



Uso de tests diagnósticos en la práctica clínica (Parte 2). Aplicación clínica y utilidad de un test diagnóstico

Lorena Cifuentes A. y Jaime Cerda L.

Pontificia Universidad
Católica de Chile.

Facultad de Medicina
Departamento de Pediatría (LCA).
Departamento de Salud Pública
(JCL).
Programa de Salud Basada en la
Evidencia (LCA, JCL).

Recibido: 15 de septiembre de
2009

Aceptado: 28 de abril de 2010

Correspondencia a:

Lorena Cifuentes A.
lcifuen@med.puc.cl

Clinical use of diagnostic tests (Part 2). Clinical application and usefulness of a diagnostic test

The clinical usefulness of a diagnostic test is determined by the extent to which it helps to modify the pretest probability of occurrence of a certain diagnosis. For this purpose, the calculation and application of likelihood ratios (LR) is a very useful tool. Every possible result of a test has a LR, which allows for an estimation of the magnitude by which the pretest probability increases or decreases, thereby helping the physician rule out, confirm, or continue investigating a diagnosis with new tests. In the first part of this article the properties of a diagnostic test (validity, reproducibility and safety) were analyzed. This second part illustrates the theoretical background underlying the concept of LR, explains the methodology for calculating them and describes their clinical application.

Key words: Diagnostic tests, clinical epidemiology, likelihood ratios.

Palabras clave: Tests diagnósticos, epidemiología clínica, razones de verosimilitud.

Introducción

En la primera parte del presente artículo (Parte 1: Análisis de las propiedades de un test diagnóstico)¹ se ilustró la forma en que se analizan las propiedades de un test o prueba diagnóstica (validez, reproducibilidad y seguridad) en base a la información aportada por su sensibilidad, especificidad y valores predictores. Un segundo aspecto lo constituye la evaluación de su utilidad clínica, para la cual resulta fundamental el cálculo y aplicación de estimadores conocidos como *likelihood ratios* (i.e. razones de verosimilitud, LR's). El presente artículo explica, en términos generales, los fundamentos teóricos subyacentes al concepto de LR, su forma de cálculo y su aplicación clínica. Para tales efectos se ejemplificarán algunos conceptos en base a los resultados de dos estudios.

Concepto y cálculo de *likelihood ratios*

El uso de LR's constituye una herramienta de gran utilidad para la toma de decisiones clínicas frente a un paciente en que se ha decidido solicitar algún test diagnóstico, por cuanto: los LR's son valores inherentes al test diagnóstico, sus valores son independientes de la prevalencia de la enfermedad en la población objeto de estudio (a diferencia de los valores predictores) y finalmente, contribuyen a dar respuestas a preguntas clínicas en el mismo sentido en que se desarrolla el proceso diagnóstico².

Cabe destacar que cada resultado posible de un test posee un LR, es decir, podremos calcular y aplicar tantos LR's

como resultados tenga una prueba diagnóstica. Para efectos prácticos, muchos resultados de pruebas diagnósticas se expresan como resultado positivo o negativo, en cuyo caso se obtendrán sólo dos LR's: LR para resultado positivo y LR para resultado negativo. Sin embargo, en numerosas ocasiones no es posible separar los resultados simplemente en positivo o negativo. En otras ocasiones, el "dicotomizar" resultados continuos o de más de dos categorías conduce a una pérdida de valiosa información científica, dificultando la toma de decisiones clínicas cuando en el uso cotidiano de ese test diagnóstico se obtienen resultados no sólo positivos o negativos. Por consiguiente, analizaremos por separado el cálculo de los LR's para resultados dicotómicos y politómicos (i.e. más de dos niveles).

Cálculo de LR's para tests diagnósticos de resultados dicotómicos

Los resultados de un test diagnóstico pueden ser de tipo binario o dicotómico (i.e. dos categorías, e.g. urocultivo positivo o negativo), politómicos (i.e. más de dos categorías, e.g. neutropenia leve, moderada o severa) o continuos (e.g. recuento de leucocitos)³. Los tests diagnósticos cuyos resultados son de tipo dicotómico poseen únicamente dos LR's, uno para el resultado positivo (LR+) y otro para el resultado negativo (LR-). Estos se pueden calcular a partir de los valores de sensibilidad y especificidad del test, empleando las siguientes fórmulas, (a) $LR+ = (\text{sensibilidad} / 1 - \text{especificidad})$ y (b) $LR- = (1 - \text{sensibilidad} / \text{especificidad})$.



