Evidence Report/Technology Assessment
Number 119

Perinatal Depression: Prevalence, Screening Accuracy, and Screening Outcomes

Prepared for:
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
540 Gaither Road
Rockville, MD 20850
www.ahrq.gov

Contract No. 290-02-0016

Prepared by:
RTI-University of North Carolina Evidence-based Practice Center
Research Triangle Park, North Carolina

Investigators
Bradley N. Gaynes, MD, MPH
Norma Gavin, PhD
Samantha Meltzer-Brody, MD, MPH
Kathleen N. Lohr, PhD
Tammeka Swinson, BA
Gerald Gartlehner, MD, MPH
Seth Brody, MD, MPH
William C. Miller, MD, PhD

AHRQ Publication No. 05-E006-2
February 2005

This report may be used, in whole or in part, as the basis for development of clinical practice guidelines and other quality enhancement tools, or a basis for reimbursement and coverage policies. AHRQ or U.S. Department of Health and Human Services endorsement of such derivative products may not be stated or implied.

AHRQ is the lead Federal agency charged with supporting research designed to improve the quality of health care, reduce its cost, address patient safety and medical errors, and broaden access to essential services. AHRQ sponsors and conducts research that provides evidence-based information on health care outcomes; quality; and cost, use, and access. The information helps health care decisionmakers—patients and clinicians, health system leaders, and policymakers—make more informed decisions and improve the quality of health care services.

Suggested Citation:
Gaynes BN, Gavin N, Meltzer-Brody S, Lohr KN, Swinson T, Gartlehner G, Brody S, Miller WC. Perinatal Depression: Prevalence, Screening Accuracy, and Screening Outcomes. Evidence Report/Technology Assessment No. 119. (Prepared by the RTI-University of North Carolina Evidence-based Practice Center, under Contract No. 290-02-0016.) AHRQ Publication No. 05-E006-2. Rockville, MD: Agency for Healthcare Research and Quality. February 2005.

Preface

The Agency for Healthcare Research and Quality (AHRQ), through its Evidence-Based Practice Centers (EPCs), sponsors the development of evidence reports and technology assessments to assist public- and private-sector organizations in their efforts to improve the quality of health care in the United States. The reports and assessments provide organizations with comprehensive, science-based information on common, costly medical conditions and new health care technologies. The EPCs systematically review the relevant scientific literature on topics assigned to them by AHRQ and conduct additional analyses when appropriate prior to developing their reports and assessments.

This report on perinatal depression was requested and funded by the Safe Motherhood Group (SMG). The SMG consists of representatives from several agencies within the U.S. Department of Health and Human Services (DHHS): the DHHS Office on Women's Health; Centers for Disease Control and Prevention; Health Resources and Services Administration; Maternal and Child Health Bureau; National Institutes of Health, National Institute of Mental Health, National Institute of Child Health and Human Development, National Institute on Drug Abuse; Food and Drug Administration; Substance Abuse and Mental Health Services Administration; and Agency for Healthcare Research and Quality.

To bring the broadest range of experts into the development of evidence reports and health technology assessments, AHRQ encourages the EPCs to form partnerships and enter into collaborations with other medical and research organizations. The EPCs work with these partner organizations to ensure that the evidence reports and technology assessments they produce will become building blocks for health care quality improvement projects throughout the Nation. The reports undergo peer review prior to their release.

AHRQ expects that the EPC evidence reports and technology assessments will inform individual health plans, providers, and purchasers as well as the health care system as a whole by providing important information to help improve health care quality.

We welcome comments on this evidence report. They may be sent by mail to the Task Order Officer named below at: Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850, or by e-mail to **epc@ahrq.gov**.

Carolyn M. Clancy, M.D.
Director
Agency for Healthcare Research and Quality

Kenneth S. Fink, M.D., M.G.A., M.P.H.
Director, EPC Program
Agency for Healthcare Research and Quality

Jean Slutsky, P.A., M.S.P.H.
Director, Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Marian D. James, M.A., Ph.D.
EPC Program Task Order Officer
Agency for Healthcare Research and Quality

Structured Abstract

Context. Depression during pregnancy or the first year postpartum is impressively common and can have devastating consequences for the woman, her children, and other family members.

Objectives. We systematically review the evidence on (1) the prevalence and incidence of perinatal depression, (2) the accuracy of different screening instruments, and (3) the effectiveness of interventions for women screened as high risk for perinatal depression

Data Sources. MEDLINE, CINAHL, PsycINFO, Sociofile, and the Cochrane Library (1980 through March 2004); bibliographic hand searches; and experts.

Study Selection. The English-language studies assessed women for major depression alone or for major or minor depression. Studies of the prevalence and incidence of depression and the accuracy of screening tools had to include diagnostic confirmation by a reference standard. Studies involving interventions required a comparison group. Two reviewers independently evaluated each abstract to determine inclusion by consensus.

Data Extraction. A primary reviewer abstracted data on key variables from the articles directly into detailed evidence tables; a second reviewer confirmed accuracy.

Data Synthesis. We conducted a meta-analysis of the prevalence and incidence estimates to compute combined estimates for particular periods and points in time. We also conducted meta-analyses of the sensitivity and specificity of different screening instruments. For screening outcome studies, we were only able to synthesize qualitatively.

Results. We identified 30 studies of prevalence. For major depression alone, point prevalence estimates ranged from 3.1 percent to 4.9 percent at different times during pregnancy and 1.0 percent to 5.9 percent at different times during the first postpartum year. For major and minor depression, estimates of the point prevalence ranged from 8.5 percent to 11.0 percent during pregnancy and 6.5 percent to 12.9 percent during the first year postpartum. However, these prevalence estimates were not significantly different from those of similarly aged nonchildbearing women. Data on incidence were more limited.

We identified 10 studies of screening accuracy. One small study reported on accuracy during pregnancy. For postpartum depression, screeners appeared feasible, but the small number of depressed patients involved precluded identifying an optimal screener or threshold for screening. Screening instruments studied are generally good at identifying major depression alone, with accuracy consistent with reports from primary care settings, but they performed poorer for the major or minor depression category.

We found no studies directly testing whether screening improved outcomes. However, we identified 15 studies that used some sort of screening to identify women at risk of depression and for whom a subsequent intervention was provided. The results of four small studies of various psychosocial interventions during pregnancy did not demonstrate consistently superior outcomes. Results were also mixed for postpartum interventions. Six of nine studies of various psychosocial

interventions reported significant improvement in depression for the experimental group. Two studies with pharmacologic interventions provided conflicting results.

Conclusions. Although limited, the available research suggests that depression is one of the most common perinatal complications and that fairly accurate and feasible screening measures are available. Studies with larger sample sizes and a greater racial and ethnic mix are needed. Researchers also need to determine whether screening itself leads to better access to proven treatment and improved outcome relative to usual care.

