



Artículo original

Validez del Cuestionario de salud del paciente-9 (PHQ-9) para cribado de depresión en adultos usuarios de Atención Primaria en Bucaramanga, Colombia



Carlos Arturo Cassiani-Miranda^{a,*}, Angy Karina Cuadros-Cruz^b, Harold Torres-Pinzón^c, Orlando Scoppetta^d, Jhon Henry Pinzón-Tarrazona^e, Wendy Yulieth López-Fuentes^e, Andrea Paez^b, Diego Fernando Cabanzo-Arenas^f, Sergio Ribero-Marulanda^e y Elkin René Llanes-Amaya^a

^a Grupo de investigación Neurociencias, Universidad de Santander (UDES), Bucaramanga, Colombia

^b Instituto de Salud de Bucaramanga (ISABU), Hospital local del Norte, Bucaramanga, Colombia

^c Grupo de Investigación Salud ComuniUdes, Universidad de Santander (UDES), Bucaramanga, Colombia

^d Universidad Católica de Colombia, Bogotá DC, Colombia

^e Universidad Autónoma de Bucaramanga (UNAB), Floridablanca, Santander, Colombia

^f Facultad de Medicina, Universidad de Santander (UDES), Bucaramanga, Colombia

INFORMACIÓN DEL ARTÍCULO

Historia del artículo:

Recibido el 27 de febrero de 2019

Aceptado el 20 de septiembre de 2019

On-line el 25 de enero de 2020

Palabras clave:

PHQ-9

Reproducibilidad de los resultados

Cribado

Depresión

Atención primaria de salud

Colombia

R E S U M E N

El Cuestionario de salud del paciente-9 (PHQ-9) es uno de los instrumentos de autoinforme más utilizado en Atención Primaria (AP). No existe validez de criterio del PHQ-9 en Colombia. El objetivo fue realizar la validez de criterio del PHQ-9 como instrumento de cribado en AP. Se realizó un estudio trasversal de validez de criterio de una escala usando como criterio de referencia la minientrevista neuropsiquiátrica (MINI) en usuarios adultos de centros de AP de ambos sexos. Se calcularon la consistencia interna y la validez convergente y de criterio del PHQ-9 mediante el análisis de las características operativas del receptor (COR) y el área bajo la curva (ABC). Participaron 243 pacientes, 184 (75,7%) fueron de sexo femenino. El promedio de edad fue 34,05 (mediana 31 y DE = 12,47). El α de Cronbach fue 0,80 y ω de McDonald, 0,81. La rho de Spearman fue 0,64 para HADS-D ($p < 0,010$) y 0,70 para PHQ-2 ($p < 0,010$). El ABC fue 0,92 (IC del 95%: 0,880-0,963). El punto de corte óptimo del PHQ-9 fue ≥ 7 : sensibilidad de 90,38 (IC del 95%: 81,41-99,36); especificidad de 81,68 (IC del 95%: 75,93-87,42); el VPP 57,32 (IC del 95%: 46,00-68,63); el VPN 96,89 (IC del 95%: 93,90-99,88); índice de Youden 0,72 (IC del 95%: 0,62-0,82; LR+ 4,93 (IC del 95%: 3,61-6,74); LR- 0,12 (IC del 95%: 0,005-0,270). En conclusión, la versión colombiana del PHQ-9 es un instrumento válido y confiable para el cribado de depresión en AP de Bucaramanga, con un punto de corte ≥ 7 .

© 2019 Asociación Colombiana de Psiquiatría. Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

* Autor para correspondencia.

Correo electrónico: ca.cassiani@mail.udes.edu.co (C.A. Cassiani-Miranda).

<https://doi.org/10.1016/j.rcp.2019.09.001>

0034-7450/© 2019 Asociación Colombiana de Psiquiatría. Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

Validity of the Patient Health Questionnaire-9 (PHQ-9) for depression screening in adult primary care users in Bucaramanga, Colombia

A B S T R A C T

Keywords:

PHQ-9
 Reproducibility of results
 Screening
 Depression
 Primary healthcare
 Colombia

The patient health questionnaire-9 (PHQ-9) is one of the most widely used self-report instruments in primary care. There is no criterion validity of the PHQ-9 in Colombia. The objective was to validate the PHQ-9 as a screening tool in primary care. A cross-sectional, scale criterion validity study was performed using as reference criterion the mini neuropsychiatric interview (MINI) in male and female adult users of primary care centres. We calculated the internal consistency and convergent and criterion validity of the PHQ-9 by analysing the receiver operating characteristics (ROC) and the area under the curve (AUC). We analysed 243 participants; 184 (75.7%) were female. The average age was 34.05 (median of 31 and SD = 12.47). Cronbach's α was 0.80 and McDonald's ω was 0.81. Spearman's Rho was 0.64 for HADS-D ($P < 0.010$) and 0.70 for PHQ-2 ($P < 0.010$). The AUC was 0.92 (95% CI 0.880-0.963). The optimal cut-off point of PHQ-9 was ≥ 7 : sensitivity of 90.38 (95% CI: 81.41-99.36); specificity of 81.68 (95% CI: 75.93-87.42); PPV 57.32 (95% CI: 46.00-68.63); NPV 96.89 (95% CI: 93.90-99.88); Youden index 0.72 (95% CI: 0.62-0.82); LR+ 4.93 (95% CI: 3.61-6.74); LR- 0.12 (95% CI: 0.005-0.270). In sum, the Colombian version of PHQ-9 is a valid and reliable instrument for depression screening in primary care in Bucaramanga, with a cut-off point ≥ 7 .

© 2019 Asociación Colombiana de Psiquiatría. Published by Elsevier España, S.L.U. All rights reserved.

Introducción

La depresión es un problema mayor de salud pública a nivel mundial¹ y tiene un impacto significativo sobre la calidad de vida², produce altos niveles de morbilidad³, disminución de la expectativa de vida⁴ y exceso de mortalidad⁵. La prevalencia de episodio depresivo mayor (EDM) a lo largo de la vida es del 11,2%⁶. Las prevalencias tienden a ser más altas países de bajos y medianos ingresos como Pakistán, en donde se han informado prevalencias de depresión del 45,9%⁷. En Atención Primaria (AP), la prevalencia de EDM presenta significativas variaciones ubicándose en el rango entre el 4,5 y el 47,8%⁸.

En la Encuesta de salud mental de Colombia 2015, la prevalencia de depresión mayor en la población general fue de 5,4 (IC del 95%: 4,6-6,4), 2,3 (IC del 95%: 1,8-2,9) y 0,8 (IC del 95%: 0,5-1,3) a lo largo de la vida, en el último y en último mes, respectivamente⁹. En Bucaramanga, la prevalencia de síntomas depresivos con importancia clínica (SDIC) fue del 22,3% (IC del 95%: 20,0-24,6) y de TDM del 11,2% (IC del 95%: 9,7-12,9%)¹⁰. Un estudio poblacional posterior en adultos residentes de Bucaramanga (N = 266), mediante entrevista estructurada del eje I (SCID-I), informó una prevalencia del 16,5% (IC del 95%: 12,3-21,6)¹¹, lo que confirma la alta prevalencia de depresión en esta región.

A pesar de su alta carga, cronicidad y carácter recurrente, la depresión es subdiagnosticada en AP, ya que aproximadamente el 50% de los pacientes que presentan depresión no serían detectados¹². Esta brecha diagnóstica podría explicarse por el hecho de que más del 75% de los pacientes con depresión consultan inicialmente a su médico familiar o de AP con escaso entrenamiento en la identificación de trastornos depresivos¹³, las limitaciones de tiempo en entornos

congestionados de AP¹⁴ y la escasez de instrumentos de cribado validados en países de bajos y medianos ingresos¹⁵.

