Measurement of depression in the Brazilian population: validation of the *Patient Health Questionnaire* (PHQ-8)

Mensuração da depressão na população brasileira: validação do *Questionário de Saúde do Paciente* (PHQ-8)

Medición de la depresión en la población brasileña: validación del *Cuestionario de Salud del Paciente* (PHQ-8)

Iracema Lua [1]
Katia Santana Freitas [2]
Jules Ramon Brito Teixeira [2]
Michael Eduardo Reichenheim [3]
Maura Maria Guimarães de Almeida [2]
Tânia Maria de Araújo [2]

Abstract

We aimed to evaluate the psychometric properties of the Brazilian version of the Patient Health Questionnaire *(PHQ-8). A study with a sample of 4,170 individuals (≥ 15 years old) from the urban area. Conglomerate sampling was adopted in two stages (census sectors and streets), with weighting of estimates by sample weights. A structured questionnaire with sociodemographic data, the PHQ – the modules for depression, generalized anxiety disorder and panic disorder – and the* Self-Reporting Questionnaire *(SRQ-20) were used. In the evaluation of the PHQ-8, we verified the construct validity by analyzing the dimensional structure, convergent validity and internal consistency. We found a linear disorder without losses to maintain the four response categories. The factor analysis found unidimensionality of the depression construct, with strong factor loads, low residual variances, low residual correlation between items, good fit of the model, internal consistency and satisfactory convergent factorial validity (high loads and correlations with other tests/scales of similar constructs). The PHQ-8 has a one-dimensional structure with evidence of good validity and reliability, being suitable for use in the Brazilian population.*

Depression; Questionnaires; Validation Study

Correspondence
I. Lua
Av. Artemia Pires Freitas 10297, Feira de Santana, BA 44071-010, Brasil.
ira_lua@hotmail.com

[1] *Universidade Federal da Bahia, Salvador, Brasil.*
[2] *Universidade Estadual de Feira de Santana, Feira de Santana, Brasil.*
[3] *Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brasil.*

Introduction

Depression is a highly prevalent disease worldwide [1,2], with increasing trend in recent years [3]. Brazil is among the five countries with the highest rates in the world. Depression is a treatable mental disorder characterized by the presence of affective symptoms such as sadness, empty or irritable mood, accompanied or not by somatic and cognitive changes, which affect the individual's functional ability [5]. Early identification and treatment improve the prognosis and are essential to cope with this condition and to prevent the occurrence of new episodes [6].

Brazilian studies identifying mental disorders in general and depression are scarce. The few existing studies focus on groups with depression [4,7]. Thus, broader analyses with active evaluation of depression in apparently healthy populations are incipient, and information about the real dimension of mental disorders in the general population is scarce. This gap in the Brazilian context lacks data, but a study in the United States showed that less than 5% of adults are screened for depression in primary health care services [8], which is worsened by the lack of instruments with good psychometric performance and suitable for screening and diagnosis [9,10].

The difficulties in screening and diagnosis hinder the monitoring of mental disorders, although several evaluation instruments are available [11]. Incomplete assessments of the performance of these instruments or weaknesses and gaps observed in their assessments constitute barriers to the effective incorporation of health care routines and clinical practices in general, making it impossible to advance in the early diagnosis and intervention of the disease and the consequent reduction in health expenditures.

The improvement in access to mental health services and the routine use of related diagnostic tools that are reliable, easy to handle and apply, with good performance, are challenges in this field [8]. Investing efforts in the assessment of instruments that can increase safety in their use is essential.

The scientific literature describes at least 16 instruments for detection of depression used in primary health care [11], with better psychometric indicators identified for the *Beck Depression Inventory* (BDI), *Center for Epidemiologic Studies Depression Scale* (CES-D), *Medical Outcomes Study Depression Scale* (MOSD), *Primary Care Evaluation of Mental Disorders* (PRIME-MD), *Symptom-Driven Diagnostic System for Primary Care* (SDDS-PC), *General Health Questionnaire* (GHQ), *Hopkins Symptom Checklist* (HSCL), and *Patient Health Questionnaire* (PHQ).

The PHQ is used worldwide in many contexts and translated into several languages [8,12,13,14]. An instrument of easy and quick application, with less items and satisfactory estimates of sensitivity and specificity [13,15].

The PHQ is a short, self-administered version of PRIME-MD [16], based on criteria for m*ajor depressive disorder described in the Diagnostic and Statistical Manual of Mental Disorders* (DSM), organized in modules for measuring specific psychological disorders. The module depression assessment originally included nine items (PHQ-9), but subsequent assessments recommended its reduction to an eight-item version, PHQ-8 [12].

To justify the exclusion of the ninth item proposed in the PHQ-9, "Thinking about getting hurt in some way or that it would be better to be dead", experts suggest that this change does not influence the diagnosis of the condition, given that: (a) thought of self-mutilation is the last item to be endorsed in the diagnosis of depression because they are uncommon in the general population [10,12,17]; (b) in the vvalidation studies of the PHQ-9, the item does not accurately discriminate the presence or absence of depression, presenting a lower factor loading [18,19], sometimes with cross-loadings on other constructs, such as anxiety [20]; (c) its omission does not significantly alter the instrument's sensitivity and specificity [12,15,21,22,23], nor the reliability indicators [18,19], with evidence that the measures produced by the 8- or 9-item version are similar, with high correlation indicators between them [13,18,21,22].

