



Published in final edited form as:

Arch Womens Ment Health. 2013 August ; 16(4): 271–277. doi:10.1007/s00737-013-0345-z.

Analysis of brief screening tools for the detection of postpartum depression: Comparisons of the PRAMS 6-item instrument, PHQ-9, and structured interviews

Kristina Davis, PhD^a, Teri Pearlstein, MD^{b,c}, Scott Stuart, MD^d, Michael O'Hara, PhD^d, and Caron Zlotnick, PhD^{a,b}

^aWomen and Infants Hospital, Providence, RI, USA

^bWarren Alpert Medical School of Brown University, Providence, RI, USA

^cWomen's Health Collaborative, a Lifespan partner, Providence, RI, USA

^dUniversity of Iowa, Iowa City, IA

Abstract

Purpose—Postpartum depression (PPD) is an important mental health issue affecting approximately 10% of women. Self-report screening measures represent utility for detecting PPD in both clinical and research settings. The current study sought to inspect the accuracy of two screening measures compared to clinical interviews.

Methods—As part of an ongoing clinical trial, 1392 women between the ages of 18 and 45 were screened for PPD using the Patient Health Questionnaire-9 (PHQ-9) and a six item scale developed from CDC Pregnancy Risk Assessment questions (PRAMS-6). Three item subscales of the PRAMS-6 were also inspected – three depression (PRAMS-3D) and three anxiety items (PRAMS-3A).

Results—Receiver Operating Characteristics compared the diagnostic accuracy of the PHQ-9, PRAMS-6, PRAMS-3D, and PRAMS-3A to both the Structured Clinical Interview for the DSM-IV (SCID) and the Hamilton Rating Scale for Depression. The PHQ-9, PRAMS-6, and PRAMS-3D all showed moderate accuracy at diagnosing PPD. Diagnostic cut points are provided.

Conclusions—The PRAMS-6 instrument is a brief and effective screening tool for PPD. The time frame of symptom assessment may account for some variability in accuracy between the PHQ-9 and PRAMS screening instruments.

Keywords

postpartum depression; depression screening; depression

Introduction

Postpartum depression (PPD) is an important mental health issue affecting approximately 7–13% of women (Leahy-Warren, McCarthy, & Corcoran, 2011). Untreated PPD results in significant detriments to multiple domains of infant development and family functioning (Armitage et al., 2009; Fishell, 2010). Therefore, proper detection of PPD is important to ensure that affected mothers are identified and referred to appropriate treatments.

Corresponding author: Kristina Davis, z_davisk2@yahoo.com, Phone: 512-897-0037.

Other authors report no conflicts of interest. The authors alone are responsible for the content and writing of this paper.

Additionally, research on the treatment and prevention of PPD relies heavily on screening for probable PPD for recruitment.

Although the DSM-IV defines PPD as a major depressive episode within one month after delivery, many researchers regard the six-month period after delivery as the period in which PPD can occur. Perinatal women often continue obstetric care for three to eight weeks after childbirth (Declercq, Sakala, Corry, & Applebaum, 2007), but may switch to primary care after this early post-birth care. A brief PPD instrument could be used at standard primary care patient visits without adding to the burden of care, which might increase the number of cases of PPD detected after obstetric care has ended. Well baby visits afford pediatricians an opportunity to screen mothers for PPD as well (Liberto, 2012, Heneghan, 2007). Therefore, a brief screening instrument expands the reach of obstetric, primary care, and pediatric providers in detecting suspected cases of PPD and referring those cases to appropriate care.

A variety of self-report screening measures for PPD are used in research and in clinical practice to detect probable cases of PPD. The most widely used self-report tools for detection of PPD are the Edinburgh Postnatal Depression Scale (EPDS) (Cox, 1987) and the Beck Depression Inventory (BDI; Hewitt, 2010) having 10 and 21 items respectively. The Patient Health Questionnaire (PHQ-9) is a nine item screening instrument for depression that is based on the DSM-IV diagnostic criteria, is commonly used in clinical practice to detect suspected cases of depression, and has been validated in perinatal populations (for review see Kroenke, 2010).

