

Ivonne Analí Roy-García<sup>1,2a</sup>, Carlos Paredes-Manjarrez<sup>3b</sup>, Jorge Moreno-Palacios<sup>4c</sup>, Rodolfo Rivas-Ruiz<sup>1d</sup>, Andrey Arturo Flores-Pulido<sup>2,5e</sup>

### Resumen

El uso de pruebas diagnósticas para determinar la presencia o ausencia de una enfermedad es esencial en la práctica clínica. Los resultados de una prueba diagnóstica pueden corresponder a estimaciones numéricas que requieren parámetros cuantitativos de referencia para trasladarse a una interpretación dicotómica como normal o anormal y así, implementar acciones para la atención de una condición o una enfermedad. Por ejemplo, en el diagnóstico de anemia es necesario definir un punto de corte para la variable hemoglobina y crear dos categorías que distingan la presencia o no de anemia. El método utilizado para este proceso es la elaboración de curvas de rendimiento diagnóstico, mejor conocidas por sus siglas en inglés como ROC (*Receiver Operating Characteristic*). La curva ROC además es útil como marcador pronóstico, ya que permite definir el punto de corte de una variable cuantitativa que se asocia a mayor mortalidad o riesgo de complicaciones. Se han usado en distintos marcadores pronósticos en COVID-19, como el índice neutrófilos/linfocitos y dímero D, en los que se identificaron puntos de corte asociados a mortalidad y/o riesgo de ventilación mecánica. La curva ROC se utiliza para evaluar el rendimiento diagnóstico de una prueba de forma aislada, pero también se puede usar para comparar el rendimiento de dos o más pruebas diagnósticas y definir aquella que es más precisa. En este artículo se describen los conceptos básicos para el uso e interpretación de la curva ROC, la interpretación de un área bajo la curva (ABC) y la comparación de dos o más pruebas diagnósticas.

### Abstract

The use of diagnostic tests to determine the presence or absence of a disease is essential in clinical practice. The results of a diagnostic test may correspond to numerical estimates that require quantitative reference parameters to be transferred to a dichotomous interpretation as normal or abnormal and thus implement actions for the care of a condition or disease. For example, in the diagnosis of anemia it is necessary to define a cut-off point for the hemoglobin variable and create two categories that distinguish the presence or absence of anemia. The method used for this process is the preparation of diagnostic performance curves, better known by their acronym in English as ROC (*Receiver Operating Characteristic*). The ROC curve is also useful as a prognostic marker, since it allows defining the cut-off point of a quantitative variable that is associated with greater mortality or risk of complications. They have been used in different prognostic markers in COVID-19, such as the neutrophil/lymphocyte ratio and D-dimer, in which cut-off points associated with mortality and/or risk of mechanical ventilation were identified.

The ROC curve is used to evaluate the diagnostic performance of a test in isolation, but it can also be used to compare the performance of two or more diagnostic tests and define which one is more accurate. This article describes the basic concepts for the use and interpretation of the ROC curve, the interpretation of an area under the curve (AUC) and the comparison of two or more diagnostic tests.

<sup>1</sup>Instituto Mexicano del Seguro Social, Coordinación de Investigación en Salud, Centro de Adiestramiento e Investigación Clínica. Ciudad de México, México

<sup>2</sup>Instituto Politécnico Nacional, Escuela Superior de Medicina, Sección de Posgrado. Ciudad de México, México

<sup>3</sup>Instituto Mexicano del Seguro Social, Centro Médico Nacional Siglo XXI, Hospital de Especialidades, Departamento de Imagenología. Ciudad de México, México

<sup>4</sup>Instituto Mexicano del Seguro Social, Centro Médico Nacional Siglo XXI, Hospital de Especialidades, Departamento de Urología. Ciudad de México, México

<sup>5</sup>Secretaría de Salud, Centro Regional de Alta Especialidad, Hospital de Especialidades Pediátricas. Tuxtla Gutiérrez, Chiapas, México

