



Artículo Especial

Metodología científica para cirugía vascular

Validación de test diagnósticos (segunda parte). Test cuantitativos

Validation of diagnostic tests (part II). Quantitative tests

Antonio Martín Conejero

Hospital Clínico San Carlos, Madrid

Resumen

En esta segunda parte sobre la validación de los test diagnósticos vamos a estudiar cómo analizar la bondad de los test diagnósticos cuantitativos. El análisis de estos test se realiza a partir de las curvas ROC (*Receiver Operating characteristic Curve*). Constituye un método para determinar la exactitud diagnóstica de test que utilizan escalas continuas. A partir de estas curvas analizaremos dos conceptos: la AUC (*Area Under Curve*) y el índice de Youden, que nos ayudará a una mejor interpretación de la validez de una prueba diagnóstica y su punto de corte.

Palabras clave:

Test cuantitativo.
Curva ROC. AUC.
Índice de Youden.

Abstract

In this second part on the validation of diagnostic tests we are going to study how to analyze the goodness of quantitative diagnostic tests. The analysis of these tests is performed from the ROC curves (*Receiver Operating characteristic Curve*). It constitutes a method to determine the diagnostic accuracy of tests that use continuous scales. From these curves we will analyze two concepts: the AUC (*Area Under Curve*) and the Youden index, which will help us to better interpret the validity of a diagnostic test and its cut-off point.

Keywords:

Quantitative test. ROC
curve. AUC. Youden's J
statistic.

Recibido: 31/10/2022 • Aceptado: 31/10/2022

Conflictos de interés: el autor declara no tener conflictos de interés.

Martín Conejero A. Validación de test diagnósticos (segunda parte). Test cuantitativos. *Angiología* 2023;75(3):181-185

DOI: <http://dx.doi.org/10.20960/angiologia.00475>

Correspondencia:

Antonio Martín Conejero. Hospital Clínico San Carlos.
C/ del Prof. Martín Lagos, s/n. 28040 Madrid
e-mail: amartinconejero.cto@gmail.com

INTRODUCCIÓN

En un artículo anterior (1) hemos analizado cómo podemos evaluar la bondad de una prueba diagnóstica cualitativa frente a una prueba de referencia (*gold standard*). Sin embargo, en la práctica diaria de todos nosotros empleamos test diagnósticos cuantitativos, como pueden ser las velocidades en el eco Doppler que valora una estenosis carotídea, la cifra de dímero D en el diagnóstico de TVP, el diámetro de un aneurisma de aorta como predictor de crecimiento de dicho aneurisma o el índice tobillo-brazo como predictor de enfermedad cardiovascular. Es decir, en este artículo vamos a recordar cómo analizar la bondad de los test diagnósticos cuantitativos.

El análisis de estos test se realiza a partir de las curvas ROC (*Receiver Operating Characteristic Curve*). Constituye un método para determinar la exactitud diagnóstica de los test que utilizan escalas continuas. Se utilizan con tres propósitos específicos:

1. Determinar, de entre varios test cuantitativos, cuál es el mejor en el diagnóstico de una enfermedad. Por ejemplo, en el diagnóstico de estenosis carotídea severa, ¿es mejor la velocidad picosistólica que la velocidad telediastólica?
2. Evaluar la capacidad discriminativa del test diagnóstico. Es decir, si tenemos un test cuantitativo determinado, ¿es buen test? Por ejemplo, ¿es buen test el dímero D en el diagnóstico de TVP?
3. Calcular el mejor valor del test cuantitativo para separar al paciente del sujeto sano. Por ejemplo, identificar que el valor 300 es el que determina la diferencia entre estenosis severa y no severa en la anastomosis de un *bypass*.

Este artículo pretende ayudar a entender cómo las curvas ROC permiten validar test diagnósticos cuantitativos. Para ello interpretaremos juntos los resultados de un artículo recientemente publicado que relaciona los niveles de dímero D como biomarcador de aneurisma de aorta abdominal (AAA) (2).

LA CURVA ROC

Esa gran desconocida

Las curvas ROC surgen de lo aprendido en situaciones tan desagradables como los conflictos bélicos.

Aparecida una señal en un radar, ¿podríamos a partir de la frecuencia del sonido (variable cuantitativa) saber si la señal corresponde a una bomba (enfermo) o a cualquier otra cosa (sujeto sano)?

