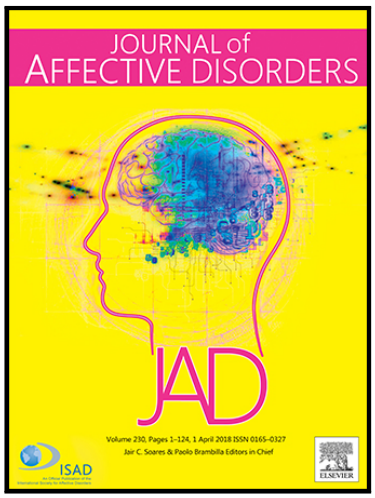

Validity of the Patient Health Questionnaire-2 (PHQ-2) for the Detection of Depression in Primary Care in Colombia

Orlando Scoppetta , Carlos Arturo Cassiani-Miranda , Karen Nicolle Arocha-Díaz , Diego Fernando Cabanzo-Arenas , Adalberto Campo-Arias

PII: S0165-0327(20)32792-0
DOI: https://doi.org/10.1016/j.jad.2020.09.096
Reference: JAD 12499

To appear in: *Journal of Affective Disorders*

Received date: 24 December 2019
Revised date: 1 August 2020
Accepted date: 21 September 2020

Please cite this article as: Orlando Scoppetta , Carlos Arturo Cassiani-Miranda , Karen Nicolle Arocha-Díaz , Diego Fernando Cabanzo-Arenas , Adalberto Campo-Arias , Validity of the Patient Health Questionnaire-2 (PHQ-2) for the Detection of Depression in Primary Care in Colombia, *Journal of Affective Disorders* (2020), doi: https://doi.org/10.1016/j.jad.2020.09.096

Highlights

- The Patient Health Questionnaire-2 (PHQ-2) was tested in a Colombian sample.
- The gold standard was the MINI interview
- The PHQ-2 demonstrates acceptable psychometric performance.
- We recommended to consider screening tests with better indicators

Validity of the Patient Health Questionnaire-2 (PHQ-2) for the Detection of Depression in Primary Care in Colombia.

Orlando Scoppetta [a], Carlos Arturo Cassiani-Miranda * [b], Karen Nicolle Arocha-Díaz [b], Diego Fernando Cabanzo-Arenas [b], Adalberto Campo-Arias [c].

a. Universidad Católica de Colombia, Faculty of Psychology, GAEM Group (Research methods applied to behavioral sciences), Bogotá, Colombia.
b. Universidad de Santander, Faculty of Health Sciences, medicine program, Neurosciences Research Group. Bucaramanga, Colombia.
c. Universidad del Magdalena, Faculty of Health Sciences, Health Psychology and Psychiatry Research Group. Santa Marta, Colombia.

*School of Health Sciences, Calle 70 No 55-210, Bucaramanga, Colombia, Tel: 057-6516500. Fax: 57-7-6516500 cel: 57-3184455273. E-mail address: Kassio30@hotmail.com, ca.cassiani@mail.udes.edu.co

Abstract

Background: The PHQ-2 is a screening test for major depressive disorder (MDD) derived from PHQ-9, which has shown to be useful in the detection of cases of clinical

significance in previous studies. The psychometric properties of PHQ-2 in the Colombian population are unknown.

Methods: PHQ-2 were assessed in 243 patients in general medical consultations using the Mini-International Neuropsychiatric Interview (MINI) as the gold standard. Internal consistency, convergent validity and criterion validity were calculated by analyzing the Receptor Operating Characteristics (ROC) and the area under the curve (AUC).

Results: Cronbach's alpha and McDonald's omega coefficients were 0.71. Spearman's rho coefficients for correlations with PHQ-9 and HADS-D scores were 0.63 and 0.59 ($p<0.01$). AUC was 0.89. The optimal cut point was ≥2 with the following indicators: sensitivity 0.87 specificity 0.74; Youden index 0.60; PPV 0.47; NPV 0.95; (LR+) 3.24; (LR-) 0.18 (95% CI 0.09-0.37). Kappa coefficient between PHQ-2 and MINI for depression was .458 and 0.46 for HADS-D.

Limitations: Since this study was done with people attending hospital, which could has implications for the prevalence of depression, affecting the validity indicators of the instrument.

Conclusions: PHQ-2 show an acceptable test performance in the context of the study. However, the test could yield a considerable amount of false positives that would require specialized evaluation to establish a reliable diagnosis.

Key words: PHQ-2; reproducibility of results; screening; depression; primary health care; Colombia.

1. Introduction

Major depressive disorder (MDD) is a major public health problem that significantly contributes to the global burden of the disease worldwide (König et al., 2019; Steel et al., 2018). The prevalence of MDD is significantly higher in primary care (PC) than in the general population (Mitchell et al., 2009). Thus, the lifetime prevalence of a major depressive episode (MDE) is 11.2% in the general population (Kessler et al., 2015), while it can be as high as 48% in PC (Wang et al., 2017). In Colombia, last-month prevalence of MDE in users of PC services was 21.1%, which was assessed using clinical interviews (Cassiani-Miranda et al., 2020).

The highest number of MDE cases is managed in PC centers (Ferenchick et al., 2019). Only the most complicated cases, with little response or comorbidity, reach other levels of health care (Spijker et al., 2002). It has been estimated that more than 8 million medical consultations are for MDE in the United States and more than half of them are delivered at PC centers (Centers for Disease Control and Prevention, 2017). Similarly, around 7% of all PC consultations are for MDE worldwide (Shah, 1992).

Most MDE cases are treated by general practitioners (Harman et al., 2006). However, PC physicians often fail to make an accurate diagnosis. Consequently, symptoms may be underestimated leading to a high proportion of unidentified and untreated cases (Demyttenaere et al., 2004; Zastrow et al., 2008), with negative consequences on the service users' quality of life, costs to the healthcare system and country productivity (Luo et al., 2010). Failing at identifying MDE is partly due to the lack of trained professionals on

mental health issues and appropriate screening instruments (Campo-Arias and Barliza, 2018; Patel et al., 2008; Patel and Sartorius, 2008)

Available evidence suggests that the routine use of screening instruments can potentially improve case detection and treatment delivery for MDE cases at PC (Johnson et al., 2014; Mitchell et al., 2016; Romera et al., 2013; Thombs et al., 2017).

The use of short questionnaires (< 5 minutes) and ultra-short questionnaires (< 2 minutes) may improve MDE case detection at PC (Mitchell et al., 2011; Mitchell and Coyne, 2007). The growing interest in the use of these short instruments to detect depression is due to the advantages that these can offer in overcrowded PC settings where consultation time is limited (Mitchell and Coyne, 2007).

The Patient Health Questionnaire (PHQ-9) is the instrument of choice for assessing depression at PC (Ferenchick et al., 2019), showing excellent reliability, validity and great acceptance in different cultural contexts (El-Den et al., 2018; Pettersson et al., 2015), with a sensitivity and specificity of 78% (CI95% 0.70-0.84) and 87% (CI95% 0.84-0.90), respectively, at a cut-off point (CP) of ≥10 (Moriarty et al., 2015).