Both the EPDS and BDI appear to be accurate and effective screening tools for PPD. A recent study by Ji and colleagues (2011) looked at the accuracy of the BDI in 534 perinatal women over multiple time points ranging from pre-conception, first through third trimester, and early to late postpartum. The BDI was comparable to diagnostic interview for detecting PPD showing good accuracy with ROC values from 0.8 to 0.9 depending on the time period inspected (Ji, 2011). However, this study also found that the optimum cut-point for the BDI varied depending on the perinatal time frame. In a meta-analysis of seven studies, the EPDS showed positive predictive value of 62% in perinatal populations (Milgrom, Mendelsohn, and Gemmill, 2011). The authors of the meta-analysis concluded that the EPDS was an effective screening tool for PPD. Neither the EPDS nor the BDI is ultra-brief and may be too lengthy for quick screening in obstetric, primary care, or pediatric settings.

In adolescents and emerging adults, three-item and two-item subscales of the EPDS were compared to the full 10 items (Kabir, 2008). In this population, the three-item subscale, which measured anxiety symptoms, was a more accurate detector of PPD, but the ultra-brief subscale of the EPDS has not been validated in adult populations or perinatal populations to date.

The PHQ-9 has been validated in a variety of populations and languages (Arroll, 2010; Liu, 2011; Yeung, 2008; Merz, 2011; for review see Kroenke, 2010). Some alterations in diagnostic cut-points are necessary for the elderly with cognitive impairment (Boyle, 2010) and in adolescents (Richardson, 2010) indicating that alterations in its use might be warranted for other populations like perinatal women.

Studies comparing the PHQ-9 to the EPDS found that both measures exhibit similar performance in detecting PPD. One study found that the EPDS was more accurate at detecting PPD but that both the EPDS and the PHQ-9 had moderate accuracy in a sample of 135 women (Hanusa, 2008). Another study reported that the EPDS and PHQ-9 were concordant in 399 postpartum women (Yawn, 2009). In perinatal women (N = 185), the PHQ-9 and EPDS were moderately accurate at detecting PPD when compared to unstructured clinical diagnosis (Flynn, 2010).

A repeated measures study with the PHQ-9 for 506 women undergoing PPD treatment showed that the somatic items (e.g., sleep, appetite, and energy) were less effective predictors of PPD than the other items on that scale (Gjerdingen, Crow, McGovern, Miner, & Center, 2011). The items on the PHQ-9 cover all the diagnostic symptoms of depression; however, somatic symptoms may be common for all postpartum women, independent of mood status, because they are recovering from birth and adjusting to the sleep demands of a new baby. All nine diagnostic items for depression may not be relevant for PPD, which may have slightly different features from depression outside the postpartum period, such as comorbid anxiety symptoms which are not assessed in the PHQ-9.

Each of these validated instruments for detecting PPD vary in their time scale and length of the instrument. The BDI and EPDS ask about symptoms in the previous week, and the PHQ-9 addresses symptoms of the last two weeks. Diagnostic interview addresses symptoms in the last month. Because screening opportunities for PPD may vary, an expanded time frame for symptom screening may improve detection and treatment of subclinical and clinical depression symptoms. Each of these measures varies from 9 to 21 items, and a diagnostic interview can require hours to complete. A brief screening instrument that can detect symptom profiles over a broad time frame in the postpartum period would be helpful for triggering a full clinical assessment of mood and referral to appropriate treatments.

The Pregnancy Risk Assessment Monitoring System (PRAMS) was developed by the Centers for Disease Control (CDC) to assess prevalence of perinatal health issues (www.cdc.gov/prams/Questionnaire.htm). The original use of the PRAMS was to implement state-wide surveillance of pregnancy-related health issues. Prior to 2009, the PRAMS Core Questionnaire did not include questions that assessed postpartum depression and anxiety (O'Hara, et al. 2012). At the time of inception of this study, one of the co-authors (MO) was conducting a validation study of depression and anxiety items that had the highest prediction values for assessing PPD related symptoms and would potentially be incorporated into future versions of the PRAMS Questionnaire (O'Hara, et al., 2012). Based upon the items with the best validity, depression and anxiety items were chosen to create the six-item PRAMS (PRAMS-6). The PRAMS-6 was chosen over other brief screening tools, like the three item EPDS, because it covered the entire postpartum period, included both depression and anxiety symptoms common to postpartum women, and would potentially allow future comparisons to CDC surveillance populations.

