



RINCÓN DEL INVESTIGADOR

Artículo bilingüe español / inglés

Rev Esp Podol. 2024;35(1):69-70

DOI: <http://dx.doi.org/10.20986/revesppod.2024.1698/2024>

Sensibilidad, especificidad y valores predictivos (Parte II)

Sensitivity, specificity, and predictive values (Part II)

Javier Pascual Huerta

Clínica del Pie Elcano. Bilbao, España

En el anterior número de esta sección del Rincón del Investigador, introdujimos los términos de sensibilidad, especificidad y valores predictivos en pruebas diagnósticas dicotómicas utilizadas en salud. Cuando los estudios se refieren a estos conceptos para describir las características de un test, la simplicidad y familiaridad con la que se utilizan estas métricas enmascaran la existencia de un número de complejidades en las mismas que no suelen ser tenidas en cuenta. En esta sección hablaremos de dos ideas en la interpretación de estos conceptos: la interpretación y confusión que generan sensibilidad y especificidad, y cómo la prevalencia de la enfermedad afecta al valor predictivo positivo y negativo de un test.

La sensibilidad mide la proporción de personas con enfermedad que tienen un resultado positivo con el test de estudio o test de *screening* (sensibilidad = $VP / (VP + FN) \times 100$) (Tabla I). El valor de sensibilidad no puede aportar una recomendación certera en la toma de una decisión de un paciente concreto a pesar de que el resultado del test sea positivo, porque el test tiene falsos positivos que no son tenidos en cuenta en el cálculo de la sensibilidad. De hecho, los falsos positivos son ignorados para calcular la sensibilidad (solo las celdas de VP y FN se usan para calcular sensibilidad). Un resultado positivo, por sí mismo, incluso cuando el test tenga alta sensibilidad, no es realmente útil para decidir si una enfermedad está presente en un paciente concreto. De igual forma, la especificidad mide la proporción de personas sin enfermedad que tienen un resultado negativo del test de estudio (especificidad = $VN / (VN + FP) \times 100$). La especificidad de un test no aporta una indicación adecuada sobre un paciente con resultado negativo en el test porque los resultados negativos del test pueden contener falsos negativos que se ignoran a la hora de determinar la especificidad de dicho test (solo las celdas de VN y FP se usan para el cálculo de especificidad). Un resultado negativo en un test con alta especificidad no es, de ninguna manera, definitivo para descartar una enfermedad en un individuo particular. Estas ideas reflejan el error que comúnmente existe al creer que un resultado positivo en un test altamente sensitivo indica la presencia de una enfermedad o condición y que un resultado negativo en un test altamente específico indica la ausencia de la enfermedad o condición.

A pesar de esto, ambos conceptos de sensibilidad y especificidad pueden ser muy útiles si los valores son altos. Existe una relación inversa entre

verdaderos positivos y falsos negativos de tal forma que un test con muy alta sensibilidad indica un test en el que existen muchos verdaderos positivos y muy pocos falsos negativos. Es por esto que, en un test con alta sensibilidad, si da un resultado negativo, es muy raro que el paciente tenga la enfermedad o condición. Dicho de otra manera, cuando un test con alta sensibilidad da un resultado negativo permite descartar con bastante grado de seguridad que el individuo tenga la enfermedad. Esto ha llevado a la regla nemotécnica en inglés SNOOUT (*Sensibility, Negative, OUT*; nótese que la N en SNOOUT se refiere tanto a sensibilidad como a negativo). De la misma forma, en el caso de la especificidad existe una relación inversa entre verdaderos negativos y falsos positivos, de tal forma que un test con muy alta especificidad indica un test en el que existen muchos verdaderos negativos y muy pocos falsos positivos. Individuos que han resultado positivos en un test con alta especificidad es muy probable que tengan la enfermedad o condición. Dicho de otra manera, cuando un test con alta especificidad da un resultado positivo,

Tabla I. Tabla de contingencia obtenida de Tardáliga-García y cols. (n = 52)¹.

