Postpartum Depression Screening at Well-Child Visits: Validity of a 2-Question Screen and the PHQ-9

Dwenda Gjerdingen, MD, MS[1]

Scott Crow, MD[2]

Patricia McGovern, PhD, MPH[3]

Michael Miner, PhD[1]

Bruce Center, PhD[1]

[1]Department of Family Medicine & Community Health, University of Minnesota, Minneapolis, Minnesota

[2]Department of Psychiatry, University of Minnesota, Minneapolis, Minnesota

[3]School of Public Health, University of Minnesota, Minneapolis, Minnesota

ABSTRACT

PURPOSE Postpartum depression affects up to 22% of women who have recently given birth. Most mothers are not screened for this condition, and an ideal screening tool has not been identified. This study investigated (1) the validity of a 2-question screen and the 9-item Patient Health Questionnaire (PHQ-9) for identifying postpartum depression and (2) the feasibility of screening for postpartum depression during well-child visits.

METHODS Study participants were English-literate mothers registering their 0- to 1-month-old infants for well-child visits at 7 family medicine or pediatric clinics. They were asked to complete questionnaires during well-child visits at 0 to 1, 2, 4, 6, and 9 months postpartum. Each questionnaire included 2 depression screens: the 2-question screen and the PHQ-9. The mothers also completed the depression component of the Structured Clinical Interview for *DSM-IV* (SCID) initially, and again at a subsequent interval if either screening result was positive for depression.

RESULTS The response rate was 33%. Of the 506 women who participated, 45 (8.9%) had major depression (ie, they had a positive result on the SCID). The screen sensitivities/specificities over the course of the study were 100%/44% with the 2-question screen, 82%/84% with the PHQ-9 using simple scoring, and 67%/92% with the PHQ-9 using complex scoring. In addition, the corresponding values for the first 2 items of the PHQ-9 (ie, the 2-item Patient Health Questionnaire or PHQ-2) were 84%/79%. Some 38% of women completed their 2- to 6-month questionnaires during well-child visits; the rest completed them by mail (29%) or telephone (33%).

CONCLUSIONS The 2-question screen was highly sensitive and the PHQ-9 was highly specific for identifying postpartum depression. These results suggest the value of a 2-stage procedure for screening for postpartum depression, whereby a 2-question screen that is positive for depression is followed by a PHQ-9. These screens can be easily administered in primary care clinics; feasibility of screening during well-child visits was moderate but may be better in clinics using a mass-screening approach.

CORRESPONDING AUTHOR

Dwenda Gjerdingen, MD, MS
580 Rice St
St Paul, MN 55103
dgjerdin@umphysicians.umn.edu

INTRODUCTION

Postpartum depression is increasingly recognized as a unique and serious complication of childbirth, with an estimated prevalence in the 12-month postpartum period of up to 21.9%.[1] Mothers' depressive symptoms—diminished mood, pleasure, energy, concentration, and self-worth; psychomotor retardation; changes in appetite and sleep; and suicidal ideation—can markedly impair their sense of well-being, marital and other key relationships,[2] work performance and productivity,[3] relationships with their infants,[4] and infants' behavioral and cognitive development.[5] Recognizing the seriousness of this disorder, the US Preventive

Services Task Force has recommended routine depression screening for adults in practices that have systems in place to ensure accurate diagnosis, effective treatment, and follow-up.[6] Most primary care practices do not have such systems in place, for either general or postpartum depression, however.

Studies of postpartum depression screening demonstrate that it is feasible in outpatient clinical settings, either during mothers' postpartum visits[7,8] or during infants' well-child visits,[9-11] with the use of screens such as the Edinburgh Postnatal Depression Scale[7-10] or the 2-item Patient Health Questionnaire (PHQ-2).[11] In 2 recent studies in which maternal depression screening was performed during well-child visits conducted in pediatric offices, mothers' response rates were 55% to 74%.[10,11] Two additional large studies involving a total of 860,479 mothers found that more than 80% said they were comfortable with the idea of being screened for postpartum depression.[12,13] Unknown is the degree of comfort with screening among the 41% to 48% of women who did not respond to the questionnaire, or the level of compliance with screening among those who say they are comfortable. In reality, fewer than 50% of women with infants are currently being screened for postpartum depression.[14-16] Several studies have confirmed that informal assessment or nonassessment for postpartum depression identifies fewer than one-half of cases or potential cases.[7,8,17,18]

