

# EVIDENCIAS EN PEDIATRÍA

Toma de decisiones clínicas basadas en las mejores pruebas científicas  
[www.evidenciasenpediatria.es](http://www.evidenciasenpediatria.es)

## Fundamentos de medicina basada en la evidencia

### Pruebas diagnósticas con resultados continuos o politómicos. Curvas ROC

Molina Arias M<sup>1</sup>, Ochoa Sangrador C<sup>2</sup>

<sup>1</sup>Servicio de Gastroenterología. Hospital Infantil Universitario La Paz. Madrid (España).

<sup>2</sup>Servicio de Pediatría. Hospital Virgen de la Concha. Zamora (España).

Correspondencia: Manuel Molina Arias, [mma1961@gmail.com](mailto:mma1961@gmail.com)

---

**Palabras clave en inglés:** sensitivity; specificity; area under curve; ROC curve.

**Palabras clave en español:** sensibilidad; especificidad; área bajo la curva; curva ROC.

**Fecha de recepción:** 27 de enero de 2017 • **Fecha de aceptación:** 31 de enero de 2017  
**Fecha de publicación del artículo:** 8 de febrero de 2017

---

Evid Pediatr. 2017;13:12.

#### CÓMO CITAR ESTE ARTÍCULO

Molina Arias M, Ochoa Sangrador C. Pruebas diagnósticas con resultados continuos o politómicos. Curvas ROC. Evid Pediatr. 2017;13:12.

Para recibir Evidencias en Pediatría en su correo electrónico debe darse de alta en nuestro boletín de novedades en <http://www.evidenciasenpediatria.es>

---

Este artículo está disponible en: <http://www.evidenciasenpediatria.es/EnlaceArticulo?ref=2017;13:12>.

©2005-17 • ISSN: 1885-7388

# Pruebas diagnósticas con resultados continuos o politómicos. Curvas ROC

Molina Arias M<sup>1</sup>, Ochoa Sangrador C<sup>2</sup>

<sup>1</sup>Servicio de Gastroenterología. Hospital Infantil Universitario La Paz. Madrid (España).

<sup>2</sup>Servicio de Pediatría. Hospital Virgen de la Concha. Zamora (España).

Correspondencia: Manuel Molina Arias, mma1961@gmail.com

En números anteriores vimos cómo se valoraba el comportamiento de las pruebas diagnósticas cuando estas tenían como resultado un valor positivo o negativo. Calculábamos la sensibilidad (S) y especificidad (E) de la prueba<sup>1</sup>, los valores predictivos<sup>2</sup> y los cocientes de probabilidades<sup>3</sup>, todo ello examinado a saber la probabilidad posprueba.

Pues bien, existen pruebas diagnósticas en las que el resultado no es positivo o negativo, sino un valor cuantitativo de tipo continuo. Pensemos, por ejemplo, en la glucemia, el colesterol sérico, el número de neutrófilos totales, etc. En estos casos, la S y E de la prueba van a depender del punto de corte que consideremos por encima del cual la prueba será positiva y por debajo del cual será negativa.

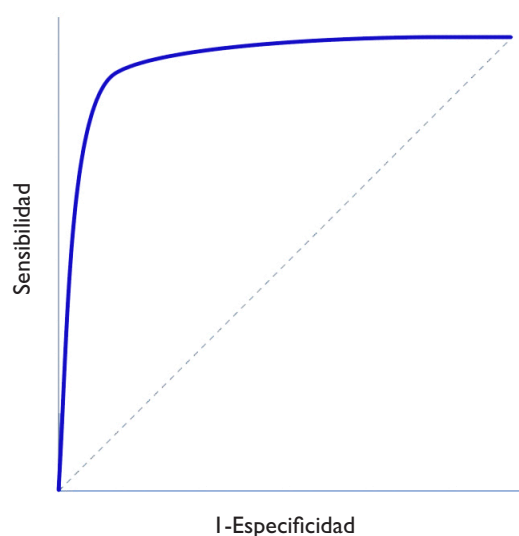
Veamos un ejemplo. Pensemos que utilizamos el valor de la procalcitonina (PCT) para distinguir si un lactante con fiebre sin foco tiene una infección vírica o bacteriana. Si elegimos un punto de corte muy bajo, a partir del cual consideremos que la infección es bacteriana, detectaremos la mayor parte de los niños con infección bacteriana (pocos tendrán la PCT por debajo de ese valor), pero estaremos diagnosticando de infección bacteriana muchos niños con infección vírica (falsos positivos [FP]). En este caso, la prueba será muy sensible, pero poco específica.

