Validity of Depression Rating Scales during Pregnancy and the Postpartum Period: Impact of Trimester and Parity

Shuang Ji[1], **Qi Long**[1,4], **D. Jeffrey Newport**[2], **Hyeji Na**[2], **Bettina Knight**[2], **Elizabeth B. Zach**[2], **Natalie J. Morris**[2], **Michael Kutner**[1], and **Zachary N. Stowe**[2,3]

[1] Department of Biostatistics and Bioinformatics, Emory University Rollins School of Public Health, 1518 Clifton Rd NE, Room 322, Atlanta, GA 30322

[2] Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, 1365 Clifton Rd NE, Suite B6100, Atlanta, GA 30322

[3] Department of Gynecology and Obstetrics, Emory University School of Medicine

Abstract

The objective of the current study was to delineate the optimal cutpoints for depression rating scales during pregnancy and the postpartum period and to assess the perinatal factors influencing these scores. Women participating in prospective investigations of maternal mental illness were enrolled prior to 28 weeks gestation and followed through 6 months postpartum. At each visit, subjects completed self-rated depression scales – Edinburgh Postnatal Depression Scale (EPDS) and Beck Depression Inventory (BDI) and clinician-rated scales – Hamilton Rating Scale for Depression ($HRSD_{17}$ and $HRSD_{21}$). These scores were compared to the SCID Mood Module for the presence of fulfilling diagnostic criteria for a major depressive episode (MDE) during 6 perinatal windows: preconception; first trimester; 2nd trimester; 3rd trimester; early postpartum; and later postpartum. Optimal cutpoints were determined by maximizing the sum of each scale's sensitivity and specificity. Stratified ROC analyses determined the impact of previous pregnancy and comparison of initial to follow-up visits. A total of 534 women encompassing 640 pregnancies and 4025 follow-up visits were included. ROC analysis demonstrated that all 4 scales were highly predictive of MDE. The AUCs ranged from 0.857 to 0.971 and were all highly significant ($p<0.0001$). Optimal cutpoints were higher at initial visits and for multigravidas and demonstrated more variability for the self-rated scales. These data indicate that both clinician-rated and self-rated scales can be effective tools in identifying perinatal episodes of major depression. However, the results also suggest that prior childbirth experiences and the use of scales longitudinally across the perinatal period influence optimal cutpoints.

Keywords

[4]To whom correspondence should be addressed: Qi Long, Ph.D., Assistant Professor, Department of Biostatistics and Bioinformatics, 1518 Clifton Rd. NE, Room 322, Rollins School of Public Health, Emory University, Atlanta, GA 30322, Office: (404) 712-9975, Fax: (404) 727-1370, qlong@emory.edu.

INTRODUCTION

Maternal depression during pregnancy and the postpartum period, i.e. perinatal depression, is a common problem that has been the focus of extensive investigation. Studies examining the prevalence of perinatal depression have demonstrated considerable variability that is a consequence, at least in part, of the assessment method used to identify the presence of depression, the timing of the assessment, and population characteristics (Gaynes et al 2005). Authors of one review recommended that more precise determinants of the occurrence of perinatal depression are needed to estimate disease burden more accurately (Gaynes et al 2005).

Depressive symptoms are common in pregnancy with most studies reporting rates comparable to non-gravid women (Cutrona 1983; Kumar & Robson 1984; Watson et al 1984; Gotlib et al 1989; O'Hara et al 1991). A meta-analysis of depression during pregnancy (Bennett et al 2004), utilizing data encompassing 19,284 gravidas from 21 studies in which depression was assessed by a structured clinical interview or self-rated scale such as the Beck Depression Inventory (BDI) (Beck et al 1961), or the Edinburgh Postnatal Depression Scale (EPDS) (Cox et al 1987), estimated the prevalence of depression as 7.4% in the first trimester, 12.8% in the second trimester, and 12.0% in the third trimester. However, the data were inadequate to render conclusions regarding comparative risk between trimesters. Furthermore, the authors reported that the BDI produced significantly higher prevalence estimates, whereas EPDS estimates were statistically equivalent to those of structured clinical interviews.

Depression during the postpartum period has also garnered considerable attention. An earlier meta-analysis by O'Hara and Swain (1996), encompassing 12,810 postpartum women from 59 studies utilizing a clinical interview or self-report scale, estimated the prevalence of depression in the postpartum period at 13.0%. Similar to the pregnancy data, self-report measures yielded higher estimates of postpartum depression than clinician-administered assessments. The postpartum timing of the assessment did not significantly affect the prevalence estimates in this meta-analysis. A review of the prevalence studies found that 7.1% may experience a major depressive episode (MDE) during the first 3 months postpartum (Gavin et al 2005). Despite the historical assumption of increased vulnerability to depression in the postpartum period, the literature has not definitively demonstrated an increased risk (Gavin et al 2005). In contrast, a recent large-scale epidemiological study provided evidence of increased risk for major depression in the postpartum period compared with non-pregnant/non-postpartum women (adjusted odds ratio: 1.52; 95% CI: 1.07–2.15) (Vesga-Lopez et al 2008). Moreover, women are more likely to require psychiatric admission for depression during the postpartum period than outside the puerperium (Kendell et al 1987; Munk-Olsen et al 2006).

