

# Factor structure of the Spanish version of the Patient Health Questionnaire-9 in Mexican women

ITZIAR FAMILIAR,<sup>1</sup> EDUARDO ORTIZ-PANOZO,<sup>2</sup> BRIAN HALL,<sup>3</sup> ISABEL VIEITEZ,<sup>2</sup> ISABELLE ROMIEU,<sup>4</sup> RUY LOPEZ-RIDAURA<sup>2</sup> & MARTIN LAJOUS<sup>5,2</sup>

1 Department of Psychiatry, Michigan State University, Lansing, MI, USA

2 Center for Population Health Research, Instituto Nacional de Salud Pública

3 Department of Psychology, University of Macau, People's Republic of China

4 Section of Nutrition and Metabolism, International Agency for Cancer Research

5 Department of Epidemiology, Harvard School of Public Health

---

## Key words

depression assessment, Patient Health Questionnaire-9, factor structure, women, Mexico

## Correspondence

Eduardo Ortiz-Panozo, Instituto Nacional de Salud Pública, Av. Universidad 655, Col. Santa María Ahuacatitlán. CP 62100. Cuernavaca, Mor., México, Phone: +52 777 3293000 ext. 1709  
Email: eduardo.ortiz@insp.mx

## Abstract

Structure of the Spanish version of the nine-item Patient Health Questionnaire (PHQ-9) has been inconclusive. We report the factor structure of the PHQ-9 in 55,555 women from the Mexican Teachers' Cohort (MTC). Factor structure of the PHQ-9 was assessed by exploratory and confirmatory factor analyses in two sub-samples ( $n = 27,778$  and  $27,777$  respectively). A one-factor model of the PHQ-9 was the solution with the best fit to the data, exhibiting strong factor loadings (0.71 to 0.90) and high internal consistency (Cronbach's  $\alpha = 0.89$ ). A prevalence rate of moderate to high severity of depressive symptoms of 12.6% was identified. Results suggest that a global score is an appropriate measure of depressive symptoms and commend the use of the Spanish PHQ-9 as a measure of depression for research and clinical purposes. Copyright © 2014 John Wiley & Sons, Ltd.

Received 16 January 2014;

revised 5 March 2014;

accepted 8 May 2014

---

## Introduction

Several depression-screening instruments are available, differing in number of items, mode of administration, and link to diagnostic criteria. The nine-item Patient Health Questionnaire (PHQ-9; Spitzer *et al.*, 1999) is currently one of the most widely used measures across

different countries and settings (Gilbody *et al.*, 2007; Kroenke *et al.*, 2010; Manea *et al.*, 2012; Wittkamp *et al.*, 2007). It differs from other instruments for depression screening in that it was constructed to operationalize most of the Diagnostic and Statistical Manual for Mental Disorders, Fourth Edition (DSM-IV-TR) criteria for major depression.

Spanish versions of the PHQ-9 have been reported to be reliable and valid measures of depression in clinical settings from Spain (Diez-Quevedo *et al.*, 2001), Honduras (Wulsin *et al.*, 2002) and Chile (Baader *et al.*, 2012). The factor structure of the Spanish version of the PHQ-9 has been explored primarily in Spanish-speaking populations living in the United States, including primary care (Huang *et al.*, 2006), indigenous Mexican migrant farmworkers (Donlan and Lee, 2010) and Hispanic American community-based (Merz *et al.*, 2011) samples. However, results of exploratory and confirmatory factor analyses of the PHQ-9 from these and other studies have been inconclusive. Merz *et al.* (2011) found support and equivalency in a one-factor structure for the Spanish and English versions of the PHQ-9. In a study carried out in female college students in the United States that included Latinas, the English version of the PHQ-9 showed a two-factor structure describing somatic and affective dimensions (Granillo, 2012), challenging the unidimensionality of the measure and convenience of the global score derived.

An official version of the PHQ-9 for Mexico is available online (Spitzer *et al.*, 2009), yet no studies have examined the factor structure of the Mexican-Spanish version of the PHQ-9 in a population-based sample, warranting further research into the underlying dimensionality of the scale. Although self-reporting tools convey useful information for screening or research purposes, it cannot be assumed that the measure will accurately assess the construct intended when translated and applied in a different population (Ramada-Rodilla *et al.*, 2013). The objective of the study was to examine the structure of the PHQ-9 through factor analyses among women enrolled in the Mexican Teachers' Cohort (MTC). Originally aimed at evaluating the impact of dietary and lifestyle factors on the incidence of chronic diseases, such as breast cancer, diabetes and cardiovascular disease, MTC offers a unique sample from which to assess the structure of the PHQ-9 due to its large sample size, ample range in age-groups (25–84 years at enrollment), multi-state sampling frame, and non-clinical setting.