Por el contrario, si elegimos un punto de corte muy alto, nos equivocaremos muy poco cuando diagnosticuemos una infección bacteriana (pocas tendrán valores por debajo del punto de corte), pero se nos pasarán muchas que diagnosticaremos como víricas (falsos negativos [FN]). En este caso, la prueba tendrá poca sensibilidad y mucha especificidad.

Para solucionar el problema de saber cuál es el punto de corte que más nos conviene disponemos de una herramienta denominada curva de características operativas para el receptor, conocidas como curvas ROC<sup>4</sup> por sus siglas en inglés (*receiver operating characteristic*).

En la figura 1 se representa en ordenadas (eje y) la S y en abscisas el complementario de la E (1-E) y se traza una curva según la S y E de cada valor que se tome como posible punto de corte. Así, cada punto representa la probabilidad de diagnosticar correctamente a sanos y enfermos. La diagonal del gráfico representaría la "curva" si la prueba no tuviese capacidad discriminatoria.

FIGURA 1. REPRESENTACIÓN DE UNA CURVA ROC

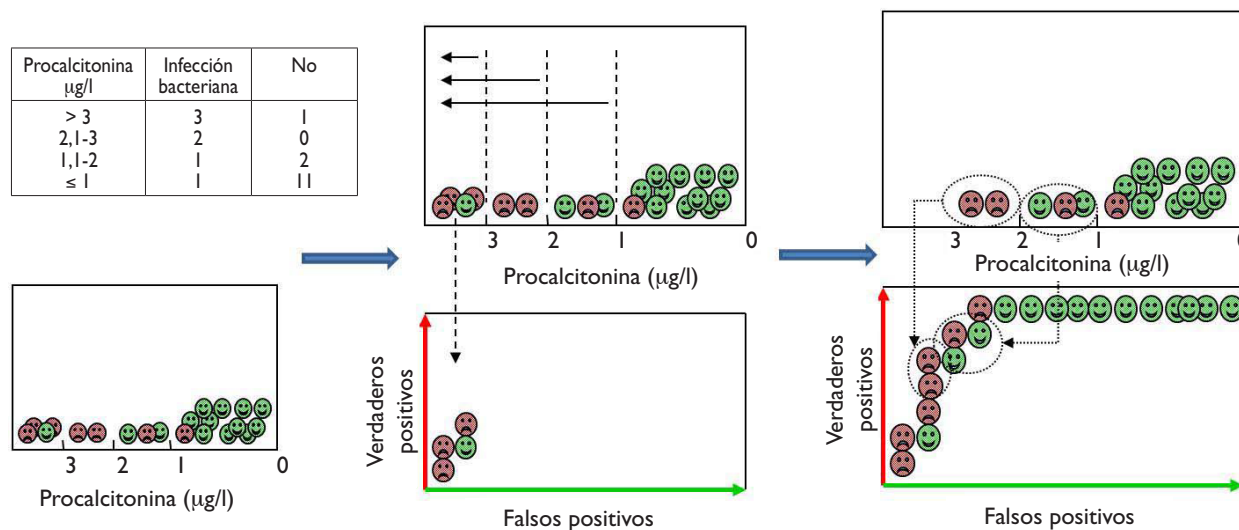


Veamos cómo construir una curva ROC con un ejemplo ficticio del uso de PCT para distinguir entre infección vírica y bacteriana, cuyos resultados aparecen en la tabla incluida en la figura 2. Para visualizar de forma gráfica cómo hacer una curva ROC, dentro de cada intervalo de los valores de PCT comenzamos a colocar los casos de infección bacteriana (verdaderos positivos) sobre el eje vertical (hacia arriba en el gráfico) y los casos de infección viral (falsos positivos) hacia la derecha en horizontal, tal como se muestra en la figura 2. En cada intervalo, los verdaderos positivos nos acercan a la esquina superior izquierda del gráfico, mientras que los falsos positivos nos alejan. Obtenemos así la curva para este ejemplo.

