



Published in final edited form as:

J Womens Health (Larchmt). 2008 May ; 17(4): 585–596. doi:10.1089/jwh.2006.0248.

Screening for Depression in the Postpartum Period: A Comparison of Three Instruments

BARBARA H. HANUSA, Ph.D.¹, SARAH HUDSON SCHOLLE, Dr.P.H.¹, ROGER F. HASKETT, M.D.¹, KATHLEEN SPADARO, M.Ed., R.N., C.N.S.², KATHERINE L. WISNER, M.D., M.S.^{1,3}

¹Department of Psychiatry, School of Medicine, School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania.

²School of Nursing, School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania.

³Departments of Psychiatry and Obstetrics and Gynecology and Reproductive Sciences, School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania.

Abstract

Objectives: Postpartum depression, the most prevalent complication of childbirth, is often unrecognized. Our objective was to compare the effectiveness of three screening instruments—Edinburgh Postnatal Depression Scale (EPDS), Patient Health Questionnaire (PHQ-9), and the 7-item screen of the Postpartum Depression Screening Scale (PDSS)—for identifying women with postpartum depression in the first 6 months after delivery.

Methods: We administered the three instruments via telephone to women who were 18 years and had delivered infants 6–8 weeks earlier. We arranged home interviews to confirm DSM-IV criteria current major depressive disorder (MDD) in women who had an above-threshold score on any of the instruments. For women who screened negative on the 6–8 week call, we repeated the screening at 3 months and 6 months to identify emergent symptoms. The primary outcome measures were the screening scores and DSM-IV diagnoses.

Results: Of 135 women reached, 123 (91%) were screened, 29 (24%) had home visits, and 13 (11%) had an MDD within 6 months of delivery. Analyses of the scores at 6–8 weeks postpartum and the DSM-IV diagnoses indicated the EPDS at a cutoff point of 10 identified 8 (62%) of cases, the PHQ-9 at a cutoff point of 10 identified 4 (31%), and the PDSS 7-item Short Form (PDSS_SF) at a cutoff point of 14 identified 12 (92%). However, 15 of 16 (94%) women without current MDD screened positive on the PDSS_SF. The EPDS was significantly more accurate ($p = 0.01$) than the PDSS_SF and PHQ-9 with the cutoff points used. After correcting for verification bias, we found the EPDS and the PDSS_SF were significantly more accurate than the PHQ-9 ($p < 0.03$).

Conclusions: Administering the EPDS by phone at 6–8 weeks postpartum is an efficient and accurate way to identify women at high risk for postpartum depression within the first 6 months after delivery.

INTRODUCTION

IN THE UNITED STATES, about 14%–15% of women experience depression in the first 3 months after childbirth.^{1,2} Even higher percentages have been reported among innercity women,³ mothers of preterm infants,⁴ and adolescents.^{3–5} Although maternal depression is associated with negative consequences for the mother, her infant, and other family members, low rates of diagnosis and treatment for postpartum depression are common in medical settings.^{6,7} Two critical symptoms of postpartum depression (low energy levels and lack of motivation) coupled with the physical demands of delivery and caring for a new infant interfere with a woman's ability to seek help.⁸ Without treatment, postpartum depression persists for months to years, with limitations in physical and mental functioning lingering even after recovery.^{7–9} Rapid treatment is essential because the episodes are lengthy and psychosocial sequelae increase with the duration of the disorder. Screening is important because it is the first step in the pathway to treatment.^{10,11} The primary objective of this study was to compare the effectiveness of the Edinburgh Postnatal Depression Scale (EPDS), Patient Health Questionnaire (PHQ-9), and 7-item screen from the Postpartum Depression Screening Scale (PDSS) (PDSS_SF) for identifying women who met the criteria for major depressive disorder (MDD), according to the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV), within 6 months of delivery. Secondary objectives were to test the acceptability of phone interviews to screen for depression in the postpartum period and to assess whether a single screening at 6–8 weeks was sufficient to identify the prevalence and incidence of MDD in a sample of women who were depressed within 6 months of delivery.

The EPDS,¹² and the PDSS.¹³ are commonly used to identify women at high risk for postpartum depression. In addition, we included the 9-item depression module of the PHQ-9¹⁴ to determine if this widely used, well-validated measure was an accurate screen for depression in postpartum women. The accuracy of this last measure is of interest because postpartum women may be seen in general healthcare settings as well as obstetrical/gynecological settings.

MATERIALS AND METHODS

Participants and procedures

This study was approved by the Institutional Review Board at the University of Pittsburgh. Women were initially eligible if they were enrolled in the pregnancy/postpartum care management program of a health plan located in Western Pennsylvania that includes members with private insurance and Medicaid. The care management program was coordinated by a nurse case manager who contacted potential women for the study at 2–4 weeks after delivery. This case manager obtained permission for the study team to contact the women. The insurance plan provided information about the type of insurance (private or Medicaid), date of delivery, and phone numbers. At the initial research contact between 6 and 8 weeks postdelivery, our research nurse informed the women about the goals of the study, obtained verbal informed consent, administered the screening tests, and asked a brief series of questions about demographics, delivery, plan to return to work, and current or past treatment for depression, bipolar disorder, or alcohol or drug abuse or dependence. Women were excluded if they were <18 years old, delivered an infant >8 weeks prior to the contact,

delivered before 29 weeks of gestation, reported having bipolar disorder, or reported currently abusing alcohol or illicit drugs. This screening study was part of a larger study that piloted a telephone-based care management program for postpartum depression. This report does not include information on the telephone-based care management program.