## Methods

### Participants

Data from the present study comes from the MTC, an ongoing longitudinal study of female teachers employed by the Ministry of Education across 12 culturally and economically diverse states in Mexico. The process of subject selection, administration of instruments, and data collection for MTC can be found in detail elsewhere (Romieu *et al.*, 2012). Briefly, in 2006–2008, eligible participants were identified through the Teachers' Incentives Program – an academic achievement-

based economic incentives program – administrative database. An initial baseline questionnaire was sent between 2006 and 2008 to 180,723 women to assess nutrition, general health, lifestyle, and risk factors for chronic diseases. In 2011, participants who responded to the baseline questionnaire ( $n = 115,346$ ) received the first follow-up questionnaire that included a self-administered screener for depression; the PHQ-9. At the time of this analysis, PHQ-9 information was available for 66,947 women, among them, 83% ( $n = 55,555$ ) completed all the nine items and were included in this analysis.

The study protocol was approved by the Ethics, Research and Biosecurity Committees of the National Institute of Public Health, Mexico. Participants gave informed consent.

### Measures

The PHQ-9 is a self-report questionnaire in which each of the nine PHQ depression items corresponds to one of the DSM-IV Diagnostic Criterion A for major depressive disorder (American Psychiatric Association, 2000). The measure refers to symptoms occurring within the previous two-week interval. Based on several studies, validity has been reported within satisfactory ranges, with a sensitivity of 88% and specificity of 88% for major depression (Spitzer *et al.*, 1999, 2000). Response options to the PHQ-9 items are in a four-point scale; including zero for *not at all*, one for *several days*, two if *more than half the days*, and three to *nearly every day*. Scoring is obtained by simply adding the response to each item yielding a range of zero to 27 (Spitzer *et al.*, 1999). We categorized depression severity in the PHQ-9 by the frequently used cutoff scores as: none or minimal (0–4), mild (5–9), moderate (10–14), moderately severe (15–19) and severe (20 or greater). To determine prevalence rates, a cutoff score  $\geq 10$  was used on the PHQ-9 according to Kroenke *et al.* (2001) and because the range of  $\geq 10$  up to 27 correlates well with moderate to severe levels of depression in Spanish-speaking populations (Baader *et al.*, 2012; Diez-Quevedo *et al.*, 2001; Wulsin *et al.*, 2002). We used the Spanish for Mexico version of the PHQ-9 available free online (Spitzer *et al.*, 2009). After piloting this version of the PHQ-9 within a small sample of women, the team of researchers only added “have you had” at the beginning of items 1, 3 and 7, and deleted the word “hopeless” from item 2.

### Covariates

We categorized age reported in the 2011 questionnaire in four groups (25–34, 35–44, 45–54, and 55 or more

years). Data on education and marital status was derived from the baseline questionnaire (2006–2008). The baseline questionnaire also included a yes/no response on the ownership of seven household assets: telephone, car, computer, vacuum cleaner, microwave oven, cell phone and Internet access. We defined socio-economic level by the number of owned household assets categorized in tertiles: low,  $\leq 4$ ; medium, 5–6; and high, 7 household assets.

### Data analysis

We conducted two main analyses in this study to assess the PHQ-9 factor structure: exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). To cross-validate findings from the EFA with CFA, we generated two random sub-samples from the original sample of participants. One sub-sample was assigned to the EFA group ( $n = 27,778$ ) and the remaining observations were assigned to the CFA group ( $n = 27,777$ ). In each sub-sample, we conducted descriptive analyses of demographic characteristics, PHQ-9 items and scores. We used chi-squared tests for comparisons between sub-samples. We considered a two-tailed  $p$ -value of 0.05 as statistically significant for these tests. We calculated Cronbach's alpha ( $\alpha$ ) to measure internal consistency of items in the full measure and in each underlying factor. Also we explored item-test and item-rest correlations as well as changes in  $\alpha$  if an item was removed from the full measurement.

The factor structure of the PHQ-9 was initially examined using EFA. We used the eigenvalues and their graphic representation to determine the number of factors retained (Ford *et al.*, 1986). We dropped factors with eigenvalues of less than one and/or all further factors after the one starting the elbow of the scree plot. One- and two-factor models were fitted. Items with factor loading of 0.4 or more and no cross-loadings were assigned to the corresponding factor. To obtain the final solution, orthogonal and oblique rotations of factors were compared. Alternative EFA models were compared through the root mean square error of approximation (RMSEA) and the root mean square residual (RMSR), where lower values indicate better fit and good models are considered to have a RMSEA and a RMSR of 0.05 or less (Muthén, 1998–2004).

