



EDUCACIÓN MÉDICA

INTERPRETACIÓN DEL DESEMPEÑO OPERATIVO DE LAS PRUEBAS DE TAMIZAJE Y DE DIAGNÓSTICO EN ENFERMEDADES EN OBSTETRICIA Y GINECOLOGÍA

Interpreting obstetric and gynaecological disease screening and diagnostic test performance

Hernando Gaitán-Duarte, M.D., MSc., Jorge Andrés Rubio-Romero, M.D., MSc.**,
Margarita Gómez-Chantraine, M.D.****

Recibido: noviembre 3/09 – Aceptado: noviembre 23/09

RESUMEN

Objetivo: explicar tanto los conceptos que subyacen a la evaluación de la validez de una prueba diagnóstica empleada para realizar tamizaje o para confirmar o descartar una enfermedad y aclarar cómo se aplican las características operativas de una prueba diagnóstica para disminuir la incertidumbre acerca de la presencia de la enfermedad para la cual ésta es utilizada.

Metodología: a partir de un caso hipotético, se emplean los estimadores del desempeño de una prueba diagnóstica, los conceptos que subyacen a la evaluación e interpretación de las pruebas de tamizaje y de diagnóstico y los métodos propios para evaluar las características operativas. Finalmente, se concluye con la puesta en práctica de los estimadores de validez para resolver un problema planteado.

Resultados: los conceptos de la historia natural de la enfermedad, la definición de normalidad, las

proporciones y las razones son primordiales para el uso y la interpretación adecuada de las características operativas de una prueba de tamizaje o de diagnóstico. La determinación de la probabilidad preprueba y el conocimiento de la razón de probabilidades positiva y negativa de las pruebas son necesarios para reducir el nivel de incertidumbre en el proceso de diagnóstico.

Conclusión: la interpretación adecuada de las características operativas de las pruebas de tamizaje y de diagnóstico es útil para dar solución de manera explícita a los problemas que se enfrentan en la práctica clínica diaria.

Palabras clave: cribado, diagnóstico precoz, sensibilidad y especificidad.

SUMMARY

Objectives: explaining how diagnostic test performance is assessed, whether used for screening or confirming a certain disease, and explaining how a diagnostic test's operational characteristics are used for reducing uncertainty about the absence or presence of the disease for which it is being used.

Methodology: the main concepts underlying the assessment of the performance, validity, interpretation and the use of screening and diagnostic tests' operational characteristics are addressed within

* Profesor Titular, Departamento de Obstetricia y Ginecología, Grupo de Evaluación de Tecnologías y Políticas en Salud, Instituto de Investigaciones Clínicas Universidad Nacional de Colombia. Bogotá (Colombia). Correo electrónico: hggaitand@unal.edu.co

** Profesor Asociado, Universidad Nacional de Colombia, Grupo de Evaluación de Tecnologías y Políticas en Salud, Instituto de Investigaciones Clínicas Universidad Nacional de Colombia. Bogotá (Colombia).

*** Especialista en Obstetricia y Ginecología. Estudiante de Maestría en Epidemiología Clínica, Universidad Javeriana. Bogotá (Colombia).

a clinical context. A sample clinical scenario is then resolved using the tools provided for the reader.

Results: the natural history of disease, a definition of normality, percentages and ratios are essential concepts for the proper use and interpretation of a screening or a diagnostic test's operational characteristics. Knowledge of pre-test probability and positive and negative likelihood ratios for a test being used is necessary for reducing uncertainty in a given diagnostic process.

Conclusions: interpreting a given diagnostic test's operational characteristics determines its optimal use for screening or accurately ruling out disease and thereby solving the problems we face daily in our clinical practice.

Key words: screening, early diagnosis, sensitivity and specificity.

INTRODUCCIÓN

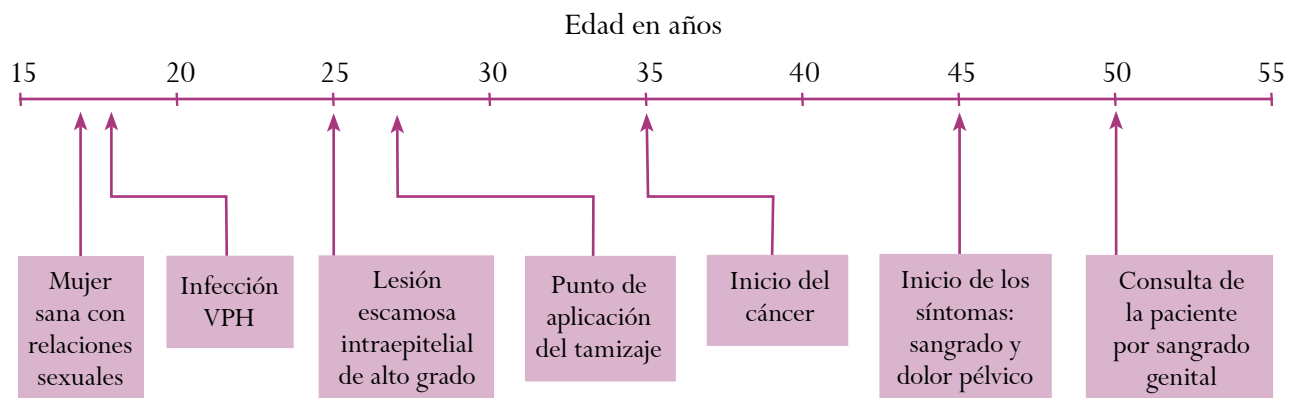
Las pruebas de tamizaje y las pruebas diagnósticas son herramientas muy utilizadas en la aplicación de programas de prevención o de diagnóstico temprano de enfermedades^{1,2} y en la atención individual de las pacientes.

