

# Concordance of Edinburgh Postnatal Depression Scale (EPDS) and Patient Health Questionnaire (PHQ-9) to Assess Increased Risk of Depression among Postpartum Women

Barbara P. Yawn, MD, MSc, Wilson Pace, MD, Peter C. Wollan, PhD, Susan Bertram, BSN, MSN, Margary Kurland, BSN, Deborah Graham, MSPH, and Allen Dietrich, MD

**Objectives:** To compare the Edinburgh Postnatal Depression Scale (EPDS) and Patient Health Questionnaire (PHQ-9) as screening tools for postpartum depression.

**Methods:** This study population included the first 500 women to enroll and return their packets during an ongoing study of postpartum depression.

**Results:** The primary outcome of this study was to find rates of concordance and discordance in the EPDS and PHQ-9 categories of “normal” and “increased risk for major depressive disorder.” Overall, 97% of eligible women enrolled and 70% returned the packets that included the EPDS and PHQ-9. Four hundred eighty-one of the first 500 packets had complete data, with elevated EPDS or PHQ-9 scores in 138 and 132 women, respectively. Concordance of the EPDS and PHQ-9 were present in 399 women (83%): 326 (67.8%) had “normal” score on both, and 73 (15.2%) had elevated scores for both. Discordant scores in 82 women included 17 with elevated PHQ-9 scores but normal EPDS scores and 65 with elevated EPDS scores and PHQ-9 scores <10. In multivariate logistic regression modeling, only age >30 and low education level were predictive of discordant scores, using EPDS and PHQ-9 scores of  $\geq 10$  as elevated (odds ratio, 1.9 and  $P = .02$ ; and odds ratio, 2.3 and  $P = .01$ , respectively). PHQ-9 scores of 5 to 9 have been referred to as consistent with “mild depressive symptoms” and appropriate for “watchful waiting” and repeat PHQ-9 at follow-up. Using this follow-up approach would require re-evaluation of 120 (25%) of the women screened.

**Conclusions:** Postpartum depression screening is feasible in primary care practices, and for most women the EPDS and PHQ-9 scores were concordant. Further work is required to identify reasons for the 17% discordant scores as well as to provide definitive recommendations for PHQ-9 scores of 5 to 9. (J Am Board Fam Med 2009;22:483–491.)

Primary care office-based screening for postpartum depression (PPD) has been shown to increase recognition and treatment of PPD<sup>1–8</sup> but has not been shown to improve outcomes—such as lower levels of depressive symptoms among the women, greater

parenting comfort, or increased relationship satisfaction between the parents—at 12 months postpartum. Although inadequate outcome data have prevented national recommendations for routine screening for PPD, some large health care and professional organizations, plus a few state legislatures, are recommending or requiring routine PPD screening.<sup>9–12</sup> Most organizations are recommending use of the well-validated Edinburgh Postnatal Depression Scale (EPDS), developed specifically for PPD screening.<sup>13–18</sup>

Routine depression screening has been recommended for all adults<sup>19–24</sup> using tools, such as the Patient Health Questionnaire (PHQ-9),<sup>24</sup> that have been validated in primary care practices.<sup>24–28</sup> None of the tools used for adults have been adequately assessed during the postpartum period. A

This article was externally peer reviewed.

Submitted 23 July 2008; revised 12 November 2008; accepted 21 November 2008.

From the Department of Research, Olmsted Medical Center, Rochester, MN (BPY, PCW, SB, MK); the National Research Network, American Academy of Family Physicians, Leawood, KS (WP, DG); and the Department of Community and Family Medicine, Dartmouth Hitchcock Medical Center, Lebanon, NH (AD).

**Funding:** Funding provided by Agency for Healthcare Research and Quality grant R01 HS014744-01.

**Conflict of interest:** none declared.

**Corresponding author:** Barbara P. Yawn, MD, MSc, Department of Research, Olmsted Medical Center, 210 Ninth Street SE, Rochester, MN 55904 (E-mail: yawnx002@umn.edu).

recent report suggests that a modification of the first 2 questions of the PHQ-9 that had dichotomized responses could be used as a prescreening tool for PPD, but the number of patients in the study was modest and the outcomes inconclusive.<sup>29</sup> If the PHQ-9 could be used for PPD screening, physicians and health care systems might be able to use a single tool for screening all adults for depression.

