

Estudios de exactitud diagnóstica: Herramientas para su Interpretación

Dres. Sebastián Bravo-Grau⁽¹⁾, Juan Pablo Cruz Q⁽²⁾.

1. Radiólogo, Fellow de Neurorradiología, P. Universidad Católica de Chile. Santiago - Chile.

2. Neurorradiólogo, Profesor Asistente, P. Universidad Católica de Chile. Santiago - Chile.

Diagnostic accuracy studies: Tools for its Interpretation

Abstract: The aim of this paper is to provide basic tools for the interpretation of diagnostic accuracy studies, by analyzing the data that the contingency table provides. Concepts are given in the application and meaning of sensitivity, specificity, positive and negative predictive values. Furthermore, by analyzing the pre-test and post-test probabilities, construction and implementation of the likelihood ratio are explained. The diagnostic odds ratio and its limitations are described. Finally, the importance of calculating confidence intervals is explained. Understanding the contingency table and the construction of the parameters that we can calculate from this, is essential for the critical evaluation and development of diagnostic accuracy studies, as well as their application in daily clinical work.

Keywords (MeSH): Biostatistics, Diagnostic, Diagnostic accuracy, Likelihood ratio, Predictive value, Sensibility, Specificity.

Resumen: El objetivo del presente artículo es entregar herramientas básicas para la interpretación de los estudios de exactitud diagnóstica, mediante el análisis de los datos que entrega la tabla de contingencia. Se entregan conceptos en la aplicación y significado de sensibilidad, especificidad, valores predictivos positivo y negativo. Además, mediante el análisis de las probabilidades pre-test y post-test, se explica la construcción e implementación de los likelihood ratio o razones de verosimilitud. Se describe el odds ratio diagnóstico y sus limitaciones. Por último, se explica la importancia del cálculo de intervalos de confianza. Entender la tabla de contingencia y la construcción de los parámetros que de esta podemos calcular es fundamental para la valoración crítica y desarrollo de estudios de exactitud diagnóstica, así como para su aplicación en el quehacer clínico diario.

Palabras clave (MeSH): Bioestadística, Diagnóstico, Especificidad, Exactitud diagnóstica, Índice de probabilidad, Sensibilidad, Valores predictivos.

Bravo-Grau S., Cruz JP. Estudios de exactitud diagnóstica: Herramientas para su Interpretación. Rev Chil Radiol 2015; 21(4): 158-164.

Correspondencia: Sebastián Bravo-Grau / sebastianbravo@gmail.com / sebravo1@uc.cl

Trabajo recibido el 30 de agosto de 2015. Aceptado para publicación el 16 de noviembre de 2015.

Introducción

Una prueba diagnóstica se refiere a cualquier método para obtener información adicional del estado de salud del paciente. El tipo de información adquirida mediante la utilización de un test diagnóstico no sólo incluye a la presencia o ausencia de una determinada enfermedad, sino que también a la etapificación de una enfermedad conocida o bien a establecer la existencia de determinada condición, no necesariamente patológica⁽¹⁾. La prueba diagnóstica o test pretende

sacar de la 'zona de incertidumbre' a un determinado paciente⁽²⁾. La 'zona de incertidumbre' se encuentra limitada entre el 'umbral de diagnóstico', bajo el cual la enfermedad es descartada y el 'umbral terapéutico', sobre el cual la enfermedad se confirma y el tratamiento se inicia⁽²⁻⁴⁾. Fuera de estos dos umbrales, vale decir, fuera de la 'zona de incertidumbre', no se requiere un test diagnóstico para establecer la presencia o ausencia de la condición o enfermedad sospechada. Desde ahora en adelante nos referiremos a condición

o enfermedad en forma intercambiable.

