

Metodología científica

Instituto Nacional de Endocrinología

ANÁLISIS DE LAS CURVAS *RECEIVER-OPERATING CHARACTERISTIC*: UN MÉTODO ÚTIL PARA EVALUAR PROCEDERES DIAGNÓSTICOS

Dra. Emma Domínguez Alonso¹ y Dr. Roberto González Suárez²

RESUMEN

Se sabe que los procedimientos diagnósticos son parte esencial y crítica en la toma de decisiones clínicas, de ahí la importancia de evaluar su precisión diagnóstica. El análisis de las curvas *Receiver-Operating Characteristic* aporta, de manera sencilla, un magnífico indicador de la precisión de una prueba diagnóstica. En este trabajo se consideraron aspectos, fundamentalmente prácticos, relativos a la metodología *Receiver-Operating Characteristic* como: utilidad, modo de interpretar los resultados y ventajas en relación con otros métodos. Se concluyó que el conocimiento sobre esta metodología resulta de gran utilidad para todo profesional de la salud encargado de desarrollar, validar, indicar e interpretar resultados de pruebas diagnósticas.

DeCS: CURVA ROC; TECNICAS Y PROCEDIMIENTOS DIAGNOSTICOS; METODOS Y PROCEDIMIENTOS ESTADISTICOS; SENSIBILIDAD Y ESPECIFICIDAD.

Los procedimientos diagnósticos permiten clasificar correctamente a los pacientes en determinadas categorías, en relación con una enfermedad, de modo que se pueda aplicar una terapéutica adecuada y estimar su pronóstico. Sin embargo, ese proceso es a menudo difícil. Los pacientes pueden tener diversos procesos patológicos concurrentes

y los síntomas y signos clínicos de muchas enfermedades en ocasiones carecen de especificidad. Además, los resultados de exámenes de laboratorio y otros procedimientos diagnósticos de individuos saludables y enfermos a menudo se superponen.

Es necesario aplicar una amplia gama de métodos estadísticos para evaluar y

¹ Especialista de I Grado en Bioestadística. Investigadora Agregada.

² Doctor en Ciencias Médicas. Especialista de II Grado en Bioquímica. Investigador Titular.

manejar la información diagnóstica, lo cual facilita la adecuada atención de los pacientes.¹

El desempeño clínico de una prueba diagnóstica puede ser medido en términos de precisión diagnóstica, o sea, de su habilidad para clasificar correctamente a los sujetos en subgrupos clínicamente relevantes.²

Aunque la sensibilidad, la especificidad y los valores predictivos han sido usados por mucho tiempo como indicadores de la precisión de una prueba diagnóstica, existen métodos modernos, como las curvas de operación característica del receptor (*Receiver-Operating Characteristic* [ROC]), la regresión logística y la razón de verosimilitud, que son considerados indicadores más robustos, pues superan muchas de las limitaciones de los índices tradicionales.^{3,4}

Las curvas ROC proporcionan un buen índice de la capacidad de una prueba diagnóstica para discriminar entre estados alternativos de salud cuando los resultados son medidos en escala ordinal, por intervalo o continua.^{2,5} Son útiles también para comparar distintos procedimientos diagnósticos^{4, 6-10} y seleccionar umbrales de decisión (puntos de corte entre los resultados positivos y negativos de la prueba).^{2, 10-13}

Las pruebas diagnósticas son una parte crítica del proceso clínico; estrategias diagnósticas inapropiadas ponen en riesgo al paciente y ocasionan grandes pérdidas de recursos, esto evidencia la importancia de evaluar adecuadamente la precisión de las pruebas. Con este trabajo nos hemos propuesto un acercamiento, fundamentalmente práctico, a un método muy útil, y aceptablemente sencillo, para evaluar el poder discriminatorio de un proceder diagnóstico, para esto haremos una breve descripción teórica del método y de las

situaciones en las que está indicado su uso, expondremos el modo de interpretar los resultados del gráfico y del área bajo la curva ROC, las ventajas y las desventajas del método.