Por lo anterior, se han desarrollado programas para el reconocimiento de la depresión^{16,17} que recomiendan instrumentos estandarizados. Existen diversos instrumentos para identificar casos de depresión; sin embargo, sus beneficios no se han determinado plenamente y la literatura muestra resultados contradictorios¹⁸. Una revisión sistemática reciente sugiere que dentro de los instrumentos de cribado, solo el Cuestionario de salud del paciente-9 (PHQ-9) satisface el nivel óptimo de exactitud para depresión¹⁹. El PHQ-9 es una escala de tipo adjetival derivada de la Primary Care Evaluation of Mental Disorders (PRIME-MD) para evaluar síntomas depresivos con los criterios del DSM-IV²⁰. El PHQ-9 es más corto que la mayoría de las escalas de cribado de depresión^{21,22} y se considera como el mejor instrumento de cribado para depresión en AP por su exactitud, brevedad, ser de dominio público y multipropósito, facilidad de aplicación, puntuación e interpretación^{19,23}. El PHQ-9 cuenta con traducción en más de 20 idiomas y se ha utilizado en muchos países y contextos²⁴. En AP, la sensibilidad del PHQ-9 estuvo entre 0,71 a 0,84 (media de 0,77) y la especificidad entre 0,90 y 0,97 (media de 0,94)²⁵, lo que confirma un adecuado desempeño psicométrico en AP, aunque con algunas variaciones en el punto de corte (PC) y los parámetros psicométricos que se explican por la influencia de aspectos culturales en el patrón de respuesta²³. Su uso extendido se sustenta además por los hallazgos de Williams et al., quienes concluyeron, en un análisis de más de 38 estudios con más de 32.000 pacientes de AP, que el PHQ-9 fue igual o superior a otras mediciones de depresión²². Además, el grupo de trabajo para TDM del DSM-5 y las guías NICE consideran el PHQ-9 como la medición preferida para evaluar la presencia de depresión y cuantificar su severidad^{21,22,26}.

El PHQ-9 se ha evaluado en Colombia en estudiantes universitarios²⁷; sin embargo, no se comparó con patrón de oro, por lo tanto es necesario realizar la validez de criterio del PHQ-9 en AP en Colombia empleando un patrón de oro, sobre todo por la oportunidad que representan los servicios de AP en la detección temprana de la depresión²⁸. Por consiguiente, el objetivo de este estudio fue evaluar validez de criterio del PHQ-9, comparado con la Minientrevista Estructurada Neuropsiquiátrica (MINI) para el cribado de sintomatología depresiva en adultos usuarios de AP del área metropolitana de Bucaramanga.

Materiales y métodos

Diseño

Este estudio fue diseñado y analizado siguiendo las recomendaciones de la declaración Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2)²⁹. Se realizó un estudio observacional analítico de validez de criterio de una escala usando criterio de referencia.

Participantes

Se incluyó a usuarios de AP de la localidad de ambos sexos de 18 a 65 años. Los centros de AP pertenecen al Instituto de Salud de Bucaramanga (ISABU), que es una empresa social del Estado que coordina la prestación de servicios primarios de salud del área metropolitana de Bucaramanga.

Se excluyó a los sujetos con síntomas psicóticos, con deterioro cognitivo, delirium o discapacidad intelectual que impidiera responder los instrumentos, estar bajo efectos de sustancias psicoactivas, con alteración funcional de la visión o audición que impidiera comprender el contenido de la encuesta, y aquellos que no comprendieran el idioma español. Se calculó el tamaño de muestra para evaluar hipótesis sobre características de una prueba diagnóstica³⁰:

$$N = \frac{\left\{ Z_{1-\alpha/2} \sqrt{[\pi_1(1-\pi)]} | Z_{1-\beta} \sqrt{[\pi_2(1-\pi_2)]} \right\}^2}{\delta^2}$$

donde, π_1 es la sensibilidad del estándar (0,96) y π_2 es la sensibilidad esperada del PHQ-9 (0,88); $Z_{1-\alpha/2}$ estableció en 1,96 y $Z_{1-\beta}$ en 1,28; δ se fijó en 0,08 ($\pi_1-\pi_2$). Su resultado fue 214. Se seleccionó a los participantes de manera consecutiva a medida que acudían a los centros de salud hasta encuestar el número máximo de sujetos por encima de 214.

Procedimientos

El estudio fue aprobado por el comité de ética del ISABU y de la Universidad de Santander teniendo en cuenta las normas internacionales³¹ y nacionales³², vigentes para investigación en humanos.

El PHQ-9 fue traducido siguiendo las recomendaciones para la adaptación de pruebas de autoinforme³³. Se realizó la traducción directa a partir de la escala original por 2 traductores bilingües certificados independientes; se discutieron los puntos discordantes de las 2 traducciones, luego se hizo

la retrotraducción al inglés, que fue revisada por el equipo investigador para evaluar la coincidencia con la escala original. Posteriormente, la escala traducida fue revisada por 10 psiquiatras expertos en investigación o experiencia clínica para verificar si los ítems eran coherentes con el constructo de depresión y ellos, además, realizaron observaciones sobre la comprensión y la redacción de los ítems. Diez personas de la población general con antecedentes de depresión también opinaron sobre la comprensión de las preguntas. El grupo investigador analizó e incorporó las observaciones de los pacientes y los expertos para la obtención de la nueva versión colombiana (fig. 1). Con la nueva versión de la escala, se realizó una prueba piloto con 21 sujetos con características similares a los sujetos de estudio pero en otros centros, los que respondieron las preguntas sin dificultades y no fueron necesarios ajustes en la estructura gramatical.

Se hizo un entrenamiento del equipo investigador en entrevista psiquiátrica estructurada (MINI) y heteroaplicación del PHQ-9. Las personas encargadas de aplicar las escalas y las entrevistas estructuradas eran profesionales con experiencia clínica (4 psicólogos, 2 residentes de medicina familiar y un psiquiatra), los cuales recibieron un entrenamiento de 8 h a cargo del primer autor, con sesiones teórico-prácticas, juego de roles y observación de entrevistas piloto con retroalimentación. Los participantes del estudio fueron contactados en la sala de espera a medida que llegaban a consulta externa de cualquier causa con médico general. Uno de los integrantes del grupo investigador explicó la naturaleza del estudio y les entregó el consentimiento informado. Las escalas de cribado fueron leídas por miembros entrenados del equipo investigador. Después de completar el PHQ-9, el mismo día, cada participante era evaluado en otro consultorio por otro miembro del equipo (psicólogo o psiquiatra) entrenado y sin conocimiento del resultado del PHQ-9, para aplicar el módulo de depresión de la MINI. Las encuestas fueron revisadas por 2 revisores independientes y almacenadas en un formato generado en Excel.

Instrumentos

PHQ-9

El PHQ-9 es una escala de cribado que mide la presencia y severidad de síntomas depresivos³⁴. El PHQ-9³⁵ está constituido por los 9 síntomas del criterio A de EDM del DSM-IV²⁰. Estos 9 ítems están dispuestos en forma de una escala de tipo adjetival que evalúa la presencia del síntoma en las 2 últimas semanas («nada en absoluto», «varios días», «más de la mitad de los días» y «casi todos los días»), que se puntúan de 0 a 3, llegando a una puntuación de entre 0 a 27³⁶.

Puede ser auto o heteroadministrado y se usa ya sea de manera algorítmica para realizar un diagnóstico probable de un EDM o como una medida continua de puntajes que van de 0 a 27, y puntos de corte (PC) de 5, 10, 15 y 20, representando los niveles de síntomas depresivos como leve, moderado, moderadamente severo y severo³⁴. Estos puntajes también se pueden usar de manera dicotómica a partir de un PC y clasifica a los sujetos con o sin SDIC³⁷. Las características psicométricas del PHQ-9, según Kroenke et al., presentan una sensibilidad del 88% y una especificidad del 88%, una consistencia interna adecuada (α de Cronbach de 0,86-0,89), un

Durante las últimas 2 semanas, ¿qué tan a menudo le han afectado alguno de los siguientes problemas? (Use «\» para indicar su respuesta)	Para nada	Varios días	Más de la mitad de los días	Casi todos los días
1. Poco interés o placer para hacer las cosas				
2. Sentirse triste, deprimido o sin esperanzas				
3. Dificultad para conciliar o mantener el sueño, o dormir demasiado				
4. Sentirse cansado o tener poca energía				
5. Mal apetito o comer en exceso				
6. Sentirse mal sobre sí mismo (o que es un fracaso o se ha decepcionado a usted o a su familia)				
7. Dificultad para concentrarse en actividades, como leer el periódico o ver televisión				
8. Moverse o hablar tan despacio que otras personas pueden haberlo notado. O lo contrario, sentirse tan nervioso e inquieto que ha estado moviéndose más de lo usual				
9. Pensamientos de que estaría mejor muerto o de lastimarse a usted mismo de alguna manera				

Figura 1 – Cuestionario sobre la salud del paciente (PHQ-9), versión para Atención Primaria, Bucaramanga, Colombia.

puntaje prueba-reprueba de 0,84, una concordancia entre el test autoadministrado y el realizado por el evaluador del 84% y un área bajo la curva (ABC) de 0,95³⁴. Para la identificación de casos de SDIC, en este estudio se utilizó el PC de 8 o más basados en el metaanálisis de Manea et al.²³ y el estudio de Rancans et al. en AP³⁸.