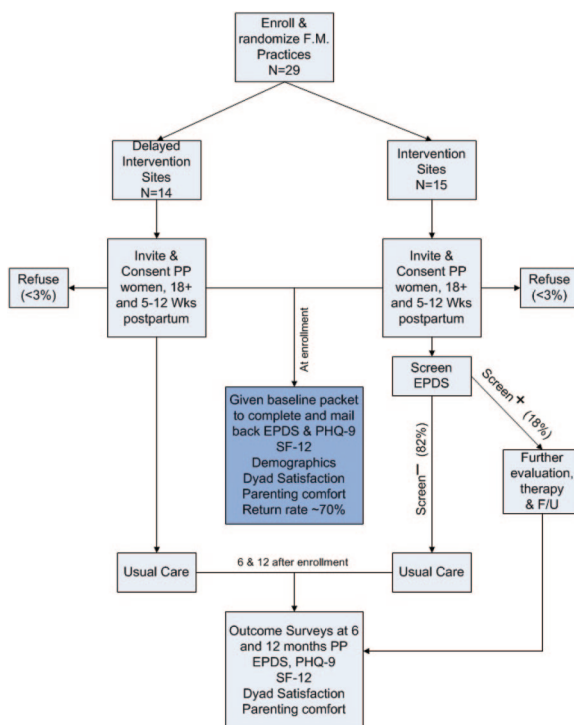
Here we report methods and initial results from the Translating Research into Practice for Postpartum Depression (TRIPPD) study, a large randomized, controlled trial of PPD screening and follow-up in primary care practice-based research network (PBRN) practices. These initial results focus on the concordance and discordance of EPDS and PHQ-9 scores. In addition to the total PHQ-9 and EPDS scores, concordance and discordance of suicidal ideation answers are also discussed. These findings identify issues to consider when using the PHQ-9 for PPD screening.

## Methods

TRIPPD is a randomized, controlled trial assessing the feasibility and impact of a practice change for screening, diagnosis, and follow-up of PPD in 29 PBRN practices associated with the American Academy of Family Physicians National Research Network.<sup>30</sup> Institutional Review Board approval was obtained for all investigators and all sites.

## Design

The 29 practices were selected from PBRN practices that provide maternity or newborn care for at least 50 women or infants each year. Practice sites were randomized to continue usual care or to introduce a PPD practice change that included routine screening of all women 5 to 12 weeks postpartum. The PPD screening led to either a simple review of a low score on the EPDS and return to usual care or additional diagnostic evaluation when the EPDS score was elevated and, if depression was diagnosed, selection of appropriate therapy and a recommended follow-up program (see Figure 1) At the time of enrollment, women were given a packet that included both the EPDS and PHQ-9 for them to complete and return by mail to the central study site, the Olmsted Medical Center. Return rate for the enrollment packet was 70.4%.



**Figure 1. Study diagram. EPDS, Edinburgh Postnatal Depression Scale; PHQ, Patient Health Questionnaire; PP, postpartum; F/U, follow-up.**

## Setting

The 29 practices in the TRIPPD study are all family medicine-based practices, including 15 rural and 14 metropolitan sites. Sixteen of the sites were family medicine residency sites. The 29 practices were in 20 states from Oregon to Mississippi, representing all regions of the country.

## Patients

In both intervention and usual care practices, enrollment was offered to all women 18 years of age or older who spoke and read either English or Spanish (self-reported) and who were 5 to 12 weeks postpartum at the time of the visit. Women who did not return the packet were called and asked if it would be helpful to have someone read the packet to them to allow those women with low reading levels to participate without having to admit they were unable to read either English or Spanish. Women who did not intend to continue care at the participating practice, those who had an emergency condition, those who were outside the window of interest (5 to 12 weeks postpartum), or those who could not read English or Spanish were excluded.

**Table 1. Categories of Edinburgh Postnatal Depression Scale (18) and Patient Health Questionnaire-9 (42) scores**

EPDS		PHQ-9	
Score	Interpretation	Score	Interpretation
0-9	“Normal”	0-4	“Normal”
10, 11	“Slightly increased risk”	5-9	“Mild depressive symptoms”; “repeat PHQ-9 at follow-up”
12-15	“Increased risk”	10-14	“Moderate depressive symptoms”
≥16	“Likely depression”	≥15	“Moderately severe to severe depressive symptoms”

EPDS, Edinburgh Postnatal Depression Scale; PHQ, Patient Health Questionnaire.

The requirement of continuity care was necessary to assure that sites could provide depression management for any mothers identified with PPD.

### Intervention

In preparation for implementation of the intervention, the intervention practices received education about diagnosing, treating, and monitoring PPD, as well as information about the specific intervention procedures and tools that were to be used in this study. Physicians were advised to follow elevated screening scores by a short physician diagnostic interview.<sup>31</sup> Recommendations regarding depression follow-up care and monitoring were based on the national depression management guidelines<sup>32,33</sup> and our work with adult depression in primary care practices<sup>34–38</sup> (Figure 1). Practices assigned to usual care continued to follow their usual approach to recognizing and managing PPD; eg, they did not use standardized screening tools or follow-up procedures.

