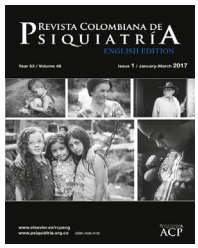
Asociación Colombiana de Psiquiatría ACP

Revista Colombiana de Psiquiatría

www.elsevier.es/rcp



Original Article

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

Carlos Arturo Cassiani-Miranda[a,*], *Angy Karina Cuadros-Cruz*[b], *Harold Torres-Pinzón*[c], *Orlando Scoppetta*[d], *Jhon Henrry Pinzón-Tarrazona*[e], *Wendy Yulieth López-Fuentes*[e], *Andrea Paez*[b], *Diego Fernando Cabanzo-Arenas*[f], *Sergio Ribero-Marulanda*[e], *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*

ARTICLE INFO

Article history:
Received 27 February 2019
Accepted 20 September 2019
Available online 20 February 2021

Keywords:
PHQ-9
Reproducibility of results
Screening
Depression
Primary healthcare
Colombia

ABSTRACT

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

DOI of original article: https://doi.org/10.1016/j.rcp.2019.09.001.

* Please cite this article as: Cassiani-Miranda CA, Cuadros-Cruz AK, Torres-Pinzón H, Scoppetta O, Pinzón-Tarrazona JH, López-Fuentes WY, et al. Validez del Cuestionario de salud del paciente-9 (PHQ-9) para cribado de depresión en adultos usuarios de Atención Primaria en Bucaramanga, Colombia. Rev Colomb Psiquiat. 2021;50:11–21.

* *Corresponding author.*
E-mail address: ca.cassiani@mail.udes.edu.co (C.A. Cassiani-Miranda).

https://doi.org/10.1016/j.rcpeng.2019.09.002

2530-3120/© 2019 Asociación Colombiana de Psiquiatría. Published by Elsevier España, S.L.U. All rights reserved.

(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.

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

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.

Introduction

Depression is a major public health problem worldwide[1] and has a significant impact on quality of life,[2] high morbidity levels,[3] reduced life expectancy[4] and excess mortality.[5] The lifetime prevalence of major depressive disorder (MDE) is 11.2%.[6] Prevalences tend to be higher in low and middle income countries such as Pakistan, where depression prevalences of 45.9% have been reported.[7] In primary care (PC), the prevalence of MDE varies significantly in a range from 4.5% to 47.8%.[8]

In the 2015 Colombian National Mental Health Survey, the prevalence of major depression in the general population was 5.4 (95% CI: 4.6–6.4), 2.3 (95% CI: 1.8–2.9) and 0.8 (95% CI: 0.5–1.3) for lifetime, the last year and the last month, respectively.[9] In Bucaramanga, the prevalence of clinically significant depressive symptoms (CSDS) was 22.3% (95% CI: 20.0–24.6) and 11.2% for major depressive disorder (MDD) (95% CI: 9.7–12.9%).[10] A later population study in adults living in Bucaramanga (N = 266), using the Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I), reported a prevalence of 16.5% (95% CI: 12.3–21.6),[11] which confirms the high prevalence of depression in this region.

In spite of its high burden, chronicity and recurrent nature, depression is underdiagnosed in PC, as approximately 50% of patients who present depression will not be detected.[12] This diagnostic gulf could be explained by the fact that more than 75% of patients with depression initially consult a family or PC physician with little training in the identification of depressive disorders,[13] time constraints in busy PC environments[14] and the lack of validated screening tools in low and middle income countries.[15]

Because of the above, programmes have been developed to recognise depression,[16,17] which recommend standardised tools. A number of tools exist to identify cases of depression; however, their benefits have not been fully determined and the literature reveals contradictory results.[18] A recent systematic review suggests that of the screening tools, only the Patient Health Questionnaire-9 (PHQ-9) attains the optimum accuracy level for depression.[19] The PHQ-9 is an adjectival scale derived from the Primary Care Evaluation of Mental Disorders (PRIME-MD) to assess depressive symptoms using the DSM-IV criteria.[20] The PHQ-9 is shorter than most of the depression screening scales[21,22] and is considered the best screening tool for depression in PC due to its accuracy, brevity, being in the public domain and multipurpose, and ease of administration, scoring and interpretation.[19,23] The PHQ-9 has been translated into more than 20 languages and used in many countries and