Hemos incluido algunos ejemplos de la aplicación de esta técnica estadística en la evaluación de métodos diagnósticos en el campo de la Endocrinología.

DESCRIPCIÓN DEL MÉTODO

La metodología ROC fue desarrollada en el contexto de la detección de señales electrónicas en los inicios de la década de los 50. A mediados de los 60 se habían usado las curvas ROC en psicología y psicofísica experimental.¹⁴

Leo Lusted, un radiólogo, fue el primero en usarlas en el proceso de toma de decisiones médicas, en 1967, y comenzó a aplicarlas en estudios con imágenes, en 1969.^{15,16}

Las pruebas diagnósticas son utilizadas para ayudar a responder preguntas en relación con la atención de los pacientes.

La precisión es la principal característica de un proceder diagnóstico como recurso de clasificación, esta mide la habilidad de la prueba para distinguir entre estados alternativos de salud, lo cual incluye la distinción entre enfermedad y salud, entre enfermedad benigna y maligna, entre sujetos que responden o no a una determinada terapia y predecir quién enfermará o no. Esta habilidad de discriminar es el punto de comienzo cuando estimamos qué contribución puede hacer una prueba diagnóstica al proceso de atención del paciente.

Si bien la precisión de una prueba diagnóstica es el indicador fundamental de su valor en la atención al paciente, otro

aspecto que debemos considerar al evaluarlas es su utilidad, el valor práctico de la información que obtendremos. Una prueba puede tener una habilidad considerable para discriminar y, sin embargo, poco valor práctico para la atención a los pacientes, por su costo y carácter invasivo, entre otros aspectos.^{17,18}

La precisión de un proceder diagnóstico ha sido definida como su habilidad para discriminar entre 2 subclases de sujetos cuando hay alguna razón clínica relevante para hacerlo.²

La precisión de una prueba diagnóstica puede ser medida en términos de sensibilidad y especificidad, estas son definidas como la proporción de sujetos con enfermedad y sin ella, correctamente clasificados por la prueba.¹⁹

La importancia de estos conceptos es absolutamente reconocida, pero reportar un solo valor de sensibilidad y especificidad es una simplificación del problema; un proceder diagnóstico no tiene un solo valor de sensibilidad y especificidad, sino muchos. Como el umbral de decisión usado para clasificar a los sujetos como positivos o negativos varía a través del rango de resultados posibles, la sensibilidad y la especificidad se moverán en direcciones opuestas, cuando una aumenta la otra disminuye, para cada umbral de decisión hay una combinación de sensibilidad y especificidad. Consecuentemente, solo el rango completo de los pares sensibilidad/especificidad proporciona un cuadro íntegro de la precisión de la prueba.

FIGURA DE LA CURVA ROC

Para que una prueba diagnóstica sea usada en la atención de pacientes es imprescindible seleccionar un umbral de decisión, pero no es necesario hacerlo para

estimar la precisión. La figura de la curva ROC proporciona un cuadro completo de la habilidad de una prueba para discriminar, se examinan todos los posibles umbrales de decisión; representa los pares de sensibilidad/especificidad para todo el rango de resultados observados.

En el eje "Y" se representa la sensibilidad o la fracción de verdaderos positivos definido como:

$$\frac{\text{Número de verdaderos positivos (VP)}}{\text{Número de verdaderos positivos (VP) + Número de falsos negativos (FN)}}$$

VP: Sujetos enfermos que el proceder diagnóstico clasifica como tales.

FN: Sujetos enfermos que el proceder diagnóstico clasifica como sanos.

Esta es también definida como la positividad en presencia de una enfermedad o condición y es calculada solo del grupo afectado. En el eje "X" está la fracción de falsos positivos o 1-especificidad, definida como:

$$\frac{\text{Número de falsos positivos (FP)}}{\text{Número de falsos positivos (FP) + Número de verdaderos negativos (VN)}}$$

FP: Sujetos sanos que el proceder diagnóstico clasifica como enfermos.