### Main Instruments

The EPDS is considered a sensitive (96%) but only moderately specific (82%) screening tool for PPD (positive predictive value, 23%) when a score of ≥10 is used as an indication for further assessment.<sup>39</sup> The EPDS has been validated against in-depth interview and mental health assessment for use in postpartum women in all types of care settings and is responsive to improvement in depressive symptoms.<sup>27,40,41</sup>

The PHQ-9 has been validated against in-depth mental health interviews<sup>24,31,41,42</sup> and is reported to be specific (>86% at scores of >10) for identification of people with major depressive disorders (MDD).<sup>24,31,41,42</sup> No studies or subanalyses of published studies have assessed use of the PHQ-9 during the postpartum period. The suggested interpretations of PHQ-9 scores in the general primary

care adult population are shown in Tables 1 and 2.<sup>28,42</sup> The PHQ-9 has been described as both a screening and severity measure for MDD that is responsive to change over time.<sup>42</sup> To make PHQ-9 scoring comparable to EPDS scoring, we explored use of PHQ-9 cut points of ≥5 and ≥10 to trigger further assessment by the clinician without requiring functional impairment. We also explored the impact of scoring the PHQ-9 with and without the requirement that either sadness or anhedonia be present more than half the time during the past 2 weeks.

### Data Collection

Both an EPDS and a PHQ-9 were included in a packet of questions completed by the enrolled women at both control and intervention sites when the women came for their 5- to 12-week postpartum visits. After completing the packets women

**Table 2. Response to Suicidal Ideation Questions on Screeners**

Screening Tool	Suicidal Issues	Response
EPDS	Never	Normal
	Hardly ever	Suicidal ideation requiring evaluation
	Sometimes	Suicidal ideation requiring evaluation
	Often	Suicidal ideation requiring evaluation
PHQ-9	Not at all	Normal
	Several days	Suicidal ideation requiring evaluation
	>Half the days	Suicidal ideation requiring evaluation
	Nearly every day	Suicidal ideation requiring evaluation

EPDS, Edinburgh Postnatal Depression Scale; PHQ, Patient Health Questionnaire.

**Table 3. Demographics of Women Completing EPDS and PHQ-9 (n = 481)**

Age	Married*	Completed High School <sup>†</sup>	Employed Outside Home <sup>‡</sup>	Income >\$30,000/year <sup>§</sup>
18 to 19 (n = 57)	21 (36.8)	46 (80.7)	18 (31.6)	6 (10.5)
20 to 24 (n = 154)	78 (50.6)	125 (81.1)	71 (46.1)	29 (18.8)
25 to 29 (n = 137)	97 (70.8)	119 (86.9)	65 (47.4)	72 (52.6)
30 to 34 (n = 86)	71 (82.6)	81 (94.2)	50 (58.1)	58 (67.4)
≥35 (n = 46)	33 (71.7)	42 (91.3)	28 (60.9)	28 (60.9)

All values provided as n (%). Percents are by age group.

\* $P < .0001$  ( $\chi^2$  test of row variation).

<sup>†</sup> $P = .04$  ( $\chi^2$  test of row variation).

<sup>‡</sup> $P = .01$  ( $\chi^2$  test of row variation).

<sup>§</sup> $P < .0001$  ( $\chi^2$  test of row variation).

EPDS, Edinburgh Postnatal Depression Scale; PHQ, Patient Health Questionnaire.

mailed them directly to the central study site so they would not interfere with the control arm of the study. This process allowed us to compare EPDS and PHQ-9 scores for all enrolled women in both arms of the study.

Other data for the larger study aims were also gathered through the enrollment packet, including demographic information and measures of parent-parent dyad relationship satisfaction (Dyad Assessment of Satisfaction)<sup>43</sup> and parenting comfort (Parenting Stress Inventory),<sup>44</sup> and re-assessed at 6 and 12 months postpartum. To assure patient safety, all sites were informed by telephone and email of very high EPDS and PHQ-9 scores (>19) or a positive response to the suicidal ideation questions that were included in both the EPDS and PHQ-9.

The demographic information and EPDS and PHQ-9 scores collected from the enrollment packets were machine scored and linked for each woman. A 10% sample was hand scored to assess reliability of the machine scoring. No scoring differences were found.

### Data Analysis

For the aims of this study we developed a comparison of EPDS and PHQ-9 scoring categories (Table 1) and used it to determine concordance and discordance in the EPDS and PHQ-9 scores for each woman. Concordance means that both scores are “normal” or that both scores are in the elevated range. Discordant scores have one in the elevated range and one in the “normal” range.

