

¹Juan O. Talavera,
²Niels H. Wachter-Rodarte,
³Rodolfo Rivas-Ruiz

¹Centro de Adiestramiento en Investigación Clínica
²Unidad de Investigación en Epidemiología Clínica,
Hospital de Especialidades
³Servicio de Escolares y Adolescentes, Hospital de Pediatría

Autores 1 y 2, Coordinación de Investigación en Salud

Centro Médico Nacional Siglo XXI,
Instituto Mexicano del Seguro Social, Distrito Federal, México

Resumen

Una prueba diagnóstica tiene como fin establecer la presencia de salud o enfermedad, en cuyo caso puede, incluso, graduar el grado de afección. Las pruebas diagnósticas suelen evaluarse matemáticamente. Así, se estiman la sensibilidad y la especificidad una vez conocida la existencia o no de enfermedad; en la práctica clínica suele actuarse en forma inversa: de la positividad o negatividad de una prueba hacia la presencia o no de la enfermedad y, por lo tanto, se utilizan los valores predictivos positivos y negativos. Las estrategias matemáticas permiten cuantificar lo observado, pero se requiere juicio clínico para determinar la calidad de esa observación; en consecuencia se deben considerar algunas características: a) la selección bajo los mismos criterios para casos y testigos; b) la inclusión de todo el espectro de severidad de la enfermedad (procurando que todos los estratos cuente con un número importante de sujetos); c) la interpretación del estándar de oro y de la prueba en estudio debe ser a ciegas y por expertos; d) la interpretación de los resultados debe mostrar la aplicación de prueba en la práctica cotidiana; e) se debe comprobar la reproducibilidad de la prueba. No debe olvidarse que habitualmente se atiende a un paciente a la vez, por lo tanto es indispensable un conocimiento pleno del desempeño de la prueba diagnóstica y considerar la parte clínica para su correcta aplicación.

Palabras clave

investigación
proyectos de investigación
técnicas y procedimientos diagnósticos

Introducción

En la parte 1 de esta serie [Rev Med Inst Seguro Soc 2011; 49(1):53-58] se mencionaron los distintos enfoques desde los cuales se abordan los problemas clínicos: *enfoque arquitectónico*, basado en el fenómeno natural; *enfoque metodológico*, basado en la jerarquía de la información; *enfoque clínico*, basado en los objetivos de la práctica médica. Se desglosaron

Investigación clínica II. Estudios de proceso (prueba diagnóstica)

Comunicación con: Juan O. Talavera.
Correo electrónico: jtalaverap@cis.gob.mx

Summary

A diagnosis test is carried out to establish the presence of health or illness. In the latter it could grade the severity. Due to its importance in clinical decisions, the diagnosis test is evaluated by mathematical strategies. We estimate the sensitivity and specificity once we know the existence or not of the disease, but we act in the reverse direction; with the presence «X» test positive or negative we estimate the presence of the disease, therefore, we use the positive and negative predictive values. Mathematical strategy allow us to quantify the observation, but it requires judgment to determine the quality making use of a minimum of features: a) selection under the same criteria for cases and controls; b) the inclusion of the full spectrum of disease severity (from mild to the most serious, ensuring that all levels have an enough number of subjects); c) the interpretation of both, the gold standard and the new tool of diagnosis, it must be blind and conducted by experts; d) the interpretation of results should show us what is their application in everyday clinical practice; e) the reproducibility must be checked. Do not forget that usually, we treat only one patient at once, what enforce us to have full knowledge of the performance of the diagnostic test, and to consider all clinical aspects for its proper implementation.

Key words

research
research design
diagnostic techniques and procedures

las características principales del enfoque metodológico y, además, se mencionaron los estudios de integración.

Sin embargo, en la práctica clínica, las preguntas suelen estar relacionadas con la necesidad de establecer un diagnóstico o atribuir causalidad, ya sea a través de un estudio pronóstico, de tratamiento o al intentar identificar lo que provocó determinada alteración o enfermedad. Es aquí donde el enfoque arquitectónico se empata con el enfoque por objetivos.

Entre los estudios de proceso, de acuerdo con el enfoque arquitectónico se encuentra la prueba diagnóstica (enfoque por objetivos). Asimismo, dentro de los estudios de causalidad se sitúan los de pronóstico, tratamiento y de factores de riesgo o agente causal (enfoque por objetivos). En el presente artículo describiremos las herramientas más comúnmente utilizadas en las pruebas diagnósticas.

