



OPEN

New cut-off points of PHQ-9 and its variants, in Costa Rica: a nationwide observational study

Armando González-Sánchez^{1,2✉}, Raúl Ortega-Moreno³, Greibin Villegas-Barahona⁴, Eva Carazo-Vargas⁴, Harold Arias-LeClaire⁴ & Purificación Vicente-Galindo^{1,2}

The PHQ-9 questionnaire is a screening test worldwide used to measure depression. But it cannot be used in Costa Rica, due to the fact that it has not previously been validated for its population. The present study aims to show the validation of the PHQ-9 questionnaire and its variants (PHQ-2, PHQ-4, PHQ-8) in a population sample of adults residing in Costa Rica. A sample was collected ($n = 1162$) using a self-administered questionnaire. Confirmatory Factor Analysis (CFA), Receiver Operating Characteristic (ROC) curve, and Multiple Group Confirmatory Factor Analysis (MG-CFA) were tested. One factor was found that explained 73.33% of the variance with excellent internal consistency ($\alpha = 0.928$). Goodness-of-fit measures were adequate (RMSEA = 0.107; CFI = 0.948), as was diagnostic power at a cut-off of 10 (78.60 for Sensitivity and 27.95 for 1-Specificity). External validation indices were good ($r = 0.843$ with GAD-7, $r = -0.647$ with RS14, and $r = 0.301$ with FCV19S), and the model showed invariance by sex ($\Delta\chi^2 = 27.90$; $df = 27$; $p < 0.001$). Additionally, new cut-off points were proposed for PHQ-9 and its variants for Costa Rican male, female, and general populations. The PHQ-9 and its variants (PHQ-2, 4, and 8) are valid tools for detecting depression (and anxiety for PHQ-4) in Costa Rican population. In addition, new cut-off points differentiated by sex are proposed.

Depression is a deterioration of a person's mental health that implies a lower ability of the individual to function, characterized by sadness, emptiness, or irritability¹. By influencing the ability to perform tasks, the depression can produce a reduction in productivity and impairment that implies disability^{2,3}. The age of onset for depressive disorders is between 30 and 35 years old⁴. Among the factors involved in the tendency to suffer from depression are job instability and low social support⁵⁻⁷.

Depressive symptoms constitute one of the main mental health problems; thus, most psychiatric care is devoted to treating depression⁸. Worldwide, it is estimated that 12.26% of the population suffers from mental disorders, with depression at 3.44%; in Central America, the prevalence of depressive disorders stands at 3.19%⁹. In Costa Rica, the prevalence of concurrent symptoms with major depressive disorder was 9.5% at pre-pandemic dates due to COVID-19 ($n = 797$), while during the pandemic it was 61.0% ($n = 6786$)⁷. In addition, it has a rate of 7.6 (95% CI 5.29–10.53) deaths by suicide per 100,000 inhabitants in the country, ranking above most countries in the Americas¹⁰. Although one of the characteristics of depression is suicidal ideation, it is difficult to predict imminent suicide¹¹.

There are sexual differences in the comorbidity of mental disorders, since women tend to internalize disorders, while men tend to develop external disorders such as substance misuse⁶. As in the rest of the world, in Costa Rica women have a higher prevalence of depression than men^{7,12}, although the reasons for these differences are still being studied^{4,13}.

This has consequences to the economy, due to inability to carry out work activities, in the quality of life due to the disability it produces, loss of life, the reduction of educational opportunities, work and social connections, affecting future generations^{3,14,15}. In Costa Rica, 8.0% of disability is due to exclusively depressive disorders, excluding substance-related disorders, self-harm, and suicide; mainly affecting working-age people³. This implies that mental health depends in part on the social and economic conditions in which the person finds him- or herself.

¹Department of Statistics and Operative Research, University of Salamanca, Campus Miguel de Unamuno. C/ Alfonso X El Sabio, S/N, 37007 Salamanca, Spain. ²Instituto de Investigación Biomédica de Salamanca (IBSAL), Hospital Virgen de la Vega, 10ª Planta, Paseo de San Vicente, 58-182, 37007 Salamanca, Spain. ³Psychology School, National University, Costa Rica (UNA), Heredia, Costa Rica. ⁴Distance State University, Costa Rica (UNED), San Jose, Costa Rica. ✉email: armando_gonzalez@usal.es

Being a treatable and preventable disease, its early identification avoids health complications. Therefore, prevention, early detection, and treatment efforts are essential¹⁵. With adequate detection of depression, \$11,134 per year could be saved per woman and \$34,065 per man¹⁵. Thus, routine detection of depression in populations is cost-effective regardless of gender or age¹⁵. Population screenings are effective both through self-administered surveys and during the COVID-19 pandemic¹⁶.

The benefits of detecting depression early can only be achieved with proper studies, since making erroneous diagnoses not only makes detection more difficult but can also lead to great emotional distress, worse prognosis, and discrimination by social characteristics or group belonging^{17–20}.

Therefore, it would be necessary to consider the importance of screening the population properly. If a tool is used in circumstances other than those for which it was designed, such as a different culture or language, adaptations must be made so that the tools are equivalent²¹. The equivalence of them in populations is closely linked to validity, especially when administered across cultures²². Conceptual equivalence (the meaning attributed to a given concept), items (guaranteeing relevance, clarity and relevance), semantics (cross-cultural meaning), operation (administration, qualification and interpretation under the same conditions for quality assessment), measurement (referring to identical psychometric properties between cultures) and functional equivalence (fulfillment of test purpose: adequate measurement of construct while respecting group independence) are all important considerations²³.

The PHQ-9 is a test used for assessing the severity of depression in adults. It can be completed within 5 min and is self-administered. It can be used for monitoring the evolution of depression in response to treatment^{24,25}. It can be applied to several patients and have results in a short time. It is often used in primary care to diagnose depression and its severity quickly and efficiently²⁶. It has been studied for its high relationship with the anxiety test GAD-7^{27,28}, and Fear of Covid-19²⁹, and indirect relationship with resilience tested by RS14^{30,31}.

PHQ-9 has been translated into many different languages^{32–35}. It has been validated in Latin American countries such as Peru^{36,37} or Chile^{38,39}, and has been widely used as a tool for identifying symptoms associated with depression during the COVID-19 pandemic throughout the world^{40–43}. However, as of the date of the presentation of this document, there is no reference to a validation in Costa Rica.

There have only been a few studies of depression in Costa Rica, offering prevalence data that is not close to other populations⁴⁴. In this way, prevalence studies place depression with very different values: 1.08%⁴⁵, 12.6%⁴⁶ and 27.3%⁴⁷. Likewise, in women it manifests itself with a ratio of 3 to 1. Culturally, men do not express their feelings and it is only after 30 years of age that the first diagnoses of depression are given⁴⁷.