Besides, the exclusion of this ninth item, related to suicidal ideation, has been suggested in population-based epidemiological studies holding low risk of severe suicidal ideation, that is, in populations with no indication of previous mental illness. Given the rather sensitive nature of the item, its exclusion has also been indicated in contexts of insufficient financial resources, where mental health specialists are likely to be unavailable [12]. These issues strengthen the relevance of using the PHQ-8 to measure depression in the Brazilian context in population-based epidemiological studies.

The PHQ-8 has been both self-administered [24,25] and applied by an interviewer [26]. Validation studies have found adequate psychometric indicators that support its use [12,21,23,27,28,29,30,31,32,33,34].

Despite the metric qualities of PHQ-8 in the international context, its psychometric properties have not yet been evaluated in the Brazilian context. Valid and reliable instruments are necessary for screening and clinical confirmation and can be used in the routines of health care services and in studies with large samples. Thus, this study aimed to evaluate the psychometric properties of the Brazilian version of the PHQ-8 in an urban population.

Method

Study design and sampling

This study is an offshoot of the project *Surveillance in Mental Health and Work: a Cohort of the Population of Feira de Santana, Bahia, Brazil* steered by the Epidemiology Center (NEPI) of the State University of Feira de Santana (UEFS).

Data concerns a probabilistic sample of a population aged ≥ 15 years, living in urban area of Feira de Santana in 2007, based on face-to-face interviews in which a structured questionnaire was applied. In the original study, with the aim of estimating the prevalence of mental disorders in urban populations, a sample size of 1,868 participants was estimated, considering the urban population aged ≥ 15 years (N = 422,282), prevalence of common mental disorders (CMD) of 24%[35], 95% confidence interval (95%CI), 3% accuracy, design effect study 2 and an increase of 20% predicting possible refusals and losses. A total of 4,170 individuals were interviewed, enabling the estimation of a model with up to 83 parameters (50 observations per parameter).

Participants were selected through a complex sampling procedure comprising two clustering stages (census tracts and streets) and stratified according to subdistricts. The details of these procedures are found elsewhere [36]. The sample weight was estimated to compensate for different selection probabilities at each sampling stage, considering different weightings for each element of the sample.

Data collection instruments

A structured questionnaire containing eight blocks of questions was used in data collection. Information on sociodemographic, economic, and mental health assessment characteristics was used for this study. The questionnaire was applied at the participants' household by a trained team after the participant signed the informed consent form.

The Portuguese adapted version of PHQ-8 was used to evaluate depression. Depressive symptoms are assessed considering the two-week recall (Box 1).

The literature [37,38] and previous application of the instrument indicated that the temporal gradient of the response categories of the original version was unable to elucidate the respondent's adequate positioning to the measured item (e.g., "several days" is similar to "more than half the days"). The research team organized workshops with mental health and psychometrics experts to discuss and propose changes. The changes aimed to discriminate the frequency of events, respecting the meanings of the original proposition. Thus, we changed the response category "1" to "a few days." In this way, the response categories used were: "0" (none); "1" (a few days), "2" (more than half the days); "3" (almost every day). The pre-test, conducted in the first applications of the instrument, showed a better understanding of the response categories and adequate distinction of the frequency of the event.

Besides PHQ-8, other mental health outcomes were assessed to offer parameters for comparison with depression results.

CMD were measured by the *Self-Reporting Questionnaire* (SRQ-20), validated for the Brazilian context, presenting satisfactory performance indicators [39]. This instrument consists of 20 items with dichotomous response categories. A cut-off point of five or more positive responses is adopted for men, and seven or more positive responses for women, for screening for suspicion of CMD [40].

Box 1

Items from the *Patient Health Questionnaire* (PHQ-8) for measuring depression.

ITEM		CONTENT
p1	Little interest or pleasure in doing things.	Anhedonia
p2	Feeling sad, depressed or hopeless.	Depressed mood
p3	Difficulty to fall asleep or stay asleep, or sleeping more than usual.	Sleep problems
p4	Feeling tired or having little energy.	Tiredness or lack of energy
p5	Poor appetite or overeating.	Change in appetite or weight
p6	Feeling bad about yourself; thinking that you are a failure or that you were disappointing yourself or your family.	Guilt or worthlessness
p7	Trouble concentrating on things, such as reading the newspaper or watching television.	Concentration problems
p8	Slowness to move or talk to the point that other people notice, or the opposite, you have been so agitated or restless that you have been moving around a lot more than usual.	Feeling slow or restless

Anxiety disorders were assessed using the 22-item PHQ module, a scale with adequate psychometric performance in other populations [9]. The Brazilian version of the instrument is available free of charge (https://www.phqscreeners.com), which was previously evaluated to ensure its proper use in this population (data not shown in tables). This scale evaluates two types of specific disorders:

(a) Panic disorder: a total of 15 items that assess anxiety and panic attack symptoms with dichotomous response categories (yes or no). Panic disorder is diagnosed in case of positive responses to the four items of anxiety symptoms (presence of anxiety crisis or recurring and unexpected panic in the last month; apprehension or persistent concern about a new attack), associated with at least four of the 11 panic attack symptoms (severe anxiety crisis) [5,13]. The scale showed a one-dimensional structure with factor loadings > 0.58, with adequate adjustment (comparative fit index – CFI and Tucker-Lewis index – TLI > 0.95 and root mean square error of aproximation – RMSEA < 0.05) and adequate reliability (CC = 0.90).