En la práctica clínica, una prueba diagnóstica tiene como objetivo identificar la condición de salud o de enfermedad del sujeto en estudio. Frecuentemente, ante enfermedad permite establecer la severidad del padecimiento, por ejemplo: en un paciente con déficit neurológico súbito la tomografía permite definir el diagnóstico (accidente vascular cerebral isquémico), mientras que si ya se cuenta con el diagnóstico la tomografía hace posible conocer la extensión de la lesión.

El uso de las matemáticas durante el proceso diagnóstico tiene como fin estimar el grado de eficacia y certidumbre de las pruebas en la práctica clínica. A continuación se describen las principales características de toda prueba diagnóstica, tanto con el uso de datos clínicos como de hallazgos de laboratorio y de gabinete.

Características de una prueba diagnóstica

La forma de evaluar la eficacia de una prueba diagnóstica depende del tipo de dato (variable) a utilizar. Por lo tanto, resulta

importante identificar el tipo de variable. Las variables básicas son las que conocemos como *cualitativas de tipo nominal o dicotómicas*, y se refiere a aquellas en las que solo nos percatamos de su presencia o en las que solo existen dos opciones (por ejemplo nacionalidad, presencia o no de enfermedad, hombre o mujer). Las variables *cualitativas ordinales* son aquellas en las que solo se puede identificar qué lugar ocupan en el grupo pero no conocemos de qué tamaño es la diferencia entre unas y otras (por ejemplo: el grado de afección de una enfermedad —leve, moderada o grave—, o la intensidad de un dato clínico identificada con cruces, en cuyo caso aun cuando se reconoce que + es menor que ++ y en consecuencia menor que +++, no puede afirmarse que ++ sea el doble de +). Y, finalmente, las *variables cuantitativas*, aquellas en las que se conoce la distancia entre una intensidad y otra; se denominan *discretas o discontinuas* cuando no pueden fraccionarse (por ejemplo: cuántos hijos tiene una familia [0, 1, 2, 3]), y *continuas* cuando pueden identificarse fracciones entre uno y otro valor (por ejemplo: peso de 52.0 kg, 52.2 kg o 52.250 kg).

La sensibilidad y la especificidad son características propias de toda prueba diagnóstica e indican su eficacia. La *sensibilidad* se refiere a la proporción de individuos enfermos que tienen la prueba positiva. La *especificidad* se refiere a la proporción de individuos no enfermos que tienen la prueba negativa.

El cálculo de sensibilidad y especificidad utiliza *datos nominales o dicotómicos* y se basa en el uso de una tabla de 2 x 2,

		Tomografía computarizada (estándar de oro)		
		+	-	
Rigidez de nuca	+	13 a	10 b	23
	-	9 c	165 d	174
		22	175	197

Figura 1 | Estimación de sensibilidad y especificidad de la rigidez de nuca en el diagnóstico de la hemorragia subaracnoidea

Sensibilidad	$a/a + c = 0.59$ (59 %)	Especificidad	$d/b + d = 0.94$ (94 %)
Falsos positivos	$b/b + d = 0.6$ (6 %)	Falsos negativos	$c/a + c = 0.41$ (41 %)
Valor predictivo positivo	$a/a + b = 0.57$ (57 %)	Valor predictivo negativo	$d/c + d = 0.95$ (95 %)
Prevalencia	$a + c / a + b + c + d = 0.11$ (11 %)	Certeza diagnóstica	$a + d / a + b + c + d = 90$ (90 %)

en la que se contrasta el dato a prueba contra el diagnóstico definitivo obtenido a través de un parámetro ideal llamado *estándar de oro*, que representa la prueba con mayor confiabilidad para demostrar una enfermedad, por ejemplo: el resultado histopatológico (seminoma de testículo), los hallazgos quirúrgicos (colecistitis) o la interpretación de los estudios de imagen (accidente vascular cerebral por tomografía o resonancia magnética) o los estudios de imagen intervencionista (tipo de cardiopatía congénita por cateterismo cardíaco) o de laboratorio (insuficiencia renal por depuración de creatinina).

En la figura 1 se muestra el cálculo de la sensibilidad y la especificidad de la rigidez de nuca para el diagnóstico de hemorragia subaracnoidea, entre pacientes con déficit neurológico de inicio súbito de probable causa vascular. Se observa una sensibilidad de 59 % con una especificidad de 94 %; lo que significa que 59 % de los pacientes con hemorragia subaracnoidea pueden presentar rigidez de nuca y entre los que no tienen hemorragia subaracnoidea 94 % no presenta rigidez de nuca.