The main goal of this article is to evaluate the psychometric properties of the Spanish version of the PHQ-9 and its variants: PHQ-8, PHQ-4, and PHQ-2 questionnaires, in a sample of Costa Rican adults by testing the factorial structure⁴⁸. We expect it to have a unifactorial structure, as in the original validation of the tool in Spanish⁴⁹. The second objective is to assess the cross-sectional multigroup invariance (MGCFI) (female versus male participants) of the PHQ-9 test and its variants.

To do this, the descriptive statistics of the tests will be verified, their psychometric properties evaluated, new cut-off points established for both the general population and differentiated by sex, MGCFI multigroup invariance verified, and relationships between the tests established through multivariate analysis.

Methods

Study population

Costa Rica is the Latin American country listed as the happiest⁵⁰. The country lacks robust health coverage and mental health research is not advanced; This implies that data on the characteristics of depression are unknown⁵¹. This circumstance is more accentuated in rural areas where mental health professionals are scarcer⁵¹. Additionally, addresses are often described in relation to local landmarks rather than street names or house numbers, leading to complications in providing mental healthcare.

Costa Rica population has a total of 84.7% of internet access (87.0% urban and 78.7% rural), also 96% have phones, 47% have computer, and 16.7% have tablet⁵². Costa Rica has a low level of upper secondary and post-secondary educational level (18%) and the lowest percentage of doctoral achievement 0.1% of the OECD countries⁵³.

Its ethnic distribution is made up of: Caucasians, who constitute 83.6%, mixed race 6.72%, South American Indians 2.42%, the rest are Undeclared or Others⁵⁴.

Costa Rican Spanish recognizes twelve linguistic features spread across five large regions, forming internal sociolinguistic differences⁵⁵, there being a complex compendium of unique social characteristics such as *gypsisms* such as the *achará*⁵⁶, making it difficult to understand and express the language among other cultures. Where other versions of Spanish are spoken⁵⁷, such as Spanish or Mexican the questionnaire was validated previously.

The inclusion criteria were adults residing in Costa Rica at the time of evaluation, ability to read, and understand Spanish, and having expressed their explicit consent to participate in the study.

A total of 14,702 responses were obtained. Of these, 13,506 responses had some missing data and were eliminated. To facilitate responses, completing the survey was not made mandatory. Data such as age and sex, or incomplete responses to the tool were used as criteria for eliminating responses. Given that we had sufficient data, a decision was made to perform data cleaning of this magnitude. An additional 34 responses were removed due to being completed too fast. Our sample consisted of 1162 completed surveys. The age, gender, civil status, education level, province, and employment status of each participant were recorded (Table 1).

In a previous analysis, a minimum sample size of 1100 subjects was required to calculate structural equation models⁵⁸.

Sociodemographic variable	Response option	Value (n. %)
Gender	Male	247 (21.3)
	Female	915 (78.7)
Civil status	Single	561 (48.3)
	Married	361 (31.1)
	Coupled	116 (10.0)
	Divorced	94 (8.1)
	Separated	19 (1.6)
	Widowed	9 (0.8)
	Missing	2 (0.2)
Education level	Non universitarian	598 (51.5)
	Universitarian	378 (32.5)
	Master	167 (14.4)
	PhD	19 (1.6)
Province	San José	459 (39.5)
	Alajuela	176 (15.1)
	Cartago	205 (17.6)
	Heredia	149 (12.8)
	Guanacaste	94 (8.1)
	Puntarenas	42 (3.6)
	Limón	37 (3.2)
Labor condition	Employed	651 (56.0)
	Unemployed	105 (9.0)
	Others	406 (34.9)

Table 1. Sociodemographic characteristics of the sample.

Instruments

The use of the PHQ-9 during the pandemic allows for reaching adequate conclusions⁵⁹. The PHQ-9 was administered, and is used to detect symptomatology concurrent with depression by screening, which has excellent psychometric properties, both its internal consistency ($\alpha = 0.89$) and its test–retest reliability ($ICC = 0.84$)^{26,60}. Based on Classical Test Theory, its items have scores ranging from 0 to 3. The PHQ-9 scale has total scores ranging from 0 to 27 and is obtained by adding the score of each item. The test correction has cut-off points of 5, 10, 15, and 20, resulting in the following classification: minimal depression (0–4); mild depression (5–9); moderate depression (10–14); moderate to severe depression (15–19); and severe depression (20–27). The PHQ-9 test consists of nine main items plus an additional three items that ask if you have experienced discomfort due to any of the problems mentioned in the main items. These additional items refer to the difficulty caused by these problems in three contexts: work, household tasks, or social relations. The last three items have response options ranging from 0 (no difficulty) to 3 (extreme difficulty).

For this research, the team made minor modifications taking from example various versions of the PHQ-9 test. We selected items from different versions of the test that had already been validated in other Spanish-speaking countries. These cultural modifications were made to improve comprehension while preserving the original meaning of the items (see Supplementary material).

The PHQ-9 has several variants. The PHQ-2 version is considered within the PHQ-9 as it corresponds to its first two items, forming a reduced version of the main version. Additionally, these two items are considered the core criterion for depressive disorder. Its score range goes from 0 to 6, with scores greater than or equal to 3 indicating depression^{61–64}. The PHQ-8 version includes all items of the PHQ-9 except for the ninth item related to suicide and indicates current depression at a cut-off point ≥ 10 ⁶⁵. The PHQ-4 version consists of the first two items of the PHQ-9 combined with the first two items of the GAD-7 (General Anxiety Disorder-7), thus examining both depression and anxiety⁶⁶.

For external validity, we used tests to assess anxiety, resilience, and fear of Covid. The GAD-7 is a screening scale for measuring generalized anxiety disorder⁶⁷. It has strong internal consistency ($\alpha = 0.94$) and good test–retest reliability ($ICC = 0.83$) and uses seven items with response options ranging from 0 to 3. The total score of the test results from the sum of its items, with minimum and maximum scores ranging from 0 to 21. The thresholds proposed in its original scale are: minimal anxiety (0–4), mild anxiety (5–9), moderate anxiety (10–14), and severe anxiety (15–21). The GAD-2 questionnaire is included in the GAD-7 and corresponds to its first two items. A score ≥ 3 indicates clinically relevant anxiety disorder^{62,67–69}. The RS-14 Resilience test measures an individual's degree of adaptation to adverse situations. This scale consists of 14 items with theoretical scores ranging from 98 to 14, indicating different levels of resilience according to this score range: very high resilience (98–82), high resilience (81–64), fair (63–49), low resilience (48–31), and very low resilience (30–14)⁷⁰. The Fear of Covid-19 Scale measures fear of Coronavirus disease (FCV-19S). With theoretical scores ranging from 7 to 35^{71,72}, it indicates a direct relationship between fear of illness and test score.

Finally, as an external criterion, questions about current depression and anxiety were included in the tool: suffering from depression; suffering from anxiety; taking medication in the last 30 days for depression, sleep, or anxiety; or being in psychiatric or psychological treatment in the last 30 days.