Desde un punto de vista numérico calcularíamos las parejas de S y E para cada uno de los posibles puntos de corte y los representaríamos gráficamente, tal como se muestra en la figura 3.

Como puede verse en el gráfico, la curva suele tener un segmento de gran pendiente donde aumenta rápidamente la S sin que apenas varíe la E: si nos desplazamos hacia arriba podemos aumentar la S sin que prácticamente nos aumenten los FP. Pero llega un momento en que nos acercamos a la parte plana. Si seguimos desplazándonos hacia la derecha llegará un

**FIGURA 2. EJEMPLO GRÁFICO DE CONSTRUCCIÓN DE LA CURVA ROC A PARTIR DEL DIAGNÓSTICO DE LOS PACIENTES PARA LOS DIFERENTES PUNTOS DE CORTE DE LA PRUEBA**



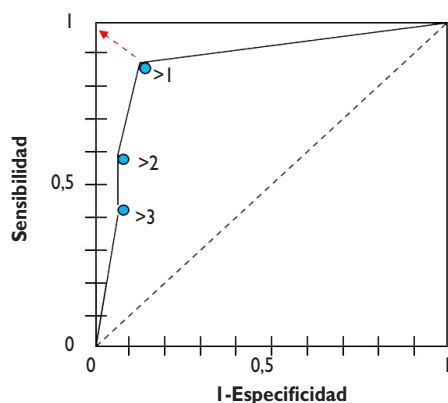
punto a partir del cual la S ya no aumentará más, pero comenzarán a aumentar los FP.

Así, podemos utilizar esta curva para calcular cuál es el punto de S y E que más nos convenga según nos interese primar una u otra. En general, en aquellos casos en que los inconvenientes de los FP sean menores que los de los FN nos interesará una prueba muy sensible, por lo que elegiremos puntos de corte situados más a la derecha de la curva. Por otro lado,

cuando sea preferible tener FN que FP nos interesará que la prueba sea más específica, por lo que elegiremos puntos de corte más a la izquierda (menos FP). Por último, en los casos en que queramos maximizar S y E, el mejor punto de corte será el punto más próximo al ángulo superior izquierdo de la gráfica<sup>5</sup>.

**FIGURA 3. REPRESENTACIÓN GRÁFICA DE LAS PAREJAS DE S Y E PARA CONSTRUIR LA CURVA ROC DE LA PRUEBA DIAGNÓSTICA**

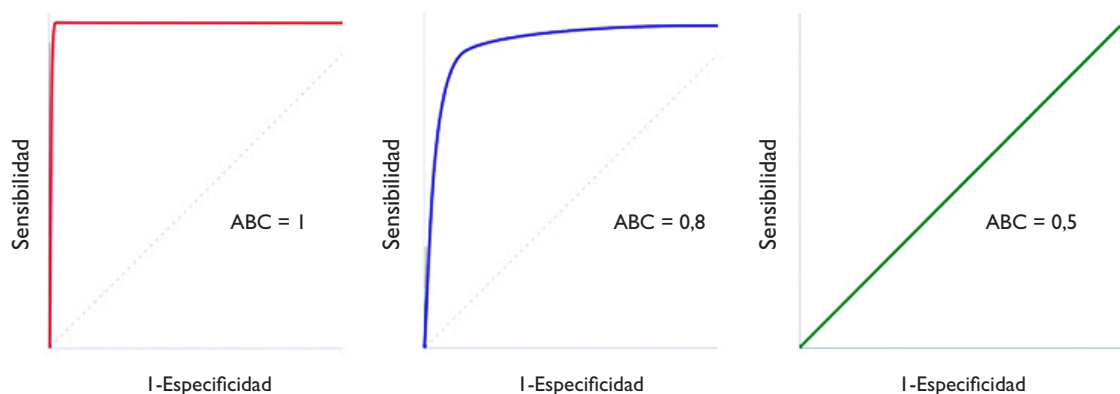
Procalcitonina	Sensibilidad	Especificidad
>3	0,43	0,92
>2	0,57	0,92
>1	0,86	0,87



