NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Comparative Performance of Patient Health Questionnaire-9 and Edinburgh Postnatal Depression Scale for Screening Antepartum Depression

Qiuyue Zhong, Sc.M.[1], **Bizu Gelaye, Ph.D.**[1], **Marta Rondon, M.D.**[2], **Sixto E Sánchez, M.D., Ph.D.**[3], **Pedro J García, M.D., M.P.H.**[3], **Elena Sánchez**[4], **Yasmin V Barrios, M.P.H.**[1], **Gregory E. Simon, M.D., M.P.H.**[5], **David C. Henderson, M.D.**[6], **Swee May Cripe, Ph.D.**[7], and **Michelle A Williams, Sc.D.**[1]

[1]Department of Epidemiology, Harvard School of Public Health, Boston, MA

[2]Department of Medicine, Cayetano Heredia Peruvian University, Lima, Peru

[3]Universidad San Martin de Porres, Lima, Peru

[4]Asociación Civil PROESA, Lima, Peru

[5]Group Health Research Institute, Seattle, WA

[6]Department of Psychiatry, Massachusetts General Hospital, Boston, MA

[7]School of Medicine, Perdana University, Kuala Lumpur, Malaysia

Abstract

Objective—We sought to evaluate the psychometric properties of two widely used screening scales: the Patient Health Questionnaire (PHQ-9) and Edinburgh Postnatal Depression Scale (EPDS) among pregnant Peruvian women.

Methods—This cross-sectional study included 1,517 women receiving prenatal care from February 2012 to March 2013. A structured interview was used to collect data using PHQ-9 and EPDS. We examined reliability, construct and concurrent validity between two scales using internal consistency indices, factor structures, correlations, and Cohen's kappa.

Results—Both scales had good internal consistency (Cronbach's alpha > 0.8). Correlation between PHQ-9 and EPDS scores was fair (rho=0.52). Based on exploratory factor analysis

Corresponding author: Dr. Bizu Gelaye, Department of Epidemiology, Harvard School of Public Health, 677 Huntington Ave, K505F, Boston, MA 02115 USA, Telephone: 617-432-6477, Facsimile: 617-566-7805, bgelaye@hsph.harvard.edu.

(EFA), both scales yielded a two-factor structure. EFA including all items from PHQ-9 and EPDS yielded four factors, namely, "somatization", "depression and suicidal ideation", "anxiety and depression", and "anhedonia". The agreement between the two scales was generally fair at different cutoff scores with the highest Cohen's kappa being 0.46.

Conclusions—Both the PHQ-9 and EPDS are reliable and valid scales for antepartum depression assessment. The PHQ-9 captures somatic symptoms, while EPDS detects depressive symptoms comorbid with anxiety during early pregnancy. Our findings suggest simultaneous administration of both scales may improve identification of antepartum depressive disorders in clinical settings.

Keywords

1. Background

Depression is one of the leading causes of disability and the largest contributor to the global burden of disease (Murray et al., 2012). In 2010, depression affected over 298 million individuals (Murray et al., 2012) with the highest proportion of cases being those between 25 and 34 years of age (Ferrari et al., 2013). Depression in women during the childbearing age, especially those living in low- and middle-income countries (LMICs), is twice as common as compared to men (Stewart et al., 2003).

Antepartum depression is a unipolar, non-psychotic depressive episode of mild, moderate or severe, beginning in or extending into pregnancy (Gibson et al., 2009, Töreki et al., 2013, Campagne, 2004). The prevalence of antepartum depression ranges from 10% to 41% in LMICs (Organization, 2008). In a recent study, 41% of Peruvian women reported moderate to severe depressive symptoms during their entire pregnancy (Cripe et al., 2010). Depressive disorders during pregnancy are associated with adverse obstetric (de Paz et al., 2011, Qiu et al., 2007, Sanchez et al., 2013) and neonatal outcomes (Chung et al., 2001, Lou et al., 1994, Zuckerman et al., 1989). Depressed women are also less likely to attend antenatal clinics and have an increased risk of alcohol, tobacco and drugs use, poor weight gain (Zuckerman et al., 1989), and postpartum depression (Sidebottom et al., 2012).

Early detection of depression during pregnancy and effective interventions intended to manage it are critical to prevent adverse outcomes (Adewuya et al., 2006). However, depression during pregnancy is often under recognized (Felice et al., 2006) and untreated (Sidebottom et al., 2012) by health professionals especially in LMICs. Systematic use of brief screening scales during antenatal care has been one suggested approach (Buist et al., 2002). Although there is no clear consensus as to which scale is recommended for screening antepartum depression (Flynn et al., 2011), two widely utilized screening scales are the Patient Health Questionnaire (PHQ-9) and the Edinburgh Postnatal Depression Scale (EPDS). Few studies, mostly conducted in North America, have compared the psychometric properties of PHQ-9 and EPDS (Flynn et al., 2011, Hanusa et al., 2008, Weobong et al., 2009, Yawn et al., 2009) as used during postnatal period. To date, only one study (Flynn et al., 2011) has evaluated the comparability of the two scales when used for depression

screening during pregnancy. Of the studies that compared the two scales during postpartum period, most investigators found both scales to be equally reliable (Flynn et al., 2011, Yawn et al., 2009). Weobong (Weobong et al., 2009) noted PHQ-9 was superior to EPDS among Ghanaian women while Hanusa (Hanusa et al., 2008) found EPDS was more accurate in detecting postpartum depression in the US.

Given the high prevalence of depressive symptoms during pregnancy in Peru (Cripe et al., 2010) and the lack of studies evaluating the comparability of the two scales during pregnancy, we conducted the present study to compare the psychometric properties of PHQ-9 and EPDS among pregnant Peruvian women. Specifically, as an initial step towards developing clinical intervention studies to mitigate the burden of depressive disorders among pregnant women in Peru, we sought to evaluate the construct validity of the two scales by evaluating their factor structures. We further sought to investigate the concurrent validity between two scales using factor structures, correlations, and Cohen's kappa at different cutoff scores. We expect that results from this study will lay the groundwork for establishing valid and broadly applicable depression and depressive symptom severity screening scales for use in Peruvian and possibly other South American prenatal clinics.