Informed consent was obtained from all participants.

Procedure

A self-administered survey hosted on an online survey platform was used (LimeSurvey). The access was made through a link. The survey was answered by adults from the general population of Costa Rica for 8 months (between March 22nd and September 22nd, 2021), during the second year of the pandemic in the country. The population just needed internet and a device for answering the survey, as it can be done using a mobile device, tablet, or computer. Participation in this study was made from the Ministry of Health of Costa Rica, Distance Education University of Costa Rica (UNED), the National University (UNA), and the Costa Rican Social Security Fund (CCSS). The institutions requested participation through their institutional websites, and their social networks: Facebook and Twitter.

Four attentional check items were distributed throughout the questionnaire: “If you are reading this, please select the following option: totally agree/agree/never/always”.

Participation was assured to be anonymous, confidential, voluntary, and participants could withdraw from the survey at any time. Additionally, information on national psychological care and resources was available at the end of the survey.

All subjects gave their informed consent for inclusion before they participated in the study. The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Ethics Committee of the Ministry of Health through the National Health Research Council (CONIS) under Agreement No. 22 of the ordinary session No. 38 of August 26, 2020, with the number: CONIS-242-2020 confirming that all research was performed in accordance with relevant guidelines or regulations. This study was not preregistered.

Data analysis

First, descriptive statistics were obtained for frequency, means, standard deviation, asymmetry, and kurtosis using the statistical program SPSS v.25⁷³. To estimate concurrent validity, Spearman's Rho and Pearson's correlations were used.

The FACTOR v.12.01.02 program⁷⁴ was used for internal validity analyses, descriptives, frequencies and item properties. Factor analysis was performed following recommended standards⁷⁵. Matrix analysis was conducted through polychoric correlation with the Minimum Rank Factor Analysis factor model (MRFA)⁷⁶. To determine the Kaiser–Meyer–Olkin index (KMO), the entire sample was used as two random subsamples with the Solomon method⁷⁷. In the analysis presented below, explained variance is based on eigenvalues. Additionally, the Measure of Sampling Adequacy (MSA) statistic is offered which advises eliminating items with values less than 0.50⁷⁴. For internal consistency, McDonald's Omega (ω) statistic was used under the Omega macro for SPSS beta 0.2 package⁷⁸.

To calculate sex invariance in the model, Multi-Group Confirmatory Factor Analysis (MGCFA) was used and executed using AMOS⁷⁹. Changes in χ^2 were reported with their degrees of freedom and p -values. The parameters AIC, GFI and RMSEA were reported to demonstrate goodness-of-fit to a single model factor structure for PHQ-9 test.

To calculate optimal cohort points for depression, being in psychiatric or psychological treatment for depression was an additional question, this determines robustly the two different groups: one with a depressive trait and another without it. Similarly, to establish cut-off points for PHQ-4 test, criteria such as suffering from depression while also suffering from anxiety were used. The Youden's J index and Closest-top-left criteria on ROC curve were used; when these indexes did not coincide, Youden's J statistic was chosen⁸⁰. Sensitivity, specificity, and area under ROC curve are also presented. Cut-off points for PHQ-8 were established based on its original version while respecting severity length from optimal shear point⁶⁵.

Finally, to calculate new cut-off points indicating severity of depression for each test, corresponding bases were added to cut-off points obtained using Youden J indices and Closest top left on ROC curve with respect to ranges composing each court to establish severity of depressive symptoms.

A confidence level of 0.05 was marked as significant for all analyses including confidence intervals (CI) at 95%.

Results

Participants

A total of 1162 people participated in study with a mean age of 35.52 years old (SD = 12.18) ranging from 15 to 78 years old (Table 1).

Psychometric properties

The mean and standard deviation of PHQ-2, PHQ-4, PHQ-8, and PHQ-9 scales were 2.38 (1.94), 4.97 (3.65), 9.33 (6.90) and 9.80 (7.43) respectively, with scores reach-ing theoretical range of tests.

Internal consistency of items was high, with lower scores in item 9 of depression (Table 2). Among different scales, high correlations were observed, positive for depression and fear of COVID-19 and negative for resilience (Table 2).

Eliminating any item did not increase the internal consistency of the questionnaire (Table 3). The suicide item represents an extreme condition of depression and low scores. It forms part of the general dimension of depression and is recommended to remain on the scale to assess depression. Item 4 (loss of energy) is considered the easiest item to answer even though items 1 and 2 make up the core of major depressive syndrome.

Psychometrics	No. Item									
Dimension	Parameter	1	2	3	4	5	6	7	8	9
Item properties	Fixed correlation item-total	0.78	0.83	0.74	0.76	0.76	0.79	0.74	0.66	0.59
	α if item deleted	0.92	0.91	0.92	0.92	0.92	0.92	0.92	0.93	0.93
	Load factors	0.84	0.87	0.80	0.82	0.82	0.84	0.80	0.73	0.66
	QIM	3	3	3	3	3	3	3	2	2
	RDI	0.40	0.39	0.45	0.54	0.42	0.37	0.31	0.21	0.15
	Normed MSA	0.94	0.90	0.95	0.94	0.94	0.95	0.95	0.95	0.94
Descriptives	Mean	1.20	1.18	1.36	1.62	1.27	1.12	0.94	0.63	0.46
	SD	1.01	1.06	1.13	1.06	1.10	1.14	1.00	0.92	0.86
	CI 95%	(1.12–1.28)	(1.10–1.26)	(1.27–1.44)	(1.54–1.70)	(1.19–1.36)	(1.04–1.21)	(0.87–1.02)	(0.56–0.70)	(0.40–0.52)
	Skewness	0.48	0.51	0.29	0.06	0.37	0.53	0.78	1.37	1.88
	Kurtosis	–0.85	–0.96	–1.31	–1.30	–1.19	–1.15	–0.51	0.82	2.48
	Loading weights by ULS	0.87	0.93	0.83	0.87	0.84	0.88	0.83	0.79	0.76
	Communality	0.70	0.76	0.64	0.66	0.67	0.71	0.63	0.53	0.68
Answer frequencies	Not at all	324	365	323	172	351	468	486	696	838
	Several days	459	423	386	441	381	303	384	283	186
	More than half the days	201	170	165	204	190	173	162	95	65
	Nearly every day	178	204	288	345	240	218	130	88	73

Table 3. Descriptives, item properties and frequencies. PHQ-9.