contexts.[24] In PC, the sensitivity of the PHQ-9 was between 0.71 and 0.84 (mean 0.77) and its specificity was between 0.90 and 0.97 (mean 0.94),[25] confirming its adequate psychometric performance in PC, albeit with some variations in the cut-off point (COP) and psychometric parameters that can be explained by the influence of cultural aspects in the response pattern.[23] Its broad use is also supported by the findings of Williams et al., who concluded, in an analysis of more than 38 studies with more than 32,000 PC patients, that the PHQ-9 was equal or superior to other measures of depression.[22] In addition, the DSM-5 MDD working group and the NICE guidelines consider the PHQ-9 the preferred measure to assess the presence of the depression and quantify its severity.[21,22,26]

The PHQ-9 has been evaluated in Colombia in university students[27]: however, it was not compared to a gold standard. The PHQ-9 criteria therefore need to be validated in PC in Colombia against a gold standard, in particular due to the opportunity PC services represent in the early detection of depression.[28] As a result, this study's objective was to assess the validity of the PHQ-9 criteria, comparing it with the Mini-International Neuropsychiatric Interview (MINI) for screening for depressive symptoms in adult PC users in the Bucaramanga metropolitan area.

Materials and methods

Design

This study was designed and analysed based on the recommendations of the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) declaration.[29] An analytical observational study of the validity of a scale's criteria was conducted using reference criteria.

Participants

Local PC users of both genders, aged 18–65 years were included. The PC centres belong to the Instituto de Salud de Bucaramanga [Bucaramanga Health institute] (ISABU), a State social enterprise that coordinates primary health care services in the Bucaramanga metropolitan area.

Subjects with psychoactive symptoms, cognitive decline, delirium or an intellectual disability that might prevent them from responding to the tools, those under the effects of psychoactive substances, with functional changes in vision or hearing that might prevent them from understanding the content of the survey, and those who did not understand Spanish were excluded. The sample size was calculated to evaluate the hypothesis regarding the characteristics of a diagnostic test[30]:

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

where π_1 is the sensitivity of the standard (0.96) and π_2 is the anticipated sensitivity of the PHQ-9 (0.88); $Z_{1-\alpha/2}$ was set at 1.96 and $Z_{1\text{-}B}$ at 1.28; δ was set at 0.08 ($\pi_1 - \pi_2$). The result was 214. Participants were selected consecutively as they attended the health centres until a maximum number of subjects above 214 had been surveyed.

Procedures

The study was approved the ethics committee of ISABU and the Universidad de Santander [University of Santander], taking into account the current international[31] and national norms[32] for research with human subjects.

The PHQ-9 was translated following the recommendations for adaptation of self-reporting tests.[33] A direct translation from the original scale was performed by two independent certified bilingual translators; discrepancies between the two translations were discussed, then a backtranslation was done into English, which was reviewed by the research team to assess its closeness to the original scale. The translated scale was then reviewed by 10 psychiatrists with research expertise or clinical experience to verify whether the items were consistent with the construct of depression, who also commented on the comprehension and wording of the items. Ten people from the general population with a history of depression also gave their opinions on the comprehension of the questions. The research group analysed and incorporated the patients' and experts' observations to obtain the new Colombian version (Fig. 1). With the new version of the scale, a pilot test was conducted with 21 subjects with characteristics similar to the study subjects but in other centres. They answered the questions without difficulties and no adjustments to the grammatical structure were needed.

The research team was trained in the structured psychiatric interview (MINI) and hetero-administration of the PHQ-9. The people responsible for administering the scales and structure interviews were professionals with clinical experience (four psychologists, two general practice residents and one psychiatrist) who received eight hours of training by the lead author, with theoretical and practical sessions, role playing, and observation of pilot interviews with feedback. The study participants were contacted in the waiting room as they arrived for outpatient appointments with a general practitioner for any reason. One member of the research group explained the nature of the study and gave them the informed consent form. The screening scales were read by trained members of the research team. After completing the PHQ-9, each participant was assessed on the same day in another consulting room by a different trained member of the team (psychologist or psychiatrist) who did not know their PHQ-9 result, to administer the MINI depression module. The questionnaires were reviewed by two independent reviewers and saved in a form generated in Excel.