ORCID: [0000-0002-1859-3866](https://orcid.org/0000-0002-1859-3866)<sup>a</sup>, [0000-0002-3319-7634](https://orcid.org/0000-0002-3319-7634)<sup>b</sup>, [0000-0001-9994-4922](https://orcid.org/0000-0001-9994-4922)<sup>c</sup>, [0000-0002-5967-7222](https://orcid.org/0000-0002-5967-7222)<sup>d</sup>, [0000-0002-3358-1054](https://orcid.org/0000-0002-3358-1054)<sup>e</sup>

#### Palabras clave

Técnicas y Procedimientos Diagnósticos  
Curva ROC  
Área Bajo la Curva  
Estadística

#### Keywords

Diagnostic Techniques and Procedures  
ROC Curve  
Area Under Curve  
Statistics

Fecha de recibido: 13/01/2023

Fecha de aceptado: 15/02/2023

#### Comunicación con:

Rodolfo Rivas Ruiz

 [rivasrodolfo@gmail.com](mailto:rivasrodolfo@gmail.com)

 55 5418 8712

**Cómo citar este artículo:** Roy-García IA, Paredes-Manjarrez C, Moreno-Palacios J *et al.* Curvas ROC: características generales y su uso en la práctica clínica. Rev Med Inst Mex Seguro Soc. 2023;61 Supl 3:S497-502. doi: 10.5281/zenodo.8319791

Los estudios diseñados para medir el desempeño o rendimiento de las pruebas diagnósticas son importantes para la atención del paciente y la optimización de los recursos durante la atención médica.<sup>1</sup> Las variables cuantitativas se consideran más precisas ya que proporcionan una mejor aproximación al objeto de estudio en comparación con las cualitativas, cuyas escalas son nominales u ordinales y a menudo se necesita convertir a las variables cuantitativas a una escala nominal dicotómica con fines de establecer un diagnóstico o un pronóstico.<sup>2</sup> Esto requiere establecer un punto de corte para definir la prueba positiva o negativa.<sup>3</sup> Una vez que se define este punto de corte es posible convertir la variable cuantitativa a dicotómica para poder calcular la sensibilidad, la especificidad y los valores predictivos.

La sensibilidad se refiere a la proporción de individuos enfermos que tienen el resultado de la prueba innovadora positiva; la especificidad considera la proporción de individuos no enfermos que tienen el resultado de la nueva prueba negativa, y el valor predictivo positivo representa la probabilidad que tiene un paciente con cierta prueba positiva (signo, síntoma, resultado de laboratorio, gabinete o algún índice) de cursar con una enfermedad determinada, el valor predictivo negativo representa la probabilidad que tiene un paciente con cierta prueba negativa de estar libre de la enfermedad. Por ejemplo, si queremos definir el mejor punto de corte de antígeno prostático específico (APE), que es indicativo de cáncer de próstata, se requiere hacer una

curva ROC para definir el mejor punto a partir de la variable cuantitativa, la cual posteriormente se convertirá a una dicotómica, que se contrastará con el estándar de oro, que en este ejemplo corresponde al resultado histopatológico que definirá la presencia o no de cáncer.<sup>4</sup> En la figura 1 es posible observar el diseño de investigación de este estudio de prueba diagnóstica, cuyo acrónimo es PIES (Población, prueba Innovadora, prueba Estándar y, como información de salida, Sensibilidad).

### Elección del punto de corte

Las curvas ROC son una herramienta útil para la evaluación del mejor desempeño de una prueba diagnóstica dentro de un rango de valores posibles de una variable predictora. El mejor punto de corte es aquel que garantiza la mejor área bajo la curva, que incluye la mejor sensibilidad o especificidad que garantiza el mejor desempeño de la prueba. En la práctica clínica no basta con pruebas altamente sensibles pero poco específicas, como funciona en las medidas de salud pública dirigidas a grandes grupos de individuos. En esta práctica *cara a cara*, se debe tomar en cuenta el desempeño global de la prueba, las implicaciones (acciones inmediatas) para el paciente y el costo de los resultados incorrectos frente a los beneficios de los resultados correctos.