Los estudios de exactitud diagnóstica tienen una estructura común básica, en términos generales, pueden corresponder a estudios de tipo caso-control, transversales o cohorte. En este tipo de estudios (de exactitud diagnóstica) los resultados obtenidos con la prueba diagnóstica en evaluación se comparan con los de un estándar de referencia en un mismo grupo de pacientes. El estándar de referencia, también llamado *gold standard* (estándar de oro) corresponde, en resumen, a la mejor manera disponible y ampliamente aceptada para establecer la presencia o ausencia de determinada condición. Este puede ser un único examen, o bien, una combinación de métodos (*gold standard* compuesto). El término exactitud se refiere precisamente a la concordancia entre los resultados de la prueba diagnóstica con el estándar de referencia⁽¹⁾. En el presente artículo no se ahondará en los requisitos ni en la calidad metodológica que una investigación de exactitud diagnóstica debe tener para que sus resultados sean válidos, lo que ha sido descrito ampliamente por otros autores^(5,6).

La utilidad de las pruebas diagnósticas generalmente se describe y/o cuantifica en términos de su sensibilidad, especificidad, valor predictivo positivo, valor predictivo negativo y likelihood ratios (razones de verosimilitud) positivo y negativo. Existen otros parámetros descriptivos de las propiedades diagnósticas de un test específico. La mayoría de los médicos generales reconocen las definiciones de la sensibilidad y el valor predictivo positivo, pero éstas no son necesariamente aplicadas en forma correcta⁽⁷⁾. Además, hay un déficit por parte de nuestra especialidad en la comunicación científica de estos parámetros⁽⁸⁾.

El objetivo del presente artículo es entregar herramientas básicas para la interpretación de los estudios de exactitud diagnóstica. Específicamente, se analizarán la interpretación de los datos que entrega la tabla de contingencia en este tipo de investigaciones. Se revisarán algunos de los conceptos entregados hace 11 años en esta revista, profundizando algunas temáticas, actualizando otras y aclarando algunos conceptos⁽⁹⁾.

La tabla de contingencia

Esta tabla (Tabla I) consiste de dos columnas que corresponden al resultado dicotómico positivo o negativo (presencia o ausencia) de la enfermedad o condición según el *gold standard* y en las filas según nuestra prueba diagnóstica. Forman cuatro celdas que se designan con una letra 'a' a la 'd', de izquierda a derecha y de arriba a abajo. La designación de las letras de cada celda, así como, la ubicación del *gold standard* y de la prueba diagnóstica en estudio, es por convención. Cada una de estas celdas corresponde a verdadero positivo, falso positivo, falso negativo y

verdadero negativo, respectivamente. Estos términos se definen como:

- Verdadero positivo: el paciente tiene la enfermedad y el test es positivo
- Falso positivo: el paciente no tiene la enfermedad pero el resultado del test es positivo
- Verdadero negativo: el paciente no tiene la enfermedad y el test es negativo
- Falso negativo: el paciente tiene la enfermedad pero el resultado del test es negativo.

Sensibilidad y especificidad

Sensibilidad: Corresponde a la proporción de individuos correctamente diagnosticados con la condición o enfermedad por la prueba diagnóstica (Tabla I). En otras palabras la proporción de verdaderos positivos correctamente identificados por el test del total de individuos enfermos según el estándar de referencia⁽¹⁰⁾.

Especificidad: Corresponde a la proporción de individuos correctamente diagnosticados con ausencia de la condición o enfermedad por la prueba diagnóstica en estudio. Vale decir, es la proporción de verdaderos negativos que fueron correctamente identificados por el test, del total de individuos sanos según el estándar de referencia⁽¹⁰⁾. De lo anterior podemos inferir que la especificidad es el cociente entre los verdaderos negativos dividido por la suma de verdaderos negativos y falsos positivos (Tabla I).

Estas proporciones son parámetros inherentes a la prueba diagnóstica, en términos generales, menos dependientes de la prevalencia de la enfermedad. Como se puede inferir de la definición de sensibilidad y especificidad, la orientación de estos parámetros es en el sentido opuesto del razonamiento de toma de decisión clínica y por ende su información no es necesariamente útil para el caso a caso⁽¹¹⁾.

La utilidad de los parámetros de sensibilidad y especificidad dependen en gran medida del escenario en que se apliquen. Por ejemplo, la citología cérvico uterina convencional, prueba de tamizaje o screening para la detección de cáncer de cuello uterino, tiene una muy alta especificidad y una baja sensibilidad. En este tipo de tamizaje de cáncer en población sana, al ser una enfermedad de baja prevalencia, se busca una muy alta especificidad (muy baja fracción de falsos positivos), ya que se pretenden evitar los falsos positivos, en cuyos casos su estudio proseguirá con exámenes más costosos e invasivos. Por otra parte, si el objetivo de una prueba de tamizaje es evitar la transmisión de una enfermedad prevenible, se optimizará la sensibilidad.