Tools

PHQ-9

The PHQ-9 is a screening scale that measures the presence and severity of depressive symptoms.[34] The PHQ-9[35] is made up of the nine symptoms from DSM-IV MDE criterion A.[20] These nine items are arranged in the form of an adjectival scale that assesses the presence of the symptom in the last two weeks ("not at all", "several days", "more than half the days" and

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 austed mismo de alguna manera				

Fig. 1 – Patient Health Questionnaire (PHQ-9), version for Primary Care, Bucaramanga, Colombia.

"nearly every day"), scored from 0 to 3 to give a score between 0 and 27.[36]

It can be self- or hetero-administered and is used both algorithmically to make a probable diagnosis of MDE or as a continuous measure of scores from 0 to 27 with cut-off points (COP) at 5, 10, 15 and 20 representing levels of depressive symptoms, i.e. mild, moderate, moderately severe and severe.[34] The scores can also be used dichotomously based on a COP to classify subjects with or without CSDS.[37] According to Kroenke et al., the psychometric characteristics of the PHQ-9 have a sensitivity of 88% and a specificity of 88%, adequate internal consistency (Cronbach's α of 0.86–0.89), a test–retest score of 0.84, a concordance between self-administered and evaluator-administered tests of 84% and an area under the curve (AUC) of 0.95.[34] In this study a COP of 8 or more was used to identify cases of CSDS, based on the meta-analysis by Manea et al.[23] and the study by Rancans et al. in PC.[38]

Mini-International Neuropsychiatric Interview

The MINI is a brief structured diagnostic interview that explores the diagnostic categories of the DSM-IV and the ICD-10.[39] Its original version was developed by Sheehan et al.[39] and Lecrubier et al.[40] in the United States and France. It contains 130 questions organised into modules that assess 16 disorders from axis I of the DSM-IV and one personality disorder. The original version in English has a sensitivity range between 0.46 and 0.94 and a specificity between 0.72 and 0.97,[39,40] excellent inter-rater (kappa 0.70) and test–retest reliability, and moderate validity of criteria compared to the Composite International Diagnostic Interview (CIDI) and the SCID-P.[39,40] The MINI quickly gained international acceptance,[41–43] has translated versions in 43 languages[39] and its reliability and validity have been explored in its Italian,[44] Japanese,[45] Norwegian,[46] Moroccan[47] and Portuguese[48] versions. The average administration time is 18.7 ± 11.6 min, with a mean of 15 min.[39] Together with the CIDI and the SCID-I, the MINI is considered a globally accepted gold standard for the diagnosis of mental disorders in clinical and research settings.[49]

Hospital Anxiety and Depression Scale

The Hospital Anxiety and Depression Scale (HADS) was designed by Zigmond and Snaith in 1983[50] to detect mood disorders, especially those associated with somatic symptoms. It consists of 14 items, with an anxiety subscale (odd items) and a depression subscale (even items). Each item is graded on a four-point frequency scale from 0 to 3. The HADS has been translated into most European languages, Arabic, Hebrew, Urdu, Japanese and Chinese[51] and its reliability and validity has been demonstrated in numerous studies.[52] In Colombia, it was validated in cancer patients, finding adequate internal consistency (Cronbach's α of 0.85), a COP of 8 for the anxiety subscale and 9 for the depression subscale.[53] These psychometric properties were confirmed in a populational sample

($n = 1500$) in several cities in Colombia.[54] In this work, the version adapted by Rico et al. was used.[53]

Patient Health Questionnaire-2

The Patient Health Questionnaire-2 (PHQ-2) consists of the first two items of the PHQ-9, which are necessary to suspect the presence of depression according to the DSM-IV criteria.[55] The scoring system is the same as for the PHQ-9 and scores range from 0 to 6. A COP of 3 is optimal for screening, but a recent meta-analysis suggests that a COP of 2 may increase its sensitivity.[56] Patients who score positive for CSDS should be evaluated with the PHQ-9 to determine whether they meet the MDE criteria.[57] Its clinical utility stems from the fact that it reduces the time taken in normal PC consultations, which are usually busy.[58] The PHQ-2 has been found to have psychometric performance comparable to the PHQ-9, with good reliability, validity and sensitivity to change.[56] In this work, a COP of 2 or above was used to identify patients with CSDS.[59]