Figura 1 Diseño arquitectónico de un estudio de proceso o prueba diagnóstica



APE: Antígeno prostático específico

APE: Antígeno prostático específico

La curva ROC, además, es útil como marcador pronóstico, ya que permite definir el punto de corte de una variable cuantitativa que se asocia a mayor mortalidad o riesgo de complicaciones. A manera de ejemplo, podemos mencionar el uso de distintos marcadores pronósticos para COVID-19, como el índice neutrófilos/linfocitos, deshidrogenasa láctica y dímero D, en los que se identificaron puntos de corte asociados a una mayor mortalidad y riesgo de ventilación mecánica.<sup>5,6</sup>

Definida de manera simple, una curva ROC es un gráfico de la sensibilidad (en el eje de las Y) frente a 1 - Especificidad (falsos positivos) en el eje de las X, de una prueba de diagnóstica en escala cuantitativa.<sup>7,8</sup> La figura 2 resume la ecuación de la razón de verosimilitud positiva la cual es Sensibilidad/1 - Especificidad. Asimismo, nos permite obtener el valor el área bajo la curva (ABC), es decir, el desempeño global de la prueba en cada uno de sus valores.

Para ejemplificar este proceso consideramos un conjunto hipotético de 732 pacientes.

El cuadro I muestra la distribución de sensibilidad y falsos positivos (1 - Especificidad) para diferentes valores de antígeno prostático. En este cuadro se muestra el desempeño de cada valor de forma individual, a fin de que el investigador seleccione el más conveniente. Ejemplo, para un valor de APE de 3.4 ng/mL la sensibilidad es del 100%, pero al analizar el porcentaje de falsos positivos, este es del 100%, por lo que tendríamos que descartar ese valor como punto de corte por no tener capacidad de discernir entre los enfermos y sanos. Es decir, el desempeño es igual a tirar una moneda al aire.

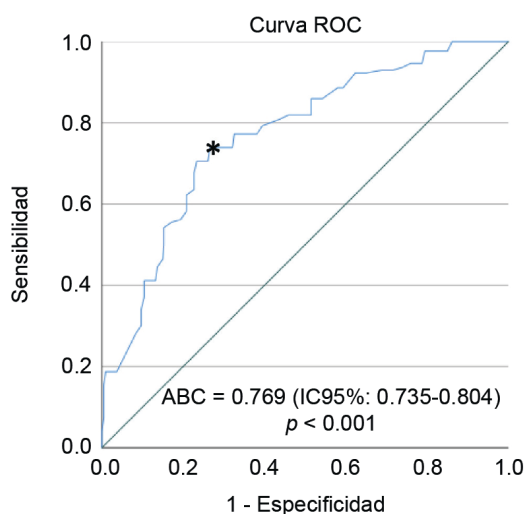
Al analizar un punto de corte de 9.1 encontramos una sensibilidad de 93.6%, pero la cantidad de falsos positivos aún es alta (73.9%). Al analizar un punto de corte de 20.7 encontramos una sensibilidad del 73.9% y un 26% de falsos positivos, por lo que en esta población que acude a un tercer nivel de atención, el punto de corte que muestra la mejor sensibilidad y la menor cantidad de falsos positivos es de 20.7.

Una vez que se establece el mejor punto de corte que ayuda a la discriminación del desenlace dicotómico, que corresponde al valor cuantitativo de la prueba que tiene mayor sensibilidad y menor cantidad de falsos positivos (1 - Especificidad) y es el punto que se sitúa más cerca del ángulo superior izquierdo de la curva, que en la figura 2 se encuentra señalado con un asterisco.<sup>9,10</sup> El mejor punto de corte también puede elegirse mediante el mayor valor del índice de Youden, que consiste en elegir aquel valor más alto, resultante de la resta: Sensibilidad – (1 – Especificidad).<sup>11</sup> Los valores de este índice van de +1 a -1. Siendo el +1 el valor ideal de este índice. Es decir, una prueba con 100% de sensibilidad y 0% de falsos positivos (o 100% de especificidad).

Una prueba similar al azar tendría un valor de 0, 0.5 (50%) de sensibilidad y 0.5 (50%) de falsos positivos. Una prueba con pésimo desempeño es aquella que tiene un índice de Youden negativo: 0% de Sensibilidad – 100% de falsos positivos = -1

Regresando a nuestro ejemplo, el mejor punto de corte con el índice de Youden más alto corresponde al valor de 20.7 para APE y a un índice de Youden de 0.479.