Minientrevista estructurada neuropsiquiátrica

La MINI es una entrevista diagnóstica estructurada breve, que explora las categorías diagnósticas del DSM-IV y la CIE-10³⁹. Su versión original fue desarrollada por Sheehan et al.³⁹ y Lecrubier et al.⁴⁰ en Estados Unidos y Francia. Contiene 130 preguntas organizadas en módulos que evalúan 16 trastornos del eje I del DSM-IV y un trastorno de personalidad. La versión original en inglés mostró un rango de sensibilidad entre 0,46 y 0,94 y de especificidad entre 0,72 y 0,97^{39,40}, una excelente confiabilidad interevaluador (κ 0,70) y prueba-reprueba, y una moderada validez de criterio comparada con la CIDI y la SCID-P^{39,40}. La MINI ganó rápida aceptación internacional⁴¹⁻⁴³, cuenta con versiones traducidas en 43 idiomas³⁹ y su confiabilidad y validez han sido exploradas en las versiones italiana⁴⁴, japonesa⁴⁵, noruega⁴⁶, marroquí⁴⁷ y portuguesa⁴⁸. El tiempo de administración promedio es 18,7 ± 11,6 min, con una media de 15 min³⁹. La MINI junto con CIDI y SCID-I son considerados patrones de oro mundialmente aceptados para el

diagnóstico de trastornos mentales en ámbitos clínicos y de investigación⁴⁹.

Escala hospitalaria de ansiedad y depresión

La Escala hospitalaria de ansiedad y depresión (HADS) fue diseñada por Zigmond y Snaith en 1983⁵⁰, para la detección de trastornos afectivos, sobre todo asociados con síntomas somáticos. Consta de 14 ítems, con una subescala de ansiedad (ítems impares) y una subescala de depresión (ítems pares). Cada ítem es valorado según una escala de 4 puntos de frecuencia que va desde 0 hasta 3. La HADS ha sido traducida a la mayoría de las lenguas europeas, al árabe, al israelí, al urdu, al japonés y al chino⁵¹, y ha demostrado fiabilidad y validez en numerosos estudios⁵². En Colombia, fue validada en pacientes oncológicos, mostrando una adecuada consistencia interna (α de Cronbach 0,85), un PC de 8 para la subescala de ansiedad y de 9 para la subescala de depresión⁵³. Estas propiedades psicométricas fueron confirmadas en una muestra poblacional (n = 1.500) en varias ciudades de Colombia⁵⁴. En esta investigación se utilizó la versión adaptada de Rico et al.⁵³.

Cuestionario de salud del paciente-2

El Cuestionario de salud del paciente-2 (PHQ-2) se compone de los 2 primeros ítems del PHQ-9, necesarios para sospechar la presencia de depresión según criterios DSM-IV⁵⁵. La forma de puntuación es igual a la del PHQ-9 y

el rango de puntaje es de 0 a 6. El PC de 3 es el óptimo para cribado, pero un metaanálisis reciente sugiere que un PC de 2 podría aumentar la sensibilidad⁵⁶. Los pacientes que puntúen positivo para SDIC deben ser evaluados con el PHQ-9 para determinar si reúnen criterios para EDM⁵⁷. Su utilidad clínica radica en que reduce el tiempo de en las consultas regulares de AP, que usualmente están congestionadas⁵⁸. El PHQ-2 ha mostrado un desempeño psicométrico comparable con el PHQ-9, con buena confiabilidad, validez y sensibilidad al cambio⁵⁶. En este trabajo, se consideró el PC de 2 o más para identificar a pacientes con SDIC⁵⁹.

Análisis estadístico

Los datos fueron analizados en SPSS versión 20.0⁶⁰, cuidadosamente verificados y revisados en 2 ocasiones. Se realizó un análisis descriptivo de las variables cualitativas y cuantitativas. Para evaluar la consistencia interna se calcularon los coeficientes α de Cronbach y ω de McDonald; para la validez concurrente se estimaron las correlaciones de Spearman o Pearson según la distribución de las variables. Para evaluar la exactitud del PHQ-9 como instrumento de cribado comparado con la MINI, se realizó el análisis de las características operativas del receptor (COR) y el ABC. Para la determinación del PC óptimo del PHQ-9 se tuvieron en cuenta los índices de validez: sensibilidad, especificidad, valores predictivos positivos (VPP) y negativos (VPN), razones de verosimilitud (LR, en inglés) positivas y negativas, el índice de Youden y el análisis de la curva COR/ABC.

Resultados

Características de los participantes

Fueron contactados 384 usuarios, de los cuales 95 se negaron a participar. De las encuestas realizadas, 46 fueron descartadas por datos faltantes e inconsistentes; por lo tanto, en el análisis se incluyó a 243 participantes, de los cuales 184 (75,7%) fueron de sexo femenino. El promedio de edad fue de 34,05 años, con una DE = 12,47 años. Para los hombres, el promedio de edad fue 33,59 años, con una DE = 12,89 años y para las mujeres, 34,20 años, con una DE = 12,37 años. Las características sociodemográficas de la muestra pueden verse en la tabla 1.

La prevalencia de SDIC según los resultados del PHQ-9 fue del 27,2% y del 21,8% según la entrevista estructurada MINI.

Consistencia interna

Se obtuvo un coeficiente α de Cronbach de 0,80 y un coeficiente ω de 0,81. En la tabla 2 se presenta la consistencia interna global de la escala si cada ítem es eliminado.

Validez convergente

Se usó la prueba de Kolmogorov-Smirnov para establecer la normalidad de las variables, con el propósito de decidir el tipo de prueba para el análisis de la validez concurrente del PHQ-9 comparado con el PHQ-2, y la subescala de depresión de

Tabla 1 – Descripción de características sociodemográficas de pacientes con o sin síntomas depresivos menores atendidos en centros de Atención Primaria en salud

Variables	N.º	%
Sexo		
Masculino	59	24,30
Femenino	184	75,72
Procedencia		
Urbana	210	86,42
Rural	33	13,58
Estado civil		
Soltero	98	40,33
Casado	53	21,88
Unión libre	83	34,16
Divorciado	5	2,06
Viudo	4	1,65
Escolaridad		
Primaria incompleta	14	5,75
Primaria completa	60	24,69
Secundaria incompleta	36	14,81
Secundaria completa	88	36,21
Técnica incompleta	18	7,41
Técnica completa	1	0,41
Tecnología completa	3	1,23
Universidad incompleta	3	1,23
Universidad completa	20	8,23
Estrato socioeconómico		
Estrato 1	103	42,39
Estrato 2	84	34,57
Estrato 3	47	19,34
Estrato 4	6	2,47
Estrato 5	3	1,23

Tabla 2 – Descripción de la consistencia interna del PHQ-9 versión colombiana: α de Cronbach y ω de McDonald por ítem

Preguntas PHQ-9	α de Cronbach	ω de McDonald
P1	0,774	0,780
P2	0,767	0,772
P3	0,795	0,800
P4	0,776	0,784
P5	0,792	0,798
P6	0,778	0,782
P7	0,788	0,794
P8	0,798	0,805
P9	0,799	0,802

HADS (HADS-D). Estas variables no mostraron una distribución normal, por lo que se utilizó la rho de Spearman. La rho de Spearman fue de 0,646 para HADS-D ($p < 0,010$) y 0,701 para PHQ-2 ($p < 0,010$).