Although physicians generally agree that it is important to recognize and treat postpartum depression, actual screening activities are quite variable. In a survey of members of the Washington Academy of Family Physicians, 70% (254 family physicians) said they always or often screened for postpartum depression at the postpartum examination, while 46% did so at well-child checkups; however, only 31% of those reported using a validated tool for screening.[15] Two national surveys of pediatricians found that most do not feel well trained in postpartum depression or confident in their abilities to diagnose the disorder.[19,20]

Clinicians who screen for postpartum depression need a valid, reliable instrument. Two recent reviews evaluated several postpartum depression screens, including the Beck Depression Inventory, Bromley Postnatal Depression Scale, Center for Epidemiological Studies Depression Scale, Edinburgh Postnatal Depression Scale, General Health Questionnaire, Inventory of Depressive Symptomatology, Postpartum Depression Screening Scale, and Zung Self-Rating Depression Scale.[1,21] These reviews found that most of the studies on postpartum depression screens were too small to effectively validate the screens, and external validity was often poor to fair. A third review looked at prenatal screening for women at risk for postpartum depression, using the Edinburgh Postnatal Depression Scale, standardized diagnostic psychiatric interviews, or both; this review concluded that none of the instruments evaluated met the criteria for routine prenatal application.[22] Additional studies with large, representative samples are therefore needed to identify the ideal postpartum depression screen.

Other promising screens not covered in these reviews are the 9-item Patient Health Questionnaire (PHQ-9) and the 2-question screen. The PHQ-9, thought by some experts to be "the best available depression screening tool for primary care,"[23] consists of the *Diagnostic and Statistical Manual,* Fourth Edition (*DSM-IV*) criteria for major depressive disorder and uses a Likert scale response.[24] When evaluated in a sample of 6,000 patients seen in 8 primary care clinics and 7 obstetrics-gynecology clinics, the PHQ-9 had a sensitivity and specificity of 88% (at a cutoff of ≥10), based on a structured interview.[24] In a recent study of postpartum depression, the PHQ-9 correctly identified 4 of 13 cases, compared with 8 of 13 and 12 of 13 identified by the Edinburgh Postnatal Depression Scale and the Postpartum Depression Screening Scale, respectively. This study was limited, however, by its small number of depressed women (n = 13) and use of the formal depression interview only in women whose findings were positive for depression.[25]

The 2-question screen asks about the 2 fundamental symptoms of depression, diminished mood and pleasure, and has simple yes/no responses. In 2 separate large studies, this screen was found to have a sensitivity of 96% and a specificity of 57% to 78%.[26,27] The addition of a help question, which asks whether "this is something with which you would like help," increased the specificity to 94%.[27] The PHQ-2, which consists of the first 2 questions of the PHQ-9, is similar to the 2-question screen except for its Likert scale response.[28] Neither the 2-question screen nor the PHQ-2 has been validated in a purely postpartum population.

The purpose of this study was to determine the validity of the 2-question screen and the PHQ-9 in a sample of postpartum women, and to investigate the feasibility of screening for postpartum depression during well-child visits in family medicine and pediatric clinics. This validation study was conducted within the context of a randomized controlled trial, the overall purpose of which was to begin to assess the impact of stepped care treatment on postpartum depression.

METHODS

Study Overview

Before its inception, the study was approved by the University of Minnesota's institutional review board. Study

participants were recruited over a 12-month period, from October 1, 2005, through September 30, 2006. Mothers registering their infants for an initial well-child visit at 0 to 1 month of age at any of 7 participating clinics were given an enrollment packet consisting of a brief description of the study, an enrollment form (on which they were asked to indicate their willingness to participate), a consent form, and an initial questionnaire. Participants were given follow-up questionnaires at subsequent 2-, 4-, and 6-month well-child visits (or, if they were unable to complete them at these visits, were offered telephone or mailed questionnaires), and were mailed a final questionnaire at 9 months postpartum. Participants were also asked to complete the Structured Clinical Interview for *DSM-IV* (SCID) interview within 2 weeks of completing the initial questionnaire, and subsequently if they had not been previously depressed, and developed a positive screening result on a follow-up questionnaire.

Study Participants

Participants were recruited from 7 Minneapolis and St. Paul metropolitan area clinics. The clinics represented a diverse cross-section of patients, as the 4 family medicine residency clinics served primarily urban, ethnically diverse, low-income groups, whereas the 3 pediatric private clinics saw mostly suburban, white, mid- to upper-income patients. To enroll in the study, mothers needed to be English literate, be aged 12 years or older, and have a 0- to 1-month-old infant who received care at any of the participating clinics.