Valores predictivos

La sensibilidad y la especificidad son medidas importantes de la exactitud diagnóstica de una prueba, pero no pueden ser usadas para estimar la probabi-

Tabla I. La Tabla de Contingencia. Explicación de la construcción de los principales parámetros en la descripción de una prueba diagnóstica.			
Estatus de la condición (resultado del gold standard)			
Resultado de la prueba diagnóstica	Presente/positivo	Ausente/negativo	Totales marginales
Test positivo	a (Verdadero positivo)	b (Falso positivo)	a+b
Test negativo	c (Falso negativo)	d (Verdadero negativo)	c+d
Totales marginales	a+c (Pacientes con la enfermedad)	b+d (Pacientes sin la enfermedad)	
Parámetro	Fórmula	Definición	
Sensibilidad	$a/(a+c)$	Proporción de pacientes con la enfermedad que tendrán test positivo	
Especificidad	$d/(b+d)$	Proporción de pacientes sin la enfermedad que tendrán test negativo	
Valor predictivo positivo	$a/(a+b)$	Probabilidad de que el paciente tenga la enfermedad dado que el test es positivo	
Valor predictivo negativo	$d/(c+d)$	Probabilidad de que el paciente no tenga la enfermedad dado que el test es negativo	
Likelihood ratio (+)	$\text{sensibilidad}/(1-\text{especificidad})$	Describe cuántas veces es más probable que reciba un resultado determinado una persona con la enfermedad que una persona sin la enfermedad	
Likelihood ratio (-)	$(1-\text{sensibilidad})/\text{especificidad}$		
Exactitud	$(a+d) / (a+b+c+d)$	La probabilidad de que el resultado del test prediga correctamente la presencia o ausencia de la enfermedad	
Odds ratio diagnóstico	$(a/c) / (b/d)$	Razón entre la odds de estar enfermo si la prueba da positivo y la odds de no estar enfermo si la prueba da negativo	

lidad de enfermedad en un paciente individual. Los valores predictivos positivos (VPP) y negativos (VPN) proporcionan estimaciones de la probabilidad de la enfermedad. Vale decir, es la probabilidad de que la prueba diagnóstica entregue el diagnóstico correcto, si esta resulta positiva o negativa.

Valor Predictivo Positivo: Corresponde a la probabilidad condicional de que el paciente tenga la enfermedad, dado que el test resultó positivo. Expresado de otra manera, es la proporción de pacientes con la prueba diagnóstica positiva que efectivamente tienen la condición (Tabla I).

Valor Predictivo Negativo: Corresponde a la probabilidad condicional de que el paciente no tenga la enfermedad, dado que la prueba diagnóstica resultó negativa. En otras palabras, es la probabilidad de que el individuo no tenga la condición en estudio luego de que el test es negativo (Tabla I). Es equivalente al inverso de la probabilidad post-test de tener la enfermedad dado que el test resultó negativo (1-VPN)⁽¹²⁾.

En la Tabla II se ilustra el ejemplo 1, que corresponde a un estudio del año 2001⁽¹³⁾, sobre la exactitud diagnóstica de la tomografía computada (TC) para predicción de daño cerebral irreversible luego de accidente cerebrovascular isquémico. Los datos en esta tabla de contingencia (Tabla II) corresponden a la exactitud diagnóstica de la TC para diagnóstico de infarto cerebral, en las primeras 6 horas luego de iniciado los síntomas, en el estudio ECASS II (European Cooperative Acute Stroke Study II) y por el panel de reporte de la TC compuesto por tres neurorradiólogos (CT reading panel)⁽¹³⁾. Estos mismos datos se modificaron para generar el ejemplo 2, con una prevalencia menor de la enfermedad (58,5%) y ligeramente disminuyendo el tamaño de la muestra (Tabla III).