To test the stability of the PHQ-9 we conducted CFA in the other sub-sample, testing the fit of one- (Model I) and two-factor (Model II) models from the EFA. Additionally, we tested the fit of a third model (Model III) as proposed by Granillo (2012) (i.e. a two-factor model with affective and somatic factors). Goodness-of-fit was assessed using the Tucker–Lewis index (TLI) (Bentler and Bonett, 1980;

Tucker and Lewis, 1973); the comparative fit index (CFI) (Bentler, 1990), both of which range from zero to one with higher values indicating better fit; and RMSEA (Steiger, 1990). Values equal to, or greater than, 0.95 for the CFI and TLI, and values lower than 0.08 for the RMSEA, were considered indicators of excellent model goodness-of-fit (Bentler, 1990; Browne and Cudeck, 1992; Vandenberg and Lance, 2000). Also a chi-square goodness-of-fit test was done, where the associated non-significant  $p$ -values indicate good model fit (Muthén, 1998–2004).

Because history of depression may increase the awareness about symptoms that the PHQ-9 assesses, we conducted a sensitivity analysis in which we replicated the procedures described above among women who reported not having a clinical diagnosis of depression in the previous two years.

Factor analyses were estimated using weighted least squares method for categorical outcomes. All analyses were conducted using Mplus 7 software (Muthén and Muthén, Los Angeles, CA).

## Results

### Sample characteristics

The characteristics of the 55,555 women included in this analysis are presented in Table 1. On average, women included in this analysis were 45–55 years old, married, had a university degree, and a medium socio-economic level. Overall, there were no significant differences between the sub-samples used for the EFA and CFA.

In the total analytic sample, the mean PHQ-9 score was 4.5 (range 0–27) and 7026 (12.6%) women had a score equivalent to moderate to high severity depression ( $\text{PHQ-9} \geq 10$ ). The frequencies of each severity category are included in Table 1. The mean score of items corresponding to somatic symptoms ranged from 0.3 (*Moving or speaking too slow or too fast*) to 0.9 (*Feeling tired, with little energy*), and the mean score of items related to an affective/cognitive dimension ranged from 0.1 (*Suicidal thoughts*) to 0.6 (*Feeling down, depressed*).

### Exploratory factor analysis (EFA)

Results from the EFA are presented in Table 2. Eigenvalues and their graphic expression suggested a one-factor solution. The eigenvalue on this single factor was 6.0 and the variance explained by this factor was 67.0%. The unidimensional model had an acceptable fit to the data (RMSEA = 0.07; RMSR = 0.04).

The EFA results for a two-factor solution are also presented in Table 2. This model displayed a slight

**Table 1.** Characteristics of women in each sample for exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). The Mexican Teachers' Cohort, Mexico, 2011

Characteristics	EFA sample ( <i>n</i> = 27,778)	CFA sample ( <i>n</i> = 27,777)	<i>p</i> -Value <sup>a</sup>
	Frequency (%)	Frequency (%)	
<i>Age category (years)</i>			
25–34	3573 (12.9)	3577 (12.9)	0.508
35–44	9418 (33.9)	9487 (34.2)	
45–54	13,029 (46.9)	12,887 (46.4)	
55+	1758 (6.3)	1826 (6.6)	
<i>Education</i>			
Less than University degree	2497 (10.7)	2442 (10.5)	0.671
University degree	16,974 (72.9)	17,046 (73.2)	
Postgraduate degree	3801 (16.3)	3785 (16.3)	
<i>Marital status</i>			
Single	4845 (17.6)	4899 (17.8)	0.509
Cohabiting	2421 (8.8)	2317 (8.4)	
Married	16,860 (61.3)	16,901 (61.4)	
Divorced or separated	2858 (10.4)	2839 (10.3)	
Widowed	544 (2.0)	570 (2.1)	
<i>Socio-economic level<sup>b</sup></i>			
Low	8801 (34.3)	8875 (34.5)	0.101
Medium	11,624 (45.3)	11,775 (45.8)	
High	5230 (20.4)	5045 (19.6)	
<i>Depressive symptoms severity (PHQ-9 score range)</i>			
None or minimal (0–4)	16,850 (60.7)	16,908 (60.9)	0.785
Mild (5–9)	7436 (26.8)	7335 (26.4)	
Moderate (10–14)	2198 (7.9)	2244 (8.1)	
Moderately severe (15–19)	858 (3.1)	839 (3.0)	
Severe ( $\geq 20$ )	436 (1.6)	451 (1.6)	

<sup>a</sup>*p*-Value for the homogeneity of distributions between exploratory and confirmatory factor analysis samples.

