

Pruebas diagnósticas: uso e interpretación

Diagnostic tests: use and interpretation

FABIÁN JAIMES • MEDELLÍN

Resumen

Las pruebas de diagnóstico son un componente fundamental de la práctica clínica; no solamente para el diagnóstico, como lo indica su nombre, sino también para el manejo diario de los pacientes, para la toma de decisiones relacionadas con el pronóstico, y aun para definir políticas de salud pública en el ámbito de las indicaciones y la utilidad de las pruebas de tamizaje. El ejercicio del diagnóstico no es un fin en sí mismo. Éste obtiene sentido solamente cuando conduce a un curso de acción específico o a una decisión clínica, decisión que no siempre implica intervenciones terapéuticas. Esta revisión muestra en forma práctica los elementos mínimos para el uso y la interpretación de las pruebas diagnósticas. Los principios generales que rigen el uso y la interpretación de las pruebas de diagnóstico son igualmente aplicables a los componentes básicos de la historia clínica como la anamnesis y los hallazgos del examen físico. (*Acta Med Colomb 2007; 32: 29-33*)

Palabras clave: *diagnóstico, sensibilidad, especificidad, razón de probabilidades.*

Abstract

The diagnostic tests are a fundamental component of the clinical practice; not only for the diagnosis but also for the patients' daily care, for making decisions related to the prognosis, and even for defining public health policies in the scope of indications and usefulness of screening tests, making the diagnosis is not a goal by itself. It is only done if it leads to a specific action or a clinical decision, and this decision does not always imply therapeutic interventions. This review shows in a very practical way the basic elements for using and interpreting the diagnostic tests. The general principles that rule the diagnostic tests, their use and interpretation, are equally applicable to the basic components of the clinical history such as anamnesis and findings in the physical examination. (*Acta Med Colomb 2007; 32: 29-33*)

Key words: *diagnosis, sensitivity, specificity, likelihood ratio*

Dr. Fabián Jaimes: MSc., Ph.D. Profesor Asociado y Jefe Departamento de Medicina Interna Universidad de Antioquia. Medellín

Correspondencia. Dr. Fabián Jaimes. Departamento de Medicina Interna Universidad de Antioquia. AA 1226. Medellín, Colombia. Teléfono y fax: 574-263-7947 E-mail: fjaimes@carios.udea.edu.co Recibido: 26/VIII/07 Aceptado: 28/VIII/07

La incertidumbre de la práctica clínica

El médico en ejercicio clínico, es decir, en contacto permanente con pacientes, enfrenta diariamente un sinnúmero de situaciones en las que pone en juego su conocimiento y su razonamiento. Desde que se escucha un motivo de consulta –un síntoma, una queja, una preocupación– se desencadena un proceso por medio del cual se debe discernir continuamente entre diferentes cursos de acción. La presencia de dolor precordial, por ejemplo, genera preguntas que deben precisar la calidad, la intensidad, la irradiación y la duración del dolor; y cada respuesta y también la suma de ellas configuran el cuadro que sugiere una u otra causa para ese dolor. El examen físico que se deriva de ese interrogatorio busca profundizar en la presencia o ausencia de determinados hallazgos, a favor o en contra,

de la causa que se había considerado para el episodio doloroso. Con base en todo ese conjunto de componentes de la historia clínica es que se decide la pertinencia de una prueba de laboratorio, un estudio de imagen o incluso de una exploración anatomopatológica. Cada uno de los anteriores pasos es simplemente un sencillo pero elegante ejercicio de raciocinio médico, en el cual se busca resolver la incertidumbre acerca de cual es la pregunta más relevante del interrogatorio, el signo más fidedigno del examen físico, la prueba de laboratorio de mayor utilidad, y por último cuál es la verdadera razón que llevó a ese paciente a buscar ayuda. De este modo, cada elemento mencionado es una prueba –un estudio o “test”– que debe ser valorada e interpretada de acuerdo con ciertos atributos denominados características operativas.