Dimension	Parameter	PHQ2	PHQ4	PHQ8	PHQ9
Psychometrics	Explained variance (1 factor)	91.47	76.84	75.43	73.39
	KMO	0.5	0.780	0.930	0.930
	Bartlett	1350.2 NS	3110.5**	8881.6**	9957.6**
	Alpha (α)	0.853	0.899	0.929	0.928
	McDonald's Omega ($O\omega$)	NA	0.898	0.930	0.931
	Items removed by MSA ^a	NA	NA	NA	NA
Goodness-of-fit (1 factor)	χ^2_{gf}	Undefined	45.75 ₂ (<0.001)	309.91 ₂₀ (<0.001)	386.71 ₂₇ (<0.001)
	AIC ^b	–	61.75	341.91	422.71
	GFI	–	0.915	0.936	0.927
	RMSEA	0.898	0.298	0.112	0.107
	$\Delta\chi^2_{gf}$ by sex (<i>p</i> -value)	–	261.22 ₄ (<0.001)	329.03 ₄₀ (<0.001)	27,90 ₂₇ (<0.001)

Table 4. Psychometric properties of the different versions of the PHQ-9. ^aMSA, Measure of Sampling Adequacy. ^bAIC were calculated using the default model. **p*-value > 0.05. ***p*-value > 0.01.

The PHQ-9 test had a determinant close to zero (0.002) and obtained a similar KMO measure for both subsamples (0.937 and 0.933) and for the complete sample (0.996). A single factor was extracted that explained 73.33% of the variance. The eigenvalue of its next component did not reach unity (0.542).

The respective indices were obtained for different versions of PHQ-9 to calculate their internal consistency and factorial structure (Table 4). McDonald's Omega could not be calculated for PHQ-2 because it only has two items; thus, a reduction in dimensions was not possible. However, results are presented for all cases where a single factor was extracted.

Cut-off-points

As shown in Fig. 1, different versions of PHQ-9 function similarly for depression prior to diagnosis criteria. Receiver Operating Characteristic (ROC) curves can be ordered by the number of items included in each questionnaire version.

When analyzing the segment in the upper left corner (Closest-top-left) on the ROC curve, tests with more items have better specificity and sensitivity indices; however, these differences are slight (Table 5).

Table 5 shows different cut-off points of PHQ versions for both the general population and the population according to sex. Optimal thresholds for the general population coincide with PHQ-2 but are disaggregated in the rest. Consequently, the optimal cohort points for men are 1 or 2 points lower than for women, depending on the version administered. In PHQ-8 and PHQ-9 for women, the optimal cut-off score is 11 points. On the other hand, for males, the optimal score is 8 and 10 points, respectively. Below is Table 6 summarizing different thresholds found as well as the severity of depression according to different segments.

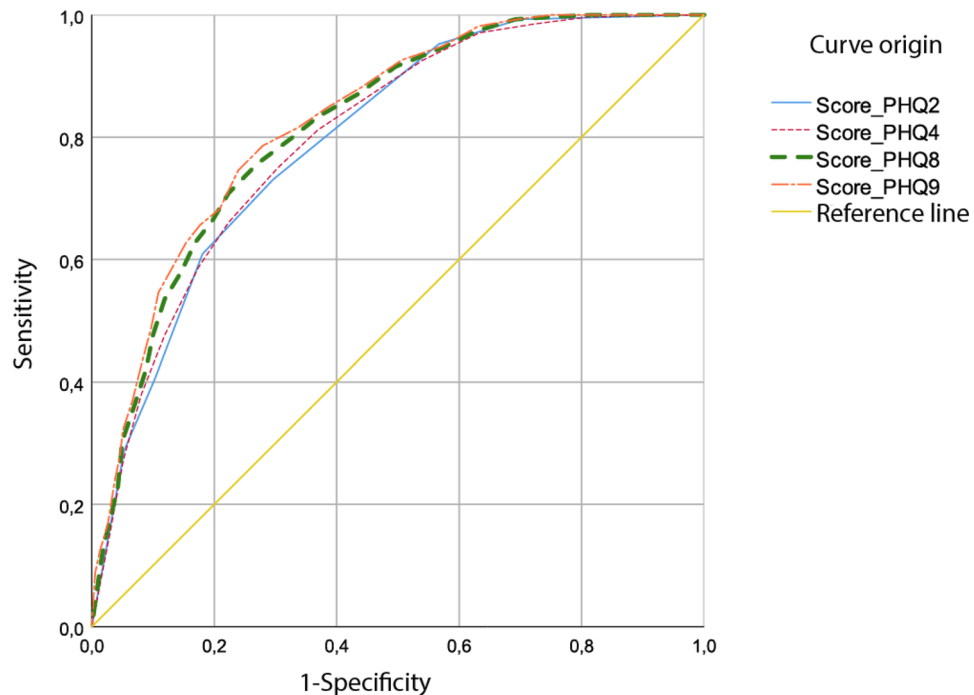


Figure 1. ROC curve for the different versions of PHQ regarding the diagnosis of depression.

Type	Version	Score	J Youden 's index	Closest top left	Sensitivity (%)	1-Specificity (%)
General	PHQ2	2.5	0.434	0.400	73.06	29.63
	PHQ4	5.5	0.446	0.393	74.90	30.30
	PHQ8	10.5	0.486	0.364	75.65	27.05
	PHQ9	<i>10.5</i>	<i>0.507</i>	0.352	78.60	27.95
	PHQ9	11.5	0.506	<i>0.349</i>	74.54	23.91
Male	PHQ2	2.5	0.472	0.374	75.00	27.80
	PHQ4	4.5	0.485	0.371	79.20	30.70
	PHQ8	8.5	0.516	0.342	76.70	25.10
	PHQ9	10.5	0.551	0.328	71.70	16.60
	PHQ9	11.5	0.512	0.346	77.30	26.10
Female	PHQ2	2.5	0.424	0.408	72.51	30.11
	PHQ4	5.5	0.450	0.394	76.80	31.80
	PHQ8	11.5	0.496	0.357	73.90	24.30
	PHQ9	11.5	0.512	0.346	77.30	26.10

Table 5. Selected sensitivity and specificity based on perceived depression. Discrepant values were obtained for PHQ-9 test for indices indicating optimal cut-off point. If Youden's J prevails over Closest Top Left, optimal cut-off point for PHQ-9 is ≥ 10 . The italics indicates the selected index and its associate score: the highest J Youden's Index and the smallest Closest top left index.

Discussion

The PHQ scale demonstrates adequate internal consistency (with an range of 0.853 to 0.929) for different PHQ versions analyzed. These values are like those found in other studies^{26,60,81}. All versions studied exhibit concurrent validity with positive and high correlations for GAD-7 (ranging from 0.767 to 0.931), positive but slight for FCV-19S (ranging from 0.246 to 0.338), and negative but moderate for RS-14 (ranging from -0.620 to -0.647), all of which are consistent with expectations.

The PHQ-9 exhibits high inter-item correlations like those found in other studies (ranging from 0.416 to 0.756). As expected, the suicide item has lower but still significant correlations (ranging from 0.416 to 0.552)⁸².