2. Methods

2.1 Study Population

This was a cross sectional study. We recruited women who attended the Instituto Nacional Materno Perinatal (INMP) in Lima, Peru, for their first prenatal care visit from February 2012 to March 2013. The INMP is the main reference establishment for maternal and perinatal care operated by the Ministry of Health of the Peruvian government. Eligible participants were pregnant women who were 18–49 years of age, under 16 weeks of gestation, and who spoke and understood Spanish. All participants were provided written informed consent. All procedures used in this study were approved by the institutional review boards of the INMP, Lima, Peru and the Harvard School of Public Health Office of Human Research Administration, Boston, MA.

2.2 Data Collection and Variable Specification

Each participant was interviewed by trained research personnel using a structured questionnaire in a private setting. The structured interview was adopted to elicit information regarding maternal socio-demographic, lifestyle characteristics, medical and reproductive histories and experience with symptoms of depression. A trained midwife conducted anthropometric measurements. Of the 1,810 women approached, 1,556 completed the interview. Due to missing information on the PHQ-9 and EPDS scales, a total of 39 participants were excluded. The final analyzed sample included 1,517 women.

Information of socio-demographic characteristics was collected during a structured interview. Age of participants was categorized as 18–20, 20–29, 30–34, and 35 years. Educational attainment was categorized as 6, 7–12, and >12 completed years of schooling. Other social-demographic variables were categorized as: marital status (married, living with husband vs. others), employment status (employed vs. not employed), difficulty paying for medical care and very basics (very hard or hard; somewhat hard; not very hard), parity

(nulliparous vs. multiparous), gravidity (primigravida vs. multigravida), race (Mestizo vs. others), and unplanned pregnancy (yes vs.no). Gestational age was based on the date of the last menstrual period and ultrasound assessment.

2.3 Scales

2.3.1 The Patient Health Questionnaire (PHQ-9)—The PHQ-9 is a nine-item depression screening scale (Kroenke et al., 2001, Spitzer et al., 1999, Kroenke and Spitzer, 2002) based on the criteria from the *Diagnostic and Statistical Manual of Mental Disorder-IV* (DSM-IV). Each item requires participants to rate the frequency of a depressive symptom experienced in the two weeks prior to evaluation. The PHQ-9 assesses nine depressive symptoms. These items include: 1) anhedonia, 2) depressed mood, 3) insomnia or hypersomnia, 4) fatigue or loss of energy, 5) appetite disturbances, 6) guilt or worthlessness, 7) diminished ability to think or concentrate, 8) psychomotor agitation or retardation, and 9) suicidal thoughts. The PHQ-9 score is calculated by assigning a score of 0, 1, 2, and 3, to the response categories of "not at all", "several days", "more than half the days" and "nearly every day" respectively with a total score ranging from 0 to 27. A score of   10 on the PHQ-9 is associated with 88 % sensitivity and 88% specificity in diagnosing possible major depressive disorder (MDD) using the DSM-IV criteria (Kroenke et al., 2001). Therefore, we defined presence of depression as a score   10 on PHQ-9 scale (Kroenke and Spitzer, 2002). Additionally, we categorized participants as exhibiting minimal (PHQ-9 score 0–4), mild (PHQ-9 score 5–9), moderate (PHQ-9 score 10–14), and moderately severe/severe (PHQ-9 score   15) depressive symptoms (Kroenke et al., 2001).

2.3.2 The Edinburgh Postnatal Depression Scale (EPDS)—The EPDS is a 10-item widely used screening scale for antepartum and postpartum depression (Cox et al., 1987). The EPDS items ask women to rate how they have felt in the previous seven days. The items include 1) ability to laugh, 2) anhedonia, 3) guilt, 4) anxiety, 5) panic attacks, 6) overwhelmed, 7) sleep disorders, 8) sadness, 9) tearfulness and 10) suicidal ideas. EPDS deemphasizes the somatic symptoms, such as changes in sleep and appetite, along with loss of energy (Eberhard-Gran et al., 2001, Flynn et al., 2011, Hanusa et al., 2008). Response categories are scored 0, 1, 2 and 3 for each item according to increased severity of the symptoms. Items 3 and 5–10 are reversed scored. Individual items are totaled to give an overall score ranging from 0 to 30. Prior validation studies suggest a cutoff score of 10 or higher for possible depressive disorder (Cox et al., 1987, Gibson et al., 2009). Particularly, it has been validated among postpartum Peruvian women. The sensitivity was 89.47% and the specificity 51.32% with a cutoff score of 10.5 (Vega-Dienstmaier et al., 2002). We categorized participants depression severity at cutoff scores of 10, 12 and 16, representing the thresholds for "normal", "slightly increased risk", "increased risk" and "marked risk" for depression (Cox et al., 1996, Yawn et al., 2009).

2.4 Statistical Analysis

First, we assessed the internal consistency of the PHQ-9 and EPDS using the Cronbach's alpha and inter-item correlations. Given the non-normal distributions, we used Spearman correlations to assess the direction and strength of correlation between the PHQ-9 and EPDS total scores. Next, we explored the factor structures of the two scales separately and

combined using exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). Prior to performing EFA, we assessed the suitability of the data for performing factor analysis. This analysis showed that it was appropriate to proceed with factor analysis (Bartlett's test of sphericity, *p*-value < 0.001 for both scales; and the Kaiser-Meyer-Olkin measure of sampling adequacy, 0.87 for PHQ-9 and 0.88 for EPDS). Then, we conducted the EFA using principal component analysis with orthogonal rotation. We used the scree plot, presenting eigenvalues associated with each factor, to identify the number of meaningful factors. Factors with relatively large eigenvalues (greater than one) were assumed to be meaningful and were retained for rotation (Kaiser, 1960). We used factor loadings of 0.4 or greater in the factor designation. To complement our EFA approaches and to evaluate the fit of the factor models identified in the literature, we conducted CFA using maximum likelihood estimation approach. We calculated the likelihood ratio chi-square (χ^2) statistics, standardized root mean square residual (SRMR), root mean square error of approximation (RMSEA) along with 90% confidence interval (90% CI), and comparative fit index (CFI) to evaluate model fit (Brown, 2006). Brown (Brown, 2006) recommended that the following criteria were evidence for reasonably good fit: 1) SRMR values close to 0.08 or below; 2) RMSEA values close to 0.06 or below; 3) CFI close to 0.95 or greater. We assessed the agreement between two scales at different cutoff scores using Cohen's kappa. All statistical analyses were performed using SAS 9.3 (SAS Institute, Cary, NC, USA). The level of statistical significance was set at *p*-values <0.05 and all tests were two-sided.