Figura 2 Curva ROC para la variable de antígeno prostático



ABC: Área bajo la curva  
Fuente: datos hipotéticos con fines didácticos

Cuadro I Rendimiento diagnóstico del antígeno prostático en diferentes puntos de corte

Antígeno prostático (ng/mL)	Sensibilidad	1 - Especificidad (falsos positivos)	Índice de Youden
3.4	1.000	1.000	0.000
5.1	1.000	0.972	0.028
6.7	1.000	0.917	0.083
7.4	1.000	0.889	0.111
8.2	0.946	0.788	0.158
9.1	0.936	0.739	0.197
10.1	0.923	0.651	0.272
12.8	0.860	0.543	0.317
14.3	0.819	0.487	0.332
19.8	0.739	0.293	0.446
20.7	0.739	0.260	0.479
22.0	0.706	0.233	0.470
24.2	0.676	0.226	0.450

\*Fuente: datos hipotéticos con fines didácticos

Una vez que se tiene el punto de corte será necesario evaluar la capacidad diagnóstica para clasificar correctamente a los pacientes con una enfermedad, lo cual se denomina área bajo la curva (ABC).<sup>11</sup> El ABC es una forma efectiva de resumir la precisión diagnóstica general de la prueba. Toma valores de 0 a 1, donde un valor de 0 indica una prueba perfectamente inexacta y un valor de 1 refleja una prueba perfectamente precisa. En general, un ABC de 0.5 sugiere que no hay discriminación (es decir, la capacidad de diagnosticar pacientes con y sin la enfermedad o afección según la prueba es similar a la de un volado o a tomar la decisión con base al azar), mientras que de 0.7 a 0.8 se considera aceptable, de 0.8 a 0.9 se considera excelente y más de 0.9 es considerado como un desempeño sobresaliente de la prueba.<sup>3</sup>

Un valor de 0.5 para ABC indica que la curva ROC caerá en diagonal (es decir, una línea de 45 grados, llamado también *toin coss* o volado) y, por lo tanto, sugiere que la prueba de diagnóstico o pronóstico no tiene capacidad de discriminación (a lo cual se le considera como hipótesis nula). Se considera que las curvas ROC por encima de esta línea diagonal tienen una capacidad de discriminación razonable para diagnosticar pacientes con y sin la enfermedad/condición (hipótesis alterna).<sup>12,13</sup>

En la figura 3 es posible observar diferentes áreas bajo la curva del lado derecho, y del lado izquierdo se observa el impacto en la capacidad de la prueba para discriminar entre

sanos y enfermos. La ABC de 0.50 no tiene la capacidad de discriminar correctamente entre sanos y enfermos, su rendimiento es igual que el azar.

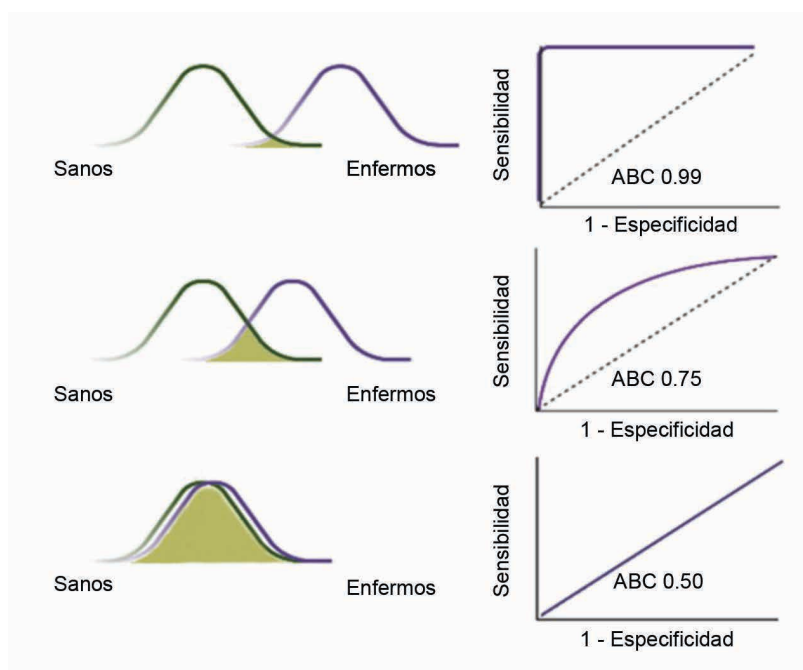
Por lo tanto, es natural hacer una prueba de hipótesis para evaluar si el ABC difiere significativamente de 0.5. Así, la hipótesis nula ( $H_0$ ) y alternativa ( $H_1$ ) se definen como  $H_0$ :  $ABC = 0.5$  frente a  $H_1$ :  $ABC > 0.5$ . Al analizar el ABC es necesario tomar en cuenta los intervalos de confianza (un intervalo de confianza nos indica el verdadero valor que toma el parámetro en el universo), para considerar a una prueba diagnóstica como útil para discriminar la condición de interés se debe tener al intervalo de confianza por arriba de 0.5.<sup>3</sup>