Screening tests were administered in the following order: EPDS, PHQ-9, and PDSS_SF. Scores were computed during the screening process. If a woman had negative results on all the instruments at 6–8 weeks after delivery, she was contacted for repeat screenings 3 and 6 months after delivery to identify emergent cases of depression. If a woman had a positive result on the EPDS, PHQ-9, or PDSS_SF at baseline or the 3 or 6 month screening call, she was asked to participate in a home visit to identify current cases of depression. During the home visit, women completed the Diagnostic Interview Schedule (DIS) modules for depression, alcohol abuse and dependence, and drug abuse and dependence.

Study instruments and scores

The 10-item EPDS is the instrument most commonly used to screen for postpartum depression.¹² This measure was designed to reduce the focus on the somatic symptoms that are common both to women in the postpartum period and to women with depression, such as poor sleep quality and weight gain or loss. The cutoff point used to identify women at high risk for MDD varies among studies.^{15,16} Most studies use either the 9/10 cutoff (with 9 for low risk and 10 for high risk) or the 12/13 cutoff (with 12 for low risk and 13 for high risk).^{1,12}

The PHQ-9¹⁴ is a 9-item screen for mood disorders. Each item is rated on how frequently each symptom has occurred over the past 2 weeks and has four possible responses: not at all, several days, more than half the days, and nearly every day. The PHQ-9 was not developed to screen for postpartum depression but was designed to screen for MDD in primary care settings. It has been widely used and validated in these settings.^{14,17} The full PHQ, the self-administered form of the Primary Care Evaluation of Mental Disorders (PRIME-MD),¹⁸ screens for MDD as well as other mood, anxiety, and substance abuse disorders. When Kroenke et al.¹⁴ compared the results from identifying patients with the PHQ-9 vs. the entire PHQ in >3000 primary care patient and >3000 gynecology/obstetrical patients, they found PHQ and PHQ-9 were equally effective at identifying patients with and without MDD. For the PHQ, they reported a specificity of 88% and a sensitivity of 88% for agreement with telephone-based clinical interviews.^{14,17} Unfortunately, they do not separate primary care patients from gynecology/obstetrical patients in the PHQ-9 report.¹⁴ The PHQ-9 has been used successfully to screen for MDD by phone.¹⁹

We used a cutoff point of 10 on the PHQ-9 to schedule a home interview. This score would be indicative of five or more ratings of “more than half the days,” one of the criteria for a DSM-IV diagnosis of MDD. Use of the nine items rather than stopping after the first two reflected our concern that the primary symptom of postpartum depression may not be depressed mood.²⁰ We also used the Spitzer scoring method for the PHQ-9.¹⁷ This method requires that the response to at least five items be “more than half the days” or “nearly every day” and that one of these five items be an item concerning either depressed mood or anhedonia.

The complete 35-item PDSS^{13,21} was developed from discussions with women about their experiences with postpartum depression. An initial screen of seven items is used to identify women at increased risk for postpartum depression. Women with scores ≥ 14 are given an additional 28 items to explore symptoms and severity. Women with scores ≥ 60 on all 35 items are considered to be at risk for minor or major postpartum depression, and those with scores ≥ 80 are considered to have current major depression. According to Beck and Gable's studies of the sensitivity and specificity of the complete PDSS, the ability of this instrument to detect postpartum depression is comparable to the ability of the Structured Clinical Interview for DSM-IV Axis I Disorders (SCID) to detect MDD.^{13,21,22} They also report that it was more accurate than the EPDS and Beck Depression Inventory (BDI) at detecting major and minor depression at 12 weeks postpartum.²¹ The PDSS has a cost per use and initially was developed with white, middle-class samples. A Spanish version was developed and validated,²³ and the PDSS has been used in an economically disadvantaged American Indian sample.²⁴ In office-based settings, an advantage of the PDSS is that it allows clinicians to identify dimensions, for example, sleeping disturbances, guilt and shame, in which women have elevated symptoms, and may help identify the more appropriate treatment options. Beck and Gable²¹ report a high correlation between the complete PDSS score and the EPDS ($r = 0.79$). A recent study has shown that the PDSS can be administered successfully over the telephone and that the PDSS_SF can identify cases of MDD as well as the complete PDSS.²⁵ In the current study, we were interested in the capacity of the PDSS_SF to accurately identify women with MDD.

Home visit criteria.

We planned to use the following cutoff points for positive screens that were recommended by the developers of the instruments: ≥ 10 in the EPDS, ≥ 14 in the PDSS_SF, and ≥ 10 in the PHQ-9. Early in the study, we observed that women who scored greater than the PDSS_SF cutoff point of ≥ 14 (and negative on the other two screens) reported that they did not consider themselves depressed and declined the home visit. Further, the first 10 women who scored positive on the PDSS_SF but none of the other measures did not meet criteria for MDD based on the home interview. After these 10 home interviews with only the PDSS_SF score above the cutoff point, we asked women with PDSS scores ≥ 14 and EPDS scores ≥ 9 and PHQ-9 score ≥ 9 whether they preferred a home visit or a repeat call at 3 and 6 months. A total of 32 women had a positive score of ≥ 14 on the PDSS_SF only; 23 (72%) were offered a home visit and 16 (70%) agreed. Independent of the PDSS_SF score, any woman with an EPDS score ≥ 10 or a PHQ-9 score ≥ 10 was offered a home visit.

Statistical analysis

We used descriptive statistics to characterize study participants in terms of sociodemographic and health-related data, screening results, and DIS results. We used chi-square or Fisher exact tests for categorical variables and analysis of variance (ANOVA) for continuous variables to compare data from women who were positive on different patterns of the three screens. Pearson correlations were used to measure the associations among the scores on the measures.

