

Aplicación del Teorema de Bayes en la Interpretación Clínica del PHQ-9 para el Diagnóstico de Depresión

Application of Bayes' Theorem in the Clinical Interpretation of the PHQ-9 for the Diagnosis of Depression

Sofía Ricaurte Peñafiel¹

Resumen

La interpretación de pruebas diagnósticas en salud mental suele realizarse sin considerar la prevalencia de la enfermedad, lo que puede afectar la estimación de la probabilidad real de diagnóstico. El presente estudio analizó la variación de los valores predictivos positivo y negativo del Patient Health Questionnaire-9 (PHQ-9) en función de diferentes niveles de prevalencia, empleando valores de sensibilidad de 0.88 y especificidad de 0.85 basados en estudios de validación y metaanálisis previos (Kroenke et al., 2001; Levis et al., 2019), con escenarios de prevalencia del 5%, 10%, 20%, 30% y 50%. Los resultados mostraron que el valor predictivo positivo aumenta conforme se incrementa la prevalencia, pasando de 0.236 a 0.854, mientras que el valor predictivo negativo disminuye de 0.993 a 0.876. Se concluye que la interpretación del PHQ-9 debe considerar el contexto epidemiológico, y que el Teorema de Bayes constituye una herramienta clave para mejorar la toma de decisiones clínicas.

Palabras clave: depresión, PHQ-9, Teorema de Bayes, pruebas diagnósticas, prevalencia.

Abstract

The interpretation of diagnostic tests in mental health is frequently conducted without considering disease prevalence, which may distort the estimation of the true probability of diagnosis. This study analyzed the variation of the positive and negative predictive values of the Patient Health Questionnaire-9 (PHQ-9) across different prevalence levels, using sensitivity and specificity values of 0.88 and 0.85, respectively, based on previous validation studies and meta-analyses (Kroenke et al., 2001; Levis et al., 2019), across five prevalence scenarios: 5%, 10%, 20%, 30%, and 50%. Results showed that the positive predictive value increases as prevalence rises, from 0.236 to 0.854, while the negative predictive value decreases from 0.993 to 0.876. It is concluded that PHQ-9 results must be interpreted in light of the epidemiological context, and that Bayes' Theorem constitutes a key tool for improving clinical decision-making.

Keywords: depression, PHQ-9, Bayes' Theorem, diagnostic tests, prevalence.

1. Filiación: Universidad ECOTEC-Guayaquil. ORCID: <https://orcid.org/0009-0009-9822-5788>. sricaurtep@ecotec.edu.ec



INTRODUCCIÓN

La depresión constituye una de las principales causas de discapacidad a nivel mundial, afectando a más de 332 millones de personas (Organización Mundial de la Salud, 2025), lo que la posiciona como un problema prioritario de salud pública.

En la práctica clínica, herramientas de tamizaje como el Patient Health Questionnaire-9 (PHQ-9) son ampliamente utilizadas para la detección de posibles casos. Sin embargo, la interpretación de sus resultados suele realizarse sin considerar el contexto de prevalencia en el que se aplican, lo que puede conducir a errores en la estimación de la probabilidad real de enfermedad.

Ante esta limitación, el Teorema de Bayes ofrece un marco para integrar la sensibilidad, la especificidad y la prevalencia, permitiendo estimar con mayor precisión la probabilidad posterior de enfermedad tras un resultado positivo (Akobeng, 2007).

El presente estudio analiza la variación de los valores predictivos positivo y negativo del PHQ-9 en distintos escenarios de prevalencia, a partir de parámetros tomados de estudios de validación y metaanálisis previos (Kroenke et al., 2001; Levis et al., 2019).

METODOLOGÍA

Se empleó un diseño de simulación determinístico con parámetros fijos tomados de la literatura:

sensibilidad de 0.88 y especificidad de 0.85, en concordancia con los reportados por Levis et al. (2019) en un metaanálisis de datos individuales, y consistentes con los valores de sensibilidad reportados por Kroenke et al. (2001) en el estudio de validación original del PHQ-9.

Se definieron cinco escenarios de prevalencia (0.05, 0.10, 0.20, 0.30 y 0.50), abarcando desde contextos de población general hasta escenarios de alta sospecha diagnóstica.

A partir de estos parámetros se estimaron el valor predictivo positivo y el valor predictivo negativo mediante la aplicación del Teorema de Bayes.

RESULTADOS

Para una prevalencia del 5%, el valor predictivo positivo del PHQ-9 fue de 0.236. Este valor se incrementa progresivamente conforme aumenta la prevalencia, alcanzando 0.395 para una prevalencia del 10%, 0.595 para el 20%, 0.715 para el 30% y 0.854 para el 50%.

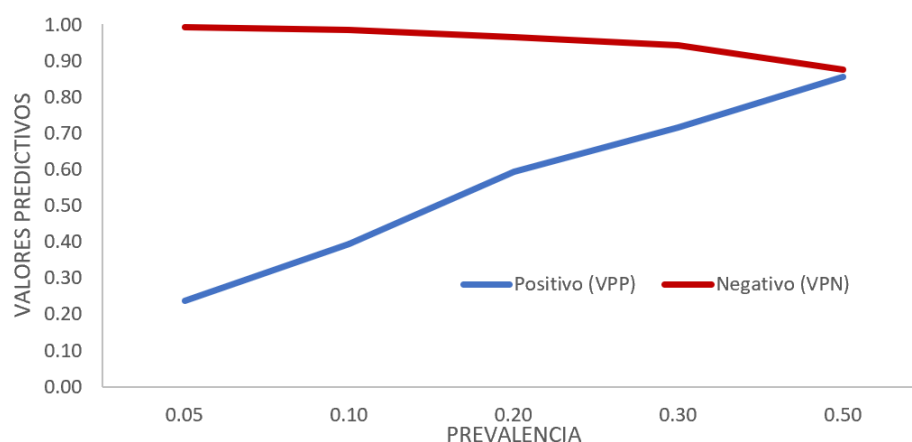
En cuanto al valor predictivo negativo, el comportamiento fue inverso: partiendo de 0.993 para una prevalencia del 5%, disminuyó progresivamente hasta 0.876 para una prevalencia del 50%.