<sup>b</sup>As defined by the number of owned household assets (telephone, car, computer, vacuum cleaner, microwave oven, cell phone, Internet access) categorized in tertiles.

improvement of fit according to indices values. Standardized factor loadings permitting assignment of an item to a specific factor exceeded 0.4, and individual factor item loadings ranged from 0.46 to 0.80, suggesting each item was strongly related to the underlying factor. The two-factor solution accounted for 73.8% of the variance. The first factor was defined by five items that mostly related to affective symptoms of depression (e.g. *Little or no pleasure doing things, Feeling depressed*) but also included two items that tap into somatic symptomatology (e.g. *Feeling tired, with little energy* and *Poor appetite/overeating*). The second factor included four items; two affective symptoms (*Feeling bad, Suicidal thoughts*) and two somatic symptoms of depression (*Trouble concentrating, Irregular activity level*). Correlation between factors was high (0.76).

Internal consistency was good for the full scale ( $\alpha = 0.89$ ), and for each factor separately ( $F1 = 0.85$ ;  $F2 = 0.77$ ) (Table 2). The item-test (i.e. correlation between the item and the total test) and item-rest (i.e. correlation between the item and the sum of the rest of the item scores) correlations of *Suicidal thoughts* were lower than those of the other items, both in the one- (0.57 versus 0.70–0.81 and 0.45 versus 0.71–0.84, respectively) and two-factor (0.70 versus 0.77–0.81 and 0.46 versus 0.58–0.64, respectively) solutions. Because the item-test correlation can be considered as a measure of item functioning associated with internal consistency, we explored if removing the item *Suicidal thoughts* from the scale would impact the test's internal consistency. However, omitting this item from the scale did not change the  $\alpha$  coefficient and was kept in subsequent analyses.

**Table 2.** Factor loadings and model fit statistics for one-factor (F1) and two-factor (F2) solutions. Exploratory factor analysis of the PHQ-9 among women from the Mexican Teachers' Cohort ( $n=27,778$ ), Mexico, 2011

Exploratory factor analysis	One factor	Two factors <sup>a</sup>	
	Depression	F1	F2
1. Little or no pleasure doing things	0.82	<i>0.67</i>	0.19
2. Feeling down, depressed	0.89	<i>0.68</i>	0.26
3. Trouble with sleep	0.75	<i>0.60</i>	0.19
4. Feeling tired, little energy	0.83	<i>0.80</i>	0.08
5. Poor appetite/overeating	0.74	<i>0.46</i>	0.33
6. Feeling bad about yourself	0.83	0.17	<i>0.73</i>
7. Trouble concentrating on things	0.80	0.15	<i>0.71</i>
8. Moving or speaking too slow or too fast	0.77	0.19	<i>0.64</i>
9. Suicidal thoughts	0.75	0.10	<i>0.71</i>
Cronbach's alpha ( $\alpha$ )	0.89	0.85	0.77
Correlation between factors	—		0.76
$\chi^2$ (degrees of freedom)	3275.4 (27)		1451.0 (19)
RMSEA	0.07		0.05
RMSR	0.04		0.02

Note: PHQ-9, Patient Health Questionnaire-9; RMSEA, root mean square error of approximation; RMSR, root mean square residual.

<sup>a</sup>Italic typeface indicates factor loading > 0.4.

**Table 3.** Standardized factor loadings and model fit statistics for one-factor (F1) and two-factor (F2) solutions. Confirmatory factor analysis of the PHQ-9 among women from the Mexican Teachers' Cohort ( $n=27,777$ ), Mexico, 2011

Confirmatory factor analysis	Model I	Model II		Model III	
	Depression	F1	F2	Affect	Somatic
1. Little or no pleasure doing things	0.71	0.71		0.71	
2. Feeling down, depressed	0.90	0.92		0.91	
3. Trouble with sleep	0.75	0.76			0.71
4. Feeling tired, little energy	0.84	0.85			0.85
5. Poor appetite/overeating	0.74	0.75			0.74
6. Feeling bad about yourself	0.83		0.71	0.83	
7. Trouble concentrating on things	0.80		0.82	0.81	
8. Moving or speaking too slow or too fast	0.78		0.79		0.78
9. Suicidal thoughts	0.75		0.76	0.75	
Cronbach's alpha ( $\alpha$ )	0.89	0.85	0.77	0.82	0.78
Correlation between factors	—	0.98		0.99	
$\chi^2$ (degrees of freedom)	6449.5 (28)	9400.0 (28)		5859.2 (28)	
RMSEA	0.09	0.11		0.09	
TLI	0.97	0.95		0.97	
CFI	0.98	0.96		0.98	

Note: Model I and Model II were based on the one- and two-factor solutions from exploratory factor analysis, respectively. Model III was based on the two-factor solution by Granillo (2012). PHQ-9, Patient Health Questionnaire-9; RMSEA, root mean square error of approximation; TLI, Tucker–Lewis Index; CFI, comparative fit index.