Un parámetro de interés es el área bajo la curva (ABC), que nos representa el comportamiento global de la prueba diagnóstica, la probabilidad de que clasifique correctamente al paciente al que se le practique, considerando todos los puntos de corte posibles. Las curvas ROC se representan siempre como un cuadrado de  $1 \times 1$  de lado. Una prueba ideal con S y E del 100% sigue el marco del gráfico y tiene un área bajo la curva de 1: siempre acierta. Sin embargo, esta situación no suele verse en la práctica habitual, ya que es excepcional encontrar una prueba con S y E de 100%. En clínica, una prueba cuya curva ROC tenga un ABC > 0,9 se considera muy exacta, entre 0,7-0,9 de exactitud moderada y entre 0,5-0,7 de exactitud baja. Así, la capacidad discriminadora de la prueba disminuye al disminuir el ABC. Cuando la curva coincide con la diagonal, el ABC es igual a 0,5, lo que significa que la capacidad discriminadora es nula: obtendríamos la misma probabilidad de acertar realizando la prueba o tirando una moneda al aire. Valores por debajo de la diagonal (ABC < 0,5) se corresponden con un error de clasificación de sanos y enfermos: la capacidad de la prueba es tan baja que toma a los sanos por enfermos, y viceversa. En la figura 4 podemos ver ejemplos de curvas con distintas ABC.

De manera ideal, debemos obtener el intervalo de confianza del ABC y comprobar que no incluye el valor 0,5, ya que en este caso la diferencia no sería estadísticamente significativa

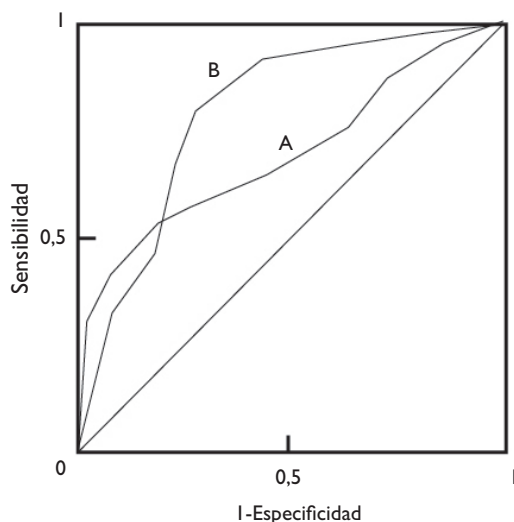
**FIGURA 4. TRES EJEMPLOS DE CURVA ROC. DISCRIMINACIÓN PERFECTA (ÁREA BAJO LA CURVA [ABC] = 1), BUENA DISCRIMINACIÓN (ABC = 0,8) Y CAPACIDAD DE DISCRIMINACIÓN SIMILAR AL AZAR (ABC = 0,5)**



y la prueba no tendría mayor capacidad discriminadora que el azar. De manera alternativa, puede hacerse un contraste de hipótesis mediante el test de Mann-Whitney, que nos proporcionará el valor de  $p$  correspondiente. El problema es que estos procedimientos son matemáticamente complejos y no están al alcance de todos los programas de estadística habitualmente empleados<sup>6</sup>.