The Patient Health Questionnaire 2 (PHQ-2) is composed of the first two items of PHQ-9 which represent the cardinal signs for the diagnosis of MDE according to the DSM (depressed mood or anhedonia) (Kroenke et al., 2003). Like PHQ-9, the PHQ-2 has also shown to be an excellent screening instrument for detecting depression in different settings; with a sensitivity of 91% (CI 95% 0.85-0.94) and a sensitivity of 70% (CI 95% 0.64-0.76) for CP ≥ 2 (Manea et al., 2016). However, it has been poorly evaluated in low-and middle-income countries (LAMIC) (Hanlon et al., 2015; Lino et al., 2014). It should be also considered that the performance of screening instruments may vary according to the demographic, social and cultural characteristics of the population. Therefore, it is always

necessary to first make adjustments to achieve greater linguistic equivalence (Ramada-Rodilla et al., 2013) and secondly, to make modifications to the cut-off point since the sensitivity and specificity of the instruments gets affected by the prevalence of the problem at stake in a particular context (Brenner and Gefeller, 1997; Lange and Lippa, 2017).

There are no data published in Colombia about the psychometric properties of PHQ-2. Under the assumption that the PHQ-2 shows adequate validity, PC healthcare professionals in Colombia will have a valid, reliable, and easy-to-use strategy that helps better identify MDE at PC. Therefore, this study is aimed at evaluating the performance of the PHQ-2 against the Mini-International Neuropsychiatric Interview (MINI) in PC adult users in Bucaramanga, Colombia.

2. Methods

An observational analytical study of criterion validity was conducted using a gold standard scale and following the recommendations of the QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies) tool (Whiting et al., 2011).

This manuscript is part of an instrument validation study to identify common mental disorders at PC (Cassiani et al, 2020).

2.1. Participants

The sample size was calculated to evaluate the proportion difference between the characteristics of a new test and a recognized gold standard (Sánchez and Echeverry, 2001). The statistical details of the sample size calculation have been previously published (Cassiani et al, 2020). 243 adults of both sexes attending primary care centers in Bucaramanga, Colombia were included. Several reasons were considered to choose primary care centers to carry out this study: Most care for depression is delivered by general

practitioners (GPs) in primary care and many GPs have considerable experience in managing depression (Harman et al., 2006) Yet, clinicians find it challenging to diagnose depression accurately and often overestimate or underestimate levels of distress of their patients that sometimes result in false-positive or false-negative diagnoses (Zastrow et al., 2008)**.** Indeed, GPs are typically able to detect about half of true cases of depression in a one-off consultation and once diagnosed, not all depressed patients receive adequate and timely care (Duhoux et al., 2012; Takayanagi et al., 2015).

Patients with psychotic symptoms, cognitive impairment, delirium or intellectual disability that may hinder response to the instruments were excluded, as well as being under the effects of psychoactive substances, with functional alteration of vision or hearing that prevented the understanding the content of the questionnaire and those who did not understand Spanish. To identify patients with psychotic symptoms, the psychotic disorders module of the Mini-International Neuropsychiatric Interview (MINI) (Sheehan et al., 1998) (Lecrubier et al., 1997) was applied, and to detect those with cognitive impairment, the Colombian version of the Mini Mental State Examination was used (Rosselli et al., 2000).

2.2. Procedure

The study was approved by the Ethics Committee of the University of Santander, following the international World Medical Association Declaration of Helsinki (2013) and the national standards (Ministry of Health, 1993) applicable for human subject research.

The PHQ-2 was translated and adapted following the recommendations for adaptation of self-report tests (Beaton et al., 2000).

The research team was trained on the structured psychiatric interview (MINI) and hetero-application of the PHQ-2. Healthcare professionals with clinical experience (four

psychologists, two family medicine fellows and a psychiatrist) were responsible for applying the structured scales and interviews, they also received eight hours of training with theoretical-practical sessions, role-playing, observation of pilot interviews with feedback. Study participants were contacted in the waiting room as they arrived for outpatient consultation for any cause with a general practitioner. After completing the PHQ-2, the participants were evaluated on the same day in another office by another trained team member (psychologist or psychiatrist), who was unaware of the result of the PHQ-2, to apply the MINI MDE module.

2.3. Instruments

2.3.1. PHQ-9

The PHQ-9 is a screening scale that measures the presence and severity of depressive symptoms (Kroenke et al., 2001). It comprises the nine symptoms of DSM-IV Criteria for Major Depressive Disorder (Kocalevent et al., 2013; Spitzer et al., 1999). These nine items are arranged in a Likert response format that evaluates the presence of the symptom over the last two weeks ("not at all", "several days", "more than half of the days" and "almost every day"), which score from 0 to 3, getting a total score from 0 to 27 (Spitzer et al., 2000). The psychometric characteristics of the PHQ-9, according to Kroenke et al., present a sensitivity of 88% and specificity of 88%, an adequate internal consistency (Cronbach's alpha of 0.86-0.89), a test-retest score of 0.84, a concordance between the self-determined test vs. the test performed by the evaluator of 84% and an area under the curve of 0.95 (Kroenke et al., 2001). The process of adapting the version used in this study has been previously described in other studies (Cassiani-Miranda et al., 2019).

2.3.2. Patient Health Questionnaire-2 (PHQ-2)

The PHQ-2 consists of the first two items of the PHQ-9. The scoring formula is equal to that of PHQ-9 and the scoring range is 0 to 6 (Kroenke et al., 2003). In the PC context, the PHQ-2 showed a sensitivity of 86% and 61%, a specificity of 78% and 92% for CPs of ≥2 and ≥3 respectively (Arroll et al., 2010; Manea et al., 2016). Its clinical usefulness lies in the fact that it reduces the length of regular PC consultations, which are usually busy, considering that general practitioners have limited time (Inagaki et al., 2013). Considering the cited authors, the cut-off point of ≥ 2 was established to identify patients experiencing depressive symptoms with clinical importance (DSCI) in the current work, following the recommendations by Mitchell et al. (Mitchell et al., 2016). Although there is a previous version of PHQ-9 and, therefore, of PHQ-2 used in Colombia (Cassiani-Miranda et al. 2018), a new adaptation was considered for this study since the previously used version was taken from Pfizer website with prior permission (http: // www.phqscreeners.com/). Besides, there was no evidence of a translation and back-translation process recommended by international standards for cross-cultural adaptation of health assessment tools (Ortiz-Gutiérrez and Cruz-Avelar 2018; Ramada-Rodilla et al., 2013).

Despite having Spanish-language versions (Rodríguez-Muñoz et al., 2017; Saldivia et al. 2019), the questionnaire was completely translated since there may be minimal linguistic details that can affect its validity (Ortiz-Gutiérrez and Cruz-Avelar, 2018).

2.3.3. Hospital Anxiety and Depression Scale (HADS)

The HADS was designed by Zigmond and Snaith for the detection of affective disorders associated with somatic symptoms, which consists of 14 items with a subscale of anxiety (odd items) and depression (even items). Each item is evaluated based on a four-point scale of frequency ranging from 0 to 3 (Zigmond and Snaith, 1983). The HADS has been translated into most European languages, Arabic, Hebrew, Urdu, Japanese, and Chinese (Snaith, 1992) and has demonstrated reliability and validity in numerous studies (Moorey et al., 1991). In Colombia, it was validated in oncological patients showing an adequate internal consistency (Cronbach's alpha of 0.85) and a CP of 8 for the anxiety subscale and 9 for the depression subscale (Rico et al., 2005). These findings were confirmed through a population sample (n 1500) conducted in several cities in Colombia (Hinz et al., 2014). The adapted version elaborated by Rico et al. was used in this research (Rico et al., 2005).

