



# REVISTA ESPAÑOLA DE PODOLOGÍA

Publicación Oficial del Consejo General de Colegios Oficiales de Podólogos

**Artículo Aceptado para su pre-publicación / Article Accepted for pre-publication**

**Título / Title:**

Sensibilidad, especificidad y valores predictivos (Parte II) / Sensitivity, specificity and predictive values (Part II)

**Autores / Authors:**

Javier Pascual Huerta

DOI: [10.20986/revesppod.2024.1698/2024](https://doi.org/10.20986/revesppod.2024.1698/2024)

**Instrucciones de citación para el artículo / Citation instructions for the article:**

Pascual Huerta Javier. Sensibilidad, especificidad y valores predictivos (Parte II) / Sensitivity, specificity and predictive values (Part II). Rev. Esp. Pod. 2024. doi: 10.20986/revesppod.2024.1698/2024.



Este es un archivo PDF de un manuscrito inédito que ha sido aceptado para su publicación en la Revista Española de Podología. Como un servicio a nuestros clientes estamos proporcionando esta primera versión del manuscrito en estado de pre-publicación. El manuscrito será sometido a la corrección de estilo final, composición y revisión de la prueba resultante antes de que se publique en su forma final. Tenga en cuenta que durante el proceso de producción se pueden dar errores lo que podría afectar el contenido final.



## 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)<sup>1</sup>.

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

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/](http://www.creativecommons.org/licenses/by/4.0/)).

Correspondencia:

Javier Pascual Huerta  
[javier.pascual@hotmail.com](mailto: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 <b>VP</b>	13 <b>FP</b>
	No	9 <b>FN</b>	9 <b>VN</b>

Sensibilidad =  $VP / (VP + FN) = 21 / (21 + 9) = 0.70$   
 Especificidad =  $VN / (VN + FP) = 9 / (9 + 13) = 0.40$   
 Valor Predictivo Positivo =  $VP / (VP + FP) = 21 / (21 + 13) = 0.62$   
 Valor Predictivo Negativo =  $VN / (VN + FN) = 9 / (9 + 9) = 0.50$

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 SNOOUT & 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 30 casos con osteomielitis (57.7 % 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 0.62 y el VPN pasa

de 0.13 a 0.50. 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 62 %. 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 50 %. 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 existen<sup>2</sup>. 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 (SNOOUT). 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.



## RESEARCHER'S CORNER

Bilingual article English/Spanish

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

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

## Sensitivity, specificity, and predictive values (Part II)

### Sensibilidad, especificidad y valores predictivos (Parte II)

Javier Pascual Huerta

Clínica del Pie Elcano. Bilbao, España

In the previous issue of this section of the Researcher's Corner, we introduced the terms of sensitivity, specificity, and predictive values in dichotomous diagnostic tests used in health. When studies refer to these concepts to describe the characteristics of a test, the simplicity and familiarity with which these metrics are used mask the existence of a number of complexities that are usually not considered. In this section, we will discuss two ideas in the interpretation of these concepts: the interpretation and confusion generated by sensitivity and specificity, and how the prevalence of the disease affects the positive and negative predictive value of a test.

Sensitivity measures the proportion of people with the disease who have a positive result with the study or screening test ( $\text{sensitivity} = \text{TP} / (\text{TP} + \text{FN}) \times 100$ ) (Table I). The sensitivity value cannot provide a definitive recommendation in making a decision for a specific patient, even if the test result is positive, because the test has false positives that are not considered in the calculation of sensitivity. In fact, false positives are ignored when calculating sensitivity (only the TP and FN cells are used to calculate sensitivity). A positive result, by itself, even when the test has high sensitivity, is not really useful for making a decision on whether a disease is present in a specific patient. Similarly, specificity measures the proportion of people without the disease who have a negative result from the study test ( $\text{specificity} = \text{TN} / (\text{TN} + \text{FP}) \times 100$ ). The specificity of a test does not provide an adequate indication for a patient with a negative test result because negative test results can contain false negatives that are ignored when determining the specificity of the test (only the TN and FP cells are used to calculate specificity). A negative result in a highly specific test is by no means definitive for ruling out a disease in a particular individual. These ideas reflect the common error of believing that a positive result in a highly sensitive test indicates the presence of a disease or condition, and that a negative result in a highly specific test indicates the absence of the disease or condition.

Despite this, both concepts of sensitivity and specificity can be very useful if the values are high. There is an inverse relationship between true positives and false negatives such that a test with very high sensitivity indicates a

test with many true positives and very few false negatives. This is why, in a test with high sensitivity, if it gives a negative result, it is very rare that the patient has the disease or condition. In other words, when a test with high sensitivity gives a negative result, it allows one to rule out with a considerable degree of certainty that the individual has the disease. This has led to the mnemonic rule SNOUT (Sensitivity, Negative, OUT; note that the N in SNOUT refers to both sensibility and negative). Similarly, in the case of specificity, there is an inverse relationship between true negatives and false positives such that a

**Table I. Contingency table obtained from Tardáliga-García et al. (n = 52)<sup>1</sup>.**

		Histopathology analysis (Gold standard)	
		YES	NO
Microbiology culture (Screening test)	YES	33 <b>TP</b>	3 <b>FP</b>
	NO	14 <b>FN</b>	2 <b>TN</b>

$\text{SENSITIVITY} = \text{TP} / (\text{TP} + \text{FN}) = 33 / (33 + 14) = 0.70$

$\text{SPECIFICITY} = \text{TN} / (\text{TN} + \text{FP}) = 2 / (2 + 3) = 0.40$

$\text{POSITIVE PREDICTIVE VALUE} = \text{TP} / (\text{TP} + \text{FP}) = 33 / (33 + 3) = 0.92$

$\text{NEGATIVE PREDICTIVE VALUE} = \text{TN} / (\text{TN} + \text{FN}) = 2 / (2 + 14) = 0.13$

Received: 20-11-2023

Accepted: 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/](http://www.creativecommons.org/licenses/by/4.0/)).