3. Results

3.1 Participant Characteristics

A summary of selected socio-demographic and reproductive characteristics of study participants is presented in Table 1. A total of 1,517 participants between the ages of 18 and 48 years (mean age=28.0 years, standard deviation=6.2 years) participated in the study. The majority of participants were Mestizos (75%), with at least 7 years of education (95%), multigravida (68%), and with an average gestational age of 9.8 (standard deviation = 3.4) weeks at interview. Approximately, half of participants were unemployed (56%), nulliparous (50%), and reported that the index pregnancy was unplanned (58%).

3.2 Reliability

The internal consistency of PHQ-9 gave a Cronbach's alpha of 0.81. The correlations between nine items of the PHQ-9 and the total scores ranged from 0.41 to 0.74 (all *p*-values <0.0001). The Cronbach's alpha for the EPDS was 0.82. The correlations between ten items of the EPDS and the total scores ranged from 0.50 to 0.75 (all *p*-values <0.0001).

3.3 Correlation

The Spearman correlation coefficient for PHQ-9 and EPDS scores was rho=0.52 (*p*-value <0.0001).

3.4 Factor Analyses

The results obtained from the EFA for PHQ-9 indicated a two-factor solution with eigenvalues of 3.64 and 1.13, corresponding to a somatic factor and a non-somatic factor

(Table 2). These two factors explained 53% of the total variance. Similarly, the EFA of EPDS resulted in two factors with eigenvalues of 3.94 and 1.06, explaining 50% of the total variance. Items of these two factors were characterized as "anxiety and depression" and "anhedonia". We performed an EFA including all 19 items from both PHQ-9 and EPDS (Table 3). It yielded four factors with eigenvalues 6.09, 1.98, 1.08 and 1.04. They together explained 54% of the total variance. Items from each scale loaded on four separate factors: "somatization", "anxiety and depression", "depressed mood and suicidal ideation "and "anhedonia."

The results of our CFA (Supplement 1) demonstrated a good fit of PHQ-9 scale (SRMR=0.040; RMSEA=0.069) and corroborated the results of our EFA where five somatic and four non-somatic items were loaded together. For the EPDS scale, a three-factor model with two anhedonia items, four anxiety items and four depressed mood items stood out as the best fitting model in the CFA (SRMR=0.022; RMSEA=0.034). The one-factor structure, which had the highest RMSEA and lowest CFI, did not fit current data well for both scales.

3.5 Severity Ratings

The mean score of PHQ-9 was 7.8 (standard deviation = 5.4) while the mean score of EPDS was 6.7 (standard deviation = 5.6) (Table 4). The PHQ-9 scale identified 29% of participants with possible depressive disorder (score 10) while 28% of women were identified as depressed using EPDS scale (score 10). Approximately 4% of participants were classified as having symptoms consistent with a diagnosis of "severe depression" on the PHQ-9 scale. In comparison, 8% of participants were classified as having symptoms consistent with being at "marked risk" of depression on the EPDS scale.

3.6 Agreement at Different Cutoff scores for Depression

When both PHQ-9 and EPDS scores were categorized into two categories (10 or <10), 1,119 (74%) women had concordant classifications (Table 5). Of note, 236 (16%) women were classified as depressed on both PHQ-9 and EPDS scales, while 883 (58%) were classified as non-depressed on both scales. The agreement (Cohen's kappa) between two scales was 0.36 (95% CI: 0.31, 0.41).

When considering the severity scores (Supplement 2), the rate of concordance was 64% (n=973). In particular, 108 (7%) women in the "moderately severe depression" or "severe depression" range (score 15) on PHQ-9 scale were in the normal range (score <10) of the EPDS scale. Additionally, 59 (4%) women had PHQ-9 scores in the normal range (score <10) but their EPDS scores were in the range of "increased risk" or "marked risk" (score 12) of depression. The agreement (weighted Cohen's kappa) of two scales using severity categories was 0.35 (95% CI: 0.31, 0.40).

Finally, we examined the agreement between the PHQ-9 and EPDS scores at different cutoff scores. Based on the recommended cutoff scores (10) for both scales, the agreement of two scale was fair (Cohen's kappa = 0.36). The best agreement between PHQ-9 and EPDS was Cohen's kappa = 0.46 at PHQ-9 19 and EPDS 19 (Supplement 3).

4. Discussion

Overall, the Spanish language versions of PHQ-9 and EPDS demonstrated high degree of internal consistency, and good construct and concurrent validity when administered to Peruvian women during early pregnancy. The EFA results showed that the PHQ-9 had a 2-factor model of "somatic" and "non-somatic" factors. Similarly, the EPDS resulted in 2-factor structure of "anxiety and depression" and "anhedonia" factors. Finally, factor analysis performed on the pooled items from the two depression scales together gave four factors.