Measures

The initial questionnaire collected demographic information about the women: age, education, race/ethnicity, total family income, health insurance, marital status, number of children, and delivery date. Both the initial and the follow-up questionnaires contained the 2 depression screens (Table 1): the 2-question screen, which consists of the 2 fundamental symptoms of depression (diminished mood and pleasure), and the PHQ-9, which contains the *DSM-IV* criteria for major depressive disorder.[24] The first 2 questions of the PHQ-9 constitute the PHQ-2,[28] which was also evaluated.

All participants were also asked to complete the depression component of the SCID[29] at 0 to 1 month postpartum and again later if they were previously not depressed but had a positive screen result on a follow-up questionnaire. The SCID interview, our reference standard for diagnosing major depression, evaluates the presence and frequency of 9 symptoms of depression, as well as other potential causes of these symptoms (eg, grief), and the result is scored as positive or negative. The SCID interviews were conducted by doctoral-level psychology students, whose training consisted of observing SCID training tapes and completing 5 practice tapes under the supervision and review of a highly experienced doctoral-level assessor, followed by weekly quality assurance assessment conferences throughout the study. Sixty-eight (13.4%) of the women could not be reached for the initial SCID interview; an additional 9 women could not be reached for a follow-up SCID interview.

Table 1. Depression Screens: 2-Question Screen and PHQ-9

2-Question Screen[26]

1. During the past month, have you often been bothered by feeling down, depressed, or hopeless? (Yes/No)

2. During the past month, have you often been bothered by having little interest or pleasure in doing things? (Yes/No)

PHQ-9[24]	Not at All	Several Days	More Than One-Half the Days	Nearly Every Day
Over the *last 2 weeks*, how often have you been bothered by any of the following problems?				
1. Little interest or pleasure in doing things[a]	0	1	2	3
2. Feeling down, depressed, or hopeless[a]	0	1	2	3
3. Trouble falling or staying asleep, or sleeping too much	0	1	2	3
4. Feeling tired or having little energy	0	1	2	3
5. Poor appetite or overeating	0	1	2	3
6. Feeling bad about yourself—or that you are a failure or have let yourself or your family down	0	1	2	3
7. Trouble concentrating on things, such as reading the newspaper or watching television	0	1	2	3
8. Moving or speaking so slowly that other people could have noticed? Or the opposite—being so fidgety or restless that you have been moving around a lot more than usual	0	1	2	3
9. Thoughts that you would be better off dead or of hurting yourself in some way	0	1	2	3

PHQ-9 = 9-item Patient Health Questionnaire.

[a] Questions 1 and 2 constitute the 2-item Patient Health Questionnaire (PHQ-2).[28]

Developed by Drs. Robert L. Spitzer, Janet B. W. Williams, Kurt Kroenke, and colleagues, with an educational grant from Pfizer, Inc. Copyright 2005, Pfizer, Inc. All rights reserved. Reproduced with permission.

Statistical Analysis

We scored the 2-question screen results as positive if the respondent answered "yes" to either or both of the 2 questions. Validity of the PHQ-9 was assessed using 2 separate scoring methods. For the simple scoring method, we summed the responses and considered a total score of 10 or greater to be positive, based on the cutoff of 10 or greater used in the previously cited PHQ-9 validation study.[24] The complex PHQ-9 scoring method, modeled after the *DSM-IV* definition of depression,[30] requires the presence of 5 or more symptoms, including symptom 1, symptom 2, or both, and each symptom must have a response of 2 or 3 with the exception of symptom 9, for which a response of 1 to 3 was acceptable. Given that the first 2 questions of the PHQ-9 constitute the PHQ-2, we also tested the validity of the PHQ-2; the result was considered to be positive if either question had a response of 2 or greater.

We conducted 2 sets of validity tests (including sensitivity, specificity, negative predictive value, and positive predictive value) for each screen. The first set was performed using only data from the questionnaire completed at 0 to 1 month, whereas the second set was performed using data from the entire study period. For the second set, a woman was said to have a positive 2-question screening result if her result was positive at any time point during the course of the study, and a negative screening result if her result was negative at every time point throughout the study.

To determine whether women who dropped out of the study (did not complete the final questionnaire) differed from women who completed the study, we used bivariate analyses (*t* tests and χ^2 tests) to compare these 2 groups on maternal age, number of children, marital status, education, family income, and positive PHQ-9.