In our study, we considered the DIS diagnoses to be the criterion standard (gold standard) for the diagnosis of MDD. Therefore, in our estimations of the positive predictive value (PPV), sensitivity, and specificity of each screening instrument, we considered a diagnosis of MDD from the DIS to be a true-positive result. Diagnosis of the 29 women with DIS interviews and dichotomized results on the three screening measures were used to construct receiver operating characteristic curves (ROC) and to compute the areas under the curves (AUC). ROC curves plot the sensitivity of a measure on the Y axis and (1 minus specificity) on the X axis. The AUC, which ranges from 0 to 1.0, is a measure of the accuracy of a test. Screening instruments that identify cases significantly better than chance have AUCs significantly >0.5 , and an AUC >0.80 is expected for an accurate test. Comparison of AUCs created by different measures on the same outcome data are compared with chi-square statistics.²⁶ We also generated the ROC curves for the continuous scores. Differences between the AUCs obtained from the dichotomized and continuous curves were used to address the question whether potential differences among scale performance were due to dichotomizing the scales at suboptimum values.

Verification bias.

Not having resources to perform DIS interviews with women who had negative screens significantly biases our estimates of sensitivity, specificity, PPV, and negative predictive value (NPV). This issue of incomplete assessment of the negative screens is a common problem in medical research and has been labeled verification bias.²⁷ A more compelling example of verification bias occurs when a screening test is used in the decision to perform invasive surgery. Only patients with positive screening tests receive the surgery, which leads to the possibility of verifying the diagnosis. We followed a method proposed by Harel and Zhou²⁷ and built 100 datasets with our observed data at 6–8 weeks postpartum and diagnostic interviews ($n = 29$) in the first 6 months postpartum. We imputed the missing gold standard results (DIS interview) for the 94 women with no home interviews. For each of the 100 datasets, the imputed values were based on the predicted probabilities of a positive result given a combination of the three continuous measures collected at the 6–8-week postpartum phone call. For each of the 100 datasets, we compared the AUCs for the three possible pairings of the measures with a chi-square statistic for comparison of nonindependent measures proposed by DeLong et al.²⁶ The means of the 100 AUCs were used to describe the AUCs that might be obtained, and the means of chi-square for the 300 comparisons²⁶ were used to explore whether our conclusions made with the verified cases about relative accuracy would be sustained if we had complete verification. Stata (Version 8, College Station, TX) was used to generate, summarize, and compare the imputed data. The Stata programs used to complete these tasks are available from the first author.

Secondary aims.

To deal with our secondary aims, we compared groups of women who (1) could be reached by telephone vs. those who could not, (2) completed a screening interview vs. those who did not, and (3) completed a home visit vs. those who did not. We used chi-square or Fisher exact tests for categorical variables and ANOVA for continuous variables to compare data from women in the defined groups.

RESULTS

Participants

Figure 1 shows the number of women involved in each stage of the study. Between July 24, 2002, and January 15, 2003, 203 women agreed to be contacted; only 135 (67%) could be reached by phone. Of the 135 reached, 123 (91%) met all eligibility requirements. Depending on the screening measure used, the prevalence of high levels of depressive symptoms at the 6–8-week postdelivery screen varied from 5% to 36%. Of the 34 women who were offered home interviews at the initial screening, 7 (21%) refused. At the 3 and 6 months screen, another 16 women scored high on the measures, 6 were offered home interviews, and 2 (33%) completed them. Overall, 40 women were offered home visits, 29 (73%) accepted, and 13 (45%) had DIS-identified current diagnosis of MDD. Results from the screening measures at 6–8 weeks after delivery and home interview status are shown in Table 1. Home interviews followed the screening phone calls 2–7 days for 26 (90%) women, 15 days for 1 woman, and 3 months and 6 months from the 6–8-week postdelivery screen for 2 women, but these visits were within 5–7 days of the elevated screens at the later interviews.

Participant characteristics

Women in the study had a mean age of 30.1 years (range 18–41 years). About 72% were white, and 31% were receiving Medicaid. Table 2 displays the characteristics of the women categorized by their pattern of screening results. The majority of women who were screened were married (84%), had 2 or more children (57%), and were planning to return to work (52%). Only 4 (3%) reported that depressive symptoms interfered with daily life. Although 5 (4%) had recently begun antidepressant treatment, 9 (7%) had taken antidepressants sometime in their adult life.

In the 123 women screened at baseline the correlations between the three measures were EPDS with PHQ-9 ($r = 0.75$, $p = 0.000$), EPDS with PDSS_SF ($r = 0.75$, $p = 0.000$), and PHQ-9 with PDSS_SF ($r = 0.71$, $p = 0.000$).

Characteristics and results of tests

Table 3 shows the results from the 29 women who completed the three screening instruments and the DIS (27 at 6–8 weeks after delivery, 1 at 3 months, and 1 at 6 months). There were 13 women who had current MDD based on the DIS interview. In the other 16, 3 had alcohol and drug dependence diagnoses and 13 had no current or past psychiatric disorders. For the EPDS, a cutoff point of 10 yielded the best combination correct positive identifications (8 of 13, 62%) and negative identifications (14 of 16, 88%), which are our best estimates of sensitivity and specificity. The PHQ-9 with a cutoff point of 10 had a specificity of 31% (4 of 13). Using Spitzer's scoring method, the PHQ-9 had a specificity of 8% (1 of 13). The PDSS_SF with a cutoff point of 14 identified 12 of 13 (92%) women with MDD and 1 of 16 (6%) without MDD correctly. Of the 27 women with PDSS_SF 14, only 12 (44%) met criteria for MDD based on the DIS. The corresponding PPVs were 73% (8 of 11) for the EPDS and 80% (4 of 5) for the PHQ-9.

The AUCs for the three scales dichotomized at the chosen cutoff points are shown in Table 4. Comparisons of the AUCs indicated a significant difference (chi-square($df=2$) 9.52, $p=0.01$) with EPDS PDSS_SF PHQ-9. With the use of continuous scales, there were no significant differences among the AUCs generated ($p>0.19$ for all comparisons). The similarity of the AUCs from the continuous scales suggests that the poorer performance of the PDSS_SF and PHQ-9 may be due to suboptimum cutoff points of the continuous scales. At baseline, 34 women also completed the additional 28 items on the PDSS. For this subset of women, the correlation between the PDSS_SF and the complete PDSS was 0.84, $p<0.00$).