Entiéndase prevalencia de enfermedad como la proporción del total de pacientes, incluidos en el estudio, con la condición según el *gold standard*. En nuestro ejemplo, utilizaron como *gold standard*

el seguimiento con TC a las 22-36 horas y a los 6-8 días de iniciado los síntomas⁽¹³⁾.

Mediante el empleo de estos ejemplos, se nota como al variar la prevalencia de la enfermedad en nuestra población, la sensibilidad y especificidad se mantienen estables. Por el contrario, el VPP y VPN, varían considerablemente en forma inversa, al aumentar la prevalencia aumenta el VPP y disminuye el VPN y vice-versa. En el ejemplo 1 (Tabla II), con una prevalencia de 86,4%, el VPP de la TC para infarto cerebral era de 96,4%. Vale decir, la probabilidad de que el paciente tenga un infarto cerebral dado que

la TC fue positiva, en este escenario clínico, es de 96,4%, mientras que en el ejemplo 2 (Tabla III) es de 85,8%.

Likelihood ratios

Los likelihood ratios (LR) o razones de verosimilitud se definen como cuántas veces es más probable que un paciente con la enfermedad tenga un determinado resultado en el test que pacientes sin la enfermedad (Tabla I). Dicho de otra manera, es la razón de probabilidad de un resultado específico en pacientes con la enfermedad versus en aquellos que

Tabla II. Ejemplo 1, con datos obtenidos de un estudio del año 2001⁽¹³⁾, sobre la exactitud diagnóstica de la tomografía computada para predicción de daño cerebral irreversible luego de accidente cerebrovascular isquémico.

TC en infarto agudo	Gold standard		TOTAL
	Positivo	Negativo	
Test positivo	433	16	449
Test negativo	246	91	337
Totales marginales	679	107	786
			IC 95%
Sensibilidad	63,8%	60,2 - 67,4	
Especificidad	85,0%	78,3 - 91,8	
VPP	96,4%	94,7 - 88,2	
VPN	27,0%	22,3 - 31,7	
LR+	4,3	2,7 - 6,7	
LR-	0,4	0,37 - 0,48	
Prevalencia	86,4%		

Tabla III. Ejemplo 2 con datos modificados, con una prevalencia menor de la enfermedad y un tamaño muestral ligeramente menor.

TC en infarto agudo -2	Gold standard		TOTAL
	Positivo	Negativo	
Test positivo	289	48	337
Test negativo	164	273	437
Totales marginales	453	321	774
			IC 95%
Sensibilidad	63,8%	59,4 - 68,2	
Especificidad	85,0%	81,1 - 88,9	
VPP	85,8%	82,0 - 89,4	
VPN	62,5%	57,9 - 67,0	
LR+	4,3	3,3 - 5,6	
LR-	0,4	0,37 - 0,49	
Prevalencia	58,5%		

no la tienen^(14,15). En el caso de resultados dicotómicos el LR positivo toma valores entre 1 y el infinito, mientras que el LR negativo toma valores entre el 1 y el 0. El LR positivo se calcula como sensibilidad dividido en 1-especificidad, o bien el cociente de verdaderos positivos dividido en falsos positivos. El LR negativo se calcula como especificidad dividido en 1-sensibilidad, o bien el cociente de los falsos negativos dividido en los verdaderos negativos. Si la razón de verosimilitud es igual a 1, la probabilidad del diagnóstico es la misma antes y después de aplicar la prueba. En este caso la prueba es inútil, no tiene capacidad discriminante. Cuanto más se aleje de 1 el valor de la razón de verosimilitud, con mayor fuerza la prueba nos sacará de la 'zona de incertidumbre' diagnóstica⁽²⁾. Como podemos ver mediante los ejemplos, los LR no se ven afectados por la prevalencia de la enfermedad (Tablas II y III).