Respecto al primer uso, la aplicación organizada de actividades de diagnóstico temprano y de tratamiento en grandes grupos de poblaciones se conoce como tamizaje poblacional.³ En general, se considera que una prueba de tamizaje es aquella que se utiliza en la fase presintomática de la enfermedad;⁴ sin embargo,

en algunas situaciones, la aplicación de esta prueba es mucho más amplia ya que se utiliza no sólo antes de que la enfermedad sea sintomática sino también antes de que ésta aparezca. Éste es el caso de la citología cervical, con la cual se busca detectar lesiones precursoras del cáncer de cuello uterino (**figura 1**). De este modo, la prueba de tamizaje disminuye tanto la incidencia de cáncer de cuello como la mortalidad asociada a ésta.⁵ Además, es muy importante que al realizar una prueba de tamizaje se tenga en cuenta que no sólo nos debe interesar conocer y darle a conocer al paciente que tiene la enfermedad sino que además debemos ofrecerle algún beneficio real para su salud a partir de este conocimiento.⁴

En cuanto a la práctica clínica, con frecuencia enfrentamos situaciones enmarcadas por un alto grado de incertidumbre que requieren la realización de pruebas orientadas a detectar un estado fisiológico como el embarazo,⁶ confirmar o descartar una enfermedad como el embarazo ectópico,⁷ definir la severidad en una entidad, por ejemplo las cifras de tensión arterial para el diagnóstico de hipertensión severa en preeclampsia,⁸ o establecer un pronóstico como el diagnóstico de retardo del crecimiento intrauterino en términos de malos resultados perinatales.⁹ Por esta razón, el uso de estas pruebas busca reducir el nivel de duda razonable para llegar al umbral en el cual se puede decidir iniciar una conducta terapéutica en una entidad o descartar una enfermedad.¹⁰

Figura 1. Historia natural de un hipotético caso de carcinoma de cuello uterino.



Por consiguiente, este documento pretende brindar los conceptos que permitan al obstetra y al ginecólogo entender cómo se interpreta, desde el punto de vista de la medicina basada en la evidencia, el desempeño operativo de las pruebas diagnósticas a la vez que se hace una aproximación práctica al uso de los estimadores de la exactitud diagnóstica para disminuir la incertidumbre acerca de la presencia o no de una enfermedad.

ESCENARIO

Paciente de 35 años, cursando su tercer embarazo, quien acude a consulta obstétrica de medicina materno-fetal para control prenatal. La mujer ha tenido dos abortos en el primer trimestre, su última menstruación fue hace 10 semanas y trae una prueba de embarazo cualitativa positiva de hace 4 semanas.

Entre los antecedentes médicos se encontró el diagnóstico de Lupus Eritematoso Sistémico (LES) desde hace 3 años, actualmente sin tratamiento. Los resultados del examen físico revelaron un peso de 67 Kg, talla 156 cm, tensión arterial de 125/80 mm/Hg, abdomen blando con útero grávido de 6 cm sobre el pubis y examen genito-urinario normal. La paciente porta una ecografía pélvica que muestra un embarazo de 10 semanas con feto único vivo y una prueba para sífilis positiva (Rapid Plasma Reagin - RPR). A partir de esto, la mujer es remitida con diagnóstico de embarazo de 10 semanas, G3P0A2, lupus eritematoso sistémico y se interroga una posible sífilis gestacional. Ante esto último, la paciente pregunta «¿Doctor es cierto que tengo una infección de transmisión sexual?».

CONCEPTOS BÁSICOS EPIDEMIOLÓGICOS DEL DIAGNÓSTICO DE UNA ENFERMEDAD

Historia natural y espectro de la enfermedad

En un sujeto “sano”, la enfermedad inicia en lo que se denomina el período de latencia,¹⁰ fase preclínica en enfermedades crónicas o etapa de incubación en

enfermedades infecciosas.¹¹ En cualquiera de estos casos, la enfermedad no presenta síntomas. Sin embargo, después de esto inicia el período sintomático en el que se reconocen estadios de la enfermedad desde las formas leves hasta las severas.

En la **figura 1** se observa la historia natural de una hipotética paciente con carcinoma de cuello. La mujer estaría sana hasta el inicio de sus relaciones sexuales, momento en el cual podría adquirir la infección por el Virus del Papiloma Humano (VPH) e iniciar la evolución hacia una lesión intraepitelial de alto grado. Durante este período, y de ser sometida a una citología cervical como prueba de tamizaje, podría detectarse la lesión preneoplásica para ser tratada. De esta manera, sería posible impedir la evolución hacia el cáncer de cuello invasivo, evento que se desarrollaría años más tarde, inicialmente de manera asintomática, y finalmente con la aparición de síntomas como sangrado genital o dolor pélvico, los cuales facilitarían el diagnóstico en un estadio avanzado de la enfermedad.

Lo normal y lo anormal en ciencias de la salud

Es importante introducir el concepto sobre lo que significa o se considera “anormal” en un examen clínico o en un estado de salud. A continuación, se explicará en detalle las diferentes aproximaciones a lo normal – anormal en ciencias de la salud:

1. Un resultado puede ser considerado anormal cuando ocurre de manera infrecuente (“inusual”). Tal es el caso de la desnutrición proteico-calórica en niños, la cual es definida al encontrar un niño cuyo peso se ubica por debajo de 2 desviaciones estándar de la curva normal de peso para su edad.¹² Otra aproximación a dicha infrecuencia es la medición del peso del recién nacido para considerarlo como pequeño para la edad gestacional cuando se encuentra por debajo del percentil 10 o macrosómico cuando se encuentra por encima del percentil 90.¹³
2. La definición de anormalidad también está relacionada con la presencia de un factor que

se considera incrementa el riesgo de malos resultados en la salud. Un ejemplo es el período intergenésico menor a 12 meses, considerado anormal porque aumenta el riesgo de malos resultados maternos y perinatales.¹⁴

3. También se define anomalía a la probabilidad de estar frente a una enfermedad dado que se tiene el resultado positivo de una prueba diagnóstica. Un ejemplo de esto lo constituye la translucencia nucal en las semanas 11 a 14 del embarazo, en la cual un resultado positivo a un punto de corte del percentil 95 indica que la oportunidad de tener un feto con trisomía 21 a no tenerla es de 16:1.¹⁵
4. Por último, se observa la definición de anomalía en la que se considera que ante una situación dada se obtendrá un mejor resultado en salud si se aplica tratamiento que si no se practica ninguno. Esta es la aproximación por la cual se considera anormal la presencia de las lesiones precancerosas del cuello uterino, aunque sólo un porcentaje del 30% al 50% puede evolucionar a lesiones neoplásicas.¹⁶

Probabilidades y razones

Las probabilidades y las razones son de uso frecuente en la aplicación de las pruebas diagnósticas. En el presente artículo se hablará permanentemente de la probabilidad preprueba, la razón de probabilidades, la oportunidad postprueba, entre otros.