As an indicator of internal validity, a single factor was obtained in all PHQ-9 variants^{63–65} including the PHQ-4, which also yielded a single factor despite comprising two different theoretical dimensions⁶⁶.

The thresholds for the general population coincide with those proposed in its original version. To assess depression in men, cut-off points have been lowered for both PHQ-4 and PHQ-8. In contrast, to assess depression

Group	Version	Non major depressive disorder	Major depressive disorder	None to minimal	Mild	Moderate	Moderately severe	Severe
Original	PHQ2	0–2	3–6	–	–	–	–	–
	PHQ4	–	–	0–2	3–5	6–8	–	9–12
	PHQ8	–	–	0–4	5–9	10–14	15–19	20–24
	PHQ9	–	–	0–4	5–9	10–14	15–19	20–27
General population	PHQ2	0–2	3–6	–	–	–	–	–
	PHQ4	–	–	0–2	3–5	6–8	–	9–12
	PHQ8	–	–	0–5	6–10	11–15	16–20	21–24
	PHQ9	–	–	0–5	6–10	11–15	16–20	21–27
Male	PHQ2	0–2	3–6	–	–	–	–	–
	PHQ4	–	–	0–1	2–4	5–7	–	8–12
	PHQ8	–	–	0–3	4–8	9–13	14–18	19–24
	PHQ9	–	–	0–5	6–10	11–15	16–20	21–27
Female	PHQ2	0–2	3–6	–	–	–	–	–
	PHQ4	–	–	0–2	3–5	6–8	–	9–12
	PHQ8	–	–	0–6	7–11	12–16	17–21	22–24
	PHQ9	–	–	0–6	7–11	12–16	17–21	22–27

Table 6. Cut-off points for the severity of depressive symptoms of the PHQ-9 versions in the Costa Rican general population, its original measure and new cut-off points split by sex.

in women, original thresholds for PHQ-2 and PHQ-4 have been maintained; however, cut-off points have been increased for both PHQ-8 and PHQ-9. If cut-off points from this study were not applied, men would be underdiagnosed, and women would be overdiagnosed.

Since scores from PHQ-8 are assimilated to those of PHQ-9 for convenience in its original article, this alters thresholds of PHQ-8 in practice leaving less scope for extreme scores. Although new cut-off points have been established to maintain the nature of PHQ-8, it continues to have less range in the severe part of depression.

The diagnostic validity of PHQ-9 with a sensitivity of 78.60% and a specificity of 72.05% presents lower but adequate indices compared to its versions in English (88% sensitivity and 88% specificity)²⁶, Spanish (87% sensitivity and 88% specificity)^{35,49}, Mexico (for PHQ-2) 80.0 and 86.9⁸³.

Despite having only two items, PHQ-2 can detect depression with slightly lower sensitivity and specificity values. Considering the reduction in items, it maintains excellent psychometric properties compared to its longer version. Since PHQ-4 also assesses anxiety, it was compared with the anxiety-depression construct, obtaining a ROC curve like that obtained by comparing it only with depression.

A limitation is although both gender and sex of each patient were assessed, a sample of people born with a gender other than the one identified was too small to form a representative group. Another limitation is the race or ethnicity of participants was not assessed. A large number of discarded tests could be not considered a limitation due to the sample size is being sufficient and data cleaning meeting valid criteria. An explanation for missing data is that this research was published on TV, radio, newspaper, and institutional web pages. The health minister and other relevant public figures asked the population to fill out the questionnaires. That phenomenon made most of the population enter the platform just to satisfy their curiosity. This could be a limitation to not retaining most of the participants to fill all the survey.

Although our sample is representative and heterogeneous in terms of age, sex, academic level, and marital status, we acknowledge that no formal approach was followed to ensure this. This is a cross-sectional study, so a future longitudinal follow-up could be more valuable. Convenience sampling represents a weakness in the generalization of the findings of the present study. Another sampling limitation is that in the country the devices for accessing the internet are not well distributed⁵², and because of this, the population who are without internet access couldn't be contacted. Despite these limitations, this study provides valuable information on the usefulness of the application of the PHQ-9 within the Costa Rican population.

Conclusions

Due to the ability to detect depression, brevity of application, and self-administered application, use of PHQ-2, PHQ-4, PHQ-8 and PHQ-9 tests are considered adequate tools for detecting depression and its severity in the general Costa Rican population. Due to the high consistency of this test, its application is considered valid in the entire Spanish-speaking community.

Proposed thresholds improve the detection of depression based on gender.

Data availability

Data are available from the correspondence author (Armando González-Sánchez. armando_gonzalez@usal.es), upon request. Data are not publicly available due to ethical issues.