Las características operativas de una prueba diagnóstica

Los valores informados al proponer o evaluar una prueba –su sensibilidad, su especificidad y sus valores predictivos–, son el fruto de la comparación de esa prueba con otra aceptada o propuesta como “modelo de oro” para el diagnóstico de la enfermedad (Gold Standard del inglés). Esta prueba de referencia usualmente es compleja, costosa, riesgosa o por cualquier razón inadecuada para el proceso usual del diagnóstico. En algunos casos incluso es imposible contar con una prueba de referencia y sólo la evaluación posterior de los individuos potencialmente afectados por la enfermedad, o el criterio clínico de un consenso de expertos, pueden definir con alguna certidumbre la condición final de enfermedad o ausencia de la misma. Dicha comparación entre la prueba en estudio y la prueba de referencia puede visualizarse y explicarse más fácilmente con la construcción de una tabla de 2 x 2 (Tabla 1).

El primer concepto de importancia es la prevalencia de la enfermedad en la población de estudio ($A+C / A+B+C+D$: la proporción de pacientes, del total estudiado, que tiene la enfermedad de acuerdo con la prueba de referencia). Este valor, expresado en términos de probabilidad al igual que todos los que se verán en esta sección, es relevante porque determina el riesgo de base o la probabilidad previa (es decir, antes de cualquier examen) de tener la enfermedad en la población estudiada.

La sensibilidad de la prueba está dada por la probabilidad de que su resultado sea positivo en pacientes con la enfermedad ($A / A+C$: la proporción de enfermos con prueba positiva). Esta característica define el hecho de que una prueba muy sensible –es decir, con una frecuencia alta de verdaderos positivos– deje muy pocos individuos que tienen la enfermedad sin un diagnóstico correcto ($C / A+C$: la proporción de enfermos con prueba negativa o proporción de falsos negativos), y por lo tanto sea muy útil para descartar la enfermedad cuando su resultado es negativo.

La especificidad de la prueba está dada por la probabilidad de que su resultado sea negativo en pacientes sin la enfermedad ($D / B+D$: la proporción de no enfermos con prueba negativa). Esta característica, a su vez, permite que una prueba muy específica –es decir, con una frecuencia alta

de verdaderos negativos– deje muy pocos individuos sin la enfermedad erróneamente diagnosticados como enfermos ($B / B+D$: la proporción de no enfermos con prueba positiva o proporción de falsos positivos), y por tanto, sea muy útil para confirmar la enfermedad cuando su resultado es positivo.

Los anteriores valores (sensibilidad y especificidad) se obtienen de apreciar el comportamiento de la prueba en individuos con y sin la enfermedad en estudio; y son, de hecho, el primer requisito para considerar la decisión de usar o no determinada prueba diagnóstica. Si el interés fundamental es descartar una enfermedad, que es el principio que orienta por ejemplo la utilización de pruebas de tamización en cáncer, es necesario contar con una prueba con la mayor sensibilidad. Por otro lado, si se parte de una presunción clínica fuerte y por lo tanto se quiere confirmar una enfermedad, el uso de una prueba de muy alta especificidad es lo indicado. Sin embargo, en la práctica cotidiana, el médico usualmente se enfrenta a una pregunta diferente: ¿qué información se puede obtener del resultado observado en una prueba? En otras palabras ¿cuál es la probabilidad de enfermedad en individuos con un resultado positivo o negativo de la prueba? Estos interrogantes se pueden resolver con el uso de los valores “predictivos”: el valor “predictivo” de una prueba positiva (VPP) y el valor “predictivo” de una prueba negativa (VPN).