RESULTS

Participants

A total of 506 women participated in the study, which represented approximately 33% of the estimated 1,556 eligible women (response rate, 28% for family medicine clinics and 36% for pediatric clinics). Numbers of eligible women were based on the estimated number of English-literate mothers of newborns seen at participating clinics during the enrollment period. The reasons for nonparticipation among eligible women are shown in Figure 1.
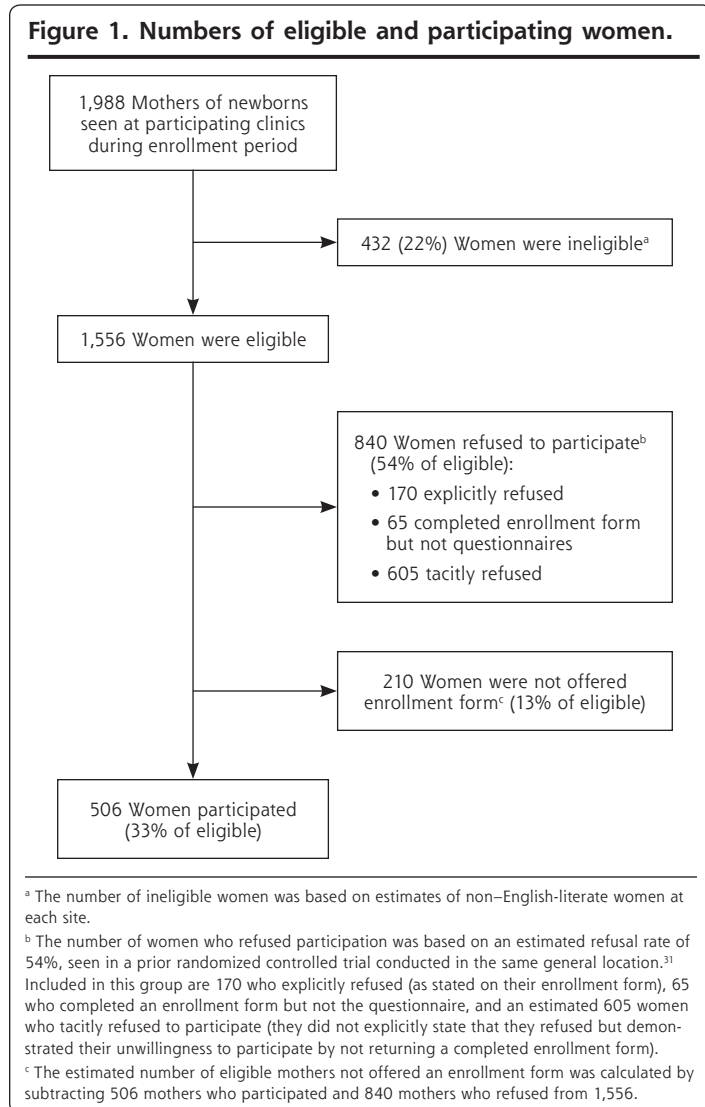
Participants' demographic characteristics, displayed in Table 2, indicate that the sample was diverse with varied ethnicity and a broad range of income and education. Overall, 167 (33%) of participants were recruited from family medicine clinics, and 339 (67%) were recruited from pediatric clinics.

Only 34 (6.7%) of the 506 participants dropped out of the study (did not complete the final questionnaire). When women who dropped out were compared with those who completed the study, the dropouts were found to be younger and less educated, had more children, were less likely to be married, had lower family incomes, and were more likely to be depressed (Table 3).

Postpartum Depression Diagnosis

Twenty women (4.6%) had a diagnosis of major depression (ie, had a positive SCID result) at 0 to 1 month

Figure 1. Numbers of eligible and participating women.

1,988 Mothers of newborns seen at participating clinics during enrollment period

432 (22%) Women were ineligible[a]

1,556 Women were eligible

840 Women refused to participate[b] (54% of eligible):
- 170 explicitly refused
- 65 completed enrollment form but not questionnaires
- 605 tacitly refused

210 Women were not offered enrollment form[c] (13% of eligible)

506 Women participated (33% of eligible)

[a] The number of ineligible women was based on estimates of non–English-literate women at each site.

[b] The number of women who refused participation was based on an estimated refusal rate of 54%, seen in a prior randomized controlled trial conducted in the same general location.[31] Included in this group are 170 who explicitly refused (as stated on their enrollment form), 65 who completed an enrollment form but not the questionnaire, and an estimated 605 women who tacitly refused to participate (they did not explicitly state that they refused but demonstrated their unwillingness to participate by not returning a completed enrollment form).