2.3.4. Mini-International Neuropsychiatric Interview (MINI)

The MINI is a brief structured diagnostic interview that explores the diagnostic categories of DSM-IV and ICD-10. Its original version was developed by Sheehan (Sheehan et al., 1998) and Lecrubier (Lecrubier et al., 1997) in the United States and France, which contains 130 questions organized into modules evaluating 16 DSM-IV axis I disorders and a personality disorder. The original English version showed a range of sensitivity between 0.46 and 0.94 and specificity between 0.72 and 0.97 (Sheehan et al., 1998), a good interrater reliability (Kappa 0.70) and test-retest and moderate criterion validity compared to CIDI and SCID-P (Lecrubier et al., 1997; Sheehan et al., 1998). The MINI has gained rapid international acceptance (Balázs et al., 2003; Pinninti et al., 2003), has been translated into 43 languages (Sheehan et al., 1998) and its reliability and validity

has been explored in different countries (de Azevedo Marques and Zuardi, 2008; Mordal et al., 2010; Otsubo et al., 2005).

The average administration time is 18.7±11.6 minutes with an average of 15 minutes (Sheehan et al., 1998). The MINI along with CIDI and SCID-I are considered gold standards that are globally accepted for the diagnosis of mental disorders in clinical and research settings (Tejada et al., 2014).

2.4. Statistical Analysis

A descriptive analysis of qualitative and quantitative variables was conducted. The following indicators were calculated: Cronbach's alpha and McDonald's coefficient omega for internal consistency. The lowest acceptable value for the test is 0.70 (Ventura-León and Caycho-Rodríguez, 2017; Watkins, 2017).

Spearman's non-parametric correlation coefficients were calculated as convergent validity indicators concerning the PHQ-9 and HADS-D tests since the PHQ-2 scores showed an asymmetric distribution. The existence of convergent validity was accepted, provided that the value of r_s was higher than 0.60 (Post, 2016).

For criterion validity, the area under the receiver-operating-characteristic (ROC) curve was calculated, which compared the PHQ-2 scores to the presence of an MDE in a MINI interview to determine the probability that a participant who met the criteria for MDE was correctly classified and to choose the cut-off point with the corresponding indicators. Sensitivity, specificity, positive and negative predictive value, and positive and negative likelihood ratios were calculated with the respective 95% confidence intervals, 95% CI (Hanley and McNeil, 1982). The probability of concordance between PHQ-2 and MINI beyond chance was calculated using Cohen's kappa coefficient (Cohen, 1960).

In order to determine a more precise CP, the CP that achieved a better balance within the other indicators was chosen, according to the Youden index, which corresponds to the difference between correct positive responses and incorrect positive responses, and whose maximum value is usually reached with the best balance between sensitivity and specificity (Cerda and Cifuentes, 2012; Youden, 1950).

The analyses were carried out using the STATA® statistical software version 13 and the application of formulas in spreadsheets.

3. Results

3.1. Sample Characteristics

The sample consisted of 243 participants of whom 184 (75.7%) were female. The average age was 34.1 years (median 31.0 and SD=12.5). The average was 33.6 for men (SD=12.9) and 34.2 for women (SD=12.4) (table 1).

3.2. Prevalence of Depression Diagnoses

PHQ-9 scores ranged between 0 and 25 with an average of 5.4 (SD=5.2) and a median of 4.0. PHQ-2 scores ranged between 0 and 6 with an average of 1.2 (SD =1.7).

Twenty-one percent of participants (21.4%) met the criteria for MDE according to the MINI. There was a significant difference between depression rates in men (15.2%) and in women (23.4%). According to the PHQ-9, 33.7% of the participants met criteria for DSCI) (37.5% for women and 22.0% for men). The prevalence of DSCI using PHQ-2 was 33.7% for the entire group of participants (38.0% for women and 20.3% for men).

3.3. Internal Consistency

In the analysis of internal consistency by calculating the Cronbach's alpha and McDonald's omega coefficient, a value of 0.71 was obtained in both cases.

3.4. Convergent Validity

To assess convergent validity, Spearman's rho coefficients (suitable for non-parametric correlation tests) were calculated for correlations with total PHQ-9 and HADS-D scores of 0.63 and 0.59, respectively ($p<0.01$).

3.4. Criterion Validity

The area under the ROC curve was 0.89 (95% CI 0.84-0.93). The estimated optimal CP was ≥2 with the following indicators: sensitivity of 0.87 (95% CI 0.74-0.94); specificity of 0.74 (95% CI 0.66-0.79); Youden index of 0.60 (95% CI 0.54-0.66); positive predictive value of 0.47 (95% CI 0.37-0.57); negative predictive value of 0.95 (95% CI 0.90-0.98); positive likelihood ratio (LR+) of 3.24 (95% CI 2.50-4.20); negative likelihood ratio (LR-) of 0.18 (95% CI 0.09-0.37). The Kappa coefficient between PHQ-2 and MINI for depression was .458 and 0.46 for HADS-D.

4. Discussion

Broadly speaking, the PHQ-2 is profiled as an acceptable screening instrument for the rapid identification of MDE cases. The instrument shows acceptable convergent validity and modest performance indicators against the structured clinical interview.

Internal consistency was 0.71 (coefficient alpha and coefficient omega) in this study, similar to the PHQ-2 for the Russian version (0.79) (Rancans et al., 2018) and the Hong-Kong's version (0.76) (Yu et al. 2011). For a self-report instrument to be reliable, it

is suggested that Cronbach's alpha be at least 0.70 (Kottner & Streiner, 2010; Ventura-León, 2018; De Vet et al., 2017).

A sensitivity of 0.83 and specificity of 0.90 for an optimal CP ≥ 3 was found in the original PHQ-2 validity study (Kroenke, Spitzer & Williams 2003). In the current study, the sensitivity and specificity of PHQ-2 were 0.87 and 0.74 respectively (for optimal PC ≥ 2), which indicate comparable results in sensitivity, although a substantial difference in specificity is observed in favor of the original version by Kroenke et al. (Kroenke, Spitzer & Williams 2003). These differences are consistent with the findings of the systematic review conducted by Manea et al., (Manea et al. 2016), suggesting that the accuracy of the PHQ-2 is lower in the involved studies than in the original validity study (Kroenke, Spitzer & Williams 2003).

Furthermore, the diagnostic meta-analysis of primary care studies (Arroll et al. 2010; Kroenke et al. 2003; Phelan et al. 2010; Zuithoff et al. 2010) found a pooled sensitivity of 0.84 (95% CI = 0.80-0.88) and a pooled specificity of 0.76 (95% CI = 074–0.79). The sensitivity and specificity of the PHQ-2 in this study (Se =0.87; Sp = 0.74) exceed the sensitivity being very close to the specificity of the PC studies included in the meta-analysis conducted by Manea et al. (Se= 0.84, Sp = 0.76) (Manea et al. 2016). These differences in the PHQ-2 psychometric indices may be due to different CPs, different prevalence in the target population and the response pattern of each population (Keszei, Novak & Streiner, 2010; Mitchell et al., 2016). This response pattern of each population can is determined by cultural, cognitive, educational, and sociodemographic factors of the population, variables that can vary from one country to another or even within the same country (Keszei et al., 2010).