Correspondence:

Javier Pascual Huerta  
[javier.pascual@hotmail.com](mailto:javier.pascual@hotmail.com)

**Table II. Hypothetical contingency table with a 60 % Prevalence (n = 52).**

Microbiology culture (Screening test)	Histopathology analysis (Gold standard)	
	YES	NO
	YES	21 <b>TP</b>
NO	9 <b>FN</b>	9 <b>TN</b>

SENSITIVITY = TP / (TP + FN) = 21 / (21 + 9) = 0.70

SPECIFICITY = TN / (TN + FP) = 9 / (9 + 13) = 0.40

POSITIVE PREDICTIVE VALUE = TP / (TP + FP) = 21 / (21 + 13) = 0.62

NEGATIVE PREDICTIVE VALUE = TN / (TN + FN) = 9 / (9 + 9) = 0.50

test with very high specificity indicates a test with many true negatives and very few false positives. Individuals who have tested positive in a highly specific test are very likely to have the disease or condition. In other words, when a test with high specificity gives a positive result, it allows one to ensure with a significant degree of confidence that the individual has the disease. This idea has led to the mnemonic rule SPIN (Specificity, Positive, IN-; note that the P in SPIN refers to both specificity and positive).

These mnemonics, SNOUT & SPIN, are a counterintuitive application of the concepts of sensitivity and specificity that only work when both values are high. A screening test with high sensitivity is not necessarily useful for finding patients. In fact, it is especially useful when the test result is negative because it provides strong evidence indicating the absence of disease. Similarly, a test with very high specificity is not useful for ruling out a disease when it is not present. In fact, it is especially useful when the result is positive for deciding that the patient most likely has the disease.

The second idea of this letter refers to how positive predictive values (PPV) and negative predictive values (NPV) are conditioned by the prevalence of the disease in the sample studied. Sensitivity is calculated using only the cases with disease, and specificity using only the cases without disease according to the reference test. Both are characteristics of the study test, and prevalence does not affect their results. However, the calculation of PPV and NPV includes individuals with and without the disease, so their calculation is affected by the prevalence of the disease in the sample. Tardáguila-García et al. conducted a study in 2021 to compare the diagnostic accuracy of microbiological culture (screening test) with histopathological analysis (Gold standard) in diabetic patients with suspected osteomyelitis. Table I presents a 2 × 2 table showing the results obtained by the authors. Each case is assigned to one of the four boxes of the table according to its result in the microbiological culture (positive or negative) and its result in the histopathological analysis (positive or negative):

The prevalence of the disease in this sample used by the authors is very high. According to the reference test (histopathological analysis), 47 of the 52 cases had osteomyelitis in the analyzed sample (90.4 % prevalence). Now let's imagine that the results had been obtained with a sample in which the prevalence of the disease was lower, for example around 60%, and calculate the statistics based on this new hypothetical prevalence. Table II shows hypothetical results of the Tardáguila-García et al. study in which the cases have been modified to decrease the disease prevalence to 30 cases with osteomy-

elitis (exactly 57.7% of prevalence) while maintaining the sensitivity (0.70) and specificity (0.40) values of the study.

In Table II, PPV and NPV change significantly compared to the authors' original study. PPV drops from 0.92 down to 0.62, while NPV goes up from 0.13 up to 0.50. For a clinician, the important point of studies evaluating the efficacy of diagnostic tests is whether patients who have tested positive (or negative) can be diagnosed with the disease. In the original study, 92 % of patients with a positive culture had osteomyelitis. In the hypothetical example, this percentage was only 62 %. Of those who had a negative test in the original study, only 13 % did not have the disease (87 % had osteomyelitis despite having a negative result), while in the invented example, this percentage was 50 %. These data illustrate the fact that the ability of a test to make a specific diagnosis based on its results depends on the discriminatory value of the test and the prevalence of the disease in the sample studied. If the disease prevalence is very high in the sample (more than in the normal population), the PPV tends to be overestimated, and the NPV tends to be underestimated, and vice versa in the opposite case.

As a final point, in studies of the efficacy of diagnostic tests, the test to be studied is compared with what is thought to be the definitive indicator, commonly referred to as the gold standard. The words "gold standard" suggest that this test provides presumably indisputable evidence of whether the disease exists or not. However, there may be doubts about the validity of the so-called gold standards, actually in the case of histopathological study in the diagnosis of OM, there are some<sup>2</sup>. This is why these tests have begun to be referred to less enthusiastically as "reference standards". In this and the previous letter, we have used the term gold standard, although the correct denomination for the reasons stated is currently "reference standard".

### Key points

- Sensitivity and specificity often tend to be confused in their interpretation when analyzing the result of a test in a specific patient.
- A negative result in a test with high sensitivity is strong evidence to exclude the disease (SNOUT). A positive result in a test with high specificity is strong evidence to diagnose the disease (SPIN).
- Sensitivity and specificity are characteristics of the study test and their results are not affected by prevalence. However, the calculation of PPN and NPV includes individuals with and without the disease, so their calculation is affected by the prevalence of the disease in the sample studied.

### References

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.

### Recommended references

- 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, Pliabilities, and Pitfalls in Research and Practice. *Front Public Health*. 2017;5:307. DOI: 10.3389/fpubh.2017.00307.