El VPP es la probabilidad de encontrar la enfermedad en individuos con un resultado positivo de la prueba ($A / A+B$: la proporción de individuos con prueba positiva que verdaderamente tienen la enfermedad), y el VPN es la probabilidad de no encontrar la enfermedad en individuos con un resultado negativo de la prueba ($D / C+D$: la proporción de individuos con prueba negativa que verdaderamente no tienen la enfermedad). En la anterior formulación de los valores “predictivos” se observa un elemento muy importante: su denominador involucra, a diferencia de la sensibilidad y la especificidad, tanto a individuos con la enfermedad como a aquellos sin la misma. Por consiguiente, cualquier modificación en la composición de la población de estudio puede afectar los resultados observados o esperados en dichos valores, aun en presencia de la misma sensibilidad y especificidad de la prueba. La causa más obvia de modificación en la composición del denominador de los valores “predictivos” es la diferencia de prevalencia (probabilidad previa) entre diferentes poblaciones. De este modo, poblaciones de estudio con una alta prevalencia de la enfermedad tienden a aumentar el VPP ($A / A+b$: \uparrow) y disminuir el VPN ($d / C+d$: \downarrow); y poblaciones con baja prevalencia tienden a disminuir el VPP ($a / a+B$: \downarrow) y aumentar el VPN ($D / c+D$: \uparrow). La magnitud de ese cambio es mayor en el VPP que en el VPN si las modificaciones en la prevalencia de la enfermedad ocurren cuando ésta es inferior al 50%. Si los cambios en la prevalencia, por el contrario, se presentan siempre en valores superiores al 50%, la modificación en la magnitud de los valores “predictivos” es mayor para el VPN que para el VPP. Esta característica determina el hecho que una vez

Tabla 1. Modelo para la interpretación de una prueba de diagnóstico.

	Enfermedad	Sin enfermedad
Prueba positiva	A Verdaderos Positivos	B Falsos Positivos
Prueba negativa	C Falsos Negativos	D Verdaderos Negativos

considerada de utilidad una prueba diagnóstica, en cuanto a su sensibilidad y especificidad, es necesario estimar además la prevalencia de la enfermedad en el medio en donde se va a utilizar –o la probabilidad previa para cada paciente individual–, con el fin de calcular y ajustar sus valores predictivos para el uso indicado (es decir, confirmar o descartar la enfermedad).

Cómo refinar el uso de una prueba diagnóstica

Como se anotó anteriormente, las características operativas tradicionales de las pruebas diagnósticas –la sensibilidad, la especificidad y los valores “predictivos”– tienen algunas limitaciones para su uso en la práctica clínica diaria. A pesar que la sensibilidad y la especificidad de una prueba son el primer –y obligatorio– criterio para juzgar su utilidad, estos valores no brindan información suficiente para tomar decisiones junto al paciente. Con base en la sensibilidad o en la especificidad de una prueba, consideradas aisladamente, es imposible concluir si un paciente tiene o no una determinada enfermedad. La interpretación del resultado de una prueba, en el contexto del diagnóstico, requiere el uso de los valores “predictivos” y éstos, a su vez, pueden tener valores diferentes en cada escenario clínico de acuerdo con la prevalencia o probabilidad previa de la enfermedad. Adicionalmente, estas medidas se aplican a pruebas diagnósticas con solamente dos resultados: positivo o negativo; y por lo tanto obligan a colapsar, de manera artificial, en dos categorías a aquellas pruebas que se pueden expresar en opciones de mayor riqueza diagnóstica como los valores de ferritina en sangre o el resultado de un estudio histopatológico en una biopsia por aspirado de mama. Como una alternativa para vencer esas limitaciones, en la literatura médica de los últimos años se ha propuesto otra herramienta con algunas ventajas prácticas para la evaluación y utilización de las pruebas diagnósticas: la razón de probabilidades (RP, Likelihood Ratio del inglés).

La RP va más allá de las características operativas tradicionales, las cuales sólo se expresan como probabilidades –o porcentajes, si se multiplica el respectivo valor por 100–, en cuanto se ocupa precisamente de comparar probabilidades. La RP compara, como su nombre lo indica, la probabilidad de encontrar el resultado de la prueba (positivo, negativo, o cualquiera que éste sea) en personas enfermas, con la probabilidad de encontrar ese mismo resultado en personas sin la enfermedad. Cuando el resultado de la prueba en estudio es positivo, este resultado debe ser más frecuente en los enfermos que en los no enfermos; y cuando el resultado de la prueba es negativo, éste debe ser más frecuente en los no enfermos que en los enfermos. De este modo, la RP dice cuántas más veces más –o menos– frecuente es el resultado de una prueba en personas enfermas, comparadas con personas sin la enfermedad.