Validez de criterio

La curva de COR (fig. 2) y los índices de exactitud del PHQ-9 arrojaron los resultados que se observan en la tabla 3. El ABC fue de 0,92 (IC del 95%: 0,88-0,963).

El PC óptimo coincide un puntaje del PHQ-9 ≥ 7 (sensibilidad de 90,38 [IC del 95%, 81,41-99,36]; especificidad de 81,68 [IC del 95%: 75,93-87,42]; VPP 57,32 [IC del 95%: 46,00-68,63];

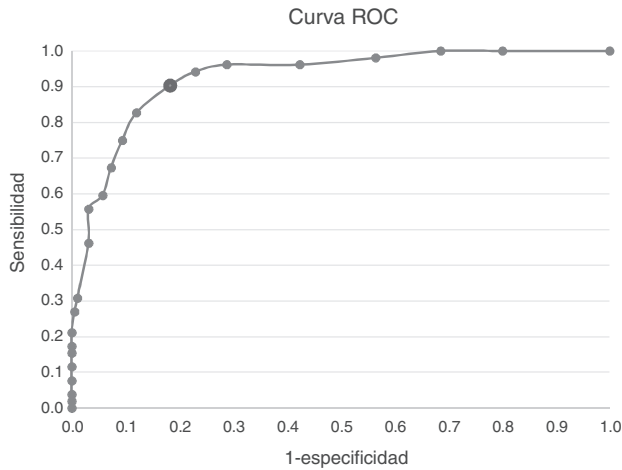


Figura 2 – Características operativas del receptor (COR) del PHQ-9 versión colombiana comparada con la MINI como patrón de referencia para depresión (N = 146).

VPN 96,89 [IC del 95%: 93,90-99,88]; índice de Youden 0,72 [IC del 95%: 0,62-0,82]; LR+ 4,93 [IC del 95%: 3,61-6,74]; LR- 0,12 [IC del 95%: 0,005-0,270]).

Discusión

Hasta donde sabemos, este es el primer estudio sobre validez de criterio del PHQ-9 en AP en Colombia. La prevalencia de EDM en este estudio fue del 21,8%. La versión colombiana del PHQ-9 mostró un excelente desempeño diagnóstico como instrumento de cribado de depresión, como lo demostraron el resultado de la curva COR y el ABC. El PHQ-9 también mostró un adecuado balance de sensibilidad y especificidad al $PC \geq 7$ al compararlo con la MINI como patrón de referencia, lo cual estableció una adecuada validez de criterio del PHQ-9. La comparación de los puntajes del PHQ-9 con los de la HADS-D y los coeficientes α y ω demostraron una buena validez convergente y una adecuada consistencia interna.

El porcentaje de sujetos clasificados con SDIC de acuerdo con el PHQ-9 al PC preestablecido fue del 27,2% (IC del 95%:

26,3-28,9), mayor a la prevalencia encontrada en Bucaramanga, del 22,3% (IC del 95%: 20,0-24,6) con la escala de Zung¹⁰, diferencia que se puede explicar por el bajo rendimiento diagnóstico de la escala de Zung en la población colombiana⁶¹. En cuanto a la prevalencia de EDM con la MINI, en esta muestra fue de 21,8 (IC del 95%: 20,8-23,5), que la ubica dentro del rango esperado de acuerdo con un metaanálisis de 41 estudios en AP con prevalencia global ajustada del 19,5% (IC del 95%: 15,7-23,7)⁶². Sin embargo, la prevalencia de EDM en este estudio es un poco más alta que lo informado en estudios previos en población general de Bucaramanga del 16,5% (IC del 95%: 12,3-21,6)¹¹, lo que puede explicarse por el hecho de que este estudio fue realizado en personas que consultan a centros de AP, en donde la prevalencia de depresión es mayor que en la población general⁶³ y por el predominio de mujeres⁶⁴.

El coeficiente α de Cronbach fue 0,80 y la ω de McDonald fue de 0,81, lo que indica una buena consistencia interna^{65,66}. Para que un instrumento de autoinforme sea confiable, el α de Cronbach y la ω de McDonald deben ser al menos de 0,70⁶⁷. La consistencia interna encontrada en este estudio concuerda con un estudio previo en Colombia³⁷ y con otros realizados en diferentes lenguas, cuyos coeficientes oscilaban entre 0,79 y 0,89^{68,70,71}.

Estudios previos han mostrado que el PHQ-9 tiene adecuada validez concurrente con varias medidas, como el inventario de depresión de Hamilton (HAMD), los formularios cortos de evaluación en salud e incluso el PHQ-2⁷². En nuestro estudio, los puntajes totales del PHQ-9 mostraron una correlación positiva estadísticamente significativa con los puntajes de la HADS-D y el PHQ-2 (ρ de Spearman 0,64 [$p < 0,01$] y 0,70 PHQ-2 [$p < 0,01$], respectivamente), consistente con estudios previos en los que los coeficientes de Pearson para el PHQ-9 con HAMD/BDI estuvieron entre 0,52 y 0,76^{68,73-75}, mientras que en un estudio de pacientes con enfermedad de Parkinson mostró que el PHQ-9 se correlacionó positivamente con la Self Depression Scale y la escala de depresión geriátrica de 15 ítems, con un coeficiente de Spearman de 0,63 para ambas escalas⁷⁶. Los coeficientes de correlación encontrados en este estudio confirman la validez convergente del PHQ-9, ya que coeficientes de correlación de Spearman entre 0,60 y 0,80 indican una correlación positiva buena o considerable⁷⁷.

Tabla 3 – Descripción de los diferentes puntos de corte del PHQ-9 versión colombiana y coeficientes de validez

Punto de corte	Sensibilidad	Especificidad	Índice de Youden	Correctamente clasificados %	VP+	VP-	LR+	LR-
≥ 3	0,98	0,43	0,42	55,14	0,25	0,99	1,73	0,04
≥ 4	0,96	0,58	0,54	65,84	0,30	0,99	2,27	0,07
≥ 5	0,96	0,71	0,67	76,54	0,39	0,99	3,34	0,05
≥ 6	0,94	0,77	0,71	80,66	0,44	0,99	4,09	0,07
$\geq 7^*$	0,90*	0,82*	0,72*	83,54*	0,48*	0,98*	4,93*	0,12*
≥ 8	0,83	0,88	0,71	86,83	0,57	0,96	6,87	0,20
≥ 9	0,75	0,91	0,66	87,24	0,60	0,95	7,96	0,28
≥ 10	0,67	0,93	0,60	87,24	0,64	0,94	9,18	0,35
≥ 11	0,60	0,94	0,54	86,83	0,66	0,92	10,35	0,43
≥ 12	0,56	0,97	0,53	88,07	0,77	0,92	17,75	0,46
≥ 13	0,46	0,97	0,43	86,01	0,74	0,90	14,69	0,56
≥ 14	0,31	0,99	0,30	84,36	0,85	0,88	29,38	0,70
≥ 15	0,27	0,99	0,26	83,95	0,91	0,88	51,42	0,73

* índices psicométricos asociados al punto de corte óptimo ≥ 7 .

Con respecto al PC, un PC de 10 en el PHQ-9 ha sido recomendado para la identificación de EDM en varios estudios³⁴. Por ejemplo, en un estudio de usuarios de AP en la China, un PC óptimo de 10 arrojó una sensibilidad de 0,87 y una especificidad de 0,8⁶⁹. Sin embargo, un reciente metaanálisis de 18 estudios demostró que el PC óptimo del PHQ-9 puede variar entre 8 y 11, dependiendo de la población estudiada; no obstante, el balance de sensibilidad y especificidad se mantiene aún para el PC de 7 (5 de los 18 estudios incluidos)²³. En nuestro estudio, el PC de 7 parece haber tenido el óptimo balance de sensibilidad y especificidad, lo cual se confirmó con una medida de exactitud adicional: el índice de Youden⁷⁸, definido como la máxima distancia vertical entre la curva COR y la línea de 45 grados, como un indicador de que tan lejos la curva está de un test no informativo⁷⁹. El índice de Youden es una función de la sensibilidad y la especificidad; se calcula como $(Se + Es - 1)$ ⁸⁰ y debe ser considerado en adición a la curva COR, ya que usualmente están correlacionados⁸¹. El rango es de 0 a 100 cuando se convierte en porcentaje. Los valores > 50% son generalmente considerados aceptables para exactitud diagnóstica⁸².