Los LR son claves en cuánto son la expresión del teorema de Bayes, de relacionar probabilidades condicionales, en particular de la probabilidad pre-test para la estimación de la probabilidad post-test⁽¹⁵⁾. La probabilidad pre-test, corresponde a la probabilidad de enfermedad del paciente en la población previo al empleo de la prueba diagnóstica, la que puede ser estimada de múltiples maneras, la consideraremos como igual a la prevalencia. La probabilidad post-test corresponde a la probabilidad de que el paciente tenga la enfermedad en nuestro escenario clínico y luego de aplicada la prueba diagnóstica⁽¹⁶⁾. Lo expuesto previamente se obtiene de la siguiente manera:

- Probabilidad pre-test= Prevalencia
- Odds pre-test= Probabilidad pre-test / (1 - Probabilidad pre-test)
- Odds post-test= Odds pre-test x LR
- Probabilidad post-test= Odds post-test / (1 + Odds post-test)

De lo anterior, entendemos la definición de odds (O) como: $O = p/(p-1)$, donde p es la prevalencia. Así, si conocemos la prevalencia podemos calcular el odds y vice-versa.

Fagan implementó un nomograma para graficar el teorema de Bayes, combinando la razón de verosimilitud (LR), la probabilidad pre-test y la probabilidad post-test⁽¹⁷⁾ (Figura 1). Este permite simplificar su utilización. Existen además múltiples sitios web y aplicaciones gratuitas para su cálculo en el día a día.

Consideremos el ejemplo en las tablas II y III, como los únicos estudios que nos describen la TC para diagnóstico de infarto. El clínico estima la probabilidad pre-test como 70%, por ende el odds pre-test es 2,33. El examen es negativo (normal), como conocemos su $LR(-) = 0,4$, entonces el odds post-test es 0,93 y finalmente la probabilidad post-test es 48,3%. Lo que sigue siendo alto y reafirma que una TC normal no descarta un infarto cerebral.

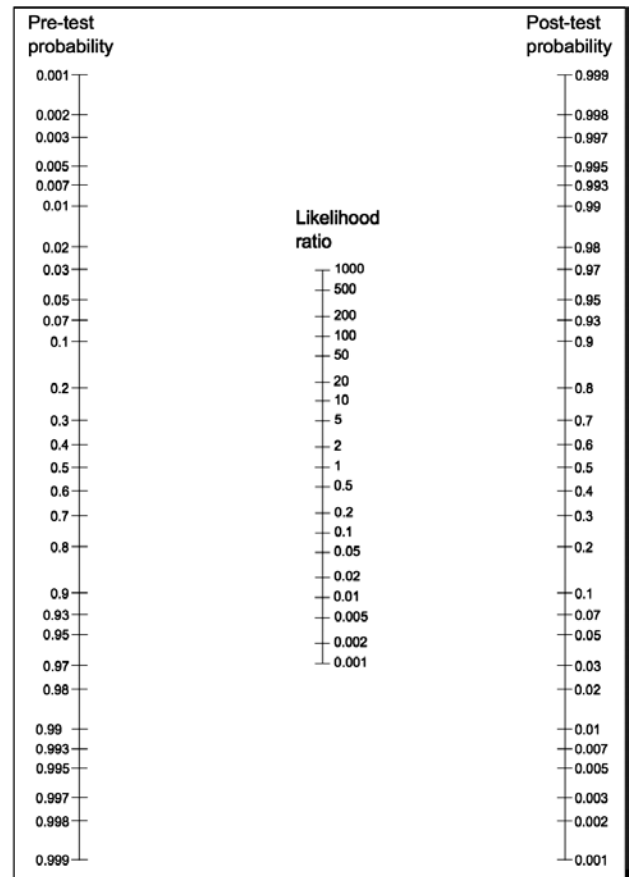


Figura 1. Nomograma de Fagan para el teorema de Bayes. Una línea recta entre la probabilidad pre-test y el likelihood ratio determina la probabilidad post-test.

Por último, si la probabilidad pre-test de alguna condición es muy baja, por ejemplo 0,001 (0,1%), con una prueba diagnóstica con un $LR(+) = 10$ (considerado alto), la probabilidad post-test con resultado de la prueba positiva es de 0,0099 (0,99%). Vale decir, la probabilidad post-test sigue siendo muy baja y no se ha salido de la 'zona de incertidumbre'.