Los cálculos de sensibilidad y especificidad se dirigen a partir de la presencia o ausencia de una determinada enfermedad, hacia la probabilidad de presentar o no cierto dato. Sin embargo, en la clínica, el acercamiento suele ser a la inversa: se parte del resultado de la prueba positiva o negativa hacia la probabilidad de tener o no determinada enfermedad. Este tipo de orientación corresponde a lo que conocemos como valores predictivos. El *valor predictivo positivo* representa la probabilidad que tiene un paciente con cierta prueba positiva (signo, síntoma, resultado de laboratorio, de gabinete o algún índice) de cursar con una enfermedad determinada; el *valor predictivo*

negativo representa la probabilidad que tiene un paciente con cierta prueba negativa de estar libre de una enfermedad determinada.

En la figura 1 se observa un valor predictivo positivo de 57 % y un valor predictivo negativo de 95 %; esto significa que entre los pacientes con cuadro clínico de accidente vascular cerebral, un sujeto con rigidez de nuca tiene una probabilidad de 57 % de presentar hemorragia subaracnoidea, mientras que un paciente sin rigidez de nuca tiene una probabilidad de 95 % de no tener hemorragia subaracnoidea.

Mientras los valores de sensibilidad y especificidad se consideran constantes, lo cual no es cierto como mencionaremos más adelante, los valores predictivos se ven afectados por la prevalencia de la enfermedad. Por ejemplo, en la figura 2, donde solo incrementamos la prevalencia de la enfermedad de 11 a 56 %, respetando la proporción de sujetos enfermos con la prueba positiva y con la prueba negativa, se puede ver como la sensibilidad y la especificidad se conservan mientras los valores predictivos se modifican: el valor predictivo positivo es de 93 % y el valor predictivo negativo de 65 %. De tal forma que un incremento en la prevalencia provoca un incremento en el valor predictivo positivo, con una disminución del valor predictivo negativo (una prueba positiva en una población con alta prevalencia de la enfermedad prácticamente hace el diagnóstico, sin embargo, una prueba negativa no lo descarta); al contrario, una disminución de la prevalencia provoca un incremento en el valor predictivo negativo y una disminución del valor predictivo positivo (una prueba negativa en una población con una baja prevalencia de la enfermedad prácticamente descarta la enfermedad).

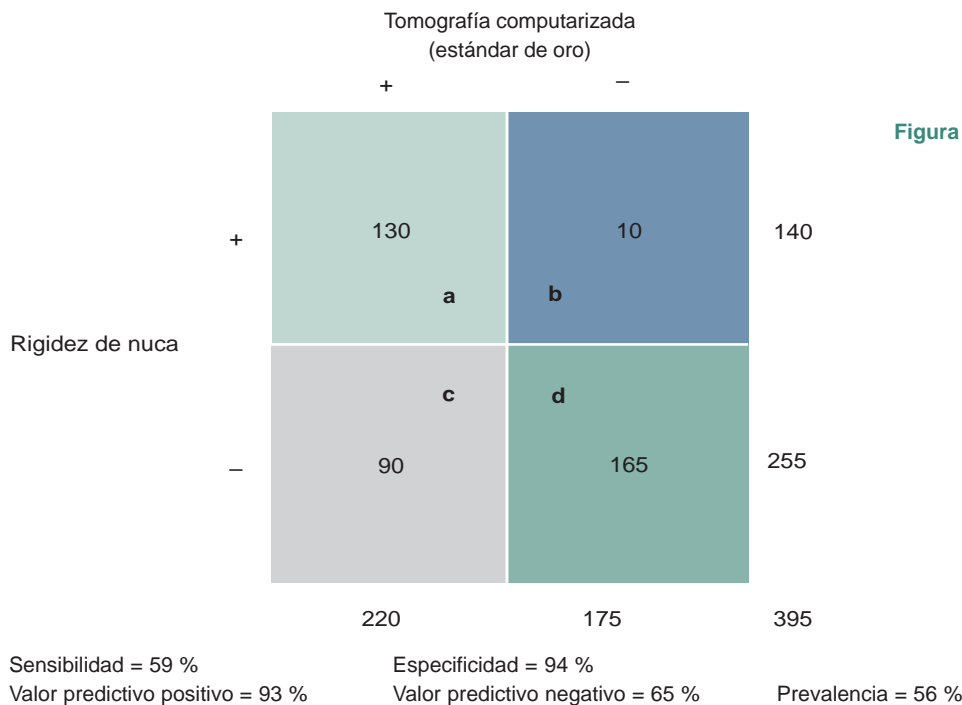


Figura 2 | Modificación de los valores predictivos de la rigidez de nuca en el diagnóstico de hemorragia subaracnoidea con el cambio en la prevalencia