Moreover, the LR+ for PHQ-2 was 3.24 and the LR- was 0.18. This means that, in a similar clinical setting, MDE patients are 3.24 times more likely to have PHQ-2 $\geq$ 2 scores compared to patients without MDE. Similarly, patients without MDE are 5.5 (1/0.18) more likely to have a negative result in PHQ-2 compared to those with MDE. This LR-value is considered desirable for a screening test (Moratalla Rodríguez, 2015).

The AUC found in this study was 0.89 which is comparable to the original version (0.93) (Kroenke et al. 2003; Mitchell et al. 2016). This indicates that the PHQ-2 showed adequate well discriminant validity (Staples et al. 2019).

Strengths and Weaknesses

To our knowledge, this is the first study that analyzes the validity of the PHQ-2 in Colombia, starting from the adaptation of the language, the comparison with a standard and the calculation of the complete set of indicators related to criterion validity.

Test application procedures were given in the context of methodological healthcare, which included the training of those responsible for test application and its detailed monitoring.

However, this study has some limitations to be considered when interpreting the results.

First, the MINI, the structured interview used as the gold standard, does not have a formal validation study in Colombia, which may compromise the validity of the results. However, since the MINI is widely used in international settings as calibration in Spanish-speaking countries (Tejada et al., 2014) (Shehan et al., 1998), this limitation was mitigated to some extent.

Secondly, compared to the MINI, the total of correctly classified cases using the PHQ-2 was 76.1%, assuming the cut-off point as ≥ 2. In raw data, 2.9% of the total of patients was classified as negative but met the criteria for MDE according to the MINI and 21% was classified as positive but according to the MINI did not meet the MDE criteria. In this regard, it cannot be ensured that when using ultrashort screening instruments such as the PHQ-2, patients who scored negative screening are true negatives and do not need any follow-up (Mulvaney-Day et al. 2018). While among the patients classified as positive, a significant proportion may not be cases of MDE, therefore they require a subsequent specialized evaluation to confirm the diagnosis, reducing the usefulness of a screening instrument.

This is why we advocate for highly sensitive instruments or for a more equal sensitivity/specificity balance for the detection of MDE in PC such as the PHQ-9 (Cassiani-Miranda et al. 2020; Mitchell et al. 2016). Screening instruments are helpful in identifying potential cases of major depressive disorders, and these are never a substitute for professional clinical evaluation. Due to the positive and negative predictive values obtained with PHQ-2, positive cases are unlikely to be lost without evaluation and treatment.

Future research

Developing tests is a long-term process that involves adaptations of the instrument to the changing social and cultural conditions. Therefore, considering that this is the first time that an analysis of the PHQ-2 properties is conducted in Colombia, it is necessary to study its behavior in other populations, for example, in a general outpatient population.

It is also important to emphasize that reliability is not a fixed property of a scale. Rather, reliability is a function of the instrument, the group with which it is being used, and its circumstances. It would be incorrect to state that the reliability of a scale is opposed to

the reliability of a scale used with a specific population for a given purpose (Keszei et al. 2010).

One of the public health implications of studies such as this one is the revival of the discussion about the impact of screening for depressive symptoms in PC. In this regard, despite the fact that some entities have recommended routine screening for depression in PC settings, the Canadian Task Force on Preventive Health Care (CTFPHC) and the UK National Screening Committee (UKNSC) (Canadian Task Force on Preventive Health Care et al. 2013; UK National Screening Committee s. f.) do not recommend routine screening for depression arguing lack of evidence of health benefits (Keshavarz et al. 2013; Thombs et al. 2014).

Thombs et al., 2014 warned about the harm of detecting depression in asymptomatic individuals such as the presence of a significant proportion of false positives and unnecessary treatment. Additionally, stigma and medicalization are situations to consider when rapidly detecting depression using a self-report scale (Smithson & Pignone, 2017). However, the US Preventive Services Task Force (USPSTF) (Siu et al., 2016) and the American Academy of Family Physicians (Maurer, 2012) recommended routine screening for depression to all adults as long as there are health system measures to ensure an subsequent accurate diagnosis, effective treatment and appropriate follow-up.

Thus, the harm of screening for depression in adults is small or null for the USPSTF (Siu et al., 2016). The controversy on this topic continues (Thombs et al., 2017) so ongoing studies are expected to provide more robust evidence on the routine evaluation of MDE in PC (Hamel et al., 2019).

On the other hand, the PHQ-2 did not classify five participants who showed suicidal behavior as people with depression when answering the question "Have you thought you

would be better off dead or of hurting yourself in some way?" at the highest degree of response ("almost every day"). It must be said that with the MINI, seven participants with this type of manifestation were no longer classified as depressed, two of them in the highest degree of response. It is interesting that with the PHQ-9, only one participant with suicidal ideation was left out of the depression classification.

Although suicidal behavior without depression is likely to occur, the weight of suicidal behavior on depression assessment instruments is an issue that requires further research, given the rising rates of suicide in the country (Instituto Nacional de Medicina Legal y Ciencias Forenses, 2019) and its proven association with depression (Hawton et al., 2013; Rotenstein et al., 2016).

5. Conclusion

The Colombian version of PHQ-2 has acceptable psychometric performance for MDE screening in adult PC users, with a CP greater than or equal to 2. However, the instrument may misclassify a significant number of patients, most of them false positives.

Authors Contributors

Orlando scoppetta : Study design, analysis of results and writing of the manuscript

Carlos Arturo Cassiani Miranda: Study design, data collection, staff training in structured interview, data analysis, writing of the manuscript and article submission

Diego Fernando Cabanzo Arenas: Data collection, information analysis and writing of the manuscript

Karen Nicolle Arocha Diaz: Data collection, information analysis and writing of the manuscript

Adalberto Campo: Statistical analysis and writing of the manuscript

Disclosures

The authors do not have conflicts of interest.

Funding source

This work was funded by the medical faculty of the University of Santander Project code: PIFE0118020041816EJ

References

Arroll, B., Goodyear-Smith, F., Crengle, S., Gunn, J., Kerse, N., Fishman, T., Falloon, K., Hatcher, S., 2010. Validation of PHQ-2 and PHQ-9 to screen for major depression in the primary care population. Ann Fam Med 8, 348–353. https://doi.org/10.1370/afm.1139

Balázs, J., Lecrubier, Y., Csiszér, N., Koszták, J., Bitter, I., 2003. 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 76, 113–119.

Beaton, D.E., Bombardier, C., Guillemin, F., Ferraz, M.B., 2000. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine 25, 3186–3191. https://doi.org/10.1097/00007632-200012150-00014

Brenner, H., Gefeller, O., 1997. Variation of sensitivity, specificity, likelihood ratios and predictive values with disease prevalence. Stat Med 16, 981–991.