Received: 13 January 2023; Accepted: 28 August 2023

Published online: 31 August 2023

References

- American Psychiatric Association. Diagnostic and statistical manual of mental disorders. *Am. Psychiatr. Assoc. Publ.* <https://doi.org/10.1176/appi.books.9780890425787> (2022).
- Wang, J., Xue, J., Jiang, Y., Zhu, T. & Chen, S. Mediating effects of depressive symptoms on social support and quality of life among rural older Chinese. *Health Qual. Life Outcomes* **18**(1), 242. <https://doi.org/10.1186/s12955-020-01490-1> (2020).
- Pan American Health Organization. The Burden of mental disorders in the region of the Americas, 2018. Washington D.C., (2018).
- Solmi, M. *et al.* Age at onset of mental disorders worldwide: Large-scale meta-analysis of 192 epidemiological studies. *Mol. Psychiatry* **27**(1), 281–295. <https://doi.org/10.1038/s41380-021-01161-7> (2022).
- Patel, V. *et al.* Income inequality and depression: A systematic review and meta-analysis of the association and a scoping review of mechanisms. *World Psychiatry* **17**(1), 76–89. <https://doi.org/10.1002/wps.20492> (2018).
- Farhane-Medina, N. Z., Luque, B., Taberner, C. & Castillo-Mayén, R. Factors associated with gender and sex differences in anxiety prevalence and comorbidity: A systematic review. *Sci. Prog.* **105**(4), 003685042211354. <https://doi.org/10.1177/00368504221135469> (2022).
- Carazo-Vargas, E., Ortega Moreno, R., Arias-LeClarire, H., González-García, N., González-Sánchez, A. & Villegas, G. Mental health and relationships with the environment in times of COVID-19. *Eredia*, (2021). [Online]. Available: <https://investiga.uned.ac.cr/wp-content/uploads/2021/01/INFORME-Salud-mental-en-tiempos-de-COVID-19.pdf>
- McCarron, R. M., Shapiro, B., Rawles, J. & Luo, J. Depression. *Ann. Intern. Med.* **174**(5), ITC65–ITC80. <https://doi.org/10.7326/AITC202105180> (2021).
- GBD 2019 Mental Disorders Collaborators. (2022) 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. *Lancet Psychiatry*, **9** (2), 137–150. [https://doi.org/10.1016/S2215-0366\(21\)00395-3](https://doi.org/10.1016/S2215-0366(21)00395-3).
- Pan American Health Organization. PAHO. The burden of suicide in the Region of the Americas. (2021). <https://www.paho.org/en/enlace/burden-suicide>
- Harmer, B., Lee, S., Vi, H., Duong, T. & Saadabadi, A. *Suicidal Ideation*. (2022).
- Morssinkhof, M. W. L. *et al.* Associations between sex hormones, sleep problems and depression: A systematic review. *Neurosci. Biobehav. Rev.* **118**, 669–680. <https://doi.org/10.1016/j.neubiorev.2020.08.006> (2020).
- Slotnick, S. D. Sex differences in the brain. *Cogn. Neurosci.* **12**(3–4), 103–105. <https://doi.org/10.1080/17588928.2021.1957808> (2021).
- Herrman, H. *et al.* Time for united action on depression: A Lancet-World Psychiatric Association Commission. *The Lancet* **399**(10328), 957–1022. [https://doi.org/10.1016/S0140-6736\(21\)02141-3](https://doi.org/10.1016/S0140-6736(21)02141-3) (2022).
- Yildirim, M., Gaynes, B. N., Keskinocak, P., Pence, B. W. & Swann, J. The cost-effectiveness of depression screening for the general adult population. *J. Affect. Disord.* **303**, 306–314. <https://doi.org/10.1016/j.jad.2022.02.044> (2022).
- Zhang, W. *et al.* Depression and psychological-behavioral responses among the general public in China during the early stages of the COVID-19 pandemic: Survey study. *J. Med. Internet Res.* **22**(9), e22227. <https://doi.org/10.2196/22227> (2020).
- Chauhan, N. A wrong diagnosis. *Indian J. Cancer* **57**(2), 212–215. https://doi.org/10.4103/ijc.IJC_418_19 (2020).
- Valença, A. M. *et al.* Case report: Depression × dementia with lewy bodies in the elderly: The importance of differential diagnosis. *Front. Psychiatry* **13**, 1059150. <https://doi.org/10.3389/fpsy.2022.1059150> (2022).
- Rodziewicz, T. L., Houseman, B. & Hipskind, J. E. in *Medical Error Reduction and Prevention*. 2023. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/23343656>
- Dubovsky, S. L., Ghosh, B. M., Serotte, J. C. & Cranwell, V. Psychotic depression: Diagnosis, differential diagnosis, and treatment. *Psychother. Psychosom.* **90**(3), 160–177. <https://doi.org/10.1159/000511348> (2021).
- de Souza, A. C. *et al.* Psychometric properties in instruments evaluation of reliability and validity. *Epidemiologia e Serviços de Saúde* **26**(3), 649–659. <https://doi.org/10.5123/S1679-49742017000300022> (2017).
- Holgado-Tello, F. P., Chacón-Moscoso, S., Sanduvete-Chaves, S. & Lozano-Lozano, J. A. Editorial: New challenges in globalized societies: Cross-cultural studies and test adaptation. *Front Psychol* **13**, 900535. <https://doi.org/10.3389/fpsy.2022.900535> (2022).
- Sánchez-Villena, A. R., de La Fuente-Figuerola, V. & Ventura-León, J. Equivalence of measurement: On life before multigroup confirmatory factor analysis. *Actas Dermosifiliogr.* **112**(5), 484–485. <https://doi.org/10.1016/j.ad.2020.06.005> (2021).
- Spitzer, R. L., Williams, J. B. & Kroenke, K. Test review: Patient Health Questionnaire–9 (PHQ-9). *Rehabil. Couns. Bull.* **57**(4), 246–248. <https://doi.org/10.1177/0034355213515305> (2014).
- Ong, C. W. *et al.* Longitudinal measurement invariance of the PHQ-9 and GAD-7. *Assessment* **29**(8), 1901–1916. <https://doi.org/10.1177/10731911211035833> (2022).
- Kroenke, K., Spitzer, R. L. & Williams, J. B. W. The PHQ-9 validity of a brief depression severity measure. *J. Gen. Intern. Med.* **16**, 606–613. <https://doi.org/10.1046/j.1525-1497.2001.016009606.x> (2001).
- Brattmyr, M., Lindberg, M. S., Solem, S., Hjemdal, O. & Havnen, A. Factor structure, measurement invariance, and concurrent validity of the Patient Health Questionnaire-9 and the generalized anxiety disorder scale-7 in a Norwegian psychiatric outpatient sample. *BMC* **22**(1), 461. <https://doi.org/10.1186/s12888-022-04101-z> (2022).
- Zhou, Y., Xu, J. & Rief, W. Are comparisons of mental disorders between Chinese and German students possible? An examination of measurement invariance for the PHQ-15, PHQ-9 and GAD-7. *BMC Psychiatry* **20**(1), 480. <https://doi.org/10.1186/s12888-020-02859-8> (2020).
- Siewe Fodjo, J. N. *et al.* Fear and depression during the COVID-19 outbreak in Cameroon: A nation-wide observational study. *BMC Psychiatry* **21**(1), 356. <https://doi.org/10.1186/s12888-021-03323-x> (2021).
- Yuan, L.-L. *et al.* Prevalence and predictors of anxiety and depressive symptoms among international medical students in China during COVID-19 pandemic. *Front. Psychiatry* **12**, 761964. <https://doi.org/10.3389/fpsy.2021.761964> (2021).
- Pappa, S., Barnett, J., Berges, I. & Sakkas, N. Tired, worried and burned out, but still resilient: A cross-sectional study of mental health workers in the UK during the COVID-19 pandemic. *Int. J. Environ. Res. Public Health* **18**(9), 4457. <https://doi.org/10.3390/ijerph18094457> (2021).
- Nallusamy, V., Afgarshe, M. & Shlosser, H. Reliability and validity of somali version of the PHQ-9 in primary care practice. *Int. J. Psychiatry Med.* **51**(6), 508–520. <https://doi.org/10.1177/0091217417696732> (2016).
- Inagaki, M. *et al.* Validity of the Patient Health Questionnaire (PHQ)-9 and PHQ-2 in general internal medicine primary care at a Japanese rural hospital: A cross-sectional study. *Gen. Hosp. Psychiatry* **35**(6), 592–597. <https://doi.org/10.1016/j.genhosppsy.2013.08.001> (2013).
- van Dijk, M. J., de Man-van Ginkel, J. M., Hafsteinsdóttir, T. B. & Schuurmans, M. J. Psychometric properties of the dutch version of the signs of depression scale. *Clin. Nurs. Res.* **27**(5), 617–637. <https://doi.org/10.1177/1054773816657798> (2018).
- Muñoz-Navarro, R. *et al.* Utility of the PHQ-9 to identify major depressive disorder in adult patients in Spanish primary care centres. *BMC Psychiatry* **17**(1), 1–9. <https://doi.org/10.1186/s12888-017-1450-8> (2017).