El antígeno prostático muestra un área bajo la curva de 0.739, con un IC95% de 0.735-0.804, al no cruzar el IC95% inferior el 0.5  $p < 0.001$ , consideramos que es una prueba que discrimina correctamente la presencia/ausencia de cáncer en el 76.9% de las veces.

## Comparación de dos curvas ROC

Si quisiéramos comparar la certeza diagnóstica de dos pruebas para cáncer de próstata, el antígeno prostático frente al volumen prostático y la elastografía, requerimos evaluar el área bajo la curva de cada una de ellas y determinar si existen diferencias significativas en el rendimiento diagnóstico entre ellas.

**Figura 3** Capacidad de discriminación de los sanos y enfermos en diferentes áreas bajo la curva (ABC)

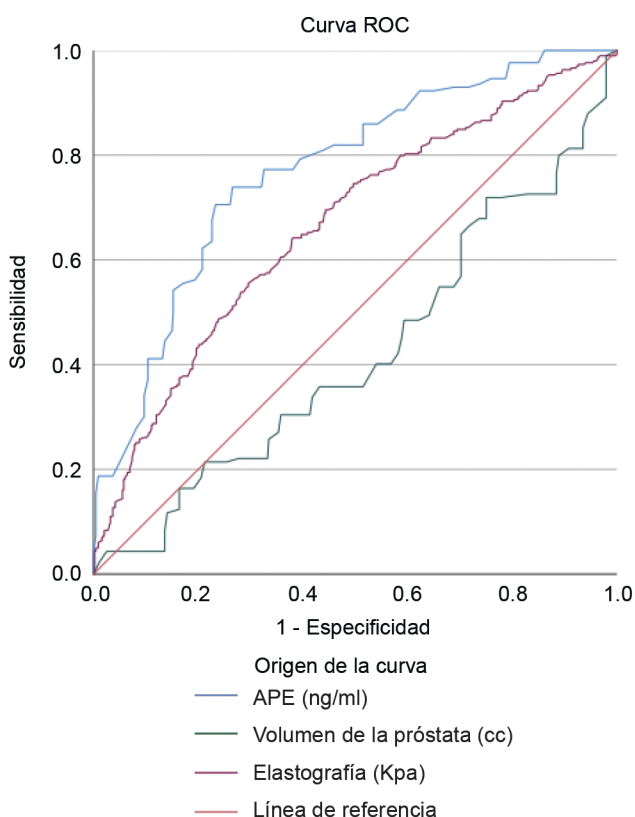


ABC: Área bajo la curva

Las curvas ROC también pueden ser útiles para comparar la capacidad diagnóstica de dos o más pruebas de detección o para evaluar la capacidad predictiva de dos o más marcadores para la misma enfermedad.<sup>14,15,16</sup> En general, la prueba con el ABC más alta puede considerarse mejor, como se muestra en la **figura 4**, en la cual se observa un mayor rendimiento para APE, pero no es suficiente, ya que resulta fundamental evaluar si existen diferencias significativas entre las pruebas diagnósticas, por lo que es necesario considerar que los intervalos de confianza entre las áreas bajo la curva entre las pruebas no se traslapen.