Tabla 1

Valor predictivo positivo y negativo del PHQ-9 según diferentes niveles de prevalencia

| Prevalencia | Sensibilidad | Especificidad | Valor Predictivo Positivo (VPP) | Valor Predictivo Negativo (VPN) |
|-------------|--------------|---------------|---------------------------------|---------------------------------|
| 0.05 | 0.88 | 0.85 | 0.236 | 0.993 |
| 0.10 | 0.88 | 0.85 | 0.395 | 0.985 |
| 0.20 | 0.88 | 0.85 | 0.595 | 0.966 |
| 0.30 | 0.88 | 0.85 | 0.715 | 0.943 |
| 0.50 | 0.88 | 0.85 | 0.854 | 0.876 |

Nota. Cálculos basados en sensibilidad = 0.88 y especificidad = 0.85.

Figura 1*Variación de los valores predictivos del PHQ-9 según la prevalencia*

Como se aprecia en la Figura 1, la brecha entre ambas curvas es máxima en contextos de baja prevalencia, donde el VPN (0.993) contrasta marcadamente con el VPP (0.236), y se reduce de forma progresiva hasta casi converger en prevalencia del 50%, con valores de 0.876 y 0.854, respectivamente. El VPN permanece por encima de 0.970 únicamente en los escenarios de 5% y 10%, descendiendo a 0.966 en el 20% y continuando su caída hasta 0.876 en el 50%, mientras que el VPP pasa de 0.236 a 0.595 en el intervalo del 5% al 20% de prevalencia, concentrando en él su mayor incremento absoluto.

DISCUSIÓN

Los resultados obtenidos confirman que el valor predictivo positivo no es una propiedad intrínseca de la prueba, sino que depende del contexto en el que se aplica. Un resultado positivo en una población de baja prevalencia tiene un significado clínico muy distinto al mismo resultado obtenido en un contexto de alta sospecha diagnóstica.

En la práctica clínica, es frecuente interpretar los resultados de tamizaje sin considerar la prevalencia, lo que puede conducir a sobreestimaciones o subestimaciones del

riesgo real. Esto es especialmente relevante en herramientas como el PHQ-9, ampliamente utilizadas en diferentes niveles de atención. En contextos de baja prevalencia, esta situación puede generar un número considerable de falsos positivos, con implicaciones no solo diagnósticas sino también en la gestión de recursos y en la calidad de la atención.

En cuanto al valor predictivo negativo, su alta confiabilidad en contextos de baja prevalencia indica que la prueba es especialmente útil para descartar la enfermedad en población general.

Sin embargo, en escenarios de alta prevalencia — como unidades especializadas en salud mental— un resultado negativo no es suficiente para excluir el diagnóstico, lo que exige cautela clínica y la consideración de evaluaciones complementarias.

Desde una perspectiva educativa, estos hallazgos resaltan la importancia de formar a los estudiantes de ciencias de la salud no solo en el uso de herramientas diagnósticas, sino también en su correcta interpretación según el contexto epidemiológico, lo que fortalece el razonamiento clínico y reduce errores en la toma de decisiones.

Este estudio presenta como limitación el uso de datos simulados y parámetros tomados

de la literatura, por lo que los resultados no corresponden a una población específica. No obstante, este enfoque permite ilustrar con claridad el impacto de la prevalencia en la interpretación de pruebas diagnósticas.

CONCLUSIONES

Los resultados del presente estudio demuestran que los valores predictivos del PHQ-9 dependen en gran medida de la prevalencia de depresión en la población evaluada.

A medida que la prevalencia aumenta, el valor predictivo positivo se incrementa de forma considerable, mientras que el valor predictivo negativo disminuye, lo que confirma que ambos indicadores deben interpretarse siempre en función del contexto epidemiológico y no como propiedades fijas de la prueba.

En consecuencia, el uso del Teorema de Bayes resulta indispensable para una interpretación clínica rigurosa del PHQ-9, ya que integra la sensibilidad, la especificidad y la prevalencia en una estimación única de la probabilidad real de enfermedad, evitando los errores derivados de considerar estos parámetros de forma aislada.

REFERENCIAS BIBLIOGRÁFICAS

Akobeng, A. K. (2007). Understanding diagnostic tests 1: Sensitivity, specificity and predictive values. *Acta Paediatrica*, 96(3), 338–341. <https://doi.org/10.1111/j.1651-2227.2006.00180.x>

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, 16(9), 606–613. <https://doi.org/10.1046/j.1525-1497.2001.016009606.x>

Levis, B., Benedetti, A., & Thombs, B. D. (2019). Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression: Individual participant data meta-analysis. *BMJ*, 365, l1476. <https://doi.org/10.1136/bmj.l1476>

Organización Mundial de la Salud. (2025, agosto 29). Trastorno Depresivo (depresión). <https://www.who.int/es/news-room/fact-sheets/detail/depression>