Los valores asociados al PC de nuestro estudio son consistentes con un estudio en adultos mayores de AP, en los cuales la validez de criterio del PHQ-9 se evaluó mediante la administración de la MINI, en donde un PC óptimo ≥ 7 (sensibilidad 0,92; especificidad 0,78) mostró las mejores características psicométricas⁸³. No obstante, este PC de 7 es más bajo que lo encontrado en la mayoría de estudios con el PHQ-9 en otras poblaciones. La características culturales y demográficas de la muestra pueden ser la razón de esta diferencia⁸⁴. El estigma es un aspecto importante que también puede influir en el patrón de respuesta de las personas a las escalas de cribado de depresión en nuestra población, causando vergüenza en personas con enfermedades mentales que limita la identificación de sus fenómenos psicopatológicos^{85,86}. Llama la atención que los PC para el PHQ-9 tienden a ser más bajos en países de medianos y bajos ingresos⁸⁷⁻⁸⁹ en comparación con los de altos ingresos^{75,90,91}. Sin embargo, no existen estudios que estudien este fenómeno. Esta diferencia en el PC óptimo resalta la importancia de validar instrumentos de cribado en diferentes contextos sociales y culturales⁹².

Para el PC ≥ 7 , la sensibilidad y la especificidad del PHQ-9 en esta muestra fueron del 90 y el 83%, respectivamente. Estos hallazgos son consistentes con el estudio de Wang et al., donde el PC ≥ 7 permitió un abalance adecuado de la sensibilidad y la especificidad (sensibilidad del 85%; especificidad del 86%)⁹³. Por lo tanto, los índices de exactitud en nuestro estudio se consideran apropiados, ya que un instrumento de cribado es considerado bueno cuando su sensibilidad se ubica entre el 79 y el 97%; y la especificidad entre el 63 y el 86%⁹⁴. Wittkamp et al. revisaron sistemáticamente las propiedades psicométricas del PHQ-9 y encontraron una sensibilidad del 77% (71-84%) y una especificidad del 94% (90-97%), incluyendo estudios en subgrupos de alta prevalencia de depresión, como usuarios de AP²⁵.

La LR+ y la LR- del PHQ-9 en nuestra muestra, para el PC ≥ 7 , fueron de 4,93 y 0,12, respectivamente. Esto quiere decir que en un contexto clínico similar, un resultado positivo en el PHQ-9 (PC ≥ 7) es 5 veces más frecuente en un paciente con depresión que en uno sin depresión y un que

un sujeto con una prueba negativa tendría una posibilidad menor del 2% de tener depresión⁹⁵. Estos resultados son comparables con los obtenidos en la versión china del PHQ-9, que con el PC ≥ 7 tuvo unas LR+ y LR- de 5,99 y 0,17, respectivamente⁹³.

El ABC de esta versión colombiana para AP del PHQ-9 fue de 0,92, lo cual demuestra una alta exactitud⁹⁶ y es consistente con estudios previos en AP y otras poblaciones^{69,71,93}.

Las principales fortalezas de este estudio incluyen el uso de un criterio de referencia clínico para evaluar la validez de criterio del PHQ-9, la adecuada proporción de respuesta de los participantes (75,3%), el entrenamiento adecuado de los entrevistadores, la adherencia a los lineamientos de QUADAS-2²⁹ y la ejecución de un plan analítico riguroso. Además, la traducción del PHQ-9 se realizó de acuerdo con los lineamientos estandarizados para la adaptación transcultural de escalas. La participación de un grupo de expertos apoyó la adaptación lingüística y garantizó una apropiada validez de contenido.

Existen varias limitaciones en este estudio. Primero, nuestro estudio fue realizado en el contexto de AP, por lo tanto, los resultados no pueden generalizarse a la población general, cuyas características determinan un patrón de respuesta diferente⁸⁴. Segundo, el estudio fue limitado a adultos. Existe creciente evidencia que los adolescentes están particularmente afectados por los trastornos depresivos⁹⁷, por lo que futuros estudios en Colombia deberían evaluar el desempeño psicométrico del PHQ-9 en esa población. En tercer lugar, este fue un estudio transversal, por consiguiente, debería considerarse, en el futuro, diseñar estudios longitudinales para establecer la sensibilidad al cambio del PHQ-9 en población colombiana, ya que existen trabajos que lo han utilizado para evaluar la respuesta al tratamiento de la depresión⁹⁸. Cuarto, el hecho de que en la muestra hubiera un predominio de mujeres (75%) podría afectar a la estimación de los índices de exactitud, ya que la prevalencia de depresión en mujeres es mayor que en hombres, obteniéndose un número mayor de casos positivos de depresión⁹⁹. En quinto lugar, una debilidad relativa es el tamaño muestral, que se calculó siguiendo las recomendaciones de Sanchez et al. para comparar la sensibilidad de la prueba tamiz con el patrón de referencia³⁰. Sin embargo, otros autores, como Buderer¹⁰⁰ y Obuchowski¹⁰¹, exigen tamaños de muestra mayores. No obstante, si se siguen los criterios de Bean para comparar la sensibilidad o la especificidad de 2 pruebas diagnósticas se obtienen tamaños de muestra similares al nuestro¹⁰².

De acuerdo con los resultados globales, la versión colombiana del PHQ-9 para AP tiene un desempeño psicométrico excelente como prueba de cribado, lo que garantiza que pueda ser usada en contextos de escasos recursos y con debilidades en el sistema de salud, donde la disponibilidad de psiquiatras es limitada¹⁰³. Dentro de las estrategias para limitar la carga de trastornos mentales en países de bajos y medianos ingresos está la integración de la salud mental en la AP¹⁰⁴. Una de las principales barreras para alcanzar esta meta es la falta de instrumentos de cribado válidos y fáciles de administrar para detectar depresión. La validación de instrumentos como el PHQ-9 en estos contextos puede ayudar a solucionar este problema¹⁰⁵. Se sabe que solo el cribado de depresión es insuficiente para suplir las

crecientes necesidades de atención de trastornos mentales en países de bajos y medianos ingresos; sin embargo, dado que la depresión contribuye significativamente a la carga de enfermedad, el tener instrumentos de cribado válidos es la primera etapa hacia la resolución de este problema¹⁰⁶. En algunos países de bajos ingresos existen programas costo-efectivos de intervención en depresión, en los cuales los instrumentos de cribado pueden ser usados para identificar los participantes apropiados¹⁰⁷. Uno de los principales componentes de las intervenciones efectivas en salud mental en AP es la monitorización de síntomas depresivos usando cuestionarios simples, breves y fáciles de administrar, como el PHQ-9¹⁰⁸.

Con la validación de esta versión del PHQ-9, los investigadores en Colombia disponen ahora de información psicométrica válida y confiable sobre el cribado de depresión en AP, que permitirá que el PHQ-9 sea utilizado en estudios en donde se requiera identificar sintomatología depresiva con un PC apropiado.

En conclusión, los resultados de este estudio indican que la versión colombiana del PHQ-9 es un instrumento válido y confiable para el cribado de depresión en el contexto de AP en Bucaramanga, con un PC de 7 o más. Se necesita evaluar las propiedades psicométricas de esta versión del PHQ-9 en poblaciones diferentes y otras regiones del país. Futuros estudios en Colombia deberían evaluar la sensibilidad al cambio del PHQ-9.

Financiación

Este trabajo fue financiado por la facultad de medicina de la universidad de Santander (UDES) y el Instituto de Salud de Bucaramanga (ISABU). Código del proyecto: PIFE0118020041816EJ.

Autoría

Carlos Arturo Cassiani-Miranda: diseño, ajuste de la escala, entrenamiento de encuestadores y entrevistadores, recolección de información, análisis estadístico, digitación, redacción y revisión del artículo.

Angy Karina Cuadros-Cruz: diseño, ajuste de la escala, recolección de información y redacción del artículo.

Harold Torres Pinzón: diseño, ajuste de la escala, análisis estadístico, digitación, redacción y revisión del artículo.