36. Baños-Chaparro, J., Ynquillay-Lima, P., Lamas Delgado, F. & Gaby Fuster Guillen, F. Psychometric evidence of the Patient Health Questionnaire-2 for the assessment of depressive symptoms in Peruvian adults. [Evidencias psicométricas del Patient Health Questionnaire-2 para la evaluación de síntomas depresivos en adultos peruanos]. *Rev Ciencias Médicas*, vol. 25 (5), 5064 (2021).
37. Calderón, M. *et al.* Validation of the Peruvian version of the PHQ-9 for diagnosing depression [Validación de la versión peruana del PHQ-9 para el diagnóstico de depresión]. *Rev. Peru. Med. Exp. Salud Pública* 29(4), 578–585 (2012).
38. Saldívar, S. *et al.* Propiedades psicométricas del PHQ-9 (Patient Health Questionnaire) en centros de atención primaria de Chile. *Rev. Med. Chil.* 147(1), 53–60. <https://doi.org/10.4067/S0034-98872019000100053> (2019).
39. Borghero, F. *et al.* Screening for depressive episode in adolescents. Validation of the PHQ-9 instrument. [Tamizaje de episodio depresivo en adolescentes. Validación del instrumento PHQ-9]. *Rev. Med. Chil.* 146(4), 479–486. <https://doi.org/10.4067/S0034-98872018000400479> (2018).
40. Shevlin, M. *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* 22(1), 154. <https://doi.org/10.1186/s12888-022-03787-5> (2022).
41. Paniagua-Ávila, A. *et al.* The mental health of health workers in Guatemala during the COVID-19 pandemic: baseline results from the HÉROES cohort study. [La salud mental de los trabajadores de salud en Guatemala durante la pandemia de COVID-19: resultados de base del estudio de cohortes HÉROES]. *Rev. Panam. Salud Pública* 46, 1. <https://doi.org/10.26633/RPSP2022.79> (2022).
42. Caycho-Rodríguez, T. *et al.* Translation into Spanish and validation of a brief measure of anxiety due to COVID-19 in Health Sciences students. [Traducción al español y validación de una medida breve de ansiedad por la COVID-19 en estudiantes de ciencias de la salud]. *Ansiedad y Estrés* 26(2–3), 174–180. <https://doi.org/10.1016/j.anyes.2020.08.001> (2020).
43. Camargo, L. *et al.* Validity of the PHQ-2 questionnaire for the detection of depression in Colombian health personnel during Covid-19. [Validez del cuestionario PHQ-2 para la detección de depresión en personal sanitario colombiano durante Covid-19]. *Rev Chil Neuropsiquiatr* 60(3), 281–288. <https://doi.org/10.4067/s0717-92272022000300281> (2022).
44. Sequeira-Cordero, A., Blanco-Vargas, M. & Brenes, J. C. Major depression in Costa Rica: epidemiological and psychobiological aspects. [La depresión mayor en Costa Rica: aspectos epidemiológicos y psicobiológicos]. 19 (2), 465–489. (2022). Accessed 01 Aug 2023. [Online]. Available: http://www.scielo.sa.cr/scielo.php?script=sci_arttext&pid=S1659-02012022000100465&lng=en
45. Instituto Nacional de Estadística y Censos (INEC). Population estimates and projections. [Estimaciones y proyecciones de población]. San José, (2021). Accessed 01 Aug 2023. [Online]. Available: <https://www.inec.cr/poblacion/estimaciones-y-proyecciones-de-poblacion>
46. Rosero-Bixby, L., Modrek, S., Domino, M. E. & Dow, W. H. Aging and mental health in a longitudinal study of elderly Costa Ricans. In *Encyclopedia of Geropsychology*, 1–11 (Springer Singapore, Singapore, 2015). https://doi.org/10.1007/978-981-287-080-3_37-1
47. Kessler, R. C. *et al.* Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the national comorbidity survey replication. *Arch. Gen. Psychiatry* 62(6), 593. <https://doi.org/10.1001/archpsyc.62.6.593> (2005).
48. 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.* 276, 220–233. <https://doi.org/10.1016/j.jad.2020.06.066> (2020).
49. González-Blanch, C. *et al.* Factor structure and measurement invariance across various demographic groups and over time for the PHQ-9 in primary care patients in Spain. *PLoS ONE* 13(2), 1–16. <https://doi.org/10.1371/journal.pone.0193356> (2018).
50. Helliwell, J. F., Layard, R., Sachs, J. D., De Neve, J.-E., Aknin, L. B. & Wang, S. World Happiness Report 2023. New York, 2023. Accessed 04 Aug 2023. [Online]. Available: <http://worldhappiness.report/>
51. Sharman, C. H. Improving mental-health care in Costa Rica. *The Lancet* 366(9480), 111–112. [https://doi.org/10.1016/S0140-6736\(05\)66849-3](https://doi.org/10.1016/S0140-6736(05)66849-3) (2005).
52. Instituto Nacional de Estadística y Censos. INEC-Costa Rica. Encuesta Nacional de Hogares, 2020. San José, 2020. Accessed 28 Jul 2023. [Online]. Available: <https://inec.cr/estadisticas-fuentes/encuestas/encuesta-nacional-hogares?filtertext=encuesta%2520nacional%2520de%2520hogares%25202020>
53. Organization for Economic Cooperation and Development (OECD), *Education at a Glance 2022*. (OECD, 2022). <https://doi.org/10.1787/3197152b-en>
54. Instituto Nacional de Estadística y Censos, “Census. 2011. Volume I: Social and demographic characteristics. Part II. [Censo. 2011. Tomo I: Características sociales y demográficas. Parte II],” San José, Jun. 2016. Accessed: Aug. 01, 2023. [Online]. Available: https://admin.inec.cr/sites/default/files/media/replaccenso2011-13_3.pdf
55. M. A. Quesada Pacheco, *New dictionary of Costa Ricanisms*. [Nuevo diccionario de costarriqueñismos]. 2007.
56. Buzek, I. Carlos Gagini’s dictionaries as sources of gypsy studies in Costa Rican Spanish at the end of the 19th century and beginning of the 20th. [Los diccionarios de Carlos Gagini como fuentes de estudios de gitanismos en el español de Costa Rica a finales de siglo XIX y principios del XX]. In *Diversidad lingüística del español*, (eds Valeš, M. & Miča S.), 45–67 (Universidad Técnica de Liberec, Liberec, 2013).
57. Constenla Umaña, A. The linguistic diversity of Costa Rica: Indigenous languages. [La diversidad lingüística de Costa Rica: Las lenguas indígenas]. *Revista de Filología y Lingüística de la Universidad de Costa Rica*, 37 (2), (2013).
58. Soper, D. S. A-priori sample size calculator for structural equation models. 2022. Accessed: Dec. 15, 2022. [Online]. Available: <https://www.danielsoper.com/statcalc>
59. Z. F. Negeri *et al.*, Accuracy of the Patient Health Questionnaire-9 for screening to detect major depression: Updated systematic review and individual participant data meta-analysis. *The BMJ*, 375 (BMJ Publishing Group, 2021). <https://doi.org/10.1136/bmj.n2183>
60. Diez-Quevedo, C., Rangil, T., Sánchez-Planell, L., Kroenke, K. & Spitzer, R. L. Validation and utility of the Patient Health Questionnaire in diagnosing mental disorders in 1003 general hospital Spanish inpatients. *Psychosom. Med.* 63(4), 679–686 (2001).
61. Kroenke, K., Spitzer, R. L. & Williams, J. B. W. The Patient Health Questionnaire-2. *Med. Care* 41(11), 1284–1292. <https://doi.org/10.1097/01.MLR.0000093487.78664.3C> (2003).
62. Kroenke, K., Spitzer, R. L., Williams, J. B. W., Monahan, P. O. & Löwe, B. Anxiety disorders in primary care: Prevalence, impairment, comorbidity, and detection. *Ann. Intern. Med.* 146(5), 317. <https://doi.org/10.7326/0003-4819-146-5-200703060-00004> (2007).
63. Mitchell, A. J., 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* 2(2), 127–138. <https://doi.org/10.1192/bjpo.bp.115.001685> (2016).
64. Richardson, L. P. *et al.* Evaluation of the PHQ-2 as a brief screen for detecting major depression among adolescents. *Pediatrics* 125(5), e1097–e1103. <https://doi.org/10.1542/peds.2009-2712> (2010).
65. Kroenke, K. *et al.* The PHQ-8 as a measure of current depression in the general population. *J. Affect. Disord.* 114(1–3), 163–173. <https://doi.org/10.1016/j.jad.2008.06.026> (2009).
66. Kroenke, K., Spitzer, R. L., Williams, J. B. W. & Löwe, B. An ultra-brief screening scale for anxiety and depression: The PHQ-4. *Psychosomatics* 50(6), 613–621. <https://doi.org/10.1176/appi.psy.50.6.613> (2009).
67. Spitzer, R. L., Kroenke, K., Williams, J. B. W. & Löwe, B. A brief measure for assessing generalized anxiety disorder: The GAD-7. *Arch. Intern. Med.* 166(10), 1092–1097. <https://doi.org/10.1001/archinte.166.10.1092> (2006).
68. Johnson, S. U., Ulvenes, P. G., Økstedalen, T. & Hoffart, A. Psychometric properties of the GAD-7 in a heterogeneous psychiatric sample. *Front. Psychol.* 10, 1–8. <https://doi.org/10.3389/fpsyg.2019.01713> (2019).
69. 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* 39, 24–31. <https://doi.org/10.1016/j.genhosppsy.2015.11.005> (2016).