Numerous scales have been developed for identifying postpartum depression or risk factors for the development of postpartum depression (Beck 1995; Ferguson et al 2002; Morris-Ruth et al 2003; Perfetti et al 2004; Austin et al 2005). The EPDS has emerged as a well-validated and widely-utilized instrument for postpartum depression screening and detection. Conversely, validated tools to assess depression during pregnancy are lacking (Gaynes et al 2005). By default, the EPDS, developed for postpartum use, has been increasingly used to identify depression during pregnancy (Adouard et al 2005; Thoppil et al 2005; Felice et al 2006) and to screen for those at risk for developing depression during pregnancy (Evans et al 2001; Rubertsson et al 2005; Gordon et al 2006). Beyond this *ad hoc* use of the EPDS, no scale exists to identify major depressive disorder during pregnancy. Moreover, only one screening test, an unvalidated scale consisting of only two items, has been developed specifically for depression in pregnancy (Campagne 2004). Our group in collaboration

(Altshuler et al 2008), recently completed an individual item analysis of the 28 item Hamilton Rating Scale for Depression (HRSD) compared to SCID Mood Module to identify the items most predictive of an accurate identification of an episode of major depression across all trimesters of pregnancy. The seven items most predictive of the presence of depression were tested as a screening tool for depression during pregnancy (Altshuler et al 2008).

The urgent need to identify reliable instruments for detecting perinatal depression is underscored by: 1) numerous reports of adverse obstetrical, neonatal, and developmental outcomes in association with maternal stress, depressive symptoms, and episodes of major depression during the perinatal period (Paton et al 1977; Zuckerman & Bresnahan 1991; Sheer et al 1992; Hedegaard et al 1993; Pritchard & Teo 1994; Orr & Miller 1995; Chung et al 2001; Andersson et al 2004; Mancuso et al 2004; Dayan et al 2006; Diego et al 2006; Neggers et al 2006); 2) accurate diagnosis of a MDE during the peripartum is complicated by the fact that purportedly normal perinatal symptoms (e.g., fatigue, sleep disturbance, appetite and weight changes, diminished libido) potentially overlap with the neurovegetative symptoms comprising part of the diagnostic criteria for major depression; 3) lower estimates of maternal mental illness during pregnancy may be in part secondary to limited recognition (Vesga-Lopez et al 2008); and 4) validated assessment tools are a requisite step in the design and completion of much needed controlled treatment studies during the perinatal period.

The overall aim of the current study was to provide clinicians and researchers alike with information regarding the sensitivity and specificity of commonly used depression rating scales during pregnancy and the postpartum period. The specific objectives of the study were: 1) to identify optimal cutpoints (maximizing the summation of sensitivity and specificity) for commonly used depression rating scales during each trimester of pregnancy and the postpartum period; 2) to determine whether previous pregnancy and childbirth experience influences the performance of the rating instruments; and 3) to determine whether repeated administration of a depression rating scale over the course of pregnancy and the postpartum period is associated with learning effects that alter the optimal cutpoints for the rating scales. With respect to these objectives, our *a priori* hypotheses were: 1) that the performance of the scales including optimal cutpoints would be altered during pregnancy, particularly during the third trimester when many of the physical symptoms of pregnancy most closely mirror the neurovegetative symptoms of depression; 2) that multigravid women (having previously experienced the physical sequelae of gestation) would be more likely to report physical symptoms of depression on a depression rating scale than primigravid women producing higher cutpoints on the scales during pregnancy; and 3) that optimal cutpoint scores would be impacted by repeated administration of both clinician-administered and self-rated depression scales.

METHODS

Subjects

The study was conducted at the Women's Mental Health Program (WMHP) at the Emory University School of Medicine. Women with a lifetime history of mental illness participating in one of two prospective longitudinal perinatal investigations of the pharmacokinetics of psychotropic medications and/or maternal stress (P50 MH 68036; P50 MH 77928) were screened for inclusion in the current analysis. The schedule and methods for assessing maternal depression were identical in the two studies. Participants were enrolled no later than week 28 of gestation and evaluated at 4–6 week intervals across pregnancy and through 26 weeks postpartum. At each visit, subjects completed the self-rated BDI and EPDS. In addition, a research interviewer masked to treatment status administered the Structured Interview Guide (Williams 1988) for the Hamilton Rating Scale for

Depression (Hamilton 1960) to obtain 17-item ($HRSD_{17}$) and 21-item ($HRSD_{21}$) scores and the Mood Module of the *Structured Clinical Interview for DSM-IV Axis I Disorders* (First et al 2002). To ensure consistent administration of the clinician-rated instruments, research interviewers were trained to use a "rate as you see" approach when scoring items, eschewing any subjective judgment as to whether symptoms were due to depression or pregnancy/ postpartum. Quarterly inter-rater reliability assessments were conducted throughout the course of both investigations to ensure maintenance of *kappa* statistics ≥ 0.8 on all clinician- administered instruments. All scales were coded with a HIPAA compliant identifier and entered into a centralized database. Subjects were included in the current analysis if they had two or more perinatal visits during which the *SCID* Mood Module and one or more of the depression rating scales were completed. The investigation was carried out in accordance with the latest version of the Declaration of Helsinki. The study was reviewed and approved by the Emory University Institutional Review Board. Informed consent of the participants was obtained after the nature of the procedures had been fully explained.

Data Analysis

Each visit was assigned to one of 6 distinct perinatal epochs including: 1) preconception; 2) 1st trimester (0–12 weeks gestation); 3) 2nd trimester (13–24 weeks gestation); 4) 3rd trimester (25 weeks gestation to delivery); 5) early postpartum (0–6 weeks); and 6) late postpartum (7–26 weeks). A completed *SCID* Mood Module plus one or more of the depression symptom scales was necessary for a visit to qualify for inclusion. At each visit, the presence (or absence) of a MDE was determined by the *SCID* Mood Module.

To assess the diagnostic accuracy of the symptom scales within each perinatal epoch and the overall perinatal period, the receiver operating characteristic curve (ROC) analysis proposed by Obuchowski (1997), which accounts for correlation due to repeatedly measured rating scales from each subject, was used. Defining a cutpoint score for a scale as the score such that greater or equal scores are considered consistent with the presence of a MDE, then the optimal cutpoint score for each scale was defined as the score at which the sum of the scale's sensitivity and specificity was maximized within each perinatal epoch. Stratified ROC analyses were conducted to examine the impact of primigravid vs. multigravid pregnancies and first visit vs. follow-up visits on the accuracy of depression scales for identifying an active MDE. Typically, the diagnostic accuracy of a symptom scale is considered poor when ROC AUC < 0.70, fair when $0.70 \leq$ ROC AUC < 0.80, good when $0.80 \leq$ ROC AUC < 0.90, and excellent when $0.90 \leq$ ROC AUC (Tape 2008). Wald tests were used to compare ROC AUCs. All statistical tests were two-tailed and conducted at a significance level of 0.05.