A partir de la situación planteada con el radar, deduciremos siempre que una curva ROC en medicina siempre necesita 2 test diagnósticos. Uno, cuantitativo, que es el que queremos validar. Es el test diagnóstico al que “ponemos a prueba”. El otro, el *gold standard* o prueba de referencia, que es dicotómico: enfermo/sano. Si seguimos utilizando como ejemplo el eco Doppler en la estenosis carotídea, el resultado velocimétrico del Doppler sería la variable cuantitativa continua que queremos validar, mientras que el angio TAC (estenosis sí/no) sería la prueba de referencia de confirmación.

Antes de entrar con la curva ROC quiero recordar que los valores de un test cuantitativo en el diagnóstico de enfermedad pueden variarse arriba/abajo, modificando así la sensibilidad y la especificidad del test. Veamos este ejemplo. Si eres un médico de cierta edad, recordarás que en la carrera nos dijeron que un sujeto se consideraba diabético cuando tenía en ayunas una glucemia superior a 140 mg/dL. Sin embargo, esto suponía un problema. Muchos sujetos con glucemias en ayunas algo menores también eran diabéticos, por lo que con el valor de corte de 140 se nos escapaban (eran falsos negativos, FN). Para que no se nos escapen estos pacientes “límite”, se modificó el punto de corte de tal forma que en la actualidad se define diabetes como un valor de glucemia en ayunas de 126 mg/dL. Es decir, bajando el punto de corte se aumenta la capacidad del test de detectar enfermos; esto es, se ha aumentado su sensibilidad. El precio que se paga con esta maniobra es que es posible que algún sujeto sano se nos cuele con una glucemia algo mayor de 126. Es decir, la mayor sensibilidad implica un mayor número de falsos positivos (FP). Podríamos decir, por tanto, que cada valor de la variable cuantitativa *glucemia* tiene su propio par, sensibilidad-especificidad, de forma que un valor de glucemia menor será más sensible y menos específico y un valor alto de glucemia será más específico, pero menos sensible.

La curva ROC se construye teniendo en cuenta esta variación en la sensibilidad y en la especificidad de cada valor de la variable cuantitativa. La representación se realiza sobre un eje de coordenadas.

En el eje de ordenadas figura la sensibilidad (S) del test diagnóstico y en el de abscisas, el complementario de la especificidad (1-E), que es la tasa de falsos positivos (3). Es decir, en el eje de ordenadas se tiene, del total de sujetos enfermos, a los verdaderos positivos mientras que en el de abscisas, del total de enfermos, a los falsos positivos. Cada sujeto de nuestra muestra tiene un valor concreto de la variable cuantitativa que, como hemos dicho, va a tener su propia S y su propia 1-E. Si unimos los valores del test cuantitativo de cada uno de los sujetos de la muestra, obtendremos una "curva", como la representamos en azul en la figura 1.

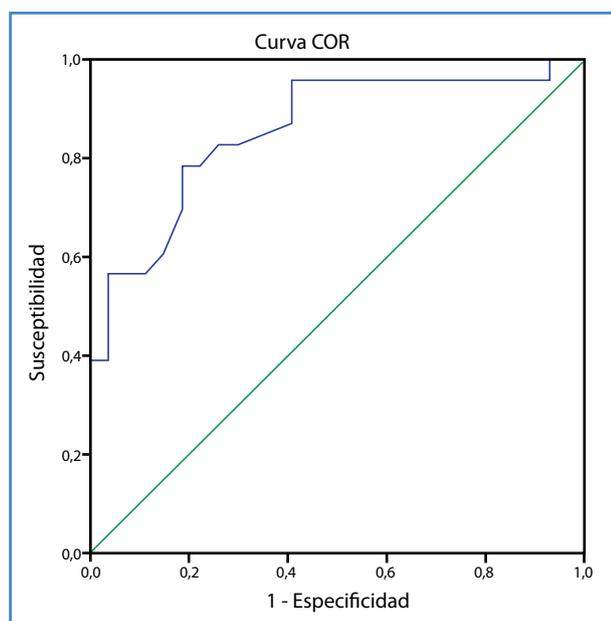


Figura 1. Curva ROC.

Si nos fijamos en la gráfica, los valores que queden arriba a la izquierda serán los que tiene más sensibilidad y más especificidad (ojo, el eje de abscisas es 1-E), por eso nos gustan las gráficas que se aproximan a este ángulo superior izquierdo. Muchas veces, cuando vemos en un artículo una representación de una curva ROC vemos una diagonal (en verde en nuestra figura). Esta diagonal divide el gráfico en dos mitades iguales, por lo que el área del gráfico que queda por encima es el 50 % y por debajo, el otro 50 %. Dicho de otra forma, si tenemos un test diagnóstico que coincidiera con esta diagonal, deberíamos concluir que la probabilidad de acierto del test es del 50 %. O sea, el peor valor posible para una prueba diagnóstica porque acertaría con la misma probabilidad que la de tirar una moneda al aire.