[c] The estimated number of eligible mothers not offered an enrollment form was calculated by subtracting 506 mothers who participated and 840 mothers who refused from 1,556.

postpartum, and a total of 45 (8.9%) had a diagnosis of major depression over the entire course of the study.

Validity of the Depression Screens

Validity test results for the 2-question screen, the PHQ-2, and the PHQ-9 at the initial visit (0-1 month postpartum), using the SCID interview as the reference standard, are displayed in Table 4. The highest sensitivity (100%) was seen with the 2-question screen, and the highest specificity (94%) was seen with the PHQ-9 using complex scoring.

Validity test results over the entire study period (0 to 9 months postpartum) are displayed in Table 5 and confirmed high sensitivity of the 2-question screen and high specificity of the PHQ-9 with complex scoring. Negative predictive values were very high (97%-100%), whereas positive predictive values were modest (15%-43%), as one might expect from this population wherein only 9% of women had a formal diagnosis of major depression.

We also considered the validity of a 2-step screening procedure as a whole over the 9-month follow-up period. In this analysis, we defined a positive result as concurrent positivity on both the 2-question screen and the PHQ-9; all other combinations were considered to be negative. With this approach, the procedure had a sensitivity of 80% (36 of 45), a specificity of 85% (392 of 461), a negative predictive value of 98%, and a positive predictive value of 34%.

Feasibility of Screening at Well-Child Visits

During the course of the study, 38% of participants completed their 2- to 6-month follow-up by completing the depression screens during well-child visits, 29% did so by mailed questionnaire, and 33% did so by telephone. Women enrolled from pediatric clinics were more likely to complete their questionnaires at the clinic site than were women enrolled from family medicine clinics (46.3% vs 27.2%, $P = .000$). Rates of in-office completion of the follow-up questionnaire were higher (46%) before August 1, 2006. After this date, incorporation of the electronic medical record in the family medicine clinics made it impossible to physically prepare participants' charts with questionnaires before their visit, so in-office completion from these clinics thereafter declined.

DISCUSSION

Using the SCID interview as the reference standard, we found the 2-question screen to be highly sensitive (100%) for identifying postpartum depression, mean-

ing that it did not miss any cases of postpartum depression—a desirable characteristic of an initial screening tool. In contrast, we found the PHQ-9 to be highly specific (92%-94%) for identifying postpartum depression, indicating a low false-positive rate in the overall sample—a desirable characteristic of a diagnostic tool. These results suggest the utility of a 2-stage screening procedure for postpartum depression, whereby the

Table 2. Participants' Demographic Characteristics (N = 506)

Characteristic	Value
Race/ethnicity, n (%)	
Native American	7 (1.4)
Asian	34 (6.7)
Black or African American	89 (17.6)
Hispanic	14 (2.8)
White	339 (67.0)
Multiracial	17 (3.4)
Other	6 (1.2)
Marital status, n (%)	
Married	328 (64.8)
Unmarried, living with partner	88 (17.5)
Unmarried, not living with partner	80 (15.9)
Other	7 (1.4)
Total family annual income, n (%)	
<$20,000	133 (27.3)
$20,000-$79,999	177 (36.3)
≥$80,000	178 (36.5)
Education, n (%)	
<High school diploma	81 (16.1)
High school diploma and some post–high school	160 (31.8)
≥4-Year degree	263 (52.2)
Health insurance, n (%)	
Private	257 (55.3)
Medical assistance	142 (30.5)
Other	60 (12.9)
None	6 (1.3)
Employment: work at a job or are on leave, n (%)	322 (63.6)
Have 1 child only, n (%)	207 (41.7)
Mother's age, mean (SD), year	29.1 (6.2)

Table 3. Comparisons Between Women Who Completed the Study and Women Who Dropped Out

Characteristic	Completers (n = 469)	Dropouts (n = 34)	P Value
Mother's age, mean, years	29.3	25.6	.001
Number of children	1.9	2.5	.000
Married, n (%)	319 (68.0)	9 (26.5)	.000
Post–high school education, n (%)	345 (73.4)	11 (32.4)	.000
Family income ≥$20,000, n (%)	343 (75.4)	12 (36.4)	.000
Positive PHQ-9 result, n (%)	30 (6.4)	7 (20.6)	.008

PHQ-9 = 9-item Patient Health Questionnaire.

Note: Values were compared by t tests and χ^2 analyses.