RESULTS

Analysis of Overall Sample

A total of 708 women were enrolled in longitudinal perinatal investigations of the pharmacokinetics of psychotropic medications and/or maternal stress. One hundred and seventy four women were excluded secondary to missing data – most commonly, lack of adequate follow-up visit data. Table 1 summarizes the characteristics of the entire study population, the subpopulation included in the analysis and the subpopulation excluded from the analysis due to missing data. The included group is demographically more diverse than the excluded group; specifically, the included group has higher minority representation, includes more unmarried subjects, has more subjects who have completed no more than a high school education, and has a higher proportion of unplanned pregnancies.

Five hundred and thirty four participants (75.4% of all enrollees), encompassing 640 pregnancies with 4025 visits, were included in the current analysis. The sample was very homogeneous: 33.1 ± 5.1 years of age, predominately Caucasian (85.8%), married or cohabitating at study entry (85.4%), and with a high school education or greater (92.0%). Because the gestational age at study entry varied (preconception through ≤ 28 weeks), the number of subjectswithin each perinatal epoch varied..

ROC analysis demonstrated that all symptom scales (BDI, EPDS, $HRSD_{17}$, $HRSD_{21}$) were highly predictive of MDE (cf. Table 2). The ROC AUCs, ranging from 0.855 to 0.971, were all statistically significant (p<0.0001). For all 4 scales, the peak ROC AUC was observed during the preconception period. As hypothesized, the ROC AUC was lowest during the third trimester for the BDI, $HRSD_{17}$, and $HRSD_{21}$. The AUC was lowest for the EPDS during the early postpartum. ROC AUCs for the EPDS suggest that it achieved excellent diagnostic validity (i.e., AUC ≥ 0.90) in 3 of the 6 epochs. The BDI achieved excellent diagnostic validity in 2 of 6 epochs, and the two HRSD scores each achieved excellent diagnostic validity in 1 of 6 epochs. The remainder of the AUCs for all 4 scales in all epochs were in the good range ($0.80 \leq AUC < 0.90$). The EPDS (preconception, first trimester, third trimester) and BDI (second trimester, early postpartum, late postpartum) each achieved the largest ROC AUC among all scales in 3 of the 6 epochs.

To compare AUCs within each epoch, the analysis was also conducted using "complete cases" only, i.e., limited to visits in which all four depression scales were completed, and the results were similar. During the 2nd trimester, BDI had the largest AUC (0.908), which was significantly larger (p=0.02) than the smallest AUC (0.852, of $HRSD_{17}$). For other epochs, there was no such significant difference between the largest and smallest AUC (data not shown).

Despite the consistently good to excellent diagnostic validity of all four scales across the 6 epochs, there was considerable variability in the optimal cutpoint scores for the self-rated instruments. The EPDS produced the greatest variation with the optimal cutpoint ranging from a low of 9 during the second trimester to a peak of 18 during preconception. Optimal cutpoints for the BDI ranged from a low of 12 during the third trimester to a high of 17 during preconception. Optimal cutpoints for the $HRSD_{17}$ varied by only 1 point (from 14 to 15) and for the HRSD21 by only 2 points (from 14 to 16). Interestingly, the optimal cutpoint scores were highest during the preconception period for all four scales.

Primigravid vs. Multigravid Pregnancies

As noted above, there were 640 pregnancies completed by the 534 subjects included in the current analysis. Among these 640 pregnancies, there were 161 primigravid pregnancies, 465 multigravid pregnancies and 14 pregnancies with missing gravidity/parity data. Demographic analysis indicates that the primigravid group was younger (31.5 vs. 34.4 years, p<.0001) and less likely to be married (80.1% vs. 89.5%, p=.0006) than the multigravid group. There were no other significant demographic differences between the groups (data not shown).

For all four depression scales, across all pregnancy stages, the overall ROC AUC, optimal cutpoint, and summation of sensitivity and specificity for primigravid pregnancies is very similar to that observed for multigravid pregnancies, suggesting that gravidity status had no discernible impact of the global performance of the scales (cf. Table 3). However, inspection of the third trimester results demonstrates that cutpoints are higher for multigravid pregnancies for all 4 scales. In addition, the sensitivity of the scales corresponding to the optimal cutpoint is lower for multigravidas than primigravidas during the third trimester for all scales but the EPDS. Furthermore, within the multigravid group, the sensitivity of the

$HRSD_{17}$, $HRSD_{21}$, and BDI at the optimal cutpoints reaches its nadir during the third trimester.

Comparison of First Visit vs. Follow-Up Visits

Among the 534 subjects included in the current analysis, 178 subjects had previously participated in other WMHP studies, and 356 completed their first research encounter at the WMHP during the current study. The comparison of first versus follow-up visits is limited to these 356 participants who completed 356 first visits and 2474 follow-up visits. Because all subjects were enrolled prior to 28 weeks gestation, none of the first visits occurred during the postpartum period. Consequently, the postpartum epochs were also excluded from the subsequent visits stratum to eliminate any confounding effect of the disparate perinatal epoch. There were no significant demographic differences between the 356 subjects included in this phase of the analysis and the 178 excluded except that the inclusion group was less likely to be Caucasian (83.2% vs. 91.0%, p= .004), reflecting the growing minority participation in WMHP research in recent years; and that the inclusion group was less likely to have a planned pregnancy (63.7% vs. 77.3%, p=.003).