El LR se puede calcular para resultados dicotómicos: positivo o negativo o para múltiples categorías, lo que le confiere una ventaja en comparación con las otras estimaciones previamente descritas⁽¹⁵⁾. Existen otros recursos gráficos para la presentación de las propiedades de una prueba diagnóstica, los que son relevantes en determinados escenarios, entre otros en pruebas diagnósticas con resultados cuantitativos continuos⁽¹⁸⁾. Entre éstos destacan las curvas ROC (Receiver Operating Characteristic Curves)⁽¹⁹⁾, la que se discutirá en más detalle en otro artículo.

Odds ratio diagnóstico

La odds ratio diagnóstica (DOR) es conocida como un índice estadístico en los estudios epidemiológicos casos/controles representando la fuerza de asociación entre el factor de riesgo y la enfermedad. Aquí podría ser útil para mostrar la fuerza de asociación

entre el resultado de una prueba y la enfermedad. Este índice único traduce las prestaciones de una prueba con un solo valor que no está influenciado por la prevalencia. Es la razón entre la odds o chance de estar enfermo si la prueba da positivo y la odds o chance de no estar enfermo si la prueba da negativo.

Los valores de la DOR varían de cero a infinito (cuantos más altos son los valores, mejor es la prueba). El valor DOR= 1 significa que la prueba no es discriminante, el test es inútil⁽²⁰⁾. Los valores mayores de 1 significan que es más probable que la prueba de positivo en el caso de enfermos que en sanos. Esta medida tiene varias limitaciones, en especial, el de combinar sensibilidad y especificidad en un único indicador, perdiendo el valor relativo de cada uno de éstos. Por consecuencia, una prueba diagnóstica con baja sensibilidad y alta especificidad puede tener el mismo DOR que una con alta sensibilidad y baja especificidad⁽²⁰⁾. Esta limitación la comparten otros parámetros únicos, como por ejemplo la exactitud (accuracy) (Tabla I).

Intervalos de confianza

La sensibilidad, especificidad, VPP y VPN, son proporciones, por lo que se les puede y debe calcular intervalos de confianza utilizando métodos estándar a partir de la proporción binomial y el teorema central del límite (Figura 2). Este método depende del tamaño del intervalo y del tamaño muestral, en la fórmula el “N”. Cabe destacar, que para tamaños muestrales menores a 30, esta fórmula de cálculo es menos confiable, recomendándose utilizar otras estimaciones. De igual manera, cuando el estimador puntual es cercano al 1 (o 100%) o al 0 (0%), algunos autores plantean la necesidad de estimar el intervalo de confianza de otras maneras para mejorar rendimiento⁽²¹⁾.

$$p \pm Z_{0.95} \sqrt{\frac{p(1-p)}{N}} \quad \begin{matrix} p = \text{proporción} \\ N = \text{tamaño muestral} \end{matrix}$$

Figura 2. Fórmula simplificada para el cálculo de intervalo de confianza de una proporción. El valor de Z se obtiene a partir de la función normal (función inversa de una normal estándar), para un intervalo de confianza de 95% con dos colas es 1,96. p corresponde a la proporción a la que se desea calcular el intervalo de confianza. N corresponde al denominador de determinada proporción, por ejemplo, en el caso de la sensibilidad, corresponde a la suma entre verdaderos positivos y falsos negativos.

El cálculo de intervalos de confianza para likelihood ratio, siendo estas razones-cocientes entre probabilidades, es más complejo, por lo que se pueden utilizar calculadoras disponibles en internet o bien programas estadísticos apropiados.

Una forma de interpretar el intervalo de confianza, es que si repetimos un mismo experimento muchas veces, el intervalo contendrá el valor del parámetro en (1-alfa) % de las veces. Se conoce el valor de alfa como el nivel de significancia, generalmente se toma 0,05 como valor de alfa, lo que equivale a un intervalo de confianza de 95%. Otros intervalos de confianza menos utilizados son de 90% y 99%. Entre más ancho el intervalo de confianza, mayor es la incertidumbre que existe sobre la estimación puntual del parámetro.