Because this is a first step in exploratory analysis, we used primarily simple summary statistics.  $\chi^2$  statistics were used to assess age trends in other demographic categories such as marital status, in-

come above the poverty level, and completion of a high school education. Demographic factors were examined for association with discordant scores using logistic regression. Results are presented in tables and plots.

### Results

From the first 500 packets returned, 481 packets (96.2%) had both the EPDS and the PHQ-9 as well as demographic information completed and could be included in the analysis. Demographic information for the women is presented in Table 3. The percentage of women who reported being married, being employed outside the home, having an income of >200% of the poverty level, having completed high school, and living in a house with more than one child increased with the increasing age of the woman ( $P < .05$  for all categories). Overall, 76% lived with at least one other adult, 77% reported being white, 23% reported being black, and 2% answered the questionnaires in Spanish.

When all EPDS and PHQ-9 scores were divided into 2 (dichotomous) categories (scores <10 or scores of ≥10), 399 (83%) of the women had concordant scores. Overall, 326 women (67.8%) had both EPDS and PHQ-9 scores in the low risk of depression range and 73 (15.2%) had both scores in the increased risk of MDD range. This left 17% of the women with discordant scores, meaning that one score was <10 and one was ≥10. In univariate modeling, age, household income, marital status, Hispanic ethnicity, and working outside the home were not found to be statistically significant predictors of discordant scores ( $P > .13$  for each). However, not having completed high school was signif-

**Table 4. EPDS Versus PHQ-9 Scores by Category (n = 481)**

EPDS Score	PHQ-9 Score			
	≤4; “Normal”	5-9; “Reevaluate”	10-14; “Increased Risk of Depression”	≥15; “Very Likely to Have Depression”
<0-9; “normal”	249 (51.8)	77 (16.0)	15 (3.1)	2 (0.4)
10, 11; “slightly increased risk”	11 (2.2)	17 (3.5)	7 (1.5)	1 (0.2)
12-15; “increased risk”	11 (2.2)	22 (4.6)	20 (4.2)	11 (2.3)
≥16; “marked risk”	0	4 (0.8)	12 (2.5)	22 (4.6)

All values provided as n (%).

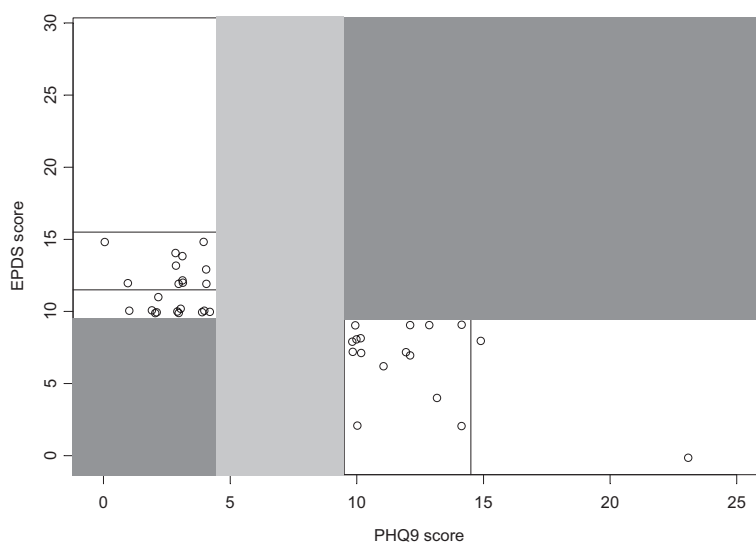
EPDS, Edinburgh Postnatal Depression Scale; PHQ, Patient Health Questionnaire.

icant (odds ratio [OR], 2.0;  $P = .03$ ). In multivariate logistic modeling, age >30 and low education level were both significant (OR, 1.9;  $P = .02$  and OR, 2.3;  $P = .01$ , respectively).

When the low risk of PHQ-9 category is further broken down into 2 groups, with “normal” (scores 0–4) and “slightly increased risk of depressive symptoms” (scores of 5–9) separated, the rate of concordance is lower (n = 322; 66.9%) (see Table 4 and Figure 2). Separation of this “watchful waiting” category<sup>43</sup> from the “normal” category for the PHQ-9 affects the concordance for “normal” (EPDS <10 and PHQ-9 <5), lowering it to only 51.8% (n = 249). In our study, 120 women (25% of all women screened) were included in this watchful waiting category (scores 5–9). Among these 120

women, most (77 of 120; 64.2%) had normal EPDS scores whereas 43 (35.8%) had EPDS scores of >10 and only 4 had scores of >15. Adding a second level of the subgroup analysis by requiring that one of the 2 major criteria for depression must be present at least half the days (questions 1 and 2 from the PHQ-9) before putting women into the watchful waiting group would remove 69 of the 77 women with PHQ-9 scores of 5 to 9 and normal EPDS scores from further evaluation. Alternatively, of the 43 women with PHQ-9 scores in the 5 to 9 range and elevated EPDS scores, 34 have at least one of the 2 major criteria for depression.