ROC analysis indicated that all 4 scales performed in the good to excellent range at both initial visits and follow-up visits (cf. Table 4). The summation of sensitivity and specificity for all 4 scales was also consistent across the perinatal epochs within both strata. However, the optimal cutpoints for the $HRSD_{17}$ and $HRSD_{21}$ were very consistent between first and follow-up visits, whereas optimal cutpoints on the self-report instruments were generally 4–6 points higher for the first visit than for the follow-up visits during each perinatal epoch (cf. Table 4) and across PC, T1, T2 and T3 periods (cf. Table 4).

DISCUSSION

It is reassuring that ROC analysis across all phases of the analysis indicates that all 4 instruments consistently performed in the good to excellent range (Tape 2008). The overall optimal cutpoints for each scale are consistent with previous reports. Moreover, the sensitivities and specificities of the 4 instruments were also very similar. It might, therefore, be inferred that any of the 4 depression rating scales evaluated in the current study are suitable candidates to identify episodes of major depression during pregnancy and the postpartum period. Given the ease of administration of self-report measures in both the clinical and research settings, it could be argued that there is no justification for using the labor intensive HRSD over the self-report BDI and EPDS. However, the stability of the cutpoint is a key consideration when screening a broad perinatal sample and conducting longitudinal follow-up. The clinician administered HRSD provided more stable cutpoints (1–2 point range) across the perinatal epochs compared to the BDI (5 point range) and EPDS (9 point range). Consequently, the present data suggests that the HRSD may be preferred when conducting longitudinal studies across pregnancy and the postpartum period, but the less costly BDI and EPDS may be preferred for cross-sectional studies.

Consistent with our previous experience (Altshuler et al 2008), the comparative performance of the $HRSD_{17}$ and $HRSD_{21}$ indicates that items 18–21 can be eliminated from the perinatal administration of the HRSD with little or no impact on the performance of the scale. Inclusion of items 18–21 elevated the optimal cutpoint within each perinatal epoch by only 0–1 points and produced no significant improvement in the ROC AUC, sensitivity, or specificity of the HRSD during pregnancy and the postpartum period.

It is noteworthy that the specificities of the scales were uniformly lower during pregnancy and the early postpartum period than during the preconception epoch. Late postpartum specificities were generally intermediate between preconception and pregnancy/early

postpartum values. Consistent with our *a priori* hypothesis, we suspect this may be a consequence of the overlap between physical symptoms of pregnancy and the neurovegetative symptoms of depression. In contrast, we had anticipated that the symptomatic overlap of pregnancy and depression would elevate the optimal cutpoints during pregnancy, the current results found the pregnancy cutpoints to be the same as or lower than the preconception cutpoints. This suggests that women incorporate their own opinions regarding the etiology of the symptoms during both interview and completion of self-rated scales. An extended item analysis may clarify the role that the neurovegetative symptom items play in the perinatal performance of these rating scales.

The primigravida versus multigravida comparison provides additional insights on this issue. As noted above, there is no evidence from the current study to indicate that the gravidity status alters the overall performance of these scales across the entirety of pregnancy and the postpartum period; however, the current data does suggest that multigravidas rate the instruments differently during the third trimester. Whereas there was no evident difference in the specificity of the scales in the third trimester, the cutpoints are consistently higher among multigravidas during the third trimester, and the scale sensitivity is lower among multigravidas for 3 of the 4 scales. We suspect that this confluence of third trimester results among multigravidas (i.e., higher cutpoint, lower sensitivity, unchanged specificity) is a consequence of multigravidas being more ready to report physical discomforts, particularly when such symptoms are incongruent with prior pregnancies, on a depression rating scale than primigravidas. Inclusion of the physical symptoms by the multigravidas would tend to elevate the scores. With the higher overall scores, higher cutpoints would be needed to forestall, at least in part, the resultant decline in sensitivity.

Finally, the analysis of first visit versus follow-up visit performance of the scales is generally reassuring, but raises concerns regarding the lack of cutpoint stability of the self-report depression measures. Overall first visit versus follow-up visit cutpoints for the HRSD only varied by 0–1 points, but the BDI and EPDS first versus follow-up visit cutpoints differed by 6 and 4 points, respectively. This apparent post-first visit learning effect has significant implications for the use of these self-report depression rating scales in longitudinal studies across the perinatal period, again suggesting that the HRSD may be preferred for longitudinal investigations.

The study is limited by the homogeneous clinical population that was able to complete participation in a longitudinal investigation; thereby limiting the generalizability of these results to community based perinatal samples where self-rated instruments are likely to be employed as screening tools. Notably, the group of women excluded from the present analyses were more demographically homogeneous than the included group. While this counter-intuitive finding runs counter to most longitudinal studies, it may reflect the benefits garnered by study participants with respect to frequent contact providing additional education and support. Indeed, the greater heterogeneity of the inclusion group potentially enhances the generalizability of the study results and increases support for further research in perinatal maternal mental illness with respect to the benefit of repeated professional contact. Similarly, the study assesses the validity of multiple scales in identifying episodes of depression that fulfill *SCID* diagnostic criteria. Limiting to episodes meeting diagnostic criteria, fails to assess the utility of scales in appropriately identifying subsyndromal symptoms warranting clinical attention.

Investigations in perinatal psychiatry continue to refine more optimal study parameters and methodology. As noted in our previous investigation, retrospective maternal reports are inadequate proxies for categorizing maternal depression during pregnancy and exposure to non-psychotropic medications (Newport et al 2008). The current study sought to define

optimal cutpoint scores and factors that influence these scores. All scales demonstrated good to excellent ROC AUCs in identification of a MDE as defined by SCID criteria across pregnancy and the postpartum period. The impact of gravidity, first visit completion versus follow up visits, and variability of optimal cutpoint scores across the perinatal period need to be considered in the application of these scales in future longitudinal and treatment investigations.