En el **cuadro II** es posible observar las ABC para cada una de las pruebas y sus IC95%. El volumen de la próstata

**Figura 4** Comparación de una curva ROC para tres pruebas diagnósticas



APE: Antígeno prostático específico  
Fuente: datos hipotéticos con fines didácticos

medido en centímetros cúbicos no muestra utilidad para la clasificación de los pacientes con cáncer de próstata, ya que la media del ABC (0.421) se encuentra por debajo de la línea de referencia (0.5). Si bien los IC95% no atraviesan la línea diagonal 0.378-0.463, al estar tan cerca del valor de 0.5 estos valores no son estadísticamente significativos ( $p = 0.421$ ).

El APE muestra mejor desempeño diagnóstico con una media de ABC de 0.769 que la elastografía con 0.666 y las diferencias entre los rendimientos son estadísticamente significativas, ya que el IC95% inferior para APE es de 0.735 y el superior para elastografía es de 0.706, es decir no se tocan los límites inferior y superior de los IC95%, por lo que consideramos que la prueba APE muestra una mayor ABC que elastografía y estas diferencias son estadísticamente significativas.

Si bien el principal interés de estos análisis es ayudar al clínico y al paciente en la toma de decisiones sobre qué prueba realizarse,<sup>17,18</sup> también son de utilidad para la implementación de programas médicos que busquen mejorar la calidad de la atención de nuestros pacientes.<sup>19,20</sup>

### Conclusión

Las curvas ROC son una herramienta útil en la evaluación del desempeño de una prueba diagnóstica, permiten, de modo sencillo, establecer un punto de corte para una variable cuantitativa, donde se obtiene la mejor sensibilidad y especificidad, dependiendo del objetivo de la prueba; además, el área bajo la curva es una forma simple y gráfica de resumir la precisión diagnóstica general de la prueba. Mediante el uso de curvas ROC se puede realizar la comparación del rendimiento diagnóstico de dos o más pruebas, el cual se deberá definir considerando el área bajo la curva y los intervalos de confianza de cada una de las pruebas.

**Declaración de conflicto de interés:** los autores han completado y enviado la forma traducida al español de la declaración de conflictos potenciales de interés del Comité Internacional de Editores de Revistas Médicas, y no fue reportado alguno que tuviera relación con este artículo.

**Cuadro II** Comparación del ABC para tres pruebas diagnósticas

Prueba diagnóstica	ABC	IC95%	$p$
APE (ng/ml)	0.769	0.735-0.804	< 0.001
Volumen de la próstata (cc)	0.421	0.378-0.463	0.421
Elastografía (Kpa)	0.666	0.626-0.706	0.020

APE: Antígeno prostático específico; ABC: Área bajo la curva; IC95%: Intervalo de confianza al 95%