A cutoff score of 10 on PHQ-9 identified 29% of pregnant women as depressed while the same cutoff score on EPDS detected 28% of pregnant women as depressed. Both the PHQ-9 and EPDS identified similar proportions of participants as non-depressed. However, the two scales were discordant in identifying those with mild to moderate depressive disorders. We also examined the agreement between two scales at different cutoff scores. Based on the recommended cutoff scores for these two scales (10), the agreement between them was fair with Cohen's kappa = 0.36. The highest agreement was Cohen's kappa = 0.46 with cutoff scores at 19 for both scales. The Cohen's kappa was still fair when changing cutoff scores confirming that two scales were comparable but did not completely identify the same patients. Although our study has not assessed the agreement of the scales with a diagnostic gold standard, prior studies that compared the PHQ-9 with Structured Clinical Interview for Depression (SCID) noted the Cohen's kappa ranged from 0.40 to 0.88 (Thompson et al., 2011, Fann et al., 2005, Bian et al., 2011). Similarly, the agreement between EPDS and SCID has been reported to range from 0.46 to 0.79 (Brandon et al., 2008, Rochat et al., 2013). The agreement derived from our study, compared with these Cohen's kappa, between the PHQ-9 and EPDS in this study is acceptable (White et al., 2008).

The following explanations may account for the observed variation between two scales. First, EPDS, unlike the PHQ-9, does not include somatic symptoms. Somatic symptoms are diagnostic symptoms for MDD. However, postpartum somatic symptoms overlap with depressive somatic symptoms (Yawn et al., 2009). Some investigators have suggested that exclusion of these items may lessen the confounding by somatic symptoms (Yonkers et al., 2009). For this reason, the EPDS does not include somatic symptoms, in part, to avoid the over-diagnosis perinatal depression (Yawn et al., 2009). In this study, a somatic factor was consistently extracted from the EFA whether the PHQ-9 scale was tested alone (explaining 40% of the total variance) or when the two scales were combined (explaining 32% of the total variance). Such large proportions of variance explained by the somatic factor may be attributed to the predominance of somatic symptoms in early pregnancy. Investigative teams have shown that the endorsement rates of most somatic symptoms, such as changes of sleep, energy, and appetite, are considerably higher in the first trimester compared to later trimester (Yonkers et al., 2009). It is important to note that the results of our factor analysis suggest that somatic symptoms play an important role in identifying antepartum depressive symptoms, especially in early pregnancy. Inclusion of somatic items results in lengthy scales (Zimmerman et al., 2006). It also may increase the false positive rate and thus may over identify women with possible depressive disorders. However, women who respond affirmatively to cardinal symptoms (depressed mood and anhedonia) for MDD are more likely to endorse somatic symptoms compared to those who are do not endorse the cardinal

symptoms (Yonkers et al., 2009). The presence of a somatic factor does not suggest they should be deleted (Krause et al., 2008). Of importance, elimination of those items may result in a loss of potential relevant information, such as depression severity (Yonkers et al., 2009).

Second, items on EPDS are selected from several depression and anxiety scales, such as hospital anxiety and depression scale (HADS) (Zigmond and Snaith, 1983) and the irritability, depression and anxiety scale (Cox et al., 1987, King, 2012, Snaith et al., 1978). However, PHQ-9 is developed strictly based on the DSM-IV criteria. EPDS does not exclusively measure depression (Brouwers et al., 2001, Navarro et al., 2007). The existence of an anxiety subscale has been confirmed in other studies (Brouwers et al., 2001, Green, 1998, King, 2012, Navarro et al., 2007, Pop et al., 1992, Tuohy and McVey, 2008). Additionally, significantly positive correlations between anxiety items (item 3 "guilt", item 4 "anxiety" and item 5 "panic attacks") with other anxiety scales such as State Trait Anxiety Inventory (STAI) (Brouwers et al., 2001) and HADS (Tuohy and McVey, 2008) are found.

Furthermore, the total score of EPDS demonstrates a higher correlation with STAI compared with the anxiety scale of EPDS [50]. Our EFA extracted a two-factor structure where one of the factors was "depression and anxiety". However, our CFA confirmed that a three-factor ("anhedonia", "anxiety" and "depression") structure had a better fit. Studies have shown that depression and anxiety are common comorbid disorders and more stable than anxiety alone or depression alone (Merikangas et al., 2003). Anxiety is a prominent feature of postpartum depression (Hendrick et al., 2000, Misri et al., 2000, Ross et al., 2003, Stuart et al., 1998). Nearly half of clinically depressed pregnant and postpartum women have clinically significant comorbid anxiety (Ross et al., 2003). Compared to women with non-perinatal forms of depression, women with perinatal depression complain about anxiety more frequently (Heron et al., 2004, Ross et al., 2003, Teixeira et al., 2009). Bowen et al in their study of Canadian pregnant women showed that the anxiety factor of EPDS accounts for the greatest variance of the overall score (Bowen et al., 2008). The symptoms of anxiety are particularly relevant if EPDS is administered prenatally (Ross et al., 2003). The proportion that the anxiety subscale contributes to total EPDS score is statistically significantly higher in pregnancy than in postpartum. On balance, available evidence suggests that anxiety may be an important feature of antenatal depression, which may be a distinction between antenatal depression and MDD in non-pregnant women. Of note, EPDS would be helpful in detecting depression comorbid with anxiety. However, the routine use of EPDS should be accompanied by further evaluation of the contribution of depression and anxiety subscales, separately (Tuohy and McVey, 2008).

Despite the observed differences, our study results provided strong evidence that the PHQ-9 and EPDS were comparable, but not identical, in identifying antepartum depressive symptoms. However, they captured two distinct aspects of antepartum depressive symptoms. The PHQ-9 was more likely to select women with somatic symptoms while the EPDS was likely to detect depressed women with comorbid anxiety during pregnancy.

