BMJ Open

Sensitivity and specificity of the Patient Health Questionnaire (PHQ-9, PHQ-8, PHQ-2) and General Anxiety Disorder scale (GAD-7, GAD-2) for depression and anxiety diagnosis: A cross-sectional study in a Peruvian hospital population

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2023-076193
Article Type:	Original research
Date Submitted by the Author:	01-Jun-2023
Complete List of Authors:	Villarreal-Zegarra, David; Instituto Peruano de Orientación Psicológica, ; Universidad Peruana Cayetano Heredia, CRONICAS Center of Excellence in Chronic Diseases Barrera, Juan ; Instituto Peruano de Orientación Psicológica Otazú-Alfaro, Sharlyn; Instituto Peruano de Orientación Psicológica Mayo-Puchoc, Nikol; Instituto Peruano de Orientación Psicológica Bazo-Alvarez, Juan Carlos ; Universidad Norbert Wiener Huarcaya-Victoria , Jeff; Universidad de San Martín de Porres Facultad de Medicina Humana,
Keywords:	Sensitivity and Specificity, Depression & mood disorders < PSYCHIATRY, Anxiety disorders < PSYCHIATRY

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Sensitivity and specificity of the Patient Health Questionnaire (PHQ-9, PHQ-8, PHQ-2) and General Anxiety Disorder scale (GAD-7, GAD-2) for depression and anxiety diagnosis: A cross-sectional study in a Peruvian hospital population

Authors

David Villarreal-Zegarra [1, 2]
Juan Barrera [2]
Sharlyn Otazú-Alfaro [2]
Nikol Mayo-Puchoc [2]
Juan Carlos Bazo-Alvarez [3]
Jeff Huarcaya-Victoria [4,5]

Affiliation

[1] Universidad César Vallejo, Escuela de Medicina. Trujillo, Peru
[2] Instituto Peruano de Orientación Psicológica, Lima, Peru
[3] Universidad Privada Norbert Wiener, Lima, Peru
[4] Escuela Profesional de Medicina Humana, Universidad Privada San Juan Bautista, Filial Ica, Peru
[5] Unidad de Psiquiatría de Enlace, Departamento de Psiquiatría, Hospital Nacional Guillermo Almenara Irigoyen, EsSalud, Lima, Peru

Email

David Villarreal-Zegarra davidvillarreal@ipops.pe
Juan Barrera juanbarrera@ipops.pe
Sharlyn Otazú-Alfaro sharlynotazu@ipops.pe
Nikol Mayo-Puchoc nikolmayo@ipops.pe
Juan Carlos Bazo-Alvarez juan.bazo@uwiener.edu.pe
Jeff Huarcaya-Victoria jhuarcayav@usmp.pe

ORCID

David Villarreal-Zegarra https://orcid.org/0000-0002-2222-4764

Juan Barrera https://orcid.org/0000-0002-6641-6266
Sharlyn Otazú-Alfaro https://orcid.org/0000-0002-6462-8716
Nikol Mayo-Puchoc https://orcid.org/0000-0002-6182-7605
Juan Carlos Bazo-Alvarez https://orcid.org/0000-0002-6169-8049
Jeff Huarcaya-Victoria https://orcid.org/0000-0003-4525-9545

Corresponding author

Jeff Huarcaya-Victoria
jeff.huarcaya@upsjb.edu.pe

Declarations

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Availability of data and materials

The database can be requested from the corresponding author.

Ethics approval and consent to participate

The Hospital Nacional Guillermo Almenara Irigoyen's Institutional Review Board (Nota N°52 CIEI-OIyD-GRPA-Essalud-2023) approved the original study and the current analysis. Throughout the study, the researchers did not have access to any identifying information about the participants. In addition, participants gave informed consent. All participants were users of the hospital's Liaison Psychiatry Unit, so all received psychological or psychiatric care as needed.

Consent for publication

Not applicable.

Authors' contributions

David Villarreal-Zegarra: Conceptualization, Methodology, Software, Validation, Formal analysis, Data Curation, Writing - Original Draft, Writing - Original Draft, Visualization.
Juan Barrera: Formal analysis, Investigation, Visualization.
Sharlyn Otazú-Alfaro: Investigation, Writing - Original Draft.
Nikol Mayo-Puchoc: Formal analysis, Investigation, Writing - Original Draft.
Juan Carlos Bazo: Methodology, Investigation, Writing - Review & Editing, Supervision.
Jeff Huarcaya-Victoria: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing - Review & Editing, Project administration.

Sensitivity and specificity of the Patient Health Questionnaire (PHQ-9, PHQ-8, PHQ-2) and General Anxiety Disorder scale (GAD-7, GAD-2) for depression and anxiety diagnosis: A cross-sectional study in a Peruvian population

Abstract

Background: The Patient Health Questionnaire (PHQ) and Generalized Anxiety Disorder Scale (GAD) are widely used screening tools, but their sensitivity and specificity in low- and middle-income countries are lower than in high-income countries. We conducted a study to determine the sensitivity and specificity of different versions of these scales in a Peruvian hospital population.

Method: Our study has a cross-sectional design. Our participants are hospitalised patients in a Peruvian hospital. The gold standard was a clinical psychiatric interview following ICD-10 criteria for depression (F32.0, F32.1, F32.2, and F32.3) and anxiety (F41.0 and F41.1).

Results: We found that a cut-off point of ≥7 in the PHQ-9 and PHQ-8 had the best balance between sensitivity and specificity. The raw score for PHQ-9 with cohort point ≥7 showed better results than the algorithm method or the adjusted algorithm. A cut-off point of ≥8 had the best performance for GAD-7. The PHQ-2 and GAD-2 had the best cut-off point of ≥2 points. All scales had good internal consistency (>0.70), and the PHQ-9, PHQ-8, and GAD-7 had appropriate goodness-of-fit indices.

Limitations: Our study is non-probabilistic.

Conclusions: The PHQ and GAD have adequate measurement properties in their different versions. We present specific cut-offs for each version.

Keywords: Depression; Anxiety; Patient Health Questionnaire; Sensitivity and Specificity; Peru.

Background

Until 2019, approximately 280 million people worldwide suffered from depression and 302 million from anxiety [1]. These data reveal that both mental disorders are the most common in the world and lead to the causes of the global burden of mental health disability-adjusted life years[2, 3]. With the onset of the COVID-19 pandemic, the worldwide prevalence of both disorders increased by around 25% [4]. In Peru, during the COVID-19 pandemic, the prevalence of moderate depressive symptoms also increased by approximately 0.17% in each quarter [5]. However, no population-level evidence has been found about the prevalence of anxious symptomatology or the diagnosis of anxiety in Peru. In this context, the impact of COVID-19 on the prevalence and burden of major depression and anxiety disorders was measured using screening tools [6].

Screening tools assist in early diagnosis and intervention that can prevent disease progression and reduce years lost to disability [7]. They are beneficial in contexts with limited mental health professionals providing care to large populations, such as in Peru. The opportune identification of people at risk of depression reduces treatment costs and disease burden [8-10]. Depressive symptom screening is also helpful in national surveys and epidemiological research [11] since, unlike diagnostic instruments, screening measures are typically brief, quick, and easy to administer [12, 13]. Internationally, the most used screening instruments for depressive and anxious symptomatology are the PHQ-9 [14], PHQ-8 [15], PHQ-2 [16], GAD-7 [17], GAD-2 [17], DASS-21, Kessler-10, HADS-A [18], HADS, WHO-5 [9]. Most have been validated in several countries, but only the PHQ and GAD have been validated in the Peruvian context [19, 20].

In particular, the PHQ versions (PHQ-9, PHQ-8, PHQ-2) and GAD versions (GAD-7, GAD-2) are the most widely used, having extensive evidence of their validity and reliability [21-23]. However, correctly identifying people at risk of depression or anxiety requires more than internal/externally valid and reliable screening measures; defining an accurate cut-off point for their raw scales (i.e., to reach valid interpretations) is also necessary. Such a cut-off point can vary across cultures and sub-populations (e.g., general versus clinical), so a local calibration is usually needed [24]. Studies of the different versions of the PHQ and GAD have yielded heterogeneous cut-offs, as they vary between different cultures[20, 25-28] and populations, such as clinical and general populations[29-31]. However, several systematic reviews suggest that cut-off 10 is most appropriate for the PHQ-9, PHQ-8, and GAD-7 [32-36], and cut-offs 2-3 for the PHQ-2 and GAD-2 [34, 36]. Furthermore, concerning the PHQ-9 correctness, the summed item score method is the most used compared to the algorithm. However, other forms of correction using diagnostic algorithms are available [37, 38].

Sensitivity and specificity studies have been barely performed in low- and middle-income countries [39]. Several of these populations do not count with verified cut-off points from calibration studies (including Peruvian populations). determine the optimal cut-off point for the PHQ-9, PHQ-8, PHQ-2, GAD-7, and GAD-2 to discriminate a formal depression and anxiety diagnosis in the Peruvian hospital population. In addition, as secondary objectives, we assessed these scales' internal structure and reliability.

Methods

Study design

This study has a cross-sectional design, and we used the STARD 2015 guidelines for reporting diagnostic accuracy studies [40].

Participants

The participants were patients from the Liaison Psychiatry Unit of a hospital in Lima, Peru. Psychiatric liaison services provide psychiatric consultation to hospitalised patients with medical or surgical conditions that have a co-existing psychiatric illness or need for psychiatric assessment and management. The total number of participants in our study is similar to the proportion of people who were hospitalised in 2022 in our setting (see Supplementary Appendix 1). The evaluation period started in September 2020 and finished in August 2022. Sampling was non-probabilistic and applied to all participants arriving at the Liaison Psychiatry Unit. The inclusion criteria were that they had complete PHQ-9 and GAD-7 data and were of legal age (>18 years). Participants with missing data were excluded.

The sample size calculation for the PHQ versions was based on an estimated sensitivity of 0.88 and specificity of 0.85 [32], a confidence level of 95%, a prevalence of 6.4% [41, 42], and a drop-out rate of 10%, giving an estimate of 705 participants. The sample size calculation for the GAD versions was based on an estimated sensitivity of 0.83 and specificity of 0.84 [17], a confidence level of 95%, a prevalence of 8.7% [43], and a drop-out rate of 10%, giving an estimate of 694 participants. The web program based on the paper by Burderer (1996) was used to calculate the sample size. [44].

Setting

The Guillermo Almenara Irigoyen National Hospital (HNGAI) was the study site, a highly complex hospital in Lima-Peru (capital city). HNGAI is one of the three largest hospitals of the Social Security system in Peru based on the number of beds (960 hospital beds) and is also a tertiary referral centre for all medical specialities, including psychiatry (http://www.essalud.gob.pe/estadistica-institucional/). It provides health care services to 1,547,840 individuals from social insurance. Because it attends to virtually all pathologies, from the simplest to the most complex, it was classified in 2015 as a Specialized Health Institute III-2, the highest level awarded by the Ministry of Health of Peru to hospital establishments.

The Liaison Psychiatry Unit at HNGAI is responsible for responding to consultation requests from different clinical-surgical services at HNGAI [45]. As part of the evaluation of each patient, in addition to the clinical interview and psychiatric diagnosis, standardised assessments such as the PHQ-9 and GAD-7 are used to ensure adequate monitoring and assess response to the established treatment. Since September 2020, the services provided by the Liaison Unit have been recorded in a Google Form to track better the patients treated.

Instruments and variables

Patient Health Questionnaire (PHQ-9, PHQ-8, and PHQ-2)

The Patient Health Questionnaire is an instrument designed to measure depressive symptoms over the past two weeks, according to the diagnostic criteria of the DSM-IV, criteria that were retained in the DSM-5. The scale has four response options (0=no days, 1=some days, 2=more than half of the days, 3=almost every day) [14]. The scale had many versions, including the PHQ-9, the full version with nine items and scores ranging from 0 to 27. In Peru, the PHQ-9 had good psychometric properties in terms of structural validity (CFI = 0.936; RMSEA = 0.089; SRMR = 0.039), internal consistency (α = ω = 0.87) and invariance between age and sex (ΔCFI<0.01) [19].

In addition, PHQ-9 had scoring versions related to the DSMV-5 indicators, which state that for a case to be positive, there must be at least five depressive symptoms present, and at least one of them must be core depressive symptoms (item 1 and item 2). First, the PHQ-9 algorithm suggests that a symptom is positive if it scores two or more, except the ninth item, suicidal ideation, which is positive if it scores one or more [46]. Second, the PHQ-9 adjusted algorithm proposes that a symptom was positive if it scored 1 or more for any of the items in the instrument [47].

The PHQ-8 was a shortened version of the PHQ-9 without the last item on suicidal ideation [15]. The PHQ-8 was as valuable as the PHQ-9 in detecting cases of major depression [48]. The PHQ-2 is an abbreviated version of the PHQ-9 with only two items, focusing on the first two items related to the core symptoms of depression (anhedonia and depressed mood) and providing scores between 0 and 6. The PHQ-2 was validated in Peru and showed adequate levels of internal consistency (α = .80) [49].

General Anxiety Disorder Scale (GAD-7 and GAD-2)

The General Anxiety Disorder Scale was a Likert-type rating scale with four response options ranging from 0 (not at all) to 3 (almost every day), based on DSM-IV criteria and assesses anxious symptoms during the past two weeks [50]. The GAD-7 was the version of the instrument with the original seven items and had a range of scores from 0 to 21. The GAD-7 had good psychometric properties in the Peruvian context for a one-dimensional model (CFI = .995, TLI = .992, RMSEA = .056), adequate internal consistency (ω =.89), and invariance according to sex (ΔCFI ≤ .01) [51].

The GAD-2 was adapted from the GAD-7, focusing on the emotional and cognitive expressions of DSM-IV anxiety (items 1 and 2) [52]. The GAD-2 shows good internal consistency values (ω =.80) and a relationship with its extended version (r>0.80) in Peruvian context [51].

Gold standard

The gold standard was an individual clinical psychiatric interview following ICD-10 criteria. The clinical assessments were performed by psychiatrists who are members of the Liaison Psychiatry Unit, all of whom have at least five years of clinical experience evaluating the psychiatric needs of

hospitalised patients. The interview focused on assessing whether the participants had depressive disorder (F32.0, F32.1, F32.2, and F32.3) or anxiety disorder (F41.0 and F41.1).

Sociodemographic covariates

Data was collected on sex (male, female), age, marital status (Single, Married/Cohabitant, Separated, Widowed), educational level (None, Elementary, High School, Technical, College), currently works (No, yes, retired), living alone (Yes, No), and history of psychiatric diagnosis (yes/no).

Statistical analysis

The sociodemographic covariates of the participants were described at frequency and percentage levels. The internal consistency and internal structure analyses were performed with R Studio, with the "Lavaan", "Semtools", and "Semplot" packages (see supplementary material 2). Sensitivity, specificity, and correlation analyses were analysed with Stata 15 (see supplementary material 3).

Sensibility and Specificity

The PHQ-9, PHQ-8, PHQ-9 algorithm, PHQ-9 adjusted algorithm, and PHQ-2 were evaluated as diagnostic tests and compared against the gold standard. In addition, the GAD-7 and GAD-2 were scored and compared against the diagnosis of anxiety through the clinical interview (gold standard).

We used receiver operating characteristics (ROC) curves to determine which cut-off had the best sensitivity and specificity for each scale. Also, we calculated the area under the curve (AUC), positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (+LR), negative likelihood ratio (-LR), and Youden Index. The AUC represents the predictive ability of the instrument; the higher the AUC values, the greater the predictive ability of the scale. A scale with an AUC of 1 has a predictive capacity of 100%, and an AUC of 0.5 is very low. The AUC for different cut-off points was compared using the nonparametric analysis described by Hanley & McNeil [53]. PPV and NPV refer to the proportion of patients correctly diagnosed as positive or negative, respectively [54]. The LR+ is the probability that a person with the disease will test positive given the probability that a person without the disease will test positive [55]. While the LR- is the probability that a person with the disease will test negative given the probability that a person without the disease will test negative [55]. The Youden Index is a measure that summarizes the performance of a diagnostic test by interpreting it as the probability that the selected cut-off point provides an adequate clinical decision (in terms of sensitivity and specificity), as opposed to the probability that the selected cut-off provides a random decision [54]. The maximum value of the Youden Index was used as a criterion to select the cut-off with the best diagnostic performance for each scale. Values closer to 1 were considered optimal, and those closer to 0 were considered inadequate.

Internal structure

Confirmatory factor analysis (CFA) was performed considering a one-dimensional model for the PHQ-9, PHQ-8, and GAD-7. We used the weighted least square mean and variance adjusted (WLSMV) estimator [56] and polychoric matrices as it best fits the categorical-ordinal nature of the data [57].

Models were evaluated using a set of goodness-of-fit indices such as CFI and TLI, which must be greater than 0.95 to be considered adequate [58]. In addition, the SRMR and RMSEA at 90% confidence were estimated, which must have values less than 0.08 to be considered adequate [58]. It was impossible to perform a CFA for the PHQ-2 and the GAD-2 because a minimum of three items are required for such analysis.

Internal consistency

We calculated the alpha (α) and McDonald's omega coefficients (ω). Values greater than 0.70 are considered adequate [59].

Ethics

The Hospital Nacional Guillermo Almenara Irigoyen's Institutional Review Board (Nota N°52 CIEI-OIyD-GRPA-Essalud-2023) approved the protocol of our study. Throughout the study, the researchers had no access to identifying information about the participants. In addition, participants gave informed consent. All participants were users of the hospital's Liaison Psychiatry Unit and received psychological or psychiatric care as needed.

Results

Participants

We collected data from 4979 attendances performed within the liaison psychiatry service during the study period. However, some of these attendances were not assessed with PHQ-9 or GAD-7 data or lacked sociodemographic information and were eliminated (see supplementary material 4). Thus, our study only included 1347 participants (see Table 1). Most participants were female (59.4%; n=800), married or living with a partner (57.0%; n=768), and had higher technical or university education (53.5%; n=721).

Sensibility and Specificity

In supplementary material 5, we provide the values of all cut-off points for the different versions of the PHQ. The cut-off points ≥7 in the PHQ-9 had the best balance between sensitivity and specificity of all the cut-off points evaluated in the various versions of the PHQ, as it obtained a Youden Index of 48.1, the sensitivity of 76.0 (95%CI: 71.1 - 80.5) and specificity of 72.1 (95%CI: 69.2 - 74.8) (see Table 2). In addition, the PHQ-9 with a cut-off of ≥10 points (i.e., the most used) showed lower levels of sensitivity (54.2; 95%CI: 8.7 - 59.6), but higher level of specificity (87.4; 95%CI: 85.2 - 89.3), compared to the cut-off point of ≥7.

The algorithm score method for PHQ-9 had low levels of sensitivity (34.7; 95%CI: 29.6 - 40.1) but high levels of specificity (93.4; 95%CI: 91.7 - 94.8) compared to the raw score method for PHQ-9 with ≥7 cohort points. In contrast, the adjusted algorithm method for PHQ-9 showed slightly higher sensitivity values (78.1; 95%CI: 73.3 - 82.5) and better specificity values (66.4; 95%CI: 63.4 - 69.3) compared to the raw score method for PHQ-9 with ≥7 cohort points. The raw score for PHQ-9 with cohort point ≥7

showed a better balance between sensitivity and specificity (higher Youden index) compared to the algorithm method or the algorithm adjusted for PHQ-9.

The best cut-off point found in the PHQ-8 was ≥7 points, as it had a Youden Index of 46.0, a sensitivity of 79.9 (95%CI: 75.2 - 84.1), and a specificity of 66.0 (95%CI: 63.0 - 69.0) (see Table 2). The best cut-off point found in the PHQ-2 was ≥2 points, as it had a Youden Index of 40.6, a sensitivity of 84.7 (95%CI: 80.4 - 88.4), and a specificity of 55.9 (95%CI: 52.8 - 59.0) (see Table 2).

compares the AUC of the different versions of the PHQ. PHQ-9 (AUC=0.805; 95%CI: 0.779-0.831) and PHQ-8 (AUC=0.802; 95%CI: 0.776-0.828) had the highest AUC values. The PHQ-9 algorithms (AUC=0.641; 95%CI: 0.614-0.667) and PHQ-9 adjusted algorithms (AUC=0.723; 95%CI: 0.696-0.749) performed worse in AUC values compared to the PHQ-9, PHQ-8 and PHQ-2.

In supplementary material 6, we present the values of all cut-off points for the different versions of the GAD. The cut-off point ≥8 had the best performance for GAD-7 among all the cut-off points evaluated, with sensitivity values of 53.6 (95%CI: 33.9 - 72.5), specificity of 78.8 (95%CI: 76.5 - 81.0), and Youden Index 32.4 (see Table 2). The GAD-7's cut-off point ≥10 (i.e., the most used) had lower levels of sensitivity (39.3; 95%CI: 21.5 - 59.4), but higher levels of specificity (88.4; 95%CI: 86.5 - 90.1, compared to the cut-off point of ≥8.

The best cut-off point for the GAD-2 was ≥2, as it had a Youden Index of 32.2, a sensitivity of 84.7 (95%CI: 80.4 - 88.4), and a specificity of 50.1 (95%CI: 47.4 - 52.8) (see Table 2). The value of the area under the curve of GAD-7 (AUC=0.718; 95%:0.622-0.814) was higher than that of GAD-2 (AUC=0.685; 95%CI:0.587-0.7823) (see Figure 2).

Internal structure

The PHQ-9 one-dimensional model showed adequate goodness-of-fit (X^2=251.9; df=27; CFI=0.974; TLI=0.965; SRMR=0.051; RMSEA[90%CI]=0.079[0.070-0.088]), while the PHQ-8 one-dimensional model reported a similar goodness-of-fit (X^2=202.7; df=20; CFI=0.977; TLI=0.977; SRMR=0.050; RMSEA[90%CI]=0.082[0.072-0.093]). The GAD-7 also showed adequate goodness-of-fit (X^2=122.3; df=14; CFI=0.977; TLI=0.966; SRMR=0.043; RMSEA[90%CI]=0.076[0.064-0.088]).

Reliability

The PHQ-9 (α=0.89; ω=0.86), the PHQ-8 (α=0.88; ω=0.85), and the GAD-7 (α=0.85; ω=0.81) showed optimal internal consistency values. Similarly, the PHQ-2 (α=0.83; ω=0.80) and the GAD-2 (α=0.74; ω=0.70) also showed adequate internal consistency scores. Table 3 shows the raw scores.

Discussion

Main findings

We determined the target population's optimal cut-off points for PHQ and GAD scales. The PHQ-9's ≥7 cut-off point showed the highest sensitivity and specificity when contrasted against a psychiatric diagnosis of depression (gold standard). For a similar contrast, the other optimal cut-off points were: ≥7 for the PHQ-8, ≥8 for the GAD-7 (against the anxiety gold standard), and ≥2 for the PHQ-2 and GAD-2. The scales with the best sensitivity and specificity balance were the PHQ-9 (cut-off ≥7) and the GAD-7 (cut-off ≥8) for detecting cases of depression and anxiety, respectively. In addition, the algorithm scoring or algorithm-adjusted scoring methods for the PHQ-9 had a lower balance between sensitivity and specificity scores than the PHQ-9 raw score scoring method with a cut-off ≥7. We confirmed the adequacy of a one-dimensional model for the PHQ-9, PHQ-8, and GAD-7, while all scales showed good internal consistency.

Contrast to literature

At the PHQ-9 level, evidence suggests that the raw score approach is more valuable than diagnostic algorithms [32], which is consistent with our findings. For the cut-off, different systematic reviews agree that the most commonly used cut-off is ≥10 [32, 60]. However, a meta-analysis found that cut-offs between 8 and 11 showed little difference in sensitivity and specificity [61]. The optimal cut-off reported in our study was slightly lower than that suggested by this meta-analysis, and two possible factors could explain this difference. Firstly, our population is inpatients in different areas of a high-complexity hospital. Other studies of hospitalised cancer patients [62], hospitalised neurology patients [63], and patients with coronary heart disease [64] also found an optimal cut-off between 5 and 7 points. Therefore, hospitalised individuals may be more likely to have depressive symptoms, which may require a lower cut-off on the PHQ-9. Second, several studies in populations from low- and middle-income countries have reported cut-offs between 5 and 7, for example, Pakistani migrants in the UK [65], Indian adolescents [66], and primary care in Ethiopia [67]. Therefore, factors such as social determinants of health present in such countries may influence cut-off.

Concerning the PHQ-8 and PHQ-9, we found that both scales have similar cut-off points (≥ 7). Our findings are consistent with a meta-analysis that found that the cut-offs between the two scales are identical; although sensitivity may be minimally reduced with the PHQ-8, specificity is similar between the two scales [35]. The PHQ-8 does not include the item corresponding to suicidal or self-harming ideation, and the use of this version of the PHQ is common in the general population, as suicidal ideation is less common in this group [15]. However, at the level of clinical populations, it has been found that omitting this item does not significantly alter the measurement capabilities of the PHQ, as the correlation between the PHQ-8 and PHQ-9 in clinical populations is very close to one [68].

Regarding the GAD-7, our findings are consistent with a meta-analysis that evaluated all possible cut-off points and reported that ≥8 is the most appropriate for anxiety disorder [17]. It also notes that

scores between 7 and 10 points have similar sensitivity and specificity values [17]. Other recent primary studies conducted in hospitalised populations or people with chronic diseases in hospital settings also found optimal cut-offs between 7 and 10 points [69-71].

Our results on PHQ-2 and GAD-2 are in line with meta-analyses supporting the use of the cut-off of 2 for PHQ-2 [34, 72]; however, most systematic reviews support the cut-off ≥3 for both instruments [17, 36, 73]. Some of the reasons for this heterogeneity may be due to the variety of populations included in the meta-analyses. The meta-analyses mentioned included studies in general populations (i.e., people attending primary care) and people hospitalised for non-communicable or infectious diseases. However, no meta-analyses were found that evaluated cohort points for hospitalised people only. At the level of primary studies, the evidence suggests that cohort scores vary between 2 and 3 points for the PHQ-2 and GAD-2 [74, 75].

Regarding internal validity, a systematic review examined the factor structure of the PHQ-9, noting that the one-dimensional model has been repeatedly confirmed across studies [76]. Although several studies evaluated alternative multidimensional models (e.g., two-dimensional, three-dimensional, or bifactorial models), their dimensions are often highly correlated with each other, so there may be overlapping [76]. We did not find systematic reviews on the internal structure of the GAD-7 and the PHQ-8. However, several studies support the one-dimensional model in hospitalised patients for both the PHQ-8 [77] and GAD-7 [20, 26]. In Peru, the GAD-7 and PHQ-9 have shown evidence of a one-dimensional factor structure in different populations, such as the general population [19], pregnant women [20], and university students [51, 78]. However, no studies have been found evaluating the factor structure of the PHQ-8 in the Peruvian population.

Public health implications

The evaluated instruments are widely used in clinical practice and research to measure symptoms of depression and anxiety, but from today, users will have optimal cut-off points for interpretations. This can help healthcare professionals identify people at risk of depression and anxiety more accurately while informing decisions about their formal diagnosis and consequent treatment. This is especially valuable in hospital environments, where time is crucial.

Our findings are of particular interest to the Peruvian health system, which has clinical practice guidelines for depression that recommend the PHQ-9 as a screening tool in primary care and hospital context [79]. Although our results correspond only to a hospital population, our study is the closest approximation to an evaluation of sensitivity and specificity in the Peruvian context, in the absence of similar studies in primary care. On the other hand, there is a lack of national clinical practice guidelines for screening and managing anxiety in Peru. Therefore, our study could contribute to future clinical practice guidelines for generalised anxiety disorder.

Our study recommends cut-offs for each version of the PHQ and DAG. However, health professionals and decision-makers should use the cut-offs that they consider most relevant for their purposes. There are situations where it is more important to be sure that the user does not have a mental health

problem. Therefore, an instrument with high specificity scores would be a good alternative. For example, the use of PHQ-9 with a cut-off of ≥10.

Strengths and limitations

Our results of the study have several strengths. First, to our knowledge, this is the first study in a Peruvian context that evaluates the factorial structure of all PHQ and GAD versions in a hospitalised population. Second, the scales were administered by a team of healthcare professionals with more than five years of experience in the clinical assessment of these patients. Third, the sample size was large enough to support all analyses and conclusions. Further, our sample size was larger than other recently published studies' [60]. Fourth, our study is the first Peruvian study to evaluate the sensitivity and specificity of the PHQ.

Our study has limitations. First, we conducted the study only in a hospital context in a Peruvian city, which limits its applicability to other settings in Peru or other countries. However, it could be used in other Peruvian hospital contexts with similar characteristics, which is relevant because hospital care in Peru (levels II and III of complexity) represents 58.65% of total care [80]. Secondly, the generalisability of our results may be limited because the sampling is not probabilistic, as it does not include other hospitals. However, the hospital where we conducted the study serves 1.1% of all nationally insured EsSalud patients (http://www.essalud.gob.pe/estadistica-institucional/). It is also a national referral hospital, which means that people from all over the country are referred to this hospital for treatment. Therefore, the representativeness of the results is ensured.

Conclusions

The PHQ-9's ≥7 cut-off point showed the highest simultaneous sensitivity and specificity when contrasted against a psychiatric diagnosis of depression. For a similar contrast against the gold standard, the other optimal cut-off points were: ≥7 for the PHQ-8, ≥8 for the GAD-7, and ≥2 for the PHQ-2 and GAD-2. We confirmed the adequacy of a one-dimensional model for the PHQ-9, PHQ-8, and GAD-7, while all PHQ and GAD scales showed good reliability.

Tables, Figures, and Supplementary material

Table 1. Sociodemographic characteristics (n=1347).
Table 2. Sensitivity, specificity, +LR, −LR, PPV, NPV, and Youden's index for different PHQ and GAD cut-off points compared to the gold standard.
Table 3. Raw scores and internal consistency (n=1347).

Figure 1. Receiver operator characteristics curve for Patient Health Questionnaire compared to the gold standard.
Figure 2. Receiver operator characteristics curve for Generalized Anxiety Disorder Scale compared to the gold standard.

Supplementary material 1. Sex and age of the participants in our study and of the total number of people hospitalised in 2022.
Supplementary material 2. Script of R used in our study.
Supplementary material 3. Do-file of STATA used in our study.
Supplementary material 4. Flowchart.
Supplementary material 5. Sensitivity, specificity, +LR, −LR, PPV, NPV, and Youden's index for different PHQ cut-off points compared to the gold standard.
Supplementary material 6. Sensitivity, specificity, +LR, −LR, PPV, NPV, and Youden's index for different GAD cut-off points compared to the gold standard.

Reference

1. **Global Health Data Exchange (GHDx), Institute for Health Metrics and Evaluation (IHME)** [https://microdata.worldbank.org/index.php/catalog/ghdx/?page=1&ps=15&repo=ghdx]
2. GBD 2019 Mental Disorders Collaborators: **Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019**. *The lancet Psychiatry* 2022, **9**(2):137-150.
3. GBD 2019 Diseases and Injuries Collaborators: **Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019**. *Lancet (London, England)* 2020, **396**(10258):1204-1222.
4. **COVID-19 pandemic triggers 25% increase in prevalence of anxiety and depression worldwide** [https://www.who.int/news/item/02-03-2022-covid-19-pandemic-triggers-25-increase-in-prevalence-of-anxiety-and-depression-worldwide]
5. Villarreal-Zegarra D, Reátegui-Rivera CM, Otazú-Alfaro S, Yantas-Alcantara G, Soto-Becerra P, Melendez-Torres GJ: **Estimated impact of the COVID-19 pandemic on the prevalence and treatment of depressive symptoms in Peru: an interrupted time series analysis in 2014–2021**. *Social psychiatry and psychiatric epidemiology* 2023.
6. COVID-19 Mental Disorders Collaborators: **Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic**. *Lancet (London, England)* 2021, **398**(10312):1700-1712.
7. Sagan A, McDaid D, Rajan S, Farrington J, McKee M: **European Observatory Policy Briefs**. In: *Screening: When is it appropriate and how can we get it right?* , edn. Copenhagen (Denmark): European Observatory on Health Systems and Policies © World Health Organization 2020 (acting as the host organization for, and secretariat of, the European Observatory on Health Systems and Policies). 2020.
8. Iragorri N, Spackman E: **Assessing the value of screening tools: reviewing the challenges and opportunities of cost-effectiveness analysis**. *Public health reviews* 2018, **39**:17.
9. Mulvaney-Day N, Marshall T, Downey Piscopo K, Korsen N, Lynch S, Karnell LH, Moran GE, Daniels AS, Ghose SS: **Screening for Behavioral Health Conditions in Primary Care Settings: A Systematic Review of the Literature**. *J Gen Intern Med* 2018, **33**(3):335-346.
10. Jiao B, Rosen Z, Bellanger M, Belkin G, Muennig P: **The cost-effectiveness of PHQ screening and collaborative care for depression in New York City**. *PloS one* 2017, **12**(8):e0184210.
11. Ramírez-Bontá F, Vásquez-Vílchez R, Cabrera-Alva M, Otazú-Alfaro S, Almeida-Huanca G, Ambrosio-Melgarejo J, Figueroa-Quiñones J, Romero-Cabrera A, Huaman-Santa Cruz A, Chávez-Hinostroza E *et al*: **Mental health data available in representative surveys conducted in Latin America and the Caribbean countries: A scoping review**. *In press* 2023.
12. Haberer JE, Trabin T, Klinkman M: **Furthering the reliable and valid measurement of mental health screening, diagnoses, treatment and outcomes through health information technology**. *Gen Hosp Psychiatry* 2013, **35**(4):349-353.
13. National Research Council; Division of Behavioral and Social Sciences and Education; Institute of Medicine; Board on Children Y, and Families; Committee on the Prevention of Mental Disorders and Substance Abuse Among Children, Youth and Young Adults:

Research Advances and Promising Interventions; Mary Ellen O'Connell, Thomas Boat, and Kenneth E. Warner, Editors: **Preventing Mental, Emotional, and Behavioral Disorders Among Young People: Progress and Possibilities**. Washington: National Academies Press; 2009.

14. Spitzer RL, Kroenke K, Williams JB: **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* 1999, **282**(18):1737-1744.
15. Kroenke K, Strine TW, Spitzer RL, Williams JB, Berry JT, Mokdad AH: **The PHQ-8 as a measure of current depression in the general population**. *J Affect Disord* 2009, **114**(1-3):163-173.
16. Kroenke K, Spitzer RL, Williams JB: **The Patient Health Questionnaire-2: validity of a two-item depression screener**. *Med Care* 2003, **41**(11):1284-1292.
17. Plummer F, Manea L, Trepel D, McMillan D: **Screening for anxiety disorders with the GAD-7 and GAD-2: a systematic review and diagnostic metaanalysis**. *Gen Hosp Psychiatry* 2016, **39**:24-31.
18. 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**(6):e0156939.
19. Villarreal-Zegarra D, Copez-Lonzoy A, Bernabe-Ortiz A, Melendez-Torres GJ, Bazo-Alvarez JC: **Valid group comparisons can be made with the Patient Health Questionnaire (PHQ-9): A measurement invariance study across groups by demographic characteristics**. *PloS one* 2019, **14**(9):1-15.
20. Zhong QY, Gelaye B, Zaslavsky AM, Fann JR, Rondon MB, Sanchez SE, Williams MA: **Diagnostic Validity of the Generalized Anxiety Disorder - 7 (GAD-7) among Pregnant Women**. *PloS one* 2015, **10**(4):e0125096.
21. Stochl J, Fried EI, Fritz J, Croudace TJ, Russo DA, Knight C, Jones PB, Perez J: **On Dimensionality, Measurement Invariance, and Suitability of Sum Scores for the PHQ-9 and the GAD-7**. *Assessment* 2022, **29**(3):355-366.
22. Shevlin M, Butter S, McBride O, Murphy J, Gibson-Miller J, Hartman TK, Levita L, Mason L, Martinez AP, McKay R *et al*: **Measurement invariance of the Patient Health Questionnaire (PHQ-9) and Generalized Anxiety Disorder scale (GAD-7) across four European countries during the COVID-19 pandemic**. *BMC psychiatry* 2022, **22**(1):154.
23. Kroenke K, Wu J, Yu Z, Bair MJ, Kean J, Stump T, Monahan PO: **Patient Health Questionnaire Anxiety and Depression Scale: Initial Validation in Three Clinical Trials**. *Psychosom Med* 2016, **78**(6):716-727.
24. Urtasun M, Daray FM, Teti GL, Coppolillo F, Herlax G, Saba G, Rubinstein A, Araya R, Irazola V: **Validation and calibration of the patient health questionnaire (PHQ-9) in Argentina**. *BMC psychiatry* 2019, **19**(1):291.
25. García-Campayo J, Zamorano E, Ruiz MA, Pardo A, Pérez-Páramo M, López-Gómez V, Freire O, Rejas J: **Cultural adaptation into Spanish of the generalized anxiety disorder-7 (GAD-7) scale as a screening tool**. *Health and quality of life outcomes* 2010, **8**:8.
26. Sawaya H, Atoui M, Hamadeh A, Zeinoun P, Nahas Z: **Adaptation and initial validation of the Patient Health Questionnaire - 9 (PHQ-9) and the Generalized Anxiety Disorder - 7 Questionnaire (GAD-7) in an Arabic speaking Lebanese psychiatric outpatient sample**. *Psychiatry Res* 2016, **239**:245-252.
27. Lotrakul M, Sumrithe S, Saipanish R: **Reliability and validity of the Thai version of the PHQ-9**. *BMC psychiatry* 2008, **8**:46.
28. Woldetensay YK, Belachew T, Tesfaye M, Spielman K, Biesalski HK, Kantelhardt EJ, Scherbaum V: **Validation of the Patient Health Questionnaire (PHQ-9) as a screening tool for depression in pregnant women: Afaan Oromo version**. *PloS one* 2018, **13**(2):e0191782.