Campo-Arias, A., Barliza, F.A., 2018. Diagnóstico y tratamiento psiquiátrico por médicos generales. Revista Colombiana de Psiquiatría S0034745018300489. https://doi.org/10.1016/j.rcp.2018.04.005

Canadian Task Force on Preventive Health Care, M. Joffres, A. Jaramillo, J. Dickinson, G. Lewin, K. Pottie, E. Shaw, S. C. Gorber, y M. Tonelli. 2013. «Recommendations on Screening for Depression in Adults». Canadian Medical Association Journal 185(9):775-82.

Cassiani-Miranda, C., Cuadros-Cruz, K., Torres Pinzón, H., Scoppetta, O., Pinzón-Tarrazona, J., López-Fuentes, W., 2020. 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 Psiqtr.

Cassiani-Miranda, Pérez-Aníbal, Universidad de Cartagena, María Camila Vargas-Hernández, Mariana Herazo-Bustos, y Adrián Cabarcas-Tovar. 2018. «Face validity and adaptation of the PHQ-9 for detection of depressive symptoms in university health sciences students Cartagena (Colombia)». Salud Uninorte 34(1):75-87.

Centers for Disease Control and Prevention, 2017. Ambulatory Care Use and Physician office visits. Summary Health Statistics Tables for U.S. Adults: National Ambulatory Medical Care Survey: 2016 National Summary Tables, table 1. CDC/National Center for Health Statistics, Washington DC.

Cerda, J., Cifuentes, L., 2012. Uso de curvas ROC en investigación clínica: Aspectos teórico-prácticos. Rev. chil. infectol. 29, 138–141. https://doi.org/10.4067/S0716-10182012000200003

Cohen, J., 1960. A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement 20, 37–46. https://doi.org/10.1177/001316446002000104

de Azevedo Marques, J.M., Zuardi, A.W., 2008. Validity and applicability of the Mini International Neuropsychiatric Interview administered by family medicine residents in primary health care in Brazil. Gen Hosp Psychiatry 30, 303–310. https://doi.org/10.1016/j.genhosppsych.2008.02.001

de Vet, Henrica C. W., Lidwine B. Mokkink, David G. Mosmuller, y Caroline B. Terwee. 2017. «Spearman–Brown Prophecy Formula and Cronbach's Alpha: Different Faces of Reliability and Opportunities for New Applications». Journal of Clinical Epidemiology 85:45-49.

Demyttenaere, K., Bruffaerts, R., Posada-Villa, J., Gasquet, I., Kovess, V., Lepine, J.P., Angermeyer, M.C., Bernert, S., de Girolamo, G., Morosini, P., Polidori, G., Kikkawa, T., Kawakami, N., Ono, Y., Takeshima, T., Uda, H., Karam, E.G., Fayyad, J.A., Karam, A.N., Mneimneh, Z.N., Medina-Mora, M.E., Borges, G., Lara, C., de Graaf, R., Ormel, J., Gureje, O., Shen, Y., Huang, Y., Zhang, M.,

Alonso, J., Haro, J.M., Vilagut, G., Bromet, E.J., Gluzman, S., Webb, C., Kessler, R.C., Merikangas, K.R., Anthony, J.C., Von Korff, M.R., Wang, P.S., Brugha, T.S., Aguilar-Gaxiola, S., Lee, S., Heeringa, S., Pennell, B.-E., Zaslavsky, A.M., Ustun, T.B., Chatterji, S., WHO World Mental Health Survey Consortium, 2004. Prevalence, severity, and unmet need for treatment of mental disorders in the World Health Organization World Mental Health Surveys. JAMA 291, 2581–2590. https://doi.org/10.1001/jama.291.21.2581

Duhoux, A., Fournier, L., Gauvin, L., Roberge, P., 2012. Quality of care for major depression and its determinants: a multilevel analysis. BMC Psychiatry 12, 142. https://doi.org/10.1186/1471-244X-12-142

El-Den, S., Chen, T.F., Gan, Y.-L., Wong, E., O'Reilly, C.L., 2018. The psychometric properties of depression screening tools in primary healthcare settings: A systematic review. J Affect Disord 225, 503–522. https://doi.org/10.1016/j.jad.2017.08.060

Ferenchick, E.K., Ramanuj, P., Pincus, H.A., 2019. Depression in primary care: part 1—screening and diagnosis. BMJ 365, l794. https://doi.org/10.1136/bmj.l794

Gelaye, B., Wilson, I., Berhane, H.Y., Deyessa, N., Bahretibeb, Y., Wondimagegn, D., Shibre Kelkile, T., Berhane, Y., Fann, J.R., Williams, M.A., 2016. Diagnostic validity of the Patient Health Questionnaire-2 (PHQ-2) among Ethiopian adults. Comprehensive Psychiatry 70, 216–221. https://doi.org/10.1016/j.comppsych.2016.07.011

Hamel C, Lang E, Morissette K, Beck A, Stevens A, Skidmore B, Colquhoun H, LeBlanc J, Moore A, Riva J, Thombs B, Colman I, Grigoriadis S, Gordon S, Potter B, Ritchie K, Robert J, Vasa P, Lauria-Horner B, Patten S, Vigod S, Hutton B, Shea B, Shanmugasegaram S, Little J, y Moher D. 2019. «Screening for Depression in Women during Pregnancy or the First Year Postpartum and in the General Adult Population: A Protocol for Two Systematic Reviews to Update a Guideline of the Canadian Task Force on Preventive Health Care». Systematic Reviews 8(1):27.

Hanley, J.A., McNeil, B.J., 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143, 29–36. https://doi.org/10.1148/radiology.143.1.7063747

Hanlon, C., Medhin, G., Selamu, M., Breuer, E., Worku, B., Hailemariam, M., Lund, C., Prince, M., Fekadu, A., 2015. Validity of brief screening questionnaires to detect depression in primary care in Ethiopia. Journal of Affective Disorders 186, 32–39. https://doi.org/10.1016/j.jad.2015.07.015

Harman, J.S., Veazie, P.J., Lyness, J.M., 2006. Primary Care Physician Office Visits for Depression by Older Americans. J Gen Intern Med 21, 926–930. https://doi.org/10.1111/j.1525-1497.2006.00497.x

Hawton, K., Casañas i Comabella, C., Haw, C., Saunders, K., 2013. Risk factors for suicide in individuals with depression: A systematic review. Journal of Affective Disorders 147, 17–28. https://doi.org/10.1016/j.jad.2013.01.004

Hinz, A., Finck, C., Gómez, Y., Daig, I., Glaesmer, H., Singer, S., 2014. Anxiety and depression in the general population in Colombia: reference values of the Hospital Anxiety and Depression Scale (HADS). Soc Psychiatry Psychiatr Epidemiol 49, 41–49. https://doi.org/10.1007/s00127-013-0714-y

Inagaki, M., Ohtsuki, T., Yonemoto, N., Kawashima, Y., Saitoh, A., Oikawa, Y., Kurosawa, M., Muramatsu, K., Furukawa, T.A., Yamada, M., 2013. 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 35, 592–597. https://doi.org/10.1016/j.genhosppsych.2013.08.001

Instituto Nacional de Medicina Legal y Ciencias Forenses, 2019. Forensis. Datos para la vida. 2018. INML, Bogotá DC.