Para empezar, la probabilidad es definida como el número de veces que ocurre un evento sobre el número de veces que habría podido ocurrir.¹⁷ Ésta se estima directamente a partir de una proporción.

De su parte, la proporción se define como la relación entre dos números en la que el numerador está incluido en el denominador.

$$(a/a+b) \quad (1)$$

Además de esto, la proporción se puede expresar como porcentaje y corresponde a una proporción $(a/a+b)$ multiplicada por 100. La prevalencia, la

incidencia acumulada y la sensibilidad o la especificidad de una prueba diagnóstica son ejemplos de proporciones.

La oportunidad, *Odds* en inglés, se podría definir como la cantidad de veces que ocurre un evento en relación a las veces que no ocurrió y se estima por medio de la razón.

La razón se define como la relación entre dos números en la cual el numerador no está incluido en el denominador.

$$(a/b) \quad (2)$$

Como ejemplos, se encuentran la razón de mortalidad materna, el Riesgo Relativo (RR) y la razón de probabilidades (Likelihood Ratio - LR).

Ahora bien, una proporción se puede transformar en una razón (si contienen unidades semejantes) de la siguiente manera:

$$\text{Razón} = \text{proporción} / 1 - \text{proporción} \quad (3)$$

Asimismo, una razón se puede transformar en proporción así:

$$\text{Proporción} = \text{razón} / 1 + \text{razón} \quad (4)$$

El riesgo se puede expresar como un riesgo directo o absoluto mediante una proporción; por ejemplo, el riesgo de anomalías cromosómicas en relación con la edad materna. De este modo, el riesgo de tener un hijo con síndrome de Down para una mujer con 12 semanas de gestación a los 20 años de edad es de 1/1.068, a los 35 años es de 1/249 y a los 42 años es de 1/38.¹⁸

EVALUACIÓN DE EXACTITUD DE LAS PRUEBAS DIAGNÓSTICAS

Por exactitud o validez se entiende qué tan bien funciona una prueba al compararla con otra prueba que se considera que tiene la verdad absoluta sobre la existencia o ausencia de enfermedad.

La prueba reina es conocida en la terminología médica anglosajona como el “Gold Standard” o patrón de oro.¹⁹ Es muy importante que cuando se lea un artículo sobre la evaluación de una prueba diagnóstica se determine cuán adecuada es ésta al utilizarla como patrón de oro para la comparación. Por ejemplo, la patología quirúrgica es un patrón bastante adecuado en patología tumoral, sin embargo, el patrón de oro en la enfermedad pélvica inflamatoria, compuesto por cultivos, histopatología y examen visual podría ser menos apropiado.²⁰

En el presente caso, el patrón de oro propuesto está compuesto por el uso de pruebas consecutivas de RPR (*Rapid Plasma Reagin*) y TPPA (*Treponema Pallidum Particle Agglutination Assay*) para confirmar los resultados positivos cuando se sospecha sífilis gestacional.²¹

La estrategia recomendada para determinar la validez de una prueba diagnóstica consiste en compararla a estudio contra el patrón de oro mediante una tabla de contingencia. Esta tabla contiene 4 casillas en las que se compara el resultado positivo o negativo de la prueba bajo evaluación con el resultado del patrón de oro que puede ser positivo

		Patrón de oro (RPR + TPPA)	
		(+)	(-)
Prueba diagnóstica	(+)	VP (a)	FP (b)
	(-)	FN (c)	VN (d)

para la enfermedad que queremos diagnosticar o negativo, ya sea porque el control está sano o tiene otra enfermedad.

Donde:

- **VP (Verdaderos Positivos):** aquellos que la prueba clasificó con la enfermedad y efectivamente tuvieron la enfermedad. Casilla (a).
- **VN (Verdaderos Negativos):** aquellos que la prueba clasificó sin la enfermedad y efectivamente estaban sanos o no tenían esa enfermedad. Casilla (d).

- **FP (Falsos Positivos):** aquellos que la prueba clasificó con la enfermedad y realmente estaban sanos o no tenían esa enfermedad. Casilla (b).
- **FN (Falsos Negativos):** aquellos que la prueba clasificó sin la enfermedad y realmente tenían esa enfermedad. Casilla (c).

Los elementos a determinar en las características operativas de las pruebas diagnósticas son:

Sensibilidad: es la capacidad que tiene la prueba para detectar a los sujetos realmente enfermos. Es decir, determina la probabilidad de que la prueba sea positiva para la enfermedad dado que el sujeto realmente está enfermo.

$$(VP/VP+FN) \text{ o } (a/a+c) \quad (5)$$

Especificidad: es la capacidad de la prueba para detectar los sujetos realmente sanos. Determina la probabilidad de que la prueba sea negativa dado que el sujeto está sano o no tiene la enfermedad a estudio.

$$(VN/VN+FP) \text{ o } (d/b+d) \quad (6)$$

El valor predictivo positivo (VPP): es la probabilidad de tener la enfermedad dado que la prueba dio positiva.

$$(VP/VP+FP) \text{ o } (a/a+b) \quad (7)$$

El valor predictivo negativo: es la probabilidad de estar sano o no tener la enfermedad dado que la prueba dio negativa.

$$(VN/VN+FN) \text{ o } (d/c+d) \quad (8)$$

Prevalencia: es el número de enfermos en un momento en el tiempo sobre el total de la población de estudio.

$$(a+c/a+b+c+d) \quad (9)$$

CÁLCULO DE LAS CARACTERÍSTICAS OPERATIVAS

De vuelta al escenario inicial, se hará un ejercicio que permitirá afianzar los conceptos anteriormente mencionados. Tinajeros y colegas evaluaron la exactitud de la prueba rápida para el diagnóstico de sífilis gestacional al compararla con un patrón de oro compuesto por las pruebas RPR y TPPA en mujeres gestantes en Bolivia.²¹ A continuación se presentan los resultados en una tabla de contingencia.