29. Eack SM, Greeno CG, Lee BJ: **Limitations of the Patient Health Questionnaire in Identifying Anxiety and Depression: Many Cases Are Undetected**. *Research on social work practice* 2006, **16**(6):625-631.
30. Lambert SD, Clover K, Pallant JF, Britton B, King MT, Mitchell AJ, Carter G: **Making Sense of Variations in Prevalence Estimates of Depression in Cancer: A Co-Calibration of Commonly Used Depression Scales Using Rasch Analysis**. *Journal of the National Comprehensive Cancer Network : JNCCN* 2015, **13**(10):1203-1211.
31. Liu SI, Yeh ZT, Huang HC, Sun FJ, Tjung JJ, Hwang LC, Shih YH, Yeh AW: **Validation of Patient Health Questionnaire for depression screening among primary care patients in Taiwan**. *Comprehensive psychiatry* 2011, **52**(1):96-101.
32. Levis B, Benedetti A, Thombs BD: **Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression: individual participant data meta-analysis**. *BMJ : British Medical Journal* 2019, **365**:l1476.
33. Neupane D, Levis B, Bhandari PM, Thombs BD, Benedetti A: **Selective cutoff reporting in studies of the accuracy of the Patient Health Questionnaire-9 and Edinburgh Postnatal Depression Scale: Comparison of results based on published cutoffs versus all cutoffs using individual participant data meta-analysis**. *International journal of methods in psychiatric research* 2021, **30**(3):e1873.
34. Levis B, Sun Y, He C, Wu Y, Krishnan A, Bhandari PM, Neupane D, Imran M, Brehaut E, Negeri Z *et al*: **Accuracy of the PHQ-2 Alone and in Combination With the PHQ-9 for Screening to Detect Major Depression: Systematic Review and Meta-analysis**. *Jama* 2020, **323**(22):2290-2300.
35. Wu Y, Levis B, Riehm KE, Saadat N, Levis AW, Azar M, Rice DB, Boruff J, Cuijpers P, Gilbody S *et al*: **Equivalency of the diagnostic accuracy of the PHQ-8 and PHQ-9: a systematic review and individual participant data meta-analysis**. *Psychological medicine* 2020, **50**(8):1368-1380.
36. 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**(4):345-359.
37. Manea L, Boehnke JR, Gilbody S, Moriarty AS, McMillan D: **Are there researcher allegiance effects in diagnostic validation studies of the PHQ-9? A systematic review and meta-analysis**. *BMJ Open* 2017, **7**(9):e015247.
38. 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**. *BJPsych open* 2016, **2**(2):127-138.
39. Mughal AY, Devadas J, Ardman E, Levis B, Go VF, Gaynes BN: **A systematic review of validated screening tools for anxiety disorders and PTSD in low to middle income countries**. *BMC psychiatry* 2020, **20**(1):338.
40. Cohen JF, Korevaar DA, Altman DG, Bruns DE, Gatsonis CA, Hooft L, Irwig L, Levine D, Reitsma JB, de Vet HC *et al*: **STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration**. *BMJ Open* 2016, **6**(11):e012799.
41. Villarreal-Zegarra D, Cabrera-Alva M, Carrillo-Larco RM, Bernabe-Ortiz A: **Trends in the prevalence and treatment of depressive symptoms in Peru: a population-based study**. *BMJ Open* 2020, **10**(7):e036777.
42. Hernández-Vásquez A, Vargas-Fernández R, Bendezu-Quispe G, Grendas LN: **Depression in the Peruvian population and its associated factors: analysis of a national health survey**. *J Affect Disord* 2020, **273**:291-297.
43. Steel Z, Marnane C, Iranpour C, Chey T, Jackson JW, Patel V, Silove D: **The global prevalence of common mental disorders: a systematic review and meta-analysis 1980-2013**. *Int J Epidemiol* 2014, **43**(2):476-493.
44. Buderer NM: **Statistical methodology: I. Incorporating the prevalence of disease into the sample size calculation for sensitivity and specificity**. *Academic emergency*

medicine : official journal of the Society for Academic Emergency Medicine 1996, **3**(9):895-900.

45. Huarcaya-Victoria J, Segura V, Cárdenas D, Sardón K, Caqui M, Podestà Á: **Analysis of the care provided over a six-month period by the liaison psychiatry unit at a general hospital in Lima, Peru**. *Rev Colomb Psiquiatr (Engl Ed)* 2022, **51**(2):105-112.
46. 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**. *General hospital psychiatry* 2015, **37**(1):67-75.
47. Zuithoff NP, Vergouwe Y, King M, Nazareth I, van Wezep MJ, Moons KG, Geerlings MI: **The Patient Health Questionnaire-9 for detection of major depressive disorder in primary care: consequences of current thresholds in a crosssectional study**. *BMC Fam Pract* 2010, **11**:98.
48. Shin C, Lee SH, Han KM, Yoon HK, Han C: **Comparison of the Usefulness of the PHQ-8 and PHQ-9 for Screening for Major Depressive Disorder: Analysis of Psychiatric Outpatient Data**. *Psychiatry investigation* 2019, **16**(4):300-305.
49. Caycho-Rodríguez T, Barboza-Palomino M, Ventura-León J, Carbajal-León C, Noé-Grijalva M, Gallegos M, Reyes-Bossio M, Vivanco-Vidal A: **[Spanish translation and validation of a brief measure of anxiety by the COVID-19 in students of health sciences]**. *Ansiedad y Estrés* 2020, **26**(2):174-180.
50. Spitzer RL, Kroenke K, Williams JW, Löwe B: **A brief measure for assessing generalized anxiety disorder: The GAD-7**. *Archives of Internal Medicine* 2006, **166**(10):1092-1097.
51. Franco-Jimenez RA, Nuñez-Magallanes A: **[Psychometric Properties of the GAD-7, GAD-2, and GAD-Mini in Peruvian College Students]**. *Propósitos y Representaciones* 2022, **10**(1):e1437.
52. Kroenke K, Spitzer RL, Williams JB, Monahan PO, Löwe B: **Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection**. *Annals of internal medicine* 2007, **146**(5):317-325.
53. Hanley JA, McNeil BJ: **A method of comparing the areas under receiver operating characteristic curves derived from the same cases**. *Radiology* 1983, **148**(3):839-843.
54. Trevethan R: **Sensitivity, Specificity, and Predictive Values: Foundations, Pliabilities, and Pitfalls in Research and Practice**. *Frontiers in public health* 2017, **5**:307.
55. Ranganathan P, Aggarwal R: **Understanding the properties of diagnostic tests - Part 2: Likelihood ratios**. *Perspectives in clinical research* 2018, **9**(2):99-102.
56. Suh Y: **The Performance of Maximum Likelihood and Weighted Least Square Mean and Variance Adjusted Estimators in Testing Differential Item Functioning With Nonnormal Trait Distributions**. *Structural Equation Modeling: A Multidisciplinary Journal* 2015, **22**(4):568-580.
57. Holgado–Tello FP, Chacón–Moscoso S, Barbero–García I, Vila–Abad E: **Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables**. *Qual Quant* 2010, **44**(1):153-166.
58. Hair JF, Anderson RE, Tatham RL, Black WC: **Análisis multivariante**, vol. 491: Prentice Hall Madrid; 1999.
59. McDonald RP: **Test theory: A unified treatment**. New York: Taylor & Francis Group; 1999.
60. Costantini L, Pasquarella C, Odone A, Colucci ME, Costanza A, Serafini G, Aguglia A, Belvederi Murri M, Brakoulias V, Amore M *et al*: **Screening for depression in primary care with Patient Health Questionnaire-9 (PHQ-9): A systematic review**. *J Affect Disord* 2021, **279**:473-483.
61. Manea L, Gilbody S, McMillan D: **Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): a meta-analysis**. *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne* 2012, **184**(3):E191-196.

62. Hartung TJ, Friedrich M, Johansen C, Wittchen HU, Faller H, Koch U, Brähler E, Härter M, Keller M, Schulz H *et al*: **The Hospital Anxiety and Depression Scale (HADS) and the 9-item Patient Health Questionnaire (PHQ-9) as screening instruments for depression in patients with cancer**. *Cancer* 2017, **123**(21):4236-4243.
63. Sun Y, Kong Z, Song Y, Liu J, Wang X: **The validity and reliability of the PHQ-9 on screening of depression in neurology: a cross sectional study**. *BMC psychiatry* 2022, **22**(1):98.
64. Gholizadeh L, Shahmansouri N, Heydari M, Davidson PM: **Assessment and detection of depression in patients with coronary artery disease: validation of the Persian version of the PHQ-9**. *Contemporary nurse* 2019, **55**(2-3):185-194.
65. Husain N, Waheed W, Tomenson B, Creed F: **The validation of personal health questionnaire amongst people of Pakistani family origin living in the United Kingdom**. *J Affect Disord* 2007, **97**(1-3):261-264.
66. Ganguly S, Samanta M, Roy P, Chatterjee S, Kaplan DW, Basu B: **Patient health questionnaire-9 as an effective tool for screening of depression among Indian adolescents**. *The Journal of adolescent health : official publication of the Society for Adolescent Medicine* 2013, **52**(5):546-551.
67. Hanlon C, Medhin G, Selamu M, Breuer E, Worku B, Hailemariam M, Lund C, Prince M, Fekadu A: **Validity of brief screening questionnaires to detect depression in primary care in Ethiopia**. *J Affect Disord* 2015, **186**:32-39.
68. Razykov I, Ziegelstein RC, Whooley MA, Thombs BD: **The PHQ-9 versus the PHQ-8--is item 9 useful for assessing suicide risk in coronary artery disease patients? Data from the Heart and Soul Study**. *J Psychosom Res* 2012, **73**(3):163-168.
69. Konkan R, Senormancı O, Guclu O, Aydin E, Sungur M: **Validity and reliability study for the Turkish adaptation of the Generalized Anxiety Disorder-7 (GAD-7) scale**. *Archives of Neuropsychiatry* 2013, **50**:53-58.
70. Gong Y, Zhou H, Zhang Y, Zhu X, Wang X, Shen B, Xian J, Ding Y: **Validation of the 7-item Generalized Anxiety Disorder scale (GAD-7) as a screening tool for anxiety among pregnant Chinese women**. *J Affect Disord* 2021, **282**:98-103.
71. Snijkers JTW, van den Oever W, Weerts Z, Vork L, Mujagic Z, Leue C, Hesselink MAM, Kruimel JW, Muris JWM, Bogie RMM *et al*: **Examining the optimal cutoff values of HADS, PHQ-9 and GAD-7 as screening instruments for depression and anxiety in irritable bowel syndrome**. *Neurogastroenterology and motility : the official journal of the European Gastrointestinal Motility Society* 2021, **33**(12):e14161.
72. Manea L, Gilbody S, Hewitt C, North A, Plummer F, Richardson R, Thombs BD, Williams B, McMillan D: **Identifying depression with the PHQ-2: A diagnostic meta-analysis**. *J Affect Disord* 2016, **203**:382-395.
73. Luo Z, Li Y, Hou Y, Zhang H, Liu X, Qian X, Jiang J, Wang Y, Liu X, Dong X *et al*: **Adaptation of the two-item generalized anxiety disorder scale (GAD-2) to Chinese rural population: A validation study and meta-analysis**. *Gen Hosp Psychiatry* 2019, **60**:50-56.
74. Giuliani M, Gorini A, Barbieri S, Veglia F, Tremoli E: **Examination of the best cut-off points of PHQ-2 and GAD-2 for detecting depression and anxiety in Italian cardiovascular inpatients**. *Psychology & health* 2021, **36**(9):1088-1101.
75. Bentley KH, Sakurai H, Lowman KL, Rines-Toth L, McKowen J, Pedrelli P, Evins AE, Yule AM: **Validation of brief screening measures for depression and anxiety in young people with substance use disorders**. *J Affect Disord* 2021, **282**:1021-1029.
76. Lamela D, Soreira C, Matos P, Morais A: **Systematic review of the factor structure and measurement invariance of the patient health questionnaire-9 (PHQ-9) and validation of the Portuguese version in community settings**. *J Affect Disord* 2020, **276**:220-233.

77. Schantz K, Reighard C, Aikens JE, Aruquipa A, Pinto B, Valverde H, Piette JD: **Screening for depression in Andean Latin America: Factor structure and reliability of the CES-D short form and the PHQ-8 among Bolivian public hospital patients**. *International journal of psychiatry in medicine* 2017, **52**(4-6):315-327.
78. Huarcaya-Victoria J, De-Lama-Morán R, Quiros M, Bazán J, López K, Lora D: **Propiedades psicométricas del Patient Health Questionnaire (PHQ-9) en estudiantes de medicina en Lima, Perú**. *Revista de Neuro-Psiquiatria* 2020, **83**(2):72-78.
79. Beatrice M-F, Carla M-C, Matilde L-M, Víctor G-M, Amelia M-M, Jessica Hanae Z-T, Sergio G-L, Raisa NM-R, Wendy N-G, Fabian F-S *et al*: **Clinical practice guideline for the screening and management of the mild depressive episode at the first level of care for the Peruvian Social Security (EsSalud)**. *ACTA MEDICA PERUANA* 2020, **37**(4).
80. EsSalud: **Análisis ejecutivo nacional de a nivel las prestaciones de salud 2016**. In. Lima: Gerencia central de planeamiento y presupuesto, EsSalud; 2017.

Table 1. Sociodemographic characteristics (n=1347).

	n	%
Sex		
Men	547	40.6%
Women	800	59.4%
Age (categories)		
18-29	107	7.9%
30-39	164	12.2%
40-49	214	15.9%
50-59	284	21.1%
60-69	294	21.8%
70-79	203	15.1%
80 to more	81	6.0%
Civil status		
Single	329	24.4%
Married or Cohabitant	768	57.0%
Separated	133	9.9%
Widowed	117	8.7%
Education level		
None	13	1.0%
Elementary school	135	10.0%
High school	478	35.5%
Technical	246	18.2%
University	475	35.3%
Currently works		
No	330	24.5%
Yes	778	57.8%
Retired	239	17.7%
Living alone		
Yes	99	7.3%
No	1248	92.7%
History of psychiatric diagnosis		
Yes	388	28.8%
No	959	71.2%
Diagnosis of depression		
No	1013	75.2%
Yes	334	24.8%
Diagnosis of anxiety		
No	1319	97.9%
Yes	28	2.1%

Note: n=number. %=Percentage.

Table 2. Sensitivity, specificity, +LR, −LR, PPV, NPV, and Youden's index for different PHQ and GAD cut-off points compared to the gold standard.

Method	Cut-off Points	n (%)	Sensitivity (95%CI)	Specificity (95%CI)	+LR (95%CI)	-LR (95%CI)	PPV (95%CI)	NPV (95%CI)	Youden Index
PHQ-9	**≥7**	**537 (39.9)**	**76.0 (71.1 - 80.5)**	**72.1 (69.2 - 74.8)**	**2.72 (2.42 - 3.06)**	**0.33 (0.27 - 0.40)**	**47.3 (43.0 - 51.6)**	**90.1 (87.9 - 92.1)**	**48.1**
PHQ-9	≥10	309 (22.9)	54.2 (48.7 - 59.6)	87.4 (85.2 - 89.3)	4.29 (3.55 - 5.18)	0.52 (0.47 - 0.59)	58.6 (53.9 - 63.1)	85.3 (83.7 - 86.7)	41.6
PHQ-9 algorithm	-	183 (13.6)	34.7 (29.6 - 40.1)	93.4 (91.7 - 94.8)	5.25 (3.99 - 6.91)	0.70 (0.65 - 0.76)	63.4 (56.0 - 70.4)	81.3 (78.9 - 83.5)	28.1
PHQ-9 adjusted algorithm	-	601 (44.6)	78.1 (73.3 - 82.5)	66.4 (63.4 - 69.3)	2.33 (2.10 - 2.58)	0.33 (0.27 - 0.41)	43.4 (39.4 - 47.5)	90.2 (87.9 - 92.3)	44.5
PHQ-8	**≥7**	**521 (38.7)**	**74.3 (69.2 - 78.9)**	**73.1 (70.2 - 75.8)**	**2.76 (2.44 - 3.10)**	**0.35 (0.29 - 0.42)**	**47.6 (43.2 - 52.0)**	**89.6 (87.3 - 91.6)**	**47.4**
PHQ-2	**≥2**	**730 (54.2)**	**84.7 (80.4 - 88.4)**	**55.9 (52.8 - 59.0)**	**1.92 (1.77 - 2.09)**	**0.27 (0.21 - 0.35)**	**38.8 (35.2 - 42.4)**	**91.7 (89.3 - 93.8)**	**40.6**
GAD-7	**≥8**	**295 (21.9)**	**53.6 (33.9 - 72.5)**	**78.8 (76.5 - 81.0)**	**2.52 (1.76 - 3.62)**	**0.59 (0.40 - 0.88)**	**5.1 (2.9 - 8.3)**	**98.8 (97.9 - 99.3)**	**32.4**
GAD-7	≥10	164 (12.2)	39.3 (21.5 - 59.4)	88.4 (86.5 - 90.1)	3.39 (2.09 - 5.50)	0.69 (0.51 - 0.93)	6.7 (3.4 - 11.7)	98.6 (97.7 - 99.2)	27.7
GAD-2	**≥2**	**681 (50.6)**	**82.1 (63.1 - 93.9)**	**50.1 (47.4 - 52.8)**	**1.65 (1.37 - 1.97)**	**0.36 (0.16 - 0.79)**	**3.4 (2.2 - 5.0)**	**99.2 (98.3 - 99.8)**	**32.2**

Note: +LR = Positive likelihood ratio. −LR = Negative likelihood ratio. PPV= positive predictive value. NPV= negative predictive value. Youden's Index = Sensitivity+Specificity-1. 95%CI = 95% confidence interval. GAD =Generalized Anxiety Disorder Scale. PHQ = Patient Health Questionnaire. Bold values represent the cohort point with the highest Youden Index. All cut-off points for the PHQ and GAD versions can be found in supplementary materials 3 and 4, respectively.

Table 3. Raw scores and internal consistency (n=1347).

	M	SD	Min	Max	α	ω
(1) PHQ-9 score	6.4	5.0	0	27	0.89	0.86
(2) PHQ-8 score	6.1	4.7	0	24	0.88	0.85
(3) PHQ-2 score	1.9	1.6	0	6	0.83	0.8
(4) GAD-7 score	5.1	3.9	0	21	0.85	0.81
(5) GAD-2 score	1.7	1.4	0	6	0.74	0.7

Note: All correlation coefficients were significant ($p<0.001$). α = Classical alpha. ω = Mcdonald's omega

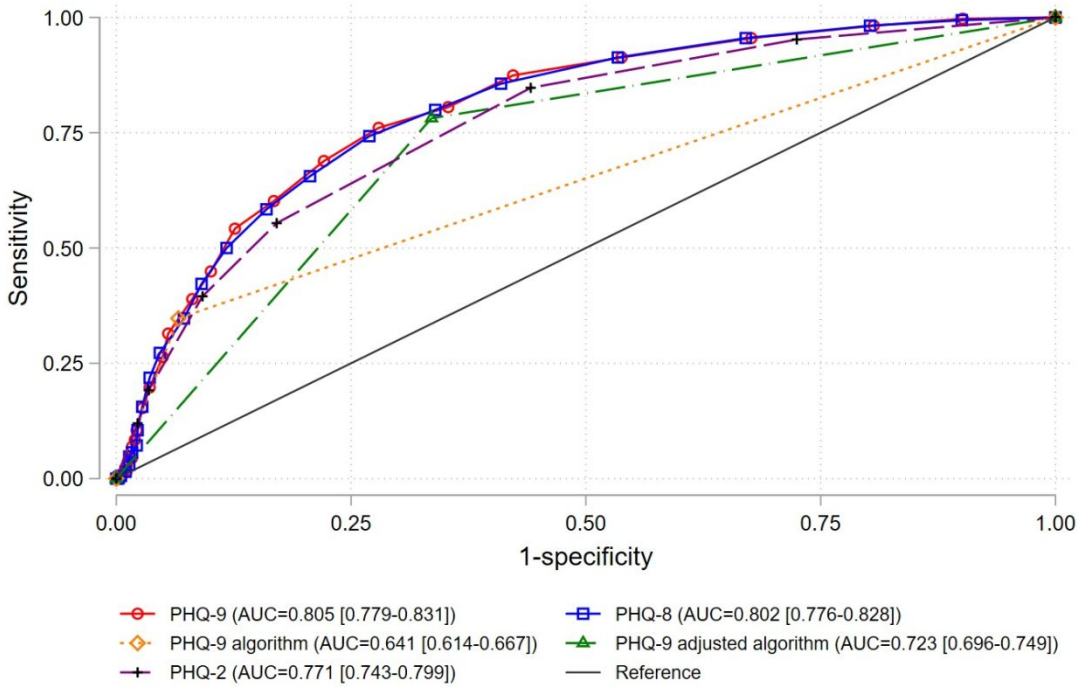


Figure 1. Receiver operator characteristics curve for Patient Health Questionnaire compared to the gold standard.

Note: Standard Error = 0.01.

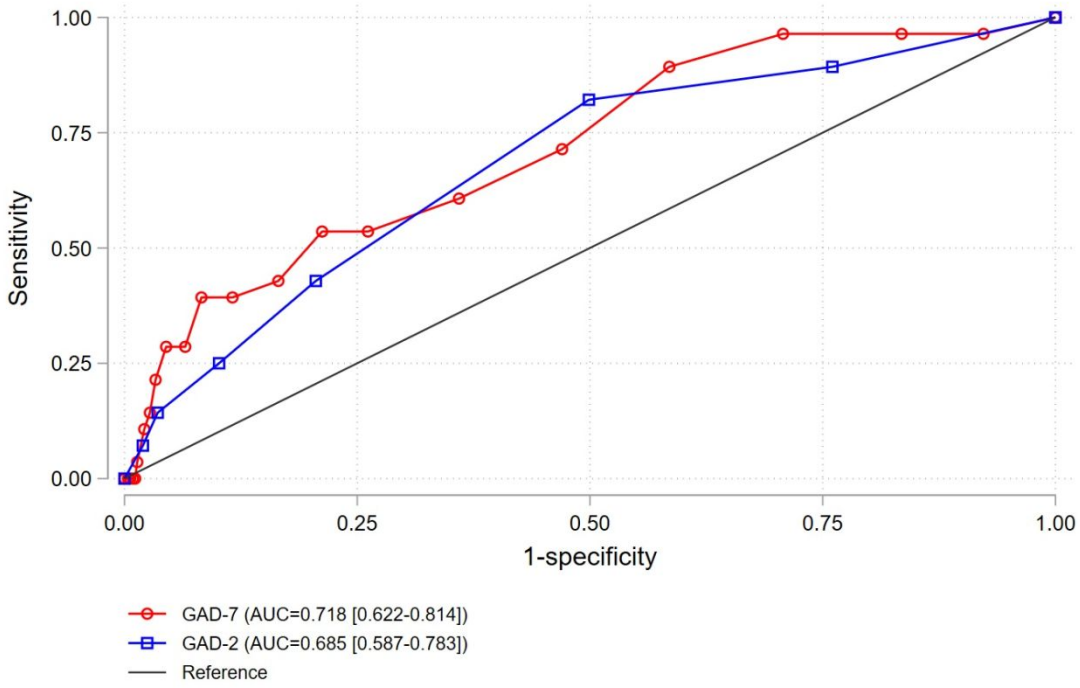


Figure 2. Receiver operator characteristics curve for Generalized Anxiety Disorder Scale compared to the gold standard.
Note: Standard Error = 0.05.

Supplementary material 1. Sex and age of the participants in our study and of the total number of people hospitalised in 2022.

	Our study		Total inpatients in 2022		
	n	%			p
Sex					
Men	547	40.60%	9677	43.2%	0.971
Women	800	59.40%	12732	56.8%	
Age (categories)					
18-29	107	7.90%	1962	8.8%	1.000
30-39	164	12.20%	3486	15.6%	
40-49	214	15.90%	3145	14.0%	
50-59	284	21.10%	3497	15.6%	
60-69	294	21.80%	4276	19.1%	
70-79	203	15.10%	3806	17.0%	
80 to more	81	6.00%	2227	9.9%	

Note: p = chi-square test.

Supplementary material 2. Script of R used in our study.

```
library(lavaan)
library(semPlot)
library(semTools)
library(psych)
library(haven)
Database <- read_dta("E:/Database_v1.dta")

#PHQ-9
model.PHQ9 <- "
F1 =~ PHQ_1 + PHQ_2 + PHQ_3 + PHQ_4 + PHQ_5 + PHQ_6 + PHQ_7 + PHQ_8 + PHQ_9"

fit.model.PHQ9 <- cfa(model.PHQ9, data=Database,estimator="WLSMV", missing = "listwise",
ordered = c("PHQ_1", "PHQ_2", "PHQ_3", "PHQ_4", "PHQ_5", "PHQ_6", "PHQ_7", "PHQ_8", "PHQ_9"))

summary(fit.model.PHQ9, fit.measures=TRUE, standardized=TRUE)

reliability(fit.model.PHQ9)

#PHQ-8
model.PHQ8 <- "
F1 =~ PHQ_1 + PHQ_2 + PHQ_3 + PHQ_4 + PHQ_5 + PHQ_6 + PHQ_7 + PHQ_8"

fit.model.PHQ8 <- cfa(model.PHQ8, data=Database, estimator="WLSMV", missing = "listwise",
ordered = c("PHQ_1", "PHQ_2", "PHQ_3", "PHQ_4", "PHQ_5", "PHQ_6", "PHQ_7", "PHQ_8", "PHQ_9"))

summary(fit.model.PHQ8, fit.measures=TRUE, standardized=TRUE)

reliability(fit.model.PHQ8)

#PHQ-2
model.PHQ2 <- "
F1 =~ PHQ_1 + PHQ_2"

fit.model.PHQ2 <- cfa(model.PHQ2, data=Database, estimator="WLSMV", missing = "listwise",
ordered = c("PHQ_1", "PHQ_2", "PHQ_3", "PHQ_4", "PHQ_5", "PHQ_6", "PHQ_7", "PHQ_8", "PHQ_9"))

summary(fit.model.PHQ2, fit.measures=TRUE, standardized=TRUE)

reliability(fit.model.PHQ2)

#GAD-7
model.GAD7 <- "
F1 =~ GAD1 + GAD2 + GAD3 + GAD4 + GAD5 + GAD6 + GAD7"

fit.model.GAD7 <- cfa(model.GAD7, data=Database, estimator="WLSMV", missing = "listwise",
ordered = c("GAD1", "GAD2", "GAD3", "GAD4", "GAD5", "GAD6", "GAD7"))

summary(fit.model.GAD7, fit.measures=TRUE, standardized=TRUE)

reliability(fit.model.GAD7)

#GAD-2
model.GAD2 <- "
F1 =~ GAD1 + GAD2"

fit.model.GAD2 <- cfa(model.GAD2, data=Database, estimator="WLSMV", missing = "listwise",
ordered = c("GAD1", "GAD2", "GAD3", "GAD4", "GAD5", "GAD6", "GAD7"))

summary(fit.model.GAD2, fit.measures=TRUE, standardized=TRUE)

reliability(fit.model.GAD2)
```

Supplementary material 3. Do-file of STATA used in our study.

```
use "E:\Database_v1.dta", clear

*Table 1 - Socio-demographic analysis

global catvars_table1 sex agecat civilstatus educationcat work Livingalone
Historypsychiatricdx depression anxiety

tabout $catvars_table1 depression using Table1.xlsx, ///
  replace c(freq col) clab(No. %) f(0c 1p) style(xlsx) font(bold) ///
  ptotal(none) stats(chi2) stpos(col) ppos(only) plab(P value) ///
  title(Table 1. Sociodemographic characteristics) ///
  fn(Note: n=number, %=Percentage.) twidth(14) sheet(Table1)

* Table 2. Sensitivity, specificity, +LR, -LR, PPV, NPV, and Youden's index for different PHQ
and GAD cut-off points compared to the gold standard.
tab cutoffPHQ9_7
diagt depression cutoffPHQ9_7
tab cutoffPHQ9_10
tab PHQ9algorithm
diagt depression PHQ9algorithm
tab PHQ9ajus_algorithm
diagt depression PHQ9ajus_algorithm
diagt depression cutoffPHQ9_10
tab cutoffPHQ8_7
diagt depression cutoffPHQ8_7
tab cutoffPHQ2_2
diagt depression cutoffPHQ2_2
tab cutoffGAD7_8
diagt anxiety cutoffGAD7_8
tab cutoffGAD7_10
diagt anxiety cutoffGAD7_10
tab cutoffGAD2_2
diagt anxiety cutoffGAD2_2



* Supplementary material 3. Sensitivity, specificity, +LR, -LR, PPV, NPV, and Youden's index
for different PHQ cut-off points compared to the gold standard.
* PHQ-9 - Sensitivity, specificity, +LR, -LR, PPV, NPV, and Youden's index

tab cutoffPHQ9_1
diagt depression cutoffPHQ9_1
tab cutoffPHQ9_2
diagt depression cutoffPHQ9_2
tab cutoffPHQ9_3
diagt depression cutoffPHQ9_3
tab cutoffPHQ9_4
diagt depression cutoffPHQ9_4
tab cutoffPHQ9_5
diagt depression cutoffPHQ9_5
tab cutoffPHQ9_6
diagt depression cutoffPHQ9_6
tab cutoffPHQ9_7
diagt depression cutoffPHQ9_7
tab cutoffPHQ9_8
diagt depression cutoffPHQ9_8
tab cutoffPHQ9_9
diagt depression cutoffPHQ9_9
tab cutoffPHQ9_10
diagt depression cutoffPHQ9_10
tab cutoffPHQ9_11
diagt depression cutoffPHQ9_11
```

```
tab cutoffPHQ9_12
diagt depression cutoffPHQ9_12
tab cutoffPHQ9_13
diagt depression cutoffPHQ9_13
tab cutoffPHQ9_14
diagt depression cutoffPHQ9_14
tab cutoffPHQ9_15
diagt depression cutoffPHQ9_15
tab cutoffPHQ9_16
diagt depression cutoffPHQ9_16
tab cutoffPHQ9_17
diagt depression cutoffPHQ9_17
tab cutoffPHQ9_18
diagt depression cutoffPHQ9_18
tab cutoffPHQ9_19
diagt depression cutoffPHQ9_19
tab cutoffPHQ9_20
diagt depression cutoffPHQ9_20
tab cutoffPHQ9_21
diagt depression cutoffPHQ9_21
tab cutoffPHQ9_22
diagt depression cutoffPHQ9_22
tab cutoffPHQ9_23
diagt depression cutoffPHQ9_23
tab cutoffPHQ9_24
diagt depression cutoffPHQ9_24
tab cutoffPHQ9_25
diagt depression cutoffPHQ9_25
tab cutoffPHQ9_26
diagt depression cutoffPHQ9_26
tab cutoffPHQ9_27
diagt depression cutoffPHQ9_27
```

***PHQ-8 - Sensitivity, specificity, +LR, −LR, PPV, NPV, and Youden's index**

```
tab cutoffPHQ8_1
diagt depression cutoffPHQ8_1
tab cutoffPHQ8_2
diagt depression cutoffPHQ8_2
tab cutoffPHQ8_3
diagt depression cutoffPHQ8_3
tab cutoffPHQ8_4
diagt depression cutoffPHQ8_4
tab cutoffPHQ8_5
diagt depression cutoffPHQ8_5
tab cutoffPHQ8_6
diagt depression cutoffPHQ8_6
tab cutoffPHQ8_7
diagt depression cutoffPHQ8_7
tab cutoffPHQ8_8
diagt depression cutoffPHQ8_8
tab cutoffPHQ8_9
diagt depression cutoffPHQ8_9
tab cutoffPHQ8_10
diagt depression cutoffPHQ8_10
tab cutoffPHQ8_11
diagt depression cutoffPHQ8_11
tab cutoffPHQ8_12
diagt depression cutoffPHQ8_12
tab cutoffPHQ8_13
diagt depression cutoffPHQ8_13
tab cutoffPHQ8_14
diagt depression cutoffPHQ8_14
tab cutoffPHQ8_15
```

```
diagt depression cutoffPHQ8_15
tab cutoffPHQ8_16
diagt depression cutoffPHQ8_16
tab cutoffPHQ8_17
diagt depression cutoffPHQ8_17
tab cutoffPHQ8_18
diagt depression cutoffPHQ8_18
tab cutoffPHQ8_19
diagt depression cutoffPHQ8_19
tab cutoffPHQ8_20
diagt depression cutoffPHQ8_20
tab cutoffPHQ8_21
diagt depression cutoffPHQ8_21
tab cutoffPHQ8_22
diagt depression cutoffPHQ8_22
tab cutoffPHQ8_23
diagt depression cutoffPHQ8_23
tab cutoffPHQ8_24
diagt depression cutoffPHQ8_24
```

*** PHQ-9 algorithm - Sensitivity, specificity, +LR, −LR, PPV, NPV, and Youden's index**

```
tab PHQ9algorithm
diagt depression PHQ9algorithm
```

*** PHQ-9 adjusted algorithm - Sensitivity, specificity, +LR, −LR, PPV, NPV, and Youden's index**

```
tab PHQ9ajus_algorithm
diagt depression PHQ9ajus_algorithm
```

*** PHQ-2 - Sensitivity, specificity, +LR, −LR, PPV, NPV, and Youden's index**

```
tab cutoffPHQ2_1
diagt depression cutoffPHQ2_1
tab cutoffPHQ2_2
diagt depression cutoffPHQ2_2
tab cutoffPHQ2_3
diagt depression cutoffPHQ2_3
tab cutoffPHQ2_4
diagt depression cutoffPHQ2_4
tab cutoffPHQ2_5
diagt depression cutoffPHQ2_5
tab cutoffPHQ2_6
diagt depression cutoffPHQ2_6
```

***Supplementary material 4. Sensitivity, specificity, +LR, −LR, PPV, NPV, and Youden's index for different GAD cut-off points compared to the gold standard.**
***GAD-7 - Sensitivity, specificity, +LR, −LR, PPV, NPV, and Youden's index**

```
tab cutoffGAD7_1
diagt anxiety cutoffGAD7_1
tab cutoffGAD7_2
diagt anxiety cutoffGAD7_2
tab cutoffGAD7_3
diagt anxiety cutoffGAD7_3
tab cutoffGAD7_4
diagt anxiety cutoffGAD7_4
tab cutoffGAD7_5
diagt anxiety cutoffGAD7_5
tab cutoffGAD7_6
diagt anxiety cutoffGAD7_6
tab cutoffGAD7_7
diagt anxiety cutoffGAD7_7
tab cutoffGAD7_8
```

```
diagt anxiety cutoffGAD7_8
tab cutoffGAD7_9
diagt anxiety cutoffGAD7_9
tab cutoffGAD7_10
diagt anxiety cutoffGAD7_10
tab cutoffGAD7_11
diagt anxiety cutoffGAD7_11
tab cutoffGAD7_12
diagt anxiety cutoffGAD7_12
tab cutoffGAD7_13
diagt anxiety cutoffGAD7_13
tab cutoffGAD7_14
diagt anxiety cutoffGAD7_14
tab cutoffGAD7_15
diagt anxiety cutoffGAD7_15
tab cutoffGAD7_16
diagt anxiety cutoffGAD7_16
tab cutoffGAD7_17
diagt anxiety cutoffGAD7_17
tab cutoffGAD7_18
diagt anxiety cutoffGAD7_18
tab cutoffGAD7_19
diagt anxiety cutoffGAD7_19
tab cutoffGAD7_20
diagt anxiety cutoffGAD7_20
tab cutoffGAD7_21
diagt anxiety cutoffGAD7_21
```

*** GAD-2 - Sensitivity, specificity, +LR, −LR, PPV, NPV, and Youden's index**

```
tab cutoffGAD2_0
diagt anxiety cutoffGAD2_0
tab cutoffGAD2_1
diagt anxiety cutoffGAD2_1
tab cutoffGAD2_2
diagt anxiety cutoffGAD2_2
tab cutoffGAD2_3
diagt anxiety cutoffGAD2_3
tab cutoffGAD2_4
diagt anxiety cutoffGAD2_4
tab cutoffGAD2_5
diagt anxiety cutoffGAD2_5
tab cutoffGAD2_6
diagt anxiety cutoffGAD2_6
```

***Figures 2 - ROC curve**

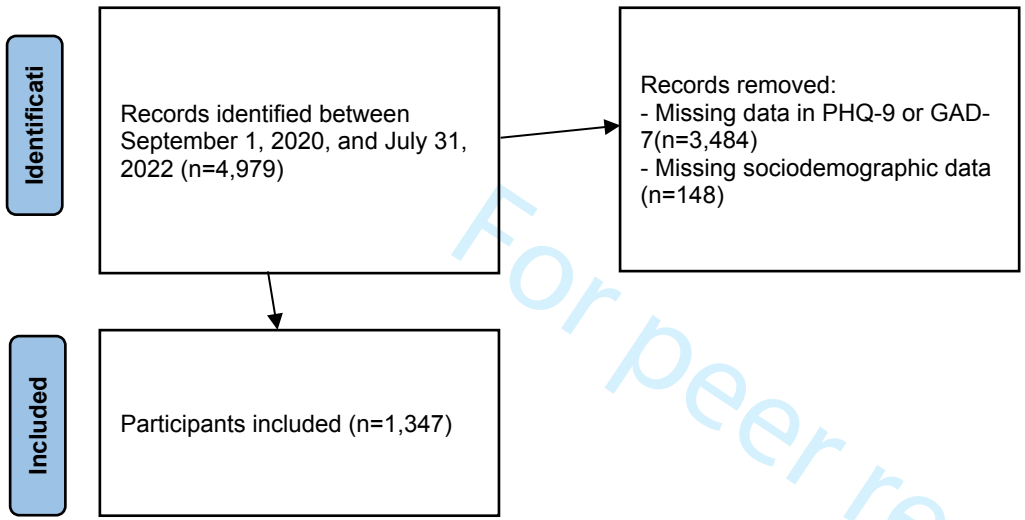
```
roccomp depression PHQ9TOTAL PHQ8TOTAL PHQ9algorithm PHQ9ajus_algorithm PHQ2TOTAL, graph
summary plot1opts(mcolor(red) lcolor(red)) plot2opts(mcolor(blue) lcolor(blue))
plot3opts(mcolor(orange) lcolor(orange)) plot4opts(mcolor(green) lcolor(green))
plot5opts(mcolor(purble) lcolor(purple)) legend(order(1 "PHQ-9 (AUC=0.805 [0.779-0.831])" 2
"PHQ-8 (AUC=0.802 [0.776-0.828])" 3 "PHQ-9 algorithm (AUC=0.641 [0.614-0.667])" 4 "PHQ-9
adjusted algorithm (AUC=0.723 [0.696-0.749])" 5 "PHQ-2 (AUC=0.771 [0.743-0.799])" 6
"Reference") size(2.5) position(7) cols(2) rows(3))

graph export "E:\Figure2.tif", as(tif) replace
```

***Figures 3 - ROC curve**

```
roccomp anxiety GAD7TOTAL GAD2TOTAL, graph summary plot1opts(mcolor(red) lcolor(red))
plot2opts(mcolor(blue) lcolor(blue)) legend(order(1 "GAD-7 (AUC=0.718 [0.622-0.814])" 2 "GAD-2
(AUC=0.685 [0.587-0.783])" 3 "Reference") size(2.5) position(7) cols(2) rows(3))

graph export "E:\Figure3.tif", as(tif) replace
```



Supplementary material 4. Flowchart.

Supplementary material 5. Sensitivity, specificity, +LR, −LR, PPV, NPV, and Youden's index for different PHQ cut-off points compared to the gold standard.