		Análisis histopatología (Gold standard)	
		Sí	No
Cultivo microbiología (Screening test)	Sí	33 VP	3 FP
	No	14 FN	2 VN

Sensibilidad = $VP / (VP + FN) = 33 / (33 + 14) = 0.70$
Especificidad = $VN / (VN + FP) = 2 / (2 + 3) = 0.40$
Valor predictivo positivo = $VP / (VP + FP) = 33 / (33 + 3) = 0.92$
Valor predictivo negativo = $VN / (VN + FN) = 2 / (2 + 14) = 0.13$

Recibido: 20-11-2023

Aceptado: 05-12-2023



0210-1238 © El autor, 2024.
Editorial: INSPIRA NETWORK GROUP S.L.
Este es un artículo Open Access bajo la licencia CC Reconocimiento 4.0 Internacional
(www.creativecommons.org/licenses/by/4.0/).

Correspondencia:

Javier Pascual Huerta
javier.pascual@hotmail.com

Tabla II. Tabla de contingencia hipotética con una prevalencia del 60 % (n = 52).

		Resultado histopatología (Gold standard)	
		Sí	No
Cultivo microbiología (Screening test)	Sí	21 VP	9 FP
	No	9 FN	13 VN

Sensibilidad = $VP / (VP + FN) = 21 / (21 + 9) = 0.70$
 Especificidad = $VN / (VN + FP) = 9 / (9 + 13) = 0.40$
 Valor Predictivo Positivo = $VP / (VP + FP) = 21 / (21 + 9) = 0.70$
 Valor Predictivo Negativo = $VN / (VN + FN) = 9 / (9 + 13) = 0.42$

permite asegurar con un grado importante de confianza que el individuo tiene la enfermedad. Esta idea ha derivado en la regla nemotécnica en inglés SPIN (*Specificity, Positive, IN-*; nótese que la P en SPIN se refiere tanto a especificidad como a positivo).

Estos nemotécnicos SNOUT & SPIN son una aplicación contraintuitiva de los conceptos de sensibilidad y especificidad que funcionan únicamente cuando los valores de ambos son altos. Un test de *screening* con alta sensibilidad no es necesariamente útil para encontrar enfermos. De hecho, es especialmente útil cuando el resultado es negativo en el test, ya que aporta una prueba fuerte que indica ausencia de enfermedad. Igualmente, un test con muy alta especificidad no es útil para descartar una enfermedad cuando no está presente. De hecho, es especialmente útil cuando el resultado es positivo para decidir que dicho paciente muy probablemente tenga la enfermedad.

La segunda idea de esta carta hace referencia a cómo los valores predictivos positivos (VPP) y negativos (VPN) están condicionados por la prevalencia de la enfermedad de la muestra estudiada. La sensibilidad se calcula utilizando únicamente los casos con enfermedad y especificidad utilizando únicamente los casos sin enfermedad según la prueba de referencia. Ambas son características del test de estudio y la prevalencia no afecta a sus resultados. Sin embargo, el cálculo de los VPP y VPN incluye individuos con enfermedad y sin enfermedad, por lo que su cálculo está afectado por la prevalencia de la enfermedad en la muestra. Tardáguila-García y cols. realizaron un estudio en 2021 para comparar la precisión diagnóstica del cultivo microbiológico (test de *screening*) con respecto al análisis histopatológico (*gold standard*) en pacientes diabéticos con sospecha de osteomielitis. La Tabla I presenta una tabla 2 x 2 mostrando los resultados obtenidos por los autores. Cada caso es asignado a uno de los cuatro recuadros de la tabla según haya sido su resultado en el cultivo microbiológico (positivo o negativo) y su resultado en el análisis histopatológico (positivo o negativo).

La prevalencia de la enfermedad en esta muestra utilizada por los autores es muy alta. Según el test de referencia (análisis histopatológico), 47 de los 52 casos presentaba osteomielitis en la muestra analizada (90.4 % de prevalencia). Imaginemos ahora que los resultados se hubieran hecho con una muestra en la que la prevalencia de la enfermedad hubiera sido más baja, por ejemplo en torno al 60 %, y calculemos los estadísticos en base a esta nueva hipotética prevalencia. La Tabla II muestra unos resultados hipotéticos del estudio de Tardáguila-García y cols. en los que se ha modificado los casos para disminuir la prevalencia de la enfermedad a 31 casos con osteomielitis (59.6 % de prevalencia) pero manteniendo los valores de sensibilidad (0.70) y especificidad (0.40) del estudio.