VN: Sujetos sanos que el proceder diagnóstico clasifica como tales.

Este es un índice de especificidad y solo se calcula en el grupo no afectado.

Las fracciones de verdaderos y falsos positivos se calculan por separado para cada subgrupo, por tanto, la curva ROC es independiente de la prevalencia de la enfermedad en la muestra. Cada punto en la figura representa un par de sensibilidad/especificidad correspondiente

a un umbral de decisión particular. Una prueba diagnóstica con una discriminación perfecta tiene una curva que pasa a través de la esquina superior izquierda, donde la fracción de verdaderos positivos es 1 ó 100 % (sensibilidad perfecta) y la fracción de falsos positivos es 0 (especificidad perfecta). Una figura teórica para una prueba que no discrimina (distribución idéntica de los resultados para ambos grupos) es una línea diagonal de 45° desde la esquina inferior izquierda hasta la superior derecha. La mayoría de las representaciones caen entre estos extremos.

Cuanto más cerca esté la línea de la esquina superior izquierda, mayor será la precisión de la prueba.

COMPARACIÓN DE DIFERENTES PRUEBAS DIAGNÓSTICAS MEDIANTE LA FIGURA DE LA CURVA ROC

Cuando se han obtenido resultados de múltiples pruebas en un mismo grupo de pacientes, pueden ser representados juntos. La posición relativa de las líneas indica la precisión relativa de las pruebas. Una curva que cae encima y a la izquierda de otra indica mayor precisión.

VENTAJAS DE LA FIGURA DE LA CURVA ROC

- Es una representación simple, y fácilmente comprensible, de la precisión de una prueba, o sea, de su habilidad de discriminar a través de todo el rango de valores.
- No requiere seleccionar un umbral de decisión particular porque es incluido todo el rango de posible umbrales.
- Es independiente de la prevalencia, no necesita obtener muestras con prevalencia representativa, de hecho usualmente es preferible tener igual

número de sujetos con ambas condiciones. Sin embargo, se ha planteado que estudios en los que se recluta a los pacientes con la enfermedad y sin ella, por separado, sobrestiman la precisión, en relación con aquellos en que los sujetos son obtenidos como una muestra representativa de la población en la cual el proceder diagnóstico fue realizado, sin selección previa, según el estado de la enfermedad.²⁰

- Proporciona una comparación visual directa entre pruebas sobre una escala común.
- Puede ser aplicado para pruebas diagnósticas cuyos resultados son medidos en escala tanto ordinal, como por intervalo o continua.¹¹

DESVENTAJAS DE LA FIGURA DE LA CURVA ROC

- No se muestra los umbrales de decisión reales.
- No se muestra el número de sujetos, y a medida que el tamaño de la muestra decrece, la representación gráfica tiende a volverse progresivamente mellada y desigual.²
- La generación de la figura y el cálculo de los parámetros es difícilmente manejable sin programas de computación, los que no están ampliamente disponibles.
- No tiene aplicación cuando los resultados de la prueba son medidos en una escala dicotómica.

ÁREA BAJO LA CURVA ROC

El área bajo la curva ROC es el mejor indicador global de la precisión de una prueba diagnóstica.⁴ Hace factible expresar el desempeño de una prueba mediante un número simple.

Esta área es siempre mayor o igual a 0,5. El rango de valores se mueve entre

1 (discriminación perfecta) y 0,5 (no hay diferencias en la distribución de los valores de la prueba entre los 2 grupos). La interpretación del valor del área sería del modo siguiente: un área de 0,8 significa que un individuo seleccionado aleatoriamente del grupo de enfermos tiene un valor de la prueba mayor que uno seleccionado aleatoriamente del grupo de sanos en el 80 % de las veces.