Method	Cut-off Points	n (%)	Sensitivity (95%CI)	Specificity (95%CI)	+LR (95%CI)	-LR (95%CI)	PPV (95%CI)	NPV (95%CI)	Youden Index
PHQ-9	≥1	1246 (92.5)	99.7 (98.3 - 100.0)	9.9 (8.1 - 11.9)	1.11 (1.08 - 1.13)	0.03 (0.00 - 0.22)	26.7 (24.3 - 29.3)	99.0 (94.6 - 100.0)	9.6
	≥2	1145 (85.0)	98.2 (96.1 - 99.3)	19.3 (17.0 - 21.9)	1.22 (1.18 - 1.26)	0.93 (0.04 - 0.21)	28.6 (26.0 - 31.4)	97.0 (93.6 - 98.9)	17.5
	≥3	1004 (74.5)	95.5 (92.7 - 97.5)	32.4 (29.5 - 35.4)	1.41 (1.35 - 1.48)	0.14 (0.84 - 0.23)	31.8 (28.9 - 34.8)	95.6 (92.9 - 97.5)	27.9
	≥4	850 (63.1)	91.3 (87.8 - 94.1)	46.2 (43.1 - 49.3)	1.70 (1.59 - 1.81)	0.19 (0.13 - 0.27)	35.9 (32.7 - 39.2)	94.2 (91.7 - 96.1)	37.5
	≥5	720 (53.5)	87.4 (83.4 - 90.8)	57.7 (54.6 - 60.8)	2.07 (1.90 - 2.25)	0.22 (0.16 - 0.29)	40.6 (36.9 - 44.2)	93.3 (91.1 - 95.1)	45.1
	≥6	627 (46.6)	80.5 (75.9 - 84.6)	64.7 (61.6 - 67.6)	2.28 (2.07 - 2.52)	0.30 (0.24 - 0.38)	42.9 (39.0 - 46.9)	91.0 (88.6 - 93.0)	45.2
	≥7	**537 (39.9)**	**76.0 (71.1 - 80.5)**	**72.1 (69.2 - 74.8)**	**2.72 (2.42 - 3.06)**	**0.33 (0.27 - 0.40)**	**47.3 (43.0 - 51.6)**	**90.1 (87.9 - 92.1)**	**48.1**
	≥8	454 (33.7)	68.9 (63.6 - 73.8)	77.9 (75.2 - 80.4)	3.11 (2.72 - 3.57)	0.40 (0.34 - 0.47)	50.7 (46.0 - 55.4)	88.4 (86.1 - 90.4)	46.8
	≥9	371 (27.5)	60.2 (54.7 - 65.5)	83.2 (80.8 - 85.5)	3.59 (3.05 - 4.22)	0.48 (0.42 - 0.55)	54.2 (49.0 - 59.3)	86.4 (84.1 - 88.5)	43.4
	≥10	309 (22.9)	54.2 (48.7 - 59.6)	87.4 (85.2 - 89.3)	4.29 (3.55 - 5.18)	0.52 (0.47 - 0.59)	58.6 (53.9 - 63.1)	85.3 (83.7 - 86.7)	41.6
	≥11	252 (18.7)	44.9 (39.5 - 50.4)	89.9 (87.9 - 91.7)	4.46 (3.58 - 5.55)	0.61 (0.56 - 0.68)	59.5 (53.2 - 65.6)	83.2 (80.8 - 85.4)	34.8
	≥12	212 (15.7)	38.9 (33.7 - 44.4)	91.9 (90.1 - 93.5)	4.81 (3.76 - 6.16)	0.67 (0.61 - 0.73)	61.3 (54.4 - 67.9)	82.0 (79.7 - 84.2)	30.8
	≥13	161 (12.0)	31.4 (26.5 - 36.7)	94.5 (92.9 - 95.8)	5.69 (4.21 - 7.67)	0.73 (0.67 - 0.78)	65.2 (57.3 - 72.5)	80.7 (78.3 - 82.9)	25.9
	≥14	138 (10.2)	26.3 (21.7 - 31.4)	95.1 (93.5 - 96.3)	5.34 (3.86 - 7.38)	0.78 (0.73 - 0.83)	63.8 (55.2 - 71.8)	79.7 (77.3 - 81.9)	21.4
	≥15	102 (7.6)	19.8 (15.6 - 24.4)	96.4 (95.1 - 97.5)	5.56 (3.78 - 8.19)	0.83 (0.79 - 0.88)	64.7 (54.6 - 73.9)	78.5 (76.1 - 80.7)	16.2
	≥16	79 (5.9)	15.3 (11.6 - 19.6)	97.2 (96.0 - 98.2)	5.52 (3.54 - 8.61)	0.87 (0.83 - 0.91)	64.6 (53.0 - 75.0)	77.7 (75.3 - 79.9)	12.5
	≥17	57 (4.2)	10.5 (7.4 - 14.3)	97.8 (96.7 - 98.6)	4.83 (2.87 - 8.11)	0.92 (0.88 - 0.95)	61.4 (47.6 - 74.0)	76.8 (74.4 - 79.1)	8.3
	≥18	48 (3.6)	8.4 (5.6 - 11.9)	98.0 (97.0 - 98.8)	4.25 (2.42 - 7.44)	0.94 (0.90 - 0.97)	58.3 (43.2 - 72.4)	76.4 (74.0 - 78.7)	6.4
	≥19	40 (3.0)	6.9 (4.4 - 10.2)	98.3 (97.3 - 99.0)	4.10 (2.22 - 7.59)	0.95 (0.92 - 0.98)	57.5 (40.9 - 73.0)	76.2 (73.8 - 78.5)	5.2
	≥20	35 (2.4)	4.5 (2.5 - 7.3)	98.3 (97.3 - 99.0)	2.68 (1.35 - 5.30)	0.97 (0.95 - 1.00)	46.9 (29.1 - 65.3)	75.7 (73.3 - 78.0)	2.8
	≥21	26 (1.9)	3.9 (2.1 - 6.6)	98.7 (97.8 - 99.3)	3.03 (1.42 - 6.48)	0.97 (0.95 - 1.00)	50.0 (29.9 - 70.1)	75.7 (73.3 - 78.0)	2.6
	≥22	20 (1.5)	2.7 (1.2 - 5.1)	98.9 (98.1 - 99.5)	2.48 (1.04 - 5.94)	0.98 (0.97 - 1.00)	45.0 (23.1 - 68.5)	75.5 (73.1 - 77.8)	1.6
	≥23	16 (1.2)	1.5 (0.5 - 3.5)	98.9 (98.1 - 99.5)	1.38 (0.48 - 3.94)	1.00 (0.98 - 1.01)	31.3 (11.0 - 58.7)	75.3 (72.9 - 77.6)	0.4
	≥24	10 (0.7)	0.9 (0.2 - 2.6)	99.3 (98.6 - 99.7)	1.30 (0.34 - 5.00)	1.00 (0.99 - 1.01)	30.0 (6.7 - 65.2)	75.2 (72.8 - 77.5)	0.2
	≥25	4 (0.3)	0.3 (0.0 - 1.7)	99.7 (99.1 - 99.9)	1.01 (0.11 - 9.69)	1.00 (0.99 - 1.01)	25.0 (0.6 - 80.6)	75.2 (72.8 - 77.5)	0.0
	≥26	2 (0.2)	0.0 (0.0 - 1.1)	99.8 (99.3 - 100.0)	0.00 (0.00 - 0.00)	1.00 (1.00 - 1.00)	0.0 (0.0 - 84.2)	75.2 (72.8 - 77.5)	-0.2
	≥27	2 (0.2)	0.0 (0.0 - 1.1)	99.8 (99.3 - 100.0)	0.00 (0.00 - 0.00)	1.00 (1.00 - 1.00)	0.0 (0.0 - 84.2)	75.2 (72.8 - 77.5)	-0.2
PHQ-8	≥1	1244 (92.4)	99.4 (97.9 - 99.9)	10.0 (8.2 - 12.0)	1.10 (1.08 - 1.13)	0.06 (0.01 - 0.24)	26.7 (24.2 - 29.2)	98.1 (93.2 - 99.8)	9.4
	≥2	1141 (84.7)	98.2 (96.1 - 99.3)	19.7 (17.3 - 22.3)	1.22 (1.18 - 1.27)	0.09 (0.04 - 0.20)	28.7 (26.1 - 31.5)	97.1 (93.8 - 98.9)	17.9
	≥3	998 (74.1)	95.5 (92.7 - 97.5)	33.0 (30.1 - 36.0)	1.42 (1.36 - 1.50)	0.14 (0.08 - 0.23)	32.0 (29.1 - 35.0)	95.7 (93.0 - 97.6)	28.5
	≥4	846 (62.8)	91.3 (87.8 - 94.1)	46.6 (43.5 - 49.7)	1.71 (1.60 - 1.83)	0.19 (0.13 - 0.27)	36.1 (32.8 - 39.4)	94.2 (91.8 - 96.1)	37.9
	≥5	701 (52.0)	85.6 (81.4 - 89.2)	59.0 (55.9 - 62.1)	2.09 (1.92 - 2.28)	0.24 (0.19 - 0.32)	40.8 (37.1 - 44.5)	92.6 (90.3 - 94.5)	44.7
	≥6	611 (45.4)	79.9 (75.2 - 84.1)	66.0 (63.0 - 69.0)	2.35 (2.13 - 2.61)	0.30 (0.24 - 0.38)	43.7 (39.7 - 47.7)	90.9 (88.6 - 92.9)	46.0
	≥7	**521 (38.7)**	**74.3 (69.2 - 78.9)**	**73.1 (70.2 - 75.8)**	**2.76 (2.44 - 3.10)**	**0.35 (0.29 - 0.42)**	**47.6 (43.2 - 52.0)**	**89.6 (87.3 - 91.6)**	**47.4**

	≥8	428 (31.8)	65.6 (60.2 - 70.7)	79.4 (76.7 - 81.8)	3.18 (2.75 - 3.67)	0.43 (0.37 - 0.50)	51.2 (46.3 - 56.0)	87.5 (85.2 - 89.6)	44.9
	≥9	357 (26.5)	58.4 (52.9 - 63.7)	84.0 (81.6 - 86.2)	3.65 (3.09 - 4.32)	0.50 (0.44 - 0.56)	54.6 (49.3 - 59.9)	86.0 (83.6 - 88.1)	42.4
	≥10	286 (21.2)	50.0 (44.5 - 55.5)	88.3 (86.1 - 90.2)	4.26 (3.48 - 5.20)	0.57 (0.51 - 0.63)	58.4 (52.4 - 64.2)	84.3 (81.9 - 86.4)	38.3
	≥11	233 (17.3)	42.2 (36.9 - 47.7)	90.9 (89.0 - 92.6)	4.65 (3.69 - 5.86)	0.64 (0.58 - 0.70)	60.5 (53.9 - 66.8)	82.7 (80.3 - 84.9)	33.1
	≥12	189 (14.0)	34.7 (29.6 - 40.1)	92.8 (91.0 - 94.3)	4.82 (3.70 - 6.28)	0.70 (0.65 - 0.76)	61.4 (54.0 - 68.4)	81.2 (78.8 - 83.4)	27.5
	≥13	138 (10.2)	27.2 (22.5 - 32.4)	95.4 (93.9 - 96.6)	5.87 (4.22 - 8.17)	0.76 (0.71 - 0.82)	65.9 (57.4 - 73.8)	79.9 (77.5 - 82.1)	22.6
	≥14	103 (8.1)	21.9 (17.5 - 26.7)	96.4 (95.1 - 97.5)	6.15 (4.21 - 8.99)	0.81 (0.76 - 0.86)	67.0 (57.3 - 75.7)	78.9 (76.5 - 81.2)	18.3
	≥15	80 (5.9)	15.6 (11.9 - 19.9)	97.2 (96.0 - 98.2)	5.63 (3.62 - 8.77)	0.87 (0.83 - 0.91)	65.0 (53.5 - 75.3)	77.7 (75.3 - 80.0)	12.8
	≥16	58 (4.3)	10.5 (7.4 - 14.3)	97.7 (96.6 - 98.6)	4.62 (2.77 - 7.70)	0.92 (0.88 - 0.95)	60.3 (46.6 - 73.0)	76.8 (74.4 - 79.1)	8.2
	≥17	46 (3.4)	7.2 (4.7 - 10.5)	97.8 (96.7 - 98.6)	3.31 (1.88 - 5.82)	0.95 (0.92 - 0.98)	52.2 (36.9 - 67.1)	76.2 (73.8 - 78.5)	5.0
	≥18	36 (2.7)	5.7 (3.5 - 8.7)	98.3 (97.3 - 99.0)	3.39 (1.78 - 6.44)	0.96 (0.93 - 0.99)	52.8 (35.5 - 69.6)	76.0 (73.6 - 78.3)	4.0
	≥19	30 (2.2)	4.8 (2.8 - 7.7)	98.6 (97.7 - 99.2)	3.47 (1.71 - 7.03)	0.97 (0.94 - 0.99)	53.3 (34.3 - 71.7)	75.9 (73.4 - 78.1)	3.4
	≥20	24 (1.8)	3.0 (1.4 - 5.4)	98.6 (97.7 - 99.2)	2.17 (0.97 - 4.83)	0.98 (0.96 - 1.00)	41.7 (22.1 - 63.4)	75.5 (73.1 - 77.8)	1.6
	≥21	15 (1.1)	1.5 (0.5 - 3.5)	99.0 (98.2 - 99.5)	1.52 (0.52 - 4.41)	0.99 (0.98 - 1.01)	33.3 (11.8 - 61.6)	75.3 (72.9 - 77.6)	0.5
	≥22	7 (0.5)	0.6 (0.1 - 2.1)	99.5 (98.9 - 99.8)	1.21 (0.24 - 6.22)	1.00 (0.99 - 1.01)	28.6 (3.7 - 71.0)	75.2 (72.8 - 77.5)	0.1
	≥23	4 (0.3)	0.3 (0.0 - 1.7)	99.7 (99.1 - 99.9)	1.01 (0.11 - 9.69)	1.00 (0.99 - 1.01)	25.0 (0.6 - 80.6)	75.2 (72.8 - 77.5)	0.0
	≥24	2 (0.2)	0.0 (0.0 - 1.1)	99.8 (99.3 - 100.0)	0.00 (0.00 - 0.00)	1.00 (1.00 - 1.00)	0.0 (0.0 - 84.2)	75.2 (72.8 - 77.5)	-0.2
PHQ-9 algorithm		183 (13.6)	34.7 (29.6 - 40.1)	93.4 (91.7 - 94.8)	5.25 (3.99 - 6.91)	0.70 (0.65 - 0.76)	63.4 (56.0 - 70.4)	81.3 (78.9 - 83.5)	28.1
PHQ-9 adjusted algorithm		601 (44.6)	78.1 (73.3 - 82.5)	66.4 (63.4 - 69.3)	2.33 (2.10 - 2.58)	0.33 (0.27 - 0.41)	43.4 (39.4 - 47.5)	90.2 (87.9 - 92.3)	44.5
PHQ-2	≥1	1052 (78.1)	95.2 (92.3 - 97.2)	27.5 (24.8 - 30.4)	1.31 (1.26 - 1.37)	0.17 (0.11 - 0.28)	30.2 (27.5 - 33.1)	94.6 (91.3 - 96.9)	22.7
	≥2	**730 (54.2)**	**84.7 (80.4 - 88.4)**	**55.9 (52.8 - 59.0)**	**1.92 (1.77 - 2.09)**	**0.27 (0.21 - 0.35)**	**38.8 (35.2 - 42.4)**	**91.7 (89.3 - 93.8)**	**40.6**
	≥3	358 (26.6)	55.4 (49.9 - 60.8)	82.9 (80.5 - 85.2)	3.24 (2.75 - 3.83)	0.54 (0.48 - 0.61)	51.7 (46.4 - 57.0)	84.9 (82.6 - 87.1)	38.3
	≥4	225 (16.7)	39.5 (34.2 - 45.0)	90.8 (88.9 - 92.5)	4.30 (3.40 - 5.44)	0.67 (0.61 - 0.73)	58.7 (51.9 - 65.2)	82.0 (79.6 - 84.2)	30.3
	≥5	99 (7.4)	19.2 (15.1 - 23.8)	96.5 (95.2 - 97.6)	5.55 (3.74 - 8.22)	0.84 (0.79 - 0.88)	64.6 (54.4 - 74.0)	78.4 (76.0 - 80.6)	15.7
	≥6	63 (4.7)	12.0 (8.7 - 15.9)	97.7 (96.6 - 98.6)	5.27 (3.21 - 8.68)	0.90 (0.87 - 0.94)	63.5 (50.4 - 75.3)	77.1 (74.7 - 79.4)	9.7

Note: +LR = Positive likelihood ratio. −LR = Negative likelihood ratio. PPV= positive predictive value. NPV= negative predictive value. Youden's Index = Sensitivity+Specificity-1. 95%CI = 95% confidence interval. PHQ = Patient Health Questionnaire. Bold values represent the cohort point with the highest Youden Index.

Supplementary material 6. Sensitivity, specificity, +LR, −LR, PPV, NPV, and Youden's index for different GAD cut-off points compared to the gold standard.

Method	Cut-off Points	n (%)	Sensitivity (95%CI)	Specificity (95%CI)	+LR (95%CI)	-LR (95%CI)	PPV (95%CI)	NPV (95%CI)	Youden Index
GAD-7	≥1	1244 (92.4)	96.4 (81.7 - 99.9)	7.7 (6.4 - 9.3)	1.05 (0.97 - 1.12)	0.46 (0.07 - 3.19)	2.2 (1.4 - 3.1)	99.0 (94.7 - 100.0)	4.1
	≥2	1128 (83.7)	96.4 (81.7 - 99.9)	16.5 (14.6 - 18.6)	1.16 (1.07 - 1.25)	0.22 (0.03 - 1.49)	2.4 (1.6 - 3.5)	99.5 (97.5 - 100.0)	12.9
	≥3	960 (71.3)	96.4 (81.7 - 99.9)	29.3 (26.8 - 31.8)	1.36 (1.26 - 1.48)	0.12 (0.02 - 0.84)	2.8 (1.9 - 4.1)	99.7 (98.6 - 100.0)	25.7
	≥4	797 (59.2)	89.3 (71.8 - 97.7)	41.5 (38.8 - 44.2)	1.53 (1.33 - 1.75)	0.26 (0.09 - 0.75)	3.1 (2.0 - 4.6)	99.5 (98.4 - 99.9)	30.8
	≥5	640 (47.5)	71.4 (51.3 - 86.8)	53.0 (50.3 - 55.7)	1.52 (1.19 - 1.93)	0.54 (0.30 - 0.97)	3.1 (1.9 - 4.8)	98.9 (97.8 - 99.5)	24.4
	≥6	491 (36.5)	60.7 (40.6 - 78.5)	64.1 (61.4 - 66.7)	1.69 (1.24 - 2.30)	0.61 (0.39 - 0.97)	3.5 (2.0 - 5.5)	98.7 (97.7 - 99.4)	24.8
	≥7	360 (26.7)	53.6 (33.9 - 72.5)	73.8 (71.4 - 76.2)	2.05 (1.43 - 2.93)	0.63 (0.42 - 0.94)	4.2 (2.4 - 6.8)	98.7 (97.8 - 99.3)	27.4
	≥8	**295 (21.9)**	**53.6 (33.9 - 72.5)**	**78.8 (76.5 - 81.0)**	**2.52 (1.76 - 3.62)**	**0.59 (0.40 - 0.88)**	**5.1 (2.9 - 8.3)**	**98.8 (97.9 - 99.3)**	**32.4**
	≥9	230 (17.1)	42.9 (24.5 - 62.8)	83.5 (81.4 - 85.4)	2.59 (1.66 - 4.04)	0.69 (0.50 - 0.94)	5.2 (2.7 - 8.9)	98.6 (97.7 - 99.2)	26.4
	≥10	164 (12.2)	39.3 (21.5 - 59.4)	88.4 (86.5 - 90.1)	3.39 (2.09 - 5.50)	0.69 (0.51 - 0.93)	6.7 (3.4 - 11.7)	98.6 (97.7 - 99.2)	27.7
	≥11	120 (8.9)	39.3 (21.5 - 59.4)	91.7 (90.1 - 93.2)	4.75 (2.90 - 7.79)	0.66 (0.49 - 0.89)	9.2 (4.7 - 15.8)	98.6 (97.8 - 99.2)	31.0
	≥12	94 (7.0)	28.6 (13.2 - 48.7)	93.5 (92.0 - 94.8)	4.38 (2.36 - 8.15)	0.76 (0.60 - 0.97)	8.5 (3.8 - 16.1)	98.4 (97.5 - 99.0)	22.1
	≥13	67 (5.0)	28.6 (13.2 - 48.7)	95.5 (94.3 - 96.6)	6.39 (3.39 - 12.10)	0.75 (0.59 - 0.95)	11.9 (5.3 - 22.2)	98.4 (97.6 - 99.0)	24.1
	≥14	50 (3.7)	21.4 (8.3 - 41.0)	96.7 (95.5 - 97.6)	6.42 (2.98 - 13.80)	0.81 (0.67 - 0.99)	12.0 (4.5 - 24.3)	98.3 (97.4 - 98.9)	18.1
	≥15	40 (3.0)	14.3 (4.0 - 32.7)	97.3 (96.2 - 98.1)	5.23 (2.00 - 13.70)	0.88 (0.76 - 1.03)	10.0 (2.8 - 23.7)	98.2 (97.3 - 98.8)	11.6
	≥16	31 (2.3)	10.7 (2.3 - 28.2)	97.9 (96.9 - 98.6)	5.05 (1.63 - 15.60)	0.91 (0.80 - 1.04)	9.7 (2.0 - 25.8)	98.1 (97.2 - 98.8)	8.6
	≥17	19 (1.4)	3.6 (0.1 - 18.3)	98.6 (97.9 - 99.2)	2.62 (0.36 - 18.90)	0.98 (0.91 - 1.05)	5.3 (0.1 - 26.0)	98.0 (97.1 - 98.7)	2.2
	≥18	15 (1.1)	0.0 (0.0 - 12.3)	98.9 (98.1 - 99.4)	0.00 (0.00 - 0.00)	1.01 (1.01 - 1.02)	0.0 (0.0 - 21.8)	97.9 (97.0 - 98.6)	-1.1
	≥19	12 (0.9)	0.0 (0.0 - 12.3)	99.1 (98.4 - 99.5)	0.00 (0.00 - 0.00)	1.01 (1.00 - 1.01)	0.0 (0.0 - 26.5)	97.9 (97.0 - 98.6)	-0.9
	≥20	8 (0.6)	0.0 (0.0 - 12.3)	99.4 (98.8 - 99.7)	0.00 (0.00 - 0.00)	1.01 (1.00 - 1.01)	0.0 (0.0 - 36.9)	97.9 (97.0 - 98.6)	-0.6
	≥21	5 (0.4)	0.0 (0.0 - 12.3)	99.6 (99.1 - 99.9)	0.00 (0.00 - 0.00)	1.00 (1.00 - 1.01)	0.0 (0.0 - 52.2)	97.9 (97.0 - 98.6)	-0.4
GAD-2	≥1	1028 (76.3)	89.3 (71.8 - 97.7)	24.0 (21.7 - 26.4)	1.17 (1.03 - 1.34)	0.45 (0.15 - 1.31)	2.4 (1.6 - 3.6)	99.1 (97.3 - 99.8)	13.3
	≥2	**681 (50.6)**	**82.1 (63.1 - 93.9)**	**50.1 (47.4 - 52.8)**	**1.65 (1.37 - 1.97)**	**0.36 (0.16 - 0.79)**	**3.4 (2.2 - 5.0)**	**99.2 (98.3 - 99.8)**	**32.2**
	≥3	283 (21.0)	42.9 (24.5 - 62.8)	79.5 (77.2 - 81.6)	2.09 (1.34 - 3.24)	0.72 (0.52 - 0.99)	4.2 (2.2 - 7.3)	98.5 (97.6 - 99.1)	22.4
	≥4	141 (10.5)	25.0 (10.7 - 44.9)	89.8 (88.1 - 91.4)	2.46 (1.27 - 4.77)	0.84 (0.67 - 1.03)	5.0 (2.0 - 10.0)	98.3 (97.4 - 98.9)	14.8
	≥5	51 (3.8)	14.3 (4.0 - 32.7)	96.4 (95.3 - 97.4)	4.01 (1.55 - 10.40)	0.89 (0.76 - 1.03)	7.8 (2.2 - 18.9)	98.1 (97.3 - 98.8)	10.7
	≥6	28 (2.1)	7.1 (0.9 - 23.5)	98.0 (97.1 - 98.7)	3.62 (0.90 - 14.50)	0.95 (0.86 - 1.05)	7.1 (0.9 - 23.5)	98.0 (97.1 - 98.7)	5.1

Note: +LR = Positive likelihood ratio. −LR = Negative likelihood ratio. PPV= positive predictive value. NPV= negative predictive value. Youden's Index = Sensitivity+Specificity-1. 95%CI = 95% confidence interval. GAD =Generalized Anxiety Disorder Scale. Bold values represent the cut-off with the highest Youden Index.

BMJ Open

Sensitivity and specificity of the Patient Health Questionnaire (PHQ-9, PHQ-8, PHQ-2) and General Anxiety Disorder scale (GAD-7, GAD-2) for depression and anxiety diagnosis: A cross-sectional study in a Peruvian hospital population

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2023-076193.R1
Article Type:	Original research
Date Submitted by the Author:	08-Aug-2023
Complete List of Authors:	Villarreal-Zegarra, David; Instituto Peruano de Orientación Psicológica, ; Universidad Peruana Cayetano Heredia, CRONICAS Center of Excellence in Chronic Diseases Barrera-Begazo, Juan ; Instituto Peruano de Orientación Psicológica Otazú-Alfaro, Sharlyn; Instituto Peruano de Orientación Psicológica Mayo-Puchoc, Nikol; Instituto Peruano de Orientación Psicológica Bazo-Alvarez, Juan Carlos ; Universidad Norbert Wiener Huarcaya-Victoria , Jeff; Universidad de San Martín de Porres Facultad de Medicina Humana,
Primary Subject Heading:	Mental health
Secondary Subject Heading:	Mental health
Keywords:	Sensitivity and Specificity, Depression & mood disorders < PSYCHIATRY, Anxiety disorders < PSYCHIATRY

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Sensitivity and specificity of the Patient Health Questionnaire (PHQ-9, PHQ-8, PHQ-2) and General Anxiety Disorder scale (GAD-7, GAD-2) for depression and anxiety diagnosis: A cross-sectional study in a Peruvian hospital population

Authors

David Villarreal-Zegarra [1, 2]
Juan Barrera [2]
Sharlyn Otazú-Alfaro [2]
Nikol Mayo-Puchoc [2]
Juan Carlos Bazo-Alvarez [3]
Jeff Huarcaya-Victoria [4,5]

Affiliation

[1] Universidad César Vallejo, Escuela de Medicina. Trujillo, Peru
[2] Instituto Peruano de Orientación Psicológica, Lima, Peru
[3] Universidad Privada Norbert Wiener, Lima, Peru
[4] Escuela Profesional de Medicina Humana, Universidad Privada San Juan Bautista, Filial Ica, Peru
[5] Unidad de Psiquiatría de Enlace, Departamento de Psiquiatría, Hospital Nacional Guillermo Almenara Irigoyen, EsSalud, Lima, Peru

Email

David Villarreal-Zegarra davidvillarreal@ipops.pe
Juan Barrera juanbarrera@ipops.pe
Sharlyn Otazú-Alfaro sharlynotazu@ipops.pe
Nikol Mayo-Puchoc nikolmayo@ipops.pe
Juan Carlos Bazo-Alvarez juan.bazo@uwiener.edu.pe
Jeff Huarcaya-Victoria jhuarcayav@usmp.pe

Corresponding author

Jeff Huarcaya-Victoria
jeff.huarcaya@upsjb.edu.pe

Abstract

Objectives : The Patient Health Questionnaire (PHQ) and Generalized Anxiety Disorder Scale (GAD) are widely used screening tools, but their sensitivity and specificity in low- and middle-income countries are lower than in high-income countries. We conducted a study to determine the sensitivity and specificity of different versions of these scales in a Peruvian hospital population.

Design: Our study has a cross-sectional design.

Setting: Our participants are hospitalised patients in a Peruvian hospital. The gold standard was a clinical psychiatric interview following ICD-10 criteria for depression (F32.0, F32.1, F32.2, and F32.3) and anxiety (F41.0 and F41.1).

Participants: The sample included 1347 participants. A total of 334 participants (24.8%) were diagnosed with depression, and 28 participants (2.1%) were diagnosed with anxiety.

Results: We found that a cut-off point of ≥7 in the PHQ-9 and PHQ-8 had the best balance between sensitivity and specificity. The raw score for PHQ-9 with cohort point ≥7 showed better results than the algorithm method or the adjusted algorithm. The PHQ-2 had the best cut-off point of ≥2 points. We present the sensitivity and specificity values of each cut-off point in GAD-7 and GAD-2. All scales had good internal consistency (>0.70), and the PHQ-9, PHQ-8, and GAD-7 had appropriate goodness-of-fit indices.

Conclusions: The PHQ and GAD have adequate measurement properties in their different versions. We present specific cut-offs for each version.

Keywords: Anxiety; Depression; Patient Health Questionnaire; Peru; Sensitivity and Specificity.

Strengths and limitations of this study

- Study methods allowed us to establish clinically meaningful cut-off points for PHQ and GAD.
- Sample size was larger than in other similar studies and large enough to support all analyses and conclusions.
- Research findings may not be directly applicable to some hospital or primary care settings due to the specific context of our study population.

Background

Until 2019, approximately 280 million people worldwide suffered from depression and 302 million from anxiety [1]. These data reveal that both mental disorders are the most common in the world and lead to the causes of the global burden of mental health disability-adjusted life years[2, 3]. With the onset of the coronavirus disease (COVID-19) pandemic, the worldwide prevalence of both disorders increased by around 25% [4]. In Peru, during the COVID-19 pandemic, the prevalence of moderate depressive symptoms also increased by approximately 0.17% in each quarter [5]. However, no population-level evidence has been found about the prevalence of anxious symptomatology or the diagnosis of anxiety in Peru. In this context, the impact of COVID-19 on the prevalence and burden of major depression and anxiety disorders was measured using screening tools [6]. In addition, it was noted that during the pandemic, there was a reduction in the number of mental health service users being seen [7].

Screening tools assist in early diagnosis and intervention that can prevent disease progression and reduce years lost to disability [8]. They are beneficial in contexts with limited mental health professionals providing care to large populations, such as in Peru. The opportune identification of people at risk of depression reduces treatment costs and disease burden [9-11]. Depressive symptom screening is also helpful in national surveys and epidemiological research [12] since, unlike diagnostic instruments, screening measures are typically brief, quick, and easy to administer [13, 14]. Internationally, the most used screening instruments for depressive and anxious symptomatology are the Patient Health Questionnaire (PHQ-9) [15], PHQ-8 [16], PHQ-2 [17], Generalized Anxiety Disorder (GAD-7) [18], GAD-2 [18], Depression, Anxiety and Stress Scale (DASS-21), Kessler scale-10, Hospital Anxiety and Depression Scale (HADS) [19], Five Well-Being Index (WHO-5) [10]. Most have been validated in several countries, but only the PHQ and GAD have been validated in the Peruvian context [20, 21].

In particular, the PHQ versions (PHQ-9, PHQ-8, PHQ-2) and GAD versions (GAD-7, GAD-2) are the most widely used, having extensive evidence of their validity and reliability [22-24]. However, correctly identifying people at risk of depression or anxiety requires more than internal/externally valid and reliable screening measures; defining an accurate cut-off point for their raw scales (i.e., to reach valid interpretations) is also necessary. Such a cut-off point can vary across cultures and sub-populations (e.g., general versus clinical), so a local calibration is usually needed [25]. Studies of the different versions of the PHQ and GAD have yielded heterogeneous cut-offs, as they vary between different cultures [21, 26-29] and populations, such as clinical and general populations [30-32]. However, several systematic reviews suggest that cut-off 10 is most appropriate for the PHQ-9, PHQ-8, and GAD-7 [33-37], and cut-offs 2-3 for the PHQ-2 and GAD-2 [35, 37]. Furthermore, concerning the PHQ-9 correctness, the summed item score method is the most used compared to the algorithm. However, other forms of correction using diagnostic algorithms are available [38, 39].

Sensitivity and specificity studies have been barely performed in low- and middle-income countries [40]. Several of these populations do not count with verified cut-off points from calibration studies (including Peruvian populations), in particular, the inpatient population is particularly vulnerable as

they have physical comorbidities that may influence the establishment of cohort points. Therefore, our aim was to determine the optimal cut-off point for the PHQ-9, PHQ-8, PHQ-2, GAD-7, and GAD-2 to discriminate a formal depression and anxiety diagnosis in the Peruvian hospital population. In addition, as secondary objectives, we assessed these scales' internal structure and reliability.

Methods

Study design

This study has a cross-sectional design, and we used the STARD 2015 guidelines for reporting diagnostic accuracy studies [41].

Participants

The participants were patients from the Liaison Psychiatry Unit of a hospital in Lima, Peru. Psychiatric liaison services provide psychiatric consultation to hospitalised patients with medical or surgical conditions that have a co-existing psychiatric illness or need for psychiatric assessment and management. The total number of participants in our study is similar to the proportion of people who were hospitalised in 2022 in our setting (see supplementary material 1). The evaluation period started in September 2020 and finished in August 2022. Sampling was non-probabilistic and applied to all participants arriving at the Liaison Psychiatry Unit. The inclusion criteria were that they had complete PHQ-9 and GAD-7 data and were of legal age (>18 years). Participants with missing data were excluded.

The sample size calculation for the PHQ versions was based on an estimated sensitivity of 0.88 and specificity of 0.85 [33], a confidence level of 95%, a prevalence of 6.4% [42, 43], and a drop-out rate of 10%, giving an estimate of 705 participants. The sample size calculation for the GAD versions was based on an estimated sensitivity of 0.83 and specificity of 0.84 [18], a confidence level of 95%, a prevalence of 8.7% [44], and a drop-out rate of 10%, giving an estimate of 694 participants. The web program based on the paper by Burderer (1996) was used to calculate the sample size [45].

Setting

The Guillermo Almenara Irigoyen National Hospital (HNGAI) was the study site, a highly complex hospital in Lima-Peru (capital city). HNGAI is one of the three largest hospitals of the Social Security system in Peru based on the number of beds (960 hospital beds) and is also a tertiary referral centre for all medical specialities, including psychiatry (http://www.essalud.gob.pe/estadistica-institucional/). It provides health care services to 1,547,840 individuals from social insurance. Because it attends to virtually all pathologies, from the simplest to the most complex, it was classified in 2015 as a Specialized Health Institute III-2, the highest level awarded by the Ministry of Health of Peru to hospital establishments.

The Liaison Psychiatry Unit at HNGAI is responsible for responding to consultation requests from different clinical-surgical services at HNGAI [46]. As part of the evaluation of each patient, in addition to the clinical interview and psychiatric diagnosis, standardised assessments such as the PHQ-9 and GAD-7 are used to ensure adequate monitoring and assess response to the established treatment. Since September 2020, the services provided by the Liaison Unit have been recorded in a Google Form to track better the patients treated.

Instruments and variables

Patient Health Questionnaire (PHQ-9, PHQ-8, and PHQ-2)

The Patient Health Questionnaire is an instrument designed to measure depressive symptoms over the past two weeks, according to the diagnostic criteria of the DSM-IV, criteria that were retained in the DSM-5. The scale has four response options (0=no days, 1=some days, 2=more than half of the days, 3=almost every day) [15]. The scale had many versions, including the PHQ-9, the full version with nine items and scores ranging from 0 to 27. In Peru, the PHQ-9 had good psychometric properties in terms of structural validity (CFI = 0.936; RMSEA = 0.089; SRMR = 0.039), internal consistency ($\alpha = \omega = 0.87$) and invariance between age and sex ($\Delta CFI<0.01$) [20].

In addition, PHQ-9 had scoring versions related to the DSM-5 indicators, which state that for a case to be positive, there must be at least five depressive symptoms present, and at least one of them must be core depressive symptoms (item 1 and item 2). First, the PHQ-9 algorithm suggests that a symptom is positive if it scores two or more, except the ninth item, suicidal ideation, which is positive if it scores one or more [47]. Second, the PHQ-9 adjusted algorithm proposes that a symptom was positive if it scored 1 or more for any of the items in the instrument [48].

The PHQ-8 was a shortened version of the PHQ-9 without the last item on suicidal ideation [16]. The PHQ-8 was as valuable as the PHQ-9 in detecting cases of major depression [49]. The PHQ-2 is an abbreviated version of the PHQ-9 with only two items, focusing on the first two items related to the core symptoms of depression (anhedonia and depressed mood) and providing scores between 0 and 6. The PHQ-2 was validated in Peru and showed adequate levels of internal consistency ($\alpha = .80$) [50].

General Anxiety Disorder Scale (GAD-7 and GAD-2)

The General Anxiety Disorder Scale was a Likert-type rating scale with four response options ranging from 0 (not at all) to 3 (almost every day), based on DSM-IV criteria and assesses anxious symptoms during the past two weeks [51]. The GAD-7 was the version of the instrument with the original seven items and had a range of scores from 0 to 21. The GAD-7 had good psychometric properties in the Peruvian context for a one-dimensional model (CFI = .995, TLI = .992, RMSEA = .056), adequate internal consistency ($\omega = .89$), and invariance according to sex ($\Delta CFI \leq .01$) [52].

The GAD-2 was adapted from the GAD-7, focusing on the emotional and cognitive expressions of DSM-IV anxiety (items 1 and 2) [53]. The GAD-2 shows good internal consistency values (ω =.80) and a relationship with its extended version (r>0.80) in Peruvian context [52].

Gold standard

The gold standard was an individual clinical psychiatric interview following ICD-10 criteria. The clinical assessments were performed by psychiatrists who are members of the Liaison Psychiatry Unit, all of whom have at least five years of clinical experience evaluating the psychiatric needs of hospitalised patients. The interview focused on assessing whether the participants had depressive disorder (F32.0, F32.1, F32.2, and F32.3) or anxiety disorder (F41.0 and F41.1), with a duration between 25 to 30 minutes. The individual clinical psychiatric interview and the psychometric instruments (i.e., PHQ and GAD) were independently applied on the same day, the latter by a mental health nurse or a psychologist and the former by a psychiatrist. The average time between both measurements was 15 minutes (standard deviation = 4.5 minutes), and the order (i.e., psychometric instruments before or after the interview) was randomly assigned.

Sociodemographic covariates

Data was collected on sex (male, female), age, marital status (Single, Married/Cohabitant, Separated, Widowed), educational level (None, Elementary, High School, Technical, College), currently works (No, yes, retired), living alone (Yes, No), and history of psychiatric diagnosis (yes/no). In addition, information was collected on the physical diagnosis of the participants based on the ICD-10.

Statistical analysis

The sociodemographic covariates of the participants were described at frequency and percentage levels. The internal consistency and internal structure analyses were performed with R Studio, with the “Lavaan”, “Semtools”, and “Semplot” packages (see supplementary material 2). Sensitivity, specificity, and correlation analyses were analysed with Stata 15 (see supplementary material 3).

Sensibility and Specificity

The PHQ-9, PHQ-8, PHQ-9 algorithm, PHQ-9 adjusted algorithm, and PHQ-2 were evaluated as diagnostic tests and compared against the gold standard. In addition, the GAD-7 and GAD-2 were scored and compared against the diagnosis of anxiety through the clinical interview (gold standard).