Patrón de oro (RPR + TPPA)

		(+)	(-)	
Prueba rápida de sífilis (POC)	(+)	VP: 314	FP: 128	442
	(-)	FN: 28	VN: 8.422	8.450
		Enfermos (342)	No enfermos (8.550)	8.892

Con base en las fórmulas antes presentadas se puede calcular:

- La prevalencia de la enfermedad en la población estudiada:
 $(a+c)/(a+b+c+d) = 342/8.892 = 3,8\%$
- La sensibilidad:
 $(a)/(a+c) = 314/342 = 91,8\%$
- La especificidad:
 $(d)/(b+d) = 8.422/8.550 = 98,5\%$
- El valor predictivo positivo:
 $(a/a+b) = 314/442 = 71\%$
- El valor predictivo negativo:
 $(d/c+d) = 8.422/8.450 = 99,6\%$

De igual manera, existen otros dos estimadores de las características operativas de una prueba que son razones entre las proporciones antes descritas. Estos son la razón de probabilidades positiva (LR+) y la razón de probabilidades negativa (LR-).

La razón de probabilidades para prueba positiva (*Likelihood Ratio for Positive Test - LR+*) se define como la probabilidad de que la prueba

sea positiva dado que el sujeto está enfermo en relación con la probabilidad de que la prueba esté positiva dado que el sujeto está sano. Esto equivale a:

$$(a/a+c)/(b/b+d)$$

Si remplazamos por la fórmula (5) tenemos que el numerador corresponde a la sensibilidad y en el denominador tenemos $(b/b+d)$. Ahora bien, según la fórmula (6), $(d/b+d)$ corresponde a la especificidad, por lo tanto $(b/b+d)$ corresponde al complemento de la especificidad (1-especificidad) o tasa de falsos positivos. De este modo:

$$LR (+) = \text{Sensibilidad} / 1 - \text{especificidad} \quad (10)$$

De la misma manera, la razón de probabilidades negativa (*Likelihood Ratio for Negative Test - LR-*) corresponde a la probabilidad de que la prueba sea negativa dado que el sujeto está enfermo en relación con la probabilidad de que la prueba dé negativa dado que el sujeto está sano. Esto equivale a:

$$(c/a+c)/(d/b+d)$$

En el numerador $(c/a+c)$ se encuentra el complemento de $(a/a+c)$. Según la fórmula (5) $(a/a+c)$ corresponde a la sensibilidad; por lo tanto, $(c/a+c)$ corresponde al complemento de la sensibilidad (1 - sensibilidad). Igualmente, en el denominador tenemos $(d/b+d)$, el cual al ser reemplazado por la fórmula (6) corresponde a la especificidad.

$$LR- = 1 - \text{sensibilidad} / \text{especificidad} \quad (11)$$

Ahora bien, al estimar el valor del LR+ y del LR- en el ejemplo anterior de la evaluación de la prueba rápida de sífilis se obtiene lo siguiente:

Según los datos obtenidos previamente, la sensibilidad fue de 91,8% (0,918) y la especificidad alcanzó 98,5% (0,985); por lo tanto, y de acuerdo a la fórmula (10), la $LR+ = 0,918/1 - 0,985 = 0,918/0,015 = 61,2$.

Este resultado se interpreta de la siguiente forma: ante un resultado positivo de la prueba rápida, la oportunidad de que la paciente tenga sífilis gestacional o no es de 62:1. Dicho en términos de apuestas, el médico estaría tan seguro de que la paciente tiene sífilis que estaría dispuesto a apostar 62 pesos por cada peso que otra persona apueste a que no la tiene.

En cuanto a la razón de probabilidades negativa, y de acuerdo con la fórmula (11), $LR- = 1 - 0,918/0,985 = 0,082/0,985 = 0,083$.

Con respecto a lo expuesto, se interpreta que ante un resultado negativo de la prueba rápida, la oportunidad de que la paciente tenga sífilis gestacional a que no la tenga es de 0,08:1. Esto quiere decir que el médico estaría tan seguro de que la paciente no tiene sífilis que estaría dispuesto a apostar sólo 8 centavos de peso por cada persona que apueste que la paciente tiene sífilis.

LA PREVALENCIA Y SU RELACIÓN CON LOS VALORES PREDICTIVOS

La prevalencia, según la definición dada, podría interpretarse como la probabilidad de que una persona esté enferma en un momento dado en una población determinada. Específicamente, ésta podría considerarse como la primera probabilidad de que una persona esté enferma antes de realizar cualquier tipo de prueba diagnóstica o primera probabilidad preprueba.