For 8.1% of the women the scores were highly discordant, meaning that the EPDS was ≥10 but the PHQ-9 was in the normal range of <5 (n = 22;



**Figure 2. Comparison of EPDS and PHQ-9 scores by categories of risk. Darkly shaded areas are concordant between the level of risk identified by the EPDS and the PHQ-9. The lightly shaded area represents those who fall into the PHQ-9 group that have “mild symptoms and require repeat PHQ-9.”<sup>42</sup> EPDS, Edinburgh Postnatal Depression Scale; PHQ, Patient Health Questionnaire.**

**Table 5. Comparison of Suicidal Ideation from EPDS and PHQ-9**

EPDS “The thought of harming myself has occurred to me”	PHQ-9 “Thoughts that you would be better off dead, or of hurting yourself in some way”			
	Not at All	Several Days	More Than Half the Days	Nearly Every Day
Never	435	10	1	0
Hardly ever	12	10	0	0
Sometimes	3	5	4	0
Often	0	0	0	1

EPDS, Edinburgh Postnatal Depression Scale; PHQ, Patient Health Questionnaire.

4.6%) or the PHQ-9 was in the increased risk of depression range ( $\geq 10$ ) but the EPDS was “normal” ( $n = 17$ ; 3.5% of all women).

Both the EPDS and the PHQ-9 have questions that address suicidal ideation (Tables 2 and 5). The question in the EPDS is active, asking about self harm, whereas the PHQ-9 question is passive, asking if “you have thought you would be better off dead.” These suicidal ideation questions were as likely to be completed as any of the other questions (96% vs. 97%, respectively). Overall, 435 women (90.4%) answered “never” to suicidal ideation questions on both the EPDS and PHQ-9; 5 responded positively to both screening questions. Thirty women (6.2%) had answers that seemed clinically disparate on the 2 instruments.

### Discussion

Postpartum women in this study were willing and able to complete both the PHQ-9 and the EPDS without significant missing data. When separated into only 2 categories of normal versus increased risk of MDD, the EPDS and PHQ-9 score were concordant for the vast majority of women screened (83%). The number of women who would require additional evaluation after screening varied from 138 (28.7%) when using the EPDS to 90 to 210 women (18.7% to 43.7%) when using the PHQ-9, depending on how the scores of 5 to 9 are categorized (as part of “normal” vs “watchful waiting”). Limited data have been published on the outcomes of these 2 different approaches to the PHQ-9 scores of 5 to 9, yet the decision of categorization is crucial to determining the number and burden of the follow-up evaluation of women with “abnormal” scores as well as the rates of false normal screening results. Questions regarding suicidal ideation did not seem to be a barrier to completion of the screening tools; levels of concordance were

very high for the active suicide questions on the EPDS and the passive suicide questions on the PHQ-9, with only 30 women (6.2%) having clinically discordant responses.

The decision of how to deal with PHQ-9 scores in the range between 5 and 9 becomes very important in determining how many women the PHQ-9 recommends for further evaluation. If a score of  $< 10$  is considered normal, 90 women in our study would be recommended for further evaluation of depression. However, if PHQ-9 scores (5–9) are followed by repeat assessment, then 210 women in our study would be appropriate for further evaluation. Kroenke and Spitzer<sup>42</sup> state that an elevated score cut off of  $\geq 10$  provides a sensitivity of 88% for MDD, but they offer little discussion of what is meant by “watchful waiting and possible repeat PHQ-9 at follow-up” for scores of 5 to 9. Suggesting that each woman with a score of 5 to 9 be reassessed adds a substantial burden to PPD screening. Because many postpartum women do not plan another visit to their physician for 12 months or more, the reassessment would require additional health care utilization. Adding a requirement of positive response to PHQ-9 questions about anhedonia or questions about feeling sad (first 2 questions of the PHQ-9) to the 5 to 9 scores would cut the percentage of required reassessments by approximately 45%. Although this additional requirement has been reported to be used in practice, little data about using just the PHQ-2 as a screening tool are published.<sup>42</sup>

The EPDS was developed specifically to avoid over-identification of PPD based on “physical” symptoms such as fatigue, weight and appetite changes, and problems with sleeping that can be suggestive of depression but are a normal part of postpartum recovery.<sup>13,18,45</sup> The PHQ-9 includes these physical symptoms and may therefore over-

identify women in this early period of motherhood.<sup>18</sup> Our data are compatible with this explanation. The comparison of the EPDS and PHQ-9 scores should be assessed later in the postpartum period, when most women have completed breastfeeding and have returned to normal eating and near-normal sleeping habits. These data may also help to determine the value of the “watchful waiting” category of the PHQ-9 for postpartum women. Future data from the TRIPPD study will allow us to complete this long-term comparison.