Contents

Evidence Report

Chapter 1. Introduction 3

Chapter 2. Methods 7
Role of the Technical Expert Advisory Group 7
Literature Search Strategy 8
Inclusion and Exclusion Criteria 8
Literature Search and Retrieval Process 9
Data Collection and Assessment 12
Meta-analysis 12
Quality of Individual Articles 12
Strength of Overall Evidence 13
External Peer Review 13

Chapter 3. Prevalence and Incidence of Perinatal Depression 15
Introduction 15
Methods 16
Evaluation of the Quality and Strength of the Evidence 17
Best Estimates of Prevalence and Incidence 17
Analysis of Confounders 18
Comparison with Other Women 19
Results 19
Study Characteristics 19
Results from Prospective Studies 30
Analysis of Confounders 39
Comparison with Other Women 41
Results from Retrospective Studies 41
Discussion 46

Chapter 4. Screening Accuracy 49
Introduction 49
Methods 49
Inclusion and Exclusion Criteria 50
Data Analysis 50
Evaluation of Quality and Strength of Evidence 50
Results 51
Study Characteristics 51
Screening Instruments Used 53
Reference Standards Used 54
Classifications of Depressive Illness 54
Quality Rating 55
Prenatal Screening Results 55
Postpartum Screening Results 58

Summary of Results of Screening Instrument Review 66
Interpretation of Results 68
Discussion 69
Conclusions 69
Future Research 69
Use in Clinical Settings 70

Chapter 5. Impact of Depression Screening and Interventions on Patient Outcomes 71
Introduction 71
Methods 72
Results 73
Prenatal Studies 73
Postpartum Studies 79
Discussion 84
Conclusions 84
Study Limitations 84
Future Research 85

Chapter 6. Conclusions and Recommendations 87
Conclusions 87
Key Question 1: Prevalence and Incidence of Perinatal Depression 87
Key Question 2: Accuracy of Perinatal Depression Screening Tools 88
Key Question 3: Screening and Treatment Outcomes 89
Recommendations for Future Research 90

References and Included Studies 93

Glossary 99

Tables

Table 1. Key questions for the evidence report on perinatal depression 4
Table 2. Inclusion/exclusion criteria by key question 8
Table 3. Literature search strategies and yield 10
Table 4. Major characteristics of studies of prevalence and incidence of perinatal depression 21
Table 5. Quality rating of studies of the prevalence and incidence of perinatal depression 28
Table 6. Original estimates of prevalence and incidence of major and minor depression 32
Table 7. Original estimates of prevalence and incidence of major depression 33
Table 8. Best estimates of prevalence and incidence of major and minor depression 35
Table 9. Best estimates of prevalence and incidence of major depression 36
Table 10. Meta–regression results for log odds of a diagnosis of major and minor depression 42
Table 11. Meta–regression results for log odds of a diagnosis of major depression 44

Table 12. Odds ratios for studies with comparison groups of women during nonchildbearing periods 46
Table 13. Major characteristics of studies of screening for perinatal depression 52
Table 14. Key features of screening instruments for perinatal depression 53
Table 15. Quality rating of studies of screening for perinatal depression 56
Table 16. Sensitivity and specificity of perinatal depression screens 62
Table 17. Major characteristics of studies of screening with interventions for prenatal or postpartum depression 75
Table 18. Quality rating of studies of screening with interventions for prenatal or postpartum depression 76
Table 19. Major outcomes of studies of screening and interventions for perinatal depression 77

Figures

Figure 1. Causal pathway for the screening and treatment of perinatal depression 5
Figure 2. Perinatal depression article disposition 11
Figure 3. Estimates of point prevalence of major and minor depression by time of assessment 25
Figure 4. Estimates of point prevalence of major depression by time of assessment 25
Figure 5. Estimates of period prevalence of major and minor depression by time period of assessment 26
Figure 6. Estimates of period prevalence of major depression by time period of assessment 26
Figure 7. Estimates of incidence of major and minor depression by time period of assessment 27
Figure 8. Estimates of incidence of major depression by time period of assessment 27
Figure 9. Best estimates of point prevalence of major and minor depression 34
Figure 10. Best estimates of point prevalence of major depression 37
Figure 11. Best estimates of period prevalence of depression 38
Figure 12. Best estimates of incidence of major and minor depression 39
Figure 13. Cumulative meta-analysis for point prevalence of depression at 2 months postpartum 40
Figure 14a. Sensitivity of screening by Edinburgh Postnatal Depression Scale: prenatal period, major depression alone 57
Figure 14b. Specificity of Edinburgh Postnatal Depression Scale: prenatal period, major depression alone 57
Figure 15a. Sensitivity of screening by Edinburgh Postnatal Depression Scale: prenatal period, major or minor depression 58
Figure 15b. Specificity of screening by Edinburgh Postnatal Depression Scale: prenatal period, major or minor depression 58
Figure 16a. Sensitivity of screening by Edinburgh Postnatal Depression Scale: postpartum period, major depression alone 59
Figure 16b. Specificity of screening by Edinburgh Postnatal Depression Scale: postpartum period, major depression alone 60

Figure 17a. Sensitivity of screening by Edinburgh Postnatal Depression Scale: postpartum period, major or minor depression 64
Figure 17b. Specificity of screening by Edinburgh Postnatal Depression Scale: postpartum period, major or minor depression 65
Figure 18a. Sensitivity of screening by BDI: postpartum period, major depression 65
Figure 18b. Specificity of screening by BDI: postpartum period, major depression alone 66
Figure 19. Causal Pathway for Key Question 3 on Screening and Treatment Outcomes..... 71

Appendixes

Appendix A. Exact Search Strings A-1
Appendix B. Quality Rating Forms..... B-1
Appendix C. Evidence Tables C-1
Appendix D. Excluded Articles..... D-1
Appendix E. Acknowledgments.....E-1

Appendixes and Evidence Tables are provided electronically at http://www.ahrq.gov/clinic/epcindex.htm

Agency for Healthcare Research and Quality

Evidence Report/Technology Assessment

Number 119

Perinatal Depression: Prevalence, Screening Accuracy, and Screening Outcomes

Summary

Authors: Gaynes BN, Gavin N, Meltzer-Brody S, Lohr KN, Swinson T, Gartlehner G, Brody S, Miller WC

Introduction

Depression is the leading cause of disease-related disability among women.[1] In particular, women of childbearing age are at high risk for major depression.[2-4] Pregnancy and new motherhood may increase the risk of depressive episodes. Depression during the perinatal period can have devastating consequences, not only for the women experiencing it but also for the women's children and family.[5-8]