### Confirmatory factor analysis (CFA)

The fit statistics and standardized factor loadings for the competing CFA models tested in the split sample of 27,777 women are presented in Table 3. The one-factor model (Model I) displayed good fit to the data, was comparable to the results obtained with EFA (RMSEA = 0.09; TLI = 0.97; CFI = 0.98), and obtained the highest value of internal consistency ( $\alpha = 0.89$ ). Compared to Model I, the two-factor empirical model from the EFA (Model II) showed a modest decline in all fit indices; and factor loadings were above 0.70. Correlation between the two factors was 0.98.

In Model III the items were forced to load on two factors as proposed by Granillo (2012), each characterizing affective and somatic dimensions of depression. Fit indices indicated a good fit to the data and were almost the same as those of Model I. The standardized factor loadings of Model III ranged from 0.71–0.91 with *Little or no pleasure doing things* and *Trouble with sleep* as the items with the lowest loadings (Table 3). The correlation between these affective somatic factors was very high (0.99) and internal consistency values were 0.82 and 0.78, respectively.

When we excluded women who reported clinical depression ( $n = 1237$ ; 2.2%), results remained essentially the same.

### Discussion

This study represents the first to assess the structure of the Spanish version of the PHQ-9 among a large sample of Mexican women. Our results suggest that the Mexican-Spanish version of the PHQ-9 is unidimensional and has good internal consistency.

Results from our factor analytical study suggest that the Mexican-Spanish version of the PHQ-9 measures a single construct corresponding to depression. In contrast with previous studies (Granillo, 2012; Krause *et al.*, 2010) the two-factor solution from EFA in our sample produced two factors that represented a mixture of somatic and affective symptoms, precluding a conceptually reasonable interpretation of factors as a separate constructs. When using factor analyses, the nature of the items loading into a factor is an important issue to interpret how a factor is determined. Specifically, interpretation of factor analysis results calls for an *a priori* sense, based on past evidence and theory, of the number of plausible factors and which items are related to which factors. Further support for a single factor solution was provided by the high correlation between the two-factors identified in all two-factor solutions

tested (0.76–0.99), suggesting that factors are not well differentiated.

Reliability coefficients obtained in our study (0.77–0.89) are within the range considered acceptable and are similar to those reported by others for the Spanish version of the PHQ-9 in mostly or exclusively female-based samples (Baader *et al.*, 2012; Huang *et al.*, 2006; Merz *et al.*, 2011). High reliability coefficients also favor a unidimensional conceptualization of the PHQ-9.

We found only one report from a Latin American country examining the factor structure of the Spanish version of the PHQ-9 (Baader *et al.*, 2012). This study also found a one-factor structure of the PHQ-9 by EFA in a predominantly female (79%) sample of patients from primary care in Chile ( $n = 1327$ ). Three additional studies carried out in Latino populations living in the United States further support a one-factor structure of the PHQ-9. In a predominantly female (98%) and Spanish-speaking (74%) sample recruited from a primary care setting ( $n = 974$ ), Huang *et al.* (2006) showed a unidimensional structure of the PHQ-9 by EFA. Merz *et al.* (2011) compared the structure of English and Spanish versions of the PHQ-9 among two sub-samples of Hispanic American women from a community-based study and found that both English and Spanish versions of the PHQ-9 had an underlying unidimensional structure. In the sub-sample that completed the PHQ-9 in Spanish ( $n = 234$ ), 88% of the women were born in Mexico. Finally, Donlan and Lee (2010) also found support for a single factor for the Spanish version of the PHQ-9 by CFA among predominantly male (70%) indigenous Mexican migrant farmworkers in the United States ( $n = 123$ ), suggesting that the core features of depression are common and culturally appropriate among Mexicans.

The CFA conducted in a split-sample allowed us to evaluate the fit of one- and two-factor models. Granillo's (2012) two-factor model, based on a sample of Latin female college students ( $n = 1455$ ) did not provide a significant improvement in model fit to our data when compared to a one-factor model. A possible explanation is that in the sample used by Granillo (2012) women reported somatic symptoms more frequently than in our sample (mean of somatic scores of 1.0 and 0.6, respectively). The same could apply to Krause *et al.*'s (2010) findings, where a two-factor solution for the PHQ-9 is presented among patients with spinal-cord injury reporting a high mean of the somatic items. Interestingly, Merz *et al.* (2011) reported a mean score of somatic symptoms on the PHQ-9 that was the same as ours and obtained a one-factor solution.