Our regression analysis suggests that older age and lower levels of completed education may also lead to greater discordance between the EPDS and the PHQ-9. Bennett et al<sup>29</sup> also reported that lower levels of education attained (specifically “did not complete high school”) seemed to increase discordance between the EPDS results and a modification of the first 2 questions of the PHQ-9 used as a “prescreening tool.” These differences require further assessment to determine whether the differences were based on the concepts presented in the questions or the language used to address that content.

The EPDS has been compared with screening tools for depression other than the PHQ-9.<sup>46</sup> Lee et al<sup>47</sup> found that several tools used for non-PPD screening were comparable for identification of women at a high risk of depression. The Chinese cohort was small ( $n = 145$ ) and the study used the Chinese version of all the instruments, but comparisons were based on results of the Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders III-R. In comparison to the Beck Depression Inventory, the EPDS was reported to be “superior.”<sup>16</sup> We found no studies that compared the EPDS and PHQ-9 for postpartum women or any studies that used the PHQ-9 as a screening tool for PPD. A recent study reported on the sensitivity and specificity of modifications of the first 2 PHQ-9 questions (anhedonia and feeling sad or down) for prediction of EPDS scores. The rate of positive response to these modified PHQ-2 questions was reported as compared with rates of EPDS scores of  $\geq 13$  versus  $\leq 12$ . Thirteen is a high cutoff score for risk of depression in community-based screening<sup>18</sup> and the assessment of the PHQ-2 “sensitivity” is based on identifying 4 of 5 women with high EPDS scores who were called “depressed.”<sup>29</sup> The results do not provide information about the possibility of replacing the EPDS

with the PHQ-9 or the PHQ-2, which could allow all depression screening for adults in primary care to use a single tool and might facilitate greater integration of routine depression screening into daily primary care practice.<sup>48</sup>

Addressing suicidal ideation is often uncomfortable for health professionals. Among the 481 women who did complete and return their packets, it was reassuring that we did not find any apparent reluctance to answer the suicidal ideation questions. Using either the PHQ-9 or the EPDS may help facilitate discussions between patients and their health professionals, who might otherwise be reluctant or unsure of how to broach this important topic. Although questions about suicidal ideation were answered at the same rate as the other questions in the depression screeners, it is possible that some women did not return their packets at all because they did not want to answer the suicidal ideation questions. The 30 discordant responses to the suicidal ideation questions may be because of the difference in the questions: the EPDS asks about harming oneself (an active question) and the PHQ-9 asks about being better off dead (a passive approach to suicidal ideation).

Certain limitations of this study should be recognized. Although the study population may not have been representative of the full spectrum of postpartum women, a substantial proportion reported incomes of  $< \$30,000$  a year, the group believed to have the greatest group of stressors. Further, forms could be completed in English or Spanish and women who did not return the packet within 21 days received a call offering help with completion of the forms. This provided those uncomfortable with form completion because of literacy or cultural concerns the opportunity to participate. Finally, identification of depression risk was by means of PHQ-9 and EPDS only. No formal diagnostic interviews for depression were required. The addition of a formal interview, such as the Structured Clinical Interview for DSM Disorders, is impractical in this type of translational study. To make PHQ-9 scoring comparable to EPDS scoring for this analysis we used PHQ-9 scores of  $\geq 5$  as well as  $\geq 10$  without regard to other Diagnostic and Screening Manual IV<sup>49</sup> criteria to trigger clinician assessment of depression. This is in keeping with the focus of this translational study, completed in community practices. Both the EPDS

and the PHQ-9 have been validated against formal psychiatric interviews in several previous studies.<sup>13-18,25,26,28,39,41,42,46</sup>

## Conclusion

These preliminary data from a large randomized control trial of PPD screening and follow-up show that the PHQ-9 and the EPDS have good concordance in identifying those women not at increased risk of PPD. The large number of women (25%) who may require further follow-up because of PHQ-9 scores in the range of 5 to 9 seems excessive but would be reduced substantially if, to be considered positive, the PHQ-9 results were required to include anhedonia or feeling sad more than half the days during the past 2 weeks. Whether the cutoff for the PHQ-9 is 5 or 10 when used for screening and whether the reevaluation of the women with scores in the range of 5 to 9 is a benefit or an additional cost without a benefit requires further evaluation.