Si la prevalencia de la enfermedad en la población de la cual se obtuvieron los valores predictivos de la prueba diagnóstica es diferente a la prevalencia de la enfermedad en nuestra población, no es posible hacer uso de dichos valores predictivos. Sin embargo, el teorema de Bayes permite estimar los valores predictivos con el uso de la sensibilidad y especificidad de la prueba, y la prevalencia de la entidad en estudio en nuestra población. En el cuadro I se muestra cómo el incremento de la prevalencia de 11 a 56 % provoca un incremento en el valor predictivo positivo de 57 a 94 %. En este ejemplo se observa claramente como una prueba positiva en una población con prevalencia baja (11 %) tiene una probabilidad aproximada de 50 % de tener la enfermedad, mientras que con una prevalencia alta (56 %) prácticamente hace el diagnóstico.

Otra estrategia práctica para estimar la probabilidad de la enfermedad ante una prueba positiva, pero a distintos valores de prevalencia, es el uso del nomograma de Fagan y la razón de verosimilitud (RV, *likelihood ratio*). La RV positiva

(RVP) se obtiene del cociente entre la fracción de verdaderos positivos (sensibilidad) y la fracción de falsos positivo (1-especificidad). Por su parte, la RV negativa (RVN) se obtiene del cociente de especificidad/1-sensibilidad. El nomograma de Fagan se divide en tres partes. En la primera columna está la posibilidad preprueba (prevalencia). En medio se encuentran los valores de la RV y en la última columna la probabilidad posprueba. La probabilidad posprueba para una RVP se refiere a la probabilidad de obtener un resultado positivo cuando la prueba es positiva y corresponde al VPP; la probabilidad posprueba para una RVN se refiere a la probabilidad de obtener un resultado positivo cuando la prueba es negativa; equivalente a 1-VPN. En la figura 3 se encuentran los ejemplos para una prevalencia de 11 y 56 %.

Se dijo que la sensibilidad y especificidad de una prueba no son dependientes de la prevalencia de la enfermedad, sin embargo, los valores varían de acuerdo con el grado de severidad predominante de la enfermedad en el grupo en estudio. Por ejemplo, diagnosticar un cáncer pulmonar en un estadio avanzado con una radiografía de tórax es sencillo y rara vez pasará desapercibido, o sea, rara vez existirán falsos negativos y la sensibilidad será elevada, sin embargo, difícilmente se detectará si intentamos diagnosticarlo en sujetos asintomáticos, en estadio temprano, provocando un alto porcentaje de falsos negativos y una baja sensibilidad. Por lo tanto, considerar que la sensibilidad obtenida de una prueba en una población es aplicable a otra población implica que la distribución de la severidad de la enfermedad es igual en ambos lugares, ya que si en el primero predomina la proporción de sujetos en estadios avanzados, la sensibilidad será alta, y si en el segundo predomina un estadio inicial, la sensibilidad será baja. Contar con los mismos criterios de inclusión entre diferentes estudios de poblaciones distintas no garantiza que la distribución de los sujetos conserve una proporción similar de sujetos en cada estadio de la enfermedad y, consecuentemente, la sensibilidad puede ser distinta.

Cuadro I | Teorema de Bayes

$p(E+/P+) = \frac{p(P+/E+) p(E+)}{p(P+/E+) p(E+) + p(P+/E-) p(E-)}$	
p (E+/P+)	= probabilidad <i>a posteriori</i> de tener cierta enfermedad ante una prueba positiva; corresponde al <i>valor predictivo positivo</i> (VPP).
p (P+/E+)	= probabilidad de resultado positivo de la prueba cuando el paciente padece la enfermedad; corresponde a la <i>sensibilidad</i> .
p (E+)	= probabilidad <i>a priori</i> de tener la enfermedad de acuerdo con la población a la que se pertenece; corresponde a <i>prevalencia</i> .
p (P+/E-)	= probabilidad de un resultado positivo de la prueba cuando el paciente no padece la enfermedad; equivale a los <i>falsos positivo</i> o <i>1-especificidad</i> .
p (E-)	= probabilidad <i>a priori</i> de no tener la enfermedad y corresponde a <i>1-prevalencia</i> [1 - p (E+)].
Prevalencia	11 % 56 %
Sensibilidad	59 % 59 %
Especificidad	94 % 94 %
VPP	57 % 94 %
VPN	95 % 64 %

La estimación del valor predictivo negativo se realiza de la misma forma invirtiendo los signos de la fórmula [por ejemplo: p (E-/P+) cambia a p (E-/P-)]

Uso de datos ordinales y cuantitativos

A diferencia de los datos nominales, cuando la prueba en estudio corresponde a datos ordinales o cuantitativos (con más de un punto de corte), es necesaria la elaboración de una curva ROC (*receiver operator characteristic*), la cual permite determinar en cuál de los puntos de corte se obtiene la certeza diagnóstica más elevada.