We calculated the positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (+LR), negative likelihood ratio (-LR), and Youden Index. PPV and NPV refer to the proportion of patients correctly diagnosed as positive or negative, respectively [54]. The LR+ is the probability that a person with the disease will test positive given the probability that a person without the disease will test positive [55]. While the LR- is the probability that a person with the disease will test negative given the probability that a person without the disease will test negative [55]. The Youden Index is a measure that summarizes the performance of a diagnostic test by interpreting it as the probability that the selected cut-off point provides an adequate clinical decision (in terms of sensitivity and specificity), as

opposed to the probability that the selected cut-off provides a random decision [54]. The maximum value of the Youden Index was used as a criterion to select the cut-off with the best diagnostic performance for each scale. Values closer to 1 were considered optimal, and those closer to 0 were considered inadequate.

Internal structure

Confirmatory factor analysis (CFA) was performed considering a one-dimensional model for the PHQ-9, PHQ-8, and GAD-7. We used the weighted least square mean and variance adjusted (WLSMV) estimator [56] and polychoric matrices as it best fits the categorical-ordinal nature of the data [57]. Models were evaluated using a set of goodness-of-fit indices such as CFI and TLI, which must be greater than 0.95 to be considered adequate [58]. In addition, the SRMR and RMSEA at 90% confidence were estimated, which must have values less than 0.08 to be considered adequate [58]. It was impossible to perform a CFA for the PHQ-2 and the GAD-2 because a minimum of three items are required for such analysis.

Internal consistency

We calculated the alpha (α) and McDonald's omega coefficients (ω). Values greater than 0.70 are considered adequate [59].

Ethics

The Hospital Nacional Guillermo Almenara Irigoyen's Institutional Review Board (Nota N°52 CIEI-OIyD-GRPA-Essalud-2023) approved the protocol of our study. Throughout the study, the researchers had no access to identifying information about the participants. In addition, participants gave informed consent. All participants were users of the hospital's Liaison Psychiatry Unit and received psychological or psychiatric care as needed.

Patient and Public Involvement

No patient involved.

Results

Participants

We collected data from 4979 attendances performed within the liaison psychiatry service during the study period. However, some of these attendances were not assessed with PHQ-9 or GAD-7 data (n=3484) or lacked sociodemographic information (n=148) and were eliminated (see supplementary material 4). Thus, our study only included 1347 participants (see Table 1). Most participants were female (59.4%; n=800), married or living with a partner (57.0%; n=768), and had higher technical or university education (53.5%; n=721). A total of 334 participants (24.8%) were diagnosed with

depression, and 28 participants (2.1%) were diagnosed with anxiety, as determined through individual psychiatric interviews conducted based on the ICD-10 criteria.

The most common physical morbidities were cardiovascular diseases (n=111; 8.2%), endocrine, nutritional and metabolic diseases (n=130; 9.7%) and neoplasms, diseases of the blood and haematopoietic organs and other diseases affecting the mechanism of immunity (n=348; 25.8%).

Sensibility and Specificity

In supplementary material 5, we provide the values of all cut-off points for the different versions of the PHQ. The cut-off points ≥7 in the PHQ-9 had the best balance between sensitivity and specificity of all the cut-off points evaluated in the various versions of the PHQ, as it obtained a sensitivity of 76.0 (95%CI: 71.1 - 80.5) and specificity of 72.1 (95%CI: 69.2 - 74.8) (see supplementary material 6). In addition, the PHQ-9 with a cut-off of ≥10 points (i.e., the most used) showed lower levels of sensitivity (54.2; 95%CI: 8.7 - 59.6), but higher level of specificity (87.4; 95%CI: 85.2 - 89.3), compared to the cut-off point of ≥7.

The algorithm score method for PHQ-9 had low levels of sensitivity (34.7; 95%CI: 29.6 - 40.1) but high levels of specificity (93.4; 95%CI: 91.7 - 94.8) compared to the raw score method for PHQ-9 with ≥7 cohort points. In contrast, the adjusted algorithm method for PHQ-9 showed slightly higher sensitivity values (78.1; 95%CI: 73.3 - 82.5) and better specificity values (66.4; 95%CI: 63.4 - 69.3) compared to the raw score method for PHQ-9 with ≥7 cohort points. The raw score for PHQ-9 with cohort point ≥7 showed a better balance between sensitivity and specificity compared to the algorithm method or the algorithm adjusted for PHQ-9.

The best cut-off point found in the PHQ-8 was ≥7 points, as it had a sensitivity of 79.9 (95%CI: 75.2 - 84.1), and a specificity of 66.0 (95%CI: 63.0 - 69.0) (see supplementary material 6). The best cut-off point found in the PHQ-2 was ≥2 points, as it had a sensitivity of 84.7 (95%CI: 80.4 - 88.4), and a specificity of 55.9 (95%CI: 52.8 - 59.0) (see supplementary material 6).

Because we have a small number of cases with truly anxious people, any changes in the scores of these people could lead to large changes in sensitivity and specificity. Therefore, it is not possible to give an optimal cohort score over the rest, but we present all cohort scores in Supplementary Material 7. In particular, the cut-off point ≥8 had good performance for GAD-7 with sensitivity values of 53.6 (95%CI: 33.9 - 72.5) and specificity of 78.8 (95%CI: 76.5 - 81.0), (see supplementary material 6). The GAD-7's cut-off point ≥10 (i.e., the most used) had lower levels of sensitivity (39.3; 95%CI: 21.5 - 59.4), but higher levels of specificity (88.4; 95%CI: 86.5 - 90.1, compared to the cut-off point of ≥8. In addition, the cut-off point for the GAD-2 was ≥2 had a sensitivity of 84.7 (95%CI: 80.4 - 88.4), and a specificity of 50.1 (95%CI: 47.4 - 52.8) (see supplementary material 6).

Internal structure

The PHQ-9 one-dimensional model showed adequate goodness-of-fit (X^2=251.9; df=27; CFI=0.974; TLI=0.965; SRMR=0.051; RMSEA[90%CI]=0.079[0.070-0.088]), while the PHQ-8 one-dimensional model reported a similar goodness-of-fit (X^2=202.7; df=20; CFI=0.977; TLI=0.977; SRMR=0.050; RMSEA[90%CI]=0.082[0.072-0.093]). The GAD-7 also showed adequate goodness-of-fit (X^2=122.3; df=14; CFI=0.977; TLI=0.966; SRMR=0.043; RMSEA[90%CI]=0.076[0.064-0.088]).

Reliability

The PHQ-9 (α=0.89; ω=0.86), the PHQ-8 (α=0.88; ω=0.85), and the GAD-7 (α=0.85; ω=0.81) showed optimal internal consistency values. Similarly, the PHQ-2 (α=0.83; ω=0.80) and the GAD-2 (α=0.74; ω=0.70) also showed adequate internal consistency scores. Table 2 shows the raw scores.

Discussion

Main findings

We determined the target population's optimal cut-off points for PHQ scale. The PHQ-9's ≥7 cut-off point showed the highest sensitivity and specificity when contrasted against a psychiatric diagnosis of depression (gold standard). For a similar contrast, the other optimal cut-off points were: ≥7 for the PHQ-8, and ≥2 for the PHQ-2. In addition, the algorithm scoring or algorithm-adjusted scoring methods for the PHQ-9 had a lower balance between sensitivity and specificity scores than the PHQ-9 raw score scoring method with a cut-off ≥7. In the case of GAD, the small number of participants with actual anxiety made it impossible to determine an optimal cut-off point. However, we present the sensitivity and specificity of each cut-off point. We confirmed the adequacy of a one-dimensional model for the PHQ-9, PHQ-8, and GAD-7, while all scales showed good internal consistency.

Contrast to literature

At the PHQ-9 level, evidence suggests that the raw score approach is more valuable than diagnostic algorithms [33], which is consistent with our findings. For the cut-off, different systematic reviews agree that the most commonly used cut-off is ≥10 [33, 60]. The optimal cut-off reported in our study was slightly lower than that suggested by the other studies, and two possible factors could explain this difference. Firstly, our population is inpatients in different areas of a high-complexity hospital. Other studies of hospitalised cancer patients [61], hospitalised neurology patients [62], and patients with coronary heart disease [63] also found an optimal cut-off between 5 and 7 points. Therefore, hospitalised individuals may be more likely to have depressive symptoms, which may require a lower cut-off on the PHQ-9. Second, several studies in populations from low- and middle-income countries have reported cut-offs between 5 and 7, for example, Pakistani migrants in the UK [64], Indian adolescents [65], and primary care in Ethiopia [66]. One reason for the difference in cut-off points between high and low-income countries may be due to cultural factors, as culturally diverse groups do

not achieve invariance between the PHQ-9 and the GAD-7 [67]. Therefore, factors such as social determinants of health present in such countries may influence cut-off.

Concerning the PHQ-8 and PHQ-9, we found that both scales have similar cut-off points (≥ 7). Our findings are consistent with a meta-analysis that found that the cut-offs between the two scales are identical; although sensitivity may be minimally reduced with the PHQ-8, specificity is similar between the two scales [36]. The PHQ-8 does not include the item corresponding to suicidal or self-harming ideation, and the use of this version of the PHQ is common in the general population, as suicidal ideation is less common in this group [16]. However, at the level of clinical populations, it has been found that omitting this item does not significantly alter the measurement capabilities of the PHQ, as the correlation between the PHQ-8 and PHQ-9 in clinical populations is very close to one [68].

Regarding the GAD-7, our findings are consistent with a meta-analysis that evaluated all possible cut-off points and reported that ≥8 is the most appropriate for anxiety disorder [18]. It also notes that scores between 7 and 10 points have similar sensitivity and specificity values [18]. Other recent primary studies conducted in hospitalised populations or people with chronic diseases in hospital settings also found optimal cut-offs between 7 and 10 points [69-71].

Our results on PHQ-2 was in line with meta-analyses supporting the use of the cut-off of 2 for PHQ-2 [35, 72]. Also, the values most frequents for GAD-2 are cut-off ≥2 and ≥3 [18, 37, 73]. The meta-analyses mentioned included studies in general populations (i.e., people attending primary care) and people hospitalised for non-communicable or infectious diseases. However, no meta-analyses were found that evaluated cut-off for hospitalised people only. At the level of primary studies, the evidence suggests that cut-off vary between 2 and 3 points for the PHQ-2 and GAD-2 [74, 75].

Regarding internal validity, a systematic review examined the factor structure of the PHQ-9, noting that the one-dimensional model has been repeatedly confirmed across studies [76]. Although several studies evaluated alternative multidimensional models (e.g., two-dimensional, three-dimensional, or bifactorial models), their dimensions are often highly correlated with each other, so there may be overlapping [76]. We did not find systematic reviews on the internal structure of the GAD-7 and the PHQ-8. However, several studies support the one-dimensional model in hospitalised patients for both the PHQ-8 [77] and GAD-7 [21, 27]. In Peru, the GAD-7 and PHQ-9 have shown evidence of a one-dimensional factor structure in different populations, such as the general population [20], pregnant women [21], and university students [52, 78]. However, no studies have been found evaluating the factor structure of the PHQ-8 in the Peruvian population.

Our study focuses on a hospital-based clinical population with one or more physical morbidities, it is important to consider that our finding of a different cut-off point, equal to or greater than 10 points for PHQ, may be influenced by the characteristics of this specific population. It is relevant to note that other studies conducted in hospital settings have found cut-off points lower than the recommendation of equal to or greater than 10 [79, 80]. It is important to bear in mind that the cut-off point may vary depending on the reference group and the context in which it is applied.

Our study used the Youden index to determine the optimal cut-off, but it is important to consider that the cut-off may vary depending on the sample size. A recent simulation study found that for large samples of more than 1000 participants, the optimal sensitivity and specificity values can vary by up to approximately 2 points from the optimal cut-off in cross-sectional studies [81]. Therefore, while a sample size calculation was performed to ensure adequate power, we cannot rule out the use of a cut-off of 10 or more for the Peruvian population. However, within the study, we present the sensitivity and specificity found for such a cut-off.

Public health implications

The evaluated instruments are widely used in clinical practice and research to measure symptoms of depression and anxiety, but from today, users will have optimal cut-off points for interpretations. This can help healthcare professionals identify people at risk of depression and anxiety more accurately while informing decisions about their formal diagnosis and consequent treatment. This is especially valuable in hospital environments, where time is crucial.

Our findings are of particular interest to the Peruvian health system, which has clinical practice guidelines for depression that recommend the PHQ-9 as a screening tool in primary care and hospital context [82]. Although our results correspond only to a hospital population, our study is the closest approximation to an evaluation of sensitivity and specificity in the Peruvian context, in the absence of similar studies in primary care. On the other hand, there is a lack of national clinical practice guidelines for screening and managing anxiety in Peru. Therefore, our study could contribute to future clinical practice guidelines for generalised anxiety disorder.

Although our study found alternative cut-off points to the standard (cut-off ≥10) for the PHQ-9 and PHQ-8 questionnaires, it is important to note that in certain contexts, higher specificity values (cut-off ≥10) may be necessary. These higher values enable a more accurate identification of individuals without depression or anxiety, thereby reducing the likelihood of false-positive results. This reduction in false positives is particularly crucial for alleviating the burden on the healthcare system. A screening tool with high specificity avoids unnecessary diagnoses and optimizes the use of healthcare resources. Therefore, utilizing a cut-off point of 10 or higher for the PHQ-9, PHQ-8, and GAD-7 can facilitate the early and accurate identification of true cases of depression and anxiety, ensuring that resources are appropriately focused on those who need care and treatment.

Strengths and limitations

Our results of the study have several strengths. First, to our knowledge, this is the first study in a Peruvian context that evaluates the factorial structure of all PHQ and GAD versions in a hospitalised population. Second, the scales were administered by a team of healthcare professionals with more than five years of experience in the clinical assessment of these patients. Third, the sample size was large enough to support all analyses and conclusions. Further, our sample size was larger than other

recently published studies' [60]. Fourth, our study is the first Peruvian study to evaluate the sensitivity and specificity of the PHQ.

Our study has limitations. First, we conducted the study only in a hospital context in a Peruvian city, which limits its applicability to other settings in Peru or other countries. However, it could be used in other Peruvian hospital contexts with similar characteristics, which is relevant because hospital care in Peru (levels II and III of complexity) represents 58.65% of total care [83]. Secondly, the generalisability of our results may be limited because the sampling is not probabilistic, as it does not include other hospitals. However, the hospital where we conducted the study serves 1.1% of all nationally insured EsSalud patients (http://www.essalud.gob.pe/estadistica-institucional/). It is also a national referral hospital, which means that people from all over the country are referred to this hospital for treatment. Therefore, the representativeness of the results is ensured. Thirdly, we used an individual psychiatric interview according to the ICD-10 criteria as a gold standard. We were not able to use the Composite International Diagnostic Interview (CIDI) or the Standardised Clinical Assessment (SCID), more typical gold standards, because of the time constraints involved in conducting such interviews. In Peru, health systems are overburdened, and it is not feasible to have lengthy sessions with highly specialised professionals to conduct such structured interviews. However, based on our experience, we believe that a psychiatric interview is a sufficient benchmark in this context. Fourthly, our study identified a limited number of individuals (n=28) with a diagnosed anxiety condition. Consequently, minor variations in the study cohort could potentially impact the sensitivity or specificity [81]. Nonetheless, we have ensured sufficient statistical power for our analysis based on our sample size calculation. Moreover, all cohort scores on the GAD scale are provided, which can be valuable for future research involving larger numbers of individuals diagnosed with anxiety (refer to supplementary material 7). Fifth, our study allows us to obtain sensitivity and specificity values for users in inpatient mental health settings; however, our findings are not generalisable to physical outpatients.

Conclusions

The PHQ-9's ≥7 cut-off point showed the highest simultaneous sensitivity and specificity when contrasted against a psychiatric diagnosis of depression. For a similar contrast against the gold standard, the other optimal cut-off points were: ≥7 for the PHQ-8, and ≥2 for the PHQ-2. Also, we present the sensitivity and specificity values of each cut-off point in GAD-7 and GAD-2. We confirmed the adequacy of a one-dimensional model for the PHQ-9, PHQ-8, and GAD-7, while all PHQ and GAD scales showed good reliability.

Tables, Figures, and Supplementary material

Table 1. Sociodemographic characteristics (n=1347).
Table 2. Raw scores and internal consistency (n=1347).

Supplementary material 1. Sex and age of the participants in our study and of the total number of people hospitalised in 2022.
Supplementary material 2. Script of R used in our study.
Supplementary material 3. Do-file of STATA used in our study.
Supplementary material 4. Flowchart.
Supplementary material 5. Sensitivity, specificity, +LR, −LR, PPV, NPV, and Youden's index for different PHQ cut-off points compared to the gold standard.
Supplementary material 6. Sensitivity, specificity, PPV, NPV, and Youden's index for different PHQ and GAD cut-off points compared to the gold standard.
Supplementary material 7. Sensitivity, specificity, +LR, −LR, PPV, NPV, and Youden's index for different GAD cut-off points compared to the gold standard.

Contributions

DVZ contributed to the conceptualizing the study, designing the methodology, developing the software tools, validating the results, conducting formal analyses, curating, and managing the data, and contributed to the initial drafting and visualization of the manuscript. JB contributed to the formal analysis, performed investigations, and aided in visualizing the findings. SOA participated in the investigation phase and contributed to the initial drafting of the manuscript. NMP engaged in formal analysis, conducted investigations, and contributed to the initial drafting of the manuscript. JCB contributed to the methodology, conducted investigations, provided critical input for the manuscript in the review and editing stages, and played a supervisory role. JHV contributed to the conceptualizing the study, designing the methodology, developing software tools, validating the results, conducting investigations, managing resources, curating data, project administration responsibilities, and participated in reviewing and editing the manuscript.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Competing interests

The authors declare that they have no conflicts of interest.

Patient consent for publication

Not applicable.

Ethics approval

The Hospital Nacional Guillermo Almenara Irigoyen's Institutional Review Board (Nota N°52 CIEI-OIyD-GRPA-Essalud-2023) approved the original study and the current analysis. Throughout the study, the researchers did not have access to any identifying information about the participants. In addition, participants gave informed consent. All participants were users of the hospital's Liaison Psychiatry Unit, so all received psychological or psychiatric care as needed.

Data availability statement

Data are available upon reasonable request.

ORCID iDs

David Villarreal-Zegarra https://orcid.org/0000-0002-2222-4764

Juan Barrera https://orcid.org/0000-0002-6641-6266
Sharlyn Otazú-Alfaro https://orcid.org/0000-0002-6462-8716
Nikol Mayo-Puchoc https://orcid.org/0000-0002-6182-7605
Juan Carlos Bazo-Alvarez https://orcid.org/0000-0002-6169-8049
Jeff Huarcaya-Victoria https://orcid.org/0000-0003-4525-9545

Reference

1. **Global Health Data Exchange (GHDx), Institute for Health Metrics and Evaluation (IHME)** [https://microdata.worldbank.org/index.php/catalog/ghdx/?page=1&ps=15&repo=ghdx]
2. GBD 2019 Mental Disorders Collaborators: **Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019**. *The lancet Psychiatry* 2022, **9**(2):137-150.
3. GBD 2019 Diseases and Injuries Collaborators: **Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019**. *Lancet (London, England)* 2020, **396**(10258):1204-1222.
4. **COVID-19 pandemic triggers 25% increase in prevalence of anxiety and depression worldwide** [https://www.who.int/news/item/02-03-2022-covid-19-pandemic-triggers-25-increase-in-prevalence-of-anxiety-and-depression-worldwide]
5. Villarreal-Zegarra D, Reátegui-Rivera CM, Otazú-Alfaro S, Yantas-Alcantara G, Soto-Becerra P, Melendez-Torres GJ: **Estimated impact of the COVID-19 pandemic on the prevalence and treatment of depressive symptoms in Peru: an interrupted time series analysis in 2014–2021**. *Social psychiatry and psychiatric epidemiology* 2023.
6. COVID-19 Mental Disorders Collaborators: **Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic**. *Lancet (London, England)* 2021, **398**(10312):1700-1712.
7. Villarreal-Zegarra D, Segovia-Bacilio P, Paredes-Angeles R, Vilela-Estrada AL, Cavero V, Diez-Canseco F: **Provision of community mental health care before and during the COVID-19 pandemic: A time series analysis in Peru**. *International Journal of Social Psychiatry* 2023.
8. Sagan A, McDaid D, Rajan S, Farrington J, McKee M: **European Observatory Policy Briefs**. In: *Screening: When is it appropriate and how can we get it right?* , edn. Copenhagen (Denmark): European Observatory on Health Systems and Policies © World Health Organization 2020 (acting as the host organization for, and secretariat of, the European Observatory on Health Systems and Policies). 2020.
9. Iragorri N, Spackman E: **Assessing the value of screening tools: reviewing the challenges and opportunities of cost-effectiveness analysis**. *Public health reviews* 2018, **39**:17.
10. Mulvaney-Day N, Marshall T, Downey Piscopo K, Korsen N, Lynch S, Karnell LH, Moran GE, Daniels AS, Ghose SS: **Screening for Behavioral Health Conditions in Primary Care Settings: A Systematic Review of the Literature**. *J Gen Intern Med* 2018, **33**(3):335-346.
11. Jiao B, Rosen Z, Bellanger M, Belkin G, Muennig P: **The cost-effectiveness of PHQ screening and collaborative care for depression in New York City**. *PloS one* 2017, **12**(8):e0184210.
12. Ramírez-Bontá F, Vásquez-Vílchez R, Cabrera-Alva M, Otazú-Alfaro S, Almeida-Huanca G, Ambrosio-Melgarejo J, Figueroa-Quiñones J, Romero-Cabrera A, Huaman-Santa Cruz A, Chávez-Hinostroza E *et al*: **Mental health data available in representative surveys conducted in Latin America and the Caribbean countries: A scoping review**. *In press* 2023.
13. Haberer JE, Trabin T, Klinkman M: **Furthering the reliable and valid measurement of mental health screening, diagnoses, treatment and outcomes through health information technology**. *Gen Hosp Psychiatry* 2013, **35**(4):349-353.

14. National Research Council; Division of Behavioral and Social Sciences and Education; Institute of Medicine; Board on Children Y, and Families; Committee on the Prevention of Mental Disorders and Substance Abuse Among Children, Youth and Young Adults: Research Advances and Promising Interventions; Mary Ellen O'Connell, Thomas Boat, and Kenneth E. Warner, Editors: **Preventing Mental, Emotional, and Behavioral Disorders Among Young People: Progress and Possibilities**. Washington: National Academies Press; 2009.
15. Spitzer RL, Kroenke K, Williams JB: **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* 1999, **282**(18):1737-1744.
16. Kroenke K, Strine TW, Spitzer RL, Williams JB, Berry JT, Mokdad AH: **The PHQ-8 as a measure of current depression in the general population**. *J Affect Disord* 2009, **114**(1-3):163-173.
17. Kroenke K, Spitzer RL, Williams JB: **The Patient Health Questionnaire-2: validity of a two-item depression screener**. *Med Care* 2003, **41**(11):1284-1292.
18. Plummer F, Manea L, Trepel D, McMillan D: **Screening for anxiety disorders with the GAD-7 and GAD-2: a systematic review and diagnostic metaanalysis**. *Gen Hosp Psychiatry* 2016, **39**:24-31.
19. 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**(6):e0156939.
20. Villarreal-Zegarra D, Copez-Lonzoy A, Bernabe-Ortiz A, Melendez-Torres GJ, Bazo-Alvarez JC: **Valid group comparisons can be made with the Patient Health Questionnaire (PHQ-9): A measurement invariance study across groups by demographic characteristics**. *PloS one* 2019, **14**(9):1-15.
21. Zhong QY, Gelaye B, Zaslavsky AM, Fann JR, Rondon MB, Sanchez SE, Williams MA: **Diagnostic Validity of the Generalized Anxiety Disorder - 7 (GAD-7) among Pregnant Women**. *PloS one* 2015, **10**(4):e0125096.
22. Stochl J, Fried EI, Fritz J, Croudace TJ, Russo DA, Knight C, Jones PB, Perez J: **On Dimensionality, Measurement Invariance, and Suitability of Sum Scores for the PHQ-9 and the GAD-7**. *Assessment* 2022, **29**(3):355-366.
23. Shevlin M, Butter S, McBride O, Murphy J, Gibson-Miller J, Hartman TK, Levita L, Mason L, Martinez AP, McKay R *et al*: **Measurement invariance of the Patient Health Questionnaire (PHQ-9) and Generalized Anxiety Disorder scale (GAD-7) across four European countries during the COVID-19 pandemic**. *BMC psychiatry* 2022, **22**(1):154.
24. Kroenke K, Wu J, Yu Z, Bair MJ, Kean J, Stump T, Monahan PO: **Patient Health Questionnaire Anxiety and Depression Scale: Initial Validation in Three Clinical Trials**. *Psychosom Med* 2016, **78**(6):716-727.
25. Urtasun M, Daray FM, Teti GL, Coppolillo F, Herlax G, Saba G, Rubinstein A, Araya R, Irazola V: **Validation and calibration of the patient health questionnaire (PHQ-9) in Argentina**. *BMC psychiatry* 2019, **19**(1):291.
26. García-Campayo J, Zamorano E, Ruiz MA, Pardo A, Pérez-Páramo M, López-Gómez V, Freire O, Rejas J: **Cultural adaptation into Spanish of the generalized anxiety disorder-7 (GAD-7) scale as a screening tool**. *Health and quality of life outcomes* 2010, **8**:8.
27. Sawaya H, Atoui M, Hamadeh A, Zeinoun P, Nahas Z: **Adaptation and initial validation of the Patient Health Questionnaire - 9 (PHQ-9) and the Generalized Anxiety Disorder - 7 Questionnaire (GAD-7) in an Arabic speaking Lebanese psychiatric outpatient sample**. *Psychiatry Res* 2016, **239**:245-252.
28. Lotrakul M, Sumrithe S, Saipanish R: **Reliability and validity of the Thai version of the PHQ-9**. *BMC psychiatry* 2008, **8**:46.
29. Woldetensay YK, Belachew T, Tesfaye M, Spielman K, Biesalski HK, Kantelhardt EJ, Scherbaum V: **Validation of the Patient Health Questionnaire (PHQ-9) as a screening**

tool for depression in pregnant women: Afaan Oromo version. *PloS one* 2018, **13**(2):e0191782.

30. Eack SM, Greeno CG, Lee BJ: **Limitations of the Patient Health Questionnaire in Identifying Anxiety and Depression: Many Cases Are Undetected**. *Research on social work practice* 2006, **16**(6):625-631.
31. Lambert SD, Clover K, Pallant JF, Britton B, King MT, Mitchell AJ, Carter G: **Making Sense of Variations in Prevalence Estimates of Depression in Cancer: A Co-Calibration of Commonly Used Depression Scales Using Rasch Analysis**. *Journal of the National Comprehensive Cancer Network : JNCCN* 2015, **13**(10):1203-1211.
32. Liu SI, Yeh ZT, Huang HC, Sun FJ, Tjung JJ, Hwang LC, Shih YH, Yeh AW: **Validation of Patient Health Questionnaire for depression screening among primary care patients in Taiwan**. *Comprehensive psychiatry* 2011, **52**(1):96-101.
33. Levis B, Benedetti A, Thombs BD: **Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression: individual participant data meta-analysis**. *BMJ : British Medical Journal* 2019, **365**:l1476.
34. Neupane D, Levis B, Bhandari PM, Thombs BD, Benedetti A: **Selective cutoff reporting in studies of the accuracy of the Patient Health Questionnaire-9 and Edinburgh Postnatal Depression Scale: Comparison of results based on published cutoffs versus all cutoffs using individual participant data meta-analysis**. *International journal of methods in psychiatric research* 2021, **30**(3):e1873.
35. Levis B, Sun Y, He C, Wu Y, Krishnan A, Bhandari PM, Neupane D, Imran M, Brehaut E, Negeri Z *et al*: **Accuracy of the PHQ-2 Alone and in Combination With the PHQ-9 for Screening to Detect Major Depression: Systematic Review and Meta-analysis**. *Jama* 2020, **323**(22):2290-2300.
36. Wu Y, Levis B, Riehm KE, Saadat N, Levis AW, Azar M, Rice DB, Boruff J, Cuijpers P, Gilbody S *et al*: **Equivalency of the diagnostic accuracy of the PHQ-8 and PHQ-9: a systematic review and individual participant data meta-analysis**. *Psychological medicine* 2020, **50**(8):1368-1380.
37. 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**(4):345-359.
38. Manea L, Boehnke JR, Gilbody S, Moriarty AS, McMillan D: **Are there researcher allegiance effects in diagnostic validation studies of the PHQ-9? A systematic review and meta-analysis**. *BMJ Open* 2017, **7**(9):e015247.
39. 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**. *BJPsych open* 2016, **2**(2):127-138.
40. Mughal AY, Devadas J, Ardman E, Levis B, Go VF, Gaynes BN: **A systematic review of validated screening tools for anxiety disorders and PTSD in low to middle income countries**. *BMC psychiatry* 2020, **20**(1):338.
41. Cohen JF, Korevaar DA, Altman DG, Bruns DE, Gatsonis CA, Hooft L, Irwig L, Levine D, Reitsma JB, de Vet HC *et al*: **STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration**. *BMJ Open* 2016, **6**(11):e012799.
42. Villarreal-Zegarra D, Cabrera-Alva M, Carrillo-Larco RM, Bernabe-Ortiz A: **Trends in the prevalence and treatment of depressive symptoms in Peru: a population-based study**. *BMJ Open* 2020, **10**(7):e036777.
43. Hernández-Vásquez A, Vargas-Fernández R, Bendezu-Quispe G, Grendas LN: **Depression in the Peruvian population and its associated factors: analysis of a national health survey**. *J Affect Disord* 2020, **273**:291-297.
44. Steel Z, Marnane C, Iranpour C, Chey T, Jackson JW, Patel V, Silove D: **The global prevalence of common mental disorders: a systematic review and meta-analysis 1980-2013**. *Int J Epidemiol* 2014, **43**(2):476-493.

45. Buderer NM: **Statistical methodology: I. Incorporating the prevalence of disease into the sample size calculation for sensitivity and specificity**. *Academic emergency medicine : official journal of the Society for Academic Emergency Medicine* 1996, **3**(9):895-900.
46. Huarcaya-Victoria J, Segura V, Cárdenas D, Sardón K, Caqui M, Podestà Á: **Analysis of the care provided over a six-month period by the liaison psychiatry unit at a general hospital in Lima, Peru**. *Rev Colomb Psiquiatr (Engl Ed)* 2022, **51**(2):105-112.
47. 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**. *General hospital psychiatry* 2015, **37**(1):67-75.
48. Zuithoff NP, Vergouwe Y, King M, Nazareth I, van Wezep MJ, Moons KG, Geerlings MI: **The Patient Health Questionnaire-9 for detection of major depressive disorder in primary care: consequences of current thresholds in a crosssectional study**. *BMC Fam Pract* 2010, **11**:98.
49. Shin C, Lee SH, Han KM, Yoon HK, Han C: **Comparison of the Usefulness of the PHQ-8 and PHQ-9 for Screening for Major Depressive Disorder: Analysis of Psychiatric Outpatient Data**. *Psychiatry investigation* 2019, **16**(4):300-305.
50. Caycho-Rodríguez T, Barboza-Palomino M, Ventura-León J, Carbajal-León C, Noé-Grijalva M, Gallegos M, Reyes-Bossio M, Vivanco-Vidal A: **[Spanish translation and validation of a brief measure of anxiety by the COVID-19 in students of health sciences]**. *Ansiedad y Estrés* 2020, **26**(2):174-180.
51. Spitzer RL, Kroenke K, Williams JW, Löwe B: **A brief measure for assessing generalized anxiety disorder: The GAD-7**. *Archives of Internal Medicine* 2006, **166**(10):1092-1097.
52. Franco-Jimenez RA, Nuñez-Magallanes A: **[Psychometric Properties of the GAD-7, GAD-2, and GAD-Mini in Peruvian College Students]**. *Propósitos y Representaciones* 2022, **10**(1):e1437.
53. Kroenke K, Spitzer RL, Williams JB, Monahan PO, Löwe B: **Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection**. *Annals of internal medicine* 2007, **146**(5):317-325.
54. Trevethan R: **Sensitivity, Specificity, and Predictive Values: Foundations, Pliabilities, and Pitfalls in Research and Practice**. *Frontiers in public health* 2017, **5**:307.
55. Ranganathan P, Aggarwal R: **Understanding the properties of diagnostic tests - Part 2: Likelihood ratios**. *Perspectives in clinical research* 2018, **9**(2):99-102.
56. Suh Y: **The Performance of Maximum Likelihood and Weighted Least Square Mean and Variance Adjusted Estimators in Testing Differential Item Functioning With Nonnormal Trait Distributions**. *Structural Equation Modeling: A Multidisciplinary Journal* 2015, **22**(4):568-580.
57. Holgado–Tello FP, Chacón–Moscoso S, Barbero–García I, Vila–Abad E: **Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables**. *Qual Quant* 2010, **44**(1):153-166.
58. Hair JF, Anderson RE, Tatham RL, Black WC: **Análisis multivariante**, vol. 491: Prentice Hall Madrid; 1999.
59. McDonald RP: **Test theory: A unified treatment**. New York: Taylor & Francis Group; 1999.
60. Costantini L, Pasquarella C, Odone A, Colucci ME, Costanza A, Serafini G, Aguglia A, Belvederi Murri M, Brakoulias V, Amore M *et al*: **Screening for depression in primary care with Patient Health Questionnaire-9 (PHQ-9): A systematic review**. *J Affect Disord* 2021, **279**:473-483.
61. Hartung TJ, Friedrich M, Johansen C, Wittchen HU, Faller H, Koch U, Brähler E, Härter M, Keller M, Schulz H *et al*: **The Hospital Anxiety and Depression Scale (HADS) and the 9-item Patient Health Questionnaire (PHQ-9) as screening instruments for depression in patients with cancer**. *Cancer* 2017, **123**(21):4236-4243.

62. Sun Y, Kong Z, Song Y, Liu J, Wang X: **The validity and reliability of the PHQ-9 on screening of depression in neurology: a cross sectional study**. *BMC psychiatry* 2022, **22**(1):98.
63. Gholizadeh L, Shahmansouri N, Heydari M, Davidson PM: **Assessment and detection of depression in patients with coronary artery disease: validation of the Persian version of the PHQ-9**. *Contemporary nurse* 2019, **55**(2-3):185-194.
64. Husain N, Waheed W, Tomenson B, Creed F: **The validation of personal health questionnaire amongst people of Pakistani family origin living in the United Kingdom**. *J Affect Disord* 2007, **97**(1-3):261-264.
65. Ganguly S, Samanta M, Roy P, Chatterjee S, Kaplan DW, Basu B: **Patient health questionnaire-9 as an effective tool for screening of depression among Indian adolescents**. *The Journal of adolescent health : official publication of the Society for Adolescent Medicine* 2013, **52**(5):546-551.
66. Hanlon C, Medhin G, Selamu M, Breuer E, Worku B, Hailemariam M, Lund C, Prince M, Fekadu A: **Validity of brief screening questionnaires to detect depression in primary care in Ethiopia**. *J Affect Disord* 2015, **186**:32-39.
67. Harry ML, Coley RY, Waring SC, Simon GE: **Evaluating the Cross-Cultural Measurement Invariance of the PHQ-9 between American Indian/Alaska Native Adults and Diverse Racial and Ethnic Groups**. *Journal of affective disorders reports* 2021, **4**.
68. Razykov I, Ziegelstein RC, Whooley MA, Thombs BD: **The PHQ-9 versus the PHQ-8--is item 9 useful for assessing suicide risk in coronary artery disease patients? Data from the Heart and Soul Study**. *J Psychosom Res* 2012, **73**(3):163-168.
69. Konkan R, Senormancı O, Guclu O, Aydin E, Sungur M: **Validity and reliability study for the Turkish adaptation of the Generalized Anxiety Disorder-7 (GAD-7) scale**. *Archives of Neuropsychiatry* 2013, **50**:53-58.
70. Gong Y, Zhou H, Zhang Y, Zhu X, Wang X, Shen B, Xian J, Ding Y: **Validation of the 7-item Generalized Anxiety Disorder scale (GAD-7) as a screening tool for anxiety among pregnant Chinese women**. *J Affect Disord* 2021, **282**:98-103.
71. Snijkers JTW, van den Oever W, Weerts Z, Vork L, Mujagic Z, Leue C, Hesselink MAM, Kruimel JW, Muris JWM, Bogie RMM *et al*: **Examining the optimal cutoff values of HADS, PHQ-9 and GAD-7 as screening instruments for depression and anxiety in irritable bowel syndrome**. *Neurogastroenterology and motility : the official journal of the European Gastrointestinal Motility Society* 2021, **33**(12):e14161.
72. Manea L, Gilbody S, Hewitt C, North A, Plummer F, Richardson R, Thombs BD, Williams B, McMillan D: **Identifying depression with the PHQ-2: A diagnostic meta-analysis**. *J Affect Disord* 2016, **203**:382-395.
73. Luo Z, Li Y, Hou Y, Zhang H, Liu X, Qian X, Jiang J, Wang Y, Liu X, Dong X *et al*: **Adaptation of the two-item generalized anxiety disorder scale (GAD-2) to Chinese rural population: A validation study and meta-analysis**. *Gen Hosp Psychiatry* 2019, **60**:50-56.
74. Giuliani M, Gorini A, Barbieri S, Veglia F, Tremoli E: **Examination of the best cut-off points of PHQ-2 and GAD-2 for detecting depression and anxiety in Italian cardiovascular inpatients**. *Psychology & health* 2021, **36**(9):1088-1101.
75. Bentley KH, Sakurai H, Lowman KL, Rines-Toth L, McKowen J, Pedrelli P, Evins AE, Yule AM: **Validation of brief screening measures for depression and anxiety in young people with substance use disorders**. *J Affect Disord* 2021, **282**:1021-1029.
76. Lamela D, Soreira C, Matos P, Morais A: **Systematic review of the factor structure and measurement invariance of the patient health questionnaire-9 (PHQ-9) and validation of the Portuguese version in community settings**. *J Affect Disord* 2020, **276**:220-233.
77. Schantz K, Reighard C, Aikens JE, Aruquipa A, Pinto B, Valverde H, Piette JD: **Screening for depression in Andean Latin America: Factor structure and reliability of the CES-D**

short form and the PHQ-8 among Bolivian public hospital patients. *International journal of psychiatry in medicine* 2017, **52**(4-6):315-327.