Orlando Scoppetta: ajuste de la escala, análisis estadístico, redacción y revisión del artículo.

Jhon Henry Pinzón-Tarrazona: recolección de información, digitación y redacción del artículo.

Wendy Yulieth López-Fuentes: recolección de información, digitación y redacción del artículo.

Andrea Paez: recolección de información y redacción del artículo.

Diego Fernando Cabanzo-Arenas: digitación, redacción y revisión del artículo.

Sergio Ribero-Marulanda: recolección de información, redacción y revisión del artículo.

Elkin René Llanes-Amaya: recolección de información y redacción del artículo.

Conflicto de intereses

Ninguno.

Agradecimientos

Al panel de expertos por sus aportes en validez de apariencia y contenido: Astrid I. Arrieta, Jaider A. Barros, Adalberto Campo-Arias, Mauricio Castaño, Jenny García, Luis A. Montenegro, Jorge A. Niño, Heidi C. Oviedo, Andrés M. Rangel, Jorge J. Téllez-Vargas.

BIBLIOGRAFÍA

1. Steel N, Ford JA, Newton JN, Davis ACJ, Vos T, Naghavi M, et al. Changes in health in the countries of the UK and 150 English Local Authority areas 1990-2016: A systematic analysis for the Global Burden of Disease Study 2016. *Lancet*. 2018;392(10158):1647-61.
2. Yang D, Hur JW, Kwak YB, Choi SW. A systematic review and meta-analysis of applicability of web-based interventions for individuals with depression and quality of life impairment. *Psychiatry Investig*. 2018;15:759-66.
3. Van Eck van der Sluijs JF, Castelijns H, Eijsbroek V, Rijnders CAT, van Marwijk HWJ, van der Feltz-Cornelis CM. Illness burden and physical outcomes associated with collaborative care in patients with comorbid depressive disorder in chronic medical conditions: A systematic review and meta-analysis. *Gen Hosp Psychiatry*. 2018;50:1-14.
4. Laursen TM, Musliner KL, Benros ME, Vestergaard M, Munk-Olsen T. Mortality and life expectancy in persons with severe unipolar depression. *J Affect Disord*. 2016;193:203-7.
5. Brandão D, Fontenelle LF, da Silva SA, Menezes PRP-VM. Depression and excess mortality in the elderly living in low- and middle-income countries: Systematic review and meta-analysis. *Int J Geriatr Psychiatry*. 2019;34:22-30.
6. Kessler RC, Sampson NA, Berglund P, Gruber MJ, Al-Hamzawi A, Andrade L, et al. Anxious and non-anxious major depressive disorder in the World Health Organization World Mental Health Surveys. *Epidemiol Psychiatr Sci*. 2015;24:210-26.
7. Muhammad Gadit AA, Mugford G. Prevalence of depression among households in three capital cities of Pakistan: Need to revise the mental health policy. *PLoS One*. 2007;2:1-5.
8. Wang J, Wu X, Lai W, Long E, Zhang X, Li W, et al. Prevalence of depression and depressive symptoms among outpatients: A systematic review and meta-analysis. *BMJ Open*. 2017;7:1-14.
9. Gómez-Restrepo C, Bohórquez A, Tamayo Martínez N, Rondón M, Bautista N, Rengifo H, et al. Trastornos depresivos y de ansiedad y factores asociados en la población de adolescentes colombianos, Encuesta Nacional de Salud Mental 2015. *Rev Colomb Psiquiatr*. 2016;45 Supl. 1:50-7.
10. Rueda-Sánchez M, Díaz-Martínez LA, Rueda-Jaimes GE. Prevalencia del trastorno depresivo mayor y factores asociados: un estudio poblacional en Bucaramanga (Colombia). *Rev Colomb Psiquiatr*. 2008;37:159-68.
11. Cadena L, del P, Díaz L, Rueda G, Hernández N, Campo A. Prevalencia actual del trastorno depresivo mayor en la población de Bucaramanga, Colombia. *Rev Fac Nac Salud Pública*. 2010;28:36-41.
12. Mitchell AJ, Rao S, Vaze A. Can general practitioners identify people with distress and mild depression? A meta-analysis of clinical accuracy. *J Affect Disord*. 2011;130(1-2):26-36.

13. Carey M, Yoong SL, Grady A, Bryant J, Jayakody A, Sanson-Fisher R, et al. Unassisted detection of depression by GPs: Who is most likely to be misclassified? *Fam Pract.* 2015;32:282-7.
14. Mitchell AJ, Rao S, Vaze A. International comparison of clinicians' ability to identify depression in primary care: Meta-analysis and meta-regression of predictors. *Br J Gen Pract.* 2011;61:72-80.
15. Ali GC, Ryan G, de Silva MJ. Validated screening tools for common mental disorders in low and middle income countries: A systematic review. *PLoS One.* 2016;11:1-14.
16. Paykel ES, Tylee A, Wright A, Priest RG, Rix S, Hart D. The defeat depression campaign: Psychiatry in the public arena. *Am J Psychiatry.* 1997;154 6 Suppl:59-66.
17. Mitchell C, Dwyer R, Hagan T, Mathers N. Impact of the QOF and the NICE guideline in the diagnosis and management of depression: A qualitative study. *Br J Gen Pract.* 2011;61:279-89.
18. Gilbody S, Sheldon T, House A. Screening and case-finding instruments for depression: A meta-analysis. *CMAJ.* 2008;178:997-1003.
19. Pettersson A, Boström KB, Gustavsson P, Ekselius L. Which instruments to support diagnosis of depression have sufficient accuracy? A systematic review. *Nord J Psychiatry.* 2015;69:497-508.
20. Kocalevent RD, Hinz A, Brähler E. Standardization of the depression screener Patient Health Questionnaire (PHQ-9) in the general population. *Gen Hosp Psychiatry.* 2013;35:551-5.
21. Sung SC, Low CCH, Fung DSS, Chan YH. Screening for major and minor depression in a multiethnic sample of Asian primary care patients: A comparison of the nine-item Patient Health Questionnaire (PHQ-9) and the 16-item Quick Inventory of Depressive Symptomatology - Self-Report (QIDS-SR16). *Asia-Pacific Psychiatry.* 2013;5:249-58.
22. Williams JW, Pignone M, Ramirez G, Perez Stellato C. Identifying depression in primary care: A literature synthesis of case-finding instruments. *Gen Hosp Psychiatry.* 2002;24:225-37.
23. Manea L, Gilbody S, McMillan D. Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): A meta-analysis. *CMAJ.* 2012;184:191-6.
24. Kroenke K. Enhancing the clinical utility of depression screening. *CMAJ.* 2012;184:281-2.
25. Wittkampf KA, Naeije L, Schene AH, Huyser J, van Weert HC. Diagnostic accuracy of the mood module of the Patient Health Questionnaire: A systematic review. *Gen Hosp Psychiatry.* 2007;29:388-95.
26. Deneke DE, Schultz H, Fluent TE. Screening for depression in the primary care population. *Prim Care.* 2014;41:399-420.
27. Cassiani-Miranda CA, Scopetta O. Factorial structure of the Patient Health Questionnaire-9 as a depression screening instrument for university students in Cartagena, Colombia. *Psychiatry Res.* 2018;269:425-9.
28. Smithson S, Pignone MP. Screening adults for depression in primary care. *Med Clin North Am.* 2017;101:807-21.
29. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Deeks JJ, et al. QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med.* 2011;155:529-36.
30. Sánchez-Pedraza R, Echeverry-Raad J. Aspectos sobre diseño y tamaño de muestra en estudios de pruebas diagnósticas. *Rev Fac Med.* 2001;49:175-80.
31. Asociación Médica Mundial. Declaración de Helsinki de la AMM. Principios éticos para las investigaciones médicas en seres humanos [Internet]. 2015 [consultado 20 Feb 2019]. Disponible en: <http://www.wma.net/es/30publications/10policies/b3/>.
32. Ministerio de Salud. Resolución 8430 de 1993. República de Colombia: Ministerio de Salud; 1993 p. 1-12.
33. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine (Phila Pa 1976).* 2000;25:23186-91.
34. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: Validity of a brief depression severity measure. *J Gen Intern Med.* 2001;16:606-13.
35. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD. *JAMA.* 1999;282:1737-44.
36. Spitzer RL, Williams JBW, Kroenke K, Hornyak R, McMurray J. Validity and utility of the PRIME-MD Patient Health Questionnaire in assessment of 3000 obstetric-gynecologic patients: The PRIME-MD Patient Health Questionnaire Obstetrics-Gynecology Study. *Am J Obstet Gynecol.* 2000;183:759-69.
37. Cassiani-Miranda C, Vargas-Hernández M, Pérez-Anibal E, Herazo-Bustos M, Hernández-Carrillo M. Confiabilidad y dimensionalidad del PHQ-9 para el cribado de sintomatología depresiva en estudiantes de ciencias de la salud de Cartagena, 2014. *Biomédica.* 2017;37 Supl. 1: 112-20.
38. Rancans E, Trapencieris M, Ivanovs R, Vrublevska J. Validity of the PHQ-9 and PHQ-2 to screen for depression in nationwide primary care population in Latvia. *Ann Gen Psychiatry.* 2018;17:1-9.
39. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry.* 1998;59 Suppl 20:22-33.
40. Lecrubier Y, Sheehan DV, Weiller E, Amorim P, Bonora I, Sheehan KH, et al. The Mini International Neuropsychiatric Interview (MINI). A short diagnostic structured interview: Reliability and validity according to the CIDI. *Eur Psychiatry.* 1997;12:224-31.
41. Balázs J, Lecrubier Y, Csiszér N, Koszták J, Bitter I. Prevalence and comorbidity of affective disorders in persons making suicide attempts in Hungary: Importance of the first depressive episodes and of bipolar II diagnoses. *J Affect Disord.* 2003;76(1-3):113-9.
42. Juang KD, Wang SJ, Fuh JL, Lu SR, Su TP. Comorbidity of depressive and anxiety disorders in chronic daily headache and its subtypes. *Headache.* 2000;40:818-23.
43. Pinninti NR, Madison H, Musser E, Rissmiller D. MINI International Neuropsychiatric Schedule: Clinical utility and patient acceptance. *Eur Psychiatry.* 2003;18:361364.
44. Rossi A, Alberio R, Porta A, Sandri M. TMA. The reliability of the Mini-International Neuropsychiatric Interview —Italian version. *J Clin Psychopharmacol.* 2004;24:561-3.
45. Otsubo T, Tanaka K, Koda R, Shinoda J, Sano N, Tanaka S, Aoyama H, et al. Reliability and validity of Japanese version of the Mini-International Neuropsychiatric Interview. *Psychiatry Clin Neurosci.* 2005;59:517-26.
46. Mordal J, Gundersen, Bramness JG. Norwegian version of the Mini-International Neuropsychiatric Interview: Feasibility, acceptability and test-retest reliability in an acute psychiatric ward. *Eur Psychiatry.* 2010;25:172-7.
47. Kadri N, Agoub M, El Gnaoui S, Mchichi Alami K, Hergueta T, Moussaoui D. Moroccan colloquial Arabic version of the Mini International Neuropsychiatric Interview (MINI): Qualitative and quantitative validation. *Eur Psychiatry.* 2005;20:193-5.
48. De Azevedo Marques JM, Zuardi AW. Validity and applicability of the Mini International Neuropsychiatric Interview administered by family medicine residents in primary health care in Brazil. *Gen Hosp Psychiatry.* 2008;30:303-10.