Perinatal depression encompasses major and minor depressive episodes that occur either during pregnancy or within the first 12 months following delivery. When referring to depression in this population, researchers and clinicians frequently have not been clear about whether they are referring to major depression alone or to both major and minor depression. Major depression is a distinct clinical syndrome for which treatment is clearly indicated,[9] whereas the definition and management of minor depression are less clear. In this report, we refer to major depression alone by identifying it discretely as major depression. Minor depression is an impairing, yet less severe, constellation of depressive symptoms[10] for which controlled trials have not consistently indicated whether or not particular interventions are more effective than placebo.[11,12] In this report, we refer to this grouping as major or minor depression or by the more general terms "depression" or "depressive illness." Perinatal depression, whether one is referring to major depression alone or to either major or minor depression, often goes unrecognized because many of the discomforts of pregnancy and the puerperium are similar to symptoms of depression.[13,14]

Another mental disorder that can occur in the perinatal period is postpartum psychosis. Unlike postpartum depression, postpartum psychosis is a relatively rare event with a range of estimated incidence of 1.1 to 4.0 cases per 1,000 deliveries.[15] The onset of postpartum psychosis is usually acute, within the first 2 weeks of delivery, and appears to be more common in women with a strong family history of bipolar or schizoaffective disorder.[16] Postpartum psychosis is an important disorder in its own right, but it is not addressed specifically in this report.

The precise level of the prevalence and incidence of perinatal depression is uncertain. Published estimates of the rate of major and minor depression in the postpartum period range widely—from 5 percent to more than 25 percent of new mothers, depending on the assessment method, the timing of the assessment, and population characteristics.[17-19]

In addition, although many screening instruments have been developed or modified to detect major and minor depression in pregnant and newly delivered women, the evidence on their screening accuracy relative to a reference standard has yet to be systematically reviewed and assessed.[20] Evidence on the effectiveness of screening all pregnant women and providing a preventive intervention to those scoring at high risk has not been systematically investigated and evaluated either.[20]

To address these gaps, the Agency for Healthcare Research and Quality (AHRQ), in collaboration with the Safe Motherhood Group (SMG), commissioned this evidence report from the RTI International-University of North Carolina's (RTI-UNC's) Evidence-based Practice Center (EPC) for a systematic review of the evidence on three questions related to perinatal depression. These questions address the prevalence and incidence of perinatal depression, the accuracy of screening instruments for

Agency for Healthcare Research and Quality
Advancing Excellence in Health Care • www.ahrq.gov
Evidence-Based Practice

perinatal depression, and the effectiveness of interventions for women screened as high risk for developing perinatal depression. The three key questions (KQs) are:

1. What are the incidence and prevalence of depression (major and minor) during pregnancy and during the postpartum period? Are they increased during pregnancy and the postpartum period compared to nonchildbearing periods?
2. What is the accuracy of different screening tools for detecting depression during pregnancy and the postpartum period?
3. Does prenatal or early postnatal screening for depressive symptoms with subsequent intervention lead to improved outcomes?

Methods

In conducting this systematic review, we followed standardized procedures developed by AHRQ in collaboration with all its EPCs for such reviews. Throughout the project we enlisted the assistance of a Technical Expert Advisory Group (TEAG) to react to work in progress and advise us on substantive issues and overlooked areas of research. The TEAG included four individuals who, collectively, have expertise in obstetrics, psychiatry, psychology, and research methods, along with clinical and research experience in perinatal depression.

Inclusion and Exclusion Criteria

We made the inclusion and exclusion criteria fairly restrictive to ensure that our conclusions were based on the highest quality data available with the lowest risk of bias. Some criteria were common across the three key questions; others were specific to the question.

For all key questions, studies had to report on original data, be in English, and be published from January 1980 through March 2004. The study also had to be from a developed country to increase the likelihood of its being generalizable to the U.S. population. We excluded studies of women with bipolar disorder, primary psychotic disorders, or maternity blues (a mild mood disturbance experienced by approximately half of childbearing women within 3 to 6 days after delivery that resolves within a few hours to a few days) in which the outcomes of interest were not distinguishable from those for women with major or minor depression. For KQs 2 and 3, we excluded studies that enrolled women with known depressive disorders at the outset because screening would not be necessary for a patient already known to have a current depressive episode.

In addition, studies for all key questions had to assess women for depression during pregnancy or in the first year postpartum. Diagnostic confirmation, by means of a clinical assessment or structured clinical interview, was required for KQs 1 and 2. For KQ 1, we excluded studies of the prevalence and incidence of perinatal depression that relied solely on self-report screens to identify depression. In KQ 2, study investigators used the clinical assessment or structured clinical interview to assess the properties of the screening instrument.

In KQ 3, we required that patients had to have been screened, whether by formal instrument or by another type of screen that identified women as being at risk of having a depressive illness (e.g., prior history of postpartum depression). As the screening process was the focus of interest here, for KQ 3 we excluded studies in which a reference standard confirmation of depression was required for enrollment.

For the first part of KQ 1, we included both prospective and retrospective studies of the prevalence and incidence of perinatal depression; for the second part, we included clinical trials and case-control studies comparing the incidence or prevalence of depression among pregnant women and newly delivered mothers to prevalence among women of similar age during nonchildbearing periods of their lives. We included only prospective studies in those reviewed for KQs 2 and 3 and only controlled trials to provide evidence of the effectiveness of interventions among women at high risk of perinatal depression for KQ 3.

Literature Search and Retrieval Process

We used three strategies to identify studies providing evidence related to the key questions: systematic searches of electronic databases using both a list of Medical Subject Heading (MeSH®) search terms and author names, hand searches of reference lists of included articles, and consultation with the TEAG. We searched standard electronic databases, including MEDLINE®, Cumulative Index to Nursing & Allied Health Literature (CINAHL), PsycINFO, Sociofile, and the Cochrane Library. We found a total of 837 citations in the electronic searches and picked up an additional 9 citations through the hand searches and discussion with the TEAG, for a total of 846 citations.

Three senior reviewers with clinical expertise in perinatal depression reviewed the abstracts of articles identified during the literature search. Two clinicians evaluated each abstract against the inclusion criteria and resolved any differences in inclusion by consensus. In several instances, the abstracts did not provide enough information to make an inclusion decision; we pulled full articles to review for those studies. Of the 846 articles identified, 729 did not meet the inclusion criteria for any of the key questions and were therefore excluded, 8 studies were pulled for background only, and the remaining 109 articles were pulled for a full review.

Among the studies pulled for full review, 50 did not meet our inclusion/exclusion criteria for any of the three key questions. The most common reason for exclusion was the absence of a gold standard (i.e., either a clinical assessment or structured clinical interview) for assessing depression, which eliminated 26 studies. We excluded 10 of the studies pulled for the evaluation of the properties of screening instruments because they did not report sensitivity and specificity or data that we could use to compute those measures. Other reasons

for exclusion were restriction of the study sample to specific population subgroups (e.g., teenagers, patients of psychiatric hospitals), depression assessed after the first year postpartum, no depression outcome measured, and a retrospective study design.