Mediante una prueba de hipótesis y/o de la estimación del intervalo de confianza para el área, podemos evaluar la precisión de un proceder diagnóstico. Rechazar la hipótesis de que el área teórica es igual a 0,5 ($p < 0,05$ y/o intervalo de confianza que no contiene al 0,5), proporciona evidencia de que la prueba diagnóstica tiene la habilidad para distinguir entre los 2 subgrupos.

Utilizando una prueba de hipótesis es posible comparar varias áreas bajo la curva ROC, lo que permite hacer distinciones entre el poder discriminatorio de 2 o más procedimientos diagnósticos, cuando estos se han realizado en el mismo grupo de pacientes.²

ALGUNOS EJEMPLOS DEL USO DE LA CURVAS ROC EN EL CAMPO DE LA ENDOCRINOLOGÍA

El primer ejemplo muestra los resultados de la aplicación de las curvas ROC para evaluar la precisión diagnóstica de un método desarrollado en el Instituto Nacional de Endocrinología (INEN) para determinar la progesterona en plasma. Se empleó como referencia el diagnóstico

de la ovulación realizado con un kit comercial (prog-ctria, cib bio internacional) de uso en el país.

En la tabla 1 se observa que el área es 0,981, con un intervalo de confianza que no contiene al 0,5 y una $p = 0,00$, todos estos aspectos indican que estamos ante un proceder diagnóstico de alta precisión, lo que se corrobora con una curva muy cercana al extremo superior izquierdo (figura 1).

En el segundo ejemplo se comparó el desempeño, como indicadores pronóstico de la evolución de la diabetes, de 3 índices relacionados con la secreción o la sensibilidad a la insulina, medidos en sujetos con tolerancia a la glucosa alterada. El resultado evaluado fue si el sujeto era diabético o no, 18 años después de realizado el estudio inicial.

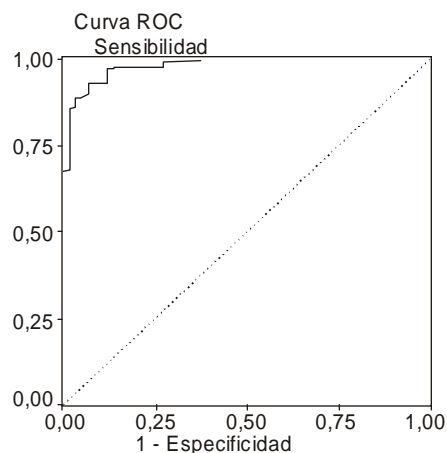


FIG.1. Gráfico de la Curva ROC. Evaluación de un método para la determinación de progesterona.

TABLA 1. Área bajo la curva ROC. Evaluación de un método para determinar progesterona

Variable	Área	Error estándar	Significación	Límite de confianza 95 %	
				Límite inferior	Límite superior
Proginen	0,981	0,008	0,00	0,965	0,996

Los índices fueron:

- Índice insulínogénico a los 30 min (Ind030in), que consiste en el cociente del incremento de la insulinemia en relación con el incremento de la glucemia en los primeros 30 min de la PTG oral.^{21,22}
- Índice de resistencia a la insulina (Irhomain) e índice de actividad de la célula Beta (Betahoin), ambos del modelo homeostático de regulación de la glucemia, calculado a partir de la

glucemia e insulinemia en ayunas, como se ha descrito anteriormente.²³

De los índices evaluados solo uno (Ind030in) tiene un área significativamente diferente de 0,5 (área 0,642, $p = 0,025$ e intervalo de confianza que no contiene al 0,5); por lo tanto, es el único de los 3 índices analizados que resulta de utilidad para establecer pronóstico de mala evolución de la tolerancia a la glucosa (tabla 2). Los índices restantes tienen áreas

TABLA 2. Evaluación de 3 índices como indicadores pronósticos de la evolución de la diabetes mellitus