Statistical analysis

The data were analysed in SPSS version 20.0,[60] carefully verified and reviewed twice. A descriptive analysis of the qualitative and quantitative variables was performed. Cronbach's α and McDonald's ω coefficients were calculated to assess internal consistency; for concurrent validity, Spearman or Pearson correlations were estimated depending on the distribution of the variables. To assess the accuracy of the PHQ-9 as a screening tool compared to the MINI, the receiver operating characteristics (ROC) and AUC were analysed. The optimum COP for the PHQ-9 was determined taking into account validity indices: sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), positive and negative likelihood ratios (LR), Youden's index and ROC curve/AUC analysis.

Results

Participant characteristics

Three hundred and eighty-four users were contacted, of whom 95 did not agree to participate. Of the surveys conducted, 46 were discarded due to missing and inconsistent data; the analysis therefore included 243 participants, of whom 184 (75.7%) were female. The average age was 34.05 years with a standard deviation (SD) of 12.47 years. For male participants the average age was 33.59 years with a SD of 12.89 years, while for female participants the average was 34.20 years with a SD of 12.37 years. The sociodemographic characteristics of the sample can be seen in Table 1.

The prevalence of CSDS was 27.2% according to the results of the PHQ-9 and 21.8% according to the MINI structured interview.

Internal consistency

A Cronbach's α coefficient of 0.80 and an ω coefficient of 0.81 were obtained. The overall internal consistency of the scale if each item is eliminated is shown in Table 2.

Table 1 – Description of the sociodemographic characteristics of patients with or without minor depressive symptoms receiving healthcare at Primary Care centres.

Variables	No.	%
Gender		
Male	59	24.30%
Female	184	75.72
Area of origin		
Urban	210	86.42
Rural	33	13.58
Marital status		
Single	98	40.33
Married	53	21.88
Cohabiting	83	34.16
Divorced	5	2.06
Widowed	4	1.65
Education		
Primary, not completed	14	5.75
Primary, completed	60	24.69
Secondary, not completed	36	14.81
Secondary, completed	88	36.21
Vocational, not completed	18	7.41
Vocational, completed	1	0.41
Technological, completed	3	1.23
University, not completed	3	1.23
University, completed	20	8.23
Socio-economic stratum		
Stratum 1	103	42.39
Stratum 2	84	34.57
Stratum 3	47	19.34
Stratum 4	6	2.47
Stratum 5	3	1.23

Table 2 – Description of the internal consistency of the PHQ-9 Colombian version: Cronbach's α and McDonald's ω by item.

PHQ-9 questions	Cronbach's α	McDonald's ω
Q1	0.774	0.780
Q2	0.767	0.772
Q3	0.795	0.800
Q4	0.776	0.784
Q5	0.792	0.798
Q6	0.778	0.782
Q7	0.788	0.794
Q8	0.798	0.805
Q9	0.799	0.802

Convergent validity

The Kolmogorov–Smirnov test was used to establish the normality of the variables, for the purpose of deciding the type of test to use for the concurrent validity analysis of the PHQ-9 against the PHQ-2 and the HADS depression subscale (HADS-D). These variables were not found to have a normal distribution, to Spearman's Rho was used. The Spearman's Rho was 0.646 for the HADS-D ($P < 0.010$) and 0.701 for the PHQ-2 ($P < 0.010$).

Table 3 – Description of the various cut-off points for the PHQ-9 Colombian version and validity coefficients.

Cut-off point	Sensitivity	Specificity	Youden's index	% correctly classified	PPV	NPV	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
≥7[a]	0.90[a]	0.82[a]	0.72[a]	83.54[a]	0.48[a]	0.98[a]	4.93[a]	0.12[a]
≥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

[a] Psychometric indices associated with the optimum cut-off point of ≥7.