Table 4. Validity of the Depression Screens at 0 to 1 Month Postpartum (n = 438 Women, 20 With Major Depression)

Depression Screen and Result	Major Depression[a]			Validity, % (n/n)			
	No	Yes	Total	Sensitivity	Specificity	PPV	NPV
2-Question screen (yes/no)				100 (20/20)	62 (259/418)	11 (20/179)	100 (259/259)
Negative	259	0	259				
Positive	159	20	179				
Total	418	20	438				
PHQ-2 (Likert response)[b]				75 (15/20)	88 (368/416)	24 (15/63)	99 (368/373)
Negative	368	5	373				
Positive	48	15	63				
Total	416	20	436				
PHQ-9 with simple scoring[c]				75 (15/20)	91 (380/418)	28 (15/53)	99 (380/385)
Negative	380	5	385				
Positive	38	15	53				
Total	418	20	438				
PHQ-9 with complex scoring[d]				55 (11/20)	94 (394/418)	31 (11/35)	98 (394/403)
Negative	394	9	403				
Positive	24	11	35				
Total	418	20	438				

NPV = negative predictive value; PHQ-2 = 2-item Patient Health Questionnaire; PHQ-9 = 9-item Patient Health Questionnaire; PPV = positive predictive value.

Note: Values are numbers of women unless otherwise indicated.

[a] Women were defined as having major depression if they had a positive result on the Structured Clinical Interview for the *Diagnostic and Statistical Manual of Mental Disorders - DSM-IV* (SCID).
[b] Consists of the first 2 questions of the PHQ-9.
[c] Simple scoring: result is positive if sum of numbered responses is ≥10.
[d] Complex scoring: result is positive if at least 5 symptoms are present, including symptom 1, symptom 2, or both, and each symptom present has a response score of 2 to 3, except for symptom 9, for which a response score of 1 to 3 was acceptable.

Table 5. Validity of the Depression Screens Over Entire Course of Study (N = 506 Women, 45 With Major Depression)

Depression Screen and Result	Major Depression[a]			Validity, % (n/n)			
	No	Yes	Total	Sensitivity	Specificity	PPV	NPV
2-Question screen (yes/no)				100 (45/45)	44 (203/461)	15 (45/303)	100 (203/203)
Negative	203	0	203				
Positive	258	45	303				
Total	461	45	506				
PHQ-2 (Likert response)[b]				84 (38/45)	79 (363/461)	28 (38/136)	98 (363/370)
Negative	363	7	370				
Positive	98	38	136				
Total	461	45	506				
PHQ-9 with simple scoring[c]				82 (37/45)	84 (386/461)	33 (37/112)	98 (386/394)
Negative	386	8	394				
Positive	75	37	112				
Total	461	45	506				
PHQ-9 with complex scoring[d]				67 (30/45)	92 (422/461)	43 (30/69)	97 (422/437)
Negative	422	15	437				
Positive	39	30	69				
Total	461	45	506				

NPV = negative predictive value; PHQ-2 = 2-item Patient Health Questionnaire; PHQ-9 = 9-item Patient Health Questionnaire; PPV = positive predictive value.

Note: Values are numbers of women unless otherwise indicated.

[a] Women were defined as having major depression if they had a positive result on the Structured Clinical Interview for *DSM-IV* (SCID).
[b] Consists of the first 2 questions of the PHQ-9.
[c] Simple scoring: result is positive if the sum of the numbered responses is ≥10.
[d] Complex scoring: result is positive if at least 5 symptoms are present, including symptom 1, symptom 2, or both, and each symptom present has a response score of 2 to 3, except for symptom 9, for which a response score of 1 to 3 was acceptable.

2-question screen is used as the initial screening test and the PHQ-9 is used as a confirmatory test. Busy primary care professionals would undoubtedly find these 2 screens to be easier to administer than the formal, lengthier SCID interview, which requires additional specialized training.