Además, la prevalencia podrá variar según el sitio donde se atiende a la paciente. De este modo, si el médico trabaja en un hospital de cuidado primario, tendrá pocos sujetos con la enfermedad en relación con una gran cantidad de personas a riesgo. Por el contrario, si éste se encuentra en un hospital de tercer nivel de atención, centro de referencia, es posible que tenga muchos más casos de enfermos con relación a todos los pacientes vistos. Por ejemplo, Aghaizu y colaboradores encontraron una prevalencia de 0,5% de *Neisseria gonorrhoeae* en un estudio de base poblacional en el sur de Londres.²² Por otra parte, Rao G e investigadores informaron

una prevalencia de 4,1% en mujeres jóvenes atendidas en una clínica comunitaria para la atención de infecciones de transmisión sexual ubicada en esa misma ciudad.²³

A continuación se evalúa el efecto de la prevalencia en la estimación de los valores predictivos. Se asume que la prevalencia de sífilis en la presente institución es similar a la provista por Aghaizu para la *Neisseria gonorrhoeae*. De igual modo, se calcula el VPP y el VPN con base en la sensibilidad y la especificidad descritas por Tinajeros para la prueba RPR en una población base de 10.000 mujeres.

Con una prevalencia de sífilis de 0,5% se construiría la siguiente tabla de contingencia

Patrón de oro (RPR + TPPA)

		(+)	(-)	
Prueba rápida de sífilis (POC)	(+)	VP: 46	FP: 150	196
	(-)	FN: 4	VN: 9.800	9.024
		Enfermos (50)	No enfermos (9.950)	10.000

El valor predictivo positivo (VPP) calculado fue de 23,4% y el valor predictivo negativo (VPN) alcanzó el 99,9%.

Ahora bien, a partir de una suposición en la que un médico trabaja en una institución de referencia para ITS, cuya prevalencia de 4% para sífilis es similar a la informada por Rao para *Neisseria gonorrhoeae*, bajo la misma sensibilidad y especificidad y en 10.000 sujetos se construye la siguiente tabla.

Patrón de oro (RPR + TPPA)

		(+)	(-)	
Prueba rápida de sífilis (POC)	(+)	VP: 368	FP: 144	512
	(-)	FN: 32	VN: 9.456	9.488
		Enfermos (400)	No enfermos (9.600)	10.000

A partir de esto, el VPP será de 71,8% % y el VPN será de 99,6%. Esto quiere decir que debido a que el paciente está en centro de referencia, la probabilidad de estar enfermo, dado que se tiene un resultado positivo, es mayor. La información anterior revela la importancia de que el médico conozca la prevalencia de la entidad que diagnostica en su sitio de trabajo, ya que ésto modifica sustancialmente los valores predictivos de las pruebas que utiliza.

Comparación gráfica entre la exactitud de dos pruebas o diferentes puntos de corte: curvas ROC

En algunas ocasiones se brinda la comparación de dos pruebas o varios puntos de corte de una prueba para ver cuál es el que tiene mejores características operativas (sensibilidad y especificidad). Por punto de corte se entiende un nivel dado de una prueba cuantitativa u ordinal que se establece como umbral para delimitar lo normal de lo anormal.²⁴

Para esto, se ha utilizado la presentación gráfica de los resultados mediante la llamada curva operador-receptor o curva ROC. La interpretación gráfica de esta curva se deriva del punto de intersección entre la sensibilidad y la tasa de falsos positivos (1 – especificidad) de la

prueba en cuestión. Como resultado, el punto de corte con mayor sensibilidad y menor tasa de falsos positivos será el de mejor desempeño en la prueba y estará ubicado en el punto más alto y a la izquierda de la gráfica.²⁴ De la misma manera, se puede calcular el área bajo la curva, la cual establece la precisión diagnóstica de la prueba y comparar las diferentes curvas por medio de métodos estadísticos.²⁵

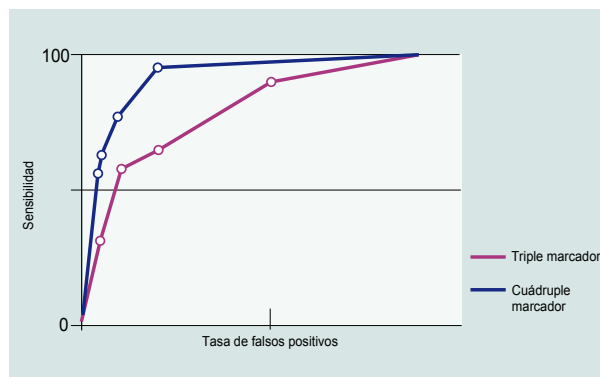
Por ejemplo, Bahado-Singh junto con sus compañeros de trabajo compararon el desempeño operativo del triple marcador clásico: la alfa-feto-proteína, la gonadotropina coriónica humana y el estriol sérico sumado a la edad materna contra la alfa-feto-proteína, la gonadotropina coriónica humana, la translucencia nuchal y la longitud del húmero para el diagnóstico de la trisomía 21.²⁶ En la **tabla 1** se observa la sensibilidad y la tasa de falsos positivos del triple marcador y del cuádruple marcador en diferentes puntos de corte dados por el riesgo según la edad materna.

Con esta información se puede construir una curva ROC que compare las dos pruebas en los diferentes puntos de corte dados por el riesgo de síndrome de Down. La **figura 2** presenta el mejor desempeño de los cuatro marcadores (curva más arriba y a la izquierda con mayor área bajo la curva).

Tabla 1. Sensibilidad y tasa de falsos positivos del triple y cuádruple marcador para la trisomía 21.²⁶

Riesgo de síndrome de Down	Triple marcador		Cuádruple marcador	
	Sensibilidad	Tasa de falsos positivos	Sensibilidad	Tasa de falsos positivos
(1/20)	37%	5%	51%	2%
(1/40)	51%	7%	66%	4%
(1/80)	63%	18%	77%	6%
(1/250)	85%	54%	94%	20%

Figura 2. Curva operador-receptor de las características operativas del triple y cuádruple marcador para el síndrome de Down.²⁵



USO DE LAS RAZONES DE PROBABILIDADES (LR) PARA EL CÁLCULO DE LAS PROBABILIDADES DIAGNÓSTICAS

Concepto de la probabilidad preprueba

Las preguntas que surgen a partir del problema inicial de la paciente con serología RPR (+) y antecedente de lupus eritematoso sistémico son: «¿Estamos ante un caso verdadero de sífilis gestacional?» o «¿un falso positivo de la prueba?». En la presente situación, un resultado positivo de la prueba específica incrementará la probabilidad del diagnóstico de sífilis hasta un umbral en el que iniciar el tratamiento traerá mayores beneficios que riesgos potenciales; o por el contrario, indicará la necesidad de solicitar una prueba confirmatoria de sífilis. Por otra parte, ante el caso de un resultado negativo se pospondrá la decisión hasta hacer nuevas pruebas tales como los anticuerpos antifosfolípidos y la evaluación cuantitativa seriada de la prueba de tamizaje de esta entidad.

