

Estadísticas para el clínico

Curvas ROC y su uso en radiología

Dra. Camila De La Barra E.

Residente de 3er año de Programa de Especialidad en Radiología
Facultad de Medicina Clínica Alemana – Universidad del Desarrollo

Dr. Esteban Hebel N.

Residente de 3er año de Programa de Especialidad en Radiología
Facultad de Medicina Clínica Alemana – Universidad del Desarrollo

Dr. Claudio Silva F-A

Médico radiólogo. Magister en Epidemiología Clínica
Departamento de Imágenes
Clínica Alemana de Santiago, Facultad de Medicina Clínica Alemana,
Universidad del Desarrollo, Santiago, Chile.

Contacto: cdelabarra@alemana.cl

Introducción

Al analizar una prueba diagnóstica, el primer paso a realizar es la evaluación de su exactitud. Las medidas de exactitud diagnóstica permiten identificar apropiadamente cuando una determinada enfermedad o condición se encuentra presente o no, en comparación a otro método considerado como el estándar de referencia ⁽¹⁾. Existen parámetros ampliamente difundidos como sensibilidad, también conocida como razón de verdaderos positivos, que mide la proporción de estudios positivos que son correctamente identificados como tales ⁽¹⁻³⁾ y especificidad, también conocida como la razón de verdaderos negativos, que mide la proporción de estudios negativos correctamente identificados como tales ⁽¹⁻³⁾. Tanto la sensibilidad como la especificidad son características independientes de la prevalencia de la enfermedad o condición dentro de la población, al ser valores intrínsecos de la prueba diagnóstica ⁽²⁾.

Una forma convencional de describir una prueba diagnóstica es por medio de una tabla de doble entrada ⁽²⁾

o tabla "2x2" (Tabla 1), que se obtiene cuando los resultados son registrados de forma dicotómica (positivos/negativos). Como se observa en la Tabla 1, las columnas representan el verdadero estatus de la enfermedad, que es evaluado por medio del estándar de referencia ⁽²⁾. Las filas demuestran de forma dicotómica los resultados del test que se está comparando. De este modo, se define por sensibilidad como $a/(a+c)$ y especificidad como $d/(d+b)$.

Es importante recordar que tanto la sensibilidad como la especificidad son directamente dependientes del punto de corte elegido para definir un estudio como positivo o negativo.

Al mover el umbral la sensibilidad y la especificidad se modifican de manera inversa, mientras la sensibilidad aumenta, la especificidad disminuye y viceversa ⁽⁴⁾.

Los análisis convencionales utilizan la sensibilidad y especificidad de una prueba diagnóstica como los índices primarios de su exactitud, dado que estos indicadores son independientes de la prevalencia de enfermedad. Sin

embargo, usar exclusivamente la sensibilidad y especificidad como única medida de exactitud es problemático, ya que como se mencionó previamente, estos valores dependen de un punto de corte elegido arbitrariamente para definir un resultado como positivo ⁽²⁾, y la modificación de este punto de corte tiene como efecto un cambio en el valor de estos parámetros. El uso de curvas ROC permite evaluar en presencia de una variable continua, la efectividad de distintos puntos de corte y globalmente analizar la fortaleza de la prueba en evaluación ⁽²⁾.

Curvas ROC

Las curvas ROC (del inglés *Receiver Operating Characteristic*) son la representación gráfica cualitativa que permite evaluar el resultado dicotómico de un estándar de referencia contra un instrumento de diagnóstico con umbrales variables de discriminación. Desarrollada originalmente para la

detección de objetos enemigos en el campo de batalla por medio de radar ⁽²⁾, fue propuesta para ser incorporada en la práctica radiológica por Lusted en 1971 ⁽⁵⁾. Se representan como la sensibilidad (eje y) versus la tasa de falsos positivos, TFP (o 1-especificidad, ver Tabla 1) (eje x). Cada punto del gráfico se obtiene al cotejar la sensibilidad y TFP obtenidos usando diferentes puntos de corte ^(3,6). Las coordenadas se conectan por líneas para construir la curva ROC empírica ⁽³⁾. La Figura 1 muestra la elaboración de una curva ROC a partir de los datos obtenidos de un estudio con radiografías de tórax en 70 pacientes con nódulos pulmonares ⁽⁶⁾ (ver Tabla 2). Una curva ROC comienza en las coordenadas (0,0), que representan el umbral más estricto donde todos los resultados son negativos para enfermedad y termina en las coordenadas (1,1) donde el umbral es el más laxo y todos los resultados son positivos ⁽³⁾.