Results of ROCs with imputed data

The AUCs and 95% confidence limits for the categorized and continuous measures are displayed in Table 4. For the dichotomized data, comparisons indicated that the EPDS = PDSS_SF ($z=0.47$, $p=0.64$), and both were greater than the PHQ-9 ($z=2.28$, $p=0.02$ and $z=2.06$, $p<0.03$), respectively. Comparisons of the continuous measures AUCs indicated EPDS = PDSS_SF = PHQ-9 ($p>0.23$ for all comparisons). These results mirror the results found with our 29 interviewed cases except that the inclusion of the entire group of women in the analyses the PDSS_SF was as accurate as the EPDS. With the imputed data, we also found the cutoff points for each of the scales that maximized the AUCs possible if three categories of risk for MDD were used. Figure 2 shows graphically the increase in AUCs if three group rather than two group categorization is used to identify women at risk for MDD in the first 6 months postpartum.

Acceptability of telephone screening and home interviews

Women were willing to participate by telephone and home visits. Screening by phone was refused by only 3 of 135 women (2%). Home visits were refused by 10 of 40 women (25%), 5 (13%) who declined any further contact and 5 (13%) who reported that they did not consider themselves to be depressed. We found no significant differences in sociodemographic characteristics of women whom we were and were not able to contact. Women who were reached by phone but did not complete the screening interview were more likely to be nonwhite than were women who completed the screening interview (67% vs. 30%, $p<0.003$). A comparison of women who did ($n=29$) and did not ($n=11$) complete a home visit showed no significant differences in demographic characteristics. Only 1 of the 13 women with MDD (<8%) was identified after baseline. Women with MDD were less likely to be planning to return to work outside the home (8% vs. 44%, $p=0.04$) and were more likely to experience numerous days in which depressive symptoms made functioning difficult (31% vs. 0%; $p=0.01$) than women without a diagnosis of MDD.

DISCUSSION

We administered three depression screening tests via telephone to 123 women and found that from 2% to 36% (depending on the measure) of women had high levels of depressive symptoms at 6–8 weeks after delivery. When the tests were scored at cutoff points recommended by the developers and we used only the women who provided home interviews to measure accuracy, the EPDS was significantly more accurate than the PHQ-9

or the PDSS_SF. In our imputed datasets, the categorized EPDS and the PDSS_SF were equivalently accurate and more accurate than the PHQ-9. Scores on the three continuous measures were highly correlated. Only 2% of the women reached by phone refused the phone screen at 6–8 weeks postpartum, and 12 of 13 (92%) women with MDD in the 6 months after delivery were identified at the 6–8-week call.

Some of our findings are contrary to those reported; others are equivalent. The correlations between the continuous EPDS and PDSS are 0.75 and 0.79¹³ and equivalent. Correlations between the short and complete forms of the PDSS are 0.84 and 0.91.²⁵ For the continuous scales, all three forms rank the women in increasing risk of having MDD, but our results on the accuracy of measures differ. Differences between the reported accuracy of the PHQ-9 in Kroenke et al.¹⁴ and in our study may reflect differences in the women interviewed. For example, Kroenke's validation sample for the PHQ-9 in obstetrical/gynecological patients likely had more gynecological and pregnant women than postpartum women.¹⁴ Differences between the accuracy of the PDSS_SF may reflect our reliance on the 7-item screener rather than the complete PDSS; however, given the high correlation between the short and complete PDSS as well as high correlations reported by Mitchell et al.,²⁵ this seems an incomplete explanation. The results from the imputed data offer a better explanation in that when the likely true negatives were included in the calculations of accuracy, we found that the PDSS_SF and the EPDS were equivalently accurate. This finding supports the possibility that when PDSS_SF is used as a screener, a higher cutoff point would decrease the numbers of false positives without increasing the false negatives.

The EPDS has several advantages. It consists of only 10 items and can be administered in a short period, it can be used free of charge by investigators, it has been translated into 23 languages, and it has been used in various socioeconomic and ethnic groups. In a large study of Australian women, 93% of women found it easy to complete.²⁸ The complete version of the PDSS has advantages in clinical settings where it can be used to identify specific areas that need to be addressed. The primary focus on screening, however, is to identify women at high risk and provide support to seek or continue treatment. Our study suggests that the EPDS accurately identifies women who are depressed or will experience depression in the first 6 months postpartum. Our analyses of continuous screening scores for all instruments indicated that the three measures were equivalently accurate. This suggests that any of these measures is useful for placing women in order of increasing risk for MDD. We did not find that using more than one screening measure led to more accurate identification than using one of the measures with an appropriate cutoff point.

In settings where the PHQ-9 or PDSS_SF is used to screen for postpartum depression, our results suggest that the cutoff points need to be modified. For the PHQ-9, a lower cutoff point of (PHQ-9 = 6) may be more appropriate. In our results, the PDSS with a cutoff point of 14 yielded a high false positive identification rate. One solution suggested by our results is to raise the cut off point to PDSS_SF = 18. Using only the PDSS_SF was also suggested by Mitchell et al.²⁵ These revised cutoff points need to be used with caution before they are validated on larger samples.

Based on our imputed data, splitting each of the measures into three categories, no risk, elevated risk, and high risk for MDD, would increase the diagnostic capacity of all of the measures. Women with the highest scores have a higher risk of current MDD and should be referred for treatment; women with the moderate scores have an elevated risk of MDD, but watchful waiting with further screening and psycho-education may be most appropriate and cost-effective. Finally, the very low scores on the EPDS (< 6) or PDSS_SF (< 13) are excellent indicators that the woman will not meet criteria for MDD in the first 6 months postdelivery. These guidelines for the EPDS are close to the guidelines suggested by Peindl et al.²⁹ based on their high-risk sample of postpartum women with episodes of MDD in previous postpartum periods. In this high-risk sample, the cutoff point for the women with minimal risk was < 4 on the EPDS. Our results raise concerns about using the PHQ-9 as a screen for postpartum depression because of its accuracy. However, the accuracy of this measure with a lower cutoff point and three categories is as high as that of the other measures.