49. Tejada P, Jaramillo LE, Sánchez-Pedraza R. Revisión crítica sobre los instrumentos para la evaluación psiquiátrica en atención primaria. *Rev Fac Med.* 2014;62:101-10.
50. Zigmond AS, Snaith RP. The Hospital Anxiety and Depression Scale. *Acta Psychiatr Scand.* 1983;67:361-70.
51. Snaith RP. Availability of the hospital anxiety and depression (HAD) scale. *Br J Psychiatry.* 1992;161:422.
52. Moorey S, Greer S, Watson M, Gorman C, Rowden L, Tunmore R, et al. The factor structure and factor stability of the hospital anxiety and depression scale in patients with cancer. *Br J Psychiatry.* 1991;158:255-9.
53. Rico JL, Restrepo M, Molina M. Adaptación y validación de la escala hospitalaria de ansiedad y depresión (HAD) en una muestra de pacientes con cáncer del instituto nacional de cancerología de Colombia. *Avances en Medicación.* 2005;3:73-86.
54. Hinz A, Finck C, Gómez Y, Daig I, Glaesmer H, Singer S. Anxiety and depression in the general population in Colombia: Reference values of the Hospital Anxiety and Depression Scale (HADS). *Soc Psychiatry Psychiatr Epidemiol.* 2014;49:41-9.
55. Kroenke K, Spitzer RL, Williams JBW. The Patient Health Questionnaire-2: Validity of a two-item depression screener. *Med Care.* 2003;41:1284-92.
56. Löwe B, Kroenke K, Gräfe K. Detecting and monitoring depression with a two-item questionnaire (PHQ-2). *J Psychosom Res.* 2005;58:163-71.
57. Manea L, Gilbody S, Hewitt C, North A, Plummer F, Richardson R, et al. Identifying depression with the PHQ-2: A diagnostic meta-analysis. *J Affect Disord.* 2016;203:382-95.
58. Inagaki M, Ohtsuki T, Yonemoto N, Kawashima Y, Saitoh A, Oikawa Y, et al. Validity of the Patient Health Questionnaire (PHQ)-9 and PHQ-2 in general internal medicine primary care at a Japanese rural hospital: A cross-sectional study. *Gen Hosp Psychiatry.* 2013;35:592-7.
59. Mitchell AJ, Yadegarfar M, Gill J, Stubbs B. Case finding and screening clinical utility of the Patient Health Questionnaire (PHQ-9 and PHQ-2) for depression in primary care: A diagnostic meta-analysis of 40 studies. *Br J Psychiatry Open.* 2016;2:127-38.
60. SPSS Inc. PASW Statistics for Windows. Chicago: SPSS Inc.; 2009.
61. Lezama-Meneses S. Propiedades psicométricas de la escala de Zung para síntomas depresivos en población adolescente escolarizada colombiana. *Psychol Av Discip.* 2012;6:91-101.
62. Mitchell AJ, Vaze A, Rao S. Clinical diagnosis of depression in primary care: A meta-analysis. *Lancet.* 2009;374:609-19.
63. Löwe B, Spitzer RL, Williams JB, Mussell M, Schellberg D, Kroenke K. Depression, anxiety and somatization in primary care: syndrome overlap and functional impairment. *Gen Hosp Psychiatry.* 2008;30:191-9.
64. Stromberg R, Wernering E, Aberg-Wistedt A, Furhoff AK, Johansson SE, Backlund LG. Screening and diagnosing depression in women visiting GPs' drop in clinic in Primary Health Care. *BMC Fam Pract.* 2008;9:1-11.
65. Campo-Arias A, Oviedo HC. Propiedades psicométricas de una escala: la consistencia interna. *Rev Salud Pública.* 2008;10:831-9.
66. Lucke JF. The α and the ω of congeneric test theory: An extension of reliability and internal consistency to heterogeneous tests. *Appl Psychol Meas.* 2005;29:65-81.
67. Ventura-León JL. ¿Es el final del alfa de Cronbach? *Adicciones.* 2019;31:80-1.
68. Baader M, Molina J, Venezian S, Rojas C, Farías R, Fierro-Freixeneta C, et al. Validación y utilidad de la encuesta PHQ-9 (Patient Health Questionnaire) en el diagnóstico de depresión en pacientes usuarios de atención primaria en Chile. *Rev Chil Neuropsiquiatr.* 2012;50:10-22.
69. Chen S, Fang Y, Chiu H, Fan H, Jin T, Conwell Y. Validation of the nine-item Patient Health Questionnaire to screen for major depression in a Chinese primary care population. *Asia-Pacific Psychiatry.* 2013;5:61-8.
70. Kohrt BA, Luitel NP, Acharya P, Jordans MJ. Detection of depression in low resource settings: Validation of the Patient Health Questionnaire (PHQ-9) and cultural concepts of distress in Nepal. *BMC Psychiatry.* 2016;16:1-14.
71. Gelaye B, Williams MA, Lemma S, Deyessa N, Bahretibeb Y, Shibre T, et al. Validity of the patient health questionnaire-9 for depression screening and diagnosis in East Africa. *Psychiatry Res.* 2013;210:653-61.
72. Arrieta J, Aguerrebere M, Raviola G, Flores H, Elliott P, Espinosa A, et al. Validity and utility of the Patient Health Questionnaire (PHQ)-2 and PHQ-9 for screening and diagnosis of depression in rural Chiapas, Mexico: A cross-sectional study. *J Clin Psychol.* 2017;73:1076-90.
73. Lotrakul M, Sumrithe S, Saipanish R. Reliability and validity of the Thai version of the PHQ-9. *BMC Psychiatry.* 2008;8:1-7.
74. Liu SI, Yeh ZT, Huang HC, Sun FJ, Tjung JJ, Hwang LC, et al. Validation of Patient Health Questionnaire for depression screening among primary care patients in Taiwan. *Compr Psychiatry.* 2011;52:96-101.
75. Wittkampf K, van Ravesteijn H, Baas K, van de Hoogen H, Schene A, Bindels P, et al. The accuracy of Patient Health Questionnaire-9 in detecting depression and measuring depression severity in high-risk groups in primary care. *Gen Hosp Psychiatry.* 2009;31:451-9.
76. Chagas MH, Tumas V, Rodrigues GR, Machado-de-Sousa JP, Filho AS, Hallak JE, et al. Validation and internal consistency of patient health questionnaire-9 for major depression in parkinson's disease. *Age Ageing.* 2013;42:645-9.
77. Astivia OLO, Zumbo BD1. Population models and simulation methods: The case of the Spearman rank correlation. *Br J Math Stat Psychol.* 2017;70:347-67.
78. Youden WJ. Index for rating diagnostic tests. *Cancer.* 1950;3:32-5.
79. Xu T, Wang J, Fang Y. A model-free estimation for the covariate-adjusted Youden index and its associated cut-point. *Stat Med.* 2014;33:4963-74.
80. Inácio de Carvalho V, de Carvalho M, Branscum AJ. Nonparametric Bayesian covariate-adjusted estimation of the Youden index. *Biometrics.* 2017;73:1279-88.
81. Yin J, Tian L. Joint confidence region estimation for area under ROC curve and Youden index. *Stat Med.* 2014;33:985-1000.
82. Li C, Chen J, Qin G. Partial Youden index and its inferences. *J Biopharm Stat.* 2019;29:385-99.
83. Lamers F, Jonkers CC, Bosma H, Penninx BW, Knottnerus JA, van Eijk JT. Summed score of the Patient Health Questionnaire-9 was a reliable and valid method for depression screening in chronically ill elderly patients. *J Clin Epidemiol.* 2008;61:679-87.
84. Kleinman A. Culture and depression. *N Engl J Med.* 2004;351:951-3.
85. Campo-Arias A, Oviedo HC, Herazo E. Estigma: barrera de acceso a servicios en salud mental. *Rev Colomb Psiquiatr.* 2014;43:162-7.
86. Phillips MR, Pearson V, Li F, Xu M, Yang L. Stigma and expressed emotion: A study of people with schizophrenia and their family members in China. *Br J Psychiatry.* 2002;181:488-93.
87. Nakku JEM, Rathod SD, Kizza D, Breuer E, Mutyaba K, Baron EC, et al. Validity and diagnostic accuracy of the Luganda version of the 9-item and 2-item Patient Health Questionnaire for detecting major depressive disorder in rural Uganda. *Glob Ment Heal.* 2016;3:e20.