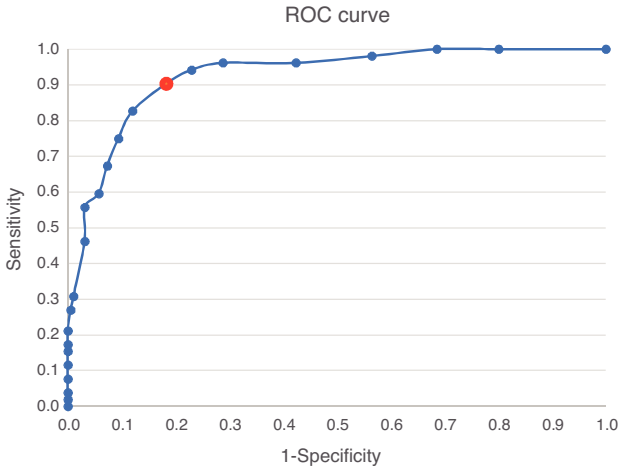


Fig. 2 – Receptor operating characteristics (ROC) of the PHQ-9 Colombian version compared with the MINI as a reference standard for depression (N = 146).

Criterion validity

The ROC curve (Fig. 2) and accuracy indices for the PHQ-9 produced the results shown in Table 3. The AUC was 0.92 (95% CI: 0.88–0.963).

The optimum COP was a PHQ-9 score ≥7 (sensitivity 90.38 [95% CI: 81.41–99.36]; specificity 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's 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]).

Discussion

To the best of our knowledge, this is the first study on PHQ-9 criterion validity in PC in Colombia. The prevalence of MDE in this study was 21.8%. The Colombian version of the PHQ-9 demonstrated excellent diagnostic performance as a depression screening tool, as can be seen from the ROC curve and AUC. The PHQ-9 also demonstrated an adequate balance of sensitivity and specificity at the COP of ≥7 when compared with the MINI as a reference standard, establishing the PHQ-9's adequate criterion validity. The comparison of PHQ-9 scores against those from the HADS-D and the α and ω coefficients demonstrated good convergent validity and adequate internal consistency.

The percentage of subjects classified as having CSDS based on the PHQ-9 with the pre-established COP was 27.2% (95% CI: 26.3–28.9), higher than the prevalence of 22.3% (95% CI: 20.0–24.6) found in Bucaramanga using the Zung Self-rating Depression Scale.[10] This difference can be explained by the poor diagnostic performance of the Zung scale in the Colombian population.[61] With regard to the prevalence of MDE based on the MINI, in this sample it was 21.8% (95% CI: 20.8–23.5), which is within the expected range based on a meta-analysis of 41 studies in PC with an adjusted global prevalence of 19.5% (95% CI: 15.7–23.7).[62] However, the prevalence of MDE in this study is a little higher than the 16.5% (95% CI: 12.3–21.6) reported in previous studies in the general population in Bucaramanga,[11] which can be explained by the fact that this study was conducted in people attending PC centres, where the prevalence of depression is higher than in the general population[63] and by the large proportion of women.[64]

Cronbach's α coefficient was 0.80 and McDonald's ω coefficient was 0.81, indicating good internal consistency.[65,66] For a self-reporting tool to be reliable, Cronbach's α and McDonald's ω need to be at least 0.70.[67] The internal consistency found in this study is in keeping with a previous study in Colombia[37] and with others conducted in different languages, whose coefficients ranged from 0.79 to 0.89.[68,70,71]

Previous studies have demonstrated that the PHQ-9 has adequate concurrent validity with various measured, including the Hamilton Depression Rating Scale (HAM-D), short health assessment forms and even the PHQ-2.[72] In our study, total PHQ-9 scores showed a statistically significant positive correlation with HADS-D and PHQ-2 scores (Spearman's Rho of 0.64 [$P < 0.01$] for HAM-D and 0.70 [$P < 0.01$] for PHQ-2), in keeping with previous studies in which the Pearson's coefficients for the PHQ-9 with the HAM-D and the Beck Depression Inventory (BDI) were 0.52 and 0.76, respectively.[68,73–75] Meanwhile, a study of patients with Parkinson's disease showed that the PHQ-9 correlated positively with the Self-rating Depression Scale and the 15-item Geriatric Depression Scale, with

a Spearman's coefficient of 0.63 in both cases.[76] The correlation coefficients found in this study confirm the convergent validity of the PHQ-9, with Spearman's correlation coefficients between 0.60 and 0.80 indicating a good or considerable positive correlation.[77]