The aims of this investigation were 1) to inspect the diagnostic accuracy of the PRAMS-6 as an ultra-brief screening tool for PPD that covers symptoms over the entire postpartum period, 2) to inspect three-item subscales of the PRAMS-6 in detecting PPD, and 3) to compare the PRAMS screening accuracy to the PHQ-9, a previously validated and brief PPD screening instrument. The sensitivity and specificity of the screening tools of interest were compared to two clinician-rated diagnostic measures of PPD, the Structured Clinical Interview for the DSM-IV MDE module (SCID; First, Spitzer, Gibbon & Williams, 2002) and the 17-item Hamilton Rating Scale for Depression (HRSD-17) (Hamilton, 1960).

Method

The data for this study was collected as part of a larger clinical treatment trial for PPD. The trial is ongoing at Women and Infants Hospital in Providence, RI and at the University of Iowa, Iowa City, IA. At the RI site, women were referred to the study by practitioners from hospital and community clinics, obstetric-gynecological clinics, primary care physicians, and mental health care providers. At the Iowa site, women were recruited for the study largely using the state birth records information. Women identified from the birth records

who had a baby in the past year and were between the ages of 18 and 50 were sent a letter inviting them to participate in a study about their emotional experiences after delivery. Women who were interested in participating included their contact information on a postcard and returned it to study personnel who contacted potential participants by phone for the screening. Potential participants in Iowa were also referred to the study from the University of Iowa Women's Wellness and Counseling Service and community clinics. All procedures were approved and monitored by the Women and Infants IRB and the University of Iowa IRB (for each site respectively).

At both sites, women were eligible for the clinical trial if they were within 12 months postpartum following a live birth, were between 18 and 45 years of age, could speak and read English, were not currently undergoing treatment for PPD, and were willing to be randomly assigned to treatment. Breast feeding women were not excluded. Women in the trial received free treatments, and they were paid for the time spent attending assessment interviews. Participants were paid \$50 if they completed the initial screening interview. If they completed all assessments throughout the course of the trial, women in RI earned \$205, and women in Iowa earned \$105.

The PHQ-9 and PRAMS-6 were administered over the phone by trained research staff to potential participants. The PHQ-9 has been administered over the phone in both perinatal population (Hanusa, 2008) and in other populations (Razykov, 2012).

Suspected PPD was indicated by a PHQ-9 score greater than or equal to 10. Women with suspected PPD based on the PHQ-9 were scheduled for an interview using the SCID and the 17-item Hamilton Rating Scale for Depression (HRSD-17), which were administered by trained research staff either over the phone or in-person. All interviews were audiotaped.

Women continued in the trial if they met criteria for depression using the SCID MDE module and had a HRSD-17 score greater than or equal to 12. Women were excluded from the trial if the SCID interview detected bipolar disorder, substance abuse, psychosis or suicidal behavior. In these cases, women were referred to more appropriate care.

Women who were screened using the PHQ-9 and PRAMS and who had the clinical intake interview were included in the analyses, regardless of whether or not they continued in the larger trial.

Measures

The PHQ-9 contains nine items that assess each of the symptoms that comprise DSM-IV diagnostic criteria. Specifically, women are asked to rate items pertaining to loss of interest, depressed mood, sleep disruption, fatigue, changes in appetite, guilt and feelings of worthlessness, changes in concentration, psychomotor retardation/agitation, and suicide. Women are asked to rate each item based on the prevalence of the symptom within the past two weeks. Rating options are: not at all, several days, more than half the days, nearly every day. Item scores are added to form a composite score.

The PRAMS-6 contains six items that are rated on a likert scale from 1 (never) to 5 (always). Responses are scored according to the prevalence of the symptoms since the delivery of the baby. The questions on this instrument are: 1) I have felt down, depressed, or sad, 2) I have felt hopeless, 3) I have felt slowed down physically, 4) I have felt panicky, 5) I have felt restless or fidgety, and 6) I have felt fearful.