A drawback of the studies discussed earlier is their reliance on moderate-sized samples of adults in clinical settings and minority groups living in the United States, limiting the generalizability of results. We sought to address these limitations and contribute evidence supporting a one-factor structure of the measure specifically among women, further advocating the use of the Spanish version of the PHQ-9 as a research and clinical tool to measure depression among Mexicans.

As in our study, previous reports have noted low factor loadings for the suicidality item (Huang *et al.*, 2006; Granillo, 2012; Merz *et al.*, 2011; Baader *et al.*, 2012). One explanation could be that suicidality does not discriminate well between depressed and non-depressed persons. The lower item-test and item-rest correlations of suicidality found in the current study also support this claim. Suicidality is rarely endorsed but should be retained given its clinical importance. Regardless of the global score, any affirmative response to this item should prompt immediate clinical attention.

Depression represents the leading cause of disease-related disability among women worldwide (Murray *et al.*, 1996). Studies have consistently showed a higher prevalence of depression among women than in men. According to a population-based study in the United States, the lifetime prevalence of major depression in women (21%) is almost twice that of men (12.7%) (Kessler *et al.*, 2005). Population-based studies from Mexico portray a similar epidemiological picture for major depression. The Mexican National Comorbidity Survey, carried out in a representative adult population across the country, showed a lifetime prevalence of major depression of 7.2% and 4.8% in the last 12 months. In this same survey, depression was more frequent among women (10.4%) than men (5.4%) (Medina-Mora *et al.*, 2007).

The mean PHQ-9 score in this study was 4.5, which is similar to reports from a community sample of Latina female (mean [ $M$ ]=4.6) (Merz *et al.*, 2011) and another predominantly Latino sample ( $M$ =4.7) (Huang *et al.*, 2006). However, it is worthy to mention that over 7000 women from the MTC had PHQ-9 scores  $\geq 10$  suggesting mild to moderate symptoms of depression. Using this cutoff score to define depression, the estimated depression prevalence of 12.6% is higher than reports for the Mexican population (1.3% prevalence in the previous month) (Medina-Mora *et al.*, 2007), and of women in other Latin American countries such as Argentina (6.0%), Chile (8%), and Brazil (10.2%) (Kohn *et al.*, 2005). Although caution should be exercised when interpreting depression prevalence

established with a screening tool such as the PHQ-9, our findings suggest that depression symptoms among educated women in Mexico might be extensive and merit further investigation.

The Mexican Comorbidity Study reported low service utilization (11.7%) of formal and informal services despite depression ranking first in terms of disability for women in Mexico (Medina-Mora *et al.*, 2007). Importantly, there was a trend of lower treatment seeking in women when compared to men (Rafful *et al.*, 2012), making it imperative to increase and strengthen screening, outreach, and primary care services for women. This treatment gap stresses the importance to incorporate valid and reliable measures into primary care such as the PHQ-9 that proactively identify individuals with depression and serve as gate openers to appropriate care.

Findings from this study can bear important research implications. To reduce the burden of disease associated with major depression, the World Health Organization suggests generating epidemiological data with a gender perspective. This emphasizes the need to validate screening instruments that can effectively capture depressive symptomatology among women (World Health Organization, 2000). In less than a decade, the PHQ-9 has had considerable uptake both clinically and by researchers as standard measure of depression. If researchers are to follow the current trend and continue to use the Spanish version of the PHQ-9 as the preferred measure of depression symptoms, then it is important that the reliability and construct validity of the scale are well established. The present study, which included the analysis of more than 55,000 women, not only gives evidence that the Spanish version of the PHQ-9 is a reliable and valid self-report measure for depression, but also sheds light into the depression construct in this population. Future studies should empirically derive the most valid cutoffs to demarcate severity gradations and produce a framework of normative data for Mexicans in which to interpret and compare depression symptoms with other populations.