78. Huarcaya-Victoria J, De-Lama-Morán R, Quiros M, Bazán J, López K, Lora D: **Propiedades psicométricas del Patient Health Questionnaire (PHQ-9) en estudiantes de medicina en Lima, Perú**. *Revista de Neuro-Psiquiatria* 2020, **83**(2):72-78.
79. Inagaki M, Ohtsuki T, Yonemoto N, Kawashima Y, Saitoh A, Oikawa Y, Kurosawa M, Muramatsu K, Furukawa TA, Yamada M: **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**(6):592-597.
80. Le Hoang Ngoc T, Le MT, Nguyen HT, Vo HV, Le NQ, Tang LNP, Tran TT, Le TV: **Patient Health Questionnaire (PHQ-9): A depression screening tool for people with epilepsy in Vietnam**. *Epilepsy & behavior : E&B* 2021, **125**:108446.
81. Bhandari PM, Levis B, Neupane D, Patten SB, Shrier I, Thombs BD, Benedetti A: **Data-driven methods distort optimal cutoffs and accuracy estimates of depression screening tools: a simulation study using individual participant data**. *J Clin Epidemiol* 2021, **137**:137-147.
82. Beatrice M-F, Carla M-C, Matilde L-M, Víctor G-M, Amelia M-M, Jessica Hanae Z-T, Sergio G-L, Raisa NM-R, Wendy N-G, Fabian F-S *et al*: **Clinical practice guideline for the screening and management of the mild depressive episode at the first level of care for the Peruvian Social Security (EsSalud)**. *ACTA MEDICA PERUANA* 2020, **37**(4).
83. EsSalud: **Análisis ejecutivo nacional de a nivel las prestaciones de salud 2016**. In. Lima: Gerencia central de planeamiento y presupuesto, EsSalud; 2017.

Table 1. Sociodemographic characteristics (n=1347).

	n	%
Sex		
Men	547	40.6%
Women	800	59.4%
Age (categories)		
18-29	107	7.9%
30-39	164	12.2%
40-49	214	15.9%
50-59	284	21.1%
60-69	294	21.8%
70-79	203	15.1%
80 to more	81	6.0%
Civil status		
Single	329	24.4%
Married or Cohabitant	768	57.0%
Separated	133	9.9%
Widowed	117	8.7%
Education level		
None	13	1.0%
Elementary school	135	10.0%
High school	478	35.5%
Technical	246	18.2%
University	475	35.3%
Currently works		
No	330	24.5%
Yes	778	57.8%
Retired	239	17.7%
Living alone		
Yes	99	7.3%
No	1248	92.7%
History of psychiatric diagnosis		
Yes	388	28.8%
No	959	71.2%
Diagnosis of depression		
No	1013	75.2%
Yes	334	24.8%
Diagnosis of anxiety		
No	1319	97.9%
Yes	28	2.1%
Physical illnesses		
A00-B99 Certain infectious and parasitic diseases	109	8.1%
C00-D48 Neoplasms, and diseases of the blood and haematopoietic organs and other disorders affecting the mechanism of immunity	348	25.8%
E00-E90 Endocrine, nutritional and metabolic diseases	130	9.7%
G00-G99 Diseases of the nervous system	96	7.1%
H00-H59 Diseases of the eye and adnexa	17	1.3%
H60-H95 Diseases of the ear and mastoid process	17	1.3%
I00-I99 Diseases of the circulatory system	111	8.2%
J00-J99 Diseases of the respiratory system	107	7.9%
K00-K93 Diseases of the gastro-intestinal tract	106	7.9%
L00-L99 Diseases of the skin and subcutaneous tissues	74	5.5%
M00-M99 Diseases of the musculo-skeletal system and connective tissue	97	7.2%
N00-N99 Diseases of the genito-urinary system	97	7.2%
O00-O99 Pregnancy, childbirth and puerperium	10	0.7%
P00-P96 Certain conditions originating in the perinatal period	0	0.0%
Q00-Q99 Congenital malformations, deformities and chromosome anomalies	6	0.4%
R00-R99 Symptoms, signs and abnormal clinical and laboratory	46	3.4%

findings, not elsewhere classified		
S00-T98 Trauma, poisoning and certain other consequences of external cause	50	3.7%
V01-Y98 External causes of morbidity and mortality	4	0.3%
Z00-Z99 Factors influencing health status and contact with health care services	90	6.7%
U00-U99 Codes for special situations	28	2.1%

Note: n=number. %=Percentage.

Table 2. Raw scores and internal consistency (n=1347).

	M	SD	Min	Max	α	ω
(1) PHQ-9 score	6.4	5.0	0	27	0.89	0.86
(2) PHQ-8 score	6.1	4.7	0	24	0.88	0.85
(3) PHQ-2 score	1.9	1.6	0	6	0.83	0.80
(4) GAD-7 score	5.1	3.9	0	21	0.85	0.81
(5) GAD-2 score	1.7	1.4	0	6	0.74	0.70

Note: α = Classical alpha. ω = Mcdonald's omega.

Supplementary material 1. Sex and age of the participants in our study and of the total number of people hospitalised in 2022.

	Our study		Total inpatients in 2022		
	n	%			p
Sex					
Men	547	40.60%	9677	43.2%	0.971
Women	800	59.40%	12732	56.8%	
Age (categories)					
18-29	107	7.90%	1962	8.8%	1.000
30-39	164	12.20%	3486	15.6%	
40-49	214	15.90%	3145	14.0%	
50-59	284	21.10%	3497	15.6%	
60-69	294	21.80%	4276	19.1%	
70-79	203	15.10%	3806	17.0%	
80 to more	81	6.00%	2227	9.9%	

Note: p = chi-square test.

Supplementary material 2. Script of R used in our study.

```
library(lavaan)
library(semPlot)
library(semTools)
library(psych)
library(haven)
Database <- read_dta("E:/Database_v1.dta")

#PHQ-9
model.PHQ9 <- "
F1 =~ PHQ_1 + PHQ_2 + PHQ_3 + PHQ_4 + PHQ_5 + PHQ_6 + PHQ_7 + PHQ_8 + PHQ_9"

fit.model.PHQ9 <- cfa(model.PHQ9, data=Database,estimator="WLSMV", missing = "listwise",
ordered = c("PHQ_1", "PHQ_2", "PHQ_3", "PHQ_4", "PHQ_5", "PHQ_6", "PHQ_7", "PHQ_8",
"PHQ_9"))

summary(fit.model.PHQ9, fit.measures=TRUE, standardized=TRUE)

reliability(fit.model.PHQ9)

#PHQ-8
model.PHQ8 <- "
F1 =~ PHQ_1 + PHQ_2 + PHQ_3 + PHQ_4 + PHQ_5 + PHQ_6 + PHQ_7 + PHQ_8"

fit.model.PHQ8 <- cfa(model.PHQ8, data=Database, estimator="WLSMV", missing =
"listwise", ordered = c("PHQ_1", "PHQ_2", "PHQ_3", "PHQ_4", "PHQ_5", "PHQ_6", "PHQ_7",
"PHQ_8", "PHQ_9"))

summary(fit.model.PHQ8, fit.measures=TRUE, standardized=TRUE)

reliability(fit.model.PHQ8)

#PHQ-2
model.PHQ2 <- "
F1 =~ PHQ_1 + PHQ_2"

fit.model.PHQ2 <- cfa(model.PHQ2, data=Database, estimator="WLSMV", missing =
"listwise", ordered = c("PHQ_1", "PHQ_2", "PHQ_3", "PHQ_4", "PHQ_5", "PHQ_6", "PHQ_7",
"PHQ_8", "PHQ_9"))

summary(fit.model.PHQ2, fit.measures=TRUE, standardized=TRUE)

reliability(fit.model.PHQ2)

#GAD-7
model.GAD7 <- "
F1 =~ GAD1 + GAD2 + GAD3 + GAD4 + GAD5 + GAD6 + GAD7"

fit.model.GAD7 <- cfa(model.GAD7, data=Database, estimator="WLSMV", missing =
"listwise", ordered = c("GAD1", "GAD2", "GAD3", "GAD4", "GAD5", "GAD6", "GAD7"))

summary(fit.model.GAD7, fit.measures=TRUE, standardized=TRUE)

reliability(fit.model.GAD7)

#GAD-2
model.GAD2 <- "
F1 =~ GAD1 + GAD2"

fit.model.GAD2 <- cfa(model.GAD2, data=Database, estimator="WLSMV", missing =
"listwise", ordered = c("GAD1", "GAD2", "GAD3", "GAD4", "GAD5", "GAD6", "GAD7"))

summary(fit.model.GAD2, fit.measures=TRUE, standardized=TRUE)

reliability(fit.model.GAD2)
```

Supplementary material 3. Do-file of STATA used in our study.

```
use "E:\Database_v1.dta", clear

*Table 1 - Socio-demographic analysis

global catvars_table1 sex agecat civilstatus educationcat work Livingalone
Historypsychiatricdx depression anxiety

tabout $catvars_table1 depression using Table1.xlsx, ///
  replace c(freq col) clab(No. %) f(0c 1p) style(xlsx) font(bold) ///
  ptotal(none) stats(chi2) stpos(col) ppos(only) plab(P value) ///
  title(Table 1. Sociodemographic characteristics) ///
  fn(Note: n=number, %=Percentage.) twidth(14) sheet(Table1)

* Table 2. Sensitivity, specificity, +LR, -LR, PPV, NPV, and Youden's index for different
PHQ and GAD cut-off points compared to the gold standard.
tab cutoffPHQ9_7
diagt depression cutoffPHQ9_7
tab cutoffPHQ9_10
tab PHQ9algorithm
diagt depression PHQ9algorithm
tab PHQ9ajus_algorithm
diagt depression PHQ9ajus_algorithm
diagt depression cutoffPHQ9_10
tab cutoffPHQ8_7
diagt depression cutoffPHQ8_7
tab cutoffPHQ2_2
diagt depression cutoffPHQ2_2
tab cutoffGAD7_8
diagt anxiety cutoffGAD7_8
tab cutoffGAD7_10
diagt anxiety cutoffGAD7_10
tab cutoffGAD2_2
diagt anxiety cutoffGAD2_2



* Supplementary material 3. Sensitivity, specificity, +LR, -LR, PPV, NPV, and Youden's
index for different PHQ cut-off points compared to the gold standard.
* PHQ-9 - Sensitivity, specificity, +LR, -LR, PPV, NPV, and Youden's index

tab cutoffPHQ9_1
diagt depression cutoffPHQ9_1
tab cutoffPHQ9_2
diagt depression cutoffPHQ9_2
tab cutoffPHQ9_3
diagt depression cutoffPHQ9_3
tab cutoffPHQ9_4
diagt depression cutoffPHQ9_4
tab cutoffPHQ9_5
diagt depression cutoffPHQ9_5
tab cutoffPHQ9_6
diagt depression cutoffPHQ9_6
tab cutoffPHQ9_7
diagt depression cutoffPHQ9_7
tab cutoffPHQ9_8
diagt depression cutoffPHQ9_8
tab cutoffPHQ9_9
diagt depression cutoffPHQ9_9
tab cutoffPHQ9_10
diagt depression cutoffPHQ9_10
tab cutoffPHQ9_11
diagt depression cutoffPHQ9_11
tab cutoffPHQ9_12
diagt depression cutoffPHQ9_12
```

```
tab cutoffPHQ9_13
diagt depression cutoffPHQ9_13
tab cutoffPHQ9_14
diagt depression cutoffPHQ9_14
tab cutoffPHQ9_15
diagt depression cutoffPHQ9_15
tab cutoffPHQ9_16
diagt depression cutoffPHQ9_16
tab cutoffPHQ9_17
diagt depression cutoffPHQ9_17
tab cutoffPHQ9_18
diagt depression cutoffPHQ9_18
tab cutoffPHQ9_19
diagt depression cutoffPHQ9_19
tab cutoffPHQ9_20
diagt depression cutoffPHQ9_20
tab cutoffPHQ9_21
diagt depression cutoffPHQ9_21
tab cutoffPHQ9_22
diagt depression cutoffPHQ9_22
tab cutoffPHQ9_23
diagt depression cutoffPHQ9_23
tab cutoffPHQ9_24
diagt depression cutoffPHQ9_24
tab cutoffPHQ9_25
diagt depression cutoffPHQ9_25
tab cutoffPHQ9_26
diagt depression cutoffPHQ9_26
tab cutoffPHQ9_27
diagt depression cutoffPHQ9_27
```

***PHQ-8 - Sensitivity, specificity, +LR, −LR, PPV, NPV, and Youden's index**

```
tab cutoffPHQ8_1
diagt depression cutoffPHQ8_1
tab cutoffPHQ8_2
diagt depression cutoffPHQ8_2
tab cutoffPHQ8_3
diagt depression cutoffPHQ8_3
tab cutoffPHQ8_4
diagt depression cutoffPHQ8_4
tab cutoffPHQ8_5
diagt depression cutoffPHQ8_5
tab cutoffPHQ8_6
diagt depression cutoffPHQ8_6
tab cutoffPHQ8_7
diagt depression cutoffPHQ8_7
tab cutoffPHQ8_8
diagt depression cutoffPHQ8_8
tab cutoffPHQ8_9
diagt depression cutoffPHQ8_9
tab cutoffPHQ8_10
diagt depression cutoffPHQ8_10
tab cutoffPHQ8_11
diagt depression cutoffPHQ8_11
tab cutoffPHQ8_12
diagt depression cutoffPHQ8_12
tab cutoffPHQ8_13
diagt depression cutoffPHQ8_13
tab cutoffPHQ8_14
diagt depression cutoffPHQ8_14
tab cutoffPHQ8_15
diagt depression cutoffPHQ8_15
tab cutoffPHQ8_16
diagt depression cutoffPHQ8_16
tab cutoffPHQ8_17
diagt depression cutoffPHQ8_17
```

```
tab cutoffPHQ8_18
diagt depression cutoffPHQ8_18
tab cutoffPHQ8_19
diagt depression cutoffPHQ8_19
tab cutoffPHQ8_20
diagt depression cutoffPHQ8_20
tab cutoffPHQ8_21
diagt depression cutoffPHQ8_21
tab cutoffPHQ8_22
diagt depression cutoffPHQ8_22
tab cutoffPHQ8_23
diagt depression cutoffPHQ8_23
tab cutoffPHQ8_24
diagt depression cutoffPHQ8_24
```

*** PHQ-9 algorithm - Sensitivity, specificity, +LR, −LR, PPV, NPV, and Youden's index**

```
tab PHQ9algorithm
diagt depression PHQ9algorithm
```

*** PHQ-9 adjusted algorithm - Sensitivity, specificity, +LR, −LR, PPV, NPV, and Youden's index**

```
tab PHQ9ajus_algorithm
diagt depression PHQ9ajus_algorithm
```

*** PHQ-2 - Sensitivity, specificity, +LR, −LR, PPV, NPV, and Youden's index**

```
tab cutoffPHQ2_1
diagt depression cutoffPHQ2_1
tab cutoffPHQ2_2
diagt depression cutoffPHQ2_2
tab cutoffPHQ2_3
diagt depression cutoffPHQ2_3
tab cutoffPHQ2_4
diagt depression cutoffPHQ2_4
tab cutoffPHQ2_5
diagt depression cutoffPHQ2_5
tab cutoffPHQ2_6
diagt depression cutoffPHQ2_6
```

***Supplementary material 4. Sensitivity, specificity, +LR, −LR, PPV, NPV, and Youden's index for different GAD cut-off points compared to the gold standard.**
***GAD-7 - Sensitivity, specificity, +LR, −LR, PPV, NPV, and Youden's index**

```
tab cutoffGAD7_1
diagt anxiety cutoffGAD7_1
tab cutoffGAD7_2
diagt anxiety cutoffGAD7_2
tab cutoffGAD7_3
diagt anxiety cutoffGAD7_3
tab cutoffGAD7_4
diagt anxiety cutoffGAD7_4
tab cutoffGAD7_5
diagt anxiety cutoffGAD7_5
tab cutoffGAD7_6
diagt anxiety cutoffGAD7_6
tab cutoffGAD7_7
diagt anxiety cutoffGAD7_7
tab cutoffGAD7_8
diagt anxiety cutoffGAD7_8
tab cutoffGAD7_9
diagt anxiety cutoffGAD7_9
tab cutoffGAD7_10
diagt anxiety cutoffGAD7_10
tab cutoffGAD7_11
diagt anxiety cutoffGAD7_11
```

```
tab cutoffGAD7_12
diagt anxiety cutoffGAD7_12
tab cutoffGAD7_13
diagt anxiety cutoffGAD7_13
tab cutoffGAD7_14
diagt anxiety cutoffGAD7_14
tab cutoffGAD7_15
diagt anxiety cutoffGAD7_15
tab cutoffGAD7_16
diagt anxiety cutoffGAD7_16
tab cutoffGAD7_17
diagt anxiety cutoffGAD7_17
tab cutoffGAD7_18
diagt anxiety cutoffGAD7_18
tab cutoffGAD7_19
diagt anxiety cutoffGAD7_19
tab cutoffGAD7_20
diagt anxiety cutoffGAD7_20
tab cutoffGAD7_21
diagt anxiety cutoffGAD7_21
```

*** GAD-2 - Sensitivity, specificity, +LR, -LR, PPV, NPV, and Youden's index**

```
tab cutoffGAD2_0
diagt anxiety cutoffGAD2_0
tab cutoffGAD2_1
diagt anxiety cutoffGAD2_1
tab cutoffGAD2_2
diagt anxiety cutoffGAD2_2
tab cutoffGAD2_3
diagt anxiety cutoffGAD2_3
tab cutoffGAD2_4
diagt anxiety cutoffGAD2_4
tab cutoffGAD2_5
diagt anxiety cutoffGAD2_5
tab cutoffGAD2_6
diagt anxiety cutoffGAD2_6
```

***Figures 2 - ROC curve**

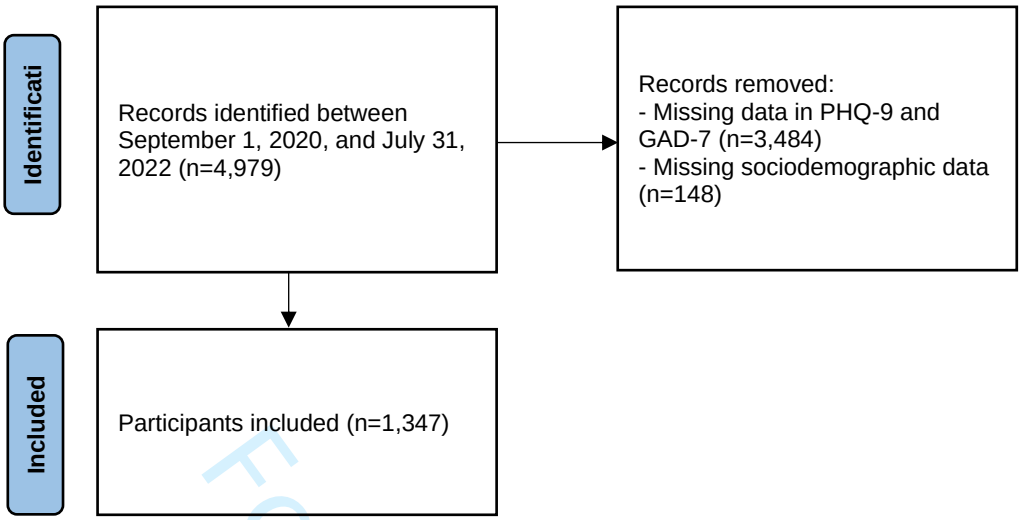
```
roccomp depression PHQ9TOTAL PHQ8TOTAL PHQ9algorithm PHQ9ajus_algorithm PHQ2TOTAL, graph
summary plot1opts(mcolor(red) lcolor(red)) plot2opts(mcolor(blue) lcolor(blue))
plot3opts(mcolor(orange) lcolor(orange)) plot4opts(mcolor(green) lcolor(green))
plot5opts(mcolor(purble) lcolor(purple)) legend(order(1 "PHQ-9 (AUC=0.805 [0.779-
0.831])" 2 "PHQ-8 (AUC=0.802 [0.776-0.828])" 3 "PHQ-9 algorithm (AUC=0.641 [0.614-
0.667])" 4 "PHQ-9 adjusted algorithm (AUC=0.723 [0.696-0.749])" 5 "PHQ-2 (AUC=0.771
[0.743-0.799])" 6 "Reference") size(2.5) position(7) cols(2) rows(3))

graph export "E:\Figure2.tif", as(tif) replace
```

***Figures 3 - ROC curve**

```
roccomp anxiety GAD7TOTAL GAD2TOTAL, graph summary plot1opts(mcolor(red) lcolor(red))
plot2opts(mcolor(blue) lcolor(blue)) legend(order(1 "GAD-7 (AUC=0.718 [0.622-0.814])" 2
"GAD-2 (AUC=0.685 [0.587-0.783])" 3 "Reference") size(2.5) position(7) cols(2) rows(3))

graph export "E:\Figure3.tif", as(tif) replace
```



Supplementary material 4. Flowchart.

Supplementary material 5. Sensitivity, specificity, +LR, −LR, PPV, NPV, and Youden's index for different PHQ cut-off points compared to the gold standard.

Method	Cut-off Points	n (%)	Sensitivity (95%CI)	Specificity (95%CI)	+LR (95%CI)	-LR (95%CI)	PPV (95%CI)	NPV (95%CI)	Youden Index
PHQ-9	≥1	1246 (92.5)	99.7 (98.3 - 100.0)	9.9 (8.1 - 11.9)	1.11 (1.08 - 1.13)	0.03 (0.00 - 0.22)	26.7 (24.3 - 29.3)	99.0 (94.6 - 100.0)	9.6
	≥2	1145 (85.0)	98.2 (96.1 - 99.3)	19.3 (17.0 - 21.9)	1.22 (1.18 - 1.26)	0.93 (0.04 - 0.21)	28.6 (26.0 - 31.4)	97.0 (93.6 - 98.9)	17.5
	≥3	1004 (74.5)	95.5 (92.7 - 97.5)	32.4 (29.5 - 35.4)	1.41 (1.35 - 1.48)	0.14 (0.84 - 0.23)	31.8 (28.9 - 34.8)	95.6 (92.9 - 97.5)	27.9
	≥4	850 (63.1)	91.3 (87.8 - 94.1)	46.2 (43.1 - 49.3)	1.70 (1.59 - 1.81)	0.19 (0.13 - 0.27)	35.9 (32.7 - 39.2)	94.2 (91.7 - 96.1)	37.5
	≥5	720 (53.5)	87.4 (83.4 - 90.8)	57.7 (54.6 - 60.8)	2.07 (1.90 - 2.25)	0.22 (0.16 - 0.29)	40.6 (36.9 - 44.2)	93.3 (91.1 - 95.1)	45.1
	≥6	627 (46.6)	80.5 (75.9 - 84.6)	64.7 (61.6 - 67.6)	2.28 (2.07 - 2.52)	0.30 (0.24 - 0.38)	42.9 (39.0 - 46.9)	91.0 (88.6 - 93.0)	45.2
	≥7	**537 (39.9)**	**76.0 (71.1 - 80.5)**	**72.1 (69.2 - 74.8)**	**2.72 (2.42 - 3.06)**	**0.33 (0.27 - 0.40)**	**47.3 (43.0 - 51.6)**	**90.1 (87.9 - 92.1)**	**48.1**
	≥8	454 (33.7)	68.9 (63.6 - 73.8)	77.9 (75.2 - 80.4)	3.11 (2.72 - 3.57)	0.40 (0.34 - 0.47)	50.7 (46.0 - 55.4)	88.4 (86.1 - 90.4)	46.8
	≥9	371 (27.5)	60.2 (54.7 - 65.5)	83.2 (80.8 - 85.5)	3.59 (3.05 - 4.22)	0.48 (0.42 - 0.55)	54.2 (49.0 - 59.3)	86.4 (84.1 - 88.5)	43.4
	≥10	309 (22.9)	54.2 (48.7 - 59.6)	87.4 (85.2 - 89.3)	4.29 (3.55 - 5.18)	0.52 (0.47 - 0.59)	58.6 (53.9 - 63.1)	85.3 (83.7 - 86.7)	41.6
	≥11	252 (18.7)	44.9 (39.5 - 50.4)	89.9 (87.9 - 91.7)	4.46 (3.58 - 5.55)	0.61 (0.56 - 0.68)	59.5 (53.2 - 65.6)	83.2 (80.8 - 85.4)	34.8
	≥12	212 (15.7)	38.9 (33.7 - 44.4)	91.9 (90.1 - 93.5)	4.81 (3.76 - 6.16)	0.67 (0.61 - 0.73)	61.3 (54.4 - 67.9)	82.0 (79.7 - 84.2)	30.8
	≥13	161 (12.0)	31.4 (26.5 - 36.7)	94.5 (92.9 - 95.8)	5.69 (4.21 - 7.67)	0.73 (0.67 - 0.78)	65.2 (57.3 - 72.5)	80.7 (78.3 - 82.9)	25.9
	≥14	138 (10.2)	26.3 (21.7 - 31.4)	95.1 (93.5 - 96.3)	5.34 (3.86 - 7.38)	0.78 (0.73 - 0.83)	63.8 (55.2 - 71.8)	79.7 (77.3 - 81.9)	21.4
	≥15	102 (7.6)	19.8 (15.6 - 24.4)	96.4 (95.1 - 97.5)	5.56 (3.78 - 8.19)	0.83 (0.79 - 0.88)	64.7 (54.6 - 73.9)	78.5 (76.1 - 80.7)	16.2
	≥16	79 (5.9)	15.3 (11.6 - 19.6)	97.2 (96.0 - 98.2)	5.52 (3.54 - 8.61)	0.87 (0.83 - 0.91)	64.6 (53.0 - 75.0)	77.7 (75.3 - 79.9)	12.5
	≥17	57 (4.2)	10.5 (7.4 - 14.3)	97.8 (96.7 - 98.6)	4.83 (2.87 - 8.11)	0.92 (0.88 - 0.95)	61.4 (47.6 - 74.0)	76.8 (74.4 - 79.1)	8.3
	≥18	48 (3.6)	8.4 (5.6 - 11.9)	98.0 (97.0 - 98.8)	4.25 (2.42 - 7.44)	0.94 (0.90 - 0.97)	58.3 (43.2 - 72.4)	76.4 (74.0 - 78.7)	6.4
	≥19	40 (3.0)	6.9 (4.4 - 10.2)	98.3 (97.3 - 99.0)	4.10 (2.22 - 7.59)	0.95 (0.92 - 0.98)	57.5 (40.9 - 73.0)	76.2 (73.8 - 78.5)	5.2
	≥20	35 (2.4)	4.5 (2.5 - 7.3)	98.3 (97.3 - 99.0)	2.68 (1.35 - 5.30)	0.97 (0.95 - 1.00)	46.9 (29.1 - 65.3)	75.7 (73.3 - 78.0)	2.8
	≥21	26 (1.9)	3.9 (2.1 - 6.6)	98.7 (97.8 - 99.3)	3.03 (1.42 - 6.48)	0.97 (0.95 - 1.00)	50.0 (29.9 - 70.1)	75.7 (73.3 - 78.0)	2.6
	≥22	20 (1.5)	2.7 (1.2 - 5.1)	98.9 (98.1 - 99.5)	2.48 (1.04 - 5.94)	0.98 (0.97 - 1.00)	45.0 (23.1 - 68.5)	75.5 (73.1 - 77.8)	1.6
	≥23	16 (1.2)	1.5 (0.5 - 3.5)	98.9 (98.1 - 99.5)	1.38 (0.48 - 3.94)	1.00 (0.98 - 1.01)	31.3 (11.0 - 58.7)	75.3 (72.9 - 77.6)	0.4
	≥24	10 (0.7)	0.9 (0.2 - 2.6)	99.3 (98.6 - 99.7)	1.30 (0.34 - 5.00)	1.00 (0.99 - 1.01)	30.0 (6.7 - 65.2)	75.2 (72.8 - 77.5)	0.2
	≥25	4 (0.3)	0.3 (0.0 - 1.7)	99.7 (99.1 - 99.9)	1.01 (0.11 - 9.69)	1.00 (0.99 - 1.01)	25.0 (0.6 - 80.6)	75.2 (72.8 - 77.5)	0.0
	≥26	2 (0.2)	0.0 (0.0 - 1.1)	99.8 (99.3 - 100.0)	0.00 (0.00 - 0.00)	1.00 (1.00 - 1.00)	0.0 (0.0 - 84.2)	75.2 (72.8 - 77.5)	-0.2
	≥27	2 (0.2)	0.0 (0.0 - 1.1)	99.8 (99.3 - 100.0)	0.00 (0.00 - 0.00)	1.00 (1.00 - 1.00)	0.0 (0.0 - 84.2)	75.2 (72.8 - 77.5)	-0.2
PHQ-8	≥1	1244 (92.4)	99.4 (97.9 - 99.9)	10.0 (8.2 - 12.0)	1.10 (1.08 - 1.13)	0.06 (0.01 - 0.24)	26.7 (24.2 - 29.2)	98.1 (93.2 - 99.8)	9.4
	≥2	1141 (84.7)	98.2 (96.1 - 99.3)	19.7 (17.3 - 22.3)	1.22 (1.18 - 1.27)	0.09 (0.04 - 0.20)	28.7 (26.1 - 31.5)	97.1 (93.8 - 98.9)	17.9
	≥3	998 (74.1)	95.5 (92.7 - 97.5)	33.0 (30.1 - 36.0)	1.42 (1.36 - 1.50)	0.14 (0.08 - 0.23)	32.0 (29.1 - 35.0)	95.7 (93.0 - 97.6)	28.5
	≥4	846 (62.8)	91.3 (87.8 - 94.1)	46.6 (43.5 - 49.7)	1.71 (1.60 - 1.83)	0.19 (0.13 - 0.27)	36.1 (32.8 - 39.4)	94.2 (91.8 - 96.1)	37.9
	≥5	701 (52.0)	85.6 (81.4 - 89.2)	59.0 (55.9 - 62.1)	2.09 (1.92 - 2.28)	0.24 (0.19 - 0.32)	40.8 (37.1 - 44.5)	92.6 (90.3 - 94.5)	44.7
	≥6	611 (45.4)	79.9 (75.2 - 84.1)	66.0 (63.0 - 69.0)	2.35 (2.13 - 2.61)	0.30 (0.24 - 0.38)	43.7 (39.7 - 47.7)	90.9 (88.6 - 92.9)	46.0
	≥7	**521 (38.7)**	**74.3 (69.2 - 78.9)**	**73.1 (70.2 - 75.8)**	**2.76 (2.44 - 3.10)**	**0.35 (0.29 - 0.42)**	**47.6 (43.2 - 52.0)**	**89.6 (87.3 - 91.6)**	**47.4**

	≥8	428 (31.8)	65.6 (60.2 - 70.7)	79.4 (76.7 - 81.8)	3.18 (2.75 - 3.67)	0.43 (0.37 - 0.50)	51.2 (46.3 - 56.0)	87.5 (85.2 - 89.6)	44.9
	≥9	357 (26.5)	58.4 (52.9 - 63.7)	84.0 (81.6 - 86.2)	3.65 (3.09 - 4.32)	0.50 (0.44 - 0.56)	54.6 (49.3 - 59.9)	86.0 (83.6 - 88.1)	42.4
	≥10	286 (21.2)	50.0 (44.5 - 55.5)	88.3 (86.1 - 90.2)	4.26 (3.48 - 5.20)	0.57 (0.51 - 0.63)	58.4 (52.4 - 64.2)	84.3 (81.9 - 86.4)	38.3
	≥11	233 (17.3)	42.2 (36.9 - 47.7)	90.9 (89.0 - 92.6)	4.65 (3.69 - 5.86)	0.64 (0.58 - 0.70)	60.5 (53.9 - 66.8)	82.7 (80.3 - 84.9)	33.1
	≥12	189 (14.0)	34.7 (29.6 - 40.1)	92.8 (91.0 - 94.3)	4.82 (3.70 - 6.28)	0.70 (0.65 - 0.76)	61.4 (54.0 - 68.4)	81.2 (78.8 - 83.4)	27.5
	≥13	138 (10.2)	27.2 (22.5 - 32.4)	95.4 (93.9 - 96.6)	5.87 (4.22 - 8.17)	0.76 (0.71 - 0.82)	65.9 (57.4 - 73.8)	79.9 (77.5 - 82.1)	22.6
	≥14	103 (8.1)	21.9 (17.5 - 26.7)	96.4 (95.1 - 97.5)	6.15 (4.21 - 8.99)	0.81 (0.76 - 0.86)	67.0 (57.3 - 75.7)	78.9 (76.5 - 81.2)	18.3
	≥15	80 (5.9)	15.6 (11.9 - 19.9)	97.2 (96.0 - 98.2)	5.63 (3.62 - 8.77)	0.87 (0.83 - 0.91)	65.0 (53.5 - 75.3)	77.7 (75.3 - 80.0)	12.8
	≥16	58 (4.3)	10.5 (7.4 - 14.3)	97.7 (96.6 - 98.6)	4.62 (2.77 - 7.70)	0.92 (0.88 - 0.95)	60.3 (46.6 - 73.0)	76.8 (74.4 - 79.1)	8.2
	≥17	46 (3.4)	7.2 (4.7 - 10.5)	97.8 (96.7 - 98.6)	3.31 (1.88 - 5.82)	0.95 (0.92 - 0.98)	52.2 (36.9 - 67.1)	76.2 (73.8 - 78.5)	5.0
	≥18	36 (2.7)	5.7 (3.5 - 8.7)	98.3 (97.3 - 99.0)	3.39 (1.78 - 6.44)	0.96 (0.93 - 0.99)	52.8 (35.5 - 69.6)	76.0 (73.6 - 78.3)	4.0
	≥19	30 (2.2)	4.8 (2.8 - 7.7)	98.6 (97.7 - 99.2)	3.47 (1.71 - 7.03)	0.97 (0.94 - 0.99)	53.3 (34.3 - 71.7)	75.9 (73.4 - 78.1)	3.4
	≥20	24 (1.8)	3.0 (1.4 - 5.4)	98.6 (97.7 - 99.2)	2.17 (0.97 - 4.83)	0.98 (0.96 - 1.00)	41.7 (22.1 - 63.4)	75.5 (73.1 - 77.8)	1.6
	≥21	15 (1.1)	1.5 (0.5 - 3.5)	99.0 (98.2 - 99.5)	1.52 (0.52 - 4.41)	0.99 (0.98 - 1.01)	33.3 (11.8 - 61.6)	75.3 (72.9 - 77.6)	0.5
	≥22	7 (0.5)	0.6 (0.1 - 2.1)	99.5 (98.9 - 99.8)	1.21 (0.24 - 6.22)	1.00 (0.99 - 1.01)	28.6 (3.7 - 71.0)	75.2 (72.8 - 77.5)	0.1
	≥23	4 (0.3)	0.3 (0.0 - 1.7)	99.7 (99.1 - 99.9)	1.01 (0.11 - 9.69)	1.00 (0.99 - 1.01)	25.0 (0.6 - 80.6)	75.2 (72.8 - 77.5)	0.0
	≥24	2 (0.2)	0.0 (0.0 - 1.1)	99.8 (99.3 - 100.0)	0.00 (0.00 - 0.00)	1.00 (1.00 - 1.00)	0.0 (0.0 - 84.2)	75.2 (72.8 - 77.5)	-0.2
PHQ-9 algorithm		183 (13.6)	34.7 (29.6 - 40.1)	93.4 (91.7 - 94.8)	5.25 (3.99 - 6.91)	0.70 (0.65 - 0.76)	63.4 (56.0 - 70.4)	81.3 (78.9 - 83.5)	28.1
PHQ-9 adjusted algorithm		601 (44.6)	78.1 (73.3 - 82.5)	66.4 (63.4 - 69.3)	2.33 (2.10 - 2.58)	0.33 (0.27 - 0.41)	43.4 (39.4 - 47.5)	90.2 (87.9 - 92.3)	44.5
PHQ-2	≥1	1052 (78.1)	95.2 (92.3 - 97.2)	27.5 (24.8 - 30.4)	1.31 (1.26 - 1.37)	0.17 (0.11 - 0.28)	30.2 (27.5 - 33.1)	94.6 (91.3 - 96.9)	22.7
	≥2	**730 (54.2)**	**84.7 (80.4 - 88.4)**	**55.9 (52.8 - 59.0)**	**1.92 (1.77 - 2.09)**	**0.27 (0.21 - 0.35)**	**38.8 (35.2 - 42.4)**	**91.7 (89.3 - 93.8)**	**40.6**
	≥3	358 (26.6)	55.4 (49.9 - 60.8)	82.9 (80.5 - 85.2)	3.24 (2.75 - 3.83)	0.54 (0.48 - 0.61)	51.7 (46.4 - 57.0)	84.9 (82.6 - 87.1)	38.3
	≥4	225 (16.7)	39.5 (34.2 - 45.0)	90.8 (88.9 - 92.5)	4.30 (3.40 - 5.44)	0.67 (0.61 - 0.73)	58.7 (51.9 - 65.2)	82.0 (79.6 - 84.2)	30.3
	≥5	99 (7.4)	19.2 (15.1 - 23.8)	96.5 (95.2 - 97.6)	5.55 (3.74 - 8.22)	0.84 (0.79 - 0.88)	64.6 (54.4 - 74.0)	78.4 (76.0 - 80.6)	15.7
	≥6	63 (4.7)	12.0 (8.7 - 15.9)	97.7 (96.6 - 98.6)	5.27 (3.21 - 8.68)	0.90 (0.87 - 0.94)	63.5 (50.4 - 75.3)	77.1 (74.7 - 79.4)	9.7

Note: +LR = Positive likelihood ratio. −LR = Negative likelihood ratio. PPV= positive predictive value. NPV= negative predictive value. Youden's Index = Sensitivity+Specificity-1. 95%CI = 95% confidence interval. PHQ = Patient Health Questionnaire. Bold values represent the cohort point with the highest Youden Index.

Supplementary material 6. Sensitivity, specificity, PPV, NPV, and Youden's index for different PHQ and GAD cut-off points compared to the gold standard.