Data Analyses

Composite scores were used in analyses for all the screening measures. The first three items of the PRAMS-6 screen for depression symptoms, and the last three items screen for anxiety symptoms. In order to investigate the individual screening accuracy of the depression items and the anxiety items, separate analyses were performed on the sums of the three-item PRAMS subscales for depression (PRAMS-3D) and the subscale for anxiety (PRAMS-3A) in addition to the sum of all six items (PRAMS-6). The sums of all items of the PHQ-9 and the HRSD-17 were used in analysis. For all measures, higher scores indicate increased morbidity.

To discern the diagnostic accuracy of the PHQ-9, PRAMS-3A, PRAMS-3D, and PRAMS-6, Receiver Operating Characteristic (ROC) analysis using the SCID and the HRSD-17 as the diagnostic standard were computed. When the HRSD-17 was the standard of depression, a score of 15 or greater determined depression (Frank, 1991). ROC curves were computed using SPSS 17.1. For each ROC, the Area Under the Curve (AUC) was computed.

Using the standards put forth by Streiner and Cairney (2007), an AUC less than 0.70 is low accuracy, AUC between 0.70 – .90 shows moderate accuracy for the test, and an AUC higher than 0.90 indicates excellent test accuracy.

Cut-point scores for screening measures can be determined by selecting a sensitivity level, which is appropriate for some cases, or by finding the value that optimizes sensitivity and specificity. A cut-point score using 80% sensitivity, regardless of specificity, minimizes false negatives while moderating false positives. The use of 80% sensitivity to determine cut-points is warranted for screening measures in psychological research because false negatives are often more detrimental than false positives (Sharifi, et al., 2008). A cut-point score that optimizes diagnostic ability by balancing sensitivity and specificity provides the score at which false positives are minimized and true positives are maximized thereby representing the score at which the least amount of detection errors are present. Using the data generated by the ROC analysis, diagnostic cut-points were determined for both the 80% sensitivity level and the optimized balance of sensitivity and specificity.

Results

Participants

Data from 1392 women screened for the clinical trial were included in the data set. Not all women had scores for all assessments. Women included for each analysis had a score on the standard test and the screening measure; sample N for each analysis is provided below.

Women in Iowa and Rhode Island included in the sample (N = 1392) were on average 28.52 years old (SD = 5.45) and 719 were breastfeeding (51.7%). In this sample, 8 were American Indian/Alaska Native (0.57%), 25 were Asian (1.80%), 73 were African American (5.24%), 4 were Native Hawaiian or Pacific Islander (0.29%), 1222 were White (87.79%), and 60 were of unknown/not reported race (4.31%). Women in this sample were predominantly non-hispanic (99.14%). There were 100 women who identified as Hispanic (7.2%), and 15 women for which ethnicity was unknown (1.1%).

ROC Analyses Using the SCID MDE as the Standard

A total of 1011 women had scores for the PHQ-9, PRAMS-6, and SCID and were included.

Table 1 show the AUC values, 95% confidence levels, and p values for each tool when the SCID was used as the gold standard. Using the SCID as the diagnostic standard, all the screening instruments were significantly better than chance at detecting PPD.

According to the AUC values obtained here, the PRAMS-6, PRAMS-3D, and PHQ-9 had AUC values with moderate accuracy compared to the SCID. The PRAMS-3A exhibited a low-level of accuracy with the SCID.

Cut-point scores determined by this analysis for each screening tool are presented in Table 2. The maximum possible score for each measure is also included for reference.

ROC Analyses using the HRSD-17 as the Standard

A total of 914 women had scores for the HRSD-17, the PHQ-9, and the PRAMS-6 and were included in analysis.

Table 3 shows the AUC and 95% confidence intervals when the HRSD-17 with a score greater than or equal to 15 as the standard for depression. The PHQ-9, PRAMS-6, PRAMS-3D, and PRAMS-3A were all significantly better than chance at detecting PPD using the HRSD-17 as the standard. The AUC values each of these tests shows that they were all moderately accurate at detecting PPD.

Additional Observations

To investigate the relative utility of each of these screening tools within the context of the cut-points determined by the preceding analyses, the diagnostic capability of each instrument was inspected by calculating the percentage of cases of PPD that were categorized by each instrument for both the 80% sensitivity level and the optimized cut-points. The PRAMS-3A subscale was excluded from this inspection because it did not show consistent accuracy using both standards. Table 4 presents the percentage of women in this sample with suspected depression using both cut-point scores for each tool, the percentages of diagnosis obtained by the gold standards (SCID and HRSD-17), and the positive predictive value (PPV) for each cut-point as determined by the SCID.