De nuevo, estos valores pueden apreciarse mejor usando los componentes de la tabla de 2 x 2 que construimos pre-

viamente: $A / A+C$ (la sensibilidad) es la probabilidad de la prueba positiva en individuos con la enfermedad y $B / B+D$ (el complemento de la especificidad: [1-especificidad]) es la probabilidad de la prueba positiva en individuos *sin* la enfermedad. Por lo tanto, $A / A+C \div B / B+D$ (sensibilidad / [1-especificidad]) es la razón de probabilidades de la prueba positiva (RP+ o LR+), y esperamos en ella un valor superior a 1 dado que nos debe decir cuántas veces *más* frecuente es la prueba positiva en individuos enfermos comparados con individuos no enfermos. Del mismo modo, $C / A+C$ (el complemento de la sensibilidad: [1-sensibilidad]) es la probabilidad de la prueba negativa en individuos *con* la enfermedad y $D / B+D$ (la especificidad) es la probabilidad de la prueba negativa en individuos *sin* la enfermedad. Por consiguiente, $C / A+C \div D / B+D$ ([1-sensibilidad] / especificidad) es la razón de probabilidades de la prueba negativa (RP- o LR-), y esperamos en ella un valor inferior a 1 dado que nos debe decir cuántas veces *menos* frecuente es la prueba negativa en individuos enfermos comparados con individuos no enfermos. Teóricamente, por sus propiedades matemáticas, una razón de probabilidades podría ir desde cero hasta infinito. Lo anterior, no obstante, implica una prueba “perfecta” de sensibilidad 100% y especificidad 100%. En un escenario más factible, aunque igualmente idealizado, una prueba con sensibilidad y especificidad del 99% tendría una razón de probabilidades para la prueba positiva (RP+) de $99 / (0.99 / [1-0.99])$, y una razón de probabilidades para la prueba negativa (RP-) de $0.01 / ([1-0.99] / 0.99)$. Esto implica, como se explicó anteriormente, que es 99 veces más probable encontrar esa prueba positiva en individuos enfermos que en los no enfermos; y es 100 veces menos frecuente (1/0.01) encontrar esa prueba negativa en individuos enfermos comparados con aquellos que no tienen la enfermedad.

Si bien, la razón de probabilidades de una prueba diagnóstica se puede “transportar” entre diferentes escenarios de probabilidad de la enfermedad –es decir, el mismo valor de RP+ o RP- de la prueba se puede aplicar sin modificaciones en situaciones con diferente prevalencia–, esto no significa que la interpretación y el uso clínico de los valores de RP ignoren las características clínicas que le confieren a cada paciente la probabilidad de tener dicha enfermedad. La utilidad final de una prueba no puede aislarse del contexto de esa probabilidad previa. Las situaciones clínicas de extrema prevalencia, alta (mayor del 90%) o baja (menor del 10%), usualmente no se modifican por el resultado de ninguna prueba, aunque ésta sea del mejor rendimiento. Las pruebas son de mayor utilidad en el área de mayor incertidumbre diagnóstica, cuando la probabilidad previa de la enfermedad está alrededor del 50%, y son útiles precisamente para llevar esa probabilidad previa a la prueba a una probabilidad posterior que permita tomar decisiones clínicas con respecto a la presencia o ausencia de enfermedad. Este traslado de la probabilidad previa a la probabilidad posterior por medio de la RP de la prueba requiere de una corta operación algebraica dada por el teorema de Bayes: (Probabilidad previa / [1 – Probabilidad

previa) x RP = (Probabilidad posterior / [1 - Probabilidad posterior]). El mismo cálculo anterior puede realizarse de una manera mucho más sencilla, afortunadamente, con el uso del siguiente nomograma, el cual se encuentra disponible ampliamente en numerosos textos de epidemiología clínica, en revistas médicas o en Internet (Figura 1)