Mientras más grande es la muestra, menor es el intervalo de confianza, lo que se puede observar en los intervalos de confianza de los ejemplos (Tabla II y Tabla III). Una pequeña variación en el tamaño muestral (en el ejemplo 2 se redujo en 12, ver Tabla III) va ampliando el intervalo de confianza, sin cambios en la estimación puntual de la sensibilidad y especificidad. Lo anterior toma especial relevancia en estudios con pequeño tamaño muestral⁽²²⁾.

Discusión

Los conceptos revisados son de especial importancia en la actual era de la radiología basada en la evidencia^(9,23,24). La validez de los análisis anteriormente descritos toman como requisito que provengan de estudios con metodologías de calidad mínima aceptable⁽¹⁻⁵⁾. Además, en el presente artículo se ha hecho énfasis en el análisis de datos dicotómicos, por motivos de extensión, sin abordar la interpretación de los datos continuos o semi-cuantitativos, lo que será analizado en otro artículo en forma separada.

Los descriptores de sensibilidad y especificidad se pueden resumir como parámetros intrínsecos a la prueba diagnóstica. De esta forma, son menos aplicables a la decisión clínica específica de determinado paciente. Por otra parte, los valores predictivos positivo y negativo, son aplicables al algoritmo de decisión de nuestro paciente en una situación clínica determinada, sin embargo, son dependientes de la prevalencia de la enfermedad en la población estudiada. Los LR son los parámetros más útiles para evaluar una prueba diagnóstica, ya que no dependen de la prevalencia, y son aplicables en la toma de decisión clínica del paciente. Nos ayudan a entender por qué no es apropiado efectuar pruebas diagnósticas a poblaciones con una muy baja probabilidad pre-test de enfermedad.

Es importante considerar la forma de estimación de la prevalencia, en especial cuando esta se asigna como la probabilidad pre-test en la implementación de los LR. Se debe ser cuidadoso en la extrapolación de la prevalencia de la enfermedad a partir de estudios de exactitud diagnóstica en otras poblaciones ya que esta no necesariamente es la prevalencia en nuestra población.

Algunos autores plantean que al aumentar la prevalencia en estudios diagnósticos también tendría un efecto sobre la sensibilidad de la herramienta, en el

específico caso de la radiología, al ser la enfermedad encontrada más frecuente que en la práctica diaria⁽²⁵⁻²⁷⁾. Lo anterior sería válido en el exclusivo caso en que el aumento de la prevalencia conlleve a un aumento en los casos graves. Por el contrario, si el espectro de enfermedad cambia a un escenario de screening poblacional, con una consecuente menor prevalencia de la enfermedad y probable presentación más leve, los resultados de la prueba diagnóstica serán más cercanos al de los pacientes sanos, por lo tanto, reduciendo su sensibilidad⁽²⁸⁾. Este y otros aspectos de la metodología de las investigaciones de pruebas diagnósticas en imagenología no han sido completamente estudiados. Profundizar en estas temáticas está fuera de los objetivos del presente artículo, no obstante, será importante estar alerta a las limitaciones y avances en esta materia.

Conclusión

En el análisis de los estudios de exactitud diagnóstica, la tabla de contingencia, y los estadígrafos que de esta podemos calcular (sensibilidad, especificidad, valor predictivo positivo, valor predictivo negativo, razones de verosimilitud), son muy importantes en cuánto a la caracterización y valoración de las pruebas diagnósticas. Entender su significado y forma de construcción es fundamental, no solo para la valoración crítica de los estudios de exactitud diagnóstica, sino que en su aplicación en el quehacer clínico diario y en el desarrollo de investigaciones originales⁽²²⁾. En este último punto, existen esfuerzos internacionales en la estandarización del reporte y análisis crítico de estudios de exactitud diagnóstica indispensables de ser considerados y adoptados⁽¹⁻⁵⁾.