References

Adouard F, Glangeaud-Freudenthal NM, Golse B. Validation of the Edinburgh postnatal depression scale (EPDS) in a sample of women with high-risk pregnancies in France. Archives of Womens Mental Health 2005;8:89–95.

Altshuler LL, Cohen LS, Vitonis AF, Faraone SV, Harlow BL, Suri R, Frieder R, Stowe ZN. The Pregnancy Depression Scale (PDS): a screening tool for depression in pregnancy. Archives of Womens Mental Health 2008;11:277–285.

Andersson L, Sundstrom-Poromaa I, Wulff M, Astrom M, Bixo M. Implications of antenatal depression and anxiety for obstetric outcome. Obstetrics & Gynecology 2004;104:467–476. [PubMed: 15339755]

Austin MP, Hadzi-Pavlovic D, Saint K, Parker G. Antenatal screening for the prediction of postnatal depression: validation of a psychosocial Pregnancy Risk Questionnaire. Acta Psychiatrica Scandinavica 2005;112:310–317. [PubMed: 16156839]

Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. Archives of General Psychiatry 1961;4:561–571. [PubMed: 13688369]

Beck CT. Screening methods for postpartum depression. Journal of Obstetrical Gynecological & Neonatal Nursing 1995;24:308–312.

Bennett HA, Einarson A, Taddio A, Koren G, Einarson TR. Prevalence of depression during pregnancy: Systematic review. Obstetrics & Gynecology 2004;103(4):698–709. [PubMed: 15051562]

Campagne DM. The obstetrician and depression during pregnancy. European Journal of Obstetrics & Gynecology and Reproductive Biology 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. Psychosomatic Medicine 2001;63:830–834. [PubMed: 11573032]

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

Cutrona CE. Causal attributions and perinatal depression. Journal of Abnormal Psychology 1983;92:161–172. [PubMed: 6863731]

Dayan J, Creveuil C, Marks MN, Conroy S, Herlicoviez M, Dreyfus M, Tordjman S. Prenatal depression, prenatal anxiety, and spontaneous preterm birth: a prospective cohort study among women with early and regular care. Psychosomatic Medicine 2006;68:938–946. [PubMed: 17079701]

Diego MA, Jones NA, Field T, Hernandez-Reif M, Schanberg S, Kuhn C, Gonzalez-Garcia A. Maternal psychological distress, prenatal cortisol, and fetal weight. Psychosomatic Medicine 2006;68:747–753. [PubMed: 17012528]

Evans J, Heron J, Francomb H, Oke S, Golding J. Cohort study of depressed mood during pregnancy and after childbirth. British Medical Journal 2001;323:257–260. [PubMed: 11485953]

Felice E, Saliba J, Grech V, Cox J. Validation of the Maltese version of the Edinburgh Postnatal Depression Scale. Archives of Womens Mental Health 2006;9:75–80.

Fergerson SS, Jamieson DJ, Lindsay M. Diagnosing postpartum depression: can we do better? American Journal of Obstetrics & Gynecology 2002;186:899–902. [PubMed: 12015507]

First, MB.; Spitzer, RL.; Gibbon, M.; Williams, JBW. Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Patient Edition (SCID-IP, 11/2002 Revision). Washington, DC: American Psychiatric Press; 2002.

Gavin NI, Gaynes BN, Lohr KN, Meltzer-Brody S, Gartlehner G, Swinson T. Perinatal depression: A systematic review of prevalence and incidence. Obstetrics & Gynecology 2005;106:1071–1083. [PubMed: 16260528]

Gaynes BN, Gavin N, Meltzer-Brody S, Lohr KN, Swinson T, Gartlehner G, Brody S, Miller WC. Perinatal depression: prevalence, screening accuracy, and screening outcomes. Evidence Report Technology Assessment 2005;119:1–8. [PubMed: 15760246]

Gordon TE, Cardone IA, Kim JJ, Gordon SM, Silver RK. Universal perinatal depression screening in an Academic Medical Center. Obstetrics & Gynecology 2006;107:342–347. [PubMed: 16449122]

Gotlib IH, Whiffen VE, Mount JH, Milne K, Cordy NI. Prevalence rates and demographic characteristics associated with depression in pregnancy and the postpartum. Journal of Consulting & Clinical Psychology 1989;57:269–274. [PubMed: 2785127]

Hamilton M. A rating scale for depression. Journal of Neurology, Neurosurgery & Psychiatry 1960;23:56–62.

Hedegaard M, Henriksen TB, Sabroe S, Secher NJ. Psychological distress in pregnancy and preterm delivery. British Medical Journal 1993;307:234–239. [PubMed: 8369684]

Kendell RE, Chalmers JC, Platz C. Epidemiology of puerperal psychoses. British Journal of Psychiatry 1987;150:662–673. [PubMed: 3651704]

Kumar R, Robson KM. A prospective study of emotional disorders in childbearing women. British Journal of Psychiatry 1984;144:35–47. [PubMed: 6692075]

Mancuso RA, Schetter CD, Rini CM, Roesch SC, Hobel CJ. Maternal prenatal anxiety and corticotropin-releasing hormone associated with timing of delivery. Psychosomatic Medicine 2004;66:762–769. [PubMed: 15385704]

Morris-Rush JK, Freda MC, Bernstein PS. Screening for postpartum depression in an inner-city population. American Journal of Obstetrics & Gynecology 2003;188:1217–1219. [PubMed: 12748483]

Munk-Olsen T, Laursen TM, Pedersen CB, Mors O, Mortensen PB. New parents and mental disorders: a population-based register study. JAMA 2006;296:2582–2589. [PubMed: 17148723]

Neggers Y, Goldenberg R, Cliver S, Hauth J. The relationship between psychosocial profile, health practices, and pregnancy outcomes. Acta Obstetricia et Gynecologica Scandinavica 2006;85:277–285. [PubMed: 16553174]

Newport DJ, Brennan PA, Green P, Ilardi D, Whitfield TH, Morris N, Knight BT, Stowe ZN. Maternal depression and medication exposure during pregnancy: Comparison of maternal retrospective recall to prospective documentation. British Journal of Obstetrics & Gynecology 2008;115:681–688.