La línea de la izquierda representa diversos valores de probabilidad previa a la prueba, la línea del medio corresponde a las diferentes RP (positivas o negativas) que podrían encontrarse en una prueba, y la línea de la derecha muestra las probabilidades de tener la enfermedad después del uso de la prueba. El procedimiento consiste sencillamente en trazar una línea recta entre la probabilidad previa de cada caso y el RP de la prueba que se está utilizando, y la continuación de esa línea recta hacia la derecha se cruzará con el valor correspondiente a la probabilidad de tener la enfermedad posterior al uso de la prueba. En este nomograma se puede apreciar, igualmente, que los cambios más significativos en la probabilidad de la enfermedad ocurren con pruebas que tienen RP mayores de 10 o menores de 0.1. Las pruebas con RP+ mayor de 10 y las pruebas con RP- menor de 0.1 usualmente son muy útiles para confirmar o descartar, respectivamente, una enfermedad.

Un ejemplo de la literatura reciente puede ilustrar en términos prácticos todos los anteriores conceptos. Gibon y colaboradores (*Annals of Internal Medicine* 2004; 141:9-15), han propuesto el uso de los valores plasmáticos de la molécula TREM-1 (Triggering Receptor Expressed on Mye-

loid Cells-1) como un método diagnóstico para diferenciar la sepsis de las inflamaciones sistémicas severas no infecciosas en pacientes admitidos a unidad de cuidado intensivo (UCI) con sospecha de infección. Los autores estudiaron 76 pacientes críticamente enfermos con dos o más criterios de síndrome de respuesta inflamatoria sistémica (SRIS) y con sospecha de infección, de los cuales 47 (62%) finalmente tuvieron sepsis o choque séptico (estos diagnósticos fueron determinados por dos intensivistas que de manera independiente y en forma retrospectiva, cegados en cuanto a los valores de TREM-1, revisaron todos los datos pertinentes para cada paciente). Con un punto de corte de 60 ng/mL, se encontró que la sensibilidad de la prueba fue del 96%, la especificidad de 89%, la RP+ de 8,6 y la RP- de 0.04. Es decir, que en el escenario clínico de un paciente admitido a UCI con sospecha de infección, es 8.6 veces más probable encontrar un valor de TREM-1 superior a 60 ng/mL en pacientes con sepsis comparados con aquellos que no la tienen. Del mismo modo, es 0.04 veces menos probable (o 25 veces menos frecuente) encontrar valores de TREM-1 inferiores a 60 ng/mL en pacientes con sepsis comparados con aquellos que no la tienen. Si se utilizan el nomograma que se explicó anteriormente y la misma prevalencia o probabilidad previa del estudio (62%), podemos notar que un paciente de esas características con resultado de TREM-1 superior a 60 ng/mL tiene una probabilidad del 93% de tener sepsis, y un paciente con un valor de TREM-1 inferior a ese mismo punto de corte tiene una probabilidad de tener sepsis de apenas el 6%. En el caso hipotético de usar esa misma prueba en otro escenario clínico, por ejemplo en un paciente atendido en el servicio de urgencias con sólo un criterio de SRIS y en quien asumimos una probabilidad de sepsis del 20%, los resultados obtenidos son algo diferentes: mientras un resultado positivo de la prueba lleva esa probabilidad previa apenas a 68% –lo que algunos podrían considerar todavía dentro de un rango de incertidumbre– un resultado negativo (es decir, un valor de TREM-1 inferior a 60 ng/mL) convierte la probabilidad de sepsis en un valor alrededor del 1%, lo cual de hecho prácticamente descarta la enfermedad.

El arte del diagnóstico clínico sigue apoyándose, fundamentalmente, en los pilares de la historia clínica y el examen físico. Este arte, sin embargo, puede mejorarse sustancialmente para la práctica cotidiana con el uso y la interpretación apropiados de los componentes básicos de las pruebas de diagnóstico.

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. Grimes DA, Schulz Kf. Refining clinical diagnosis with likelihood ratios. *Lancet* 2005; 365: 1500-5
4. Fagan TJ. Letter: Nomogram for Bayes theorem. *N Engl J Med* 1975; 293: 257
5. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative. *Ann Intern Med* 2003; 138: 40-4.