To date, we are aware of only one published study (Flynn et al., 2011) that evaluated the comparison of the two scales in pregnancy. In their retrospective study of pregnant and postnatal women seeking psychiatric services, Flynn et al (Flynn et al., 2011) found Cohen's

kappa = 0.58 for pregnant women at cutoff scores of EPDS 13 and PHQ-9 10. Compared with Flynn's results, the Cohen's kappa in our study was generally lower. The timing of assessment may impact the prevalence rate of depression (Hewitt et al., 2009). The average gestational age of women in Flynn's sample was 21 weeks (standard deviation = 9). However, in our study, the average gestational age was only 9.8 weeks (standard deviation = 3.4). As gestational age increases, the prevalence of depression among pregnant women also increases (Bennett et al., 2004). Furthermore, the sample of Flynn's study was from women who were seeking psychiatric services. Hence, participants were more likely to experience depression compared with women who attended routine antenatal care clinics. Cohen's kappa has been criticized because it is dependent on the prevalence of disease although it is the most widely used measure of concordance in validity studies of psychiatric disorders (Byrt et al., 1993, Cook, 2005, Haro et al., 2006, Kraemer et al., 2003).

Results from our study must be interpreted while considering several possible limitations. First, we did not have information from diagnostic interviews, such as the SCID or Schedules for Clinical Assessment of Neuropsychiatry (SCAN), for depression. Consequently, we are not able to report validity measures of the two screening scales and make comparison between PHQ-9 and EPDS regarding the criterion validity. Second, all participants completed the two scales in the same order (first administered PHQ-9 and then EPDS) during the same research interview session. Hence, participants may have a better understanding of EPDS based on immediately previous experience of responding to questions pertaining to the PHQ-9. Future efforts designed to control for possible bias known as the "order effect", should randomize the order in which the two scales are administered (Weobong et al., 2009). Third, we performed this analysis within the framework of classic test theory (CTT). Further analysis comparing two scales needs to be complemented with Item Response Theory (IRT) approaches to evaluate the extent to which items may differ by respondents age and race groups. Last, current data are cross-sectional. The validity of items from two scales, especially somatic items, can be influenced by gestational age and delivery. Longitudinal data would be helpful in understanding the agreement between two scales across pregnancy trimesters.

To the best of our knowledge, this study is the first to rigorously compare the performance of the Spanish language versions the PHQ-9 and the EPDS for screening antepartum depression in early pregnancy. In addition, the performance of the PHQ-9 and EPDS has not been compared among women in a LMIC with such a large sample size. Furthermore, the evidence for the construct and concurrent validity of two scales based on both EFA and CFA has been extended.

5. Conclusion

In conclusion, the PHQ-9 and EPDS demonstrate good internal consistency, construct and concurrent validity in screening antepartum depression during early pregnancy. Similar proportions of women are identified as depressed on both scales. Although PHQ-9 and EPDS measure depression, they capture two distinct aspects of antepartum depression in early pregnancy. For example, the PHQ-9 assesses common somatic symptoms during pregnancy while the EPDS detects depressive symptoms comorbid with anxiety. Our

findings suggest that using the two scales together may improve the identification of women who are at risk of antepartum depression. However, if rapid screening for depression is required in busy clinical settings, PHQ-9, which can be completed and scored in few minutes (Kroenke et al., 2001, Spitzer et al., 1999), is helpful. Moreover, the PHQ-9 is strictly derived from DSM-IV criteria, which accounts for somatic symptoms. If one considers capturing depressed women with comorbid anxiety, which is a feature of antepartum depression, using EPDS with further calculation of the scores of depression subscale and anxiety subscale would be a choice. Future studies designed to determine the extent to which scores from each scales are predictive of adverse maternal and perinatal outcomes are needed. Longitudinal data of these two scales during and after pregnancy are warranted.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Adewuya AO, Ola BA, Dada AO, Fasoto OO. Validation of the Edinburgh Postnatal Depression Scale as a screening tool for depression in late pregnancy among Nigerian women. J Psychosom Obstet Gynecol. 2006; 27:267–272.

Bennett HA, Einarson A, Taddio A, Koren G, Einarson TR. Prevalence of depression during pregnancy: systematic review. Obstet Gynecol. 2004; 103:698–709. [PubMed: 15051562]

Bian C, Li C, Duan Q, Wu H. Reliability and validity of patient health questionnaire: Depressive syndrome module for outpatients. Sci Res Essays. 2011; 6:278–282.

Bowen A, Bowen R, Maslany G, Muhajarine N. Anxiety in a Socially High-Risk Sample of Pregnant Women in Canada. Can J Psychiatry. 2008; 53:435–440. [PubMed: 18674401]

Brandon AR, Trivedi MH, Hynan LS, Miltenberger PD, Labat DB, Rifkin JB, Stringer CA. Prenatal depression in women hospitalized for obstetric risk. J Clin Psychiatry. 2008; 69:635. [PubMed: 18312059]

Brouwers EP, Van Baar AL, Pop VJ. Does the Edinburgh Postnatal Depression Scale measure anxiety? J Psychosom Res. 2001; 51:659–663. [PubMed: 11728506]

Brown, TA. Confirmatory Factor Analysis for Applied Research. Guilford Press; New York, NY: 2006.

Buist AE, Barnett BE, Milgrom J, Pope S, Condon JT, Ellwood DA, Boyce PM, Austin MPV, Hayes BA. To screen or not to screen-that is the question in perinatal depression. Med J Aust. 2002; 177:S101. [PubMed: 12358566]

Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. J Clin Epidemiol. 1993; 46:423–429. [PubMed: 8501467]

Campagne DM. The obstetrician and depression during pregnancy. Eur J Obstet Gynecol Reprod Biol. 2004; 116:125–130. [PubMed: 15358452]

Chung TK, Lau TK, Yip AS, Chiu HF, Lee DT. Antepartum depressive symptomatology is associated with adverse obstetric and neonatal outcomes. Psychosom Med. 2001; 63:830–834. [PubMed: 11573032]

Cook, RJ. Encyclopedia of Biostatistics. Wiley; New York, NY: 2005. Kappa and Its Dependence on Marginal Rates.