(b) Generalized anxiety disorder: measured by seven items that assess the individual's frequency of anxiety-related disturbances, considering a four-week recall, with three-point Likert-type response categories: (0) "none", (1) "several days", and (2) "more than half the days". The generalized anxiety disorder is identified by the presence of four or more "more than half the days" items and a positive response to "feeling nervous, anxious, tense or very worried" [13]. This module also showed a one-dimensional structure, with loads > 0.47, with adequate fit indices (CFI and TLI > 0.95 and RMSEA = 0.05) and satisfactory reliability (CC = 0.77).

Data analysis

Initially, descriptive statistics, absolute and relative frequencies were used to characterize the sociodemographic profile of the PHQ-8 respondents. The prevalence of suicidal ideation (item 11 of SRQ-20 "thought of ending one's life") was also estimated to confirm the low frequency of this item in the studied population. The SPSS software, version 24.0 (https://www.ibm.com/), was used.

Descriptive analysis was also used to assess the distribution of the response categories of the items in the PHQ-8, using the Mplus software, version 8.4 (https://www.statmodel.com/).

An exploratory analysis was initially implemented in the assessment of the configural structure. Eigenvalues were estimated as criteria for extracting factor quantity [41]. Then, exploratory structural equation models (ESEM) [42] was specified to validate the one-dimensional structure indicated by the eigenvalues. Geomin oblique rotation was adopted.

Then, confirmatory factor analysis (CFA) was employed to evaluate the identified solution [43,44]. These analyses followed the recommendations of the *COnsensus-based Standards for the Selection of Health Measurement INstruments* (COSMIN) [45], as well as other prominent references in the area [41,43,44,46,47].

The item was considered conditionally related to a specific factor when its standardized loading was ≥ 0.5, with a residual ≤ 0.7 [46,47]. Residual correlations were evaluated to check possible semantic redundancies [44]: values between 0.3 and 0.6 suggested reassessment to aggregate their semantic contents; and values ≥ 0.7 indicated the need to remove one of the items [47].

The robust diagonally weighted least squares estimator (weighted least squares mean and variance adjusted – WLSMV) was used, considering that this is an appropriate estimator for polychotomous and ordinal items [47,48]. Model fit was evaluated by the RMSEA, CFI and TLI.

RMSEA < 0.06 suggest a good fit, while values > 0.10 indicate inadequate fit and that the model should be rejected [44]. CFI and TLI compare the target model with a null model. Both vary from zero to one, when ≥ 0.95 indicate an acceptable fit [43]. All analysis accounted for the complex sampling design.

The modification indices (MI) were evaluated for cross-load diagnostics and residual correlations. They indicate how much the model adjustment would improve if a parameter was freely estimated. Expected parameter changes (EPC) were observed to evaluate the direction and intensity of the estimates with the suggested modification [44,47]. Thus, a detailed assessment of the residual correlation and/or re-specification of the model was conducted after identifying MI ≥ 10 and EPC ≥ 0.25 [44].

Convergent factor validity was assessed by: (a) inspection of high factor loading and (b) average variance extracted (AVE) ≥ 0.50 [46,47]. External correlation of the construct "depression" with other instruments that measure theoretically related mental health outcomes was checked [45,49]. Correlations were obtained using Spearman's rank correlation test, due to the lack of normality of the scores (Shapiro-Wilk test), with a criterion of statistical significance of p ≤ 0.05. The analysis also used Stata 15.0 (https://www.stata.com).

Internal consistency was assessed using composite reliability (CR) with values ≥ 0.70 as a criterion for good consistency [41,44,46]. The 95%CI of the CR and AVE were obtained by the bootstrap method with 1,000 replications.

This research was approved by the UEFS Ethics Research Committee, under opinion 2,420,653 (CAAE: 74792617.4.0000.0053).

Results

The study sample consisted of 4,170 people living in the urban area of the studied municipality, mainly women (67.6%), aged up to 40 years (58.2%), with a mean age of 38.94 years (SD = 17.9), black race/skin color (80.7%), single (51.4%) and with up to incomplete elementary school (44.3%). Regarding insertion in the labor market, 59.1% were unemployed and 52.6% had an income of up to one minimum wage (Table 1).

The sample showed low frequency of suicidal ideation. A total of 5.3% indicated "thought of ending one's life" in the last 30 days (data not shown in the table), assessed by the SRQ-20.

The distribution of frequencies in the PHQ-8 response categories showed higher response frequencies in the "none" category (63.4% in p4 to 84.3% in p6) and lower frequencies in the category "more than half the days" (2.5% p6 to 7.4% p3). Items p3 and p5 showed the highest frequency in "almost every day"; on the other hand, item p6 showed the lowest frequency in this category (Table 2).

The results of the EFA showed a dominant dimension, with a great decrease between the first (4.283) and the second (0.709) eigenvalues and small decreases thereafter. Although two models (one-dimensional and two-dimensional) demonstrated an adequate fit, the presence of cross loads between the factors and the diagnosis of the eigenvalues suggest unidimensionality (data not shown in the table). ESEM confirmed the structure (Table 3).

In the CFA, the one-dimensional model of the PHQ-8 showed satisfactory adjustment (CFI = 0.98; TLI = 0.98; RMSEA = 0.03; 90%CI: 0.02-0.03) and internal consistency (CR = 0.88) (Table 3). All factor loadings were higher than 0.50, with lowest loading in items p5 and p8 (λi = 0.61) and highest load-

Table 1

Sociodemographic and economic characteristics of the population from the urban area of Feira de Santana, Bahia State, Brazil, 2007.