Our study had several limitations that affect its generalizability. The data were collected from a small sample of women who delivered within the same insurance plan. Although we cast a wide net and administered the DIS to any participant who had a positive result on any of the tests, we were not able to administer the DIS to those with negative screening test results. This method of interviewing only women with high symptoms is widely used¹ and cost-effective. Our novel use of simulated data, which was used to address this limitation, provided a clear indication that the EPDS and PDSS are more appropriate instruments to use to identify MDD in the postpartum period. Finally, we did not administer the screening tests in random order.

We found that women were willing to participate in the screening calls, with only 2% refusing to answer questions from all the screening measures. Since we collected our data, Mitchell et al.²⁵ have reported that their telephone screening with the PDSS was acceptable to 84% of their sample. Pinto-Meza et al.¹⁹ reported that telephone screening for MDD with the PHQ-9 was acceptable to >95% of primary care patients. In our sample, only 5 (13%) women completely refused any further contact after a positive screen, and another 5 (13%) women were not interested in the home visit because they did not consider themselves depressed.

Rescreening done at 3 and 6 months was accepted by all women reached. Our conclusion was that very few postpartum women who were expecting a call to screen for depression refused to be screened and very few refused subsequent calls. Close to 90% of the women with MDD in the first 6 months were identified with the phone call at 6–8 weeks postpartum. Telephone screening for postpartum depression at 6–8 weeks postpartum is a time-efficient, well-received method of identifying women at risk for MDD. Screening is only the first step in the recovery process; subsequent work must focus on finding appropriate and acceptable treatment strategies for this vulnerable population.

As more local, state,³⁰ national, and international agencies target the impact of postpartum depression on women, children, families, employers, and the larger society for intervention,

widespread screening is likely to be implemented.³¹ The findings from this study inform the process by identifying an effective instrument.

ACKNOWLEDGMENTS

We acknowledge the contribution of Christopher R.H. Hanusa, Ph.D., in development of the simulated datasets and an anonymous reviewer for the detailed questions that improved this paper.

This study was supported by National Institute of Mental Health grant MH30915 (D.J. Kupfer, MD, Principal Investigator) and by the Staunton Farm Foundation. R.F.H. is a member of the speaker bureaus for Lilly, GlaxoSmithKline and Wyeth. K.L.W. was supported by National Institute of Mental Health grants MH R01 60335 and MH R01 53735 and funding from the Stanley Medical Research Institute. K.L.W. was a member of the speakers bureau for GlaxoSmithKline and has support from Pfizer to study the pharmacokinetics of ziprasidone during pregnancy. No pharmaceutical companies were involved in the design and conduct of the study; the collection, management, analysis, and interpretation of the data; or the preparation, review, or approval of the manuscript.

REFERENCES

1. Gaynes BN, Gavin N, Meltzer-Brody S, et al. Perinatal depression: Prevalence, screening accuracy, and screening outcomes Evidence Report/Technology Assessment (Summary). AHRQ Publication No. 05-E006–2. Rockville, MD: Agency for Healthcare Research and Quality 2 2005.
2. O'Hara MW, Swain AM. Rates and risks of postpartum depression—A meta-analysis. *Int Rev of Psychiatry* 1996;8:37.
3. Hobfoll SE, Ritter C, Lavin J, et al. Depression prevalence and incidence among inner-city pregnant and postpartum women. *J Consult Clin Psychol* 1995;63: 445. [PubMed: 7608357]
4. Logsdon MC, Usui W. Psychosocial predictors of postpartum depression in diverse groups of women. *West J Nurs Res* 2001;23:563. [PubMed: 11569330]
5. Troutman BR, Cutrona CE. Nonpsychotic postpartum depression among adolescent mothers. *J Abnorm Psychol* 1990;99:69. [PubMed: 2307769]
6. Goodman JH. Postpartum depression beyond the early postpartum period. *J Obstet Gynecol Neonat Nurs* 2004;33:410.
7. Kendler KS, Neale MC, Kessler RC, et al. The lifetime history of major depression in women. Reliability of diagnosis and heritability. *Arch Gen Psychiatry* 1993; 50:863. [PubMed: 8215812]
8. O'Hara MW, Stuart S, Gorman LL, et al. Efficacy of interpersonal psychotherapy for postpartum depression. *Arch Gen Psychiatry* 2000;57:1039. [PubMed: 11074869]
9. Holden JM Postnatal depression: Its nature, effects, and identification using the Edinburgh Postnatal Depression Scale. *Birth* 1991;18:211. [PubMed: 1764150]
10. Georgiopoulos AM, Bryan TL, Wollen P, et al. Routine screening for postpartum depression. (Erratum appears in *J Fam Pract* 2001; 50:470). *J Fam Pract* 2001;50:117. [PubMed: 11219558]
11. Bryan TL, Georgiopoulos AM, Harms RW, et al. Incidence of postpartum depression in Olmsted County, Minnesota. A population-based, retrospective study. *J Reprod Med* 1999;44:351. [PubMed: 10319305]
12. Cox J, Holden J, Sagovsky R, et al. Detection of postnatal depression: Development of the 10-item Edinburgh Postnatal Depression Scale. *Br J Psychiatry* 1987;150:782. [PubMed: 3651732]
13. Beck CT, Gable RK. Postpartum Depression Screening Scale: Development and psychometric testing. *Nurs Res* 2000;49:272. [PubMed: 11009122]
14. Kroenke K, Spitzer R, Williams J. The PHQ-9 validity of a brief depression severity measure. *J Gen Intern Med* 2001;16:606. [PubMed: 11556941]
15. Murray L, Carothers A. The validation of the Edinburgh Post-natal Depression Scale on a community sample. *Br J Psychiatry* 1990;157:288. [PubMed: 2224383]
16. Lawrie T, Hofmeyr G, deJager M, et al. Validation of the Edinburgh Postnatal Depression Scale on a cohort of South American women. *South Afr Med J* 1998; 88:1340.