Variables	Área	Error estándar	Significación	Intervalo de confianza	
				Límite inferior	Límite superior
Irhomain	0,512	0,064	0,847	0,386	0,639
Betahoin	0,487	0,064	0,84	0,362	0,613
Ind030in	0,642	0,061	0,025	0,523	0,761

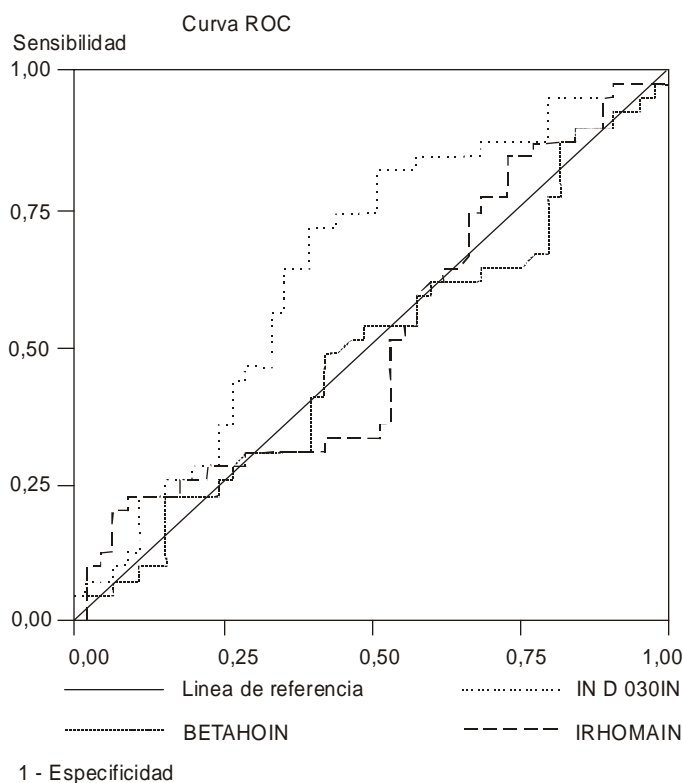


FIG.2. Gráfico de la Curva ROC. Evaluación de 3 índices como indicadores pronóstico de la evolución de la diabetes mellitus.

que no difieren significativamente de 0,5 y curvas muy cercanas a la diagonal del gráfico (figura 2).

Se concluyó que las curvas ROC proporcionan un buen índice de la capacidad de una prueba diagnóstica para discriminar entre estados alternativos de salud cuando los resultados son medidos

en escala ordinal, por intervalo o continua; son útiles para comparar procedimientos diagnósticos y seleccionar umbrales de decisión.

El conocimiento sobre esta metodología resulta de gran utilidad para todo profesional de la salud relacionado con el desarrollo y/o aplicación de pruebas diagnósticas.

SUMMARY

AS the diagnostic procedures are an essential and critical part at the time of making clinical decisions, it is very important to evaluate their diagnostic accuracy. The analysis of the Receiver-Operating Characteristic curves offers in a simple way an excellent indicator of precision of a diagnostic test. In this paper, mainly practical aspects connected with ROC methodology, such as usefulness, way of interpreting the results and advantages in relation to other methods, are considered. It is concluded that the knowledge on this methodology is very useful for every health professional in charge of developing, validating, indicating and interpreting the results of the diagnostic tests.

Subject headings: ROC CURVE; DIAGNOSTIC TECHNIQUES AND PROCEDURES; STATISTICAL METHODS AND PROCEDURES; SENSITIVITY AND SPECIFICITY.