An important benefit of this 2-stage approach is that many false-positive 2-question screen results (n = 258/506), likely more problematic in the early postpartum period when "the blues" are common, could be sorted out with the PHQ-9. False-positives (69/461 = 15% with the 2-step procedure) could be further identified during the initial visit for depression, and women with mild symptoms might initially be observed rather than treated. Clinicians also need to beware of false-negatives (9/45 = 20% with the 2-step procedure) and might advise women whose findings are positive by the 2-question screen and negative by the PHQ-9 that they need to contact their clinician if depressive symptoms persist and interfere with their function. Perhaps of even greater concern, though, is the risk of missing postpartum depression in women who do not bring their child in for well-child visits, as prior research has shown a relationship between maternal depressive symptoms and noncompliance with well-child visits.[32]

Although only 38% of the 2- to 6-month follow-up questionnaires were completed during well-child visits, the percentage was higher (46%) for questionnaires completed before August 1, 2006, when family medicine patient records were still on paper charts instead of electronic ones. Conversion to an electronic record made it impossible to physically prepare charts with screening forms before the visit. Our previous recruitment procedure was thus incompatible with the new electronic medical record, resulting in a decreased rate of completion of questionnaires in the office. Although we found the change to electronic records to be a deterrent to paper-and-pencil screening in our study wherein not all mothers were participants, routine screening via the electronic record would likely prove to be advantageous for a mass screening program that included all mothers.

The rate of completion of questionnaires in the clinic was greater in pediatric than in family medicine clinics (46% vs 27%). This difference is likely explained by the fact that the receptionists in the family medicine clinics appeared to have a larger number and range of responsibilities (eg, patients of all ages, with a greater variety of forms and needs), and usually appeared to be busier and proportionately less well staffed. These observed specialty differences would also help to explain why our in-clinic questionnaire completion rate was lower than the 55% to 74% rate of depression screen completion in 2 pediatric office-based stud-

ies.[10,11] Future studies with family medicine clinics may need to explore incentives for both clinic personnel and potential participants to increase response rates. Also, it is possible that a mass screening program would be more readily incorporated into busy family medicine clinics, as busy receptionists would not be required to differentiate participants from nonparticipants.

Although relatively few women dropped out of the study, the comparison between dropouts and completers revealed a significantly higher rate of depressive symptoms (positive PHQ-9 results) among the dropouts. If this finding is validated in future studies, subsequent investigators may want to consider participant retention strategies tailored to the demographics of the dropouts (eg, being single).

Strengths of this study include the validation of 2 depression screens—the 2-question screen and the PHQ-9—in a postpartum sample, the longitudinal study design, use of the SCID interview as a reference standard for depression diagnosis, inclusion of both family medicine and pediatric clinics, and sample diversity. The study also has some weaknesses. For example, the modest initial participation rate (33%) is likely due to a combination of factors, such as mothers' lack of time or attention, their reluctance to participate in a treatment trial, and busy receptionists' not offering women enrollment forms and questionnaires. Also, the study did not compare the validity of the 2-question screen or PHQ-9 with that of other postpartum depression screens, such as the Edinburgh Postnatal Depression Scale.

This study's findings contribute important information on scientifically based methods for identifying postpartum depression—the most common serious obstetric complication. If confirmed by additional studies, our validity test results suggest that the highly sensitive 2-question screen and highly specific PHQ-9 perform well together in a 2-stage screening procedure, whereby women initially complete the 2-question screen, and those whose results are positive complete the PHQ-9. Women who have positive findings on the PHQ-9 would be advised to see their clinician for further evaluation and treatment.

Columbia, Canada, and the Women's Health Research Conference, September 17, 2007, University of Minnesota, Minneapolis.

References

1. Gaynes BN, Gavin N, Meltzer-Brody S, et al. *Perinatal Depression: Prevalence, Screening Accuracy, and Screening Outcomes. Evidence Report/Technology Assessment No. 119.* Rockville, MD: Agency for Healthcare Research and Quality, February, 2005. AHRQ Publication No. 05-E006-2.

2. Beck CT. Predictors of postpartum depression. *Nurs Res.* 2001;50(5):275-285.

3. Stewart WF, Ricci JA, Chee E, Hahn SR, Morganstein D. Cost of lost productive work time among US workers with depression. *JAMA.* 2003;289(23):3135-3144.

4. Righetti-Veltema M, Bousquet A, Manzano J. Impact of postpartum depressive symptoms on mother and her 18-month-old infant. *Eur Child Adolesc Psychiatry.* 2003;12(2):75-83.

5. Grace SL, Evindar A, Stewart DE. The effect of postpartum depression on child cognitive development and behavior: a review and critical analysis of the literature. *Arch Womens Ment Health.* 2003;6(4):263-274.

6. U.S. Preventive Services Task Force (USPSTF). Screening for depression: recommendations and rationale. *Ann Intern Med.* 2002;136(10):760-764.

7. Georgiopoulos AM, Bryan TL, Wollan P, Yawn BP. Routine screening for postpartum depression. *J Fam Pract.* 2001;50(5):117-122.