## Referencias

1. Talavera JO, Wachter-Rodarte NH, Rivas-Ruiz R. Clinical research II. Studying the process (the diagnosis test). *Rev Med Inst Mex Seguro Soc.* 2011;49(2):163-70.
2. Obuchowski NA. An ROC-type measure of diagnostic accuracy when the gold standard is continuous-scale. *Stat Med.* 2006;25(3):481-93. doi: 10.1002/sim.2228.
3. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. *J Thorac Oncol.* 2010;5(9):1315-6. doi: 10.1097/JTO.0b013e3181ec173d.
4. Cui Y, Cao W, Li Q, Shen H, Liu C, Deng J, et al. Evaluation of prostate cancer antigen 3 for detecting prostate cancer: a systematic review and meta-analysis. *Sci Rep.* 2016;10(6):25776. doi: 10.1038/srep25776.
5. Bartzioakas K, Kostikas K. Lactate dehydrogenase, COVID-19 and mortality. *Med Clin (Barc).* 2021;156(1):37.
6. Bennouar S, Bachir Cherif A, Kessira A, Hamel H, Boudahdir A, Bouamra A, Bennouar D, Abdi S. Usefulness of biological markers in the early prediction of corona virus disease-2019 severity. *Scand J Clin Lab Invest.* 2020;80(8):611-618. doi: 10.1080/00365513.2020.1821396.
7. Nahm FS. Receiver operating characteristic curve: overview and practical use for clinicians. *Korean J Anesthesiol.* 2022;75(1):25-36. doi: 10.4097/kja.21209.
8. Yao J, Bai T, Yang B, Sun L. The diagnostic value of D-dimer in acute aortic dissection: a meta-analysis. *J Cardiothorac Surg.* 2021;16(1):343. doi: 10.1186/s13019-021-01726-1.
9. Wussler D, Kozhuharov N, Tavares-Oliveira M, Bossa A, Sabti Z, Nowak A, et al. Clinical Utility of Procalcitonin in the Diagnosis of Pneumonia. *Clin Chem.* 2019;65(12):1532-1542. doi: 10.1373/clinchem.2019.306787.
10. Ye S, Lim JY, Huang W. Statistical considerations for repeatability and reproducibility of quantitative imaging biomarkers. *BJR Open.* 2022;4(1):20210083. doi: 10.1259/bjro.20210083.
11. Martínez-Camblor P, Pardo-Fernández JC. The Youden Index in the Generalized Receiver Operating Characteristic Curve Context. *Int J Biostat.* 2019;15(1):j/ijb.2019. doi: 10.1515/ijb-2018-0060.
12. Aguglia A, Natale A, Fusar-Poli L, Gnecco GB, Lechiara A, Marino M, et al. C-Reactive Protein as a Potential Peripheral Biomarker for High-Lethality Suicide Attempts. *Life (Basel).* 2022;12(10):1557. doi: 10.3390/life12101557.
13. Wu S, Du Z, Shen S, Zhang B, Yang H, Li X, et al. Identification and Validation of a Novel Clinical Signature to Predict the Prognosis in Confirmed Coronavirus Disease 2019 Patients. *Clin Infect Dis.* 2020;71(12):3154-3162. doi: 10.1093/cid/ciaa793.
14. Tristán-Martínez PD, Doubova SV, Saucedo-Valenzuela AL. Calidad de atención en pacientes con cáncer de próstata atendidos en México. *Rev Med Inst Mex Seguro Soc.* 2022;60(3):275-282. PMID: 35759755.
15. Zhu X, Xue J, Liu Z, Dai W, Xu H, Zhou Q, et al. The Lactate/Albumin Ratio Predicts Mortality in Critically Ill Patients with Acute Kidney Injury: An Observational Multicenter Study on the eICU Database. *Int J Gen Med.* 2021;14:10511-10525. doi: 10.2147/IJGM.S339767.
16. Yu W, Park E, Chang YC. Comparison of Paired ROC Curves through a Two-Stage Test. *J Biopharm Stat.* 2015;25(5):881-902. doi: 10.1080/10543406.2014.920874.
17. Song J, Li Z, Yao G, Wei S, Li L, Wu H. Framework for feature selection of predicting the diagnosis and prognosis of necrotizing enterocolitis. *PLoS One.* 2022;17(8):e0273383. doi: 10.1371/journal.pone.0273383.
18. Chen J, Mu F, Gao K, Yan C, Chen G, Guo C. Value of abdominal ultrasonography in predicting intestinal resection for premature infants with necrotizing enterocolitis. *BMC Gastroenterol.* 2022;22(1):524. doi: 10.1186/s12876-022-02607-0.
19. Liu Z, Li D, Liu X, Zhang B, Zang Y, Ma J, et al. Elevated Serum Procalcitonin to Predict Severity and Prognosis of Extensive Burns. *J Invest Surg.* 2022;35(7):1510-1518. doi: 10.1080/08941939.2022.2073489.
20. Stocker M, van Herk W, El Helou S, Dutta S, Schuerman FABA, van den Tooren-de Groot RK, et al. C-Reactive Protein, Procalcitonin, and White Blood Count to Rule Out Neonatal Early-onset Sepsis Within 36 Hours: A Secondary Analysis of the Neonatal Procalcitonin Intervention Study. *Clin Infect Dis.* 2021;73(2):e383-e390. doi: 10.1093/cid/ciaa876.