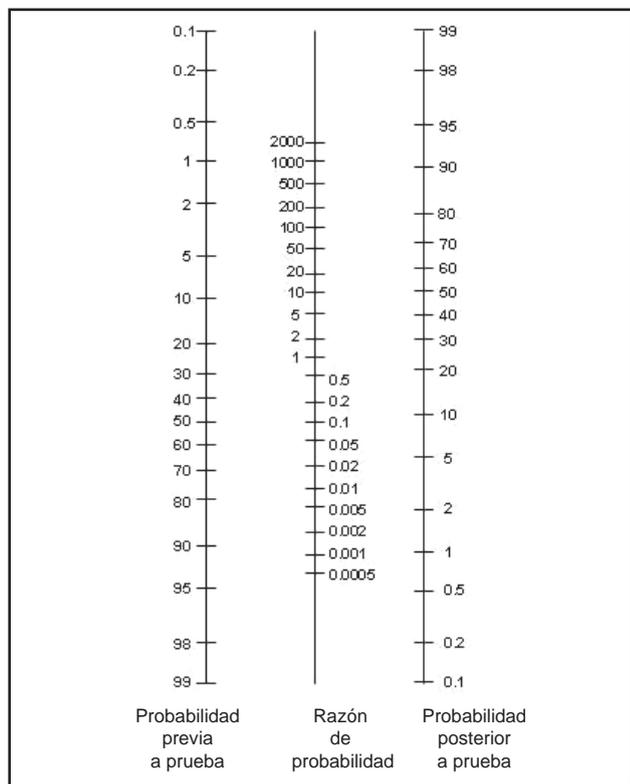


Figura 1. Nomograma para la interpretación de las razones de probabilidades (LR)

AUTOEVALUACIÓN

Una mujer de 50 años llega a la consulta con el resultado de un examen de ultrasonido de la muñeca que fue realizado gratuitamente en un centro comercial de la ciudad. Las personas que realizaron el examen le explicaron que éste tenía una sensibilidad del 98% y una especificidad del 90% para el diagnóstico de osteoporosis, y dado que el resultado fue positivo ella debía buscar tratamiento médico para su enfermedad.

1. La mujer desea que usted revise e interprete el examen y le confirme si tiene o no osteoporosis. Su respuesta es:

- A. La probabilidad de tener osteoporosis es del 98%
- B. La probabilidad de tener osteoporosis es del 90%
- C. La probabilidad de tener osteoporosis es del 88% (0.98×0.90)
- D. No es posible saberlo con la información disponible

2. Con las características operativas del examen de ultrasonido mencionado, éste parecería ser más útil en una de las siguientes situaciones:

- A. Descartar el diagnóstico de osteoporosis si el resultado del examen es negativo
- B. Confirmar el diagnóstico de osteoporosis si el resultado del examen es positivo
- C. Descartar el diagnóstico de osteoporosis si el resultado del examen es positivo
- D. Confirmar el diagnóstico de osteoporosis si el resultado del examen es negativo

3. Con la información disponible respecto al examen es posible determinar:

- A. La prevalencia de la osteoporosis en la ciudad
- B. La probabilidad de osteoporosis en la paciente
- C. Las razones de probabilidad positiva y negativa del examen
- D. Los valores predictivos positivo y negativo del examen

4. Con la información disponible respecto al examen es posible decir que:

- A. Un examen de ultrasonido positivo es aproximadamente 2 veces más frecuente en mujeres con osteoporosis que en mujeres sin osteoporosis
- B. Un examen de ultrasonido positivo es aproximadamente 10 veces más frecuente en mujeres con osteoporosis que en mujeres sin osteoporosis
- C. Un examen de ultrasonido negativo es aproximadamente 10 veces menos frecuente en mujeres con osteoporosis que en mujeres sin osteoporosis
- D. Un examen de ultrasonido negativo es aproximadamente 2 veces menos frecuente en mujeres con osteoporosis que en mujeres sin osteoporosis.

Respuestas: 1: D, 2: A, 3: C, 4: B