---

We thank Dawn Littlefield for her help in preparing this manuscript and the 29 practices and 500 women who participated.

## References

1. Lee DTS, Yip ASK, Chieu HFK, Chung TKH. Screening for Postnatal Depression Using the Double-Test Strategy. *Psychosom Med* 2000;62:258-63.
2. Georgiopoulos AM, Bryan TL, Wollan P, Yawn BP. Routine screening for postpartum depression. *J Fam Pract* 2001;50:117-22.
3. Georgiopoulos AM, Bryan TL, Yawn BP, Houston MS, Rummans TA, Therneau TM. Population-based screening for postpartum depression. *Obstet Gynecol* 1999;93:653-7.
4. Beck CT, Gable RK. Further validation of the postpartum depression screening scale. *Nurs Res* 2001;50:155-64.
5. Massoudi P, Wickberg B, Hwang P. Screening for postnatal depression in Swedish child health care. *Acta Paediatr* 2007;96:897-901.
6. Morris-Rush JK, Freda MC, Bernstein PS. Screening for postpartum depression in an inner-city population. *Am J Obstet Gynecol* 2003;188:1217-9.
7. Evins GG, Theofrastous JP, Galvin SL. Postpartum depression: a comparison of screening and routine clinical evaluation. *Am J Obstet Gynecol* 2000;182:1080-2.
8. Chaudron LH, Szilagyi PG, Kitzman HJ, Wadkins HI, Conwell Y. Detection of postpartum depressive symptoms by screening at well-child visits. *Pediatrics* 2004;113(3 Pt 1):551-8.
9. Armstrong S, Small R. Screening for postnatal depression: not a simple task. *Aust N Z J Public Health* 2007;31:57-61.
10. Peindle KS, Wisner KL, Hanusa BH. Identifying depression in the first postpartum year: guidelines for office-based screening and referral. *J Affect Disord* 2004;80:37-44.
11. Gjerdingen DK, Yawn BP. Postpartum depression screening: importance, methods, barriers, and recommendations for practice. *J Am Board Fam Med* 2007;20:280-8.
12. Goldsmith ME. Postpartum depression screening by family nurse practitioners. *J Am Acad Nurse Pract* 2007;19:321-7.
13. Cox JL, Holden JM, Sagovsky R. Detection of postnatal depression. Development of the 10-item Edinburgh Postnatal Depression Scale. *Br J Psychiatry* 1987;150:782-6.
14. Areias MEG, Kumar R, Barros H, Figueiredo E. Comparative incidence of depression in women and men. During pregnancy and after childbirth validation of the Edinburgh Postnatal Depression scale in Portuguese mothers. *Br J Psychiatry* 1996;169:30-5.
15. Boyce P, Stubbs J, Todd AN. The Edinburgh Postnatal Depression Scale: validation for an Australian sample. *Aust N Z J Psychiatry* 1993;27:472-6.
16. Wickberg B, Hwang CP. Screening for postnatal depression in a population-based Swedish sample. *Acta Psychiatr Scand* 1997;95:62-6.
17. Milgrom J, Ericksen J, Negri L, Gemmill AW. Screening for postnatal depression in routine primary care: properties of the Edinburgh Postnatal Depression Scale in an Australian sample. *Aust N Z J Psychiatry* 2005;39:833-9.
18. Cox J, Holden J, eds. Perinatal psychiatry: use and misuse of the Edinburgh Postnatal depression Scale. London: Gaskell; 1994:145-68.
19. US Department of Health and Human Services, The Agency for Healthcare Research and Quality. Topic Index: A to Z. U.S. Preventive Services Task Force. Available at <http://www.ahrq.gov/clinic/3rduspstf/depression/depressrr.htm>.
20. Pignone MP, Gaynes BN, Rushton JL, et al. Screening for depression in adults: a summary of the evidence for the U.S. Preventive Services Task Force. *Ann Intern Med* 2002;136:765-76.
21. Craig TJ, Van Natta P. Validation of the community mental health assessment interview instrument among psychiatric inpatients. Working Paper #B-27a for the Center for Epidemiologic Studies. Rockville, MD: National Institute of Mental Health; 1973.
22. Weissman MM, Prusoff B, Newberry P. Comparison of the CES-D with standardized depression rating scales at three points in time. Technical Report from Yale University under Contract ASH 74-166. Rockville, MD: Center for Epidemiologic Studies, National Institute of Mental Health; 1975.
23. Beck CT, Ward CH, Mendelson M, et al. An inven-