Ahora que hemos visto cómo se forma una curva ROC, veamos la información que nos aporta sobre los test cuantitativos que queremos validar.

EL ÁREA BAJO LA CURVA (AUC)

La decisión: elegir el mejor test entre varios cuantitativos

Seguramente en diferentes ocasiones nuestro lector se ha enfrentado a situaciones en las que contamos con más de un test cuantitativo para hacer el diagnóstico de una enfermedad. Un ejemplo clásico de nuestra especialidad es conocer qué parámetro, velocidad picosistólica (VS) o telediastólica (VD) es mejor en el diagnóstico de estenosis carotídea severa.

En un estudio que publicó nuestro grupo (4) obtuvimos la curva ROC que puede verse en la figura 2. Como podemos apreciar, de las 3 variables cuantitativas analizadas, la VD fue la que presentó mayor área bajo la curva. *A priori*, por tanto, el mejor test diagnóstico, con un área del 94,4 %.

¿Podemos afirmar que estos parámetros velocimétricos se comportarán de igual manera en otra muestra diferente a la nuestra? Probablemente no haya muchas diferencias en el comportamiento de estas velocidades en otra muestra, pero la mejor forma

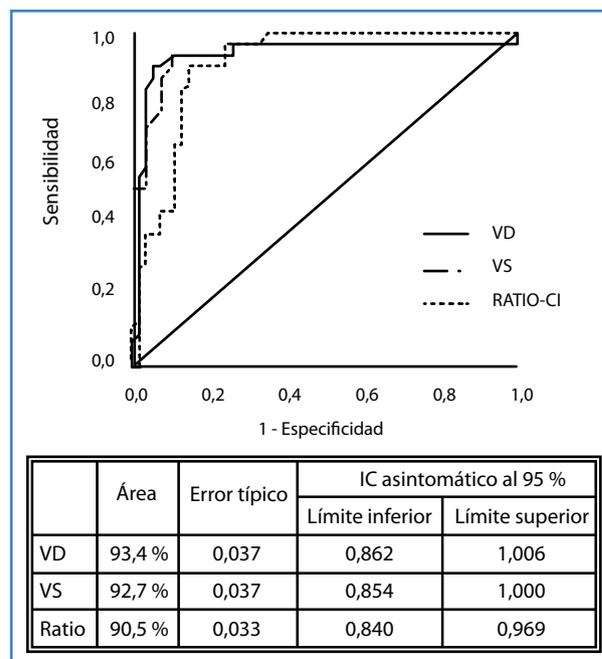


Figura 2. Comparativa entre diferentes variables cuantitativas.

de dar esta aproximación es mediante los intervalos de confianza, que también vemos en la figura 2. Vamos a fijarnos en el intervalo de confianza (IC) de la VD.

Los datos que muestra la figura son para la VD, AUC (*Area Under Curve*) del 93,4 % e IC 95 % (86,2-100 %). ¿Qué es este IC y cómo lo debemos interpretar? De forma muy resumida podemos afirmar que los intervalos de confianza son la forma que tenemos de inferir a la población los datos que se obtienen en la muestra. Como nunca tendremos certeza de que lo que inferimos es correcto, se hace habitualmente una aproximación del 95 %. Así, los datos referidos a la VD deben ser interpretados de la siguiente forma: el área bajo la curva de la VD en nuestra muestra es del 93,4 %, mientras que en la población podríamos decir que el AUC para la VD está entre el 86 % y el 100 %, con una probabilidad de que esta afirmación sea correcta del 95 %. Es decir, la probabilidad de que nos equivoquemos cuando decimos que el AUC poblacional está entre el 86 % y el 100 % es solo del 5 %.

Ahora que sabemos lo que significa un IC 95 %, las preguntas que debemos hacernos son: ¿es bueno este intervalo de confianza? ¿Nos gusta? En el caso de las curvas ROC, nos gustarán los test cuyo IC 95 % no incluye el valor del empate o valor nulo que, como hemos dicho anteriormente, es la diagonal del 50 %. Por tanto, podemos decir que nuestro IC 95 % (86,2-100 %) nos parece una buena estimación poblacional porque no incluye el valor del 50 %.

EL ÍNDICE DE YOUTEN

¿Qué valor de mi test debo elegir?