### Limitations

Results presented here need to be viewed in light of some limitations. Women included in this analysis were those who responded to the follow-up questionnaire and completed the PHQ-9 section, and on average are younger, more educated and from a higher socioeconomic level than the national average. These facts are likely to have an impact on the generalizability of

our results. However, the contrasting states included in the sample may reflect important population differences across the country. Further, women in our study had a mean age (44 years) that falls within the group identified as having the highest risk for lifetime major depressive episode (45–54 years) in Mexican women (Rafful *et al.*, 2012). While we approached the construct validity of the Mexican-Spanish version of the PHQ-9 using CFA in a split-sample and tested several models, convergent validity of the PHQ-9 with an independent criterion standard such as the World Health Organization Composite International Diagnostic Interview (CIDI) (Kessler *et al.*, 2004) or the Structured Clinical Interview for DSM Disorders (SCID) (First *et al.*, 1996) would be of value in future research. Finally, findings from this study should be replicated in men to extend the generalizability of our results.

### Acknowledgements

We are grateful to the participants who made this study possible. We wish to thank Professor Víctor Sastré from Teachers' Incentives Program (Carrera Magisterial), Ministry of Education (Mexico), for support in data collection.

### References

- American Psychiatric Association. (2000) Diagnostic and Statistical Manual of Mental Disorders (4th edition), Washington, DC, American Psychiatric Association.
- Baader T., Molina J., Venezan S., Rojas C., Fariás R., Fierro-Freixenet C., Backenstrass M., Mundt C. (2012) Validity and utility of PHQ9 (Patient Health Questionnaire) in the diagnosis of depression in user patients of primary care in Chile. *Revista Chilena de Neuro-Psiquiatría*, **50**(1), 10–22.
- Bentler, P.M. (1990) Comparative fit indexes in structural models. *Psychological Bulletin*, **107**(2), 238–246.
- Bentler P.M., Bonett D.G. (1980) Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, **88**(3), 588–606.
- Browne M.W., Cudeck R. (1992) Alternative ways of assessing model fit. *Sociological Methods Research*, **21**(2), 230–258.
- Diez-Quevedo C., Rangil T., Sanchez-Planell L., Kroenke K., Spitzer R.L. (2001) Validation and utility of the patient health questionnaire in diagnosing mental disorders in 1003 general hospital Spanish inpatients. *Psychosomatic Medicine*, **63**(4), 679–686.
- Donlan W., Lee J. (2010) Screening for depression among indigenous Mexican migrant farmworkers using the Patient Health Questionnaire-9. *Psychological Reports*, **106**(2), 419–432.
- First M.B., Williams J.B.W., Spitzer R.L., Gibbon M. (1996) Structured Clinical Interview for DSM-IV Axis I Disorders, Clinician Version (SCID-CV), Washington, DC, American Psychiatric Press.
- Ford J.K., MacCallum R.C., Tait M. (1986) The application of exploratory factor analysis in applied psychology: a critical review and analysis. *Personnel Psychology*, **39**(2), 291.
- Gilbody S., Richards D., Brealey S., Hewitt C. (2007) Screening for depression in medical settings with the Patient Health Questionnaire (PHQ): a diagnostic meta-analysis. *Journal of General Internal Medicine*, **22**(11), 1596–1602.
- Granillo T. (2012) Structure and function of the Patient Health Questionnaire-9 among Latina and non-Latina white female college students. *Journal of the Society for Social Work and Research*, **3**(2), 80–93.
- Huang F.Y., Chung H., Kroenke K., Delucchi K.L., Spitzer R.L. (2006) Using the Patient Health Questionnaire-9 to measure depression among racially and ethnically diverse primary care patients. *Journal of General Internal Medicine*, **21**(6), 547–552.
- Kessler R.C., Abelson J., Demler O., Escobar J.L., Gibbon M., Guyer M.E., Howes M.J., Jin R., Vega W.A., Walters E.E., Wang P., Zaslavsky A., Zheng H. (2004) Clinical calibration of DSM-IV diagnoses in the World Mental Health (WMH) version of the World Health Organization (WHO) Composite International Diagnostic Interview (WMHCIDI). *International Journal of Methods in Psychiatric Research*, **13**(2), 122–139.
- Kessler R.C., Chiu W.T., Demler O., Merikangas K.R., Walters E.E. (2005) Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, **62**(6), 617–627.
- Kohn R., Levav I., de Almeida J.M., Vicente B., Andrade L., Caraveo-Anduaga J.J., Saxena S., Saraceno B. (2005) Mental disorders in Latin America and the Caribbean: a public health

### Conclusions

Despite these limitations, our results support the use of the Spanish version of the PHQ-9 as a reliable instrument maintaining important psychometric properties useful to assess and monitor depression among women in Mexico. The unidimensional structure derived from our analyses supports the use of the Spanish version of the PHQ-9 with a global score to inform about depression severity and suggest caseness. The strengths of this report lie on its methodology (EFA, CFA), large sample size, and characteristics of the sample. Factor analytic studies illuminate the possible clinical and future research applications that can be derived from psychometric evaluations of self-report measures used in large epidemiological studies. This study supports future work in addressing depression in women from a public health standpoint, identifying needs and informing future development of evidence-based services for Mexican women.