El LR+ de un test corresponde a una razón entre la proporción de pacientes verdaderamente positivos (i.e. sensibilidad) y la proporción de individuos falsamente positivos (i.e. 1-especificidad). Por el contrario, el LR- de un test corresponde a una razón entre la proporción de pacientes falsamente negativos (i.e. 1-sensibilidad) y la proporción de individuos verdaderamente negativos (i.e. especificidad). Por consiguiente, un test diagnóstico será de mayor utilidad en la medida que su LR+ sea de mayor magnitud y que su LR- sea de menor magnitud. El siguiente ejemplo ilustra la metodología de cálculo de los LR's para resultados dicotómicos.

Ejemplo 1

En 1993, Steed y cols⁴, evaluaron la sensibilidad y especificidad del test directo de *Streptococcus pyogenes* (Group A *Streptococcus* Direct Test, GASD) para el diagnóstico de faringoamigdalitis estreptocócica. Para ello, se tomó un cultivo faríngeo a 277 individuos consultantes pediátricos, siendo sus resultados contrastados con aquellos aportados por el GASD (detección de cARN estreptocócico en secreción faríngea). La presencia de un cultivo positivo para *Streptococcus pyogenes* fue considerado como el estándar de oro para el diagnóstico de faringoamigdalitis estreptocócica. Por su parte, el GASD podía arrojar un resultado positivo o negativo, denotando la presencia o ausencia de material genético de la bacteria. La Tabla 1 resume los resultados de este estudio. Para el diagnóstico de faringoamigdalitis estreptocócica, el GASD presentó una sensibilidad de $51/59 = 86,4\%$ y especificidad de $208/218 = 95,4\%$. Según las fórmulas anteriormente detalladas, el $LR+ = 0,864/(1-0,954) = 19$ y $LR- = (1-0,864)/0,954 = 0,14$.

Cálculo de LR's para tests diagnósticos de resultados politómicos

Una ventaja del uso de LR's es que éstos pueden ser calculados no solamente para resultados dicotómicos, sino para tantas categorías como resultados posibles tenga un test diagnóstico. Ello ofrece la posibilidad de conocer la utilidad clínica real de tests diagnósticos politómicos sin tener que agrupar forzosamente los resultados en términos dicotómicos, operación que genera una importante pérdida de información. El siguiente ejemplo ilustra la forma en que se realiza el cálculo de LR's para tests diagnósticos de resultados politómicos. En este caso lo que se calcula el LR de cada posible categoría de resultado⁵.

Ejemplo 2

En 1998, McIsaac y cols⁶, diseñaron un puntaje clínico para estimar la probabilidad de que un individuo consultante por odinofagia presente el diagnóstico de

faringoamigdalitis estreptocócica, basándose en la presencia de diversos signos clínicos. Un total de 503 individuos (niños y adultos) fueron evaluados; para cada uno de ellos se calculó su puntaje clínico (0-4 puntos) y se tomó un cultivo faríngeo, considerado el estándar de oro para el diagnóstico de esta patología. La Tabla 2 resume los valores de puntaje clínico y los resultados del cultivo faríngeo, así como el LR de cada categoría de puntaje.

En términos generales, el LR para cada categoría de puntaje 'X' se calcula en base a una razón, correspondiendo su numerador a la proporción de pacientes con la enfermedad en estudio que obtuvo como resultado el puntaje 'X' y su denominador a la proporción de individuos sin la enfermedad en estudio que obtuvo como resultado el puntaje 'X'.

En el mencionado estudio, la presencia y ausencia de enfermedad fue definida por el resultado del cultivo faríngeo, considerado el estándar de oro para el diagnóstico de esta patología. En él, 28/65 pacientes con faringoamigdalitis estreptocócica obtuvieron 4 puntos; por su parte, 25/438 individuos sin faringoamigdalitis estreptocócica obtuvieron también 4 puntos. En consecuencia, el LR para el resultado '4 puntos' corresponde a $(28/65)/(25/438) = 7,55$. En otras palabras, la probabilidad de que un paciente con faringoamigdalitis estreptocócica obtenga '4 puntos' es 7,55 veces la probabilidad de que un individuo sin la

Tabla 1. Resultados del cultivo faríngeo y test directo de *Streptococcus pyogenes* (GASD) en 277 pacientes⁴

GASD	Cultivo faríngeo (n)		Total
	Positivo	Negativo	
Positivo	51	10	61
Negativo	8	208	216
Total	59	218	277

GASD: Group A *streptococcus* Direct Test. Prevalencia = 21,3%; Sensibilidad = 86,4%; Especificidad = 95,4%; LR+ = 19; LR- = 0,14.