Obuchowski NA. Nonparametric analysis of clustered ROC curve data. Biometrics 1997;53:567–578. [PubMed: 9192452]

O'Hara MW, Schlechte JA, Lewis DA, Wright EJ. Prospective study of postpartum blues: Biologic and psychosocial factors. Archives of General Psychiatry 1991;48:801–806. [PubMed: 1929770]

O'Hara MW, Swain AM. Rates and risk of postpartum depression-a meta-analysis. International Review of Psychiatry 1996;8:37–54.

Orr S, Miller C. Maternal depressive symptoms and the risk of poor pregnancy outcome. Epidemiologic Reviews 1995:165–170. [PubMed: 8521934]

Paton S, Kessler R, Kandel D. Depressive mood and adolescent illicit drug use: a longitudinal analysis. Journal of Genetic Psychology 1977;131(2d Half):267–289. [PubMed: 591931]

Perfetti J, Clark R, Fillmore CM. Postpartum depression: identification, screening, and treatment. Wisconsin Medical Journal 2004;103:56–63. [PubMed: 15622821]

Pritchard CW, Teo PYK. Preterm birth, low birthweight and the stressfulness of the household role for pregnant women. Social Science & Medicine 1994;38:89–96. [PubMed: 8146719]

Rubertsson C, Wickberg B, Gustavsson P, Rådestad I. Depressive symptoms in early pregnancy, two months and one year postpartum-prevalence and psychosocial risk factors in a national Swedish sample. Archives of Womens Mental Health 2005;8:97–104.

Steer R, Scholl T, Hediger M, Fischer RL. Self-reported depression and negative pregnancy outcomes. Epidemiology 1992;45:1093–1099.

Tape, TG. The area under an ROC curve. 2008 [Accessed August 25, 2008.]. Available from: http://gim.unmc.edu/dxtests/roc3.htm

Thoppil J, Riutcel TL, Nalesnik SW. Early intervention for perinatal depression. American Journal of Obstetrics & Gynecology 2005;192:1446–1448. [PubMed: 15902133]

Vesga-Lopez O, Blanco C, Keyes K, Olfson M, Grant BF, Hasin DS. Psychiatric disorders in pregnant and postpartum women in the United States. Archives of General Psychiatry 2008;65:805–815. [PubMed: 18606953]

Watson JP, Elliott SA, Rugg AJ, Brough DI. Psychiatric disorder in pregnancy and the first postnatal year. British Journal of Psychiatry 1984;144:453–462. [PubMed: 6733369]

Williams JB. A structured interview guide for the Hamilton Depression Rating Scale. Archives of General Psychiatry 1988;45:742–747. [PubMed: 3395203]

Zuckerman B, Bresnahan K. Developmental and behavioral consequences of prenatal drug and alcohol exposure. Pediatric Clinics of North America 1991;38:1387–1406. [PubMed: 1719469]

Table 1

Demographics for the Entire Population, the Subpopulation Included in the Analysis and the Subpopulation Excluded from the Analysis

Characteristic	All Enrollees (n=708)	Included in Analysis (n=534)	Excluded Group (n=174)	P-value[a]
Age (years)	33.3 (SD=5.0)	33.1(SD=5.1)	34.0(SD=4.6)	0.03
Gravidity	**682**[b]	**519**	**163**	0.29
1	217(31.8%)	157(30.3%)	60(36.8%)	
2	222(32.6%)	174(33.5%)	48(29.5%)	
3	121(17.7%)	90(17.3%)	31(19.0%)	
≥ 4	122(17.9%)	98(18.9%)	24(14.7%)	
Race	**707**	**534**	**173**	0.04
Asian	15(2.1%)	12(2.3%)	3(1.7%)	
Black/African Am.	56(7.9%)	51(9.6%)	5(2.9%)	
Multi-racial	1(0.1%)	1(0.2%)	0(0.0%)	
Native American	16(2.3%)	12(2.3%)	4(2.3%)	
White/Caucasian	619(87.4%)	458(85.8%)	161(93.1%)	
Ethnicity	**708**	**534**	**174**	0.77
Hispanic	22(3.1%)	16(3%)	6(3.5%)	
Non-Hispanic	686(96.9%)	518(97%)	168(96.5%)	
Marital Status	**708**	**534**	**174**	0.02
Divorced	22(3.1%)	19(3.6%)	3(1.7%)	
Married	619(87.4%)	456(85.4%)	163(93.7%)	
Never Married	38(5.4%)	32(6.0%)	6(3.5%)	
Unmarried/cohabitating	19(2.7%)	18(3.4%)	1(0.6%)	
Separated	9(1.3%)	9(1.7%)	0(0%)	
Widowed	1(0.1%)	0(0.0%)	1(0.6%)	
Times Married	**689**	**518**	**171**	0.26
0	56(8.1%)	48(9.3%)	8(4.7%)	
1	548(79.5%)	404(78.0%)	144(84.2%)	
2	72(10.5%)	57(11.0%)	15(8.8%)	
3	11(1.6%)	8(1.5%)	3(1.8%)	
4	2(0.3%)	1(0.2%)	1(0.6%)	
Education	**697**	**525**	**172**	0.01
≤ High School	44(6.3%)	42(8.0%)	2(1.2%)	
≤ College	403(57.8%)	301(57.3%)	102(59.3%)	
≤ Graduate School	250(35.9%)	182(34.7%)	68(39.5%)	
Pregnancy Planned	**586**	**478**	**108**	<0.0001
Yes	422(72.0%)	325(68.9%)	97(89.8%)	

Characteristic	All Enrollees (n=708)	Included in Analysis (n=534)	Excluded Group (n=174)	P-value[a]
No	164(28.0%)	153(32.0%)	11(10.2%)	

[a]P-values are for comparisons between subpopulations included and excluded in analysis.