Cox JL, Chapman G, Murray D, Jones P. Validation of the Edinburgh postnatal depression scale (EPDS) in non-postnatal women. J Affect Disord. 1996; 39:185–189. [PubMed: 8856422]

Cox JL, Holden JM, Sagovsky R. Detection of postnatal depression. Development of the 10-item Edinburgh Postnatal Depression Scale. Br J Psychiatry. 1987; 150:782–786. [PubMed: 3651732]

Cripe SM, Sanchez S, Lam N, Sanchez E, Ojeda N, Tacuri S, Segura C, Williams MA. Depressive symptoms and migraine comorbidity among pregnant Peruvian women. J Affect Disord. 2010; 122:149–153. [PubMed: 19695709]

De Paz NC, Sanchez SE, Huaman LE, Chang GD, Pacora PN, Garcia PJ, Ananth CV, Qiu C, Williams MA. Risk of placental abruption in relation to maternal depressive, anxiety and stress symptoms. J Affect Disord. 2011; 130:280–284. [PubMed: 20692040]

Eberhard-Gran M, Eskild A, Tambs K, Opjordsmoen S, Samuelsen SO. Review of validation studies of the Edinburgh Postnatal Depression Scale. Acta Psychiatr Scand. 2001; 104:243–249. [PubMed: 11722298]

Fann JR, Bombardier CH, Dikmen S, Esselman P, Warms CA, Pelzer E, Rau H, Temkin N. Validity of the Patient Health Questionnaire-9 in assessing depression following traumatic brain injury. J Head Trauma Rehabil. 2005; 20:501–511. [PubMed: 16304487]

Felice E, Saliba J, Grech V, Cox J. Validation of the Maltese version of the Edinburgh Postnatal Depression Scale. Arch Womens Ment Health. 2006; 9:75–80. [PubMed: 16172837]

Ferrari AJ, Charlson FJ, Norman RE, Flaxman AD, Patten SB, Vos T, Whiteford HA. The epidemiological modelling of major depressive disorder: application for the Global Burden of Disease Study 2010. PloS one. 2013; 8:e69637. [PubMed: 23922765]

Flynn HA, Sexton M, Ratliff S, Porter K, Zivin K. Comparative performance of the Edinburgh Postnatal Depression Scale and the Patient Health Questionnaire-9 in pregnant and postpartum women seeking psychiatric services. Psychiatry Res. 2011; 187:130–134. [PubMed: 21122923]

Gibson J, Mckenzie-Mcharg K, Shakespeare J, Price J, Gray R. A systematic review of studies validating the Edinburgh Postnatal Depression Scale in antepartum and postpartum women. Acta Psychiatr Scand. 2009; 119:350–364. [PubMed: 19298573]

Green JM. Postnatal depression or perinatal dysphoria? Findings from a longitudinal community-based study using the Edinburgh Postnatal Depression Scale. J Reprod Infant Psychol. 1998; 16:143–155.

Hanusa BH, Scholle SH, Haskett RF, Spadaro K, Wisner KL. Screening for depression in the postpartum period: a comparison of three instruments. J Womens Health (Larchmt). 2008; 17:585–596. [PubMed: 18345995]

Haro JM, Arbabzadeh-Bouchez S, Brugha TS, De Girolamo G, Guyer ME, Jin R, Lepine JP, Mazzi F, Reneses B, Vilagut G, Sampson NA, Kessler RC. Concordance of the Composite International Diagnostic Interview Version 3.0 (CIDI 3.0) with standardized clinical assessments in the WHO World Mental Health surveys. Int J Methods Psychiatr Res. 2006; 15:167–180. [PubMed: 17266013]

Hendrick V, Altshuler L, Strouse T, Grosser S. Postpartum and nonpostpartum depression: differences in presentation and response to pharmacologic treatment. Depress Anxiety. 2000; 11:66–72. [PubMed: 10812531]

Heron J, O'connor TG, Evans J, Golding J, Glover V, Team AS. The course of anxiety and depression through pregnancy and the postpartum in a community sample. J Affect Disord. 2004; 80:65–73. [PubMed: 15094259]

Hewitt C, Gilbody S, Brealey S, Paulden M, Palmer S, Mann R, Green J, Morrell J, Barkham M, Light K, Richards D. Methods to identify postnatal depression in primary care: an integrated evidence synthesis and value of information analysis. Health Technol Assess. 2009; 13:1–145. 147–230. [PubMed: 19624978]

Kaiser HF. The application of electronic computers to factor analysis. Educ Psychol Meas. 1960; 20:141–151.

King, PaL. Replicability of structural models of the Edinburgh Postnatal Depression Scale (EPDS) in a community sample of postpartum African American women with low socioeconomic status. Arch Womens Ment Health. 2012; 15:77–86. [PubMed: 22297555]

Kraemer HC, Morgan GA, Leech NL, Gliner JA, Vaske JJ, Harmon RJ. Measures of clinical significance. J Am Acad Child Adolesc Psychiatry. 2003; 42:1524–1529. [PubMed: 14627890]

Krause JS, Bombardier C, Carter RE. Assessment of depressive symptoms during inpatient rehabilitation for spinal cord injury: Is there an underlying somatic factor when using the PHQ? Rehabil Psychol. 2008; 53:513–520.

Kroenke K, Spitzer RL. The PHQ-9: a new depression diagnostic and severity measure. Psychiatr Ann. 2002; 32:1–7.

Kroenke K, Spitzer RL, Williams JBW. The PHQ-9. J Gen Intern Med. 2001; 16:606–613. [PubMed: 11556941]

Lou HC, Hansen D, Nordentoft M, Pryds O, Jensen F, Nim J, Hemmingsen R. Prenatal stressors of human life affect fetal brain development. Dev Med Child Neurol. 1994; 36:826–832. [PubMed: 7926332]

Merikangas K, Zhang H, Avenevoli S, Acharyya S, Neuenschwander M, Angst J. Longitudinal trajectories of depression and anxiety in a prospective community study: The zurich cohort study. Arch Gen Psychiatry. 2003; 60:993–000. [PubMed: 14557144]

Misri S, Kim J, Riggs KW, Kostaras X. Paroxetine levels in postpartum depressed women, breast milk, and infant serum. J Clin Psychiatry. 2000; 61:828–832. [PubMed: 11105735]

Murray CJL, Vos T, Lozano R, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. The Lancet. 2012; 380:2197–223.

