevados que desarrollan gota. Esto ha llevado a muchos médicos a afirmar que no se debe tratar a los pacientes con concentraciones ligeramente elevadas de ácido úrico. En realidad, lo que están alegando es que el límite superior de la normalidad se debe aumentar para que la concentración sérica del ácido úrico permita discriminar mejor entre los que están predispuestos a padecer gota y los que no lo están.

Incluso cuando las concentraciones de ácido úrico son elevadas, la discriminación diagnóstica no es buena, porque frecuentemente aparecen resultados positivos falsos. Las personas con insuficiencia renal en muchas ocasiones tienen concentraciones elevadas de ácido úrico, pero raramente padecen gota. Es evidente que existen otros factores, además del ácido úrico, que determinan el desarrollo de esta dolencia. Se ha demostrado que a veces la gota se desarrolla en los individuos predispuestos cuando la concentración de ácido úrico cambia rápidamente. Tanto su reducción como su aumento puede precipitar una crisis de gota. Mucha gente que padece de gota tiene concentraciones séricas de ácido úrico normales durante los episodios de la enfermedad. Estos casos se han documentado mediante la prueba de oro, la demostración de la presencia de cristales birrefringentes en el líquido articular, y no por medio de la concentración de ácido úrico en el suero. Las personas que se sospecha tienen gota a pesar

de que sus concentraciones de ácido úrico están dentro de los límites de la normalidad durante los ataques agudos, deben ser examinadas posteriormente. A menudo, sus concentraciones de ácido úrico habrán aumentado, lo cual demuestra que se obtuvo un resultado negativo falso al inicio del episodio de gota.

A causa de la débil asociación entre la concentración del ácido úrico en el suero y su excreción urinaria, se puede decir que la primera prueba tiene una baja discriminación diagnóstica de las piedras de ácido úrico. El número de resultados positivos y negativos falsos sería muy alto si se empleara la concentración de ácido úrico en el suero como criterio exclusivo de diagnóstico. La demostración de la presencia de piedras es el método de diagnóstico más definitivo. Se ha observado que la excreción elevada de ácido úrico en orina de 24 horas indica predisposición a la formación de piedras de ácido úrico. En este caso la asociación tampoco es exclusiva, porque hay otros factores que influyen en la formación de piedras. El volumen bajo de orina y el pH bajo, especialmente, son factores predisponentes que aumentan la frecuencia de formación de piedras de ácido úrico.

El ácido úrico ilustra diversos principios importantes aplicables a las pruebas diagnósticas.

1. Incluso en presencia de una prueba exacta y reproducible que refleje la anormalidad metabólica relacionada con la enfermedad, el grado de asociación entre la concentración sérica y la presencia de enfermedad puede ser bajo.
2. Es posible que la medición de concentraciones séricas, tanto del ácido úrico como de otras sustancias, no constituya una buena indicación del contenido corporal total y, en particular, de su concentración en las localizaciones patológicas.
3. Puede que la concentración sérica proporcione menos seguridad en el diagnóstico cuando más se necesita, al inicio de los síntomas.