Johnson, J.A., Al Sayah, F., Wozniak, L., Rees, S., Soprovich, A., Qiu, W., Chik, C.L., Chue, P., Florence, P., Jacquier, J., Lysak, P., Opgenorth, A., Katon, W., Majumdar, S.R., 2014. Collaborative care versus screening and follow-up for patients with diabetes and depressive symptoms: results of a primary care-based comparative effectiveness trial. Diabetes Care 37, 3220–3226. https://doi.org/10.2337/dc14-1308

Keshavarz, H., D. Fitzpatrick-Lewis, D. L. Streiner, R. Maureen, U. Ali, H. S. Shannon, y P. Raina. 2013. «Screening for Depression: A Systematic Review and Meta-Analysis». CMAJ Open 1(4):E159-67.

Kessler, R.C., Sampson, N.A., Berglund, P., Gruber, M.J., Al-Hamzawi, A., Andrade, L., Bunting, B., Demyttenaere, K., Florescu, S., de Girolamo, G., Gureje, O., He, Y., Hu, C., Huang, Y., Karam, E., Kovess-Masfety, V., Lee, S., Levinson, D., Medina Mora, M.E., Moskalewicz, J., Nakamura, Y., Navarro-Mateu, F., Browne, M.A.O., Piazza, M., Posada-Villa, J., Slade, T., Ten Have, M., Torres, Y., Vilagut, G., Xavier, M., Zarkov, Z., Shahly, V., Wilcox, M.A., 2015. Anxious and non-anxious major depressive disorder in the World Health Organization World Mental Health Surveys. Epidemiol Psychiatr Sci 24, 210–226. https://doi.org/10.1017/S2045796015000189

Keszei P., Novak M & Streiner D. 2010. «Introduction to Health Measurement Scales». Journal of Psychosomatic Research 68(4):319-23.

Kocalevent, R.-D., Hinz, A., Brähler, E., 2013. Standardization of the depression screener Patient Health Questionnaire (PHQ-9) in the general population. General Hospital Psychiatry 35, 551–555. https://doi.org/10.1016/j.genhosppsych.2013.04.006

König, H., König, H.-H., Konnopka, A., 2019. The excess costs of depression: a systematic review and meta-analysis. Epidemiol Psychiatr Sci 1–16. https://doi.org/10.1017/S2045796019000180

Kottner J, y Streiner D. 2010. «Internal Consistency and Cronbach's α: A Comment on Beeckman et al. (2010)». International Journal of Nursing Studies 47(7):926-28.

Kroenke, K., Spitzer, R.L., Williams, J., 2001. The PHQ-9: validity of a brief depression severity measure. Journal of General Internal Medicine 16, 606–13.

Kroenke, K., Spitzer, R.L., Williams, J.B.W., 2003. The Patient Health Questionnaire-2: Validity of a Two-Item Depression Screener. Medical Care 41, 1284–1292. https://doi.org/10.1097/01.MLR.0000093487.78664.3C

Lange, R.T., Lippa, S.M., 2017. Sensitivity and specificity should never be interpreted in isolation without consideration of other clinical utility metrics. Clin Neuropsychol 31, 1015–1028. https://doi.org/10.1080/13854046.2017.1335438

Lecrubier, Y., Sheehan, D., Weiller, E., Amorim, P., Bonora, I., Harnett Sheehan, K., Janavs, J., Dunbar, G., 1997. The Mini International Neuropsychiatric Interview (MINI). A short diagnostic structured interview: reliability and validity according to the CIDI. European Psychiatry 12, 224–231. https://doi.org/10.1016/S0924-9338(97)83296-8

Li, C., Friedman, B., Conwell, Y., Fiscella, K., 2007. Validity of the Patient Health Questionnaire 2 (PHQ-2) in Identifying Major Depression in Older People: Validity

of PHQ-2 in detecting depression. Journal of the American Geriatrics Society 55, 596–602. https://doi.org/10.1111/j.1532-5415.2007.01103.x

Lino, V.T.S., Portela, M.C., Camacho, L.A.B., Atie, S., Lima, M.J.B., Rodrigues, N.C.P., Barros, M.B. de L., Andrade, M.K. de N., 2014. Screening for Depression in Low-Income Elderly Patients at the Primary Care Level: Use of the Patient Health Questionnaire-2. PLoS One 9. https://doi.org/10.1371/journal.pone.0113778

Luo, Z., Cowell, A.J., Musuda, Y.J., Novak, S.P., Johnson, E.O., 2010. Course of major depressive disorder and labor market outcome disruption. J Ment Health Policy Econ 13, 135–149.

Manea, L., Gilbody, S., Hewitt, C., North, A., Plummer, F., Richardson, R., Thombs, B.D., Williams, B., McMillan, D., 2016. Identifying depression with the PHQ-2: A diagnostic meta-analysis. J Affect Disord 203, 382–395. https://doi.org/10.1016/j.jad.2016.06.003

Maurer, D 2012. «Screening for Depression». American Family Physician 85(2):139-44.

Ministry of Health, 1993. Por la cual se establecen las normas científicas, técnicas y administrativas para la investigación en salud.

Mitchell, A.J., Coyne, J.C., 2007. Do ultra-short screening instruments accurately detect depression in primary care? Br J Gen Pract 57, 144–151.

Mitchell, A.J., Vahabzadeh, A., Magruder, K., 2011. Screening for distress and depression in cancer settings: 10 lessons from 40 years of primary-care research. Psychooncology 20, 572–584. https://doi.org/10.1002/pon.1943

Mitchell, A.J., Vaze, A., Rao, S., 2009. Clinical diagnosis of depression in primary care: a meta-analysis. Lancet 374, 609–619. https://doi.org/10.1016/S0140-6736(09)60879-5

Mitchell, A.J., Yadegarfar, M., Gill, J., Stubbs, B., 2016. 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. BJPsych Open 2, 127–138. https://doi.org/10.1192/bjpo.bp.115.001685

Moorey, S., Greer, S., Watson, M., Gorman, C., Rowden, L., Tunmore, R., Robertson, B., Bliss, J., 1991. The factor structure and factor stability of the hospital anxiety and depression scale in patients with cancer. Br J Psychiatry 158, 255–259. https://doi.org/10.1192/bjp.158.2.255

Moratalla Rodríguez, G., 2015. Lectura crítica de artículos de pruebas diagnósticas II: análisis de resultados. Radiología 57, 22–28. https://doi.org/10.1016/j.rx.2014.11.004

Mordal, J., Gundersen, O., Bramness, J.G., 2010. Norwegian version of the Mini-International Neuropsychiatric Interview: feasibility, acceptability and test-retest reliability in an acute psychiatric ward. Eur. Psychiatry 25, 172–177. https://doi.org/10.1016/j.eurpsy.2009.02.004

Moriarty, A.S., Gilbody, S., McMillan, D., Manea, L., 2015. Screening and case finding for major depressive disorder using the Patient Health Questionnaire (PHQ-9): a meta-analysis. Gen Hosp Psychiatry 37, 567–576. https://doi.org/10.1016/j.genhosppsych.2015.06.012

Mulvaney-Day N, Marshall T, Downey K, Korsen N, Lynch S, Karnell L, Moran G, Daniels A, & Shoma, S. 2018. «Screening for Behavioral Health Conditions in Primary Care Settings: A Systematic Review of the Literature». Journal of General Internal Medicine 33(3):335-46.