With regard to the COP, various studies have recommended a COP of 10 in the PHQ-9 for the identification of MDE.[34] For example, in a study of PC users in China, an optimum COP of 10 produced a sensitivity of 0.87 and a specificity of 0.8.[69] However, a recent meta-analysis of 18 studies demonstrated that the optimum COP of the PHQ-9 could range from 8 to 11, depending on the population studied; nevertheless, the balance of sensitivity and specificity is maintained for a COP of 7 (5 of the 18 studies included).[23] In our study, the COP of 7 appears to have given the optimum balance between sensitivity and specificity, which was confirmed with an additional measure of accuracy: Youden's index,[78] defined as the maximum vertical distance between the ROC curve and the 45 degree line, as an indicator of how far the curve is from an uninformative test.[79] Youden's index is a function of sensitivity (Se) and specificity (Sp); it is calculated as (Se + Sp-1)[80] and should be considered alongside the ROC curve as they are usually related.[81] The range is 0–100 when converted into a percentage. Values >50% are generally considered acceptable for diagnostic accuracy.[82]

The values associated with the COP in our study are consistent with a study in older adults in PC, in whom PHQ-9 criterion validity was assessed by administering the MINI, in which an optimum COP ≥ 7 (sensitivity 0.92; specificity 0.78) demonstrated the best psychometric characteristics.[83] Nevertheless, this COP of 7 is lower than that found in most studies with the PHQ-9 in other populations. The cultural and demographic characteristics of the samples may be the reason for this difference.[84] Stigma is an important aspect that can also influence people's response pattern to depression screening scales in our population, causing shame in people with mental illnesses, which limits the identification of psychopathological phenomena.[85,86] It is worth noting that PHQ-9 COPs tend to be lower in middle and low income countries[87–89] compared with high income countries.[75,90,91] However, there are no studies looking at this phenomenon. This difference in optimum COP highlights the importance of validating screening tools in different social and cultural contexts.[92]

For a COP of ≥7, the sensitivity and specificity of the PHQ-9 in this sample were 90% and 83%, respectively. These findings are consistent with the study by Wang et al., in which the COP of ≥7 allowed an adequate balance between sensitivity and specificity (sensitivity 85%; specificity 86%).[93] The accuracy indices in our study are therefore considered appropriate, as the screening tool is considered good when its sensitivity is between 79% and 97% and its specificity is between 63% and 86%.[94] Wittkampf et al. systematically reviewed the psychometric properties of the PHQ-9 and found a sensitivity of 77% (71–84%) and specificity of 94% (90–97%), including studies in subgroups with a high prevalence of depression, such as PC users.[25]

The LR+ and LR− of the PHQ-9 in our sample, for a COP of ≥7, were 4.93 and 0.12, respectively. This means that, in a similar clinical context, a positive result in the PHQ-9 (COP ≥ 7) if five times more common in a patient with depression than in one without depression, while a subject with a negative result would have a likelihood of having depression of less than 2%.[95] These results are comparable with those obtained in the Chinese version of the PHQ-9, which with a COP of ≥7, had an LR+ and LR- of 5.99 and 0.17, respectively.[93]

The AUC of this Colombian version of the PHQ-9 for PC was 0.92, which indicates a high degree of accuracy[96] and is consistent with previous studies in PC and other populations.[69,71,93]

This study's main strengths include the use of a clinical reference criterion to assess the PHQ-9 validity criterion, the adequate participant response rate (75.3%), adequate training of interviewers, adherence to the QUADAS-2 guidelines[29] and the execution of a rigorous analysis plan. In addition, the PHQ-9 was translated in accordance with the standardised guidelines for transcultural adaptation of scales. The linguistic adaptation was supported by a group of experts, guaranteeing appropriate content validity.