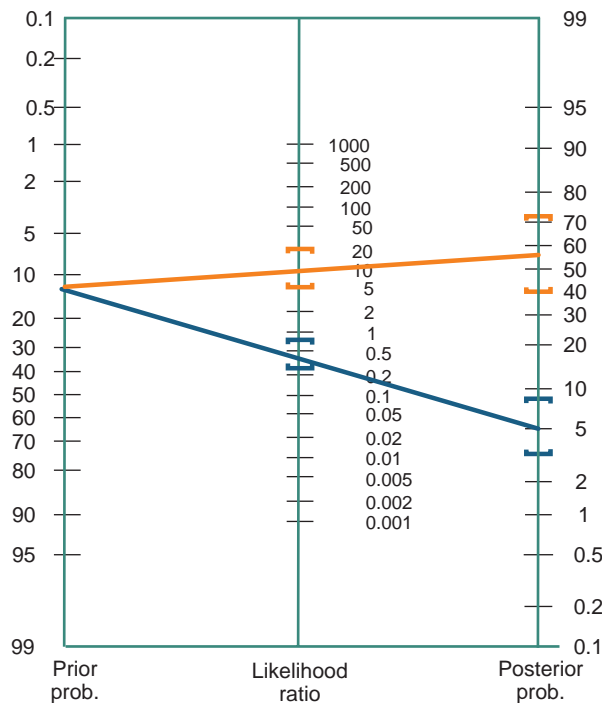
En la figura 4 se observan los diferentes rangos de valores de fosfoquinasa de creatina en líquido cefalorraquídeo expresados en µU/mL, con sus respectivas frecuencias, y se esquematiza el cálculo de sensibilidad y especificidad de acuerdo con los distintos puntos de corte a través de la elaboración de tablas de 2 x 2. En estas tablas se construyen intervalos con los distintos valores de la prueba en estudio y se

tabulan en dos columnas; la primera muestra las frecuencias de sujetos con la enfermedad en cada uno de los intervalos y la segunda muestra la frecuencia de sujetos sin la enfermedad dentro de esos mismos intervalos. Los valores más alterados se sitúan arriba (primeros intervalos) y los menos alterados abajo. Se calcula el porcentaje acumulado hacia arriba y hacia abajo de cada punto de corte, en las dos columnas. En la columna de enfermos a partir del punto de corte hacia arriba se estima la sensibilidad y en la columna de los controles el porcentaje de falsos positivos (1-especificidad).

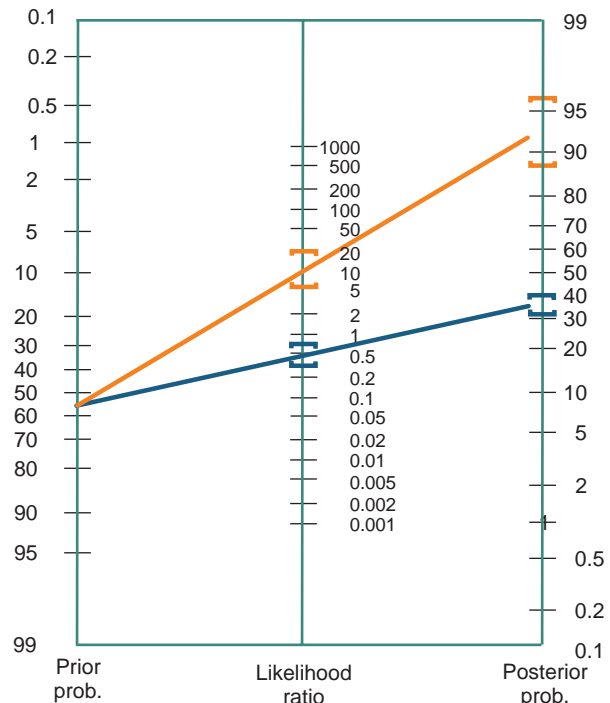
Con los valores de sensibilidad y el porcentaje de falsos positivos se grafican los resultados: los valores de la sensibilidad en el eje de la ordenada (Y), y la proporción de falsos positivos (1-especificidad) en el eje de la abscisa (X); un valor de especificidad de 90 % corresponde a 10 % de falsos positivos (figura 5). El mejor punto de corte corresponde dentro de la curva ROC al punto más cerca del ángulo superior izquierdo de

