Journal of Affective Disorders 162 (2014) 1–7

Contents lists available at ScienceDirect

Journal of Affective Disorders

journal homepage: www.elsevier.com/locate/jad

Research report

Comparative performance of Patient Health Questionnaire-9 and Edinburgh Postnatal Depression Scale for screening antepartum depression

Qiuyue Zhong [a], Bizu Gelaye [a,*], Marta Rondon [b], Sixto E. Sánchez [c], Pedro J. García [c], Elena Sánchez [d], Yasmin V. Barrios [a], Gregory E. Simon [e], David C. Henderson [f], Swee May Cripe [g], Michelle A. Williams [a]

[a] Department of Epidemiology, Harvard School of Public Health, 677 Huntington Avenue, K505F, Boston, MA 02115, United States
[b] Department of Medicine, Cayetano Heredia Peruvian University, Lima, Peru
[c] Universidad San Martin de Porres, Lima, Peru
[d] Asociación Civil PROESA, Lima, Peru
[e] Group Health Research Institute, Seattle, WA, United States
[f] Department of Psychiatry, Massachusetts General Hospital, Boston, MA, United States
[g] School of Medicine, Perdana University, Kuala Lumpur, Malaysia

ARTICLE INFO

Article history:
Received 3 March 2014
Accepted 14 March 2014
Available online 26 March 2014

Keywords:
PHQ-9
EPDS
Antepartum depression
Peru

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 1517 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 (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.

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 (WHO, 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).

* Corresponding author. Tel.: +1 617 432 6477; fax: +1 617 566 7805.
E-mail address: bgelaye@hsph.harvard.edu (B. Gelaye).

http://dx.doi.org/10.1016/j.jad.2014.03.028
0165-0327/

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 et al. (2009) noted PHQ-9 was superior to EPDS among Ghanaian women while Hanusa et al. (2008) found EPDS was more accurate in detecting postpartum depression in the United States.

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 1810 women approached, 1556 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 1517 women.

Age of participants was categorized as 18–20, 20–29, 30–34, and $\geq$35 years. Educational attainment was categorized as $\leq$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 $\geq$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 the presence of depression as a score $\geq$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 $\geq$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).