As expected in using the 80% sensitivity for cut point scores, the screening tools identified more suspected cases of PPD than were diagnosed by the SCID or HRSD-17. The screening tools identified approximately 14 – 24% more cases using the 80% sensitivity cut-point score than were confirmed by the gold standard clinical interview. For the PRAMS-6 and PHQ-9, using the optimum cut-point score reduced the number of false positives that were detected during screening. However, the optimum PRAMS-3D cut-point score was identical to the 80% sensitivity cut-point, so detection errors were not improved by using the alternate cut-point score. In summary, the 80% sensitivity cut-point score detects more false positives, while the optimum cut-point score is more specific for the PRAMS-6 and PHQ-9. The costs, benefits, and risks of both false positives and false negatives should be considered when determining which cut-point is most appropriate, within the context of the uses of the instrument.

Pearson's correlations were performed for all continuous measures (excludes the SCID binary measure). The PHQ-9 was significantly correlated with PRAMS-6 ($r=.681$, $p<.001$, $N = 1392$), PRAMS-3D ($r=.647$, $p<.001$, $N = 1392$), PRAMS-3A ($r=.549$, $p<.001$, $N = 1392$), and HRSD-17 ($r=.592$, $p<.001$, $N = 914$). The HRSD-17 was significantly correlated with the PRAMS-6 ($r=.502$, $p<.001$, $N = 914$), the PRAMS-3D ($r=.471$, $p<.001$, $N = 914$), and the PRAMS-3A ($r=.408$, $p<.001$, $N = 914$).

Discussion

The findings of the present study show that the PRAMS-6 or PRAMS-3D show a moderate level of accuracy for detecting PPD when compared to clinical interview. The results here

support the use of the PRAMS-6 as an ultra-brief screening tool in research and clinical care.

The findings of the present study support the use of the PRAMS-6 or the PRAMS-3D as brief screening instruments for PPD. All three tools, PRAMS-6, PRAMS-3D, and PHQ-9, had similar levels of accuracy (78 – 82%) for predicting PPD that was later confirmed by structured interviews (SCID and HRSD-17). The PHQ-9 exhibited slightly better accuracy than the PRAMS-6.

A recent study by Sidebottom, Harrison, Godecker, and Kim (2012) found a similar level of accuracy for the PHQ-9 in detecting prenatal depression at obstetric clinical visits. In their study, a cutoff score of 10 yielded a sensitivity of 85% when looking at women who met full criteria for PPD and a sensitivity of 75% when including women with subdiagnostic PPD (i.e., at least three depression symptoms). In our study, a cutoff score of 10 was associated with 80% sensitivity for PPD. Therefore, the accuracy of the PHQ-9 was similar for perinatal women in both clinical visits and the treatment-seeking women in our study.

The PRAMS-6 and its shorter subscale, the PRAMS-3D, are both adequate predictors of PPD. These tools show comparable accuracy to the PHQ-9 but contain only three or six items. The slightly reduced accuracy of the PRAMS scales, compared to the PHQ-9, may be reflective of the fact that the PRAMS-6 only addresses a small sample of symptoms that may be present in PPD. However, the accuracy deficit was small and resulted in more false positives, which may be less problematic than false negatives for PPD.

The PRAMS-6 showed 81% PPV using the optimized cutoff point (a score of 17) and 63% PPV using the 80% sensitivity cutoff score of 15. The PRAMS-3D had an optimized cutoff score of 9 at the 80% sensitivity level and a PPV of 70%. In both cases, some accuracy is sacrificed for a broader net - detecting more cases of PPD than were confirmed by SCID. In the case of PPD, false positives compared to the SCID may indicate sub-threshold PPD or minor depression, which may be associated with maternal impairment and warrant treatment (Weinberg, et al., 2001).