The remaining 59 studies were included in the review; some met the inclusion criteria for more than one key question. Thirty studies were abstracted for KQ 1; 23, for KQ 2; and 15, for KQ 3.

Data Abstraction and Assessment

The data collection process involved abstracting relevant information from the eligible articles and generating evidence tables that present the key details of the study design and the major findings from the articles. Each article was read and abstracted by a trained member of the study team; a second member checked the table entries for accuracy against the original article.

We also rated the quality of the studies. We developed a quality rating form for the screening accuracy (KQ 2) articles from criteria identified by the Cochrane Methods Working Group on Systematic Review of Screening and Diagnostic Tests.[21] For studies addressing KQ 1 and KQ 3, we modified the quality rating forms developed by Downs and Black for randomized controlled trials (RCTs) and observational studies.[22] The quality rating forms dealt with the reporting completeness and clarity, external validity, internal validity, and power or precision of each study. The senior abstractor completed the quality rating form for each article; another project team member then reviewed the completed form for accuracy and completeness.

In addition to the individual studies, we also rated the strength of the collective evidence on each key question. We applied four criteria: (1) the number of studies, (2) the aggregate sample sizes over the studies, (3) the quality of the individual studies, and (4) the representativeness of the study populations included in the studies.

Meta-Analysis

We conducted a meta-analysis of the different prevalence and incidence estimates from studies abstracted for KQ 1 to compute combined prevalence and incidence estimates for particular periods and points in time. We also conducted meta-analyses of the different estimates of the receiver operating characteristic (ROC) curves for screening instruments evaluated for KQ 2. Because of the diversity of screening instruments and prevention interventions in the studies found for KQ 3, we did not conduct a meta-analysis for this key question.

Key Question 1

For KQ 1, we combined all estimates with the same diagnosis, estimate type, and time period using the *meta* command in Stata. This procedure uses the inverse-variance weighting method to calculate random effects summary estimates. It also produces Q tests of the homogeneity of the estimates, forest plots of the individual study estimates, and combined estimates and their confidence intervals. To satisfy the normalcy assumptions of these methods, we first transformed the prevalence estimates into log odds estimates.

We reviewed the forest plots of the studies in each summary estimate to determine whether we could identify the source of any heterogeneity between studies. We then reran the meta-analyses excluding studies that were obvious outliers and for which we could identify the source of the bias. The new summary estimates are considered our best estimates of the prevalence and incidence of perinatal depression for the general female population in the United States and other developed countries.

To further analyze associations between the prevalence of depression and study characteristics, we conducted cumulative meta-analysis and a series of meta-regressions on the point prevalence estimates for major and minor depression together and major depression alone.

Key Question 2

For KQ 2, our main outcomes of interest were sensitivity and specificity of the screening approaches or instruments as described in the selected articles. Sensitivity refers to the proportion of patients with a disease who test positive ("true positives"); specificity refers to the proportion of patients without a disease who test negative ("true negatives").

For each reported instrument and associated cutoff, we calculated sensitivity and specificity from the published data and constructed 95-percent confidence intervals (CIs) using exact methods. For instruments with three or more estimates at a particular cutoff, we created plots of the sensitivity or specificity with associated 95-percent CIs to provide a graphic description of the degree of consistency of results. In addition, where possible, we estimated pooled sensitivity and specificity values using meta-analytic methods for fixed effects. We evaluated heterogeneity using the Q statistic test for homogeneity. In several circumstances, pooled estimates were not possible to calculate because of perfect estimates of sensitivity (i.e., 100 percent) with associated variance estimates equal to zero.

Peer Review

As is customary for all evidence reports and systematic reviews done for AHRQ, the RTI-UNC EPC requested review of the draft report from a wide array of outside experts in the field and from relevant professional societies and public organizations. AHRQ also requested review from its own staff and appropriate Federal agencies. We revised this final report on the basis of that feedback.

Results

Prevalence and Incidence of Depression

We found 30 studies providing estimates of the prevalence of perinatal depression.[14,19,23-49] Some rates were reported as point prevalences, the percentage of the population with depression at a given point in time (e.g., at 24 weeks gestational age or 9

weeks postpartum); others were reported as period prevalences, the percentage of the population with depression over a period of time (e.g., during pregnancy or from delivery to the end of the first 3 months postpartum). Only 13 studies provided estimates of the incidence of the disorder (i.e., the percentage of the population with depressive episodes that begin within a given period of time).

The studies were generally of moderate size—too small for reliable subgroup analyses. Furthermore, the study populations were typically restricted to a local community or geographic region served by one provider or a small number of providers of obstetrical services and were not representative of the racial and ethnic mix of the countries in which the studies were conducted. Other confounders included the risk status of women at study entry, their socioeconomic status, the interview methods, and the diagnostic criteria used to identify cases.

Our final combined estimates of prevalence and incidence were somewhat lower than those found in prior systematic reviews for three reasons. First, we excluded studies that assessed depression based on self-report screens alone, which have been found to overestimate prevalence. Second, we separated out estimates of major and minor depression from estimates of major depression alone. Third, we included more recent studies that use more precise criteria to identify major depression.

For major depression alone, our final combined point prevalence estimates ranged from 3.1 percent to 4.9 percent at different times during pregnancy and from 1.0 percent to 5.9 percent at different times during the first postpartum year. For major and minor depression, our final combined estimates of point prevalence ranged from 8.5 percent to 11.0 percent at different times during pregnancy and from 6.5 percent to 12.9 percent at different times during the first year postpartum. This nearly twofold higher rate suggests that approximately half of the women experience a major depressive episode and half a minor depressive episode at any given time. Confidence intervals surrounding all of these estimates remain wide, suggesting that a fair amount of uncertainty remains in the combined estimates.

Fewer estimates were available for the incidence of depression. These limited data suggest that as many as 14.5 percent of pregnant women have a new episode of major or minor depression during pregnancy and 14.5 percent have a new episode during the first 3 months postpartum. Considering only major depression, 7.5 percent may have a new episode during pregnancy, with 6.5 percent having a new episode in the first 3 months postpartum.

Prevalence estimates for perinatal depression were not significantly different from the prevalence of depression among women of similar age who were not pregnant and had not recently given birth.[45-47] However, Cox et al. found that, in the first 5 weeks postpartum, the odds of a new episode of major depression are three times that of a comparison group of females.[46] Thus, data from this one study suggest that, after an event as psychologically and physiologically stressful as labor and delivery, the likelihood of a new episode of depression may be substantially higher than in a likely less stressed group of women of similar age.