Tabla 2. Resultados del puntaje clínico, cultivo faríngeo y likelihood ratios (n = 503 pacientes)⁶

Puntaje clínico	Cultivo faríngeo (n)		LR
	Positivo	Negativo	
0	4	156	0,17
1	7	131	0,36
2	11	87	0,85
3	15	39	2,59
4	28	25	7,55
Total	65	438	-

LR+, likelihood ratio.



enfermedad obtenga ‘4 puntos’. Visto de otro modo, el LR indica que es 7,55 veces más ‘verosímil’ que un paciente con la patología obtenga 4 puntos, en comparación a que un individuo sin la enfermedad obtenga aquel puntaje (posiblemente de esta observación derive la traducción de LR’s como ‘razones de verosimilitud’)^{3,5}. Una forma más sencilla de expresar este concepto probablemente sea de manera gráfica, como se hará más adelante en este artículo.

Aplicación de los *likelihood ratios* en la práctica clínica

En términos prácticos, los LR’s indican la magnitud en la cual los resultados de un test diagnóstico modifican la probabilidad pre test, ya sea aumentándola o disminuyéndola. Mientras más se aleja un LR del valor 1, mayor es la magnitud del cambio. En distintas publicaciones se menciona que los tests diagnósticos cuyas categorías de resultados presentan LR sobre 10,0 o bajo 0,1 aportan evidencia fuerte para apoyar o rechazar hipótesis diagnósticas^{3,7}. Sin embargo, esto aplica en términos muy generales, ya que la real utilidad clínica de un test se

medirá en la medida que sus resultados, y sus respectivos LR’s, permitan abandonar la “zona de incertidumbre” (i.e. área comprendida entre los umbrales diagnóstico y terapéutico)². Esto dependerá, en última instancia, no sólo del valor de LR, sino también de la probabilidad pre test inicial.

Desde un punto de vista matemático, los LR’s corresponden a *odds* (i.e. cociente entre dos probabilidades). Este hecho no es menor, en vista que el cálculo de la probabilidad post test utilizando LR’s requiere la realización de una serie de operaciones matemáticas, a saber: transformación de la probabilidad pre test en un odds pre test, multiplicación del odds pre test por el LR y finalmente, transformación del odds post test en una probabilidad post test. La realización de estos cálculos matemáticos puede simplificarse utilizando una herramienta gráfica conocida como nomograma de Fagan (Figura 1)⁸. Este cuenta con tres columnas de números, correspondiendo la primera a la probabilidad pre test, la segunda al LR y la tercera a la probabilidad post test.

Retomando el ejemplo 1 (test diagnóstico de resultado dicotómico), un primer paso consiste en definir la probabilidad pre test de que un individuo consultante presente faringoamigdalitis estreptocócica. Para tales efectos se homologará la probabilidad pre test a la prevalencia de la enfermedad en la población en estudio (i.e. 21%) (Tabla 1). Un segundo paso consiste en calcular los valores de LR+ y LR- para el GASD, según las fórmulas (a) y (b) anteriormente detalladas, siendo LR+ = 19 y LR- = 0,14. Finalmente, un tercer paso contempla el trazo de dos líneas rectas, cuyo origen común es un punto ubicado en la primera columna, a la altura de la probabilidad pre test (21%). Cada línea debe cruzar la segunda columna a la altura del punto correspondiente al LR+ y LR-. El punto en el cual cada recta cruza la tercera columna corresponde a la probabilidad post test, expresada en términos porcentuales. En el ejemplo, para una probabilidad pre test de 21%, si el GASD es positivo la probabilidad post test aumenta a 84%; por el contrario, si el GASD es negativo, la probabilidad post test se reduce a 4% (Figura 1). En el segundo ejemplo (test diagnóstico de resultados politómicos), considerando una probabilidad pre test de faringoamigdalitis estreptocócica de 13% (i.e. prevalencia de esta patología en la población en estudio), un resultado de “0 puntos” (LR = 0,17) conduce a una probabilidad post test a 2%. Por su parte, un resultado “2 puntos” (LR = 0,85) conduce a una probabilidad post test de 11%. Por último, un resultado de “4 puntos” (LR = 7,55) conduce a una probabilidad post test de 53% (Figura 2).