Lo primero que se debe hacer para utilizar la información sobre la exactitud de una prueba diagnóstica es estimar la probabilidad basal de que la paciente tenga sífilis o probabilidad preprueba. Las fuentes de información de las probabilidades preprueba son:

1. La prevalencia nacional, regional o local de la entidad.

2. Informes que establecen las probabilidades preprueba en un paciente con un cuadro similar al tratado en el presente artículo.
3. Publicaciones en las que se evalúa la exactitud de pruebas diagnósticas en pacientes con un amplio espectro de la enfermedad, del cual se extrapolan los datos al subgrupo al cual pertenece la paciente de este estudio.
4. La experiencia clínica previa (o a partir de la historia clínica).
5. Bases de datos nacionales que contienen información con situaciones clínicas similares a las de esta investigación y en las que se describen las causas del problema.

A continuación se explicará cada una de ellas:

Cuando la paciente acude a la primera consulta prenatal y antes de iniciar la indagación del motivo de la consulta, la anamnesis, el examen físico y de que se soliciten los exámenes de tamizaje del control prenatal se obtiene la primera probabilidad preprueba de que la paciente tenga sífilis, la cual estaría dada por la prevalencia poblacional de sífilis en esta población. Asimismo, se ha informado que la prevalencia de sífilis en el embarazo en Colombia para el período 2000 - 2003 fue 2,2%.²⁷ Este dato se podría interpretar como que la probabilidad de que una mujer esté enferma de sífilis en el embarazo (sea un caso nuevo o viejo) debido a que vive en Colombia es al menos de 2,2%. Si durante la anamnesis en el primer control prenatal la paciente informa que ha tenido varios compañeros sexuales en el último año o que presenta una úlcera vulvar, se estaría ante una paciente con alto riesgo de adquirir una infección de transmisión sexual (ITS), por lo cual la probabilidad de presentar sífilis en la gestación aumentará. Por lo tanto, lo mejor sería optar por la segunda alternativa, buscar publicaciones en las que se hubiera estimado las probabilidades preprueba diagnósticas en pacientes con alto riesgo de ITS. De este modo, Wong y colegas encontraron una prevalencia de sífilis de 11,7% en 11.473 pacientes de alto riesgo atendidas en una clínica de atención de ITS en China entre 2004 y 2005.²⁸

La tercera alternativa se puede ilustrar por medio de los informes sobre la causas prevalentes de cuadros clínicos similares, especialmente síndromes mal definidos o que pueden tener su origen en múltiples fuentes. Un buen ejemplo de esta situación es el dolor abdominal bajo agudo en mujeres en edad reproductiva, que representa un especial reto diagnóstico debido a consideraciones anatómicas, fisiológicas y al embarazo. Ahondando al respecto, en un estudio con 110 mujeres con dolor abdominal bajo no específico en un hospital de atención materna y perinatal, se encontró que el 54% de los casos se debió a enfermedad pélvica inflamatoria (54%), seguido por complicaciones de quistes en los ovarios (17%) y embarazo ectópico en el 10%.²⁹ Igualmente, Morino y su equipo de trabajo observaron que las causas más frecuentes de dolor abdominal bajo no específico en pacientes atendidas en un servicio de cirugía de un hospital general son la enfermedad pélvica inflamatoria (21%), la apendicitis aguda (18%) y las complicaciones de quistes en los ovarios (12%).³⁰

Por otra parte, la probabilidad preprueba también podría ser estimada a partir de la información provista por una buena historia clínica como aquella realizada por un clínico experimentado, la cual se basa en la evaluación de casos semejantes en el pasado. Esta probabilidad se estima de acuerdo con los síntomas, signos y antecedentes; y en ella se plantean una serie de diagnósticos diferenciales (probabilidades diagnósticas) que orientan la solicitud de exámenes e imágenes de laboratorio.

Cálculo de las probabilidades preprueba y post-prueba

La estimación de las probabilidades de tener una enfermedad requiere la utilización de las fórmulas (3) y (4) con el fin de calcular la oportunidad a partir de una probabilidad y de la probabilidad a partir de la oportunidad. Los pasos que se deben llevar a cabo para estimar las probabilidades después una prueba son los siguientes:

- Paso 1: estimar la probabilidad preprueba (visto previamente).

- Paso 2: convertir a oportunidad preprueba.
- Paso 3: obtener los LR+ y LR-
- Paso 4: multiplicar la oportunidad preprueba por la razón de las probabilidades positiva o negativa para obtener la oportunidad postprueba.
- Paso 5: convertir la oportunidad postprueba en probabilidad postprueba.

En seguida, se emplea el problema del presente estudio para hacer una aproximación a los pasos descritos:

Paso 1. Estimación de la probabilidad preprueba

La primera probabilidad está dada por la prevalencia poblacional de sífilis gestacional en Colombia. La literatura disponible mostró que la prevalencia es de 2,2%. Es decir, la probabilidad preprueba que tenía la paciente de tener sífilis gestacional es de 0,022.

Paso 2. Estimación de la oportunidad preprueba

En este paso se convierte la probabilidad en una oportunidad:

$$\text{Razón (oportunidad)} = (\text{probabilidad} / 1 - \text{probabilidad}) = (0,022)/(1 - 0,022) = (0,022)/(0,978) = 0,022.$$

La oportunidad preprueba = 0,022. En este caso, la oportunidad es muy similar a la probabilidad preprueba. En probabilidades mayores, sin embargo, la probabilidad y la oportunidad muestran mayor diferencia.