- priority. *Revista Panamericana de Salud Pública*, **18**(4–5), 229–240.
- Krause J.S., Reed K.S., McArdle J.J. (2010) A structural analysis of health outcomes after spinal cord injury. *Journal of Spinal Cord Medicine*, **33**(1), 22–32.
- Kroenke K., Spitzer R.L., Williams J.B. (2001) The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine*, **16**(9), 606–613.
- Kroenke K., Spitzer R.L., Williams J.B.W., Löwe B. (2010) The Patient Health Questionnaire somatic, anxiety, and depressive symptom scales: a systematic review. *General Hospital Psychiatry*, **32**(4), 345–359.
- Manea L., Gilbody S., McMillan D. (2012) Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): a meta-analysis. *CMAJ – Canadian Medical Association Journal*, **184**(3), E191–E196.
- Medina-Mora M.E., Borges G., Benjet C., Lara C., Berglund P. (2007) Psychiatric disorders in Mexico: lifetime prevalence in a nationally representative sample. *British Journal of Psychiatry*, **190**, 521–528. doi:10.1192/bjp.pb.106.025841
- Merz E.L., Malcarne V.L., Roesch S.C., Riley N., Sadler G.R. (2011) A multigroup confirmatory factor analysis of the Patient Health Questionnaire-9 among English- and Spanish-speaking Latinas. *Cultural Diversity and Ethnic Minority Psychology*, **17**(3), 309–316.
- Murray C.J.L., Lopez A.D., Harvard School of Public Health, World Health Organization, World Bank. (1996) The Global Burden of Disease: A Comprehensive Assessment of Mortality and Disability from Diseases, Injuries, and Risk Factors in 1990 and Projected to 2020, Cambridge, MA, Published by the Harvard School of Public Health on behalf of the World Health Organization and the World Bank; Distributed by Harvard University Press.
- Muthén B.O. (1998–2004) Mplus Technical Appendices, Los Angeles, CA, Muthén & Muthén.
- Rafful C., Medina-Mora M.E., Borges G., Benjet C., Orozco R. (2012) Depression, gender, and the treatment gap in Mexico. *Journal of Affective Disorders*, **138**(1–2), 165–169.
- Ramada-Rodilla J.M., Serra-Pujadas C., Delclos-Clanchet G.L. (2013) Cross-cultural adaptation and health questionnaires validation: revision and methodological recommendations. *Salud Pública de México*, **55**(1), 57–66.
- Romieu I., Escamilla-Nunez M.C., Sanchez-Zamorano L.M., Lopez-Ridaura R., Torres-Mejia G., Yunes E.M., Lajous M., Rivera-Dommarco J.A., Lazcano-Ponce E. (2012) The association between body shape silhouette and dietary pattern among Mexican women. *Public Health Nutrition*, **15**(1), 116–125.
- Spitzer R.L., Kroenke K., Williams J.B. (1999) Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. *JAMA – Journal of the American Medical Association*, **282**(18), 1737–1744.
- Spitzer R.L., Williams J.B., Kroenke K. (2009) <http://www.phqscreeners.com/> [25 September 2013].
- Spitzer R.L., Williams J.B., Kroenke K., Hornyak R., McMurray J. (2000) Validity and utility of the PRIME-MD patient health questionnaire in assessment of 3000 obstetric-gynecologic patients: the PRIME-MD Patient Health Questionnaire Obstetrics-Gynecology Study. *American Journal of Obstetrics and Gynecology*, **183**(3), 759–769.
- Steiger J.H. (1990) Structural model evaluation and modification: an interval estimation approach. *Multivariate Behavioral Research*, **25**(2), 173–180.
- Tucker L.R., Lewis C. (1973) A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, **38**(1), 1–10.
- Vandenberg R.J., Lance C.E. (2000) A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, **3**(1), 4–69.
- Wittkamp K.A., Naeije L., Schene A.H., Huyser J., van Weert H.C. (2007) Diagnostic accuracy of the mood module of the Patient Health Questionnaire: a systematic review. *General Hospital Psychiatry*, **29**(5), 388–395.
- World Health Organization. (2000) Women's Mental Health: An Evidence Based Review, Geneva, World Health Organization.
- Wulsin L., Somoza E., Heck J. (2002) The feasibility of using the Spanish PHQ-9 to screen for depression in primary care in Honduras. *Primary Care Companion – Journal of Clinical Psychiatry*, **4**(5), 191–195.