Characteristics	n	%
Gender [n = 4,170]		
Women	2,819	67.6
Men	1,351	32.4
Age (years) [n = 4,168]		
Mean (SD)	38.94 (17.9)	
≤ 40	2,426	58.2
> 40	1,742	41.8
Race/Skin color [n = 3,953]		
White/Asian	687	17.4
Black	3,191	80.7
Indigenous	75	1.9
Marital status [n = 4,147]		
With companion	2,015	48.6
No companion	2,132	51.4
Schooling level [n = 4,156]		
Incomplete elementary school	1,842	44.3
Complete primary education	922	22.2
Complete high school	1,304	31.4
Graduated	88	2.1
Income * (minimum wage) [n = 2,385]		
Up to one	1,254	52.6
1-2	714	29.9
More than 2	417	17.5
Work situation [n = 4,167]		
Employed	1,706	40.9
Unemployed	2,461	59.1

* 2007 minimum wage: BRL 380.00.

ing in item p2 (λi = 0.82), besides low residual variance values (δi < 0.70), with the maximum residual identified in items p5 and p8 (δi = 0.63) (Table 3).

The model's diagnosis by MI assessment showed a residual correlation between items p2 (feeling sad, depressed, or hopeless) and p6 (feeling bad about yourself; thinking that you are a failure, or that you were disappointing yourself or your family) (MEP = 0.292). The free estimation of this parameter demonstrated a low residual correlation (0.25) between the items and an insignificant improvement in the adjustment of the model (ΔCFI = 0.004) (Table 3), which indicate the absence of overlap in the content and the decision to keep them in the model. MI did not indicate other changes.

The high factor loadings suggest satisfactory convergent factor validity. However, the AVE (0.48, 95%CI: 0.45-0.51) indicates 48% variance of the items explaining the depression construct (Table 3).

The positive and significant relationships (p < 0.001) of the construct measured by PHQ-8 (depression) with other similar constructs corroborated the convergent validity. CMD showed a strong correlation (r = 0.592), moderate correlation with panic disorder (r = 0.326), and generalized anxiety disorder (r = 0.274) (Table 4).

Table 2

Distribution of frequency of responses (%) of the items in the *Patient Health Questionnaire* (PHQ-8), [N = 4,168].

PHQ-8 item		Response categories (%)			
		None (0)	A few days (1)	More than half the days (2)	Almost every day (3)
p1	Little interest or pleasure in doing things.	72.1	15.1	5.1	7.7
p2	Feeling sad, depressed or hopeless.	68.1	18.2	6.8	6.9
p3	Difficulty to fall asleep or stay asleep, or sleeping more than usual.	64.0	16.1	7.4	12.5
p4	Feeling tired or having little energy.	63.4	21.8	6.9	7.9
p5	Poor appetite or overeating.	70.0	13.9	4.7	11.4
p6	Feeling bad about yourself; thinking that you are a failure or that you were disappointing yourself or your family.	84.3	9.5	2.5	3.7
p7	Trouble concentrating on things, such as reading the newspaper or watching television.	77.7	12.0	3.7	6.6
p8	Slowness to move or talk to the point that other people notice, or the opposite, you have been so agitated or restless that you have been moving around a lot more than usual.	82.4	7.3	3.3	7.0

Source: primary research database.

Discussion

The psychometric properties of the Brazilian version of PHQ-8 in the general population demonstrated that the instrument can measure the depression construct in the studied context, showing adequate configurational and metric structures (dimensional validity), as well as connection between the obtained construct and similar tests/scales (convergent validity). The findings evidence the reliability of the items.

This study evaluated the psychometric properties of populations living in an urban area in a Brazilian municipality, with a predominance of women, young people, black people, without a partner, with low schooling level and low income. Sociodemographic characteristics similar to the Brazilian population, which registers a predominance of women (52.3%), aged up to 39 years (60%), black (54%), without a partner (61%), < 15 years of study (90%) and with an income below two minimum wages (70.8%) [50]. Therefore, the general profile of this study sample is similar to that of the Brazilian population.

Dimensional validity

The dimensional validity of the Brazilian version of PHQ-8 endorsed the structure of a single factor [27,33,34] and reinforced depression as a single construct (DSM-5, 2014). This indicates that the adaptation made in the semantics of response category "1" did not change the construct's behavior. Since all four response options on the scale were used, they may represent the respondents' frequency of depressive symptoms.

Studies that indicate the unidimensionality of the PHQ-8 have been conducted in specific populations, such as outpatients of a public hospital in Bolivia [33], university students from Latin America [27] and Mexican and Central American adults [34]. In studies in which the adjustment of models with two factors was satisfactory, these proved to be highly correlated [27,34], indicating unidimensionality or a higher order factor.

The exploratory analysis in this study showed that the obtained eigenvalues corroborated findings of extraction of a single factor. A unidimensionality sustained by the presence of satisfactory adjust-

Table 3

Analysis of the factorial structure of *Patient Health Questionnaire* (PHQ-8).

Item	ESEM		CFA	
	Model unidimensionality		**Model unidimensionality**	
	λi *	**δi ****	**λi ***	**δi ****
p1	0.592	0.650	0.633	0.600
p2	0.792	0.373	0.816	0.334
p3	0.694	0.519	0.708	0.499
p4	0.713	0.492	0.700	0.510
p5	0.618	0.619	0.610	0.628
p6	0.809	0.346	0.800	0.360
p7	0.651	0.577	0.661	0.563
p8	0.618	0.618	0.607	0.631
$r_{(i2-i6)}$ ***	-		0.249	
RMSEA (90%CI)	0.043 (0.037-0.049)		0.028 (0.022-0.034)	
CFI	0.986		0.984	
TLI	0.981		0.977	
AVE (95%CI)	-		0.48 (0.45-0.51)	
CR (95%CI)	-		0.88 (0.87-0.89)	

90%CI: 90% confidence interval; 95%CI: 95% confidence interval; AVE: average variance extracted; CFA: confirmatory factor analysis; CFI: comparative fit index; CR: composite reliability; ESEM: exploratory structural equation models; RMSEA: root mean square error of approximation; TLI: Tucker-Lewis index.