Paso 3. Cálculo de la razón de las probabilidades (+) y razón de las probabilidades (-)

Según la estimación previa, la LR+ y LR- de la prueba rápida toman valores de 61,2 y 0,083, respectivamente.

Paso 4. Cálculo de la oportunidad postprueba

Dado que se obtuvo un resultado (+) se procede a calcular la oportunidad postprueba: ésta se obtiene a partir de multiplicar la oportunidad preprueba por la LR+: $0,022 \times 61,2 = 1,34$.

$$\text{Oportunidad postprueba} = 1,34$$

Paso 5. Cálculo de la probabilidad postprueba

En este paso se transforma la oportunidad postprueba a partir de la fórmula. Es importante recordar que oportunidad es sinónimo de razón:

Probabilidad = (oportunidad)/(1 + oportunidad)
 = (1,34)/(1+1,34) = (1,34)/2,34 = 0,57 = 57%.

De esta manera, la probabilidad de que la paciente tenga sífilis gestacional dado que tiene una prueba VRDL+ es de 57%.

Resolución del problema

Con base en las anteriores consideraciones, se le puede responder a la paciente que una explicación del resultado es que se trate de un falso positivo dado que ella tienen como enfermedad de base lupus eritematoso sistémico pero que para excluir o confirmar el diagnóstico deberá practicarse la prueba treponémica específica.

CONSIDERACIONES FINALES

¿Cómo escoger entre una buena sensibilidad o una buena especificidad al evaluar una prueba diagnóstica?

Se prefiere tener pruebas de alta sensibilidad cuando se enfrenta una entidad curable o se puede modificar de manera importante el pronóstico de una enfermedad al hacer la detección temprana, y como prueba de tamizaje cuando además sea fácil de aplicar y tenga bajo costo. En contraste, se opta por una alta especificidad cuando un resultado positivo se acompaña de estigma social o carga emocional por tener mal pronóstico.³¹

Por último, se debe tener en cuenta la existencia de factores que afectan la sensibilidad y especificidad de una prueba diagnóstica:

El espectro de la enfermedad: se define como el estado de la patología de base que podrá variar desde la fase de incubación, fase latente o el estadio sintomático leve, moderado o severo.³⁰

El enmascaramiento o el cegamiento del operador: idealmente, el evaluador de la segunda prueba diagnóstica no debe conocer el resultado de la primera prueba realizada (sesgo de información).³¹ Este conocimiento puede inducir a un sesgo que modifique la validez de una de estas pruebas cuando, por ejemplo, el operador es un lector de imágenes o un estudio de anatomía patológica de especímenes quirúrgicos, en el cual la sospecha por un diagnóstico clínico claro puede aumentar la

posibilidad de lecturas con dicho diagnóstico (sesgo de sospecha diagnóstica).

Asimismo, es importante tener en cuenta que la realización de la prueba patrón de comparación debe ser independiente del resultado de la prueba que se evalúa, ya que si la realización del patrón de oro depende del resultado de una prueba previa (citología y posterior biopsia de cérvix) o de la prueba en evaluación, existe el riesgo de un sesgo de verificación diferencial (biopsia sólo en el grupo de citologías alteradas) que altere la validez de la evaluación de las pruebas diagnósticas.³²

CONCLUSIÓN

La interpretación adecuada de los estimadores de la exactitud de una prueba diagnóstica, utilizada con fines de tamizaje o como prueba confirmatoria, brinda información útil al médico para tomar decisiones de manejo e informar adecuadamente a su paciente sobre la conducta a seguir en situaciones de incertidumbre diagnóstica.

REFERENCIAS

1. Lazcano-Ponce E, Palacio-Mejía LS, Allen-Leigh B, Yunes-Díaz E, Alonso P, Schiavon R, et al. Decreasing cervical cancer mortality in México: effect of papanicolaou coverage, birthrate, and the importance of diagnostic validity of cytology. *Cancer Epidemiol Biomarkers Prev* 2008;17:2808-17.
2. Schopper D, de Wolf C. How effective are breast cancer screening programmes by mammography? Review of the current evidence. *Eur J Cancer* 2009;45:1916-23.
3. Morrison A. Screening. Chapter 25. En: Rothman KJ, Greenland S, editors. *Modern epidemiology*. 2nd ed. Philadelphia: Lippincott Raven; 1998. p. 499-518.
4. Straus S, Richardson WS, Glasziou P, Haynes B. Diagnosis and screening. En: Straus S, Richardson WS, Glasziou P, Haynes B, editors. *Evidence Based Medicine: how to practice and to teach EBM*. 3rd ed. Edinburg: Elsevier; 2005.
5. Macgregor JE, Campbell MK, Mann EM, Swanson KY. Screening for cervical intraepithelial neoplasia in north east Scotland shows fall in incidence and mortality from invasive cancer with concomitant rise in preinvasive disease. *BMJ* 1994;308:1407-11.
6. McRae A, Murray H, Edmonds M. Diagnostic accuracy and clinical utility of emergency department targeted ultrasonography in the evaluation of first-trimester pelvic pain and bleeding: a systematic review. *CJEM* 2009;11:355-64.