Accuracy of Screening Tools

For our analysis of the accuracy of screening tools (KQ 2), we identified 10 studies reporting test characteristics for English-language screeners.[27,40,42,50-56] In general, studies were of fair to good quality, although external validity was only poor to fair. Specifically, the study populations were nearly entirely white, so the accuracy of these screeners in other perinatal populations is not clear. A major limitation in the available evidence is the very small number of depressed patients involved, a fact that results in substantial imprecision in the point estimate of sensitivity and prevented us from reasonably determining an ideal cutoff point.

For depression during pregnancy, we found only one study reporting on screening accuracy in a population, with 6 patients with major depression and 14 patients with either major or minor depression. For major depression, sensitivities for the Edinburgh Postnatal Depression Scale (EPDS) at all thresholds evaluated (12, 13, 14, 15) were 1.0, underscoring the markedly small number of depressed patients involved; specificities ranged from 0.79 (at EPDS $\geq$12) to 0.96 (at EPDS $\geq$15). For major or minor depression, sensitivity was much poorer (0.57 to 0.71), and specificity remained fairly high (0.72 to 0.95).

For postpartum depression, also, the small number of depressed patients involved in the studies precluded identifying an optimal screener or an optimal threshold for screening. Our ability to combine the results of different studies in a meta-analysis was limited by the use of multiple cutoffs and other differences in the studies that would have made the pooled estimate hard to interpret. Where we were able to combine the results through meta-analysis, the pooled analysis did not add to what one could conclude from individual studies.

For women with major depression alone, specificity for all screeners (the Beck Depression Inventory [BDI], the Postpartum Depression Screening Scale [PDSS], and the EPDS) was relatively high and overlapped substantially. This finding suggests that a positive screen was accurate in ruling major depression in; that is, the risk that a screen with one of these instruments would be falsely positive was low. By contrast, sensitivities varied much more. The EPDS and the PDSS appeared to be more sensitive (with estimates ranging from 0.75 to 1.0 at different thresholds) than the BDI instruments (with estimates from 0.32 to 0.68), but the wide CIs overlapped nearly completely. Thus, we could not say with confidence that the sensitivity estimates using the different tools were different.

The point estimates are consistent with what is reported for depression screeners in primary care settings.[57] Still, the imprecision is important to clarify. If falsely missing depression (a false negative) is worse than falsely identifying it (as may be the case with this disorder), clinicians must be able to feel

confident that the screen is usually positive if the disease is there and that a negative result can help rule out the illness.

For patients with major or minor depression, results were reported for EPDS, BDI, PDSS, and the Center for Epidemiologic Studies Depression Scale (CES-D). Specificity estimates remained relatively high, but sensitivity results were much lower (ranging from 0.43 to 0.71) than for major depression alone. This means that the ability of the screening instrument to score women as positive for this condition when the disease is present was poorer than for major depression alone. Again, neither any particular cutoff nor any particular screening instrument performed differently from the others. No available comparators were found for primary care populations.

Our results suggest that various screening instruments can identify perinatal depression, most accurately major depression, but clinicians need to know more about precision. If one assumes that the risk of a false-negative depression screen is worse than the risk of a false-positive screen, perinatal depression is a condition in which sensitivity is likely to be more important than specificity. Whether as a screen for major depression alone or for major or minor depression, specificities appear high and relatively precise. By contrast, sensitivity for identifying either category is imprecise and differs by diagnostic category. For major depression alone, point estimates are equivalent to those found in primary care medical settings. For major or minor depression, however, sensitivity is quite low. At this time, these screens do not appear to be useful for identifying patients in this broader category of illness.

Screening With Subsequent Intervention

KQ 3 concerned issues of whether screening ultimately leads to improved patient outcomes. Although it is the most vital question from the public health perspective, it is the one with the most limited evidence. Indeed, the studies that we identified were not designed to test whether screening for depression (versus not screening) improved patient outcomes. Such a design would randomize patients to be screened or not to be screened and then compare subsequent outcomes. We found no studies designed in this way.

Instead, we made use of studies in which women were screened by formal depression screen or the presence of a risk factor associated with perinatal depression to identify those at risk of having a depressive illness; then, for those screening positive, the investigators compared the outcomes of women receiving a treatment intervention to those in a control group. This design tests whether, among women identified as at risk of depression by a screen, an intervention improves outcomes compared to the outcomes in a control group. This is an important intermediary step, but it does not directly test whether screening itself improves outcome compared to not screening.

For patients whose screening results identified them as at risk of perinatal depression and for whom a subsequent intervention was provided, we identified 15 studies. Four small prenatal studies involved various psychosocial interventions.[58-61] Quality was poor for three of these studies and fair for one. Overall, the effects of the interventions in these perinatal studies were not consistently superior to those in the control groups.

The 11 postpartum studies were of overall fair quality and had larger sample sizes than the prenatal trials.[62-72] Study populations still reflected only a limited racial and ethnic mix, and both external validity and the power to demonstrate statistically significant differences were generally poor. Again, screening tools and interventions varied considerably; the latter involved both psychosocial and pharmaceutical interventions.

Results were mixed. Of the nine trials that employed a psychosocial intervention, six studies[62-65,67,68] reported significant benefit for depression outcomes in the experimental group compared to those in the control group. The one RCT involving pharmacologic intervention did not show benefit relative to the control group.[72] Overall, the evidence available is not sufficient to draw conclusions about this key question. These results, although limited, do suggest that providing some form of psychosocial support to pregnant women at risk of having a depressive illness may decrease depressive symptoms.

Discussion

The available research suggests that depression is one of the most common complications of the prenatal and postpartum periods, and that fairly accurate and feasible screening measures are available. The prenatal or postpartum periods are clearly not times for nonpsychiatric clinicians to ignore depression screening, which is routinely recommended for patients seen in primary care settings.[73,74] Specifics of the course of a depressive illness with onset during the perinatal period, including the severe physiologic and psychological challenges unique to this period that complicate the identification and management of perinatal depression, seem to suggest that this topic would have a substantial degree of high-quality research. We were surprised by the paucity of such evidence in this area. If one assumes that perinatal depression is a significant mental health and public health problem, then larger scale studies are needed that involve each of these domains. The small number and small size of relevant studies are not adequate to guide national policy.

Reflecting on the three key questions addressed in this report, we have concluded generally that the level of research warrants both improvement and expansion. For KQ 1, prevalence studies need to better account for the racial and ethnic mix of perinatal depression in the U.S. population. We do not have good evidence on whether perinatal depression rates differ among various ethnic groups and, if so, how. The absence of information on populations other than the white population was dramatic. A better understanding of racial and ethnic variations could help clinicians know where to target screening programs and researchers know where to target studies on screening tools, and it could help researchers clarify the need for more nationally representative perinatal depression

samples. Furthermore, researchers need to clarify whether the incidence of perinatal depression is greater than the incidence of depression in nonchildbearing women of similar ages.