Source: primary research database.

* Factor load;

** Residual variance;

*** Residual correlations.

Table 4

Spearman's correlation tests of the *Patient Health Questionnaire* (PHQ-8) score with other constructs.

Depression	Spearman's correlation (r)	p-value *
Common mental disorders (SRQ-20)	0.592	< 0.001
Panic disorder (PHQ-8)	0.326	< 0.001
Generalized anxiety disorder (PHQ-8)	0.274	< 0.001

SRQ-20: *Self-Reporting Questionnaire.*

* Spearman's rank correlation test.

ment parameters and the absence of residual correlations. The single factor model showed excellent rates of absolute and incremental adjustment [43,46], which indicate reliable and discriminating items.

Regarding the presence of residual correlation, although the MIs suggested overlapping content of depressed mood symptoms (p2) and feelings of guilt or worthlessness (p6), which may be conceptually related, the residual correlation estimate was < 0.3, which does not indicate overlap/redundancy of content or need to reevaluate the items [47]. This finding corroborated the conditional independence of the instrument's items.

Convergent factorial validity

The study of the factorial loads supported the evaluation of the convergent factorial validity of the eight items of the PHQ for measuring depression. The high loads observed in all items indicated that they converge regarding depression. The item "Do you feel sad, depressed, or hopeless" (depressed mood) directed the construct the most. This is expected due to the concept of depression as a mood disorder, in which the main characteristic is the presence of sad, empty, or irritable moods [5].

However, the AVE is borderline, and therefore, reservedly admissible [46]. Considering the upper limit of the 95%CI, the studied general population showed that 51% of the latent trait of depression can be mapped by the 8 items of PHQ-8. Factorial validity values close to our results also identified in the general German (variance = 0.50) [51] and Taiwanese and Chinese (variance = 0.42) [6] populations, using PHQ-9.

These findings suggest that the suppression of the ninth item of the instrument did not alter the convergent factorial validity in general populations. We emphasize the lack of international or national studies of performance evaluation of the PHQ-8, thus requiring future studies to reinforce the results and interpretations.

The hypothesis testing confirmed the external correlation and endorsed the convergent factorial validity of PHQ-8 in general populations [49,52], which assessed the correlation with other similar constructs of mental illness analysis. Positive correlations with the scores of instruments that assessed CMD, generalized anxiety disorder, and panic disorder (PHQ sessions) corroborated the convergent validity of PHQ-8. Studies showed similar results with undergraduate students from the United States [27], immigrants from Mexico and Central America [34] and individuals from a psychiatric department of a university hospital in the Republic of Korea [23]; the latter with positive correlations with the depression diagnosis validity scale – the *Hamilton Depression Rating Scale* (HAMD). We did not find studies using PHQ-8 in a general population that evaluated this psychometric property.

Reliability of items

The CR identifies the level of interrelation between the instrument's items [45], which evidences the homogeneous measurement of a common characteristic [49,53], in this case, the depression construct.

The CR indicator is based on the factorial loads. Since the items' factorial loads can vary, this is presented as a more robust parameter for reliability analysis than Cronbach's alpha [54]. The PHQ-8 estimate of CR (0.88) indicated satisfactory homogeneity of the items, which consistently represented depression.

The Cronbach's alpha is relatively more fragile and strongly influenced by the reduced number of items in the PHQ-8 [49]. In its psychometric history, we identified the reliability of the items in international studies in specific populations: patients with chronic heart failure [29], outpatients [33] and individuals who visited the psychiatric department of a university hospital in the Republic of Korea [23].

Similar to other studies [23,32], the removal of item 9, of suicidal ideation, did not change the internal consistency of the scale.

Limitations

We used robust methods to analyze the psychometric properties of the PHQ-8 in a representative sample of the general population of a medium-sized Brazilian municipality (largest city in the Northeast region of the country). The results showed satisfactory validity and reliability of the PHQ-8 for use in the Brazilian scenario. However, some limitations must be considered.

The data used for the analysis are from a study collected in 2007, which may represent a specific moment marked by exposures to depression different from those in the Brazilian population nowadays. However, the temporality of these data can be relativized, especially due to the relevance of psychic morbidity in the populations' general context of illness, with a continuous growth of mental disorders in the last 10 years in Brazil, mainly depression [55,56]. Besides, evidence shows that situations of vulnerability have increased in that time, including situations related to the populations' sociode-

mographic conditions. Thus, the context in which this instrument was applied, which constituted the empirical basis of this study, did not show significant changes that influence or invalidate our results.

Besides the temporal aspect, caution and considerations are necessary to analyze our results due to the Brazilian states' diversity and their significant regional differences, and because our results refer to the investigation of a single Brazilian location.

We used a cluster-sampling method in which the observations are not independent [57,58], that is, the neighborhood or homogeneity effect cannot be ruled out. We tried to correct this effect with weighting analysis of the sample weights – a procedure to compensate for the different selection probabilities at each stage of the conglomerate. The increase in the size of the studied sample was another resource used to expand the sample heterogeneity and the possibilities of inclusion of population groups.