Tabla 1. Frecuencia de resultado de prueba para n_1 pacientes con enfermedad y n_2 pacientes sin enfermedad ¹

Resultado de prueba diagnóstico	ESTADO DE LA ENFERMEDAD	
	Presente	Ausente
Positivo	a (VP)	b (FP)
Negativo	c (FN)	d (VN)
Total	$n_1 = a + c$	$n_2 = b + d$

Resumen de índices de prueba de desempeño	
TVP = Tasa de Verdaderos Positivos (Sensibilidad) = VP / (VP + FN) = a / (a + c)	
TFN = Tasa de Falsos Negativos (image006.pngSensibilidad) = FN / (VP + FN) = c / (a + c)	
TVN = Tasa de Verdaderos Negativos (Especificidad) = VN / (VN + FP) = d / (b + d)	
TFV = Tasa de Falsos Positivos (image006.png Especificidad) = FP / (VN + FP) = b / (b + d)	

Tabla 2. Sensibilidad, especificidad, y TFP para el diagnóstico de nódulos pulmonares solitarios malignos en cada nivel de corte a partir del estudio de radiografías de tórax ²

Prueba positiva si es mayor que o igual a	Sensibilidad	Especificidad	TFP
2: Probablemente benigno	0,912 (31/34)	0,222 (8/36)	0,778
3: Posiblemente maligno	0,794 (27/34)	0,528 (19/36)	0,472
4: Probablemente maligno	0,676 (23/34)	0,750 (27/36)	0,250
5: Definitivamente maligno	0,206 (7/34)	0,944 (34/36)	0,056

¹Hajian-Tilaki K (2013) / ² Park SH, Goo JM, Jo C-H (2004).

Medidas de efectividad

Se han propuesto diferentes medidas de efectividad para las curvas ROC, dentro de las que se encuentra el área bajo la curva ROC (ABC o AUC en inglés) y el área parcial bajo la curva ROC (pABC) ^(2,3,6).

ABC: una de las medidas de resumen más populares es el ABC, con valores que fluctúan entre 0.0 y 1.0. Un test diagnóstico con un ABC de 1.0 tiene una precisión perfecta, con 1.0 de sensibilidad y 0.0 de FP. La línea que conecta los puntos (0,0) y (1,1) con un ángulo de 45°, es la llamada "diagonal de probabilidad" y tiene una ABC de 0.5 ⁽⁴⁾. Ésta representa la curva ROC de un test diagnóstico sin habilidad para distinguir entre los enfermos y los libres de la enfermedad o condición (comparable con resultado obtenido por azar). Una curva ROC que cae por sobre la "diagonal de probabilidad" tiene alguna habilidad para discriminar entre sanos y enfermos ⁽⁴⁾. Mientras más se acerque la ABC a 1.0, mejor es su capacidad diagnóstica (ver Figura 2). La ABC se puede interpretar como la probabilidad de que un sujeto enfermo, elegido azarosamente, sea clasificado con mayor probabilidad como enfermo versus un sujeto libre de la condición elegido al azar ^(2,4). La otra interpretación es el valor medio de la sensibilidad para todos los valores posibles de especificidad. El uso del ABC es una excelente medida de resumen de la efectividad de una prueba diagnóstica debido a que es independiente de la prevalencia de la enfermedad, considera todos los puntos de corte posibles y finalmente permite la comparación visual de dos o más pruebas diagnósticas en el mismo gráfico ⁽²⁾.

pABC: a pesar de la gran utilidad estadística del ABC se puede argumentar que gran parte del área se ubica hacia la derecha, donde incrementa la TFP ⁽²⁾. A su vez, dos test diagnósticos pueden tener la misma ABC, pero uno de ellos puede tener una sensibilidad mayor para el rango de TFP clínicamente relevante ⁽³⁾. La Figura 3 muestra las curvas ROC de dos test diagnósticos con igual ABC, pero que en el rango de TFP clínicamente relevante (por ejemplo 0.0-0.2), son diferentes, con la curva A demostrando mayor sensibilidad.