The PRAMS-3A subscale was moderately accurate when the HRSD-17 was the standard but exhibited low accuracy when the SCID was the standard. This is an interesting finding in light of previous reports that the three-item subscale of the EPDS (containing three anxiety items) is sufficient for PPD detection in adolescents and young adults (Kabir, 2008). However, Kabir's study did not use the SCID as a standard of diagnosis and had a different population than the current study. Kabir had 199 young women from 14 to 26 years old in adolescent centered treatment program. The current study had a larger, more age-diverse sample with an average age of 28. The anxiety items in this study were not sufficiently accurate to screen for PPD. However, only one of the anxiety items, 'feeling panicky', is common to both the PRAMS and the EPDS.

A key difference in the screening instruments inspected in this study is the time period within which symptoms are assessed. The PHQ-9 covers symptoms over the last two weeks and the PRAMS addresses symptoms since the time of delivery. Women in the current study were between 0 and 12 months postpartum, so in some cases the PRAMS-6 assessed a much larger time frame than the PHQ-9. This variability in time frame could account for the decreased accuracy compared to the SCID. However, the moderate level of accuracy seen by the PRAMS may also indicate that the profile of symptoms over the entire postpartum period is at least as important for detecting PPD as the symptoms in the most previous two weeks.

The PHQ-9 addresses all the diagnostic symptoms, while the PRAMS addresses only a subset of symptoms. Together with the variable time frame of symptom assessment, the PRAMS would be expected to show a lower level of accuracy against the SCID. However, the decreased accuracy is skewed toward false positives – meaning more women were suspected of PPD that were not confirmed by interview. Previous studies support the use of screening instruments that sacrifice some accuracy in order to minimize false negatives (Gjerdingen, et al., 2011). Additionally, subclinical symptoms over the entire postpartum period may represent a need for monitoring of symptoms and possible treatment even if full clinical diagnosis is not achieved at the time of screening.

One benefit to the PRAMS-6 over other brief screening instruments is data from clinical care and research can be compared to broader populations of new mothers through CDC monitoring once the questions have been incorporated into the PRAMS Questionnaire. Research samples are often restricted regionally, ethnically, and racially. Comparison to broader populations in the CDC surveillance system may offer important insight into PPD.

The current study is limited in that many of the women in this current sample were treatment-seeking. PPD was more prevalent in our sample (about 41%) than is estimated for the general population (about 10%). The PRAMS instrument was moderately accurate in detecting PPD in our sample, but the level of accuracy could be affected by higher prevalence in the sample. Furthermore, the clinical utility of the PRAMS for postpartum women in the community at primary care, pediatric, or obstetric settings is not known. Additionally, most of the women in this study were non-Hispanic and Caucasian, so the utility of these screening instruments in more culturally diverse populations is unknown.

Acknowledgments

The treatment trial from which this data was extracted was funded by a grant from the National Institute of Mental Health (MH074919). Medication for the treatment trial was supplied by Pfizer.

Drs. Zlotnick, Pearlstein, and Stuart received research support from Pfizer. Dr. Pearlstein has acted as a consultant for Ironwood Pharmaceuticals.

References

- Armitage R, Flynn H, Hoffmann R, Vazquez D, Lopez J, Marcus S. Early developmental changes in sleep in infants: the impact of maternal depression. *Sleep*. 2009; 32(5):693–696. [PubMed: 19480236]
- Arroll B, Goodyear-Smith F, Crengle S, Gunn J, Kerse N, Fishman T, Falloon K, Hatcher S. Validation of the PHQ-2 and PHQ-9 to screen for major depression in the primary care population. *Ann of Fam Med*. 2010; 8(4):348–353.10.1370/afm.1139 [PubMed: 20644190]
- Boyle LL, Richardson TM, He H, Xia Y, Boustani M, Conwell Y. How do the phq-2, the phq-9 perform in aging services clients with cognitive impairment. *Int J Geriatr Psychiatry*. 2010; 26(9): 952–60.10.1002/gps.2632
- Cox JL, Holden JM, Sagovsky R. Detection of postnatal depression. Development of the 10-item Edinburgh Postnatal Depression Scale. *Br J Psychiatry*. 1987; 150(6):782–6.10.1192/bjp.150.6.782 [PubMed: 3651732]
- Declercq ER, Sakala C, Corry MP, Applebaum S. Listening to Mothers II: Report of the second national U.S. survey of women's childbearing experiences. *J Perinat Educ*. 2007; 16(4):9–14.10.1624/105812407X244769 [PubMed: 18769512]
- First, MB.; Spitzer, RL.; Gibbon, M.; Williams, JBW. Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Non-patient Edition. (SCID-I/NP). New York: Biometrics Research, New York State Psychiatric Institute; 2002.
- Fishell A. Depression and anxiety in pregnancy. *J Popul Ther Clin Pharmacol*. 2010; 17(3):363–369.