For KQ 2, the quality grades point to several areas in which improvements in study design and conduct are needed. In particular, future studies on the test characteristics of screeners must be designed with sample size estimates that take prevalence into account and that project a reasonably precise estimate of sensitivity for the particular illness. Moreover, samples should more closely mirror the target population; specifically, subsequent studies need to provide a more representative racial and ethnic mix. In addition, studies should incorporate a range of other demographic variables that could influence screening performance, such as socioeconomic status measures, and assess the screening tools in these subpopulations.

Furthermore, as Beck and Gable did,[51] future research should continue to assess and directly compare multiple screening instruments. This design would provide a head-to-head comparison to allow an evaluation of which screening instrument is more accurate in the setting in which the investigations are carried out. Moreover, studies evaluating the cost-effectiveness of screening—specifically assessing the relative costs of false-negative and false-positive designation, the degree of provider burden, and patient acceptability—are needed to provide insights on how to consider target sensitivity and specificity when attempting to maximize cost-effectiveness.

Diagnosis is another area of concern. Subsequent studies should carefully consider whether to target major depression alone, for which beneficial treatments clearly exist, or a combined category of major and minor depression, a heterogeneous group for which treatment benefit is unclear. Given that our results suggest that available screening tools identify major depression alone more accurately, and noting that the general benefit of interventions is more apparent for major depression alone, we believe that an evidence-based public health perspective recommends targeting major depression alone.

Timing is another factor deserving more thought in future studies. The issue involves both the need for more epidemiology to confirm prevalence rates at different times as well as the need to confirm what time point(s) would identify the greatest number of depressed women. The bulk of the few screening studies we identified had been conducted in the first 3 months postpartum. Our best estimates of prevalence suggest that depression may remain high for several more months. More studies are needed to better delineate periods of peak prevalence and incidence—to include not just 3 months but also 6 weeks, 6 months, and 12 months—and subsequent screening studies need to consider testing properties of screening at these later time periods. The very small number of adequate studies currently available hampers plans for screening and intervention programs because the best time for screening, and hence the best clinic location, is not clear. If peak prevalence and incidence occur within the first 6 weeks, the obstetrics clinic is a prime place to target resources for such a program. If, however, it peaks after this time, most postpartum women will have completed their followup care with an obstetrician, so programs in an obstetrics clinic may be less helpful. In this case, it is possible that programs targeting new mothers in family medicine, internal medicine, or pediatric clinics might be more effective.

For KQ 3, several similar or related issues emerged as well. First, studies addressing the relationship between screening and outcome need to recruit and retain sample sizes that are large enough to yield adequate power to detect relevant differences. Second, screening and outcome studies must include populations with a racial and ethnic mix that is more representative of the U.S. populations than the work we have seen to date. Third, interventions involved should be more consistent with what we know as evidence-based treatments for depression,[9] i.e., antidepressant medications[75] and/or psychotherapies such as cognitive behavioral therapy[76] or interpersonal psychotherapy.[77]

Another major issue is the types of screening measures to be used henceforth. Of the three KQ 3 studies rated as good,[62,65,72] only the one by Dennis and colleagues used a depression screener (EPDS).[65] Researchers should consider developing and using standard screening measures and using similar cutoff points, so that some elements of separate studies could more readily be compared. Screening tools with the best supporting evidence would seem to be the best candidates. While the evidence base remains quite limited and any conclusions are preliminary, at this time those instruments would appear to be the EPDS or the PDSS. For major depression alone, an EPDS cutoff of $\geq$13 or a PDSS cutoff of $\geq$81 are reasonably supported by the evidence as thresholds to use. For major or minor depression, we found the results too inconclusive to make even a preliminary recommendation.

Finally, studies should be designed to address whether the screening process itself leads to better access to proven treatment and improved outcome relative to usual care. We support additional research on interventions per se, but we conclude that important questions remain about the impact of the screening element. Reviewing studies that used screening as a means of identifying women potentially at high risk and enrolling them in interventional studies is not a sufficient approach to answering issues about the effectiveness of screening.

Availability of the Full Report

The full evidence report from which this summary was taken was prepared for the Agency for Healthcare Research and Quality (AHRQ) by the RTI-University of North Carolina Evidence-based Practice Center under Contract No. 290-02-0016. It is expected to be available in spring 2005. At that time, printed copies may be obtained free of charge from the AHRQ Publications Clearinghouse by calling 800-358-9295. Requesters should ask for Evidence Report/Technology Assessment No. 119, *Perinatal Depression: Prevalence, Screening*

Accuracy, and Screening Outcomes. In addition, Internet users will be able to access the report and this summary online through AHRQ's Web site at www.ahrq.gov.

Suggested Citation

Gaynes BN, Gavin N, Meltzer-Brody S, Lohr KN, Swinson T, Gartlehner G, Brody S, Miller WC. Perinatal Depression: Prevalence, Screening Accuracy, and Screening Outcomes. Summary, Evidence Report/Technology Assessment No. 119. (Prepared by the RTI-University of North Carolina Evidence-based Practice Center under Contract No. 290-02-0016.) AHRQ Publication No. 05-E006-1. Rockville, MD: Agency for Healthcare Research and Quality. February 2005.