[b]Bold numbers are sample sizes of available data for each baseline variable. Due to sporadic missing data, available sample sizes vary for gravidity, race, times married, education

Table 2

Performance of Depression Rating Scales across Perinatal Epochs (n=4025 observations)

Rating Scale	Epoch[a]	SCID MDE Criteria		AUC of ROC	Optimal Cut-point[b]	Sensitivity & Specificity		
		Yes	No			Total	Sens.	Spec.
HRSD$_{17}$	PC	28	265	0.908	15	1.723	0.821	0.902
	T1	72	446	0.881	14	1.597	0.819	0.778
	T2	142	860	0.884	15	1.634	0.782	0.852
	T3	81	706	0.857	15	1.579	0.741	0.839
	PP-E	42	341	0.861	14	1.619	0.810	0.809
	PP-L	115	878	0.877	14	1.618	0.774	0.844
	Overall	480	3496	0.877	14	1.614	0.806	0.807
HRSD$_{21}$	PC	28	264	0.908	16	1.759	0.857	0.902
	T1	72	441	0.873	15	1.597	0.806	0.791
	T2	141	852	0.884	15	1.629	0.823	0.806
	T3	81	705	0.858	16	1.578	0.741	0.837
	PP-E	42	341	0.857	14	1.566	0.810	0.757
	PP-L	115	877	0.874	14	1.602	0.809	0.794
	Overall	479	3480	0.875	15	1.603	0.789	0.814
BDI	PC	28	256	0.938	17	1.780	0.893	0.887
	T1	65	403	0.896	15	1.683	0.877	0.806
	T2	114	780	0.906	13	1.686	0.904	0.782
	T3	66	613	0.871	12	1.587	0.833	0.754
	PP-E	32	294	0.898	14	1.615	0.813	0.803
	PP-L	99	759	0.883	14	1.649	0.828	0.821
	Overall	404	3105	0.895	13	1.648	0.866	0.782
EPDS	PC	7	67	0.971	18	1.842	0.857	0.985
	T1	24	132	0.912	12	1.644	0.833	0.811
	T2	39	246	0.883	9	1.571	0.872	0.699
	T3	6	138	0.949	15	1.797	0.833	0.964
	PP-E	24	167	0.855	11	1.606	0.833	0.772

Rating Scale	Epoch[a]	SCID MDE Criteria		AUC of ROC	Optimal Cut-point[b]	Sensitivity & Specificity		
		Yes	No			Total	Sens.	Spec.
	PP-L	50	447	0.864	12	1.606	0.760	0.846
	Overall	150	1197	0.884	11	1.604	0.813	0.791

[a]Perinatal Epochs: PC=preconception; T1=first trimester; T2=second trimester; T3=third trimester; PP-E=early postpartum; PP-L=late postpartum

[b]Optimal Cutpoint: The cutpoint is the score such that greater or equal scores are considered consistent with the presence of a major depressive episode. The optimal cutpoint is the cutpoint score that maximizes the sum of sensitivity and specificity.

Table 3

Comparative Performance of Depression Rating Scales in Primigravid vs. Multigravid Pregnancies

| | | Primigravid Pregnancies (n=161) | | | | | | | Multigravid Pregnancies (n=465) | | | | | | |
| | | SCID MDE Criteria | | AUC of ROC | Optimal Cut-point[b] | Sensitivity & Specificity | | | SCID MDE Criteria | | AUC of ROC | Optimal Cut-point[b] | Sensitivity & Specificity | | |
Rating Scale	Epoch[a]	Yes	No			Total	Sens.	Spec.	Yes	No			Total	Sens.	Spec.
$HRSD_{17}$	PC	7	79	0.978	14	1.899	1.000	0.899	21	186	0.886	16	1.687	0.762	0.925
	T1	20	121	0.860	12	1.628	0.950	0.678	51	317	0.887	15	1.624	0.804	0.820
	T2	24	242	0.881	14	1.660	0.833	0.826	116	599	0.881	15	1.643	0.802	0.841
	T3	10	177	0.882	13	1.668	0.900	0.768	67	517	0.843	15	1.557	0.731	0.826
	PP-E	8	84	0.861	17	1.702	0.750	0.952	34	251	0.857	14	1.608	0.824	0.785
	PP-L	21	219	0.926	12	1.727	0.905	0.822	92	635	0.861	14	1.595	0.772	0.824
	Overall	90	922	0.896	12	1.655	0.900	0.755	381	2505	0.868	14	1.598	0.806	0.792
$HRSD_{21}$	PC	7	78	0.978	16	1.885	1.000	0.885	21	186	0.882	16	1.718	0.810	0.909
	T1	20	119	0.849	13	1.656	0.950	0.706	51	314	0.880	16	1.622	0.804	0.818
	T2	24	242	0.885	15	1.651	0.792	0.860	115	591	0.880	16	1.620	0.783	0.838
	T3	10	177	0.878	11	1.689	1.000	0.689	67	516	0.845	17	1.559	0.687	0.872
	PP-E	8	84	0.886	17	1.667	0.750	0.917	34	251	0.849	13	1.570	0.853	0.717
	PP-L	21	219	0.921	12	1.729	0.952	0.776	92	635	0.860	14	1.579	0.804	0.775
	Overall	90	919	0.895	15	1.637	0.778	0.860	380	2493	0.867	16	1.587	0.747	0.840
BDI	PC	7	78	0.964	20	1.936	1.000	0.936	21	178	0.928	17	1.750	0.857	0.893
	T1	18	108	0.889	13	1.704	0.944	0.759	46	290	0.900	14	1.682	0.913	0.769
	T2	20	224	0.888	13	1.671	0.850	0.821	93	543	0.906	13	1.676	0.914	0.762
	T3	9	159	0.933	12	1.786	1.000	0.786	57	443	0.854	14	1.552	0.737	0.815
	PP-E	6	75	0.820	10	1.680	1.000	0.680	26	214	0.916	15	1.707	0.885	0.822
	PP-L	15	200	0.906	12	1.692	0.867	0.825	83	539	0.873	15	1.640	0.807	0.833
	Overall	75	844	0.902	12	1.675	0.893	0.782	326	2207	0.890	13	1.637	0.871	0.766
EPDS	PC	1	13	0.769	12	1.769	1.000	0.769	6	54	0.998	18	1.981	1.000	0.981
	T1	6	33	0.932	9	1.758	1.000	0.758	18	98	0.912	13	1.635	0.778	0.857
	T2	10	79	0.908	11	1.635	0.800	0.835	29	167	0.871	9	1.561	0.897	0.665
	T3	1	38	0.737	9	1.711	1.000	0.711	5	99	0.988	15	1.970	1.000	0.970