We used PHQ-9 performance studies for part of the comparisons of the obtained results. Although the results reveal that the suppression of the ninth item did not change the psychometric properties of the PHQ to measure depression, further studies with the PHQ-8 are essential to consolidate the properties identified in this study and to advance the knowledge of its psychometric properties, example of the evaluation of the dimensional structure and measurement invariance between groups [47]. The instrument's performance must be stable in different population groups to enable more reliable comparisons. Thus, the use of Confirmatory Factor Analysis of Multiple Groups is suggested to identify aspects that may interfere in the assessment of the construct, such as aspects related to gender and work situation.

Final considerations

Despite the limitations, the findings suggest that the Brazilian version of the PHQ-8 has a one-dimensional structure, good validity and reliability, being useful and effective in contexts of research and mental health care. Based on the scientific literature and the results of this study, we concluded that the PHQ-8 evidences validity that allows suggesting its use in the general Brazilian population, but cautiously due to the study being conducted in a single location.

Thus, the results strengthen its usefulness for mental health care in Brazil, since it has psychometric properties able to diagnose depression in the general population, freely available, with few items and easy to apply, analyze and interpret. Therefore, the PHQ-8 allows the early diagnosis of the disease, with the possibility of interventions at the beginning of the illness, with greater effectiveness, and can be used in the routine of health care services at its different levels of care, especially in primary health care.

Contributors

Additional informations

ORCID: Iracema Lua (0000-0002-5760-5073); Katia Santana Freitas (0000-0002-0491-6759); Jules Ramon Brito Teixeira (0000-0002-8443-7810); Michael Eduardo Reichenheim (0000-0001-7232-6745); Maura Maria Guimarães de Almeida (0000-0001-8340-2803); Tânia Maria de Araújo (0000-0003-2766-7799).

Acknowledgments

References

1. World Health Organization. Depression and other common mental disorders: global health estimates. Geneva: World Health Organization; 2017.

2. Cartwright A, Donkin R. Knowledge of depression and malingering: an exploratory investigation. Eur J Psychol 2020; 16:32-44.

3. Bromet E, Andrade LH, Hwang I, Sampson NA, Alonso J, Girolamo G, et al. Cross-national epidemiology of DSM-IV major depressive episode. BMC Med 2011; 9:90.

4. Theme Filha MM, Souza Junior PRB, Damacena GN, Szwarcwald CL. Prevalência de doenças crônicas não transmissíveis e associação com autoavaliação de saúde: Pesquisa Nacional de Saúde, 2013. Rev Bras Epidemiol 2015; 18 Suppl 2:83-96.

5. American Psychiatry Association. Manual diagnóstico e estatístico de transtornos mentais – DSM-5. 5th Ed. Porto Alegre: Artmed; 2014.

6. Liu S-I, Yeh Z-T, Huang H-C, Sun F-J, Tjung J-J, Hwang L-C, et al. Validation of Patient Health Questionnaire for depression screening among primary care patients in Taiwan. Compr Psychiatry 2011; 52:96-101.

7. Stopa SR, Malta DC, Oliveira MM, Lopes CS, Menezes PR, Kinoshita RT. Prevalência do autorrelato de depressão no Brasil: resultados da Pesquisa Nacional de Saúde, 2013. Rev Bras Epidemiol 2015; 18 Suppl 2:170-80.

8. Maurer DM, Raymond TJ, Davis BN. Depression: screening and diagnosis. Am Fam Physician 2018; 98:508-15.

9. Karekla M, Pilipenko N, Feldman J. Patient health questionnaire: Greek language validation and subscale factor structure. Compr Psychiatry 2012; 53:1217-26.

10. Gelaye B, Williams MA, Lemma S, Deyessa N, Bahretibeb Y, Shibre T, et al. Validity of the Patient Health Questionnaire-9 for depression screening and diagnosis in East Africa. Psychiatry Res 2013; 210:653-61.

11. Williams-Junior JW, Pignone M, Ramirez G, Perez Stellato C. Identifying depression in primary care: a literature synthesis of case-finding instruments. Gen Hosp Psychiatry 2002; 24:225-37.

12. Kroenke K, Spitzer RL. The PHQ-9: a new depression diagnostic and severity measure. Psychiatr Ann 2002; 32:509-15.

13. Kroenke K, Spitzer RL, Williams JBW, Löwe B. The Patient Health Questionnaire somatic, anxiety, and depressive symptom scales: a systematic review. Gen Hosp Psychiatry 2010; 32:345-59.

14. Janevic MR, Aruquipa Yujra AC, Marinec N, Aguilar J, Aikens JE, Tarrazona R, et al. Feasibility of an interactive voice response system for monitoring depressive symptoms in a lower-middle income Latin American country. Int J Ment Health Syst 2016; 10:59.

15. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001; 16:606-13.

16. Spitzer RL, Williams JBW, Kroenke K, Linzer M, Degruy FV, Hahn SR, et al. Utility of a new procedure for diagnosing mental disorders in primary care: the PRIME-MD 1000 study. JAMA 1994; 272:1749-56.

17. Woldetensay YK, Belachew T, Tesfaye M, Spielman K, Biesalski HK, Kantelhardt EJ, et al. Validation of the Patient Health Questionnaire (PHQ-9) as a screening tool for depression in pregnant women: Afaan Oromo version. PLoS One 2018; 13:e0191782.

18. Hammash MH, Hall LA, Lennie TA, Heo S, Chung ML, Lee KS, et al. Psychometrics of the PHQ-9 as a measure of depressive symptoms in patients with heart failure. Eur J Cardiovasc Nurs 2013; 12:446-53.