- Flynn HA, Sexton M, Ratliff S, Porter K, Sivin K. Comparative performance of the Edinburgh postnatal scale and the patient health questionnaire-9 in pregnant and postpartum women seeking psychiatric services. *Psychiatr Res*. 2010; 187(1–2):130–4.10.1016/j.psychres.2010.10.022
- Frank E, Prien RF, Jarrett RB, Keller MB, Kupfer DJ, Lavori PW, et al. Conceptualization and rationale for consensus definitions of terms in major depressive disorder. Remission, recovery, relapse, and recurrence. *Arch Gen Psychiatr*. 1991; 48(9):851–855.10.1001/archpsyc.1991.01810330075011 [PubMed: 1929776]
- Gjerdengen D, Crow S, McGovern P, Miner M, Center B. Changes in the depressive symptoms over 0–9 months postpartum. *Journal of Women's Health*. 2011; 20(3):381–386.10.1089/jwh.2010.2355
- Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry*. 1960; 23(1):56–62. [PubMed: 14399272]
- Hanusa BH, Scholle SH, Haskett RF, Spadaro K, Wisner KL. Screening for depression in the postpartum period: A comparison of three instruments. *J Women's Health*. 2008; 17(4):10.1089/jwh.2006.0248
- Heneghan AM, Chaudron LH, Storfer-Isser A, Park ER, Kelleher KJ, Stein RE, Hoagwood KE, O'Connor KG, Horwitz SM. Factors associated with identification and management of maternal depression by pediatricians. *Pediatr*. 2007; 119(3):444–54.10.1542/peds.2006-0765
- Hewitt CE, Gilbody SM, Mann R, Brealey S. Instruments to identify post-natal depression: Which methods have been the most validated, in what settings and in which language? *Int J Psychiatr Clin Practice*. 2010; 14:72–76.
- Ji S, Long Q, Newport J, Knight B, Zach EB, Morris NJ, Kutner M, Stowe ZN. Validity of depression rating scales during pregnancy and the postpartum period: Impact of trimester and parity. *J Psychiatr Res*. 2011; 45(2):213–219.10.1016/j.jpsychires.2010.05.017 [PubMed: 20542520]
- Kabir K, Sheeder J, Kelly LS. Identifying postpartum depression: are 3 questions as good as 10? *Pediatr*. 2008; 122(3):e696–702.10.1542/peds.2007-1759
- Kroenke K, Spitzer RL, Williams JBW, Lowe B. The patient health questionnaire somatic, anxiety, and depressive symptom scales: a systematic review. *Genl Hosp Psychiatr*. 2010; 32:345–359.10.1016/j.genhosppsy.2010.03.006
- Leahy-Warren P, McCarthy G, Corcoran P. Postnatal depression in first-time mothers: prevalence and relationships between functional and structural social support at 6 and 12 weeks postpartum. *Arch Psychiatr Nurs*. 2011; 25(3):174–184.10.1016/j.apnu.2010.08.005 [PubMed: 21621731]
- Liberto TL. Screening for depression and help-seeking in postpartum women during well-baby pediatric visits: an integrated review. *J Pediatr Health Care*. 2012; 26(2):109–17.10.1016/j.pedhc.2010.06.012 [PubMed: 22360930]
- Liu S, Yeh Z, Huang H, Sun F, Tjung J, Hwang L, Shih Y, Yeh AW. Validation of patient health questionnaire for depression screening among primary care patients in Taiwan. *Compr Psychiatr*. 2011; 52:96–101.10.1016/j.comppsy.2010.04.013
- Merz EL, Malcarne VL, Roesch SC, Riley N, Sadler GR. A multigroup confirmatory factor analysis of the Patient Health Questionnaire-9 among English- and Spanish-speaking Latinas. *Cultur Divers Ethnic Minor Psychol*. 2011; 17(3):309–16.10.1037/a0023883 [PubMed: 21787063]
- Milgrom J, Mendelsohn J, Gemmill AW. Does postnatal depression screening work? Throwing out the bathwater, keeping the baby. *J Affect Disord*. 2010; 132(3):301–10.10.1016/j.jad.2010.09.031 [PubMed: 20952072]
- O'Hara MW, Stuart S, Watson D, Dietz PM, Farr SL, D'Angelo D. Brief scales to detect postpartum depression and anxiety symptoms. *J Womens Health (Larchmt)*. 2012; 21(12):1237–43.10.1089/jwh.2012.3612 [PubMed: 23130750]
- Razykov I, Hudson M, Baron M, Thombs BD. Canadian Scleroderma Research Group. The utility of the PHQ-9 to assess suicide risk in patients with systemic sclerosis. *Arthritis Care Res (Hoboken)*. 2012 epub ahead of print Nov 30. 10.1002/acr.21894
- Richardson LP, McCauley E, Grossman DC, McCarty CA, Richards J, Russo JE, Rockhill C, Katon W. Evaluation of the Patient Health Questionnaire-9 Item for detecting major depression among adolescents. *Pediatr*. 2010; 126(6):1117–23.10.1542/peds.2010-0852