la curva, o al punto dentro de la tabla que contiene el menor valor de $b + c$ (valores que pertenecen a la suma de falsos positivos y falsos negativos) o el mayor valor de $a + d$ (valores que pertenecen a la suma de verdaderos positivos y verdaderos negativos). En este caso, el punto de corte es de $\geq 16 \mu\text{U/mL}$, el cual permite clasificar correctamente como enfermo o sano a 79.6 % de los pacientes, con sensibilidad de 61.5 % y especificidad de 96.5. Sin embargo, de acuerdo con la utilidad que se le dé a la prueba puede seleccionarse más de un punto: donde se privilegia a la sensibilidad o a la especificidad (mayor valor predictivo negativo o positivo).

Existen casos en que no solo la prueba en estudio contiene más de dos estratos, sino incluso el estándar de oro. En estos casos se puede estimar el porcentaje de aciertos y de errores. En la figura 6 se contrasta el diagnóstico clínico de tromboembolia pulmonar considerando como estándar de oro el diagnóstico por gammagrama ventilatorio perfusorio; el porcentaje de aciertos



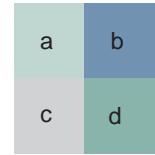
Prevalencia = 0.11 (11 %)
 Sensibilidad = 0.59 (59 %)
 Especificidad = 0.94 (94 %)
 Valor predictivo positivo = 0.57 (57 %)
 Valor predictivo negativo = 0.95 (95 %)
 Razón de verosimilitud positiva = $(a/a + c)/(d/b + d) = 9.83$
 Razón de verosimilitud negativa = $1 - (a/a + c)/(d/b + d) = 0.44$
 Probabilidad posprueba para RVP = 57 %
 Probabilidad posprueba para RVN = 5 %



Prevalencia = 56 %
 Sensibilidad = 59 %
 Especificidad = 94 %
 Valor predictivo positivo = 93 %
 Valor predictivo negativo = 65 %
 Razón de verosimilitud positiva = 9.83
 Razón de verosimilitud negativa = 0.44
 Probabilidad posprueba para RVP = 93 %
 Probabilidad posprueba para RVN = 35 %

Figura 3 | Uso del nomograma de Fagan y razones de verosimilitud

Creatinfosfocinasa (μU/mL)	Daño		≥ 76	≥ 61	≥ 46	≥ 31	≥ 16	≥ 0		
	Orgánico	Metabólico								
≥ 76	2	0	2	0						
61-75	1	0	25	27	3	0				
46-60	2	0		24	27	5	0			
31-45	6	0			22	27	11	0		
16-30	6	1				16	27	17	1	
≤ 15	10	26					10	26	27	27
								0	0	



Sensibilidad (a/a + c)	%	7	11	19	41	63	100
Especificidad (d/b + d)	%	100	100	100	100	96	0
Falsos positivos	%	0	0	0	0	4	100
Errores		25	24	22	16	11	27

Figura 4 | Estimación de sensibilidad y especificidad en distintos puntos de corte para identificar daño orgánico en pacientes con coma