70. Wagnild, G. M. & Collins, J. A. Assessing resilience. *J. Psychosoc. Nurs. Ment. Health Serv.* **47**(12), 28–33. <https://doi.org/10.3928/02793695-20091103-01> (2009).
71. Ahorsu, D. K. *et al.* The fear of COVID-19 Scale: Development and initial validation. *Int. J. Ment. Health Addict.* **20**(3), 1537–1545. <https://doi.org/10.1007/s11469-020-00270-8> (2022).
72. Martínez-Lorca, M., Martínez-Lorca, A., Criado-Álvarez, J. J., Armesilla, M. D. C. & Latorre, J. M. The fear of COVID-19 scale: Validation in Spanish university students. *Psychiatry Res.* **293**, 113350. <https://doi.org/10.1016/j.psychres.2020.113350> (2020).
73. I. SPSS Corp., IBM SPSS Statistics for Windows v.25. (IBM Corp., Armonk, NY, 2017).
74. Lorenzo-Seva, U. & Ferrando, P. J. FACTOR: A computer program to fit the exploratory factor analysis model. *Behav. Res. Methods* **38**(1), 88–91. <https://doi.org/10.3758/BF03192753> (2006).
75. Ferrando, P. J., Lorenzo-Seva, U., Hernández-Dorado, A. & Muñiz, J. Decalogue for the factor analysis of test items. *Psicothema* **34**(1), 7–17. <https://doi.org/10.7334/psicothema2021.456> (2022).
76. Timmerman, M. E. & Lorenzo-Seva, U. Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychol. Methods* **16**(2), 209–220. <https://doi.org/10.1037/a0023353> (2011).
77. Lorenzo-Seva, U. SOLOMON: A method for splitting a sample into equivalent subsamples in factor analysis. *Behav. Res. Methods* <https://doi.org/10.3758/s13428-021-01750-y> (2021).
78. Hayes, A. F. & Coutts, J. J. Use omega rather than Cronbach's Alpha for estimating reliability. *But.... Commun. Methods Meas.* **00**(00), 1–24. <https://doi.org/10.1080/19312458.2020.1718629> (2020).
79. Arbuckle, J. L. Amos (Version 4.0). IBM SPSS, Chicago, 2016. Accessed: Mar. 09, 2023. [Online]. Available: <http://amosdevelopment.com/index.html>
80. Perkins, N. J. & Schisterman, E. F. The inconsistency of 'optimal' cutpoints obtained using two criteria based on the receiver operating characteristic curve. *Am J Epidemiol* **163**(7), 670–675. <https://doi.org/10.1093/aje/kwj063> (2006).
81. Baader, T. *et al.* Validity and utility of PHQ9 (Patient Health Questionnaire) in the diagnosis of depression in user patients of primary care in Chile. *Rev. Chil. Neuro-Psiquiat* **50**(1), 10–22 (2012).
82. Pranckeviciene, A. *et al.* Validation of the Patient Health Questionnaire- 9 and the generalized anxiety disorder-7 in Lithuanian student sample. *PLoS ONE* **17**(1), e026327. <https://doi.org/10.1371/journal.pone.0263027> (2022).
83. Arrieta, J. *et al.* Validity and utility of the Patient Health Questionnaire (PHQ)-2 and PHQ-9 for screening and diagnosis of depression in rural Chiapas, Mexico: A cross-sectional study. *J. Clin. Psychol.* **73**(9), 1076–1090. <https://doi.org/10.1002/jclp.22390> (2017).

Acknowledgements

Special thanks to the Ministry of Health and the National Health Research Council (CONIS).

Author contributions

A.G., R.O., G.V.B., E.C., and H.A. design of the study, A.G. formal analysis, and data analysis; A.G. and R.O., writing and preparation of the original draft; A.G., R.O., and P.V. writing, revising and editing. All authors reviewed the manuscript.

Funding

No funding has been received for this study.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-41560-0>.

Correspondence and requests for materials should be addressed to A.G.-S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023, corrected publication 2024