- Sharifi F, Mousavinasab N, Maloomzadeh S, Jaber Y, Saeini M, Dinmohammadi M, Nagomashoaa A. Cutoff point of waist circumference for the diagnosis of metabolic syndrome in an Iranian population. *Obes Res Clin Practice*. 2008; 2(3):171–178.10.1016/j.ocrp.2008.04.004
- 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(5):367–74.10.1007/s00737-012-0295-x [PubMed: 22983357]
- Streiner DL, Cairney J. What's under the ROC? An introduction to receiver operating characteristics curves. *Can J of Psychiatr*. 2007; 52:121–128.
- Weinberg MK, Tronik EZ, Beeghly M, Olson KL, Kernan H, Riley JM. Subsyndromal depressive symptoms and major depression in postpartum women. *Amer J Orthopsychiatr*. 2001; 71(1):87–97.
- 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(5):483–91.10.3122/jabfm.2009.05.080155 [PubMed: 19734393]
- Yeung A, Fang F, Yu S, Vorono S, Ly M, Wu S, Fava M. Validation of the patient health questionnaire-9 for depression screening among Chinese americans. *Compr Psychiatr*. 2008; 49:211–217.10.1016/j.comppsy.2006.06.002

Table 1

ROC Results: SCID-IR MDE as the Standard

	AUC	p	95% Confidence	
PRAMS-6	0.777	< .001	0.748	0.806
PRAMS-3D	0.796	< .001	0.767	0.824
PRAMS-3A	0.698	< .001	0.665	0.731
PHQ-9	0.826	< .001	0.8	0.852

Table 2

Cut-point Scores for Each Measure

	80% Sensitivity	Optimized	Maximum Possible
PHQ-9	10	12	27
PRAMS-6	15	17	30
PRAMS-3D	9	9	15
PRAMS-3A	6	8	15

Table 3ROC Results: HRSD-17 as the Standard^a

	AUC	p	95% Confidence	
PRAMS-6	0.757	< .001	0.725	0.789
PRAMS-3D	0.736	< .001	0.701	0.770
PRAMS-3A	0.713	< .001	0.679	0.747
PHQ-9	0.801	< .001	0.771	0.832

^aHRSD-17 \geq 15 determined depression.

Table 4

Predictive Value of Each Measure

	%MDE (n)	%No MDE (n)	PPV%*
<i>80% Sensitivity Cut-point</i>			
PRAMS-6 (>=15)	64.2(649)	35.8(362)	63
PRAMS-3D (>=9)	57.7(583)	42.3(428)	70
PHQ-9 (>=10)	54.1(547)	45.9(464)	75
<i>Optimum Cut-point</i>			
PRAMS-6 (>=17)	50.2(508)	49.8(503)	81
PRAMS-3D (>=9)	57.7(583)	42.3(428)	70
PHQ-9 (>=12)	40.3(407)	59.7(604)	101
<i>Standards</i>			
SCID	40.7(411)	59.3(600)	
HRSD-17 (>=15)	36.3(332)	63.7(582)	

* Positive predictive values are based upon SCID diagnosis as standard