17. Spitzer RL, Kroenke K, Williams BW, et al. 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. [PubMed: 10568646]
18. Spitzer RL, Williams JB, Kroenke K, et al. Validity and utility of the PRIME-MD patient health questionnaire in assessment of 3000 obstetric-gynecologic patients: The PRIME-MD Patient Health Questionnaire Obstetrics-Gynecology Study. Am J Obstet Gynecol 2000; 183:759. [PubMed: 10992206]
19. Pinto-Meza A, Serrano-Blanco A, Peñarubia MT, et al. Assessing depression in primary care with the PHQ-9: Can it be carried out over the telephone? J Gen Intern Med 2005;20:738. [PubMed: 16050884]
20. Beck CT, Inman P. The many faces of postpartum depression. J Obst, Gynecol Neonat Nurs 2005;34:569.
21. Beck CT, and Gable RK. Further validation of the Postpartum Depression Screening Scale. Nurs Res 2001;50:155. [PubMed: 11393637]
22. Beck CT, Gable RK. Comparative analysis of the performance of the Postpartum Depression Screening Scale with two other depression instruments. Nurs Res 2001;50:242. [PubMed: 11480533]
23. Beck CT, Gable RK. Screening performance of the Postpartum Depression Screening Scale—Spanish version. J Transcultural Nurs 2005;16:331.
24. Baker L, Cross S, Greaver L, et al. Prevalence of postpartum depression in a Native American population. Maternal Child Health J 2005;9:21.
25. Mitchell AM, Mittelstaedt ME, Schott-Baer D. Postpartum depression: The reliability of telephone screening. Am J Maternal Child Nurs 2006;31:382.
26. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the area under two or more correlated receiver operating characteristic curves: A nonparametric approach. Biometrics 1988;44:837. [PubMed: 3203132]
27. Harel O, Zhou XH Multiple imputation for correcting verification bias. Statistics Med 2006;25:3769.
28. Buirst A, Condon J, Brooks J, et al. Acceptability of routine screening for perinatal depression. J Affective Disord 2006;93:233.
29. Peindl KS, Wisner KL, Hanusa BH. Identifying depression in the first postpartum year: Guidelines for office-based screening and referral. J Affective Disord 2004;80:37.
30. Office of the Governor, State of New Jersey. Corzine signs postpartum depression screening bill. www.state.nj.us/governor/news/news/approved/20060413 4 13, 2006
31. Wisner KL, Chambers CH, Sit DK. Postpartum depression: A major public health problem. JAMA 2006;296:2616. [PubMed: 17148727]