7. Condous G, Okaro E, Khalid A, Lu C, Van Huffel S, Timmerman D, et al. The accuracy of transvaginal ultrasonography for the diagnosis of ectopic pregnancy prior to surgery. *Hum Reprod* 2005;20:1404-9.
8. Report of the National High Blood Pressure Education Program Working Group On High Blood Pressure In Pregnancy. *Am J Obstet Gynecol* 2000;183:S1-S22.
9. Garite TJ, Clark R, Thorp JA. Intrauterine growth restriction increases morbidity and mortality among premature neonates. *Am J Obstet Gynecol* 2004;191:481-7.
10. Fletcher R, Fletcher S, Wagner E. Risk. Chapter 5. En: Fletcher R, Fletcher S, Wagner E, editors. *The clinical epidemiology. The Essentials*. 3rd ed. Baltimore: Williams & Wilkins; 1996. p. 94-110.
11. Gordis L. Dinámica de la transmisión de la enfermedad. Capítulo 2. En: Gordis L. *Epidemiología*. 3a ed. Madrid: Elsevier; 2005.
12. Müller O, Krawinkel M. Malnutrition and health in developing countries. *CMAJ* 2005;173:279-86.
13. De Reu PA, Smits LJ, Oosterbaan HP, Nijhuis JG. Value of a single early third trimester fetal biometry for the prediction of birth weight deviations in a low risk population. *J Perinat Med* 2008;36:324-9.
14. Conde-Agudelo A, Belizán JM, Norton MH, Rosas-Bermúdez A. Effect of the interpregnancy interval on perinatal outcomes in Latin America. *Obstet Gynecol* 2005;106:359-66.
15. Snijders RJM, Noble P, Sebire N, Souka A, Nicolaides KH. UK multicentre project on assessment of risk of trisomy 21 by maternal age and fetal nuchal translucency thickness at 10-14 weeks of gestation. Fetal Medicine Foundation First Trimester Screening Group. *Lancet* 1998;351:343-6.
16. Sankaranarayanan R, Gaffikin T, Jacob M, Sellors J, Robles SA. Critical assessment of screening methods for cervical neoplasia. *Int J Gynecol Obstet* 2005;89:S4-S12.
17. Dawson-Saunders B, Trapp RG. Probability & related topics. Chapter 5. En: Dawson-Saunders B, Trapp RG. *Basic & clinical biostatistics*. 2nd ed. Norwalk, Conn.: Appleton & Lange; 1994.
18. Nicolaides KH. Nuchal translucency and other first-trimester sonographic markers of chromosomal abnormalities. *Am J Obstet Gynecol* 2004;191:45-67.
19. Sackett D, Haynes B, Tugwell P, Guyatt G. Selección de pruebas diagnósticas. Capítulo 3. En: Sackett D, Haynes B, Tugwell P, Guyatt G, editores. *Epidemiología clínica: ciencia básica para la medicina clínica*. 2ª ed. Buenos Aires: Editorial Panamericana; 1994. p. 62-78.
20. Gaitán H, Ángel E, Díaz R, Parada A, Sánchez L, Vargas C. Accuracy of five different diagnostic techniques in mild-to-moderate pelvic inflammatory disease. *Infect Dis Obstet Gynecol* 2002;10:171-80.
21. Tinajeros F, Grossman D, Richmond K, Steele M, García SJ, Zegarra L, et al. Diagnostic accuracy of a point-of-care syphilis test when used among pregnant women in Bolivia. *Sex Transm Infect* 2006;82:v17-21.
22. Aghaizu A, Atherton H, Mallinson H, Simms I, Kerry S, Hay PE, et al. Prevalence of *Neisseria gonorrhoeae* infection in young women in South London. *Sex Transm Infect* 2008;84:570.
23. Rao GG, Bacon L, Evans J, Dejahang Y, Michalczyk P, Donaldson N, Lewisham Chlamydia and Gonococcus Screening Programme. Prevalence of *Neisseria gonorrhoeae* infection in young subjects attending community clinics in South London. *Sex Transm Infect* 2008;84:117-21.
24. Sackett D, Haynes B, Tugwell P, Guyatt G. Interpretación de pruebas diagnósticas. Capítulo 4. En: Sackett D, Haynes B, Tugwell P, Guyatt G, editores. *Epidemiología clínica: ciencia básica para la medicina clínica*. 2ª ed. Buenos Aires: Editorial Panamericana; 1994. p. 79-157.
25. Zhang DD, Zhou XH, Freeman DH Jr, Freeman JL. A non-parametric method for the comparison of partial areas under ROC curves and its application to large health care data sets. *Stat Med* 2002;21:701-15.
26. Bahado-Singh RO, Oz AU, Gómez K, Hunter D, Copel J, Baumgarten A, et al. Combined ultrasound biometry, serum markers and age for Down syndrome risk estimation. *Ultrasound Obstet Gynecol* 2000;15:199-204.
27. Organización Panamericana de la Salud. Unidad de VIH/SIDA. Eliminación de sífilis congénita en América Latina y el Caribe: marco de referencia para su implementación. Washington, D.C.; 2004. Visitado en 2008 Ago 13. Disponible en: <http://www.paho.org/Spanish/AD/FCH/AI/eliminasisifilislac.pdf>
28. Wong SP, Yin YP, Gao X, Wei WH, Shi MQ, Huang PY, et al. Risk of syphilis in STI clinic patients: a cross-sectional study of 11,500 cases in Guangxi, China. *Sex Transm Infect* 2007;83:351-6.
29. Gaitán H, Ángel E, Sánchez J, Gómez I, Sánchez L, Agudelo C. Laparoscopic diagnosis of acute lower abdominal pain in women of reproductive age. *Int J Gynaecol Obstet* 2002;76:149-58.
30. Morino M, Pellegrino L, Castagna E, Farinella E, Mao P. Acute nonspecific abdominal pain: a randomized, controlled trial comparing early laparoscopy versus clinical observation. *Ann Surg* 2006;244:881-6.
31. Fletcher R, Fletcher S, Wagner E. Risk. Chapter 3. En: Fletcher R, Fletcher S, Wagner E, editors. *The clinical Epidemiology. The Essentials*. 3rd ed. Baltimore: Williams & Wilkins; 1996. p. 43-74.
32. Leeflang MM, Deeks JJ, Gatsonis C, Bossuyt PM, Cochrane Diagnostic Test Accuracy Working Group. Systematic reviews of diagnostic test accuracy. *Ann Intern Med* 2008;149:889-97.