Ejemplos de curvas ROC en radiología

Se presenta a continuación un ejemplo de aplicación de curvas ROC en Radiología. El ejemplo se basa en el estudio de Umano y col. ⁽⁷⁾, donde se busca determinar si frente a una lesión ósea blástica identificada por tomografía computada (TC) puede hacerse el diagnóstico fiable de metástasis blástica no tratada versus foco de enostosis. Ellos plantearon que existe un umbral de atenuación media y máxima, medidas en unidades de Hounsfield (UH), bajo el cual la lesión puede ser caracterizada como metástasis blástica. Como estándar de referencia para la enfermedad ósea metastásica, se utilizó el resultado histológico y/o la progresión en TC de seguimiento. En el caso de las enostosis se utilizaron como estándar de referencia varios criterios incluyendo el estudio histológico, contar con el diagnóstico de osteopoiquiosis (caracterizada por enostosis múltiples), esclerosis tuberosa o esclerodermia (enfermedades sistémicas con asociación conocida con enostosis), o mediante las siguientes características imagenológicas: lesiones de 2 mm a 2 cm, redondeadas u ovaladas, homogéneamente escleróticas, presentes

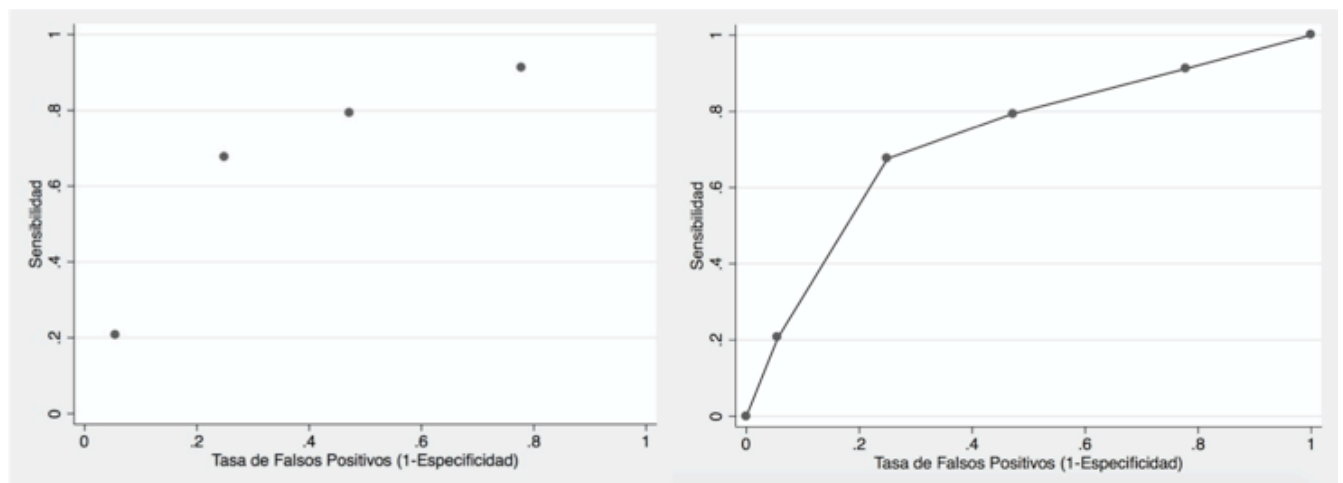


Figura 1. Curvas ROC de un estudio correspondiente a 70 pacientes con nódulos pulmonares solitarios, donde se representa la sensibilidad en el eje de las coordenadas (Y) versus la tasa de falsos positivos o 1 - especificidad en el eje de las abcisas (X). El segundo gráfico se construye a unir los puntos