Como hemos visto, las curvas ROC nos dicen qué test cuantitativo es mejor y qué bondad tiene según el AUC. En este artículo (2) se ha analizado si el dímero D podría ser un buen predictor de AAA en el grupo de pacientes con EAP. En la figura 3 se muestra la curva ROC que obtuvieron.

Podemos afirmar que el dímero D se comporta como un buen predictor, ya que el AUC es claramente superior al 50 % de la diagonal. Nos falta sin embargo una información muy importante: ¿qué valor del dímero D es el mejor punto de corte a partir del cual

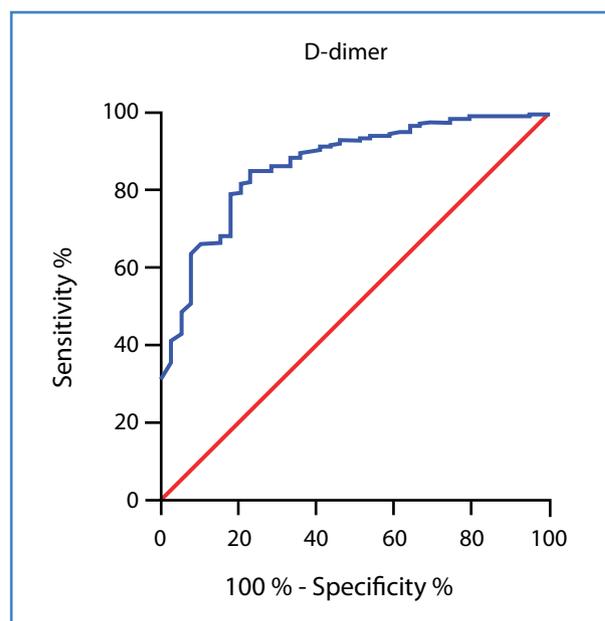


Figura 3. Curva ROC del D-dímero como predictor de AAA.

separar al sujeto sano del enfermo? Es decir, ¿hasta qué punto de la variable cuantitativa decimos que viene de un sujeto sano y cuándo de un enfermo?

Parece lógico pensar que elegiremos el test con alta sensibilidad y especificidad. Existen diferentes parámetros que miden la idoneidad de estos puntos de corte. Uno de ellos es el índice de Youden, que se define como la $S + E - 1$. Este parámetro, aunque es criticado por algunos autores porque no discrimina si el punto de corte es bueno a expensas de la sensibilidad o de la especificidad, es uno de los más empleados porque permite objetivamente decidir si el punto de corte es bueno o no.

El índice de Youden oscila de -1 a +1 (+1 es el mejor punto de corte posible), ya que se obtiene de una sensibilidad y de una especificidad teóricas del 100 %.

En la tabla I se detallan los índices para algunos de los valores obtenidos en la curva ROC de la figura 3.

En la tabla I puede apreciarse que el valor del dímero D de 0,6750 es el que tiene mejor Youden y, por tanto, el que debería emplearse como valor que divide dicotómicamente al sujeto como enfermo o sano.

Como hemos visto, la interpretación de la bondad de los test diagnósticos cuantitativos es sencilla. El análisis de la curva ROC y el cálculo del índice de Youden servirán para que el lector identifique si un test cuantitativo es el adecuado para el diagnóstico de una determinada enfermedad.

Tabla I. Valores del índice de Youden de puntos de corte de dímero D

Dímero D	Sensibilidad	Especificidad	Índice de Youden
0,6750	0,769	0,849	0,619
0,6600	0,769	0,845	0,614
0,5050	0,821	0,791	0,611
0,5550	0,795	0,816	0,611
0,6450	0,769	0,841	0,610
0,5400	0,795	0,812	0,607
0,4950	0,821	0,782	0,603

BIBLIOGRAFÍA

- Martín Conejero A. Validación de test diagnósticos (primera parte). Test categóricos. *Angiología* 2022;75(2):101-5. DOI: 10.20960/angiologia.00473
- Cai H, Pan B, Xu J, et al. D-Dimer Is a Diagnostic Biomarker of Abdominal Aortic Aneurysm in Patients with Peripheral Artery Disease. *Front Cardiovasc Med* 2022;9:890228. DOI: 10.3389/fcvm.2022.890228
- Manual CTO. Metodología Básica de la Investigación Clínica. Madrid: Grupo CTO; 2019.
- Martín-Conejero A, Sánchez-Hervás L, Moñux-Ducájú G, et al. Validación del eco Doppler carotídeo como diagnóstico único de la estenosis