8. Hearn G, Iliff A, Jones I, et al. Postnatal depression in the community. *Br J Gen Pract.* 1998;48(428):1064-1066.

9. Chaudron LH, Szilagyi PG, Kitzman HJ, Wadkins HIM, Conwell Y. Detection of postpartum depressive symptoms by screening at well-child visits. *Pediatrics.* 2004;113(3 Pt 1):551-558.

10. Freeman MP, Wright R, Watchman M, et al. Postpartum depression assessments at well-baby visits: screening feasibility, prevalence, and risk factors. *J Womens Health (Larchmt).* 2005;14(10):929-935.

11. Olson AL, Dietrich AJ, Prazar G, Hurley J. Brief maternal depression screening at well-child visits. *Pediatrics.* 2006;118(1):207-216.

12. Buist A, Condon J, Brooks J, et al. Acceptability of routine screening for perinatal depression. *J Affect Disord.* 2006;93(1-3):233-237.

13. Gemmill AW, Leigh B, Ericksen J, Milgrom J. A survey of the clinical acceptability of screening for postnatal depression in depressed and non-depressed women. *BMC Public Health.* 2006;6:211.

14. Kelly BD. St John's wort for depression: what's the evidence? *Hosp Med.* 2001;62(5):274-276.

15. Seehusen DA, Baldwin LM, Runkle GP, Clark G. Are family physicians appropriately screening for postpartum depression? *J Am Board Fam Pract.* 2005;18(2):104-112.

16. Spitzer RL, Kroenke K, Williams JBW, 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(18):1737-1744.

17. Evins GG, Theofrasstous JP, Galvin SL. Postpartum depression: a comparison of screening and routine clinical evaluation. *Am J Obstet Gynecol.* 2000;182(5):1080-1082.

18. Heneghan AM, Silver EJ, Bauman LJ, Stein REK. Do pediatricians recognize mothers with depressive symptoms? *Pediatrics.* 2000;106(6):1367-1373.

19. Olson AL, Kemper KJ, Kelleher KJ, Hammond CS, Zuckerman BS, Dietrich AJ. Primary care pediatricians' roles and perceived responsibilities in the identification and management of maternal depression. *Pediatrics.* 2002;110(6):1169-1176.

20. Wiley CC, Burke GS, Gill PA, Law NE. Pediatricians' views of postpartum depression: a self-administered survey. *Arch Womens Ment Health.* 2004;7(4):231-236.

21. Boyd RC, Le HN, Somberg R. Review of screening instruments for postpartum depression. *Arch Womens Ment Health.* 2005;8(3):141-153.

22. Austin MP, Lumley J. Antenatal screening for postnatal depression: a systematic review. *Acta Psychiatr Scand.* 2003;107(1):10-17.

23. Nease DE, Maloin JM. Depression screening: a practical strategy. *J Fam Pract.* 2003;52(2):118-124.

24. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med.* 2001;16(9):606-613.

25. Hanusa B, Scholle SH, Haskett RF, Spadaro K, Wisner KL. Screening for depression in the postpartum period: a comparison of three instruments. *J Womens Health (Larchmt).* 2008;17(4):585-596.

26. Whooley MA, Avins AL, Miranda J, Browner WS. Case-finding instruments for depression: two questions are as good as many. *J Gen Intern Med.* 1997;12(7):439-445.

27. Arroll B, Goodyear-Smith F, Kerse N, Fishman T, Gunn J. Effect of the addition of a "help" question to two screening questions on specificity for diagnosis of depression in general practice: diagnostic validity study. *BMJ.* 2005;331(7521):884.

28. Kroenke K, Spitzer RL, Williams JBW. The Patient Health Questionnaire-2. Validity of a two-item depression screener. *Med Care.* 2003;41(11):1284-1292.

29. First MG, Spitzer RL, Gibbon M, Williams JB. *Structured Clinical Interview for DSM-IV Axis I Disorder—Clinical Version, Administration Booklet.* New York, NY: Biometrics Research Department; 1997.

30. American Psychiatric Association (APA). *Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, Text Revision (DSM-IV-TR).* Washington, DC: APA; 2000.

31. Gjerdingen DK, Center B. A randomized controlled trial testing the impact of a support/work-planning intervention on first-time parents' health, partner relationship, and work responsibilities. *Behav Med.* 2002;28(3):84-91.

32. Jhanjee I, Saxeena D, Arora J, Gjerdingen DK. Parents' health and demographic characteristics predict noncompliance with well-child visits. *J Am Board Fam Med.* 2004;17(5):324-331.