Method	Cut-off Points	n (%)	Sensitivity (95%CI)	Specificity (95%CI)	PPV (95%CI)	NPV (95%CI)	Youden Index
PHQ-9	**≥7**	**537 (39.9)**	**76.0 (71.1 - 80.5)**	**72.1 (69.2 - 74.8)**	**47.3 (43.0 - 51.6)**	**90.1 (87.9 - 92.1)**	**48.1**
PHQ-9	≥8	454 (33.7)	68.9 (63.6 - 73.8)	77.9 (75.2 - 80.4)	50.7 (46.0 - 55.4)	88.4 (86.1 - 90.4)	46.8
PHQ-9	≥9	371 (27.5)	60.2 (54.7 - 65.5)	83.2 (80.8 - 85.5)	54.2 (49.0 - 59.3)	86.4 (84.1 - 88.5)	43.4
PHQ-9	≥10	309 (22.9)	54.2 (48.7 - 59.6)	87.4 (85.2 - 89.3)	58.6 (53.9 - 63.1)	85.3 (83.7 - 86.7)	41.6
PHQ-9 algorithm	-	183 (13.6)	34.7 (29.6 - 40.1)	93.4 (91.7 - 94.8)	63.4 (56.0 - 70.4)	81.3 (78.9 - 83.5)	28.1
PHQ-9 adjusted algorithm	-	601 (44.6)	78.1 (73.3 - 82.5)	66.4 (63.4 - 69.3)	43.4 (39.4 - 47.5)	90.2 (87.9 - 92.3)	44.5
PHQ-8	**≥7**	**521 (38.7)**	**74.3 (69.2 - 78.9)**	**73.1 (70.2 - 75.8)**	**47.6 (43.2 - 52.0)**	**89.6 (87.3 - 91.6)**	**47.4**
PHQ-8	≥8	428 (31.8)	65.6 (60.2 - 70.7)	79.4 (76.7 - 81.8)	51.2 (46.3 - 56.0)	87.5 (85.2 - 89.6)	44.9
PHQ-8	≥9	357 (26.5)	58.4 (52.9 - 63.7)	84.0 (81.6 - 86.2)	54.6 (49.3 - 59.9)	86.0 (83.6 - 88.1)	42.4
PHQ-8	≥10	286 (21.2)	50.0 (44.5 - 55.5)	88.3 (86.1 - 90.2)	58.4 (52.4 - 64.2)	84.3 (81.9 - 86.4)	38.3
PHQ-2	**≥2**	**730 (54.2)**	**84.7 (80.4 - 88.4)**	**55.9 (52.8 - 59.0)**	**38.8 (35.2 - 42.4)**	**91.7 (89.3 - 93.8)**	**40.6**
PHQ-2	≥3	358 (26.6)	55.4 (49.9 - 60.8)	82.9 (80.5 - 85.2)	51.7 (46.4 - 57.0)	84.9 (82.6 - 87.1)	38.3
GAD-7	≥7	360 (26.7)	53.6 (33.9 - 72.5)	73.8 (71.4 - 76.2)	4.2 (2.4 - 6.8)	98.7 (97.8 - 99.3)	27.4
GAD-7	**≥8**	**295 (21.9)**	**53.6 (33.9 - 72.5)**	**78.8 (76.5 - 81.0)**	**5.1 (2.9 - 8.3)**	**98.8 (97.9 - 99.3)**	**32.4**
GAD-7	≥9	230 (17.1)	42.9 (24.5 - 62.8)	83.5 (81.4 - 85.4)	5.2 (2.7 - 8.9)	98.6 (97.7 - 99.2)	26.4
GAD-7	≥10	164 (12.2)	39.3 (21.5 - 59.4)	88.4 (86.5 - 90.1)	6.7 (3.4 - 11.7)	98.6 (97.7 - 99.2)	27.7
GAD-2	**≥2**	**681 (50.6)**	**82.1 (63.1 - 93.9)**	**50.1 (47.4 - 52.8)**	**3.4 (2.2 - 5.0)**	**99.2 (98.3 - 99.8)**	**32.2**
GAD-2	≥3	283 (21.0)	42.9 (24.5 - 62.8)	79.5 (77.2 - 81.6)	4.2 (2.2 - 7.3)	98.5 (97.6 - 99.1)	22.4

Note: PPV= positive predictive value. NPV= negative predictive value. Youden's Index = Sensitivity+Specificity-1. 95%CI = 95% confidence interval. GAD =Generalized Anxiety Disorder Scale. PHQ = Patient Health Questionnaire. Bold values represent the cohort point with the highest Youden Index. All cut-off points for the PHQ and GAD versions can be found in supplementary materials 3 and 4, respectively.

Supplementary material 7. Sensitivity, specificity, +LR, −LR, PPV, NPV, and Youden's index for different GAD cut-off points compared to the gold standard.

Method	Cut-off Points	n (%)	Sensitivity (95%CI)	Specificity (95%CI)	+LR (95%CI)	-LR (95%CI)	PPV (95%CI)	NPV (95%CI)	Youden Index
GAD-7	≥1	1244 (92.4)	96.4 (81.7 - 99.9)	7.7 (6.4 - 9.3)	1.05 (0.97 - 1.12)	0.46 (0.07 - 3.19)	2.2 (1.4 - 3.1)	99.0 (94.7 - 100.0)	4.1
	≥2	1128 (83.7)	96.4 (81.7 - 99.9)	16.5 (14.6 - 18.6)	1.16 (1.07 - 1.25)	0.22 (0.03 - 1.49)	2.4 (1.6 - 3.5)	99.5 (97.5 - 100.0)	12.9
	≥3	960 (71.3)	96.4 (81.7 - 99.9)	29.3 (26.8 - 31.8)	1.36 (1.26 - 1.48)	0.12 (0.02 - 0.84)	2.8 (1.9 - 4.1)	99.7 (98.6 - 100.0)	25.7
	≥4	797 (59.2)	89.3 (71.8 - 97.7)	41.5 (38.8 - 44.2)	1.53 (1.33 - 1.75)	0.26 (0.09 - 0.75)	3.1 (2.0 - 4.6)	99.5 (98.4 - 99.9)	30.8
	≥5	640 (47.5)	71.4 (51.3 - 86.8)	53.0 (50.3 - 55.7)	1.52 (1.19 - 1.93)	0.54 (0.30 - 0.97)	3.1 (1.9 - 4.8)	98.9 (97.8 - 99.5)	24.4
	≥6	491 (36.5)	60.7 (40.6 - 78.5)	64.1 (61.4 - 66.7)	1.69 (1.24 - 2.30)	0.61 (0.39 - 0.97)	3.5 (2.0 - 5.5)	98.7 (97.7 - 99.4)	24.8
	≥7	360 (26.7)	53.6 (33.9 - 72.5)	73.8 (71.4 - 76.2)	2.05 (1.43 - 2.93)	0.63 (0.42 - 0.94)	4.2 (2.4 - 6.8)	98.7 (97.8 - 99.3)	27.4
	≥8	**295 (21.9)**	**53.6 (33.9 - 72.5)**	**78.8 (76.5 - 81.0)**	**2.52 (1.76 - 3.62)**	**0.59 (0.40 - 0.88)**	**5.1 (2.9 - 8.3)**	**98.8 (97.9 - 99.3)**	**32.4**
	≥9	230 (17.1)	42.9 (24.5 - 62.8)	83.5 (81.4 - 85.4)	2.59 (1.66 - 4.04)	0.69 (0.50 - 0.94)	5.2 (2.7 - 8.9)	98.6 (97.7 - 99.2)	26.4
	≥10	164 (12.2)	39.3 (21.5 - 59.4)	88.4 (86.5 - 90.1)	3.39 (2.09 - 5.50)	0.69 (0.51 - 0.93)	6.7 (3.4 - 11.7)	98.6 (97.7 - 99.2)	27.7
	≥11	120 (8.9)	39.3 (21.5 - 59.4)	91.7 (90.1 - 93.2)	4.75 (2.90 - 7.79)	0.66 (0.49 - 0.89)	9.2 (4.7 - 15.8)	98.6 (97.8 - 99.2)	31.0
	≥12	94 (7.0)	28.6 (13.2 - 48.7)	93.5 (92.0 - 94.8)	4.38 (2.36 - 8.15)	0.76 (0.60 - 0.97)	8.5 (3.8 - 16.1)	98.4 (97.5 - 99.0)	22.1
	≥13	67 (5.0)	28.6 (13.2 - 48.7)	95.5 (94.3 - 96.6)	6.39 (3.39 - 12.10)	0.75 (0.59 - 0.95)	11.9 (5.3 - 22.2)	98.4 (97.6 - 99.0)	24.1
	≥14	50 (3.7)	21.4 (8.3 - 41.0)	96.7 (95.5 - 97.6)	6.42 (2.98 - 13.80)	0.81 (0.67 - 0.99)	12.0 (4.5 - 24.3)	98.3 (97.4 - 98.9)	18.1
	≥15	40 (3.0)	14.3 (4.0 - 32.7)	97.3 (96.2 - 98.1)	5.23 (2.00 - 13.70)	0.88 (0.76 - 1.03)	10.0 (2.8 - 23.7)	98.2 (97.3 - 98.8)	11.6
	≥16	31 (2.3)	10.7 (2.3 - 28.2)	97.9 (96.9 - 98.6)	5.05 (1.63 - 15.60)	0.91 (0.80 - 1.04)	9.7 (2.0 - 25.8)	98.1 (97.2 - 98.8)	8.6
	≥17	19 (1.4)	3.6 (0.1 - 18.3)	98.6 (97.9 - 99.2)	2.62 (0.36 - 18.90)	0.98 (0.91 - 1.05)	5.3 (0.1 - 26.0)	98.0 (97.1 - 98.7)	2.2
	≥18	15 (1.1)	0.0 (0.0 - 12.3)	98.9 (98.1 - 99.4)	0.00 (0.00 - 0.00)	1.01 (1.01 - 1.02)	0.0 (0.0 - 21.8)	97.9 (97.0 - 98.6)	-1.1
	≥19	12 (0.9)	0.0 (0.0 - 12.3)	99.1 (98.4 - 99.5)	0.00 (0.00 - 0.00)	1.01 (1.00 - 1.01)	0.0 (0.0 - 26.5)	97.9 (97.0 - 98.6)	-0.9
	≥20	8 (0.6)	0.0 (0.0 - 12.3)	99.4 (98.8 - 99.7)	0.00 (0.00 - 0.00)	1.01 (1.00 - 1.01)	0.0 (0.0 - 36.9)	97.9 (97.0 - 98.6)	-0.6
	≥21	5 (0.4)	0.0 (0.0 - 12.3)	99.6 (99.1 - 99.9)	0.00 (0.00 - 0.00)	1.00 (1.00 - 1.01)	0.0 (0.0 - 52.2)	97.9 (97.0 - 98.6)	-0.4
GAD-2	≥1	1028 (76.3)	89.3 (71.8 - 97.7)	24.0 (21.7 - 26.4)	1.17 (1.03 - 1.34)	0.45 (0.15 - 1.31)	2.4 (1.6 - 3.6)	99.1 (97.3 - 99.8)	13.3
	≥2	**681 (50.6)**	**82.1 (63.1 - 93.9)**	**50.1 (47.4 - 52.8)**	**1.65 (1.37 - 1.97)**	**0.36 (0.16 - 0.79)**	**3.4 (2.2 - 5.0)**	**99.2 (98.3 - 99.8)**	**32.2**
	≥3	283 (21.0)	42.9 (24.5 - 62.8)	79.5 (77.2 - 81.6)	2.09 (1.34 - 3.24)	0.72 (0.52 - 0.99)	4.2 (2.2 - 7.3)	98.5 (97.6 - 99.1)	22.4
	≥4	141 (10.5)	25.0 (10.7 - 44.9)	89.8 (88.1 - 91.4)	2.46 (1.27 - 4.77)	0.84 (0.67 - 1.03)	5.0 (2.0 - 10.0)	98.3 (97.4 - 98.9)	14.8
	≥5	51 (3.8)	14.3 (4.0 - 32.7)	96.4 (95.3 - 97.4)	4.01 (1.55 - 10.40)	0.89 (0.76 - 1.03)	7.8 (2.2 - 18.9)	98.1 (97.3 - 98.8)	10.7
	≥6	28 (2.1)	7.1 (0.9 - 23.5)	98.0 (97.1 - 98.7)	3.62 (0.90 - 14.50)	0.95 (0.86 - 1.05)	7.1 (0.9 - 23.5)	98.0 (97.1 - 98.7)	5.1

Note: +LR = Positive likelihood ratio. −LR = Negative likelihood ratio. PPV= positive predictive value. NPV= negative predictive value. Youden's Index = Sensitivity+Specificity-1. 95%CI = 95% confidence interval. GAD =Generalized Anxiety Disorder Scale. Bold values represent the cut-off with the highest Youden Index.

STROBE Statement—Checklist of items that should be included in reports of ***cross-sectional studies***

	Item No	Recommendation	Page
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	3
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	5
Objectives	3	State specific objectives, including any prespecified hypotheses	6
Methods			
Study design	4	Present key elements of study design early in the paper	6
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	6
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants	6
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	7
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	*7*
Bias	9	Describe any efforts to address potential sources of bias	
Study size	10	Explain how the study size was arrived at	6
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	8
		(*b*) Describe any methods used to examine subgroups and interactions	8
		(*c*) Explain how missing data were addressed	8
		(*d*) If applicable, describe analytical methods taking account of sampling strategy	8
		(*e*) Describe any sensitivity analyses	9
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	9
		(b) Give reasons for non-participation at each stage	9
		(c) Consider use of a flow diagram	SUPPL.
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	10
		(b) Indicate number of participants with missing data for each variable of interest	10
Outcome data	15*	Report numbers of outcome events or summary measures	10
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	10

		(*b*) Report category boundaries when continuous variables were categorized	10
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	
Discussion			
Key results	18	Summarise key results with reference to study objectives	11
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	13
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	11
Generalisability	21	Discuss the generalisability (external validity) of the study results	11
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	16

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Sensitivity and specificity of the Patient Health Questionnaire (PHQ-9, PHQ-8, PHQ-2) and General Anxiety Disorder scale (GAD-7, GAD-2) for depression and anxiety diagnosis: A cross-sectional study in a Peruvian hospital population

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2023-076193.R2
Article Type:	Original research
Date Submitted by the Author:	18-Aug-2023
Complete List of Authors:	Villarreal-Zegarra, David; Instituto Peruano de Orientación Psicológica, ; Universidad Peruana Cayetano Heredia, CRONICAS Center of Excellence in Chronic Diseases Barrera-Begazo, Juan ; Instituto Peruano de Orientación Psicológica Otazú-Alfaro, Sharlyn; Instituto Peruano de Orientación Psicológica Mayo-Puchoc, Nikol; Instituto Peruano de Orientación Psicológica Bazo-Alvarez, Juan Carlos ; Universidad Norbert Wiener Huarcaya-Victoria , Jeff; Universidad de San Martín de Porres Facultad de Medicina Humana,
Primary Subject Heading:	Mental health
Secondary Subject Heading:	Mental health
Keywords:	Sensitivity and Specificity, Depression & mood disorders < PSYCHIATRY, Anxiety disorders < PSYCHIATRY

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Sensitivity and specificity of the Patient Health Questionnaire (PHQ-9, PHQ-8, PHQ-2) and General Anxiety Disorder scale (GAD-7, GAD-2) for depression and anxiety diagnosis: A cross-sectional study in a Peruvian hospital population

Authors

David Villarreal-Zegarra [1, 2]
Juan Barrera [2]
Sharlyn Otazú-Alfaro [2]
Nikol Mayo-Puchoc [2]
Juan Carlos Bazo-Alvarez [3]
Jeff Huarcaya-Victoria [4,5]

Affiliation

[1] Universidad César Vallejo, Escuela de Medicina. Trujillo, Peru
[2] Instituto Peruano de Orientación Psicológica, Lima, Peru
[3] Universidad Privada Norbert Wiener, Lima, Peru
[4] Escuela Profesional de Medicina Humana, Universidad Privada San Juan Bautista, Filial Ica, Peru
[5] Unidad de Psiquiatría de Enlace, Departamento de Psiquiatría, Hospital Nacional Guillermo Almenara Irigoyen, EsSalud, Lima, Peru

Email

David Villarreal-Zegarra davidvillarreal@ipops.pe
Juan Barrera juanbarrera@ipops.pe
Sharlyn Otazú-Alfaro sharlynotazu@ipops.pe
Nikol Mayo-Puchoc nikolmayo@ipops.pe
Juan Carlos Bazo-Alvarez juan.bazo@uwiener.edu.pe
Jeff Huarcaya-Victoria jhuarcayav@usmp.pe

Corresponding author

Jeff Huarcaya-Victoria
jeff.huarcaya@upsjb.edu.pe

Abstract

Objectives: The Patient Health Questionnaire (PHQ) and Generalized Anxiety Disorder Scale (GAD) are widely used screening tools, but their sensitivity and specificity in low- and middle-income countries are lower than in high-income countries. We conducted a study to determine the sensitivity and specificity of different versions of these scales in a Peruvian hospital population.

Design: Our study has a cross-sectional design.

Setting: Our participants are hospitalised patients in a Peruvian hospital. The gold standard was a clinical psychiatric interview following ICD-10 criteria for depression (F32.0, F32.1, F32.2, and F32.3) and anxiety (F41.0 and F41.1).

Participants: The sample included 1347 participants. A total of 334 participants (24.8%) were diagnosed with depression, and 28 participants (2.1%) were diagnosed with anxiety.

Results: The PHQ-9's ≥7 cut-off point showed the highest simultaneous sensitivity and specificity when contrasted against a psychiatric diagnosis of depression. For a similar contrast against the gold standard, the other optimal cut-off points were: ≥7 for the PHQ-8, and ≥2 for the PHQ-2. In particular, the cut-off point ≥8 had good performance for GAD-7 with sensitivity and specificity, and cut-off point ≥10 had lower levels of sensitivity, but higher levels of specificity, compared to the cut-off point of ≥8. Also, we present the sensitivity and specificity values of each cut-off point in PHQ-9, PHQ-8, PHQ-2, GAD-7 and GAD-2. We confirmed the adequacy of a one-dimensional model for the PHQ-9, PHQ-8, and GAD-7, while all PHQ and GAD scales showed good reliability.

Conclusions: The PHQ and GAD have adequate measurement properties in their different versions. We present specific cut-offs for each version.

Keywords: Anxiety; Depression; Patient Health Questionnaire; Peru; Sensitivity and Specificity.

Strengths and limitations of this study

- Study methods allowed us to establish clinically meaningful cut-off points for PHQ and GAD.
- Sample size was larger than in other similar studies and large enough to support all analyses and conclusions.
- Research findings may not be directly applicable to some hospital or primary care settings due to the specific context of our study population.

Background

Until 2019, approximately 280 million people worldwide suffered from depression and 302 million from anxiety [1]. These data reveal that both mental disorders are the most common in the world and lead to the causes of the global burden of mental health disability-adjusted life years[2, 3]. With the onset of the coronavirus disease (COVID-19) pandemic, the worldwide prevalence of both disorders increased by around 25% [4]. In Peru, during the COVID-19 pandemic, the prevalence of moderate depressive symptoms also increased by approximately 0.17% in each quarter [5]. However, no population-level evidence has been found about the prevalence of anxious symptomatology or the diagnosis of anxiety in Peru. In this context, the impact of COVID-19 on the prevalence and burden of major depression and anxiety disorders was measured using screening tools [6]. In addition, it was noted that during the pandemic, there was a reduction in the number of mental health service users being seen [7].

Screening tools assist in early diagnosis and intervention that can prevent disease progression and reduce years lost to disability [8]. They are beneficial in contexts with limited mental health professionals providing care to large populations, such as in Peru. The opportune identification of people at risk of depression reduces treatment costs and disease burden [9-11]. Depressive symptom screening is also helpful in national surveys and epidemiological research [12] since, unlike diagnostic instruments, screening measures are typically brief, quick, and easy to administer [13, 14]. Internationally, the most used screening instruments for depressive and anxious symptomatology are the Patient Health Questionnaire (PHQ-9) [15], PHQ-8 [16], PHQ-2 [17], Generalized Anxiety Disorder (GAD-7) [18], GAD-2 [18], Depression, Anxiety and Stress Scale (DASS-21), Kessler scale-10, Hospital Anxiety and Depression Scale (HADS) [19], Five Well-Being Index (WHO-5) [10]. Most have been validated in several countries, but only the PHQ and GAD have been validated in the Peruvian context [20, 21].

In particular, the PHQ versions (PHQ-9, PHQ-8, PHQ-2) and GAD versions (GAD-7, GAD-2) are the most widely used, having extensive evidence of their validity and reliability [22-24]. However, correctly identifying people at risk of depression or anxiety requires more than internal/externally valid and reliable screening measures; defining an accurate cut-off point for their raw scales (i.e., to reach valid interpretations) is also necessary. Such a cut-off point can vary across cultures and sub-populations (e.g., general versus clinical), so a local calibration is usually needed [25]. Studies of the different versions of the PHQ and GAD have yielded heterogeneous cut-offs, as they vary between different cultures [21, 26-29] and populations, such as clinical and general populations [30-32]. However, several systematic reviews suggest that cut-off 10 is most appropriate for the PHQ-9, PHQ-8, and GAD-7 [33-37], and cut-offs 2-3 for the PHQ-2 and GAD-2 [35, 37]. Furthermore, concerning the PHQ-9 correctness, the summed item score method is the most used compared to the algorithm. However, other forms of correction using diagnostic algorithms are available [38, 39].

Sensitivity and specificity studies have been barely performed in low- and middle-income countries [40]. Several of these populations do not count with verified cut-off points from calibration studies (including Peruvian populations), in particular, the inpatient population is particularly vulnerable as

they have physical comorbidities that may influence the establishment of cohort points. Therefore, our aim was to determine the optimal cut-off point for the PHQ-9, PHQ-8, PHQ-2, GAD-7, and GAD-2 to discriminate a formal depression and anxiety diagnosis in the Peruvian hospital population. In addition, as secondary objectives, we assessed these scales' internal structure and reliability.

Methods

Study design

This study has a cross-sectional design, and we used the STARD 2015 guidelines for reporting diagnostic accuracy studies [41].

Participants

The participants were patients from the Liaison Psychiatry Unit of a hospital in Lima, Peru. Psychiatric liaison services provide psychiatric consultation to hospitalised patients with medical or surgical conditions that have a co-existing psychiatric illness or need for psychiatric assessment and management. The total number of participants in our study is similar to the proportion of people who were hospitalised in 2022 in our setting (see supplementary material 1). The evaluation period started in September 2020 and finished in August 2022. Sampling was non-probabilistic and applied to all participants arriving at the Liaison Psychiatry Unit. The inclusion criteria were that they had complete PHQ-9 and GAD-7 data and were of legal age (>18 years). Participants with missing data were excluded.

The sample size calculation for the PHQ versions was based on an estimated sensitivity of 0.88 and specificity of 0.85 [33], a confidence level of 95%, a prevalence of 6.4% [42, 43], and a drop-out rate of 10%, giving an estimate of 705 participants. The sample size calculation for the GAD versions was based on an estimated sensitivity of 0.83 and specificity of 0.84 [18], a confidence level of 95%, a prevalence of 8.7% [44], and a drop-out rate of 10%, giving an estimate of 694 participants. The web program based on the paper by Burderer (1996) was used to calculate the sample size [45].

Setting

The Guillermo Almenara Irigoyen National Hospital (HNGAI) was the study site, a highly complex hospital in Lima-Peru (capital city). HNGAI is one of the three largest hospitals of the Social Security system in Peru based on the number of beds (960 hospital beds) and is also a tertiary referral centre for all medical specialities, including psychiatry (http://www.essalud.gob.pe/estadistica-institucional/). It provides health care services to 1,547,840 individuals from social insurance. Because it attends to virtually all pathologies, from the simplest to the most complex, it was classified in 2015 as a Specialized Health Institute III-2, the highest level awarded by the Ministry of Health of Peru to hospital establishments.

The Liaison Psychiatry Unit at HNGAI is responsible for responding to consultation requests from different clinical-surgical services at HNGAI [46]. As part of the evaluation of each patient, in addition to the clinical interview and psychiatric diagnosis, standardised assessments such as the PHQ-9 and GAD-7 are used to ensure adequate monitoring and assess response to the established treatment. Since September 2020, the services provided by the Liaison Unit have been recorded in a Google Form to track better the patients treated.

Instruments and variables

Patient Health Questionnaire (PHQ-9, PHQ-8, and PHQ-2)

The Patient Health Questionnaire is an instrument designed to measure depressive symptoms over the past two weeks, according to the diagnostic criteria of the DSM-IV, criteria that were retained in the DSM-5. The scale has four response options (0=no days, 1=some days, 2=more than half of the days, 3=almost every day) [15]. The scale had many versions, including the PHQ-9, the full version with nine items and scores ranging from 0 to 27. In Peru, the PHQ-9 had good psychometric properties in terms of structural validity (CFI = 0.936; RMSEA = 0.089; SRMR = 0.039), internal consistency ($\alpha = \omega = 0.87$) and invariance between age and sex ($\Delta CFI<0.01$) [20].

In addition, PHQ-9 had scoring versions related to the DSM-5 indicators, which state that for a case to be positive, there must be at least five depressive symptoms present, and at least one of them must be core depressive symptoms (item 1 and item 2). First, the PHQ-9 algorithm suggests that a symptom is positive if it scores two or more, except the ninth item, suicidal ideation, which is positive if it scores one or more [47]. Second, the PHQ-9 adjusted algorithm proposes that a symptom was positive if it scored 1 or more for any of the items in the instrument [48].

The PHQ-8 was a shortened version of the PHQ-9 without the last item on suicidal ideation [16]. The PHQ-8 was as valuable as the PHQ-9 in detecting cases of major depression [49]. The PHQ-2 is an abbreviated version of the PHQ-9 with only two items, focusing on the first two items related to the core symptoms of depression (anhedonia and depressed mood) and providing scores between 0 and 6. The PHQ-2 was validated in Peru and showed adequate levels of internal consistency ($\alpha = .80$) [50].

General Anxiety Disorder Scale (GAD-7 and GAD-2)

The General Anxiety Disorder Scale was a Likert-type rating scale with four response options ranging from 0 (not at all) to 3 (almost every day), based on DSM-IV criteria and assesses anxious symptoms during the past two weeks [51]. The GAD-7 was the version of the instrument with the original seven items and had a range of scores from 0 to 21. The GAD-7 had good psychometric properties in the Peruvian context for a one-dimensional model (CFI = .995, TLI = .992, RMSEA = .056), adequate internal consistency ($\omega = .89$), and invariance according to sex ($\Delta CFI \leq .01$) [52].

The GAD-2 was adapted from the GAD-7, focusing on the emotional and cognitive expressions of DSM-IV anxiety (items 1 and 2) [53]. The GAD-2 shows good internal consistency values (ω =.80) and a relationship with its extended version (r>0.80) in Peruvian context [52].

Gold standard

The gold standard was an individual clinical psychiatric interview following ICD-10 criteria. The clinical assessments were performed by psychiatrists who are members of the Liaison Psychiatry Unit, all of whom have at least five years of clinical experience evaluating the psychiatric needs of hospitalised patients. The interview focused on assessing whether the participants had depressive disorder (F32.0, F32.1, F32.2, and F32.3) or anxiety disorder (F41.0 and F41.1), with a duration between 25 to 30 minutes. The individual clinical psychiatric interview and the psychometric instruments (i.e., PHQ and GAD) were independently applied on the same day, the latter by a mental health nurse or a psychologist and the former by a psychiatrist. The average time between both measurements was 15 minutes (standard deviation = 4.5 minutes), and the order (i.e., psychometric instruments before or after the interview) was randomly assigned.

Sociodemographic covariates

Data was collected on sex (male, female), age, marital status (Single, Married/Cohabitant, Separated, Widowed), educational level (None, Elementary, High School, Technical, College), currently works (No, yes, retired), living alone (Yes, No), and history of psychiatric diagnosis (yes/no). In addition, information was collected on the physical diagnosis of the participants based on the ICD-10.

Statistical analysis

The sociodemographic covariates of the participants were described at frequency and percentage levels. The internal consistency and internal structure analyses were performed with R Studio, with the “Lavaan”, “Semtools”, and “Semplot” packages (see supplementary material 2). Sensitivity, specificity, and correlation analyses were analysed with Stata 15 (see supplementary material 3).

Sensibility and Specificity

The PHQ-9, PHQ-8, PHQ-9 algorithm, PHQ-9 adjusted algorithm, and PHQ-2 were evaluated as diagnostic tests and compared against the gold standard. In addition, the GAD-7 and GAD-2 were scored and compared against the diagnosis of anxiety through the clinical interview (gold standard).

We calculated the positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (+LR), negative likelihood ratio (-LR), and Youden Index. PPV and NPV refer to the proportion of patients correctly diagnosed as positive or negative, respectively [54]. The LR+ is the probability that a person with the disease will test positive given the probability that a person without the disease will test positive [55]. While the LR- is the probability that a person with the disease will test negative given the probability that a person without the disease will test negative [55]. The Youden Index is a measure that summarizes the performance of a diagnostic test by interpreting it as the probability that the selected cut-off point provides an adequate clinical decision (in terms of sensitivity and specificity), as

opposed to the probability that the selected cut-off provides a random decision [54]. The maximum value of the Youden Index was used as a criterion to select the cut-off with the best diagnostic performance for each scale. Values closer to 1 were considered optimal, and those closer to 0 were considered inadequate.

Internal structure

Confirmatory factor analysis (CFA) was performed considering a one-dimensional model for the PHQ-9, PHQ-8, and GAD-7. We used the weighted least square mean and variance adjusted (WLSMV) estimator [56] and polychoric matrices as it best fits the categorical-ordinal nature of the data [57]. Models were evaluated using a set of goodness-of-fit indices such as CFI and TLI, which must be greater than 0.95 to be considered adequate [58]. In addition, the SRMR and RMSEA at 90% confidence were estimated, which must have values less than 0.08 to be considered adequate [58]. It was impossible to perform a CFA for the PHQ-2 and the GAD-2 because a minimum of three items are required for such analysis.

Internal consistency

We calculated the alpha (α) and McDonald's omega coefficients (ω). Values greater than 0.70 are considered adequate [59].

Ethics

The Hospital Nacional Guillermo Almenara Irigoyen's Institutional Review Board (Nota N°52 CIEI-OIyD-GRPA-Essalud-2023) approved the protocol of our study. Throughout the study, the researchers had no access to identifying information about the participants. In addition, participants gave informed consent. All participants were users of the hospital's Liaison Psychiatry Unit and received psychological or psychiatric care as needed.

Patient and Public Involvement

No patient involved.

Results

Participants

We collected data from 4979 attendances performed within the liaison psychiatry service during the study period. However, some of these attendances were not assessed with PHQ-9 or GAD-7 data (n=3484) or lacked sociodemographic information (n=148) and were eliminated (see supplementary material 4). Thus, our study only included 1347 participants (see Table 1). Most participants were female (59.4%; n=800), married or living with a partner (57.0%; n=768), and had higher technical or university education (53.5%; n=721). A total of 334 participants (24.8%) were diagnosed with

depression, and 28 participants (2.1%) were diagnosed with anxiety, as determined through individual psychiatric interviews conducted based on the ICD-10 criteria.

The most common physical morbidities were cardiovascular diseases (n=111; 8.2%), endocrine, nutritional and metabolic diseases (n=130; 9.7%) and neoplasms, diseases of the blood and haematopoietic organs and other diseases affecting the mechanism of immunity (n=348; 25.8%).

Sensibility and Specificity

In supplementary material 5, we provide the values of all cut-off points for the different versions of the PHQ. The cut-off points ≥7 in the PHQ-9 had the best balance between sensitivity and specificity of all the cut-off points evaluated in the various versions of the PHQ, as it obtained a sensitivity of 76.0 (95%CI: 71.1 - 80.5) and specificity of 72.1 (95%CI: 69.2 - 74.8) (see supplementary material 6). In addition, the PHQ-9 with a cut-off of ≥10 points (i.e., the most used) showed lower levels of sensitivity (54.2; 95%CI: 8.7 - 59.6), but higher level of specificity (87.4; 95%CI: 85.2 - 89.3), compared to the cut-off point of ≥7.

The algorithm score method for PHQ-9 had low levels of sensitivity (34.7; 95%CI: 29.6 - 40.1) but high levels of specificity (93.4; 95%CI: 91.7 - 94.8) compared to the raw score method for PHQ-9 with ≥7 cohort points. In contrast, the adjusted algorithm method for PHQ-9 showed slightly higher sensitivity values (78.1; 95%CI: 73.3 - 82.5) and better specificity values (66.4; 95%CI: 63.4 - 69.3) compared to the raw score method for PHQ-9 with ≥7 cohort points. The raw score for PHQ-9 with cohort point ≥7 showed a better balance between sensitivity and specificity compared to the algorithm method or the algorithm adjusted for PHQ-9.

The best cut-off point found in the PHQ-8 was ≥7 points, as it had a sensitivity of 79.9 (95%CI: 75.2 - 84.1), and a specificity of 66.0 (95%CI: 63.0 - 69.0) (see supplementary material 6). The best cut-off point found in the PHQ-2 was ≥2 points, as it had a sensitivity of 84.7 (95%CI: 80.4 - 88.4), and a specificity of 55.9 (95%CI: 52.8 - 59.0) (see supplementary material 6).

Because we have a small number of cases with truly anxious people, any changes in the scores of these people could lead to large changes in sensitivity and specificity. Therefore, it is not possible to give an optimal cohort score over the rest, but we present all cohort scores in Supplementary Material 7. In particular, the cut-off point ≥8 had good performance for GAD-7 with sensitivity values of 53.6 (95%CI: 33.9 - 72.5) and specificity of 78.8 (95%CI: 76.5 - 81.0), (see supplementary material 6). The GAD-7's cut-off point ≥10 (i.e., the most used) had lower levels of sensitivity (39.3; 95%CI: 21.5 - 59.4), but higher levels of specificity (88.4; 95%CI: 86.5 - 90.1, compared to the cut-off point of ≥8. In addition, the cut-off point for the GAD-2 was ≥2 had a sensitivity of 84.7 (95%CI: 80.4 - 88.4), and a specificity of 50.1 (95%CI: 47.4 - 52.8) (see supplementary material 6).

Internal structure

The PHQ-9 one-dimensional model showed adequate goodness-of-fit (X^2=251.9; df=27; CFI=0.974; TLI=0.965; SRMR=0.051; RMSEA[90%CI]=0.079[0.070-0.088]), while the PHQ-8 one-dimensional model reported a similar goodness-of-fit (X^2=202.7; df=20; CFI=0.977; TLI=0.977; SRMR=0.050; RMSEA[90%CI]=0.082[0.072-0.093]). The GAD-7 also showed adequate goodness-of-fit (X^2=122.3; df=14; CFI=0.977; TLI=0.966; SRMR=0.043; RMSEA[90%CI]=0.076[0.064-0.088]).

Reliability

The PHQ-9 (α=0.89; ω=0.86), the PHQ-8 (α=0.88; ω=0.85), and the GAD-7 (α=0.85; ω=0.81) showed optimal internal consistency values. Similarly, the PHQ-2 (α=0.83; ω=0.80) and the GAD-2 (α=0.74; ω=0.70) also showed adequate internal consistency scores. Table 2 shows the raw scores.

Discussion

Main findings

We determined the target population's optimal cut-off points for PHQ scale. The PHQ-9's ≥7 cut-off point showed the highest sensitivity and specificity when contrasted against a psychiatric diagnosis of depression (gold standard). For a similar contrast, the other optimal cut-off points were: ≥7 for the PHQ-8, and ≥2 for the PHQ-2. In addition, the algorithm scoring or algorithm-adjusted scoring methods for the PHQ-9 had a lower balance between sensitivity and specificity scores than the PHQ-9 raw score scoring method with a cut-off ≥7. In the case of GAD, the small number of participants with actual anxiety made it impossible to determine an optimal cut-off point. However, we present the sensitivity and specificity of each cut-off point. We confirmed the adequacy of a one-dimensional model for the PHQ-9, PHQ-8, and GAD-7, while all scales showed good internal consistency.

Contrast to literature

At the PHQ-9 level, evidence suggests that the raw score approach is more valuable than diagnostic algorithms [33], which is consistent with our findings. For the cut-off, different systematic reviews agree that the most commonly used cut-off is ≥10 [33, 60]. The optimal cut-off reported in our study was slightly lower than that suggested by the other studies, and two possible factors could explain this difference. Firstly, our population is inpatients in different areas of a high-complexity hospital. Other studies of hospitalised cancer patients [61], hospitalised neurology patients [62], and patients with coronary heart disease [63] also found an optimal cut-off between 5 and 7 points. Therefore, hospitalised individuals may be more likely to have depressive symptoms, which may require a lower cut-off on the PHQ-9. Second, several studies in populations from low- and middle-income countries have reported cut-offs between 5 and 7, for example, Pakistani migrants in the UK [64], Indian adolescents [65], and primary care in Ethiopia [66]. One reason for the difference in cut-off points between high and low-income countries may be due to cultural factors, as culturally diverse groups do

not achieve invariance between the PHQ-9 and the GAD-7 [67]. Therefore, factors such as social determinants of health present in such countries may influence cut-off.

Concerning the PHQ-8 and PHQ-9, we found that both scales have similar cut-off points (≥ 7). Our findings are consistent with a meta-analysis that found that the cut-offs between the two scales are identical; although sensitivity may be minimally reduced with the PHQ-8, specificity is similar between the two scales [36]. The PHQ-8 does not include the item corresponding to suicidal or self-harming ideation, and the use of this version of the PHQ is common in the general population, as suicidal ideation is less common in this group [16]. However, at the level of clinical populations, it has been found that omitting this item does not significantly alter the measurement capabilities of the PHQ, as the correlation between the PHQ-8 and PHQ-9 in clinical populations is very close to one [68].

Regarding the GAD-7, our findings are consistent with a meta-analysis that evaluated all possible cut-off points and reported that ≥8 is the most appropriate for anxiety disorder [18]. It also notes that scores between 7 and 10 points have similar sensitivity and specificity values [18]. Other recent primary studies conducted in hospitalised populations or people with chronic diseases in hospital settings also found optimal cut-offs between 7 and 10 points [69-71].

Our results on PHQ-2 was in line with meta-analyses supporting the use of the cut-off of 2 for PHQ-2 [35, 72]. Also, the values most frequents for GAD-2 are cut-off ≥2 and ≥3 [18, 37, 73]. The meta-analyses mentioned included studies in general populations (i.e., people attending primary care) and people hospitalised for non-communicable or infectious diseases. However, no meta-analyses were found that evaluated cut-off for hospitalised people only. At the level of primary studies, the evidence suggests that cut-off vary between 2 and 3 points for the PHQ-2 and GAD-2 [74, 75].

Regarding internal validity, a systematic review examined the factor structure of the PHQ-9, noting that the one-dimensional model has been repeatedly confirmed across studies [76]. Although several studies evaluated alternative multidimensional models (e.g., two-dimensional, three-dimensional, or bifactorial models), their dimensions are often highly correlated with each other, so there may be overlapping [76]. We did not find systematic reviews on the internal structure of the GAD-7 and the PHQ-8. However, several studies support the one-dimensional model in hospitalised patients for both the PHQ-8 [77] and GAD-7 [21, 27]. In Peru, the GAD-7 and PHQ-9 have shown evidence of a one-dimensional factor structure in different populations, such as the general population [20], pregnant women [21], and university students [52, 78]. However, no studies have been found evaluating the factor structure of the PHQ-8 in the Peruvian population.

Our study focuses on a hospital-based clinical population with one or more physical morbidities, it is important to consider that our finding of a different cut-off point, equal to or greater than 10 points for PHQ, may be influenced by the characteristics of this specific population. It is relevant to note that other studies conducted in hospital settings have found cut-off points lower than the recommendation of equal to or greater than 10 [79, 80]. It is important to bear in mind that the cut-off point may vary depending on the reference group and the context in which it is applied.

Our study used the Youden index to determine the optimal cut-off, but it is important to consider that the cut-off may vary depending on the sample size. A recent simulation study found that for large samples of more than 1000 participants, the optimal sensitivity and specificity values can vary by up to approximately 2 points from the optimal cut-off in cross-sectional studies [81]. Therefore, while a sample size calculation was performed to ensure adequate power, we cannot rule out the use of a cut-off of 10 or more for the Peruvian population. However, within the study, we present the sensitivity and specificity found for such a cut-off.

Public health implications

The evaluated instruments are widely used in clinical practice and research to measure symptoms of depression and anxiety, but from today, users will have optimal cut-off points for interpretations. This can help healthcare professionals identify people at risk of depression and anxiety more accurately while informing decisions about their formal diagnosis and consequent treatment. This is especially valuable in hospital environments, where time is crucial.