Navarro P, Ascaso C, Garcia-Esteve L, Aguado J, Torres A, Martín-Santos R. Postnatal psychiatric morbidity: a validation study of the GHQ-12 and the EPDS as screening tools. Gen Hosp Psychiat. 2007; 29:1–7.

World Health Organization. Maternal Mental Health and Child Health and Development in Low and Middle Income Countries: Report of the Meeting; Geneva, Switzerland. January 30–February 1, 2008; Geneva: World Health Organization Press; 2008.

Pop VJ, Komproe IH, Van Son MJ. Characteristics of the Edinburgh Post Natal Depression Scale in The Netherlands. J Affect Disord. 1992; 26:105–110. [PubMed: 1447427]

Qiu C, Sanchez SE, Lam N, Garcia P, Williams MA. Associations of depression and depressive symptoms with preeclampsia: results from a Peruvian case-control study. BMC Womens Health. 2007; 7:15. [PubMed: 17900360]

Rochat TJ, Tomlinson M, Newell ML, Stein A. Detection of antenatal depression in rural HIV-affected populations with short and ultrashort versions of the Edinburgh Postnatal Depression Scale (EPDS). Arch Womens Ment Health. 2013; 16:401–410. [PubMed: 23615932]

Ross LE, Gilbert Evans SE, Sellers EM, Romach MK. Measurement issues in postpartum depression part 1: anxiety as a feature of postpartum depression. Arch Womens Ment Health. 2003; 6:51–57. [PubMed: 12715264]

Sanchez SE, Puente GC, Atencio G, Qiu C, Yanez D, Gelaye B, Williams MA. Risk of spontaneous preterm birth in relation to maternal depressive, anxiety, and stress symptoms. J Reprod Med. 2013; 58:25–33. [PubMed: 23447915]

Sidebottom AC, Harrison PA, Godecker A, Kim H. Validation of the Patient Health Questionnaire (PHQ)-9 for prenatal depression screening. Arch Womens Ment Health. 2012; 15:367–374. [PubMed: 22983357]

Snaith RP, Constantopoulos AA, Jardine MY, Mcguffin P. A clinical scale for the self-assessment of irritability. Br J Psychiatry. 1978; 132:164–171. [PubMed: 623950]

Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders Patient Health Questionnaire. JAMA. 1999; 282:1737–1744. [PubMed: 10568646]

Stewart, DE.; Robertson, E.; Dennis, C-L.; Grace, SL.; Wallington, T. Postpartum depression: literature review of risk factors and interventions. University Health Network Women's Health Program for Toronto Public Health; Toronto: 2003.

Stuart S, Couser G, Schilder K, O'hara MW, Gorman L. Postpartum anxiety and depression: onset and comorbidity in a community sample. J Nerv Ment Dis. 1998; 186:420–424. [PubMed: 9680043]

Teixeira C, Figueiredo B, Conde A, Pacheco A, Costa R. Anxiety and depression during pregnancy in women and men. J Affect Disord. 2009; 119:142–148. [PubMed: 19346001]

Thompson AW, Liu H, Hays RD, Katon WJ, Rausch R, Diaz N, Jacob EL, Vassar SD, Vickrey BG. Diagnostic accuracy and agreement across three depression assessment measures for Parkinson's disease. Parkinsonism Relat Disord. 2011; 17:40–45. [PubMed: 21084211]

Töreki A, Andó B, Keresztúri A, Sikovanyecz J, Dudas RB, Janka Z, Kozinszky Z, Pál A. The Edinburgh Postnatal Depression Scale: translation and antepartum validation for a Hungarian sample. Midwifery. 2013; 29:308–315. [PubMed: 22417756]

Tuohy A, Mcvey C. Subscales measuring symptoms of non-specific depression, anhedonia, and anxiety in the Edinburgh Postnatal Depression Scale. Br J Clin Psychol. 2008; 47:153–169. [PubMed: 17761026]

Vega-Dienstmaier JM, Mazzotti Suárez G, Campos Sánchez M. Validation of a Spanish version of the Edinburgh Postnatal Depression Scale. Actas Esp Psiquiatr. 2002; 30:106–111. [PubMed: 12028943]

Weobong B, Akpalu B, Doku V, Owusu-Agyei S, Hurt L, Kirkwood B, Prince M. The comparative validity of screening scales for postnatal common mental disorder in Kintampo, Ghana. J Affect Disord. 2009; 113:109–117. [PubMed: 18614241]

White, E.; Armstrong, BK.; Saracci, R.; Armstrong, B. Principles of exposure measurement in epidemiology: collecting, evaluating, and improving measures of disease risk factors. Oxford University Press; Oxford: 2008.

Yawn BP, Pace W, Wollan PC, Bertram S, Kurland M, Graham D, Dietrich A. Concordance of Edinburgh Postnatal Depression Scale (EPDS) and Patient Health Questionnaire (PHQ-9) to assess increased risk of depression among postpartum women. J Am Board Fam Med. 2009; 22:483–491. [PubMed: 19734393]

Yonkers KA, Smith MV, Gotman N, Belanger K. Typical somatic symptoms of pregnancy and their impact on a diagnosis of major depressive disorder. Gen Hosp Psychiatr. 2009; 31:327–333.

Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand. 1983; 67:361–370. [PubMed: 6880820]

Zimmerman M, Chelminski I, Mcglinchey JB, Young D. Diagnosing Major Depressive Disorder X: Can the Utility of the DSM-IV Symptom Criteria Be Improved? J Nerv Ment Dis. 2006; 194:893–897. [PubMed: 17164626]

Zuckerman B, Amaro H, Bauchner H, Cabral H. Depressive symptoms during pregnancy: relationship to poor health behaviors. Am J Obstet Gynecol. 1989; 160:1107–1111. [PubMed: 2729387]

Table 1

Socio-demographics and reproductive characteristics of the study population[a]

Characteristic	Entire population (N=1517)	
	n	%
Age (mean ± standard deviation)	28.0 ± 6.2	
Age (years)		
18–20	91	6.0
20–29	856	56.4
30–34	305	20.1
35	265	17.5
Education (years)		
6	68	4.5
7–12	851	56.1
>12	593	39.1
Married, living with husband	275	18.1
Unemployed	855	56.4
How hard to pay for medical care		
Very hard/hard	320	21.1
Somewhat hard	506	33.4
Not very hard	688	45.4
How hard to pay for the very basics		
Very hard/hard	286	18.9
Somewhat hard	492	32.4
Not very hard	738	48.7
Nulliparous	765	50.4
Primagravida	486	32.0
Mestizo	1142	75.3
Unplanned pregnancy	875	57.7
Gestational age at interview (mean ± standard deviation)	9.8 ± 3.4	

[a]Due to missing data, percentages may not add up to 100%.

Table 2

Item-level factor loadings resulting from separate exploratory factor analysis for PHQ-9 and EPDS [a]

Scale and item	Factor 1	Factor 2
PHQ-9		
1. Little interest or pleasure in doing things.	**0.76**	0.12
2. Feeling down, depressed, or hopeless.	**0.55**	**0.51**
3. Trouble falling or staying asleep, or sleeping too much.	**0.69**	0.17
4. Feeling tired or having little energy.	**0.80**	0.15
5. Poor appetite or overeating.	**0.61**	0.10
6. Feeling bad about yourself, or that you are a failure, or have let yourself or your family down.	0.22	**0.70**
7. Trouble concentrating on things, such as reading the newspaper or watching television.	0.30	**0.59**
8. Moving or speaking so slowly that other people could have noticed. Or the opposite –being so fidgety or restless that you have been moving around a lot more than usual.	**0.59**	0.32
9. Thoughts that you would be better off dead, or of hurting yourself in some way.	−0.02	**0.81**
EPDS		
1. I have been able to laugh and see the funny side of things.	0.15	**0.85**
2. I have looked forward with enjoyment to things.	0.21	**0.81**
3. I have blamed myself unnecessarily when things went wrong.	**0.67**	0.12
4. I have been anxious or worried for no good reason.	**0.66**	0.10
5. I have felt scared or panicky for no very good reason.	**0.66**	0.04
6. Things have been getting on top of me.	**0.52**	0.15
7. I have been so unhappy. I have had difficulty sleeping.	**0.59**	0.34
8. I have felt sad and miserable.	**0.71**	0.22
9. I have been so unhappy that I have been crying.	**0.73**	0.29
10. The thought of harming myself has occurred to me.	**0.50**	0.20

[a] Two scales are analyzed in separate factor analysis. Factor loadings > 0.4 are in bold.

Table 3

Item-level factor loadings resulting from pooled exploratory factor analysis for PHQ-9 and EPDS[a]

Scale and Item	Somatization	Anxiety and Depression	Depressed Mood & Suicidal Ideation	Anhedonia
PHQ-9				
1. Little interest or pleasure in doing things.	**0.74**	0.10	−0.01	0.13
2. Feeling down, depressed, or hopeless.	**0.58**	0.32	0.22	0.22
3. Trouble falling or staying asleep, or sleeping too much.	**0.67**	0.13	0.10	0.07
4. Feeling tired or having little energy.	**0.78**	0.17	0.02	0.01
5. Poor appetite or overeating.	**0.61**	0.05	0.07	−0.02
6. Feeling bad about yourself, or that you are a failure, or have let yourself or your family down.	0.32	0.31	**0.43**	0.19
7. Trouble concentrating on things, such as reading the newspaper or watching television.	**0.43**	0.16	0.28	0.26
8. Moving or speaking so slowly that other people could have noticed. Or the opposite –being so fidgety or restless that you have been moving around a lot more than usual.	**0.66**	0.11	0.14	0.08
9. Thoughts that you would be better off dead, or of hurting yourself in some way.	0.14	0.13	**0.78**	0.09
EPDS				
1. I have been able to laugh and see the funny side of things.	0.16	0.13	0.11	**0.82**
2. I have looked forward with enjoyment to things.	0.07	0.20	0.18	**0.79**
3. I have blamed myself unnecessarily when things went wrong.	0.10	**0.59**	0.34	0.08
4. I have been anxious or worried for no good reason.	0.16	**0.67**	0.13	0.11
5. I have felt scared or panicky for no very good reason.	0.15	**0.68**	0.15	0.02
6. Things have been getting on top of me.	0.18	**0.64**	−0.05	0.16
7. I have been so unhappy. I have had difficulty sleeping.	0.15	**0.41**	0.39	0.30
8. I have felt sad and miserable.	0.05	**0.52**	**0.52**	0.16
9. I have been so unhappy that I have been crying.	0.22	**0.50**	**0.52**	0.22
10. The thought of harming myself has occurred to me.	0.04	0.08	**0.81**	0.07

[a]All items from both scales are analyzed in the same factor analysis. Factor loadings > 0.4 are in bold.

Table 4

Depression severity based on PHQ-9 and EPDS

	N=1517	
Depression Severity	**n**	**%**
PHQ-9		
None (0~4)	481	31.7
Mild depression (5~9)	590	38.9
Moderate depression (10~14)	258	17.0
Moderately severe depression (15~19)	127	8.4
Severe depression (20~27)	61	4.0
Total score (mean ± standard deviation)	7.8 ± 5.4	
EPDS		
0–9; "Normal"	1093	72.1
10, 11; "Slightly increased risk"	129	8.5
12–15; "Increased risk"	174	11.5
16; "Marked risk"	121	8.0
Total score (mean ± standard deviation)	6.7 ± 5.6	

Table 5

Comparison of PHQ-9 and EPDS in identifying possible depression[a]

EPDS	PHQ-9		
	Depression	**No Depression**	**Total**
Possible Depression	236	188	424
No Depression	210	883	1093
Total	446	1071	1517

[a]For both scales, the cutoff scores identifying possible depression are   10.