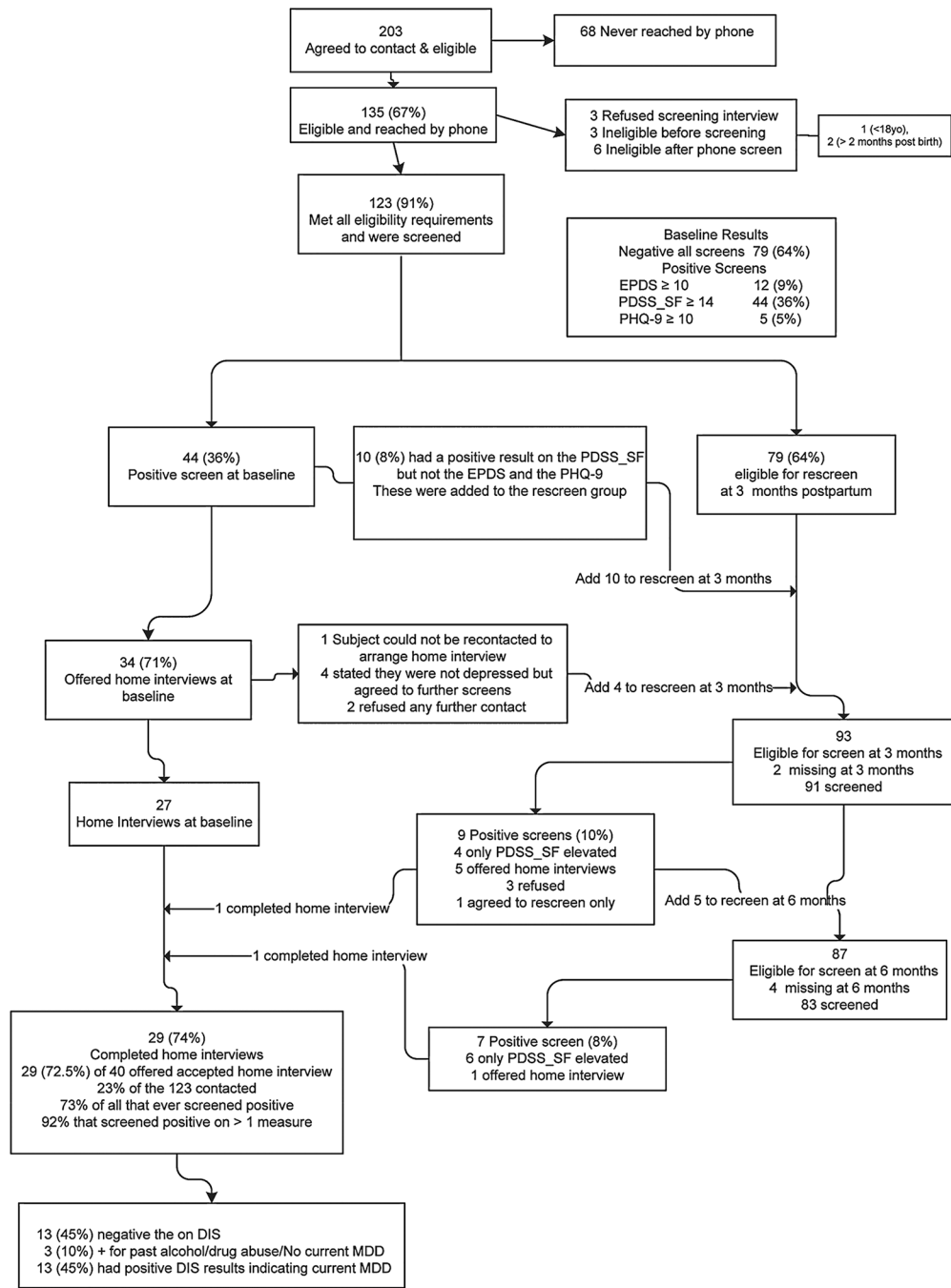


FIG. 1. Flow sheet for recruitment and participation in the study. EPDS, Edinburgh Postnatal Depression Scale; PDSS_SF, 7-item Short Postpartum Depression Screening Scale; PHQ-9, depression module of Patient Health Questionnaire; DIS, Diagnostic Interview Schedule.

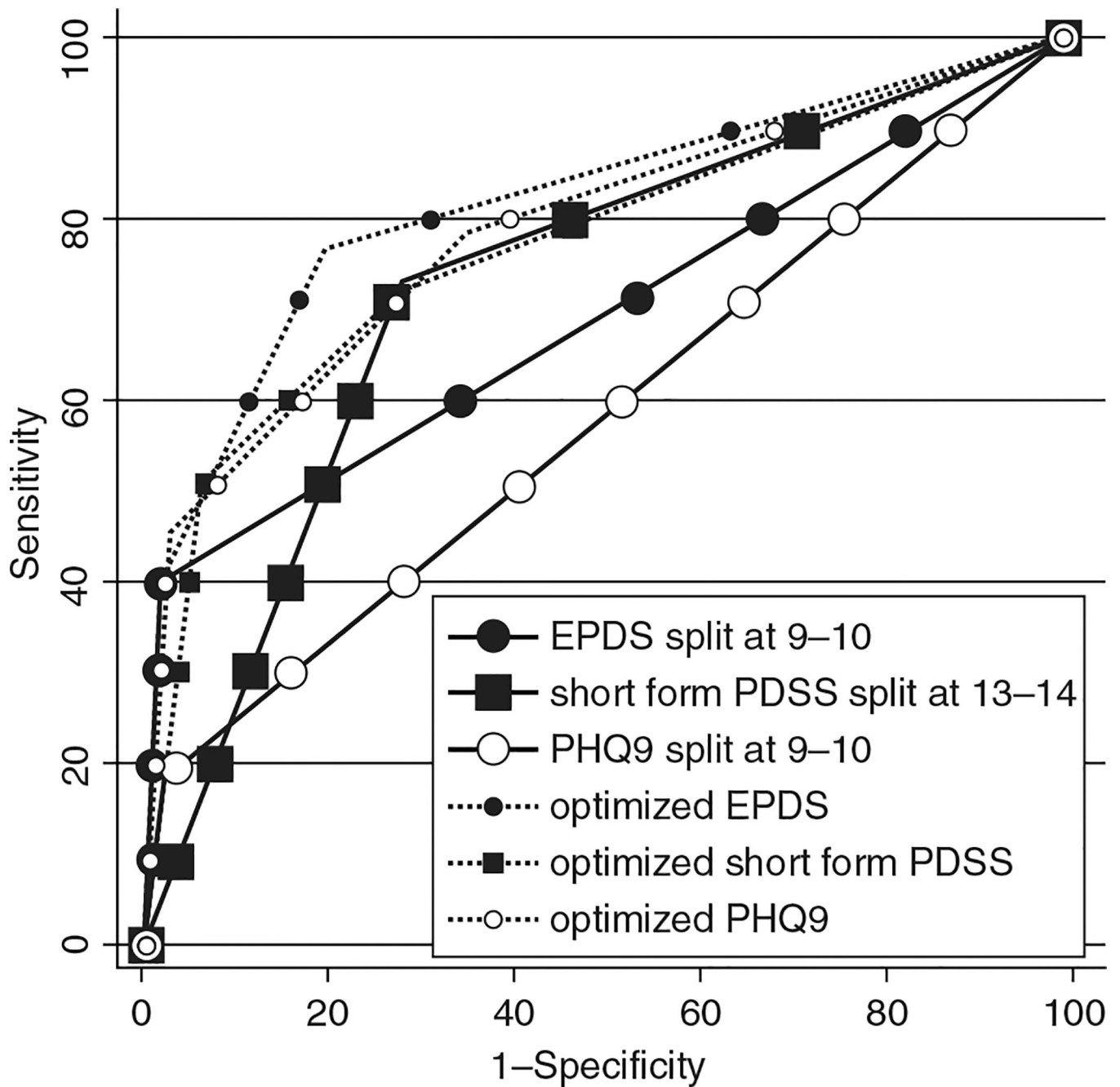


FIG. 2. ROCs for two different sets of cutoff points for identification of postpartum women at risk for MDD. EPDS, Edinburgh Postnatal Depression Scale; PDSS_SF, 7-item Short Postpartum Depression Screening Scale; PHQ-9, Depression module of Patient Health Questionnaire.

Table 1.