Our findings are of particular interest to the Peruvian health system, which has clinical practice guidelines for depression that recommend the PHQ-9 as a screening tool in primary care and hospital context [82]. Although our results correspond only to a hospital population, our study is the closest approximation to an evaluation of sensitivity and specificity in the Peruvian context, in the absence of similar studies in primary care. On the other hand, there is a lack of national clinical practice guidelines for screening and managing anxiety in Peru. Therefore, our study could contribute to future clinical practice guidelines for generalised anxiety disorder.

Although our study found alternative cut-off points to the standard (cut-off ≥10) for the PHQ-9 and PHQ-8 questionnaires, it is important to note that in certain contexts, higher specificity values (cut-off ≥10) may be necessary. These higher values enable a more accurate identification of individuals without depression or anxiety, thereby reducing the likelihood of false-positive results. This reduction in false positives is particularly crucial for alleviating the burden on the healthcare system. A screening tool with high specificity avoids unnecessary diagnoses and optimizes the use of healthcare resources. Therefore, utilizing a cut-off point of 10 or higher for the PHQ-9, PHQ-8, and GAD-7 can facilitate the early and accurate identification of true cases of depression and anxiety, ensuring that resources are appropriately focused on those who need care and treatment.

Strengths and limitations

Our results of the study have several strengths. First, to our knowledge, this is the first study in a Peruvian context that evaluates the factorial structure of all PHQ and GAD versions in a hospitalised population. Second, the scales were administered by a team of healthcare professionals with more than five years of experience in the clinical assessment of these patients. Third, the sample size was large enough to support all analyses and conclusions. Further, our sample size was larger than other

recently published studies' [60]. Fourth, our study is the first Peruvian study to evaluate the sensitivity and specificity of the PHQ.

Our study has limitations. First, we conducted the study only in a hospital context in a Peruvian city, which limits its applicability to other settings in Peru or other countries. However, it could be used in other Peruvian hospital contexts with similar characteristics, which is relevant because hospital care in Peru (levels II and III of complexity) represents 58.65% of total care [83]. Secondly, the generalisability of our results may be limited because the sampling is not probabilistic, as it does not include other hospitals. However, the hospital where we conducted the study serves 1.1% of all nationally insured EsSalud patients (http://www.essalud.gob.pe/estadistica-institucional/). It is also a national referral hospital, which means that people from all over the country are referred to this hospital for treatment. Therefore, the representativeness of the results is ensured. Thirdly, we used an individual psychiatric interview according to the ICD-10 criteria as a gold standard. We were not able to use the Composite International Diagnostic Interview (CIDI) or the Standardised Clinical Assessment (SCID), more typical gold standards, because of the time constraints involved in conducting such interviews. In Peru, health systems are overburdened, and it is not feasible to have lengthy sessions with highly specialised professionals to conduct such structured interviews. However, based on our experience, we believe that a psychiatric interview is a sufficient benchmark in this context. Fourthly, our study identified a limited number of individuals (n=28) with a diagnosed anxiety condition. Consequently, minor variations in the study cohort could potentially impact the sensitivity or specificity [81]. Nonetheless, we have ensured sufficient statistical power for our analysis based on our sample size calculation. Moreover, all cohort scores on the GAD scale are provided, which can be valuable for future research involving larger numbers of individuals diagnosed with anxiety (refer to supplementary material 7). Fifth, our study allows us to obtain sensitivity and specificity values for users in inpatient mental health settings; however, our findings are not generalisable to physical outpatients.

Conclusions

The PHQ-9's ≥7 cut-off point showed the highest simultaneous sensitivity and specificity when contrasted against a psychiatric diagnosis of depression. For a similar contrast against the gold standard, the other optimal cut-off points were: ≥7 for the PHQ-8, and ≥2 for the PHQ-2. Also, we present the sensitivity and specificity values of each cut-off point in GAD-7 and GAD-2. We confirmed the adequacy of a one-dimensional model for the PHQ-9, PHQ-8, and GAD-7, while all PHQ and GAD scales showed good reliability.

Tables, Figures, and Supplementary material

Table 1. Sociodemographic characteristics (n=1347).
Table 2. Raw scores and internal consistency (n=1347).

Supplementary material 1. Sex and age of the participants in our study and of the total number of people hospitalised in 2022.
Supplementary material 2. Script of R used in our study.
Supplementary material 3. Do-file of STATA used in our study.
Supplementary material 4. Flowchart.
Supplementary material 5. Sensitivity, specificity, +LR, −LR, PPV, NPV, and Youden's index for different PHQ cut-off points compared to the gold standard.
Supplementary material 6. Sensitivity, specificity, PPV, NPV, and Youden's index for different PHQ and GAD cut-off points compared to the gold standard.
Supplementary material 7. Sensitivity, specificity, +LR, −LR, PPV, NPV, and Youden's index for different GAD cut-off points compared to the gold standard.

Contributions

DVZ contributed to the conceptualizing the study, designing the methodology, developing the software tools, validating the results, conducting formal analyses, curating, and managing the data, and contributed to the initial drafting and visualization of the manuscript. JB contributed to the formal analysis, performed investigations, and aided in visualizing the findings. SOA participated in the investigation phase and contributed to the initial drafting of the manuscript. NMP engaged in formal analysis, conducted investigations, and contributed to the initial drafting of the manuscript. JCB contributed to the methodology, conducted investigations, provided critical input for the manuscript in the review and editing stages, and played a supervisory role. JHV contributed to the conceptualizing the study, designing the methodology, developing software tools, validating the results, conducting investigations, managing resources, curating data, project administration responsibilities, and participated in reviewing and editing the manuscript.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Competing interests

The authors declare that they have no conflicts of interest.

Patient consent for publication

Not applicable.

Ethics approval

The Hospital Nacional Guillermo Almenara Irigoyen's Institutional Review Board (Nota N°52 CIEI-OIyD-GRPA-Essalud-2023) approved the original study and the current analysis. Throughout the study, the researchers did not have access to any identifying information about the participants. In addition, participants gave informed consent. All participants were users of the hospital's Liaison Psychiatry Unit, so all received psychological or psychiatric care as needed.

Data availability statement

Data are available upon reasonable request.

ORCID iDs

David Villarreal-Zegarra https://orcid.org/0000-0002-2222-4764

Juan Barrera https://orcid.org/0000-0002-6641-6266
Sharlyn Otazú-Alfaro https://orcid.org/0000-0002-6462-8716
Nikol Mayo-Puchoc https://orcid.org/0000-0002-6182-7605
Juan Carlos Bazo-Alvarez https://orcid.org/0000-0002-6169-8049
Jeff Huarcaya-Victoria https://orcid.org/0000-0003-4525-9545

Reference

1. **Global Health Data Exchange (GHDx), Institute for Health Metrics and Evaluation (IHME)** [https://microdata.worldbank.org/index.php/catalog/ghdx/?page=1&ps=15&repo=ghdx]
2. GBD 2019 Mental Disorders Collaborators: **Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019**. *The lancet Psychiatry* 2022, **9**(2):137-150.
3. GBD 2019 Diseases and Injuries Collaborators: **Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019**. *Lancet (London, England)* 2020, **396**(10258):1204-1222.
4. **COVID-19 pandemic triggers 25% increase in prevalence of anxiety and depression worldwide** [https://www.who.int/news/item/02-03-2022-covid-19-pandemic-triggers-25-increase-in-prevalence-of-anxiety-and-depression-worldwide]
5. Villarreal-Zegarra D, Reátegui-Rivera CM, Otazú-Alfaro S, Yantas-Alcantara G, Soto-Becerra P, Melendez-Torres GJ: **Estimated impact of the COVID-19 pandemic on the prevalence and treatment of depressive symptoms in Peru: an interrupted time series analysis in 2014–2021**. *Social psychiatry and psychiatric epidemiology* 2023.
6. COVID-19 Mental Disorders Collaborators: **Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic**. *Lancet (London, England)* 2021, **398**(10312):1700-1712.
7. Villarreal-Zegarra D, Segovia-Bacilio P, Paredes-Angeles R, Vilela-Estrada AL, Cavero V, Diez-Canseco F: **Provision of community mental health care before and during the COVID-19 pandemic: A time series analysis in Peru**. *International Journal of Social Psychiatry* 2023.
8. Sagan A, McDaid D, Rajan S, Farrington J, McKee M: **European Observatory Policy Briefs**. In: *Screening: When is it appropriate and how can we get it right?* , edn. Copenhagen (Denmark): European Observatory on Health Systems and Policies © World Health Organization 2020 (acting as the host organization for, and secretariat of, the European Observatory on Health Systems and Policies). 2020.
9. Iragorri N, Spackman E: **Assessing the value of screening tools: reviewing the challenges and opportunities of cost-effectiveness analysis**. *Public health reviews* 2018, **39**:17.
10. Mulvaney-Day N, Marshall T, Downey Piscopo K, Korsen N, Lynch S, Karnell LH, Moran GE, Daniels AS, Ghose SS: **Screening for Behavioral Health Conditions in Primary Care Settings: A Systematic Review of the Literature**. *J Gen Intern Med* 2018, **33**(3):335-346.
11. Jiao B, Rosen Z, Bellanger M, Belkin G, Muennig P: **The cost-effectiveness of PHQ screening and collaborative care for depression in New York City**. *PloS one* 2017, **12**(8):e0184210.
12. Ramírez-Bontá F, Vásquez-Vílchez R, Cabrera-Alva M, Otazú-Alfaro S, Almeida-Huanca G, Ambrosio-Melgarejo J, Figueroa-Quiñones J, Romero-Cabrera A, Huaman-Santa Cruz A, Chávez-Hinostroza E *et al*: **Mental health data available in representative surveys conducted in Latin America and the Caribbean countries: A scoping review**. *In press* 2023.
13. Haberer JE, Trabin T, Klinkman M: **Furthering the reliable and valid measurement of mental health screening, diagnoses, treatment and outcomes through health information technology**. *Gen Hosp Psychiatry* 2013, **35**(4):349-353.

14. National Research Council; Division of Behavioral and Social Sciences and Education; Institute of Medicine; Board on Children Y, and Families; Committee on the Prevention of Mental Disorders and Substance Abuse Among Children, Youth and Young Adults: Research Advances and Promising Interventions; Mary Ellen O'Connell, Thomas Boat, and Kenneth E. Warner, Editors: **Preventing Mental, Emotional, and Behavioral Disorders Among Young People: Progress and Possibilities**. Washington: National Academies Press; 2009.
15. Spitzer RL, Kroenke K, Williams JB: **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* 1999, **282**(18):1737-1744.
16. Kroenke K, Strine TW, Spitzer RL, Williams JB, Berry JT, Mokdad AH: **The PHQ-8 as a measure of current depression in the general population**. *J Affect Disord* 2009, **114**(1-3):163-173.
17. Kroenke K, Spitzer RL, Williams JB: **The Patient Health Questionnaire-2: validity of a two-item depression screener**. *Med Care* 2003, **41**(11):1284-1292.
18. Plummer F, Manea L, Trepel D, McMillan D: **Screening for anxiety disorders with the GAD-7 and GAD-2: a systematic review and diagnostic metaanalysis**. *Gen Hosp Psychiatry* 2016, **39**:24-31.
19. 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**(6):e0156939.
20. Villarreal-Zegarra D, Copez-Lonzoy A, Bernabe-Ortiz A, Melendez-Torres GJ, Bazo-Alvarez JC: **Valid group comparisons can be made with the Patient Health Questionnaire (PHQ-9): A measurement invariance study across groups by demographic characteristics**. *PloS one* 2019, **14**(9):1-15.
21. Zhong QY, Gelaye B, Zaslavsky AM, Fann JR, Rondon MB, Sanchez SE, Williams MA: **Diagnostic Validity of the Generalized Anxiety Disorder - 7 (GAD-7) among Pregnant Women**. *PloS one* 2015, **10**(4):e0125096.
22. Stochl J, Fried EI, Fritz J, Croudace TJ, Russo DA, Knight C, Jones PB, Perez J: **On Dimensionality, Measurement Invariance, and Suitability of Sum Scores for the PHQ-9 and the GAD-7**. *Assessment* 2022, **29**(3):355-366.
23. Shevlin M, Butter S, McBride O, Murphy J, Gibson-Miller J, Hartman TK, Levita L, Mason L, Martinez AP, McKay R *et al*: **Measurement invariance of the Patient Health Questionnaire (PHQ-9) and Generalized Anxiety Disorder scale (GAD-7) across four European countries during the COVID-19 pandemic**. *BMC psychiatry* 2022, **22**(1):154.
24. Kroenke K, Wu J, Yu Z, Bair MJ, Kean J, Stump T, Monahan PO: **Patient Health Questionnaire Anxiety and Depression Scale: Initial Validation in Three Clinical Trials**. *Psychosom Med* 2016, **78**(6):716-727.
25. Urtasun M, Daray FM, Teti GL, Coppolillo F, Herlax G, Saba G, Rubinstein A, Araya R, Irazola V: **Validation and calibration of the patient health questionnaire (PHQ-9) in Argentina**. *BMC psychiatry* 2019, **19**(1):291.
26. García-Campayo J, Zamorano E, Ruiz MA, Pardo A, Pérez-Páramo M, López-Gómez V, Freire O, Rejas J: **Cultural adaptation into Spanish of the generalized anxiety disorder-7 (GAD-7) scale as a screening tool**. *Health and quality of life outcomes* 2010, **8**:8.
27. Sawaya H, Atoui M, Hamadeh A, Zeinoun P, Nahas Z: **Adaptation and initial validation of the Patient Health Questionnaire - 9 (PHQ-9) and the Generalized Anxiety Disorder - 7 Questionnaire (GAD-7) in an Arabic speaking Lebanese psychiatric outpatient sample**. *Psychiatry Res* 2016, **239**:245-252.
28. Lotrakul M, Sumrithe S, Saipanish R: **Reliability and validity of the Thai version of the PHQ-9**. *BMC psychiatry* 2008, **8**:46.
29. Woldetensay YK, Belachew T, Tesfaye M, Spielman K, Biesalski HK, Kantelhardt EJ, Scherbaum V: **Validation of the Patient Health Questionnaire (PHQ-9) as a screening**

tool for depression in pregnant women: Afaan Oromo version. *PloS one* 2018, **13**(2):e0191782.

30. Eack SM, Greeno CG, Lee BJ: **Limitations of the Patient Health Questionnaire in Identifying Anxiety and Depression: Many Cases Are Undetected**. *Research on social work practice* 2006, **16**(6):625-631.
31. Lambert SD, Clover K, Pallant JF, Britton B, King MT, Mitchell AJ, Carter G: **Making Sense of Variations in Prevalence Estimates of Depression in Cancer: A Co-Calibration of Commonly Used Depression Scales Using Rasch Analysis**. *Journal of the National Comprehensive Cancer Network : JNCCN* 2015, **13**(10):1203-1211.
32. Liu SI, Yeh ZT, Huang HC, Sun FJ, Tjung JJ, Hwang LC, Shih YH, Yeh AW: **Validation of Patient Health Questionnaire for depression screening among primary care patients in Taiwan**. *Comprehensive psychiatry* 2011, **52**(1):96-101.
33. Levis B, Benedetti A, Thombs BD: **Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression: individual participant data meta-analysis**. *BMJ : British Medical Journal* 2019, **365**:l1476.
34. Neupane D, Levis B, Bhandari PM, Thombs BD, Benedetti A: **Selective cutoff reporting in studies of the accuracy of the Patient Health Questionnaire-9 and Edinburgh Postnatal Depression Scale: Comparison of results based on published cutoffs versus all cutoffs using individual participant data meta-analysis**. *International journal of methods in psychiatric research* 2021, **30**(3):e1873.
35. Levis B, Sun Y, He C, Wu Y, Krishnan A, Bhandari PM, Neupane D, Imran M, Brehaut E, Negeri Z *et al*: **Accuracy of the PHQ-2 Alone and in Combination With the PHQ-9 for Screening to Detect Major Depression: Systematic Review and Meta-analysis**. *Jama* 2020, **323**(22):2290-2300.
36. Wu Y, Levis B, Riehm KE, Saadat N, Levis AW, Azar M, Rice DB, Boruff J, Cuijpers P, Gilbody S *et al*: **Equivalency of the diagnostic accuracy of the PHQ-8 and PHQ-9: a systematic review and individual participant data meta-analysis**. *Psychological medicine* 2020, **50**(8):1368-1380.
37. 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**(4):345-359.
38. Manea L, Boehnke JR, Gilbody S, Moriarty AS, McMillan D: **Are there researcher allegiance effects in diagnostic validation studies of the PHQ-9? A systematic review and meta-analysis**. *BMJ Open* 2017, **7**(9):e015247.
39. 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**. *BJPsych open* 2016, **2**(2):127-138.
40. Mughal AY, Devadas J, Ardman E, Levis B, Go VF, Gaynes BN: **A systematic review of validated screening tools for anxiety disorders and PTSD in low to middle income countries**. *BMC psychiatry* 2020, **20**(1):338.
41. Cohen JF, Korevaar DA, Altman DG, Bruns DE, Gatsonis CA, Hooft L, Irwig L, Levine D, Reitsma JB, de Vet HC *et al*: **STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration**. *BMJ Open* 2016, **6**(11):e012799.
42. Villarreal-Zegarra D, Cabrera-Alva M, Carrillo-Larco RM, Bernabe-Ortiz A: **Trends in the prevalence and treatment of depressive symptoms in Peru: a population-based study**. *BMJ Open* 2020, **10**(7):e036777.
43. Hernández-Vásquez A, Vargas-Fernández R, Bendezu-Quispe G, Grendas LN: **Depression in the Peruvian population and its associated factors: analysis of a national health survey**. *J Affect Disord* 2020, **273**:291-297.
44. Steel Z, Marnane C, Iranpour C, Chey T, Jackson JW, Patel V, Silove D: **The global prevalence of common mental disorders: a systematic review and meta-analysis 1980-2013**. *Int J Epidemiol* 2014, **43**(2):476-493.

45. Buderer NM: **Statistical methodology: I. Incorporating the prevalence of disease into the sample size calculation for sensitivity and specificity**. *Academic emergency medicine : official journal of the Society for Academic Emergency Medicine* 1996, **3**(9):895-900.
46. Huarcaya-Victoria J, Segura V, Cárdenas D, Sardón K, Caqui M, Podestà Á: **Analysis of the care provided over a six-month period by the liaison psychiatry unit at a general hospital in Lima, Peru**. *Rev Colomb Psiquiatr (Engl Ed)* 2022, **51**(2):105-112.
47. 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**. *General hospital psychiatry* 2015, **37**(1):67-75.
48. Zuithoff NP, Vergouwe Y, King M, Nazareth I, van Wezep MJ, Moons KG, Geerlings MI: **The Patient Health Questionnaire-9 for detection of major depressive disorder in primary care: consequences of current thresholds in a crosssectional study**. *BMC Fam Pract* 2010, **11**:98.
49. Shin C, Lee SH, Han KM, Yoon HK, Han C: **Comparison of the Usefulness of the PHQ-8 and PHQ-9 for Screening for Major Depressive Disorder: Analysis of Psychiatric Outpatient Data**. *Psychiatry investigation* 2019, **16**(4):300-305.
50. Caycho-Rodríguez T, Barboza-Palomino M, Ventura-León J, Carbajal-León C, Noé-Grijalva M, Gallegos M, Reyes-Bossio M, Vivanco-Vidal A: **[Spanish translation and validation of a brief measure of anxiety by the COVID-19 in students of health sciences]**. *Ansiedad y Estrés* 2020, **26**(2):174-180.
51. Spitzer RL, Kroenke K, Williams JW, Löwe B: **A brief measure for assessing generalized anxiety disorder: The GAD-7**. *Archives of Internal Medicine* 2006, **166**(10):1092-1097.
52. Franco-Jimenez RA, Nuñez-Magallanes A: **[Psychometric Properties of the GAD-7, GAD-2, and GAD-Mini in Peruvian College Students]**. *Propósitos y Representaciones* 2022, **10**(1):e1437.
53. Kroenke K, Spitzer RL, Williams JB, Monahan PO, Löwe B: **Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection**. *Annals of internal medicine* 2007, **146**(5):317-325.
54. Trevethan R: **Sensitivity, Specificity, and Predictive Values: Foundations, Pliabilities, and Pitfalls in Research and Practice**. *Frontiers in public health* 2017, **5**:307.
55. Ranganathan P, Aggarwal R: **Understanding the properties of diagnostic tests - Part 2: Likelihood ratios**. *Perspectives in clinical research* 2018, **9**(2):99-102.
56. Suh Y: **The Performance of Maximum Likelihood and Weighted Least Square Mean and Variance Adjusted Estimators in Testing Differential Item Functioning With Nonnormal Trait Distributions**. *Structural Equation Modeling: A Multidisciplinary Journal* 2015, **22**(4):568-580.
57. Holgado–Tello FP, Chacón–Moscoso S, Barbero–García I, Vila–Abad E: **Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables**. *Qual Quant* 2010, **44**(1):153-166.
58. Hair JF, Anderson RE, Tatham RL, Black WC: **Análisis multivariante**, vol. 491: Prentice Hall Madrid; 1999.
59. McDonald RP: **Test theory: A unified treatment**. New York: Taylor & Francis Group; 1999.
60. Costantini L, Pasquarella C, Odone A, Colucci ME, Costanza A, Serafini G, Aguglia A, Belvederi Murri M, Brakoulias V, Amore M *et al*: **Screening for depression in primary care with Patient Health Questionnaire-9 (PHQ-9): A systematic review**. *J Affect Disord* 2021, **279**:473-483.
61. Hartung TJ, Friedrich M, Johansen C, Wittchen HU, Faller H, Koch U, Brähler E, Härter M, Keller M, Schulz H *et al*: **The Hospital Anxiety and Depression Scale (HADS) and the 9-item Patient Health Questionnaire (PHQ-9) as screening instruments for depression in patients with cancer**. *Cancer* 2017, **123**(21):4236-4243.

62. Sun Y, Kong Z, Song Y, Liu J, Wang X: **The validity and reliability of the PHQ-9 on screening of depression in neurology: a cross sectional study**. *BMC psychiatry* 2022, **22**(1):98.
63. Gholizadeh L, Shahmansouri N, Heydari M, Davidson PM: **Assessment and detection of depression in patients with coronary artery disease: validation of the Persian version of the PHQ-9**. *Contemporary nurse* 2019, **55**(2-3):185-194.
64. Husain N, Waheed W, Tomenson B, Creed F: **The validation of personal health questionnaire amongst people of Pakistani family origin living in the United Kingdom**. *J Affect Disord* 2007, **97**(1-3):261-264.
65. Ganguly S, Samanta M, Roy P, Chatterjee S, Kaplan DW, Basu B: **Patient health questionnaire-9 as an effective tool for screening of depression among Indian adolescents**. *The Journal of adolescent health : official publication of the Society for Adolescent Medicine* 2013, **52**(5):546-551.
66. Hanlon C, Medhin G, Selamu M, Breuer E, Worku B, Hailemariam M, Lund C, Prince M, Fekadu A: **Validity of brief screening questionnaires to detect depression in primary care in Ethiopia**. *J Affect Disord* 2015, **186**:32-39.
67. Harry ML, Coley RY, Waring SC, Simon GE: **Evaluating the Cross-Cultural Measurement Invariance of the PHQ-9 between American Indian/Alaska Native Adults and Diverse Racial and Ethnic Groups**. *Journal of affective disorders reports* 2021, **4**.
68. Razykov I, Ziegelstein RC, Whooley MA, Thombs BD: **The PHQ-9 versus the PHQ-8--is item 9 useful for assessing suicide risk in coronary artery disease patients? Data from the Heart and Soul Study**. *J Psychosom Res* 2012, **73**(3):163-168.
69. Konkan R, Senormancı O, Guclu O, Aydin E, Sungur M: **Validity and reliability study for the Turkish adaptation of the Generalized Anxiety Disorder-7 (GAD-7) scale**. *Archives of Neuropsychiatry* 2013, **50**:53-58.
70. Gong Y, Zhou H, Zhang Y, Zhu X, Wang X, Shen B, Xian J, Ding Y: **Validation of the 7-item Generalized Anxiety Disorder scale (GAD-7) as a screening tool for anxiety among pregnant Chinese women**. *J Affect Disord* 2021, **282**:98-103.
71. Snijkers JTW, van den Oever W, Weerts Z, Vork L, Mujagic Z, Leue C, Hesselink MAM, Kruimel JW, Muris JWM, Bogie RMM *et al*: **Examining the optimal cutoff values of HADS, PHQ-9 and GAD-7 as screening instruments for depression and anxiety in irritable bowel syndrome**. *Neurogastroenterology and motility : the official journal of the European Gastrointestinal Motility Society* 2021, **33**(12):e14161.
72. Manea L, Gilbody S, Hewitt C, North A, Plummer F, Richardson R, Thombs BD, Williams B, McMillan D: **Identifying depression with the PHQ-2: A diagnostic meta-analysis**. *J Affect Disord* 2016, **203**:382-395.
73. Luo Z, Li Y, Hou Y, Zhang H, Liu X, Qian X, Jiang J, Wang Y, Liu X, Dong X *et al*: **Adaptation of the two-item generalized anxiety disorder scale (GAD-2) to Chinese rural population: A validation study and meta-analysis**. *Gen Hosp Psychiatry* 2019, **60**:50-56.
74. Giuliani M, Gorini A, Barbieri S, Veglia F, Tremoli E: **Examination of the best cut-off points of PHQ-2 and GAD-2 for detecting depression and anxiety in Italian cardiovascular inpatients**. *Psychology & health* 2021, **36**(9):1088-1101.
75. Bentley KH, Sakurai H, Lowman KL, Rines-Toth L, McKowen J, Pedrelli P, Evins AE, Yule AM: **Validation of brief screening measures for depression and anxiety in young people with substance use disorders**. *J Affect Disord* 2021, **282**:1021-1029.
76. Lamela D, Soreira C, Matos P, Morais A: **Systematic review of the factor structure and measurement invariance of the patient health questionnaire-9 (PHQ-9) and validation of the Portuguese version in community settings**. *J Affect Disord* 2020, **276**:220-233.
77. Schantz K, Reighard C, Aikens JE, Aruquipa A, Pinto B, Valverde H, Piette JD: **Screening for depression in Andean Latin America: Factor structure and reliability of the CES-D**

short form and the PHQ-8 among Bolivian public hospital patients. *International journal of psychiatry in medicine* 2017, **52**(4-6):315-327.

78. Huarcaya-Victoria J, De-Lama-Morán R, Quiros M, Bazán J, López K, Lora D: **Propiedades psicométricas del Patient Health Questionnaire (PHQ-9) en estudiantes de medicina en Lima, Perú**. *Revista de Neuro-Psiquiatria* 2020, **83**(2):72-78.
79. Inagaki M, Ohtsuki T, Yonemoto N, Kawashima Y, Saitoh A, Oikawa Y, Kurosawa M, Muramatsu K, Furukawa TA, Yamada M: **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**(6):592-597.
80. Le Hoang Ngoc T, Le MT, Nguyen HT, Vo HV, Le NQ, Tang LNP, Tran TT, Le TV: **Patient Health Questionnaire (PHQ-9): A depression screening tool for people with epilepsy in Vietnam**. *Epilepsy & behavior : E&B* 2021, **125**:108446.
81. Bhandari PM, Levis B, Neupane D, Patten SB, Shrier I, Thombs BD, Benedetti A: **Data-driven methods distort optimal cutoffs and accuracy estimates of depression screening tools: a simulation study using individual participant data**. *J Clin Epidemiol* 2021, **137**:137-147.
82. Beatrice M-F, Carla M-C, Matilde L-M, Víctor G-M, Amelia M-M, Jessica Hanae Z-T, Sergio G-L, Raisa NM-R, Wendy N-G, Fabian F-S *et al*: **Clinical practice guideline for the screening and management of the mild depressive episode at the first level of care for the Peruvian Social Security (EsSalud)**. *ACTA MEDICA PERUANA* 2020, **37**(4).
83. EsSalud: **Análisis ejecutivo nacional de a nivel las prestaciones de salud 2016**. In. Lima: Gerencia central de planeamiento y presupuesto, EsSalud; 2017.

Table 1. Sociodemographic characteristics (n=1347).

	n	%
Sex		
Men	547	40.6%
Women	800	59.4%
Age (categories)		
18-29	107	7.9%
30-39	164	12.2%
40-49	214	15.9%
50-59	284	21.1%
60-69	294	21.8%
70-79	203	15.1%
80 to more	81	6.0%
Civil status		
Single	329	24.4%
Married or Cohabitant	768	57.0%
Separated	133	9.9%
Widowed	117	8.7%
Education level		
None	13	1.0%
Elementary school	135	10.0%
High school	478	35.5%
Technical	246	18.2%
University	475	35.3%
Currently works		
No	330	24.5%
Yes	778	57.8%
Retired	239	17.7%
Living alone		
Yes	99	7.3%
No	1248	92.7%
History of psychiatric diagnosis		
Yes	388	28.8%
No	959	71.2%
Diagnosis of depression		
No	1013	75.2%
Yes	334	24.8%
Diagnosis of anxiety		
No	1319	97.9%
Yes	28	2.1%
Physical illnesses		
A00-B99 Certain infectious and parasitic diseases	109	8.1%
C00-D48 Neoplasms, and diseases of the blood and haematopoietic organs and other disorders affecting the mechanism of immunity	348	25.8%
E00-E90 Endocrine, nutritional and metabolic diseases	130	9.7%
G00-G99 Diseases of the nervous system	96	7.1%
H00-H59 Diseases of the eye and adnexa	17	1.3%
H60-H95 Diseases of the ear and mastoid process	17	1.3%
I00-I99 Diseases of the circulatory system	111	8.2%
J00-J99 Diseases of the respiratory system	107	7.9%
K00-K93 Diseases of the gastro-intestinal tract	106	7.9%
L00-L99 Diseases of the skin and subcutaneous tissues	74	5.5%
M00-M99 Diseases of the musculo-skeletal system and connective tissue	97	7.2%
N00-N99 Diseases of the genito-urinary system	97	7.2%
O00-O99 Pregnancy, childbirth and puerperium	10	0.7%
P00-P96 Certain conditions originating in the perinatal period	0	0.0%
Q00-Q99 Congenital malformations, deformities and chromosome anomalies	6	0.4%
R00-R99 Symptoms, signs and abnormal clinical and laboratory	46	3.4%

findings, not elsewhere classified		
S00-T98 Trauma, poisoning and certain other consequences of external cause	50	3.7%
V01-Y98 External causes of morbidity and mortality	4	0.3%
Z00-Z99 Factors influencing health status and contact with health care services	90	6.7%
U00-U99 Codes for special situations	28	2.1%

Note: n=number. %=Percentage.

Table 2. Raw scores and internal consistency (n=1347).

	M	SD	Min	Max	α	ω
(1) PHQ-9 score	6.4	5.0	0	27	0.89	0.86
(2) PHQ-8 score	6.1	4.7	0	24	0.88	0.85
(3) PHQ-2 score	1.9	1.6	0	6	0.83	0.80
(4) GAD-7 score	5.1	3.9	0	21	0.85	0.81
(5) GAD-2 score	1.7	1.4	0	6	0.74	0.70

Note: α = Classical alpha. ω = Mcdonald's omega.

Supplementary material 1. Sex and age of the participants in our study and of the total number of people hospitalised in 2022.

	Our study		Total inpatients in 2022		
	n	%			p
Sex					
Men	547	40.60%	9677	43.2%	0.971
Women	800	59.40%	12732	56.8%	
Age (categories)					
18-29	107	7.90%	1962	8.8%	1.000
30-39	164	12.20%	3486	15.6%	
40-49	214	15.90%	3145	14.0%	
50-59	284	21.10%	3497	15.6%	
60-69	294	21.80%	4276	19.1%	
70-79	203	15.10%	3806	17.0%	
80 to more	81	6.00%	2227	9.9%	

Note: p = chi-square test.

Supplementary material 2. Script of R used in our study.

```
library(lavaan)
library(semPlot)
library(semTools)
library(psych)
library(haven)
Database <- read_dta("E:/Database_v1.dta")

#PHQ-9
model.PHQ9 <- "
F1 =~ PHQ_1 + PHQ_2 + PHQ_3 + PHQ_4 + PHQ_5 + PHQ_6 + PHQ_7 + PHQ_8 + PHQ_9"

fit.model.PHQ9 <- cfa(model.PHQ9, data=Database,estimator="WLSMV", missing = "listwise",
ordered = c("PHQ_1", "PHQ_2", "PHQ_3", "PHQ_4", "PHQ_5", "PHQ_6", "PHQ_7", "PHQ_8",
"PHQ_9"))

summary(fit.model.PHQ9, fit.measures=TRUE, standardized=TRUE)

reliability(fit.model.PHQ9)

#PHQ-8
model.PHQ8 <- "
F1 =~ PHQ_1 + PHQ_2 + PHQ_3 + PHQ_4 + PHQ_5 + PHQ_6 + PHQ_7 + PHQ_8"

fit.model.PHQ8 <- cfa(model.PHQ8, data=Database, estimator="WLSMV", missing =
"listwise", ordered = c("PHQ_1", "PHQ_2", "PHQ_3", "PHQ_4", "PHQ_5", "PHQ_6", "PHQ_7",
"PHQ_8", "PHQ_9"))

summary(fit.model.PHQ8, fit.measures=TRUE, standardized=TRUE)

reliability(fit.model.PHQ8)

#PHQ-2
model.PHQ2 <- "
F1 =~ PHQ_1 + PHQ_2"

fit.model.PHQ2 <- cfa(model.PHQ2, data=Database, estimator="WLSMV", missing =
"listwise", ordered = c("PHQ_1", "PHQ_2", "PHQ_3", "PHQ_4", "PHQ_5", "PHQ_6", "PHQ_7",
"PHQ_8", "PHQ_9"))

summary(fit.model.PHQ2, fit.measures=TRUE, standardized=TRUE)

reliability(fit.model.PHQ2)

#GAD-7
model.GAD7 <- "
F1 =~ GAD1 + GAD2 + GAD3 + GAD4 + GAD5 + GAD6 + GAD7"

fit.model.GAD7 <- cfa(model.GAD7, data=Database, estimator="WLSMV", missing =
"listwise", ordered = c("GAD1", "GAD2", "GAD3", "GAD4", "GAD5", "GAD6", "GAD7"))

summary(fit.model.GAD7, fit.measures=TRUE, standardized=TRUE)

reliability(fit.model.GAD7)

#GAD-2
model.GAD2 <- "
F1 =~ GAD1 + GAD2"

fit.model.GAD2 <- cfa(model.GAD2, data=Database, estimator="WLSMV", missing =
"listwise", ordered = c("GAD1", "GAD2", "GAD3", "GAD4", "GAD5", "GAD6", "GAD7"))

summary(fit.model.GAD2, fit.measures=TRUE, standardized=TRUE)

reliability(fit.model.GAD2)
```

Supplementary material 3. Do-file of STATA used in our study.

```
use "E:\Database_v1.dta", clear

*Table 1 - Socio-demographic analysis

global catvars_table1 sex agecat civilstatus educationcat work Livingalone
Historypsychiatricdx depression anxiety

tabout $catvars_table1 depression using Table1.xlsx, ///
  replace c(freq col) clab(No. %) f(0c 1p) style(xlsx) font(bold) ///
  ptotal(none) stats(chi2) stpos(col) ppos(only) plab(P value) ///
  title(Table 1. Sociodemographic characteristics) ///
  fn(Note: n=number, %=Percentage.) twidth(14) sheet(Table1)

* Table 2. Sensitivity, specificity, +LR, -LR, PPV, NPV, and Youden's index for different
PHQ and GAD cut-off points compared to the gold standard.
tab cutoffPHQ9_7
diagt depression cutoffPHQ9_7
tab cutoffPHQ9_10
tab PHQ9algorithm
diagt depression PHQ9algorithm
tab PHQ9ajus_algorithm
diagt depression PHQ9ajus_algorithm
diagt depression cutoffPHQ9_10
tab cutoffPHQ8_7
diagt depression cutoffPHQ8_7
tab cutoffPHQ2_2
diagt depression cutoffPHQ2_2
tab cutoffGAD7_8
diagt anxiety cutoffGAD7_8
tab cutoffGAD7_10
diagt anxiety cutoffGAD7_10
tab cutoffGAD2_2
diagt anxiety cutoffGAD2_2



* Supplementary material 3. Sensitivity, specificity, +LR, -LR, PPV, NPV, and Youden's
index for different PHQ cut-off points compared to the gold standard.
* PHQ-9 - Sensitivity, specificity, +LR, -LR, PPV, NPV, and Youden's index

tab cutoffPHQ9_1
diagt depression cutoffPHQ9_1
tab cutoffPHQ9_2
diagt depression cutoffPHQ9_2
tab cutoffPHQ9_3
diagt depression cutoffPHQ9_3
tab cutoffPHQ9_4
diagt depression cutoffPHQ9_4
tab cutoffPHQ9_5
diagt depression cutoffPHQ9_5
tab cutoffPHQ9_6
diagt depression cutoffPHQ9_6
tab cutoffPHQ9_7
diagt depression cutoffPHQ9_7
tab cutoffPHQ9_8
diagt depression cutoffPHQ9_8
tab cutoffPHQ9_9
diagt depression cutoffPHQ9_9
tab cutoffPHQ9_10
diagt depression cutoffPHQ9_10
tab cutoffPHQ9_11
diagt depression cutoffPHQ9_11
tab cutoffPHQ9_12
diagt depression cutoffPHQ9_12
```

```
tab cutoffPHQ9_13
diagt depression cutoffPHQ9_13
tab cutoffPHQ9_14
diagt depression cutoffPHQ9_14
tab cutoffPHQ9_15
diagt depression cutoffPHQ9_15
tab cutoffPHQ9_16
diagt depression cutoffPHQ9_16
tab cutoffPHQ9_17
diagt depression cutoffPHQ9_17
tab cutoffPHQ9_18
diagt depression cutoffPHQ9_18
tab cutoffPHQ9_19
diagt depression cutoffPHQ9_19
tab cutoffPHQ9_20
diagt depression cutoffPHQ9_20
tab cutoffPHQ9_21
diagt depression cutoffPHQ9_21
tab cutoffPHQ9_22
diagt depression cutoffPHQ9_22
tab cutoffPHQ9_23
diagt depression cutoffPHQ9_23
tab cutoffPHQ9_24
diagt depression cutoffPHQ9_24
tab cutoffPHQ9_25
diagt depression cutoffPHQ9_25
tab cutoffPHQ9_26
diagt depression cutoffPHQ9_26
tab cutoffPHQ9_27
diagt depression cutoffPHQ9_27
```