19. Familiar I, Ortiz-Panozo E, Hall B, Vieitez I, Romieu I, Lopez-Ridaura R, et al. Factor structure of the Spanish version of the Patient Health Questionnaire-9 in Mexican women. Int J Methods Psychiatr Res 2015; 24:74-82.

20. Yazici Güleç M, Güleç H, Şimşek G, Turhan M, Aydin Sünbül E. Psychometric properties of the Turkish version of the Patient Health Questionnaire-Somatic, Anxiety, and Depressive Symptoms. Compr Psychiatry 2012; 53:623-9.

21. 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; 73:163-8.

22. Wu Y, Levis B, Riehm KE, Saadat N, Levis AW, Azar M, et al. Equivalency of the diagnostic accuracy of the PHQ-8 and PHQ-9: a systematic review and individual participant data meta-analysis. Psychol Med 2020; 50:1368-80.

23. 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 Investig 2019; 16:300-5.

24. Spitzer RL, Kroenke K, Williams JB Validation and utility of a self-report version of PRIME-MD: The PHQ primary care study. JAMA 1999; 282:1737-44.

25. Spitzer RL, Williams JBW, Kroenke K, Hornyak R, McMurray J. Validity and utility of the PRIME-MD Patient Health Questionnaire in assessment of 3000 obstetric-gynecologic patients: The PRIME-MD Patient Health Questionnaire Obstetrics-Gynecology Study. Am J Obstet Gynecol 2000; 183:759-69.

26. Pinto-Meza A, Serrano-Blanco A, Peñarrubia MT, Blanco E, Haro JM. Assessing depression in primary care with the PHQ-9: can it be carried out over the telephone? J Gen Intern Med 2005; 20:738-42.

27. Alpizar D, Plunkett SW, Whaling K. Reliability and validity of the 8-item patient health questionnaire for measuring depressive symptoms of latino emerging adults. J Lat Psychol 2018; 6:115-30.

28. Kroenke K, Strine TW, Spitzer RL, Williams JBW, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. J Affect Disord 2009; 114:163-73.

29. Pressler SJ, Subramanian U, Perkins SM, Gradus-Pizlo I, Kareken D, Kim JS, et al. Measuring depressive symptoms in heart failure: validity and reliability of the Patient Health Questionnaire-8. Am J Crit Care 2011; 20:146-52.

30. Dhingra SS, Kroenke K, Zack MM, Strine TW, Balluz LS. PHQ-8 days: a measurement option for DSM-5 Major Depressive Disorder (MDD) severity. Popul Health Metr 2011; 9:11.

31. Spangenberg L, Brähler E, Glaesmer H. Identifying depression in the general population – a comparison of PHQ-9, PHQ-8 and PHQ-2. Z Psychosom Med Psychother 2012; 58:3-10.

32. Wells TS, Horton JL, Leardmann CA, Jacobson IG, Boyko EJ. A comparison of the PRIME-MD PHQ-9 and PHQ-8 in a large military prospective study, the Millennium Cohort Study. J Affect Disord 2013; 148:77-83.

33. Schantz K, Reighard C, Aikens JE, Aruquipa A, Pinto B, Valverde H, et al. 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. Int J Psychiatry Med 2017; 52:315-27.

34. Alpizar D, Laganá L, Plunkett SW, French BF. Evaluating the eight-item Patient Health Questionnaire's psychometric properties with Mexican and Central American descent university students. Psychol Assess 2018; 30:719-28.

35. World Health Organization. Mental health: new understanding, new hope. Geneva: World Health Organization; 2001.

36. Rocha SV, Almeida MMG, Araújo TM, Virtuoso Júnior JS. Prevalência de transtornos mentais comuns entre residentes em áreas urbanas de Feira de Santana, Bahia. Rev Bras Epidemiol 2010; 13:630-40.

37. Gothwal VK, Bagga DK, Bharani S, Sumalini R, Reddy SP. The Patient Health Questionnaire-9: validation among patients with glaucoma. PLoS One 2014; 9:e101295.

38. Zhong Q, Gelaye B, Fann JR, Sanchez SE, Williams MA. Cross-cultural validity of the Spanish version of PHQ-9 among pregnant Peruvian women: a Rasch item response theory analysis. J Affect Disord 2014; 158:148-53.

39. Santos KOB, Carvalho FM, Araújo TM. Internal consistency of the self-reporting questionnaire-20 in occupational groups. Rev Saúde Pública 2016; 50:6.

40. Santos KOB, Araújo TM, Pinho PS, Silva ACC. Avaliação de um instrumento de mensuração de morbidade psíquica: estudo de validação do Self-Reporting Questionnaire (SRQ-20). Rev Baiana Saúde Pública 2010; 34:544-60.

41. Marôco J. Análise de equações estruturais – fundamentos teóricos, software e aplicações. 2nd Ed. Pêro Pinheiro: ReportNumber; 2014.

42. Marsh HW, Muthén B, Asparouhov T, Lüdtke O, Robitzsch A, Morin AJS, et al. Exploratory structural equation modeling, integrating CFA and EFA: Application to students' evaluations of university teaching. Struct Equ Modeling 2009; 16:439-76.

43. Kline RB. Principles and practice of structural equation modeling. New York: The Guilford Press; 2015.

44. Brown TA. Methodology in the social sciences. Confirmatory factor analysis for applied research. 2nd Ed. New York: The Guilford Press; 2015.

45. Mokkink LB, Prinsen CAC, Bouter LM, de Vet HCW, Terwee CB. The COnsensus-based standards for the selection of health measurement INstruments (COSMIN) and how to select an outcome measurement instrument. Braz J Phys Ther 2016; 20:105-13.