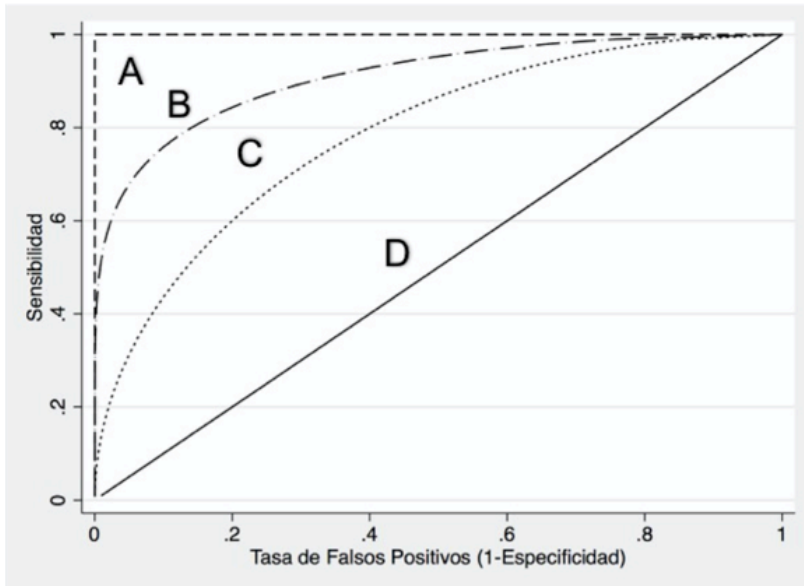


Figura 2. 4 curvas ROC con diferentes valores de área bajo la curva. Un estudio perfecto (A) tiene un área bajo la curva igual a 1. La diagonal de probabilidad (D) tiene un área bajo la curva de 0,5. Las curvas ROC permiten distinguir entre dos estudios respecto de su capacidad diagnóstica. En este ejemplo el estudio representado por la curva B es mejor que aquel representado por la curva C.

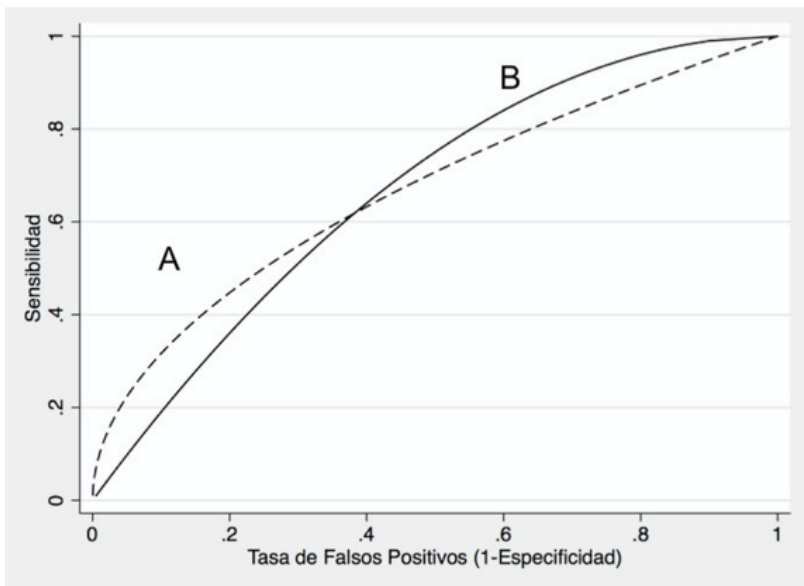


Figura 3. 2 curvas representadas A y B en el ejemplo, con áreas bajo las curvas idénticas. A pesar de esto, no son iguales, donde en el rango de alta sensibilidad el estudio B es mejor que el estudio A, mientras que el rango de baja sensibilidad el estudio A es mejor que el estudio B.