- tory for measuring depression. *Arch Gen Psychiatry* 1965;12:63–70.
24. Spitzer RL, Kroenke K, Williams JBW, et al. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. *JAMA* 1999;282:1737–44.
  25. Gilbody S, House AO, Sheldon TA. Screening and case finding instruments for depression. *Cochrane Database Syst Rev* 2005;(4):CD002792.
  26. Lowe B, Kroenke K, Herzog W, Grafe K. Measuring depression outcome with a brief self-report instrument: sensitivity to change of the Patient Health Questionnaire (PHQ-9). *J Affect Disord* 2004;81:61–6.
  27. Lowe B, Unutzer J, Callahan CM, Perkins AJ, Kroenke K. Monitoring depression treatment outcomes with the patient health questionnaire-9. *Med Care* 2004;42:1194–201.
  28. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med* 2001;16:606–13.
  29. Bennett I, Coco A, Coyne J, et al. Efficiency of a two-item pre-screen to reduce the burden of depression screening in pregnancy and postpartum: an IMPLICIT Network study. *J Am Board Fam Med* 2008;21:317–25.
  30. Westfall JM, Mold J, Fagnan L. Practice-based research—“Blue Highways” on the NIH roadmap. *JAMA* 2007;297:403–6.
  31. Schulberg H, Lee P, Bruce M, et al. Suicidal ideation and risk levels among primary care patients with uncomplicated depression. *Ann Fam Med* 2005;3:523–8.
  32. U.S. Department of Health and Human Services. Depression in primary care, volume I. Detection and diagnosis. Rockville, MD: Public Health Service; 1993.
  33. Agency for Healthcare Research and Quality: Identification of common mental disorders and management of depression in primary care. Available at [http://www.guidelines.gov/summary/summary.aspx?doc\\_id=12994&nbr=006690&string=depression](http://www.guidelines.gov/summary/summary.aspx?doc_id=12994&nbr=006690&string=depression). Accessed July 16, 2009.
  34. Kroenke K. Depression screening is not enough. *Ann Intern Med* 2001;134:418–20.
  35. Dietrich AJ, Oxman TE, Williams JW, et al. Going to scale: re-engineering systems for primary care treatment of depression. *Ann Fam Med* 2004;2:301–4.
  36. Dejesus RS, Vickers KS, Melin GJ, Williams MD. A system-based approach to depression management in primary care using the Patient Health Questionnaire-9. *Mayo Clin Proc* 2007;82:1395–402.
  37. Brody DS, Dietrich AJ, deGruy F, Kroenke K. The depression in primary care tool kit. *Int J Psychiatry Med* 2000;30:99–110.
  38. Dietrich A, Oxman T, Williams J, et al. Re-engineering systems for the treatment of depression in primary care: cluster randomized controlled trial. *BMJ* 2004;329:602.
  39. Murray L, Carothers A. The validation of the Edinburgh Post-natal Depression Scale on a community sample. *Br J Psychiatry* 1990;157:288–90.
  40. Rogers WH, Adler DA, Bungay KM, Wilson IB. Depression screening instruments made good severity measures in a cross-sectional analysis. *J Clin Epidemiol* 2005;58:370–7.
  41. Gilbody S, Richards D, Brealey S, Hewitt C. Screening for depression in medical settings with the Patient Health Questionnaire (PHQ): a diagnostic meta-analysis. *J Gen Intern Med* 2007;22:1596–602.
  42. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med* 2001;16:606–13.
  43. Spanier GB. Measuring dyadic adjustment: New scales for assessing the quality of marriage and similar dyads. *J Mar Fam* 1976;38:15–28.
  44. Lloyd BH, Abiden RR. Revision of the Parenting Stress Index. *J Pediatr Psychol* 1985;10:167–79.
  45. Wickberg B, Hwang CP. The Edinburgh Postnatal Depression Scale: validation on a Swedish community sample. *Acta Psychiatr Scand* 1996;94:181–4.
  46. Hanna B, Jarman H, Savage S. The clinical application of three screening tools for recognizing postpartum depression. *Int J Nurs Pract* 2004;10:72–9.
  47. Lee DT, Yip AS, Chiu HF, Leung TY, Chung TK. Screening for postnatal depression: are specific instruments mandatory? *J Affect Disord* 2001;63(1–3):233–8.
  48. U.S. Preventive Services Task Force. Screening for depression: recommendations and rationale. *Ann Intern Med* 2002;136:760–4.
  49. American Psychiatric Publishing, Inc. Psychiatry Online DSM library. Available at: [www.psychiatryonline.com/dsmLibrary.aspx](http://www.psychiatryonline.com/dsmLibrary.aspx). Accessed July 4, 2009.