Home Visit Status for Women Screened with EPDS, PHQ-9, and PDSS_SF^a

Home interview status	No screen positive at 6–8 weeks postdelivery	PDSS_SF 14		PDSS_SF 14	
		EPDS > 10 PHQ-9 < 10	EPDS > 10 PHQ-9 < 10	EPDS > 10 PHQ-9 < 10	EPDS > 10 PHQ-9 < 10
Not offered	83 (67%)				
All screens negative	74 (88%)	0	0	0	0
PDSS only	9 (12%)	9 (100%)	0	0	0
Offered	40 (32%)	5 ^b	23	7	5
Visit made	29 (73%)	2 (7%)	16 (55%)	6 (21%)	5 (17%)
Refused	5 (13%)	2 (40%)	2 (40%)	1 (20%)	
Lost before interview	1 (3%)	0	1 (100%)		
Refused home interview / accepted rescreen	5 (13%)	1 (20%)	4 (80%)		

^aEPDS, Edinburgh Postnatal Depression Scale; PHQ-9, depression module of the Patient Health Questionnaire; PDSS_SF, 7-item Short Postpartum Depression Screening Scale.

^bThese women had positive screens from interviews at 3 and 6 months.

Table 2.

Demographic Characteristics of Women with Positive Scores on EPDS, PDSS_SF, and PHQ-9^a at 6–8 Weeks Postdelivery

Sociodemographics	No screen positive at 6–8 weeks postdelivery	Results of baseline screening				Statistic	
		PDSS_SF 14 EPDS < 10 PHQ-9 < 10 32	PDSS_SF 14 EPDS 10 PHQ-9 < 10 7	PDSS_SF 14 EPDS 10 PHQ-9 5	PDSS_SF 14 EPDS 10 PHQ-9 10		
Age in years	30.6 (5.7)	28.7 (5.5)	29.4 (5.2)	31.6 (2.4)	$F(3,119) = 1.05, p = 0.4$		
	18–41	19–38	22–36	29–35			
Race/ethnicity							
White	57	72%	3	43%	4	80%	Chi-square = 4.52, $p = 0.52$
African American	14	18%	2	29%	1	20%	
Other	8	10%	2	29%	0	0%	
Type of Insurance							
Commercial	61	77%	1	14%	1	20%	Chi-square = 16.13, $p < 0.001$
Medicaid	18	23%	6	86%	4	80%	
Married (or long-term partner)	69	89%	4	67%	3	60%	Chi-square = 4.47, $p = 0.11$
Planning to return to work	47	60%	1	17%	1	20%	Chi-square = 8.18, $p = 0.04$
Number of children in addition to new infant							
None	32	41%	1	17%	2	40%	Chi-square = 6.48, $p = 0.35$
One	23	30%	2	33%	1	20%	
Two or more	23	30%	3	50%	2	40%	
Functional impairment							
Not at all	68	86%	2	29%	1	20%	Chi-square = 26.26, $p < 0.001$
Several days	11	14%	4	57%	2	40%	
More than half the days	0	0%	1	14%	2	40%	

^aEPDS, Edinburgh Postnatal Depression Scale; PHQ-9, depression module of the Patient Health Questionnaire; PDSS_SF, 7-item Short Postpartum Depression Screening Scale.

Table 3.

Results of DIS^a Interview by Screening Result at 6–8 Weeks

Home interview status	No screen positive at 6–8 weeks postdelivery	PDSS_SF ² 14 EPDS < 10 PHQ-9 < 10	PDSS_SF 14 EPDS 10 PHQ-9 < 10	PDSS_SF 14 EPDS 10 PHQ-9 10
Each screen separately	2 ^b (7%)	16 (55%)	6 (21%)	5 (17%)
DIS diagnostic result				
MDD positive				
No	1	12	2	1
Yes	1 (50%)	4 (25%)	4 (67%)	4 (80%)
Screens considered jointly	<i>No base screen positive</i>	<i>PDSS_SF 14</i>	<i>EPDS 10</i>	<i>PHQ-9 10</i>
DIS diagnostic result	2 ^b (7%)	27 (93%)	11 (38%)	5 (17%)
No MDD	1	15	3	1
% of true negative cases correctly identified	50%	6%	81%	94%
% of negative screen that were correct	50%	50%	72%	62%
Yes, MDD	1	12	8	4
% of true positive cases correctly identified	0	92%	62%	31%
% of positive screens that were correct		44%	72%	80%

^aDIS, Diagnostic Interview Schedule;

^bThese women had positive screens from interviews at 3 and 6 months. EPDS, Edinburgh Postnatal Depression Scale; PHQ-9, depression module of Patient Health Questionnaire; PDSS_SF, 7-item Short Postpartum Depression Screening Scale.

Table 4.

Areas under ROC Curves for Observed and Imputed Data, Continuous, Dichotomized, and Trichotomized Scales

	Continuous		Dichotomized scales used in current study		Trichotomized scales suggested by imputed ROC analyses	
	ROC	95%CI		ROC	95% CI	
Observed						
EPDS	0.84	0.70–0.99	0–9 vs. 10	0.75	0.58–0.91	
PHQ-9	0.79	0.62–0.97	0–9 vs. 10	0.59	0.44–0.75	
PDSS_SF	0.73	0.53–0.93	7–13 vs. 14	0.49	0.39–0.59	
Imputed						
	ROC	95%CI		ROC	95% CI	
EPDS	0.88	0.81–0.95	0–9 vs. 10	0.69	0.65–0.75	0–5, 6–9, 10
PHQ-9	0.80	0.72–0.88	0–9 vs. 10	0.59	0.56–0.62	0–2, 3–6, 7
PDSS_SF	0.82	0.74–0.90	7–13 vs. 14	0.73	0.67–0.80	7–13, 14–17, 18

EPDS indicates Edinburgh Postnatal Depression Scale; PDSS_SF, 7-item Short Postpartum Depression Screening Scale; PHQ-9, depression module of the Patient Health Questionnaire.