REFERENCIAS BIBLIOGRÁFICAS

1. Kazmierczak CS. Statistical Techniques for evaluating the diagnostic utility of laboratory tests. *Clin Chem Lab Med* 1999;37(11-12):1001-9.
2. Zweig M H, Campbell G. Receiver-Operating Characteristic (ROC) Plots: A fundamental evaluation tool in Clinical Medicine. *Clin Chem* 1993;39 (4):561-77.
3. Deeks JJ. Systematic reviews of evaluations of diagnostic and screening test. *Br Med J* 2001;323:157-62.
4. Boyd JC. Mathematical tools for demonstrating the clinical usefulness of biochemical markers. *Scand J Clin Lab Invest Suppl* 1997;227:46-63.
5. Smith-Bindman R, Kerlikowske K, Feldstein VA. Endovaginal ultrasound to exclude endometrial cancer and other endometrial abnormalities. *JAMA* 1998;280:1510-7.
6. Beglin FM, Firestone AR, Vig KW. A comparison of the reliability and validity of 3 occlusal indexes of orthodontic treatment need. *Am J Orthod Dentofacial* 2001;120(3):240-6.
7. Biagini RE, Krieg EF, Pinkerton LE. Receiver Operating Characteristics analyses of food and drug administration-cleared serological assays for natural rubber latex-specific immunoglobulin e antibody. *Clin Diagn Lab Immunol* 2001;8(6):1145-9.
8. Strandberg K, Bhiladvala P, Holm J. A new method to measure plasma levels of activated protein C in complex with protein C inhibitor in patients with acute coronary syndromes. *Blood Coagul Fibrinolysis* 2001;12(7):503-10.
9. Chen B, Lin G, Ni Z. Likelihood ratio and ROC curve in evaluation of iron parameters for diagnosing iron deficiency. *Zhonghua Yi Xue Za Zhi* 1999;79(2):99-103.
10. Piatt JH. Receiver-operating characteristic curves. *J Neurosurg* 2001; 95 (5): 918-9.
11. Greiner M, Pfeiffer D, Smith RD. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. *Prev Vet Med* 2000;45(1-2):23-41.

12. Farr BM, Shapiro DE. Diagnostic tests: Distinguishing good tests from bad even ugly ones. *Infect Control Hosp Epidemiol* 2000;21(4):278-84.
13. El-Halwagy HE, Gelbaya TA, El-Wahab MF. The mean third trimester postprandial blood glucose of diabetic pregnant patients and infant birth weight in the Kuwati population. *Medscape Womens Health* 2001;6(5):2.
14. Green DM, Swets JA. Signal detection theory and psychophysics. New York: John Wiley & Sons, 1966:PPP_{ic}
15. Lusted LB. Decision making studies in patient management. *N Engl J Med* 1971;284:416-24.
16. Lusted LB. Signal detectability and medical decision-making. *Science* 1971;171:1217-9.
17. Deeks JJ. Using evaluations of diagnostic tests: understanding their limitations and making the most of available evidence. *Ann Oncol* 1999;10:761-8.
18. Guyatt GH, Tugwell P, Feeny DH. A framework for clinical evaluation of diagnostic technologies. *Can Med Assoc J* 1986;134:487-594.
19. Bland JM, Altman DG. Diagnostic tests 1: Sensitivity and specificity. *Br Med J* 1994;308:1499.
20. Lijmer JC, Mol BW, Heisterkamp S. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999;282:1061-6.
21. González R, Arranz C. Secreción de insulina y sensibilidad a la insulina durante la prueba de tolerancia a la glucosa oral, en sujetos con tolerancia a la glucosa normal. *Rev Cubana Endocrinol* 2000;11:23-30.
22. Matsumoto K, Yanaguchi Y, Miyaque S. Glucose tolerance, insulin secretion and insulin sensitivity in non-obese and obese subjects. *Diabetes Care* 1997; 20:1562-8.
23. Matthews DR, Hosker JP, Rudenski AS. Homeostasis model assessment: insulin resistance and Beta cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia* 1985;28:412-9.

Recibido: 15 de marzo de 2002. Aprobado: 28 de junio de 2002.

Dra. *Emma Domínguez Alonso*. Instituto Nacional de Endocrinología. Zapata y D, El Vedado, Ciudad de La Habana, Cuba.