References and Included Studies

1. Kessler RC. Epidemiology of women and depression. J Affect Disord 2003; 74(1):5-13.
2. Robins L, Regier D. Psychiatric disorders in America. New York: Free Press; 1991.
3. Depression Guideline Panel. Depression in Primary Care: Volume 1. Detection and Diagnosis. Clinical Practice Guideline, Number 5. Rockville, Md: Agency for Health Care Policy and Research, 1993. AHCPR Publication No. 93-0550.
4. Burke KC, Burke JD Jr, Rae DS, et al. Comparing age at onset of major depression and other psychiatric disorders by birth cohorts in five US community populations. Arch Gen Psychiatry 1991; 48(9):789-95.
5. Murray L, Stein A. The effects of postnatal depression on the infant. Baillieres Clin Obstet Gynaecol 1989; 3(4):921-33.
6. Marmorstein NR, Malone SM, Iacono WG. Psychiatric disorders among offspring of depressed mothers: associations with paternal psychopathology. Am J Psychiatry 2004; 161(9):1588-94.
7. Burke L. The impact of maternal depression on familial relationships. Int Rev Psychiatry 2003; 15(3):243-55.
8. Flynn HA, Davis M, Marcus SM, et al. Rates of maternal depression in pediatric emergency department and relationship to child service utilization. Gen Hosp Psychiatry 2004; 26(4):316-22.
9. American Psychiatric Association. Practice guideline for the treatment of patients with major depression (revision). Am J Psychiatry 2000; 157(4).
10. Wagner HR, Burns BJ, Broadhead WE, et al. Minor depression in family practice: functional morbidity, comorbidity, service utilization, and outcomes. Psychol Med 2000; 30(2):1377-90.
11. Oxman TE, Sengupta A. Treatment of minor depression. Am J Geriatr Psychiatry 2002; 10(3):256-64.
12. Judd LL, Rapaport MH, Yonkers KA, et al. Randomized, placebo-controlled trial of fluoxetine for acute treatment of minor depressive disorder. Am J Psychiatry 2004; 161(10):1864-71.
13. Klein M, Essex MJ. Pregnant or depressed? The effect of overlap between symptoms of depression and somatic complaints of pregnancy on rates of major depression during the second trimester. Depression 1994; 2:1994-5.
14. O'Hara MW, Neunaber DJ, Zekoski EM. Prospective study of postpartum depression: prevalence, course, and predictive factors. J Abnorm Psychol 1984; 93(2):158-71.
15. Bloch M, Daly RC, Rubinow DR. Endocrine factors in the etiology of postpartum depression. Compr Psychiatry 2003; 44(3):234-46.
16. Jones I, Craddock N. Familiality of the puerperal trigger in bipolar disorder: results of a family study. Am J Psychiatry 2001; 158(6): 913-7.
17. O'Hara MW, Swain AM. Rates and risk of postpartum depression — a meta-analysis. Int Rev Psychiatry 1996; 8:37-54.
18. Llewellyn AM, Stowe ZN, Nemeroff CB. Depression during pregnancy and the puerperium. J Clin Psychiatry 1997; 58 Suppl 15:26-32.
19. Yonkers KA, Ramin SM, Rush AJ, et al. Onset and persistence of postpartum depression in an inner-city maternal health clinic system. Am J Psychiatry 2001; 158(11):1856-63.
20. Gaynes B, Gavin N, Meltzer-Brody S, et al. Perinatal Depression: Feasibility Study. Final Report from the RTI-International-University of North Carolina Evidence-based Practice Center to the Agency for Healthcare Research and Quality under Contract No. 290-02-0016. Research Triangle Park, NC; 2003.
21. Cochrane Methods Working Group, Working Group on Systematic Review of Screening and Diagnostic Tests. Recommended Methods; 1996.
22. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. J Epidemiol Community Health 1998; 52(6):377-84.
23. Affonso DD, Lovett S, Paul SM, et al. A standardized interview that differentiates pregnancy and postpartum symptoms from perinatal clinical depression. Birth 1990; 17(3):121-30.
24. Watson JP, Elliott SA, Rugg AJ, et al. Psychiatric disorder in pregnancy and the first postnatal year. Br J Psychiatry 1984; 144: 453-62.
25. Areias ME, Kumar R, Barros H, et al. 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(1):30-5.
26. Berle J, Aarre T, Mykletun A, et al. Screening for postnatal depression. Validation of the Norwegian version of the Edinburgh Postnatal Depression Scale, and assessment of risk factors for postnatal depression. J Affect Disord 2003; 76(1-3):151-6.
27. Campbell SB, Cohn JF. Prevalence and correlates of postpartum depression in first-time mothers. J Abnorm Psychol 1991; 100(4): 594-9.
28. Cooper PJ, Murray L, Hooper R, et al. The development and validation of a predictive index for postpartum depression. Psychol Med 1996; 26(3):627-34.
29. Cox JL, Connor Y, Kendell RE. Prospective study of the psychiatric disorders of childbirth. Br J Psychiatry 1982; 140:111-7.
30. Garcia-Esteve L, Ascaso C, Ojuel J, et al. Validation of the Edinburgh Postnatal Depression Scale (EPDS) in Spanish mothers. J Affect Disord 2003; 75(1):71-6.
31. Gotlib IH, Whiffen VE, Mount JH, et al. Prevalence rates and demographic characteristics associated with depression in pregnancy and the postpartum. J Consult Clin Psychol 1989; 57(2):269-74.
32. 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(3):445-53.
33. Kent GN, Stuckey BG, Allen JR, et al. Postpartum thyroid dysfunction: clinical assessment and relationship to psychiatric affective morbidity. Clin Endocrinol 1999; 51(4):429-38.
34. Kitamura T, Shima S, Sugawara M, et al. Psychological and social correlates of the onset of affective disorders among pregnant women. Psychol Med 1993; 23:967-75.
35. Kitamura T, Sugawara M, Shima S, et al. Temporal variation of validity of self-rating questionnaires: improved validity of repeated use of Zung's Self-Rating Depression Scale among women during the perinatal period. J Psychosom Obstet Gynecol 1999; 20(2):112-7.
36. Kumar R, Robson KM. A prospective study of emotional disorders in childbearing women. Br J Psychiatry 1984; 144:35-47.
37. Lee D, Yip A, Chiu H, et al. A psychiatric epidemiological study of postpartum Chinese women. Am J Psychiatry 2001; 158(2):220-6.