This study has several limitations. Firstly, our study was conducted in a PC context, therefore the results cannot be generalised to the general population, whose characteristics would produce a different response pattern.[84] Secondly, the study was limited to adults. There is growing evidence that adolescents are particularly affected by depressive disorders,[97] so future studies in Colombia would need to assess the psychometric performance of the PHQ-9 in this population. Thirdly, this was a cross-sectional study, and as a consequence there will be a need, in the future, to design longitudinal studies to establish the sensitivity to change of the PHQ-9 in the Colombian population, as works exist that have used in to assess response to treatment of depression.[98] Fourthly, the fact that the sample was predominantly female (75%) could affect the estimates of accuracy indices, as the prevalence of depression is higher in women than in men, giving a higher number of positive cases of depression.[99] And fifthly, one relative weakness in the sample size, which was calculated following the recommendations of Sanchez et al. for comparing the sensitivity of a screening test with a reference standard.[30] However, other authors, such as Buderer[100] and Obuchowski[101] demand larger samples. On the other hand, following Bean's criteria for comparing the sensitivity or specificity of two diagnostic tests, sample sizes similar to ours are obtained.[102]

In keeping with global results, the Colombian version of the PHQ-9 for PC has excellent psychometric performance as a screening test, which guarantees that it can be used in contexts with few resources and with weaknesses in the healthcare system, where the availability of psychiatrists is limited.[103] Among the strategies to limit the burden of mental health disorders in low and middle income countries is integration of mental health in PC.[104] One of the major barriers to achieving this goal is the lack of easy-to-administer and validated screening tools to detect depression. The validation of instruments such as the PHQ-9 in these contexts can help to solve this problem.[105] It is known that just screening for depression is insufficient to mitigate the growing care needs for mental health disorders in low and middle income countries; nevertheless, given that depression contributes significantly to the burden of disease, having validated screening tools is the first step towards solving this problem.[106] In

some low income countries, there are cost-effective depression intervention programmes, in which screening tools can be used to identify appropriate participants.[107] One of the main components of effective mental health interventions in PC is monitoring depressive symptoms using simple, brief and easy-to-administer questionnaire such as the PHQ-9.[108]

With the validation of this version of the PHQ-9, researchers in Colombia now have valid and reliable psychometric information about depression screening in PC, which will enable the PHQ-9 to be used in studies where it is necessary to identify depressive symptoms with an appropriate COP.

In conclusion, the results of this study indicate that the Colombian version of the PHQ-9 is a valid and reliable tool for screening for depression in a PC context in Bucaramanga, with a COP of 7 or above. The psychometric properties of this version of the PHQ-9 will need to be evaluated in different populations and other regions of the country. Future studies in Colombia should assess the PHQ-9's sensitivity to change.

Funding

This work was funded by the faculty of medicine of the Universidad de Santander (UDES) and the Instituto de Salud de Bucaramanga (ISABU). Project code: PIFE0118020041816EJ.

Authors' contribution

Carlos Arturo Cassiani-Miranda: design, scale adjustment, training of survey-takers and interviewer, collection of information, statistical analysis, digitisation, drafting and revision of the article.

Angy Karina Cuadros-Cruz: design, scale adjustment, collection of information and drafting of the article.

Harold Torres Pinzón: design, scale adjustment, statistical analysis, digitisation, drafting and revision of the article.

Orlando Scoppetta: scale adjustment, statistical analysis, drafting and revision of the article.

Jhon Henrry Pinzón-Tarrazona: collection of information, digitisation and drafting of the article.

Wendy Yulieth López-Fuentes: collection of information, digitisation and drafting of the article.

Andrea Paez: collection of information and drafting of the article.

Diego Fernando Cabanzo-Arenas: digitisation, drafting and revision of the article.

Sergio Ribero-Marulanda: collection of information, drafting and revision of the article.

Elkin René Llanes-Amaya: collection of information and drafting of the article.

Conflicts of interest

The authors have no conflicts of interest to declare.

Acknowledgements

To the expert panel for their contributions to the validation of appearance and content: 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.

REFERENCES

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;392:1647–61.
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 Suppl. 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;130:26–36.
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 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, Scoppetta 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; 2015 [Internet]. Available from: http://www.wma.net/es/30publications/10policies/b3/ [Accessed 20 February 2019].
32. Ministerio de Salud. Resolucion 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 Suppl. 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;76:113–9.
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, Tansella M, Amaddeo F. 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 Medición. 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.
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 randomised controlled trial. 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. 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 lowand 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.