88. Bhana A, Rathod SD, Selohilwe O, Kathree T, Petersen I. The validity of the Patient Health Questionnaire for screening depression in chronic care patients in primary health care in South Africa. *BMC Psychiatry*. 2015;15:1-9.
89. Hanlon C, Medhin G, Selamu M, Breuer E, Worku B, Hailemariam M, et al. Validity of brief screening questionnaires to detect depression in primary care in Ethiopia. *J Affect Disord*. 2015;186:32-9 (2015).
90. Muramatsu K, Miyaoka H, Kamijima K, Muramatsu Y, Tanaka Y, Hosaka M, et al. Performance of the Japanese version of the Patient Health Questionnaire-9 (J-PHQ-9) for depression in primary care. *Gen Hosp Psychiatry*. 2018;52:64-9.
91. Dros J, Wewerinke A, Bindels PJ, van Weert HC, Arroll B, Goodyear-Smith F, et al. Validation of PHQ-2 and PHQ-9 to screen for major depression in the primary care population. *Ann Fam Med*. 2010;7:348-53.
92. Kerr LK, Kerr LD Jr. Screening tools for depression in primary care: The effects of culture, gender, and somatic symptoms on the detection of depression. *West J Med*. 2001;175:349-52.
93. Wang W, Bian Q, Zhao Y, Li X, Wang W, Du J, et al. Reliability and validity of the Chinese version of the Patient Health Questionnaire (PHQ-9) in the general population. *Gen Hosp Psychiatry*. 2014;36:539-44.
94. El-Den S, Chen TF, Gan YL, Wong E, O'Reilly CL. The psychometric properties of depression screening tools in primary healthcare settings: A systematic review. *J Affect Disord*. 2018;225:503-22.
95. Grimes DA, Schulz KF. Refining clinical diagnosis with likelihood ratios. *Lancet*. 2005;365:1500-5.
96. Huang X, Qin G, Fang Y. Optimal combinations of diagnostic tests based on AUC. *Biometrics*. 2011;67:568-76.
97. Richardson LP, McCauley E, Grossman DC, McCarty CA, Richards J, Russo JE, et al. Evaluation of the Patient Health Questionnaire-9 item for detecting major depression among adolescents. *Pediatrics*. 2010;126:1117-23.
98. Salaminios G, Duffy L, Ades A, Araya R, Button KS, Churchill R, et al. A randomised controlled trial assessing the severity and duration of depressive symptoms associated with a clinically significant response to sertraline versus placebo, in people presenting to primary care with depression (PANDA trial): Study protocol for. *Trials*. 2017;18:496.
99. Oneib B, Sabir M, Abda N, Ouanass A. Epidemiological study of the prevalence of depressive disorders in primary health care in Morocco. *J Neurosci Rural Pract*. 2015;6:477.
100. Buderer NM. Statistical methodology: I. Incorporating the prevalence of disease into the sample size calculation for sensitivity and specificity. *Acad Emerg Med*. 1996;3:895-900.
101. Obuchowski NA. Sample size calculation in studies of test accuracy. *Stat methods Med Res*. 1998;7:371-92.
102. Beam A. Strategies for improving power in diagnostic radiology research. *Am J Roentgenol*. 1992;159:631-7.
103. Hanlon C, Semrau M, Alem A, Abayneh S, Abdulmalik J, Docrat S, et al. Evaluating capacity-building for mental health system strengthening in low- and middle-income countries for service users and caregivers, service planners and researchers. *Epidemiol Psychiatr Sci*. 2018;27:3-10.
104. Thornicroft G, Ahuja S, Barber S, Chisholm D, Collins PY, Docrat S, et al. Integrated care for people with long-term mental and physical health conditions in low-income and middle-income countries. *The Lancet Psychiatry*. 2018;6:174-86.
105. Manea L, Gilbody S, McMillan D. A diagnostic meta-analysis of the Patient Health Questionnaire-9 (PHQ-9) algorithm scoring method as a screen for depression. *Gen Hosp Psychiatry*. 2015;37:67-75.
106. Sharan P, Sagar R, Kumar S. Mental health policies in South-East Asia and the public health role of screening instruments for depression. *WHO South-East Asia J Public Heal [Internet]*. 2017;6:5.
107. Hanlon C, Fekadu A, Jordans M, Kigozi F, Petersen I, Shidhaye R, et al. District mental healthcare plans for five low and middle-income countries: Commonalities, variations and evidence gaps. *Br J Psychiatry*. 2016;208 Suppl 56:47-54.
108. Kroenke K, Spitzer RL, Williams JB, Löwe B. The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales: A systematic review. *Gen Hosp Psychiatry*. 2010;32:345-59.