***PHQ-8 - Sensitivity, specificity, +LR, −LR, PPV, NPV, and Youden's index**

```
tab cutoffPHQ8_1
diagt depression cutoffPHQ8_1
tab cutoffPHQ8_2
diagt depression cutoffPHQ8_2
tab cutoffPHQ8_3
diagt depression cutoffPHQ8_3
tab cutoffPHQ8_4
diagt depression cutoffPHQ8_4
tab cutoffPHQ8_5
diagt depression cutoffPHQ8_5
tab cutoffPHQ8_6
diagt depression cutoffPHQ8_6
tab cutoffPHQ8_7
diagt depression cutoffPHQ8_7
tab cutoffPHQ8_8
diagt depression cutoffPHQ8_8
tab cutoffPHQ8_9
diagt depression cutoffPHQ8_9
tab cutoffPHQ8_10
diagt depression cutoffPHQ8_10
tab cutoffPHQ8_11
diagt depression cutoffPHQ8_11
tab cutoffPHQ8_12
diagt depression cutoffPHQ8_12
tab cutoffPHQ8_13
diagt depression cutoffPHQ8_13
tab cutoffPHQ8_14
diagt depression cutoffPHQ8_14
tab cutoffPHQ8_15
diagt depression cutoffPHQ8_15
tab cutoffPHQ8_16
diagt depression cutoffPHQ8_16
tab cutoffPHQ8_17
diagt depression cutoffPHQ8_17
```

```
tab cutoffPHQ8_18
diagt depression cutoffPHQ8_18
tab cutoffPHQ8_19
diagt depression cutoffPHQ8_19
tab cutoffPHQ8_20
diagt depression cutoffPHQ8_20
tab cutoffPHQ8_21
diagt depression cutoffPHQ8_21
tab cutoffPHQ8_22
diagt depression cutoffPHQ8_22
tab cutoffPHQ8_23
diagt depression cutoffPHQ8_23
tab cutoffPHQ8_24
diagt depression cutoffPHQ8_24
```

*** PHQ-9 algorithm - Sensitivity, specificity, +LR, −LR, PPV, NPV, and Youden's index**

```
tab PHQ9algorithm
diagt depression PHQ9algorithm
```

*** PHQ-9 adjusted algorithm - Sensitivity, specificity, +LR, −LR, PPV, NPV, and Youden's index**

```
tab PHQ9ajus_algorithm
diagt depression PHQ9ajus_algorithm
```

*** PHQ-2 - Sensitivity, specificity, +LR, −LR, PPV, NPV, and Youden's index**

```
tab cutoffPHQ2_1
diagt depression cutoffPHQ2_1
tab cutoffPHQ2_2
diagt depression cutoffPHQ2_2
tab cutoffPHQ2_3
diagt depression cutoffPHQ2_3
tab cutoffPHQ2_4
diagt depression cutoffPHQ2_4
tab cutoffPHQ2_5
diagt depression cutoffPHQ2_5
tab cutoffPHQ2_6
diagt depression cutoffPHQ2_6
```

***Supplementary material 4. Sensitivity, specificity, +LR, −LR, PPV, NPV, and Youden's index for different GAD cut-off points compared to the gold standard.**
***GAD-7 - Sensitivity, specificity, +LR, −LR, PPV, NPV, and Youden's index**

```
tab cutoffGAD7_1
diagt anxiety cutoffGAD7_1
tab cutoffGAD7_2
diagt anxiety cutoffGAD7_2
tab cutoffGAD7_3
diagt anxiety cutoffGAD7_3
tab cutoffGAD7_4
diagt anxiety cutoffGAD7_4
tab cutoffGAD7_5
diagt anxiety cutoffGAD7_5
tab cutoffGAD7_6
diagt anxiety cutoffGAD7_6
tab cutoffGAD7_7
diagt anxiety cutoffGAD7_7
tab cutoffGAD7_8
diagt anxiety cutoffGAD7_8
tab cutoffGAD7_9
diagt anxiety cutoffGAD7_9
tab cutoffGAD7_10
diagt anxiety cutoffGAD7_10
tab cutoffGAD7_11
diagt anxiety cutoffGAD7_11
```

```
tab cutoffGAD7_12
diagt anxiety cutoffGAD7_12
tab cutoffGAD7_13
diagt anxiety cutoffGAD7_13
tab cutoffGAD7_14
diagt anxiety cutoffGAD7_14
tab cutoffGAD7_15
diagt anxiety cutoffGAD7_15
tab cutoffGAD7_16
diagt anxiety cutoffGAD7_16
tab cutoffGAD7_17
diagt anxiety cutoffGAD7_17
tab cutoffGAD7_18
diagt anxiety cutoffGAD7_18
tab cutoffGAD7_19
diagt anxiety cutoffGAD7_19
tab cutoffGAD7_20
diagt anxiety cutoffGAD7_20
tab cutoffGAD7_21
diagt anxiety cutoffGAD7_21
```

*** GAD-2 - Sensitivity, specificity, +LR, -LR, PPV, NPV, and Youden's index**

```
tab cutoffGAD2_0
diagt anxiety cutoffGAD2_0
tab cutoffGAD2_1
diagt anxiety cutoffGAD2_1
tab cutoffGAD2_2
diagt anxiety cutoffGAD2_2
tab cutoffGAD2_3
diagt anxiety cutoffGAD2_3
tab cutoffGAD2_4
diagt anxiety cutoffGAD2_4
tab cutoffGAD2_5
diagt anxiety cutoffGAD2_5
tab cutoffGAD2_6
diagt anxiety cutoffGAD2_6
```

***Figures 2 - ROC curve**

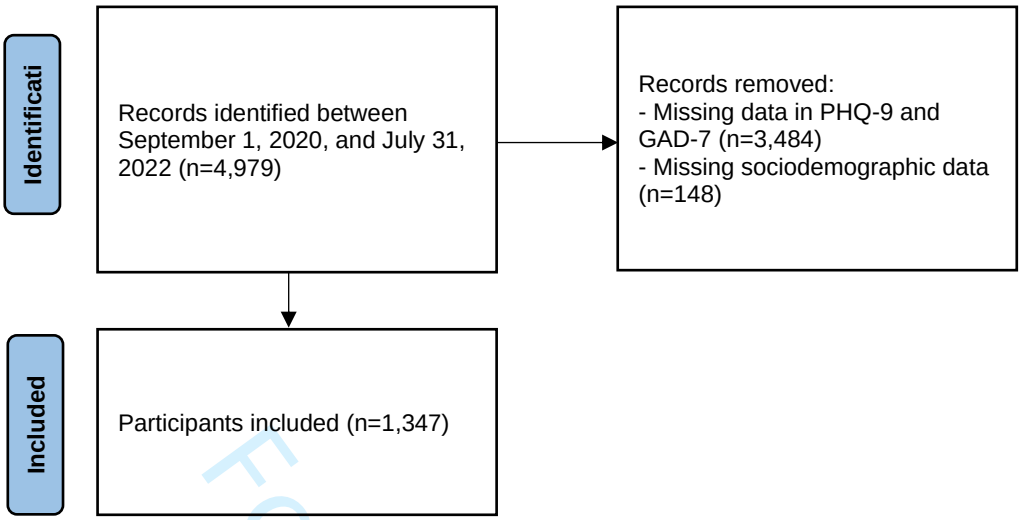
```
roccomp depression PHQ9TOTAL PHQ8TOTAL PHQ9algorithm PHQ9ajus_algorithm PHQ2TOTAL, graph
summary plot1opts(mcolor(red) lcolor(red)) plot2opts(mcolor(blue) lcolor(blue))
plot3opts(mcolor(orange) lcolor(orange)) plot4opts(mcolor(green) lcolor(green))
plot5opts(mcolor(purble) lcolor(purple)) legend(order(1 "PHQ-9 (AUC=0.805 [0.779-
0.831])" 2 "PHQ-8 (AUC=0.802 [0.776-0.828])" 3 "PHQ-9 algorithm (AUC=0.641 [0.614-
0.667])" 4 "PHQ-9 adjusted algorithm (AUC=0.723 [0.696-0.749])" 5 "PHQ-2 (AUC=0.771
[0.743-0.799])" 6 "Reference") size(2.5) position(7) cols(2) rows(3))

graph export "E:\Figure2.tif", as(tif) replace
```

***Figures 3 - ROC curve**

```
roccomp anxiety GAD7TOTAL GAD2TOTAL, graph summary plot1opts(mcolor(red) lcolor(red))
plot2opts(mcolor(blue) lcolor(blue)) legend(order(1 "GAD-7 (AUC=0.718 [0.622-0.814])" 2
"GAD-2 (AUC=0.685 [0.587-0.783])" 3 "Reference") size(2.5) position(7) cols(2) rows(3))

graph export "E:\Figure3.tif", as(tif) replace
```



Supplementary material 4. Flowchart.

Supplementary material 5. Sensitivity, specificity, +LR, −LR, PPV, NPV, and Youden's index for different PHQ cut-off points compared to the gold standard.

Method	Cut-off Points	n (%)	Sensitivity (95%CI)	Specificity (95%CI)	+LR (95%CI)	-LR (95%CI)	PPV (95%CI)	NPV (95%CI)	Youden Index
PHQ-9	≥1	1246 (92.5)	99.7 (98.3 - 100.0)	9.9 (8.1 - 11.9)	1.11 (1.08 - 1.13)	0.03 (0.00 - 0.22)	26.7 (24.3 - 29.3)	99.0 (94.6 - 100.0)	9.6
	≥2	1145 (85.0)	98.2 (96.1 - 99.3)	19.3 (17.0 - 21.9)	1.22 (1.18 - 1.26)	0.93 (0.04 - 0.21)	28.6 (26.0 - 31.4)	97.0 (93.6 - 98.9)	17.5
	≥3	1004 (74.5)	95.5 (92.7 - 97.5)	32.4 (29.5 - 35.4)	1.41 (1.35 - 1.48)	0.14 (0.84 - 0.23)	31.8 (28.9 - 34.8)	95.6 (92.9 - 97.5)	27.9
	≥4	850 (63.1)	91.3 (87.8 - 94.1)	46.2 (43.1 - 49.3)	1.70 (1.59 - 1.81)	0.19 (0.13 - 0.27)	35.9 (32.7 - 39.2)	94.2 (91.7 - 96.1)	37.5
	≥5	720 (53.5)	87.4 (83.4 - 90.8)	57.7 (54.6 - 60.8)	2.07 (1.90 - 2.25)	0.22 (0.16 - 0.29)	40.6 (36.9 - 44.2)	93.3 (91.1 - 95.1)	45.1
	≥6	627 (46.6)	80.5 (75.9 - 84.6)	64.7 (61.6 - 67.6)	2.28 (2.07 - 2.52)	0.30 (0.24 - 0.38)	42.9 (39.0 - 46.9)	91.0 (88.6 - 93.0)	45.2
	≥7	**537 (39.9)**	**76.0 (71.1 - 80.5)**	**72.1 (69.2 - 74.8)**	**2.72 (2.42 - 3.06)**	**0.33 (0.27 - 0.40)**	**47.3 (43.0 - 51.6)**	**90.1 (87.9 - 92.1)**	**48.1**
	≥8	454 (33.7)	68.9 (63.6 - 73.8)	77.9 (75.2 - 80.4)	3.11 (2.72 - 3.57)	0.40 (0.34 - 0.47)	50.7 (46.0 - 55.4)	88.4 (86.1 - 90.4)	46.8
	≥9	371 (27.5)	60.2 (54.7 - 65.5)	83.2 (80.8 - 85.5)	3.59 (3.05 - 4.22)	0.48 (0.42 - 0.55)	54.2 (49.0 - 59.3)	86.4 (84.1 - 88.5)	43.4
	≥10	309 (22.9)	54.2 (48.7 - 59.6)	87.4 (85.2 - 89.3)	4.29 (3.55 - 5.18)	0.52 (0.47 - 0.59)	58.6 (53.9 - 63.1)	85.3 (83.7 - 86.7)	41.6
	≥11	252 (18.7)	44.9 (39.5 - 50.4)	89.9 (87.9 - 91.7)	4.46 (3.58 - 5.55)	0.61 (0.56 - 0.68)	59.5 (53.2 - 65.6)	83.2 (80.8 - 85.4)	34.8
	≥12	212 (15.7)	38.9 (33.7 - 44.4)	91.9 (90.1 - 93.5)	4.81 (3.76 - 6.16)	0.67 (0.61 - 0.73)	61.3 (54.4 - 67.9)	82.0 (79.7 - 84.2)	30.8
	≥13	161 (12.0)	31.4 (26.5 - 36.7)	94.5 (92.9 - 95.8)	5.69 (4.21 - 7.67)	0.73 (0.67 - 0.78)	65.2 (57.3 - 72.5)	80.7 (78.3 - 82.9)	25.9
	≥14	138 (10.2)	26.3 (21.7 - 31.4)	95.1 (93.5 - 96.3)	5.34 (3.86 - 7.38)	0.78 (0.73 - 0.83)	63.8 (55.2 - 71.8)	79.7 (77.3 - 81.9)	21.4
	≥15	102 (7.6)	19.8 (15.6 - 24.4)	96.4 (95.1 - 97.5)	5.56 (3.78 - 8.19)	0.83 (0.79 - 0.88)	64.7 (54.6 - 73.9)	78.5 (76.1 - 80.7)	16.2
	≥16	79 (5.9)	15.3 (11.6 - 19.6)	97.2 (96.0 - 98.2)	5.52 (3.54 - 8.61)	0.87 (0.83 - 0.91)	64.6 (53.0 - 75.0)	77.7 (75.3 - 79.9)	12.5
	≥17	57 (4.2)	10.5 (7.4 - 14.3)	97.8 (96.7 - 98.6)	4.83 (2.87 - 8.11)	0.92 (0.88 - 0.95)	61.4 (47.6 - 74.0)	76.8 (74.4 - 79.1)	8.3
	≥18	48 (3.6)	8.4 (5.6 - 11.9)	98.0 (97.0 - 98.8)	4.25 (2.42 - 7.44)	0.94 (0.90 - 0.97)	58.3 (43.2 - 72.4)	76.4 (74.0 - 78.7)	6.4
	≥19	40 (3.0)	6.9 (4.4 - 10.2)	98.3 (97.3 - 99.0)	4.10 (2.22 - 7.59)	0.95 (0.92 - 0.98)	57.5 (40.9 - 73.0)	76.2 (73.8 - 78.5)	5.2
	≥20	35 (2.4)	4.5 (2.5 - 7.3)	98.3 (97.3 - 99.0)	2.68 (1.35 - 5.30)	0.97 (0.95 - 1.00)	46.9 (29.1 - 65.3)	75.7 (73.3 - 78.0)	2.8
	≥21	26 (1.9)	3.9 (2.1 - 6.6)	98.7 (97.8 - 99.3)	3.03 (1.42 - 6.48)	0.97 (0.95 - 1.00)	50.0 (29.9 - 70.1)	75.7 (73.3 - 78.0)	2.6
	≥22	20 (1.5)	2.7 (1.2 - 5.1)	98.9 (98.1 - 99.5)	2.48 (1.04 - 5.94)	0.98 (0.97 - 1.00)	45.0 (23.1 - 68.5)	75.5 (73.1 - 77.8)	1.6
	≥23	16 (1.2)	1.5 (0.5 - 3.5)	98.9 (98.1 - 99.5)	1.38 (0.48 - 3.94)	1.00 (0.98 - 1.01)	31.3 (11.0 - 58.7)	75.3 (72.9 - 77.6)	0.4
	≥24	10 (0.7)	0.9 (0.2 - 2.6)	99.3 (98.6 - 99.7)	1.30 (0.34 - 5.00)	1.00 (0.99 - 1.01)	30.0 (6.7 - 65.2)	75.2 (72.8 - 77.5)	0.2
	≥25	4 (0.3)	0.3 (0.0 - 1.7)	99.7 (99.1 - 99.9)	1.01 (0.11 - 9.69)	1.00 (0.99 - 1.01)	25.0 (0.6 - 80.6)	75.2 (72.8 - 77.5)	0.0
	≥26	2 (0.2)	0.0 (0.0 - 1.1)	99.8 (99.3 - 100.0)	0.00 (0.00 - 0.00)	1.00 (1.00 - 1.00)	0.0 (0.0 - 84.2)	75.2 (72.8 - 77.5)	-0.2
	≥27	2 (0.2)	0.0 (0.0 - 1.1)	99.8 (99.3 - 100.0)	0.00 (0.00 - 0.00)	1.00 (1.00 - 1.00)	0.0 (0.0 - 84.2)	75.2 (72.8 - 77.5)	-0.2
PHQ-8	≥1	1244 (92.4)	99.4 (97.9 - 99.9)	10.0 (8.2 - 12.0)	1.10 (1.08 - 1.13)	0.06 (0.01 - 0.24)	26.7 (24.2 - 29.2)	98.1 (93.2 - 99.8)	9.4
	≥2	1141 (84.7)	98.2 (96.1 - 99.3)	19.7 (17.3 - 22.3)	1.22 (1.18 - 1.27)	0.09 (0.04 - 0.20)	28.7 (26.1 - 31.5)	97.1 (93.8 - 98.9)	17.9
	≥3	998 (74.1)	95.5 (92.7 - 97.5)	33.0 (30.1 - 36.0)	1.42 (1.36 - 1.50)	0.14 (0.08 - 0.23)	32.0 (29.1 - 35.0)	95.7 (93.0 - 97.6)	28.5
	≥4	846 (62.8)	91.3 (87.8 - 94.1)	46.6 (43.5 - 49.7)	1.71 (1.60 - 1.83)	0.19 (0.13 - 0.27)	36.1 (32.8 - 39.4)	94.2 (91.8 - 96.1)	37.9
	≥5	701 (52.0)	85.6 (81.4 - 89.2)	59.0 (55.9 - 62.1)	2.09 (1.92 - 2.28)	0.24 (0.19 - 0.32)	40.8 (37.1 - 44.5)	92.6 (90.3 - 94.5)	44.7
	≥6	611 (45.4)	79.9 (75.2 - 84.1)	66.0 (63.0 - 69.0)	2.35 (2.13 - 2.61)	0.30 (0.24 - 0.38)	43.7 (39.7 - 47.7)	90.9 (88.6 - 92.9)	46.0
	≥7	**521 (38.7)**	**74.3 (69.2 - 78.9)**	**73.1 (70.2 - 75.8)**	**2.76 (2.44 - 3.10)**	**0.35 (0.29 - 0.42)**	**47.6 (43.2 - 52.0)**	**89.6 (87.3 - 91.6)**	**47.4**

	≥8	428 (31.8)	65.6 (60.2 - 70.7)	79.4 (76.7 - 81.8)	3.18 (2.75 - 3.67)	0.43 (0.37 - 0.50)	51.2 (46.3 - 56.0)	87.5 (85.2 - 89.6)	44.9
	≥9	357 (26.5)	58.4 (52.9 - 63.7)	84.0 (81.6 - 86.2)	3.65 (3.09 - 4.32)	0.50 (0.44 - 0.56)	54.6 (49.3 - 59.9)	86.0 (83.6 - 88.1)	42.4
	≥10	286 (21.2)	50.0 (44.5 - 55.5)	88.3 (86.1 - 90.2)	4.26 (3.48 - 5.20)	0.57 (0.51 - 0.63)	58.4 (52.4 - 64.2)	84.3 (81.9 - 86.4)	38.3
	≥11	233 (17.3)	42.2 (36.9 - 47.7)	90.9 (89.0 - 92.6)	4.65 (3.69 - 5.86)	0.64 (0.58 - 0.70)	60.5 (53.9 - 66.8)	82.7 (80.3 - 84.9)	33.1
	≥12	189 (14.0)	34.7 (29.6 - 40.1)	92.8 (91.0 - 94.3)	4.82 (3.70 - 6.28)	0.70 (0.65 - 0.76)	61.4 (54.0 - 68.4)	81.2 (78.8 - 83.4)	27.5
	≥13	138 (10.2)	27.2 (22.5 - 32.4)	95.4 (93.9 - 96.6)	5.87 (4.22 - 8.17)	0.76 (0.71 - 0.82)	65.9 (57.4 - 73.8)	79.9 (77.5 - 82.1)	22.6
	≥14	103 (8.1)	21.9 (17.5 - 26.7)	96.4 (95.1 - 97.5)	6.15 (4.21 - 8.99)	0.81 (0.76 - 0.86)	67.0 (57.3 - 75.7)	78.9 (76.5 - 81.2)	18.3
	≥15	80 (5.9)	15.6 (11.9 - 19.9)	97.2 (96.0 - 98.2)	5.63 (3.62 - 8.77)	0.87 (0.83 - 0.91)	65.0 (53.5 - 75.3)	77.7 (75.3 - 80.0)	12.8
	≥16	58 (4.3)	10.5 (7.4 - 14.3)	97.7 (96.6 - 98.6)	4.62 (2.77 - 7.70)	0.92 (0.88 - 0.95)	60.3 (46.6 - 73.0)	76.8 (74.4 - 79.1)	8.2
	≥17	46 (3.4)	7.2 (4.7 - 10.5)	97.8 (96.7 - 98.6)	3.31 (1.88 - 5.82)	0.95 (0.92 - 0.98)	52.2 (36.9 - 67.1)	76.2 (73.8 - 78.5)	5.0
	≥18	36 (2.7)	5.7 (3.5 - 8.7)	98.3 (97.3 - 99.0)	3.39 (1.78 - 6.44)	0.96 (0.93 - 0.99)	52.8 (35.5 - 69.6)	76.0 (73.6 - 78.3)	4.0
	≥19	30 (2.2)	4.8 (2.8 - 7.7)	98.6 (97.7 - 99.2)	3.47 (1.71 - 7.03)	0.97 (0.94 - 0.99)	53.3 (34.3 - 71.7)	75.9 (73.4 - 78.1)	3.4
	≥20	24 (1.8)	3.0 (1.4 - 5.4)	98.6 (97.7 - 99.2)	2.17 (0.97 - 4.83)	0.98 (0.96 - 1.00)	41.7 (22.1 - 63.4)	75.5 (73.1 - 77.8)	1.6
	≥21	15 (1.1)	1.5 (0.5 - 3.5)	99.0 (98.2 - 99.5)	1.52 (0.52 - 4.41)	0.99 (0.98 - 1.01)	33.3 (11.8 - 61.6)	75.3 (72.9 - 77.6)	0.5
	≥22	7 (0.5)	0.6 (0.1 - 2.1)	99.5 (98.9 - 99.8)	1.21 (0.24 - 6.22)	1.00 (0.99 - 1.01)	28.6 (3.7 - 71.0)	75.2 (72.8 - 77.5)	0.1
	≥23	4 (0.3)	0.3 (0.0 - 1.7)	99.7 (99.1 - 99.9)	1.01 (0.11 - 9.69)	1.00 (0.99 - 1.01)	25.0 (0.6 - 80.6)	75.2 (72.8 - 77.5)	0.0
	≥24	2 (0.2)	0.0 (0.0 - 1.1)	99.8 (99.3 - 100.0)	0.00 (0.00 - 0.00)	1.00 (1.00 - 1.00)	0.0 (0.0 - 84.2)	75.2 (72.8 - 77.5)	-0.2
PHQ-9 algorithm		183 (13.6)	34.7 (29.6 - 40.1)	93.4 (91.7 - 94.8)	5.25 (3.99 - 6.91)	0.70 (0.65 - 0.76)	63.4 (56.0 - 70.4)	81.3 (78.9 - 83.5)	28.1
PHQ-9 adjusted algorithm		601 (44.6)	78.1 (73.3 - 82.5)	66.4 (63.4 - 69.3)	2.33 (2.10 - 2.58)	0.33 (0.27 - 0.41)	43.4 (39.4 - 47.5)	90.2 (87.9 - 92.3)	44.5
PHQ-2	≥1	1052 (78.1)	95.2 (92.3 - 97.2)	27.5 (24.8 - 30.4)	1.31 (1.26 - 1.37)	0.17 (0.11 - 0.28)	30.2 (27.5 - 33.1)	94.6 (91.3 - 96.9)	22.7
	≥2	**730 (54.2)**	**84.7 (80.4 - 88.4)**	**55.9 (52.8 - 59.0)**	**1.92 (1.77 - 2.09)**	**0.27 (0.21 - 0.35)**	**38.8 (35.2 - 42.4)**	**91.7 (89.3 - 93.8)**	**40.6**
	≥3	358 (26.6)	55.4 (49.9 - 60.8)	82.9 (80.5 - 85.2)	3.24 (2.75 - 3.83)	0.54 (0.48 - 0.61)	51.7 (46.4 - 57.0)	84.9 (82.6 - 87.1)	38.3
	≥4	225 (16.7)	39.5 (34.2 - 45.0)	90.8 (88.9 - 92.5)	4.30 (3.40 - 5.44)	0.67 (0.61 - 0.73)	58.7 (51.9 - 65.2)	82.0 (79.6 - 84.2)	30.3
	≥5	99 (7.4)	19.2 (15.1 - 23.8)	96.5 (95.2 - 97.6)	5.55 (3.74 - 8.22)	0.84 (0.79 - 0.88)	64.6 (54.4 - 74.0)	78.4 (76.0 - 80.6)	15.7
	≥6	63 (4.7)	12.0 (8.7 - 15.9)	97.7 (96.6 - 98.6)	5.27 (3.21 - 8.68)	0.90 (0.87 - 0.94)	63.5 (50.4 - 75.3)	77.1 (74.7 - 79.4)	9.7

Note: +LR = Positive likelihood ratio. −LR = Negative likelihood ratio. PPV= positive predictive value. NPV= negative predictive value. Youden's Index = Sensitivity+Specificity-1. 95%CI = 95% confidence interval. PHQ = Patient Health Questionnaire. Bold values represent the cohort point with the highest Youden Index.

Supplementary material 6. Sensitivity, specificity, PPV, NPV, and Youden's index for different PHQ and GAD cut-off points compared to the gold standard.

Method	Cut-off Points	n (%)	Sensitivity (95%CI)	Specificity (95%CI)	PPV (95%CI)	NPV (95%CI)	Youden Index
PHQ-9	**≥7**	**537 (39.9)**	**76.0 (71.1 - 80.5)**	**72.1 (69.2 - 74.8)**	**47.3 (43.0 - 51.6)**	**90.1 (87.9 - 92.1)**	**48.1**
PHQ-9	≥8	454 (33.7)	68.9 (63.6 - 73.8)	77.9 (75.2 - 80.4)	50.7 (46.0 - 55.4)	88.4 (86.1 - 90.4)	46.8
PHQ-9	≥9	371 (27.5)	60.2 (54.7 - 65.5)	83.2 (80.8 - 85.5)	54.2 (49.0 - 59.3)	86.4 (84.1 - 88.5)	43.4
PHQ-9	≥10	309 (22.9)	54.2 (48.7 - 59.6)	87.4 (85.2 - 89.3)	58.6 (53.9 - 63.1)	85.3 (83.7 - 86.7)	41.6
PHQ-9 algorithm	-	183 (13.6)	34.7 (29.6 - 40.1)	93.4 (91.7 - 94.8)	63.4 (56.0 - 70.4)	81.3 (78.9 - 83.5)	28.1
PHQ-9 adjusted algorithm	-	601 (44.6)	78.1 (73.3 - 82.5)	66.4 (63.4 - 69.3)	43.4 (39.4 - 47.5)	90.2 (87.9 - 92.3)	44.5
PHQ-8	**≥7**	**521 (38.7)**	**74.3 (69.2 - 78.9)**	**73.1 (70.2 - 75.8)**	**47.6 (43.2 - 52.0)**	**89.6 (87.3 - 91.6)**	**47.4**
PHQ-8	≥8	428 (31.8)	65.6 (60.2 - 70.7)	79.4 (76.7 - 81.8)	51.2 (46.3 - 56.0)	87.5 (85.2 - 89.6)	44.9
PHQ-8	≥9	357 (26.5)	58.4 (52.9 - 63.7)	84.0 (81.6 - 86.2)	54.6 (49.3 - 59.9)	86.0 (83.6 - 88.1)	42.4
PHQ-8	≥10	286 (21.2)	50.0 (44.5 - 55.5)	88.3 (86.1 - 90.2)	58.4 (52.4 - 64.2)	84.3 (81.9 - 86.4)	38.3
PHQ-2	**≥2**	**730 (54.2)**	**84.7 (80.4 - 88.4)**	**55.9 (52.8 - 59.0)**	**38.8 (35.2 - 42.4)**	**91.7 (89.3 - 93.8)**	**40.6**
PHQ-2	≥3	358 (26.6)	55.4 (49.9 - 60.8)	82.9 (80.5 - 85.2)	51.7 (46.4 - 57.0)	84.9 (82.6 - 87.1)	38.3
GAD-7	≥7	360 (26.7)	53.6 (33.9 - 72.5)	73.8 (71.4 - 76.2)	4.2 (2.4 - 6.8)	98.7 (97.8 - 99.3)	27.4
GAD-7	**≥8**	**295 (21.9)**	**53.6 (33.9 - 72.5)**	**78.8 (76.5 - 81.0)**	**5.1 (2.9 - 8.3)**	**98.8 (97.9 - 99.3)**	**32.4**
GAD-7	≥9	230 (17.1)	42.9 (24.5 - 62.8)	83.5 (81.4 - 85.4)	5.2 (2.7 - 8.9)	98.6 (97.7 - 99.2)	26.4
GAD-7	≥10	164 (12.2)	39.3 (21.5 - 59.4)	88.4 (86.5 - 90.1)	6.7 (3.4 - 11.7)	98.6 (97.7 - 99.2)	27.7
GAD-2	**≥2**	**681 (50.6)**	**82.1 (63.1 - 93.9)**	**50.1 (47.4 - 52.8)**	**3.4 (2.2 - 5.0)**	**99.2 (98.3 - 99.8)**	**32.2**
GAD-2	≥3	283 (21.0)	42.9 (24.5 - 62.8)	79.5 (77.2 - 81.6)	4.2 (2.2 - 7.3)	98.5 (97.6 - 99.1)	22.4

Note: PPV= positive predictive value. NPV= negative predictive value. Youden's Index = Sensitivity+Specificity-1. 95%CI = 95% confidence interval. GAD =Generalized Anxiety Disorder Scale. PHQ = Patient Health Questionnaire. Bold values represent the cohort point with the highest Youden Index. All cut-off points for the PHQ and GAD versions can be found in supplementary materials 3 and 4, respectively.

Supplementary material 7. Sensitivity, specificity, +LR, −LR, PPV, NPV, and Youden's index for different GAD cut-off points compared to the gold standard.

Method	Cut-off Points	n (%)	Sensitivity (95%CI)	Specificity (95%CI)	+LR (95%CI)	-LR (95%CI)	PPV (95%CI)	NPV (95%CI)	Youden Index
GAD-7	≥1	1244 (92.4)	96.4 (81.7 - 99.9)	7.7 (6.4 - 9.3)	1.05 (0.97 - 1.12)	0.46 (0.07 - 3.19)	2.2 (1.4 - 3.1)	99.0 (94.7 - 100.0)	4.1
	≥2	1128 (83.7)	96.4 (81.7 - 99.9)	16.5 (14.6 - 18.6)	1.16 (1.07 - 1.25)	0.22 (0.03 - 1.49)	2.4 (1.6 - 3.5)	99.5 (97.5 - 100.0)	12.9
	≥3	960 (71.3)	96.4 (81.7 - 99.9)	29.3 (26.8 - 31.8)	1.36 (1.26 - 1.48)	0.12 (0.02 - 0.84)	2.8 (1.9 - 4.1)	99.7 (98.6 - 100.0)	25.7
	≥4	797 (59.2)	89.3 (71.8 - 97.7)	41.5 (38.8 - 44.2)	1.53 (1.33 - 1.75)	0.26 (0.09 - 0.75)	3.1 (2.0 - 4.6)	99.5 (98.4 - 99.9)	30.8
	≥5	640 (47.5)	71.4 (51.3 - 86.8)	53.0 (50.3 - 55.7)	1.52 (1.19 - 1.93)	0.54 (0.30 - 0.97)	3.1 (1.9 - 4.8)	98.9 (97.8 - 99.5)	24.4
	≥6	491 (36.5)	60.7 (40.6 - 78.5)	64.1 (61.4 - 66.7)	1.69 (1.24 - 2.30)	0.61 (0.39 - 0.97)	3.5 (2.0 - 5.5)	98.7 (97.7 - 99.4)	24.8
	≥7	360 (26.7)	53.6 (33.9 - 72.5)	73.8 (71.4 - 76.2)	2.05 (1.43 - 2.93)	0.63 (0.42 - 0.94)	4.2 (2.4 - 6.8)	98.7 (97.8 - 99.3)	27.4
	≥8	**295 (21.9)**	**53.6 (33.9 - 72.5)**	**78.8 (76.5 - 81.0)**	**2.52 (1.76 - 3.62)**	**0.59 (0.40 - 0.88)**	**5.1 (2.9 - 8.3)**	**98.8 (97.9 - 99.3)**	**32.4**
	≥9	230 (17.1)	42.9 (24.5 - 62.8)	83.5 (81.4 - 85.4)	2.59 (1.66 - 4.04)	0.69 (0.50 - 0.94)	5.2 (2.7 - 8.9)	98.6 (97.7 - 99.2)	26.4
	≥10	164 (12.2)	39.3 (21.5 - 59.4)	88.4 (86.5 - 90.1)	3.39 (2.09 - 5.50)	0.69 (0.51 - 0.93)	6.7 (3.4 - 11.7)	98.6 (97.7 - 99.2)	27.7
	≥11	120 (8.9)	39.3 (21.5 - 59.4)	91.7 (90.1 - 93.2)	4.75 (2.90 - 7.79)	0.66 (0.49 - 0.89)	9.2 (4.7 - 15.8)	98.6 (97.8 - 99.2)	31.0
	≥12	94 (7.0)	28.6 (13.2 - 48.7)	93.5 (92.0 - 94.8)	4.38 (2.36 - 8.15)	0.76 (0.60 - 0.97)	8.5 (3.8 - 16.1)	98.4 (97.5 - 99.0)	22.1
	≥13	67 (5.0)	28.6 (13.2 - 48.7)	95.5 (94.3 - 96.6)	6.39 (3.39 - 12.10)	0.75 (0.59 - 0.95)	11.9 (5.3 - 22.2)	98.4 (97.6 - 99.0)	24.1
	≥14	50 (3.7)	21.4 (8.3 - 41.0)	96.7 (95.5 - 97.6)	6.42 (2.98 - 13.80)	0.81 (0.67 - 0.99)	12.0 (4.5 - 24.3)	98.3 (97.4 - 98.9)	18.1
	≥15	40 (3.0)	14.3 (4.0 - 32.7)	97.3 (96.2 - 98.1)	5.23 (2.00 - 13.70)	0.88 (0.76 - 1.03)	10.0 (2.8 - 23.7)	98.2 (97.3 - 98.8)	11.6
	≥16	31 (2.3)	10.7 (2.3 - 28.2)	97.9 (96.9 - 98.6)	5.05 (1.63 - 15.60)	0.91 (0.80 - 1.04)	9.7 (2.0 - 25.8)	98.1 (97.2 - 98.8)	8.6
	≥17	19 (1.4)	3.6 (0.1 - 18.3)	98.6 (97.9 - 99.2)	2.62 (0.36 - 18.90)	0.98 (0.91 - 1.05)	5.3 (0.1 - 26.0)	98.0 (97.1 - 98.7)	2.2
	≥18	15 (1.1)	0.0 (0.0 - 12.3)	98.9 (98.1 - 99.4)	0.00 (0.00 - 0.00)	1.01 (1.01 - 1.02)	0.0 (0.0 - 21.8)	97.9 (97.0 - 98.6)	-1.1
	≥19	12 (0.9)	0.0 (0.0 - 12.3)	99.1 (98.4 - 99.5)	0.00 (0.00 - 0.00)	1.01 (1.00 - 1.01)	0.0 (0.0 - 26.5)	97.9 (97.0 - 98.6)	-0.9
	≥20	8 (0.6)	0.0 (0.0 - 12.3)	99.4 (98.8 - 99.7)	0.00 (0.00 - 0.00)	1.01 (1.00 - 1.01)	0.0 (0.0 - 36.9)	97.9 (97.0 - 98.6)	-0.6
	≥21	5 (0.4)	0.0 (0.0 - 12.3)	99.6 (99.1 - 99.9)	0.00 (0.00 - 0.00)	1.00 (1.00 - 1.01)	0.0 (0.0 - 52.2)	97.9 (97.0 - 98.6)	-0.4
GAD-2	≥1	1028 (76.3)	89.3 (71.8 - 97.7)	24.0 (21.7 - 26.4)	1.17 (1.03 - 1.34)	0.45 (0.15 - 1.31)	2.4 (1.6 - 3.6)	99.1 (97.3 - 99.8)	13.3
	≥2	**681 (50.6)**	**82.1 (63.1 - 93.9)**	**50.1 (47.4 - 52.8)**	**1.65 (1.37 - 1.97)**	**0.36 (0.16 - 0.79)**	**3.4 (2.2 - 5.0)**	**99.2 (98.3 - 99.8)**	**32.2**
	≥3	283 (21.0)	42.9 (24.5 - 62.8)	79.5 (77.2 - 81.6)	2.09 (1.34 - 3.24)	0.72 (0.52 - 0.99)	4.2 (2.2 - 7.3)	98.5 (97.6 - 99.1)	22.4
	≥4	141 (10.5)	25.0 (10.7 - 44.9)	89.8 (88.1 - 91.4)	2.46 (1.27 - 4.77)	0.84 (0.67 - 1.03)	5.0 (2.0 - 10.0)	98.3 (97.4 - 98.9)	14.8
	≥5	51 (3.8)	14.3 (4.0 - 32.7)	96.4 (95.3 - 97.4)	4.01 (1.55 - 10.40)	0.89 (0.76 - 1.03)	7.8 (2.2 - 18.9)	98.1 (97.3 - 98.8)	10.7
	≥6	28 (2.1)	7.1 (0.9 - 23.5)	98.0 (97.1 - 98.7)	3.62 (0.90 - 14.50)	0.95 (0.86 - 1.05)	7.1 (0.9 - 23.5)	98.0 (97.1 - 98.7)	5.1

Note: +LR = Positive likelihood ratio. −LR = Negative likelihood ratio. PPV= positive predictive value. NPV= negative predictive value. Youden's Index = Sensitivity+Specificity-1. 95%CI = 95% confidence interval. GAD =Generalized Anxiety Disorder Scale. Bold values represent the cut-off with the highest Youden Index.

STROBE Statement—Checklist of items that should be included in reports of ***cross-sectional studies***

	Item No	Recommendation	Page
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	3
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	5
Objectives	3	State specific objectives, including any prespecified hypotheses	6
Methods			
Study design	4	Present key elements of study design early in the paper	6
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	6
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants	6
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	7
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	*7*
Bias	9	Describe any efforts to address potential sources of bias	
Study size	10	Explain how the study size was arrived at	6
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	8
		(*b*) Describe any methods used to examine subgroups and interactions	8
		(*c*) Explain how missing data were addressed	8
		(*d*) If applicable, describe analytical methods taking account of sampling strategy	8
		(*e*) Describe any sensitivity analyses	9
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	9
		(b) Give reasons for non-participation at each stage	9
		(c) Consider use of a flow diagram	SUPPL.
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	10
		(b) Indicate number of participants with missing data for each variable of interest	10
Outcome data	15*	Report numbers of outcome events or summary measures	10
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	10

		(*b*) Report category boundaries when continuous variables were categorized	10
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	
Discussion			
Key results	18	Summarise key results with reference to study objectives	11
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	13
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	11
Generalisability	21	Discuss the generalisability (external validity) of the study results	11
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	16

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.