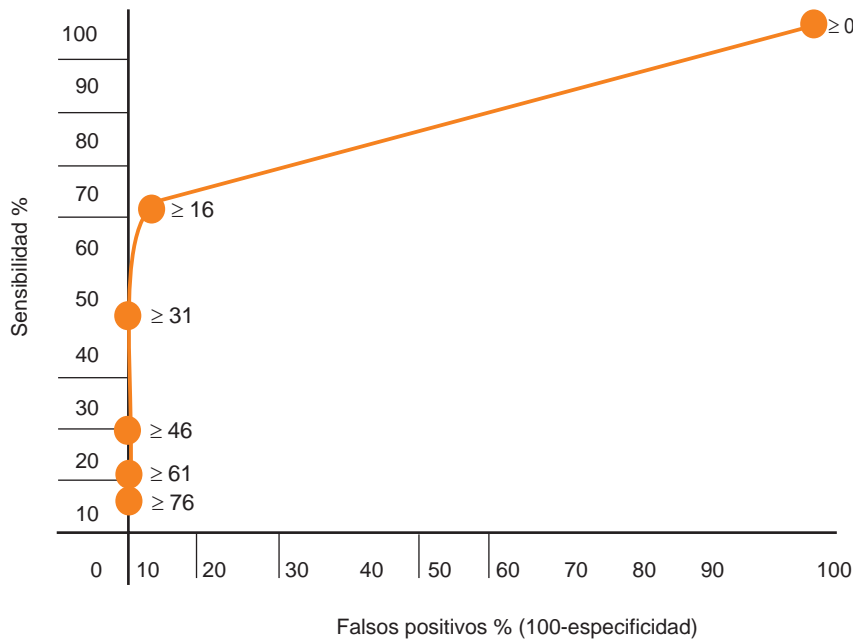


Figura 5 | Curva ROC

		Gammagrama ventilatorio perfusorio (estándar de oro)			
		Alta	Moderada	Baja	
Diagnóstico clínico	Alta	40 a	30 b	20 c	90
	Moderada	7 d	90 e	10 f	107
	Baja	3 g	30 h	70 i	103
		50	150	100	300

Figura 6 Valoración de la eficacia del diagnóstico clínico para identificar tromboembolia pulmonar

Celdas *a, e, i* = aciertos, en este caso de 66.66 %
 Celdas *b, c, f* = errores que sobrestiman el diagnóstico, en este caso de 20 %
 Celdas *d, g, h* = errores que subestiman el diagnóstico, en este caso de 13.33 %

corresponde a las casillas donde coinciden tanto el diagnóstico clínico como el estándar de oro, es decir, en las celdas *a, e, i* (40 + 90 + 70), siendo éste de 66.66 %, y nuestro porcentaje de errores que sobrestiman el diagnóstico corresponde a las celdas *b, c, f* (30 + 20 + 10), siendo éste de 20 %; finalmente, el porcentaje de error que subestima el diagnóstico lo forman las celdas *d, g, h* (7 + 30 + 3), siendo éste de 13.33 %. Sin embargo, existe la posibilidad de que se deseara manejar el desenlace con solo dos posibilidades; en tal caso se podrían agrupar los gammagramas con baja y moderada probabilidad y hablar de un gammagrama con alta probabilidad de embolia pulmonar o sin alta probabilidad, o agrupar los de alta y mediana probabilidad y dejar en un solo grupo los de baja probabilidad. Este mismo procedimiento puede realizarse con la escala clínica, de tal forma que al tener solo cuatro celdas se puedan utilizar los estimadores tradicionales de la utilidad de una prueba diagnóstica, o conservar los tres estratos de nuestra prueba en estudio y calcular una curva ROC.

Aplicaciones de la prueba diagnóstica

Debe quedar claro que la aplicación de una prueba puede tener distintos objetivos:

1. Si se desea una prueba de escrutinio se debe utilizar una prueba con alta sensibilidad, aun cuando tenga baja espe-

2. Si se desea descartar determinada enfermedad, se utiliza una prueba con alta sensibilidad y de ser posible alta especificidad (valor predictivo negativo elevado, por ejemplo: ELISA para VIH), pues si bien cuando es positiva no hace diagnóstico, cuando es negativa sí lo descarta.
3. Si se desea confirmar un diagnóstico en un paciente con sospecha de cursar con una determinada enfermedad, se utiliza una prueba con alta especificidad y de ser posible alta sensibilidad (valor predictivo positivo elevado, por ejemplo: Western-Bloot para VIH), pues aunque no siempre descarta la enfermedad cuando es negativa, si es positiva hace el diagnóstico.

Solicitar exámenes en exceso, justificados o no, genera resultados anormales incluso en personas normales, lo que a su vez desencadena una cascada de exámenes de mayor costo y riesgo, además de angustia para el paciente.

Errores comunes al elaborar una prueba diagnóstica

Ya se mencionó la forma de estimar la eficacia de una prueba diagnóstica y cómo hacer uso de ella, sin embargo, se deben vigilar posibles causas de errores sistemáticos entre los que destacan dos principalmente:

1. La selección inadecuada de pacientes.
2. La interpretación inadecuada tanto de la prueba en estudio como del estándar de oro.

La selección de un espectro inadecuado de pacientes puede suceder desde el punto de vista clínico o patológico. Por ejemplo: no es igual la eficacia del estudio citológico de esputo para detectar cáncer pulmonar en un paciente con tabaquismo intenso y prolongado, pérdida de peso, tos con expectoración hemoptoica y disnea, que en un paciente que solo tiene tos y expectoración blanquecina; tampoco es igual la eficacia de la medición del antígeno carcinoembrionario para detectar cáncer de colon en un paciente con estadio A de Dukes, en comparación a un paciente con estadio C. Es indispensable que toda prueba diagnóstica se realice con la participación de pacientes que incluyan todo el espectro de la enfermedad y que, además, se mencione la proporción de pacientes en cada estrato, de tal manera que se pueda determinar su utilidad en otras poblaciones. Por otro lado, deben considerarse las enfermedades concomitantes y las terapias utilizadas que puedan modificar la eficacia de la prueba en estudio. El grupo control debe haberse seleccionado bajo los mismos criterios que el grupo problema, es decir, bajo la misma puerta de entrada, de tal forma que la comparación tenga significación clínico-metodológica.