El ABC puede servir también para comparar el rendimiento de dos pruebas diagnósticas<sup>7</sup>. En estos casos comparamos las curvas y el ABC de cada una. Aquella que tenga un ABC mayor será la que más potencia diagnóstica tendrá. Así, lo correcto es calcular los intervalos de confianza del 95% y comprobar

**FIGURA 5. COMPARACIÓN DE LAS CURVAS DE DOS PRUEBAS DIAGNÓSTICAS. LA PRUEBA B ES MÁS POTENTE (MAYOR ÁREA BAJO LA CURVA), PERO PUEDE OBSERVARSE QUE LA PRUEBA A ES MÁS ESPECÍFICA PARA VALORES BAJOS DE SENSIBILIDAD**



si existe solapamiento (en cuyo caso la potencia de las dos pruebas será similar) o si uno es mayor que el otro (indicándonos cuál es la prueba más potente). La comparación de las curvas puede ser difícil en algunas ocasiones, por lo que existen métodos matemáticos para realizar los contrastes estadísticos y determinar si existe diferencia significativa entre las dos curvas<sup>8-9</sup>.

En cualquier caso, con independencia de la diferencia en las ABC de dos pruebas diagnósticas, la forma de las curvas puede darnos también información de interés. En la figura 5 podemos ver superpuestas las curvas ROC de dos técnicas diagnósticas, A y B. Aunque la B tiene un ABC mayor y podría considerarse como una prueba diagnóstica más potente que A, podemos fijarnos en que, a valores muy bajos de S, la prueba A tiene un valor de E más alto que la B. De esta manera, si nos interesa maximizar S y E, escogeremos la prueba B, pero si lo que realmente nos interesa es un valor alto de E, quizás nos sea más interesante utilizar la prueba A.

Para finalizar, comentar también que las curvas ROC pueden utilizarse, además de para la valoración de pruebas diagnósticas, para valorar la capacidad de un modelo de regresión logística para discriminar entre dos grupos, casos y no casos<sup>10</sup>. De manera similar a lo que hablamos anteriormente sobre pruebas diagnósticas, un ABC de 1 indica una capacidad discriminadora perfecta del modelo. Cuanto menor sea el ABC, tanto menor será el poder de discriminación, hasta llegar al ABC de 0,5, momento en que la capacidad de discriminación es similar a la del azar.

## BIBLIOGRAFÍA

1. Molina Arias M, Ochoa Sangrador C. Evaluación de la validez de las pruebas diagnósticas (I). Sensibilidad. Especificidad. Evid Pediatr. 2016; 12:34.

2. Molina Arias M, Ochoa Sangrador C. Evaluación de la validez de las pruebas diagnósticas (II). Valores predictivos. *Evid Pediatr.* 2016;12:53.
3. Molina Arias M, Ochoa Sangrador C. Evaluación de la validez de las pruebas diagnósticas (III). Valores predictivos. *Evid Pediatr.* 2016;12:69.
4. Pérez Gaxiola G, Cuello García A. Diagnóstico. Razonamiento clínico y pruebas diagnósticas. En: Cuello García C, Pérez Gaxiola G (eds.) *Medicina Basada en la Evidencia. Fundamentos y su enseñanza en el contexto clínico.* México DF: Editorial Médica Panamericana; 2015. p. 127-48.
5. López-Ratón M, Rodríguez-Álvarez MX, Cadarso-Suárez C, Gude-Sampedro F. Optimal cutpoints: an R package for selecting optimal cut points in diagnostics tests. *J Statistic Software.* 2014;61:1-36.
6. Whitley E, Ball J. Statistics review 6: nonparametric methods. *Crit Care.* 2002;6:509-13.
7. Bewick V, Cheek L, Ball J. Statistic review 13: receiver operating characteristic curves. *Crit Care.* 2004;8:508-12.
8. Martínez-González MA, Toledo E, Sánchez-Villegas A. Análisis de concordancia, validez y pronóstico. En: Martínez González MA, Sánchez-Villegas A, Toledo Atucha EA, Faulin Fajardo J (eds.) *Bioestadística amigable.* 3.ª edición. Madrid: Elsevier España; 2014. p. 455-85.
9. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology.* 1983;148:839-43.
10. Sainani KL. Logistic regression. *PM R.* 2014;6:1157-62.