Rating Scale	Epoch[a]	Primigravid Pregnancies (n=161)							Multigravid Pregnancies (n=465)						
		SCID MDE Criteria		AUC of ROC	Optimal Cut-point[b]	Sensitivity & Specificity			SCID MDE Criteria		AUC of ROC	Optimal Cut-point[b]	Sensitivity & Specificity		
		Yes	No			Total	Sens.	Spec.	Yes	No			Total	Sens.	Spec.
	PP-E	3	41	0.890	11	1.732	1.000	0.732	21	123	0.850	11	1.598	0.810	0.789
	PP-L	8	127	0.848	13	1.781	0.875	0.906	42	309	0.863	11	1.568	0.762	0.806
	Overall	29	331	0.884	11	1.649	0.828	0.822	121	850	0.883	11	1.591	0.810	0.781

[a]Perinatal Epochs: PC=preconception; T1=first trimester; T2=second trimester; T3=third trimester; PP-E=early postpartum; PP-L=late postpartum

[b]Optimal Cutpoint: The cutpoint is the score such that greater or equal scores are considered consistent with the presence of a major depressive episode. The optimal cutpoint is the cutpoint score that maximizes the sum of sensitivity and specificity.

Table 4

Comparative Performance of Depression Rating Scales at First Visit *vs.* Subsequent Visits

Rating Scale	Epoch[a]	First Visits (n=356)							Subsequent Visits (n=2474)						
		SCID MDE Criteria		AUC of ROC	Optimal Cut-point[b]	Sensitivity & Specificity			SCID MDE Criteria		AUC of ROC	Optimal Cut-point[b]	Sensitivity & Specificity		
		Yes	No			Total	Sens.	Spec.	Yes	No			Total	Sens.	Spec.
$HRSD_{17}$	PC	6	51	0.851	15	1.696	0.833	0.863	10	150	0.990	16	1.947	1.000	0.947
	T1	38	84	0.862	15	1.630	0.868	0.762	20	201	0.849	14	1.551	0.750	0.801
	T2	52	80	0.880	15	1.629	0.904	0.725	67	465	0.857	14	1.570	0.761	0.809
	T3	21	21	0.861	15	1.571	0.905	0.667	42	403	0.847	13	1.575	0.833	0.742
	Overall	117	236	0.877	15	1.648	0.889	0.759	139	1219	0.866	14	1.595	0.784	0.811
$HRSD_{21}$	PC	6	50	0.832	15	1.693	0.833	0.860	10	150	0.986	17	1.960	1.000	0.960
	T1	38	83	0.856	16	1.599	0.816	0.783	20	200	0.843	13	1.575	0.850	0.725
	T2	52	79	0.886	17	1.630	0.808	0.823	67	461	0.856	15	1.579	0.761	0.818
	T3	21	20	0.856	22	1.614	0.714	0.900	42	403	0.846	16	1.587	0.738	0.849
	Overall	117	232	0.876	15	1.610	0.889	0.721	139	1214	0.864	15	1.592	0.770	0.822
BDI	PC	6	46	0.846	12	1.652	1.000	0.652	10	148	0.982	17	1.932	1.000	0.932
	T1	33	72	0.877	18	1.654	0.848	0.806	19	187	0.871	15	1.666	0.842	0.824
	T2	42	68	0.866	19	1.626	0.714	0.912	58	423	0.901	13	1.683	0.879	0.804
	T3	17	17	0.824	12	1.588	0.941	0.647	34	353	0.857	12	1.566	0.824	0.742
	Overall	98	203	0.871	19	1.587	0.735	0.852	121	1111	0.893	13	1.646	0.851	0.795
EPDS	PC	0	8	NA^c	NA	NA	NA	NA	6	49	0.998	16	1.980	1.000	0.980
	T1	11	20	0.868	17	1.636	0.636	1.000	10	73	0.928	13	1.690	0.800	0.890
	T2	12	23	0.873	17	1.580	0.667	0.913	23	165	0.860	6	1.533	1.000	0.533
	T3	0	1	NA	NA	NA	NA	NA	6	101	0.959	15	1.814	0.833	0.980
	Overall	23	52	0.849	17	1.614	0.652	0.962	45	388	0.907	13	1.623	0.733	0.889

[a] Perinatal Epochs: PC=preconception; T1=first trimester; T2=second trimester; T3=third trimester

[b] Optimal Cutpoint: The cutpoint is the score such that greater or equal scores are considered consistent with the presence of a major depressive episode. The optimal cutpoint is the cutpoint score that maximizes the sum of sensitivity and specificity.

[c] Not applicable when the number of subjects satisfying MDE criteria is 0.