Nakku, J.E.M., Rathod, S.D., Kizza, D., Breuer, E., Mutyaba, K., Baron, E.C., Ssebunnya, J., Kigozi, F., 2016. 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. Health 3, e20. https://doi.org/10.1017/gmh.2016.14

Ortiz-Gutiérrez, S., y A. Cruz-Avelar. 2018. «Proceso de traducción y adaptación cultural de instrumentos de medición en salud». Actas Dermo-Sifiliográficas 109(3):202-6.

Otsubo, T., Tanaka, K., Koda, R., Shinoda, J., Sano, N., Tanaka, S., Aoyama, H., Mimura, M., Kamijima, K., 2005. Reliability and validity of Japanese version of the Mini-International Neuropsychiatric Interview. Psychiatry Clin. Neurosci. 59, 517–526. https://doi.org/10.1111/j.1440-1819.2005.01408.x

Patel, V., Araya, R., Chowdhary, N., King, M., Kirkwood, B., Nayak, S., Simon, G., Weiss, H.A., 2008. Detecting common mental disorders in primary care in India: a comparison of five screening questionnaires. Psychol Med 38, 221–228. https://doi.org/10.1017/S0033291707002334

Patel, V., Sartorius, N., 2008. From science to action: the Lancet Series on Global Mental Health. Curr Opin Psychiatry 21, 109–113. https://doi.org/10.1097/YCO.0b013e3282f43c7f

Pettersson, A., Boström, K.B., Gustavsson, P., Ekselius, L., 2015. Which instruments to support diagnosis of depression have sufficient accuracy? A systematic review. Nord J Psychiatry 69, 497–508. https://doi.org/10.3109/08039488.2015.1008568

Pinninti, N.R., Madison, H., Musser, E., Rissmiller, D., 2003. MINI International Neuropsychiatric Schedule: clinical utility and patient acceptance. Eur. Psychiatry 18, 361–364.

Phelan, Elizabeth, Barbara Williams, Kathryn Meeker, Katie Bonn, John Frederick, James LoGerfo, y Mark Snowden. 2010. «A Study of the Diagnostic Accuracy of the PHQ-9 in Primary Care Elderly». BMC Family Practice 11(1):63.

Post, M.W., 2016. What to Do With "Moderate" Reliability and Validity Coefficients? Archives of Physical Medicine and Rehabilitation 97, 1051–1052. https://doi.org/10.1016/j.apmr.2016.04.001

Ramada-Rodilla, J.M., Serra-Pujadas, C., Delclós-Clanchet, G.L., 2013. [Cross-cultural adaptation and health questionnaires validation: revision and methodological recommendations]. Salud Publica Mex 55, 57–66.

Rancans, Elmars, Marcis Trapencieris, Rolands Ivanovs, y Jelena Vrublevska. 2018. «Validity of the PHQ-9 and PHQ-2 to Screen for Depression in Nationwide Primary Care Population in Latvia». Annals of General Psychiatry 17(1):33.

Richardson, L.P., Rockhill, C., Russo, J.E., Grossman, D.C., Richards, J., McCarty, C., McCauley, E., Katon, W., 2010. Evaluation of the PHQ-2 as a Brief Screen for Detecting Major Depression Among Adolescents. PEDIATRICS 125, e1097–e1103. https://doi.org/10.1542/peds.2009-2712

Rico, J., Forero, M.R., Molina, M., 2005. 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 3, 73–86.

Rodríguez-Muñoz, M., Castelao, P., Olivares, M., Soto, C., Izquierdo, N., Ferrer, F., Huynh-Nhu, L., 2017. PHQ-2 como primer instrumento de cribado de la depresión prenatal. Rev Esp Salud Pública 91, 1–8.

Romera, I., Montejo, Á.L., Aragonés, E., Arbesú, J.Á., Iglesias-García, C., López, S., Lozano, J.A., Pamulapati, S., Yruretagoyena, B., Gilaberte, I., 2013. Systematic depression screening in high-risk patients attending primary care: a pragmatic cluster-randomized trial. BMC Psychiatry 13, 83. https://doi.org/10.1186/1471-244X-13-83

Rosselli, D., Ardila, A., Pradilla, G., Morillo, L., Bautista, L., Rey, O., Camacho, L., 2000. The Mini-Mental State Examination as a selected diagnostic test for dementia: a Colombian population study. GENECO. Revista de Neurología 30, 428–432.

Rotenstein, L.S., Ramos, M.A., Torre, M., Segal, J.B., Peluso, M.J., Guille, C., Sen, S., Mata, D.A., 2016. Prevalence of Depression, Depressive Symptoms, and Suicidal Ideation Among Medical Students: A Systematic Review and Meta-Analysis. JAMA 316, 2214. https://doi.org/10.1001/jama.2016.17324

Saldivia S, Aslan J, Cova F, Vicente B, Inostroza C, y Rincón P. 2019. «Propiedades Psicométricas Del PHQ-9 (Patient Health Questionnaire) En Centros de Atención Primaria de Chile». Revista Médica de Chile 147(1):53-60.

Sánchez, R., Echeverry, J., 2001. Aspectos sobre diseño y tamaño de muestra en estudios de pruebas diagnósticas. Rev Fac Med UN Col 49, 175–180.

Shah, A., 1992. The burden of psychiatric disorder in primary care. International Review of Psychiatry 4, 243–250. https://doi.org/10.3109/09540269209066324

Sheehan, D.V., Lecrubier, Y., Sheehan, K.H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., Dunbar, G.C., 1998. 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 59 Suppl 20, 22-33;quiz 34-57.

Siu, Albert L., and the US Preventive Services Task Force (USPSTF), Bibbins-Domingo K, Grossman D, Ciofu L, Davidson K, Ebell M, García F, Gillman M, Herzstein J, Kemper A, Krist A, Kurth A, Owens D, Phillips W, Phipps M, y Pignone M. 2016. «Screening for Depression in Adults: US Preventive Services Task Force Recommendation Statement». JAMA 315(4):380.

Smithson, Sarah, y Michael P. Pignone. 2017. «Screening Adults for Depression in Primary Care». Medical Clinics of North America 101(4):807-21.

Snaith, R., 1992. Availability of the hospital anxiety and depression (HAD) scale. Br J Psychiatry 161–422.

Spijker, J., de Graaf, R., Bijl, R.V., Beekman, A.T.F., Ormel, J., Nolen, W.A., 2002. Duration of major depressive episodes in the general population: results from The Netherlands Mental Health Survey and Incidence Study (NEMESIS). Br J Psychiatry 181, 208–213. https://doi.org/10.1192/bjp.181.3.208

Spitzer, R.L., Kroenke, K., Williams, J.B., 1999. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. JAMA 282, 1737–1744.

Spitzer, R.L., Williams, J.B., Kroenke, K., Hornyak, R., McMurray, J., 2000. 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. 183, 759–769.