En lo que se refiere a los errores más frecuentes durante la elaboración de una prueba diagnóstica es común que al evaluar la prueba en estudio se tenga conocimiento previo del resultado del estándar de oro; esto genera un sesgo en la interpretación debido a que el evaluador espera un cierto resultado. Ocasionalmente, la realización y evaluación de la prueba en estudio precede al estándar de oro e influye sobre la selección de los pacientes que se someten a este último, o sobre su interpretación cuando tiene un componente subjetivo y, no rara vez, la prueba en estu-

dio es parte del estándar de oro con el cual se contrasta. Todas estas desviaciones sobrestiman la utilidad de la prueba.

Estos dos grandes errores pueden evitarse durante la aplicación de una prueba diagnóstica si se consideran los valores de sensibilidad y especificidad solamente cuando:

- a) El espectro de la enfermedad en la población donde se va a aplicar sea igual al espectro de la enfermedad con la cual se elaboró el estudio.
- b) La evaluación de la prueba en estudio y del estándar de oro se haya realizado en forma ciega e independiente en todos los pacientes.

Finalmente, debe hacerse énfasis en que si la evaluación de la calidad de una prueba diagnóstica depende en parte de estrategias matemáticas, es más relevante el juicio clínico del que deriva. Y si bien la estimación de sensibilidad y especificidad se inician a partir de la presencia o no de la enfermedad, en la práctica clínica el estudio del paciente ocurre a partir de la presencia o ausencia del síntoma o signo (clínico o paraclínico).

Además, en todos los casos deberá evaluarse la reproducibilidad de la prueba, siempre y cuando los grupos en estudio sean comparables; esto significa que, además de seleccionar ambas poblaciones bajo los mismos criterios, la distribución de los sujetos dentro de los distintos grados de severidad de la enfermedad debe ser similar. No debe olvidarse que en forma cotidiana se trata a un paciente a la vez y que, por lo tanto, tener un conocimiento pleno de la severidad de la enfermedad en el grupo en estudio es indispensable para su aplicación posterior, de tal forma, que el paciente pueda ser evaluado y tratado de acuerdo con la severidad de su padecimiento y no de acuerdo con el promedio de la severidad de la enfermedad en el grupo en que se evaluó la prueba diagnóstica o el tratamiento.

Bibliografía

1. Altman DG, Bland JM. Diagnostic tests 1: sensitivity and specificity. *BMJ* 1994;308:1552.
2. Altman DG, Bland JM. Diagnostic tests 2: predictive values. *BMJ* 1994;309: 102.
3. Fagan TJ. Nomogram for Bayes's theorem. *N Engl J Med* 1975;293:257.
4. Feinstein AR. Clinical epidemiology. The architecture of clinical research. Philadelphia: W. B. Saunders Company; 1985.
5. Grund B, Sabin C. Analysis of biomarker data: logs, odds ratios, and receiver operating characteristic curves. *Curr Opin HIV AIDS* 2010;5(6):473-479.
6. Jaeschke R, Guyatt G, Lijmer J. Diagnostic tests. En: Guyatt G, Rennie D, editors. *Users' guides to the medical literature*. Chicago: AMA Press; 2002: p. 121-140.
7. Sackett DL, Straus S, Richardson WS, Rosenberg W, Haynes RB. *Evidence-based medicine. How to practice and teach EBM*. Second edition. Edinburgh: Churchill Livingstone; 2000. p. 67-93.
8. Sackett DL, Haynes RB. The architecture of diagnostic research. *BMJ* 2002;324:7336-56.
9. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. *Clinical epidemiology. A basic science for clinical medicine*. Third edition. US: Little Brown; 2009.
10. Soreide K, Korner H, Soreide JA. Diagnostic accuracy and receiver-operating characteristics curve analysis in surgical research and decision making. *Ann Surg* 2011; 253(1):27-34.
11. Tripepi G, Jager KJ, Dekker FW, Zoccali C. Diagnostic methods 2: receiver operating characteristic (ROC) curves. *Kidney Int* 2009;76(3):252-256.