Un cuarto paso, eminentemente clínico, consiste en determinar si la aplicación del test diagnóstico generó una probabilidad post test lo suficientemente alta como para superar el umbral terapéutico o lo suficientemente baja como para encontrarse por debajo del umbral diagnóstico².

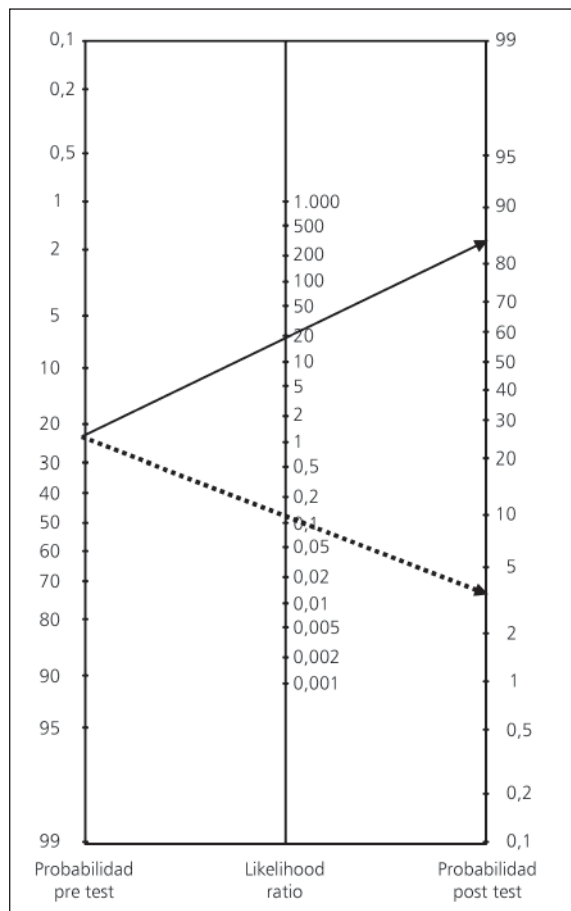


Figura 1. Nomograma de Fagan (Ejemplo 1). Probabilidad pre test = 21%. Si el GASD es positivo, la probabilidad post test aumenta a 84% (flecha continua). Si el GASD es negativo, la probabilidad post test disminuye a 4% (flecha punteada).



De no ser así, se deberá realizar un segundo test diagnóstico, mediante el cual se refinará aún más la probabilidad pre test. En tal caso, la nueva probabilidad pre test no será la inicial, sino más bien será la probabilidad post test obtenida tras la aplicación del primer test diagnóstico. Es posible aplicar seriadamente tantos tests diagnósticos como el clínico estime conveniente, refinando secuencialmente la probabilidad pre test. Sin embargo, esta conducta es muy poco eficiente, por cuanto lo ideal es descartar o bien llegar a un diagnóstico correcto realizando la menor cantidad de exámenes a cada paciente.

Conclusión

El presente artículo ilustró la forma en que se analiza la utilidad de un test diagnóstico mediante el cálculo de *likelihood ratios*. Es importante recordar que el punto de partida de todo test diagnóstico es el planteamiento de una probabilidad pre test adecuada, cuya precisión es clínico-dependiente. En la medida que éste posea mayores conocimientos teóricos y experiencia clínica, así como logre recolectar la información más relevante de la anamnesis y examen físico del paciente, más preciso será este punto de partida, optimizando el rendimiento del test. De igual forma, es importante señalar que la Medicina Basada en Evidencia ha sistematizado las preguntas que todo lector debe plantearse al analizar críticamente un artículo sobre interpretación de tests diagnósticos^{7,9,10}. El estudio en mayor profundidad de esta área temática aportará al lector herramientas de gran utilidad tanto para la actividad clínica como de investigación.