En la Tabla II, los VPP y VPN cambian de forma importante con respecto al estudio original de los autores. El VPP pasa de 0.92 a 0.70 y el VPN pasa

de 0.13 a 0.42. Para un clínico el punto importante de los estudios que valoran la eficacia de test diagnósticos es si los pacientes que han sido positivos (o negativos) en el test pueden ser diagnosticados o no de la enfermedad. En el estudio original un 92 % de los pacientes que tenían un cultivo positivo tenían osteomielitis. En el ejemplo hipotético, este porcentaje era solo de un 70 %. De los que tenían un test negativo en el estudio original solo un 13 % no tenía la enfermedad (un 87 % tenían osteomielitis a pesar de tener un resultado negativo), mientras que en el ejemplo inventado este porcentaje era del 42 %. Estos datos ilustran el hecho de que la capacidad de un test de realizar un diagnóstico concreto a partir de sus resultados depende del valor discriminatorio del test y de la prevalencia de la enfermedad en la muestra estudiada. Si la prevalencia de la enfermedad es muy alta en la muestra (más que en la población normal) el VPP tiende a sobreestimarse y el VPN tiende a infravalorarse y viceversa en el caso contrario.

Como punto final, en estudios de eficacia de pruebas diagnósticas, el test a estudiar es comparado con lo que se piensa como indicador definitivo, comúnmente referido como *gold standard*. Las palabras "*gold standard*" sugieren que dicho test aporta una evidencia presumiblemente indiscutible acerca de si la enfermedad existe o no. Sin embargo, pueden existir dudas con respecto a la validez de los llamados *gold standards* y en el caso de la validez del estudio histopatológico en el diagnóstico de OM existen². Es por esto que estos test han comenzado a denominarse de forma menos entusiasta como "standard de referencia". En esta y la anterior carta hemos usado el término *gold standard*, aunque la denominación correcta por las razones expuestas actualmente es "estándar de referencia".

Puntos clave

- La sensibilidad y la especificidad suelen tender a confundirse en su interpretación a la hora de analizar el resultado de un test en un paciente concreto
- Un resultado negativo en un test con alta sensibilidad es una prueba fuerte para excluir la enfermedad (SNOUT). Un resultado positivo en un test con alta especificidad es una prueba fuerte para diagnosticar la enfermedad (SPIN)
- La sensibilidad y la especificidad son características del test de estudio y la prevalencia no afecta a sus resultados. Sin embargo, el cálculo de los VPP y VPN incluye individuos con enfermedad y sin enfermedad, por lo que su cálculo está afectado por la prevalencia de la enfermedad en la muestra estudiada

Bibliografía

1. Tardáguila-García A, Sanz-Corbalán I, García-Morales E, García-Álvarez Y, Molines-Barroso RJ, Lázaro-Martínez JL. Diagnostic Accuracy of Bone Culture Versus Biopsy in Diabetic Foot Osteomyelitis. *Adv Skin Wound Care*. 2021;34(4):204-8. DOI: 10.1097/01.ASW.0000734376.32571.20.
2. Meyr AJ, Singh S, Zhang X, Khilko N, Mukherjee A, Sheridan MJ, Khurana JS. Statistical reliability of bone biopsy for the diagnosis of diabetic foot osteomyelitis. *J Foot Ankle Surg*. 2011;50(6):663-7. DOI: 10.1053/j.jfas.2011.08.005.

Bibliografía recomendada

- Carvajal DN, Rowe PC. Sensitivity, specificity, predictive values, and likelihood ratios. *Pediatr Rev*. 2010;31(12):511-3.
- Ghaalip Lalkhen A, McCluskey A. Clinical test: sensitivity and specificity. *Continuing Education in Anaesthesia, Critical Care & Pain*. 2008;8(6). DOI: 10.1093/bjaceaccp/mkn041. DOI: 10.1093/bjaceaccp/mkn041.
- Trevethan R. Sensitivity, Specificity, and Predictive Values: Foundations, Plausibilities, and Pitfalls in Research and Practice. *Front Public Health*. 2017;5:307. DOI: 10.3389/fpubh.2017.00307.