Bibliografía

- Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. The STARD statement for reporting studies of diagnostic accuracy: Explanation and elaboration. *Ann Intern Med* 2003; 138: W1-12.
- Pauker SG, Kassirer JP. The threshold approach to clinical decision making. *N Engl J Med* 1980; 302(20): 1109-1117.
- Graham T. Are you a "risk-minimizer" or a "test minimizer"? *Can Fam Physician* 2000; 46: 29-30, 3-5.
- Capurro D, Rada G. [The diagnostic process]. *Rev Med Chil* 2007; 135(4): 534-538.
- Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011; 155(8): 529-536.
- Valenzuela D L, Cifuentes A L. Validez de estudios de tests diagnósticos. *Rev Med Chil* 2008; 136.
- Steurer J, Fischer JE, Bachmann LM, Koller M, Riet Gt. Communicating accuracy of tests to general practitioners: a controlled study. *BMJ* 2002; 324: 824-826.
- Canals M. Ciencia e imagenología: Análisis cuantitativo de la investigación radiológica. *Rev Chil Radiol* 2008; 14(4): 221-226.
- Busel MD, Silva FC. Radiología basada en la evidencia: Estrategia conceptual focalizada para la práctica de la imagenología. *Rev Chil Radiol* 2004; 10: 109-117.
- Altman DG, Bland JM. Diagnostic tests 1: Sensitivity and specificity. *BMJ* 1994; 308(6943): 1552.
- Cerda LJ, Cifuentes AL. Uso de tests diagnósticos en la práctica clínica (Parte 1): Análisis de las propiedades de un test diagnóstico. *Rev Chilena Infectol* 2010; 27.
- Akobeng AK. Understanding diagnostic tests 1: sensitivity, specificity and predictive values. *Acta Paediatr* 2007; 96: 338-341.
- von Kummer R, Bourquain H, Bastianello S, Bozzao L, Manelfe C, Meier D, et al. Early prediction of irreversible brain damage after ischemic stroke at CT. *Radiology* 2001; 219(1): 95-100.
- Deeks JJ, Altman DG. Diagnostic tests 4: likelihood ratios. *BMJ* 2004; 329: 168-169.
- Black WC, Armstrong P. Communicating the significance of radiologic test results: the likelihood ratio. *AJR Am J Roentgenol* 1986; 147(6): 1313-1318.
- Akobeng AK. Understanding diagnostic tests 2: likelihood ratios, pre- and post-test probabilities and their use in clinical practice. *Acta Paediatr* 2007; 96: 487-491.
- Fagan TJ. Letter: Nomogram for Bayes theorem. *N Engl J Med* 1975; 293: 257.
- Whiting PF, Sterne JA, Westwood ME, Bachmann LM, Harbord R, Egger M, et al. Graphical presentation of diagnostic information. *BMC Med Res Methodol* 2008; 8: 20.
- Altman DG, Bland JM. Diagnostic tests 3: receiver operating characteristic plots. *BMJ* 1994; 309(6948): 188.
- Glas AS, Lijmer JG, Prins MH, Bossel GJ, Bossuyt PM. The diagnostic odds ratio: a single indicator of test performance. *J Clin Epidemiol* 2003; 56(11): 1129-1235.
- Berry CC. A tutorial on confidence intervals for proportions in diagnostic radiology. *AJR Am J Roentgenol* 1990; 154(3): 477-480.
- Greenhalgh T. How to read a paper. Papers that report diagnostic or screening tests. *BMJ* 1997; 315: 540-543.
- Medina LS, Blackmore CC, Applegate KE. Evidence-Based Imaging: Principles. In: Medina LS, Sanelli PC, Jarvik JG, editors. *Evidence-Based Neuroimaging Diagnosis and Treatment*. 1st ed. New York. Springer 2013; 3-18.
- García Villar C. [Evidence-based radiology for diagnostic imaging: What it is and how to practice it]. *Radiologia* 2011; 53(4): 326-334.
- Boone D, Halligan S, Mallett S, Taylor SA, Altman DG. Systematic review: Bias in imaging studies - the effect of manipulating clinical context, recall bias and reporting intensity. *Eur Radiol* 2012; 22(3): 495-505.
- Egglin TK, Feinstein AR. Context bias. A problem in diagnostic radiology. *JAMA* 1996; 276(21): 1752-1755.
- Kundel HL. Disease prevalence and radiological decision making. *Invest Radiol* 1982; 17(1): 107-109.
- Irwig L, Bossuyt P, Glasziou P, Gatsonis C, Lijmer J. Designing studies to ensure that estimates of test accuracy are transferable. *BMJ* 2002; 324(7338): 669-671.