Resumen

La utilidad clínica de un test diagnóstico se califica de acuerdo a la magnitud en la que éste contribuye a modificar la probabilidad pre test de ocurrencia de un determinado diagnóstico. Para tales efectos, el cálculo y aplicación de razones de verosimilitud (*likelihood ratios*, LR) constituye una herramienta de gran utilidad. Cada resultado posible de un test posee un LR, a partir del

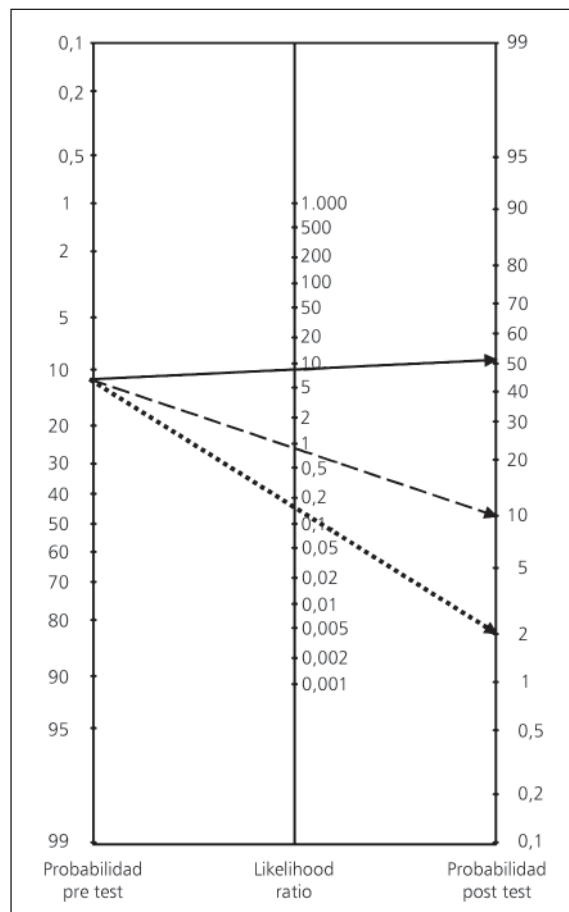


Figura 2. Nomograma de Fagan (Ejemplo 2). Probabilidad pre test = 13%. Un puntaje de "0 puntos" conduce a una probabilidad post test de 2% (flecha punteada). Un puntaje de "2 puntos" conduce a una probabilidad post test de 11% (flecha segmentada). Un puntaje de "4 puntos" conduce a una probabilidad post test de 53%.

cual es posible estimar cuánto aumenta o disminuye la probabilidad pre test, permitiendo al clínico descartar un diagnóstico, confirmarlo o bien continuar su investigación mediante la solicitud de nuevos exámenes. En la primera parte del presente artículo se analizaron las propiedades de un test diagnóstico (validez, reproducibilidad y seguridad). En esta segunda parte se ilustra el sustrato teórico subyacente al concepto de LR, se explica su metodología de cálculo y se describe su aplicación clínica.

Referencias

- 1.- Cerda J, Cifuentes L. Uso de tests diagnósticos en la práctica clínica (Parte 1). Análisis de las propiedades de un test diagnóstico. *Rev Chil Infect* 2010; 27 (3): 205-8.
- 2.- Capurro D, Rada G. El proceso diagnóstico. *Rev Méd Chile* 2007; 135: 534-8.
- 3.- Salech F, Mery V, Larrondo F, Rada G. Estudios que evalúan un test diagnóstico: interpretando sus resultados. *Rev Méd Chile* 2008; 136: 1203-8.
- 4.- Steed L L, Korgenski E K, Daly J A. Rapid detection of *Streptococcus pyogenes* in pediatric patient specimens by DNA probe. *J Clin Microbiol* 1993; 31: 2996-3000.
- 5.- Deeks J J, Altman D G. Statistics Notes: Diagnostic tests 4: likelihood ratios. *Br Med J* 2004; 329: 168-9.
- 6.- McIsaac W J, White D, Tannenbaum D, Low D E. A clinical score to reduce unnecessary antibiotic use in patients with sore throat. *Can Med Assoc J* 1998; 158: 75-83.
- 7.- Jaeschke R, Guyatt G, Lijmer J. Diagnostic Tests. *Guyatt G, Drummond R. Users's Guides to the Medical Literature. Essentials of Evidence-Based Clinical Practice.* AMA Press; 2002: 187-217.
- 8.- Fagan T J. Nomogram for Bayes theorem. *N Engl J Med* 1975; 293: 257.
- 9.- Vera C, Letelier L M, Carvajal J. Guía para el análisis crítico de estudios que evalúan exámenes diagnósticos. *Rev Chil Obstet Ginecol* 2005; 70: 196-202.
- 10.- Valenzuela L, Cifuentes L. Validez de estudios de tests diagnósticos. *Rev Méd Chile* 2008; 136: 401-4.