46. Hair JFJ, Black WC, Babin BJ, Anderson RE, Tatham RL. Análise multivariada de dados. 6th Ed. Porto Alegre: Bookman; 2009.

47. Reichenheim ME, Hökerberg YHM, Moraes CL. Assessing construct structural validity of epidemiological measurement tools: a seven-step roadmap. Cad Saúde Pública 2014; 30:927-39.

48. Muthén BO, Asparouhov T. Latent variable analysis with categorical outcomes: multiple-group and growth modeling in Mplus. Mplus Web Notes 2002. http://www.statmodel.com/download/webnotes/CatMGLong.pdf (accessed on 15/Dec/2019).

49. Souza AC, Alexandre NMC, Guirardello EB. Propriedades psicométricas na avaliação de instrumentos: avaliação da confiabilidade e da validade. Epidemiol Serv Saúde 2017; 26:649-59.

50. Instituto Brasileiro de Geografia e Estatística. Pesquisa Nacional por Amostra de Domicílios (PNAD 2014-2015). https://www.ibge.gov.br/estatisticas/sociais/populacao/19897-sintese-de-indicadores-pnad2.html?edicao=9129&t=o-que-e (accessed on 20/Mar/2019).

51. Kocalevent RD, Hinz A, Brähler E. Standardization of the depression screener Patient Health Questionnaire (PHQ-9) in the general population. Gen Hosp Psychiatry 2013; 35:551-5.

52. Carvalho MS, Travassos C, Coeli CM, Reichenheim ME. Um passo à frente na política de acesso aberto de CSP: instrumentos de aferição. Cad Saúde Pública 2014; 30:1357.

53. Dortas-Junior SD, Lupi O, Dias GAC, Guimarães MBS, Valle SOR. Adaptação transcultural e validação de questionários na área da saúde. Braz J Allergy Immunol 2016; 4:26-30.

54. Valentini F, Damásio BF. Variância média extraída e confiabilidade composta: indicadores de precisão. Psicol Teor Pesqui 2016; 32:e322225.

55. Viana MC, Andrade LH. Lifetime prevalence, age and gender distribution and age-of-onset of psychiatric disorders in the São Paulo metropolitan area, Brazil: results from the São Paulo Megacity Mental Health Survey. Rev Bras Psiquiatr 2012; 34:249-60.

56. Lopes CS. Como está a saúde mental dos brasileiros? A importância das coortes de nascimento para melhor compreensão do problema. Cad Saúde Pública 2020; 36:e00005020.

57. Luiz RR, Magnanini MMF. A lógica da determinação do tamanho da amostra em investigações epidemiológicas. Cad Saúde Colet (Rio J.) 2000; 8:9-28.

58. Szwarcwald CL, Damacena GN. Complex sampling design in population surveys: planning and effects on statistical data analysis. Rev Bras Epidemiol 2008; 11 Suppl 1:38-45.

Resumo

O objetivo deste estudo foi avaliar as propriedades psicométricas da versão brasileira do Questionário de Saúde do Paciente *(Patient Health Questionnaire; PHQ-8). Realizamos um estudo com uma amostra de 4.170 indivíduos (≥ 15 anos) em área urbana. A amostragem por conglomerado foi adotada em duas etapas (setores censitários e ruas) com ponderação das estimativas por pesos amostrais. Foi utilizado questionário estruturado com dados sociodemográficos, o PHQ – com módulos para depressão, transtorno de ansiedade generalizada e transtorno do pânico – e o* Questionário de Autorrelato *(Self-Reporting Questionnaire; SRQ-20). Uma desordem linear sem perdas foi encontrada para manter as quatro categorias de resposta. A análise fatorial revelou unidimensionalidade do construto depressão com cargas fatoriais fortes, baixas variâncias residuais, baixa correlação residual entre os itens, bom ajuste do modelo, consistência interna e validade fatorial convergente satisfatória e altas correlações com outros instrumentos de construtos semelhantes. O PHQ-8 possui estrutura unidimensional com evidência de boa validade e confiabilidade, sendo adequado para uso na população brasileira.*

Depressão; Questionários; Estudo de Validação

Resumen

El objetivo de este estudio fue evaluar las propiedades psicométricas de la versión brasileña del Cuestionario de Salud del Paciente *(Patient Health Questionnaire; PHQ-8). Se trata de un estudio con una muestra de 4.170 individuos (≥ 15 años) en un área urbana. El muestreo por conglomerados fue adoptado en dos etapas (sectores censuales y calles) con ponderación de las estimaciones por pesos muestrales. Los instrumentos utilizados fueron un cuestionario estructurado con datos sociodemográficos, el PHQ -con módulos para depresión, trastorno de ansiedad generalizada y trastorno de pánico- y el* Cuestionario de Autoinforme *(Self-Reporting Questionnaire; SRQ-20). Al evaluar el PHQ-8 se verificó la validez de su construcción a través de su estructura dimensional, validez convergente y consistencia interna. Se encontró un desorden lineal sin pérdidas para mantener las cuatro categorías de respuesta. El análisis factorizado encontró unidimensionalidad del constructo depresión con fuertes cargas factoriales, bajas varianzas residuales, baja correlación residual entre ítems, buen ajuste del modelo, consistencia interna y satisfactoria validez factorial convergente (altas cargas y correlaciones con otras pruebas/escalas de constructos similares). El PHQ-8 tiene una estructura unidimensional con evidencia de buena validez y confiabilidad, y es adecuado para su uso en la población brasileña.*

Depresión; Cuestionarios; Estudio de Validación