Staples, Lauren G., Blake F. Dear, Milena Gandy, Vincent Fogliati, Rhiannon Fogliati, Eyal Karin, Olav Nielssen, y Nickolai Titov. 2019. «Psychometric Properties and

Clinical Utility of Brief Measures of Depression, Anxiety, and General Distress: The PHQ-2, GAD-2, and K-6». General Hospital Psychiatry 56:13-18.

Steel, N., Ford, J.A., Newton, J.N., Davis, A.C.J., Vos, T., Naghavi, M., Glenn, S., Hughes, A., Dalton, A.M., Stockton, D., Humphreys, C., Dallat, M., Schmidt, J., Flowers, J., Fox, S., Abubakar, I., Aldridge, R.W., Baker, A., Brayne, C., Brugha, T., Capewell, S., Car, J., Cooper, C., Ezzati, M., Fitzpatrick, J., Greaves, F., Hay, R., Hay, S., Kee, F., Larson, H.J., Lyons, R.A., Majeed, A., McKee, M., Rawaf, S., Rutter, H., Saxena, S., Sheikh, A., Smeeth, L., Viner, R.M., Vollset, S.E., Williams, H.C., Wolfe, C., Woolf, A., Murray, C.J.L., 2018. 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. The Lancet 392, 1647–1661. https://doi.org/10.1016/S0140-6736(18)32207-4

Takayanagi, Y., Spira, A.P., Bienvenu, O.J., Hock, R.S., Carras, M.C., Eaton, W.W., Mojtabai, R., 2015. Antidepressant Use and Lifetime History of Mental Disorders in a Community Sample: Results From the Baltimore Epidemiologic Catchment Area Study. J. Clin. Psychiatry 76, 40–44. https://doi.org/10.4088/JCP.13m08824

Tejada, P.A., Jaramillo, L.E., Sánchez-Pedraza, R., 2014. Revisión crítica sobre los instrumentos para la evaluación psiquiátrica en atención primaria. Revista de la Facultad de Medicina 62, 101–110. https://doi.org/10.15446/revfacmed.v62n1.43759

Thombs, B.D., Saadat, N., Riehm, K.E., Karter, J.M., Vaswani, A., Andrews, B.K., Simons, P., Cosgrove, L., 2017. Consistency and sources of divergence in recommendations on screening with questionnaires for presently experienced health problems or symptoms: a comparison of recommendations from the Canadian Task Force on Preventive Health Care, UK National Screening Committee, and US Preventive Services Task Force. BMC Med 15, 150. https://doi.org/10.1186/s12916-017-0903-8

Thombs B, Ziegelstein C, Roseman M, Kloda L, y Ioannidis J. 2014. «There Are No Randomized Controlled Trials That Support the United States Preventive Services Task Force Guideline on Screening for Depression in Primary Care: A Systematic Review». BMC Medicine 12(1):13.

UK National Screening Committee. s. f. «Current UK NSC recommendations». Recuperado 21 de abril de 2020 (https://legacyscreening.phe.org.uk/screening-recommendations.php).

Ventura-León, J., Caycho-Rodríguez, T., 2017. El coeficiente Omega: un método alternativo para la estimación de la confiabilidad. Revista Latinoamericana de Ciencias Sociales, Niñez y Juventud 15, 625–627.

Wang, J., Wu, X., Lai, W., Long, E., Zhang, X., Li, W., Zhu, Y., Chen, C., Zhong, X., Liu, Z., Wang, D., Lin, H., 2017. Prevalence of depression and depressive symptoms among outpatients: a systematic review and meta-analysis. BMJ Open 7, e017173. https://doi.org/10.1136/bmjopen-2017-017173

Watkins, M.W., 2017. The reliability of multidimensional neuropsychological measures: from alpha to omega. The Clinical Neuropsychologist 31, 1113–1126. https://doi.org/10.1080/13854046.2017.1317364

Whiting, P.F., Rutjes, A., Westwood, M., Mallett, S., Deeks, J., Reitsma, J., Leeflang, M., Sterne, J., Bossouyt, P., QUADAS-2 Group, 2011. QUADAS-2: A Revised Tool for

the Quality Assessment of Diagnostic Accuracy Studies. Ann Intern Med 155, 529. https://doi.org/10.7326/0003-4819-155-8-201110180-00009

World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects, 2013. . JAMA 310, 2191. https://doi.org/10.1001/jama.2013.281053

Youden, W.J., 1950. Index for rating diagnostic tests. Cancer 3, 32–35. https://doi.org/10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3

Yu, Xiaonan, Sunita M. Stewart, Paul T. K. Wong, y Tai Hing Lam. 2011. «Screening for Depression with the Patient Health Questionnaire-2 (PHQ-2) among the General Population in Hong Kong». Journal of Affective Disorders 134(1-3):444-47.

Zastrow, A., Faude, V., Seyboth, F., Niehoff, D., Herzog, W., Löwe, B., 2008. Risk factors of symptom underestimation by physicians. J Psychosom Res 64, 543–551. https://doi.org/10.1016/j.jpsychores.2007.11.010

Zigmond, A.S., Snaith, R.P., 1983. The hospital anxiety and depression scale. Acta Psychiatr Scand 67, 361–370.

Zuithoff, Nicolaas PA, Yvonne Vergouwe, Michael King, Irwin Nazareth, Manja J. van Wezep, Karel GM Moons, y Mirjam I. Geerlings. 2010. «The Patient Health Questionnaire-9 for Detection of Major Depressive Disorder in Primary Care: Consequences of Current Thresholds in a Crosssectional Study». BMC Family Practice 11(1):98.

Table 1. Sociodemographic characteristics of participants.

	N (%)
Sex	
Male	59 (24,3)
Female	184 (75,7)
Place of origin	
Urban	210 (86,4)
Rural	33 (13,6)
Marital Status	
Single	98 (40,3)
Married	53 (21,9)
Civil Partnership	83 (34,2)
Divorced	5 (2,0)
Widower	4 (1,6)
Schooling	
Incomplete Elementary School Education	14 (5,7)
Complete Elementary School Education	60 (24,7)
Incomplete High School Education	36 (14,8)
Complete High School Education	88 (36,2)
Incomplete Technical Education	18 (7,4)
Complete Technical Education	1 (0,4)
Complete Technological Education	3 (1,3)
Incomplete Undergraduate Studies	3 (1,3)
Complete Undergraduate Studies	20 (8,2)
Socioeconomic strata	
1	103 (42,4)
2	84 (34,6)
3	47 (19,3)
4	6 (2,5)

5	3 (1,2)

Table 2. Description of the different cut-off points of the PHQ-2 (Colombian version) and validity coefficients

Cut-off Point	Sensitivity	Specificity	Youden Index	PV+	PV-	LR+	LR-
≥1	0,98	0,51	0,49	0,35	0,99	1,99	0,04
≥2	**0,87**	**0,73**	**0,60**	**0,46**	**0,95**	**3,24**	**0,18**
≥3	0,64	0,91	0,55	0,65	0,90	7,13	0,40
≥4	0,48	0,97	0,45	0,80	0,88	15,30	0,54
≥5	0,29	0,98	0,27	0,83	0,84	18,37	0,72
≥6	0,23	1,00	0,23	0,92	0,83	44,08	0,77