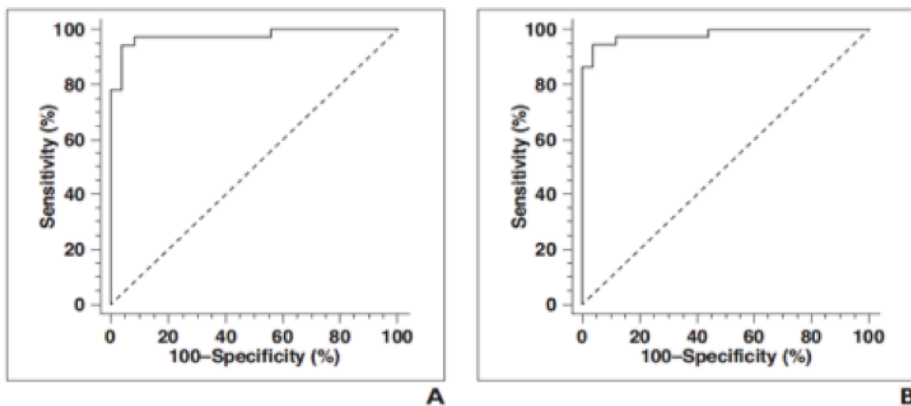


Figura 4. Reproducida del estudio de Ulano y col.⁵ El ABC del valor medio de atenuación en TC para distinguir entre enostosis y metástasis blásticas fue de 0,976 (curva B).

exclusivamente en pacientes menores de 30 años y sin primarios conocidos. La medición de la atenuación se realizó separadamente por dos radiólogos, quienes evaluaron los exámenes completos y seleccionaron en cada uno la lesión de mayor tamaño para medir la atenuación media y máxima. En total se analizaron 153 metástasis (de 25 pacientes) y 126 enostosis (de 37 pacientes). El ABC del valor medio de atenuación en TC para distinguir entre enostosis y metástasis blásticas fue de 0,982 (ver Figura 4). Usando un corte 885 UH para atenuación media se obtuvo una sensibilidad de 95% y especificidad de 96%. El ABC de la máxima atenuación fue de 0,976 (ver Figura 4), y con un corte de 1060 UH para atenuación máxima la sensibilidad y especificidad fue de 95% y 96% respectivamente. Estos valores de ABC cercanos a 1, indican que esta prueba diagnóstica tiene una elevada precisión para distinguir entre metástasis blástica y enostosis, acercándose al test perfecto.

Con estos resultados se podría plantear que frente a una lesión ósea blástica de apariencia indeterminada y un contexto clínico ausente o no categórico, el uso de estos valores de corte de atenuación por TC podrían actuar como una herramienta adicional para hacer el diagnóstico, sin necesidad de otros estudios o biopsia, evitando procedimientos invasivos.

Conclusión

El uso de curvas ROC tiene múltiples ventajas como medida de la exactitud de un test diagnóstico, entre ellas

el incluir todos los puntos de corte, mostrar la relación entre sensibilidad y especificidad, ser independiente de la prevalencia de la enfermedad y permitir calcular diferentes medidas de resumen a partir de ellas (ABC, pABC). En las últimas décadas la utilización de curvas ROC se ha incrementado significativamente en la literatura médica, especialmente en epidemiología, para la evaluación de la habilidad diagnóstica de biomarcadores y en imágenes diagnósticas, para la clasificación de sanos y enfermos. Por todo lo anterior, es fundamental para el médico y en particular para el radiólogo general conocer sus características, ventajas, interpretación y aplicaciones.

Referencias

1. Pepe MS. *The statistical evaluation of medical tests for classification and prediction* [Internet]. 2004 [citado 1 de febrero de 2017]. Disponible en: <http://public.eblib.com/choice/publicfullrecord.aspx?p=1573145>
2. Hajian-Tilaki K. Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation. *Casp J Intern Med*. 2013;4(2):627-35.
3. Obuchowski NA. Fundamentals of Clinical Research for Radiologists ROC Analysis. *American Journal of Roentgenology*. 2005;184 (2): 364-372.
4. Obuchowski NA. Receiver Operating Characteristic Curves and Their Use in Radiology. *Radiology* 2003 Oct;229(1):3-8.
5. Lusted LB. Signal detectability and medical decision-making. *Science*. 1971 Mar 26;171(3977):1217-9.
6. Park SH, Goo JM, Jo C-H. Receiver Operating Characteristic (ROC) Curve: Practical Review for Radiologists. *Korean J Radiol*. 2004;5(1):11-8.
7. Ulano A, Bredella MA, Burke P, et al. Distinguishing Untreated Osteoblastic Metastases From Enostoses Using CT Attenuation Measurements. *American Journal of Roentgenology*. 2016;207: 362-368.