38. Lee D, Yip A, Chiu H, et al. Screening for postnatal depression: are specific instruments mandatory? J Affect Disord 2001; 63(1-3):233-8.
39. Lucas A, Pizarro E, Granada ML, et al. Postpartum thyroid dysfunction and postpartum depression: are they two linked disorders? Clin Endocrinol 2001; 55(6):809-14.
40. Murray D, Cox JL. Screening for depression during pregnancy with the Edinburgh Depression Scale (EPDS). J Reprod Infant Psychol 1990; 8(2):99-107.
41. Pop VJ, Essed GG, de Geus CA, et al. Prevalence of post partum depression—or is it post-puerperium depression? Acta Obstet Gynecol Scand 1993; 72(5):354-8.
42. Whiffen V. Vulnerability of postpartum depression: a prospective multivariate study. J Abnorm Psychol 1988; 97(4):467-74.
43. Yamashita H, Yoshida K, Nakano H, et al. Postnatal depression in Japanese women. Detecting the early onset of postnatal depression by closely monitoring the postpartum mood. J Affect Disord 2000; 58(2):145-54.
44. Yoshida K, Marks M, Kibe N, et al. Postnatal depression in Japanese women who have given birth in England. J Affect Disord 1997; 43(1):69-77.
45. Cooper PJ, Campbell EA, Day A, et al. Non-psychotic psychiatric disorder after childbirth. A prospective study of prevalence, incidence, course and nature. Br J Psychiatry 1988; 152:799-806.
46. Cox JL, Murray D, Chapman G. A controlled study of the onset, duration and prevalence of postnatal depression. Br J Psychiatry 1993; 163:27-31.
47. O'Hara MW, Zekoski EM, Philipps LH, et al. Controlled prospective study of postpartum mood disorders: comparison of childbearing and nonchildbearing women. J Abnorm Psychol 1990; 99(1):3-15.
48. 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(4):351-8.
49. Georgiopoulos AM, Bryan TL, Wollan P, et al. Routine screening for postpartum depression. J Fam Pract 2001; 50(2):117-22.
50. Ballard CG, Davis R, Cullen PC, et al. Prevalence of postnatal psychiatric morbidity in mothers and fathers. Br J Psychiatry 1994; 164(6):782-8.
51. Beck CT, Gable RK. Comparative analysis of the performance of the Postpartum Depression Screening Scale with two other depression instruments. Nurs Res 2001; 50(4):242-50.
52. Boyce P, Stubbs J, Todd A. The Edinburgh Postnatal Depression Scale: validation for an Australian sample. Aust N Z J Psychiatry 1993; 27(3):472-6.
53. Cox J, Chapman G, Murray D, et al. Validation of the Edinburgh Postnatal Depression Scale (EPDS) in non-postnatal women. J Affect Disord 1996; 39(3):185-9.
54. Harris B, Huckle P, Thomas R, et al. The use of rating scales to identify post-natal depression. Br J Psychiatry 1989; 154:813-7.
55. Leverton TJ, Elliott SA. Is the EPDS a magic wand? 1. A comparison of the Edinburgh Postnatal Depression Scale and health visitor report as predictors of diagnosis on the Present State Examination. J Reprod Infant Psychol 2000; 18(4):279-96.
56. Murray L, Carothers A. The validation of the Edinburgh Post-natal Depression Scale on a community sample. Br J Psychiatry 1990; 157:288-90.
57. Williams JW Jr, Pignone M, Ramirez G, et al. Identifying depression in primary care: a literature synthesis of case-finding instruments. Gen Hosp Psych 2002; 24:225-37.
58. Brugha TS, Wheatley S, Taub NA, et al. Pragmatic randomized trial of antenatal intervention to prevent post-natal depression by reducing psychosocial risk factors. Psychol Med 2000; 30(6):1273-81.
59. Elliott SA, Leverton TJ, Sanjack M, et al. Promoting mental health after childbirth: a controlled trial of primary prevention of postnatal depression. Br J Clin Psychol 2000; 39(Pt 3):223-41.
60. Stamp GE, Williams AS, Crowther CA. Evaluation of antenatal and postnatal support to overcome postnatal depression: a randomized, controlled trial. Birth 1995; 22(3):138-43.
61. Zlotnick C, Johnson SL, Miller IW, et al. Postpartum depression in women receiving public assistance: pilot study of an interpersonal-therapy-oriented group intervention. Am J Psychiatry 2001; 158 (4):638-40.
62. Armstrong K, Fraser J, Dadds M, et al. A randomized, controlled trial of nurse home visiting to vulnerable families with newborns. J Paediatr Child Health 1999; 35(3):237-44.
63. Chabrol H, Teissedre F, Saint-Jean M, et al. Prevention and treatment of post-partum depression: a controlled randomized study on women at risk. Psychol Med 2002; 32(6):1039-47.
64. Chen CH, Tseng YF, Chou FH, et al. Effects of support group intervention in postnatally distressed women. A controlled study in Taiwan. J Psychosom Res 2000; 49(6):395-9.
65. Dennis CL. The effect of peer support on postpartum depression: a pilot randomized controlled trial. Can J Psychiatry 2003; 48(2): 115-24.
66. Fleming AS, Klein E, Corter C. The effects of a social support group on depression, maternal attitudes and behavior in new mothers. J Child Psychol Psychiatry 1992; 33(4):685-98.
67. Hiscock H, Wake M. Randomised controlled trial of behavioural infant sleep intervention to improve infant sleep and maternal mood. Br Med J 2002; 324(7345):1062-5.
68. Honey KL, Bennett P, Morgan M. A brief psycho-educational group intervention for postnatal depression. Br J Clin Psychol 2002; 41(Pt 4):405-9.
69. Horowitz JA, Bell M, Trybulski J, et al. Promoting responsiveness between mothers with depressive symptoms and their infants. J Nurs Scholarsh 2001; 33(4):323-9.
70. Onozawa K, Glover V, Adams D, et al. Infant massage improves mother-infant interaction for mothers with postnatal depression. J Affect Disord 2001; 63(1-3):201-7.
71. Wisner KL, Wheeler SB. Prevention of recurrent postpartum major depression. Hosp Community Psychiatry 1994; 45(12):1191-6.
72. Wisner KL, Perel JM, Peindl KS, et al. Prevention of recurrent postpartum depression: a randomized clinical trial. J Clin Psychiatry 2001; 62(2):82-6.
73. Pignone MP, Gaynes BN, Rushton JL, et al. Screening for depression in adults: a summary of the evidence for the US Preventive Services Task Force. Ann Intern Med 2002; 136(10):765-76.
74. U.S. Preventive Services Task Force. Screening for depression: recommendations and rationale. Ann Intern Med 2002; 136(10): 760-4.
75. Hoffbrand S, Howard L, Crawley H. Antidepressant treatment for post-natal depression. Nurs Times 2001; 97(45):35.
76. Cooper P, Murray L. The impact of psychological treatments of postpartum depression on maternal mood and infant development. In: Cooper P and Murray L, editors. Postpartum depression and child development. New York: Guilford; 1997. p. 201-20.
77. O'Hara MW, Stuart S, Gorman LL, et al. Efficacy of interpersonal psychotherapy for postpartum depression. Arch Gen Psychiatry 2000; 57(11):1039-45.

www.ahrq.gov
AHRQ Pub. No. 05-E006-1
February 2005
ISSN 1530-440X

Chapter 1. Introduction

Depressive disorders are ubiquitous and remarkably impairing; they occur throughout the lifespan. Lifetime prevalence rates of depression from community-based surveys range from 4.9 percent to 17.1 percent.[1-3] Gender plays an important role in the prevalence rates of depression; women report a history of major depression at nearly twice the rate of men.[4] In particular, women of childbearing age are at high risk for major depression.[2,3,5] Pregnancy and new motherhood may increase the risk of depressive episodes.

Depression is the leading cause of disease-related disability among women in the world.[6] It can have devastating consequences, not only for the women experiencing it but also for the women's children and family.[7-9] For example, Stein and colleagues found that the mother-child interactions of depressed mothers and their children were of lower quality than those of nondepressed mothers,[10] and Flynn et al. found that maternal depression was related to both missed pediatric appointments and greater use of emergency department services.[11] A review of other research in this area points out that parental depression has been linked to raised levels of psychiatric disturbances among children and to greater child insecurity in attachment relationships.[7,8]

The importance of detecting and treating perinatal depression has only recently been recognized. Perinatal depression encompasses major and minor depressive episodes that occur either during pregnancy or within the first 12 months following delivery. Major depression is a distinct clinical syndrome for which treatment is clearly indicated,[12] whereas the definition and management of minor depression are less clear. Minor depression is an impairing yet less severe constellation of depressive symptoms[13] for which controlled trials have not consistently indicated whether particular interventions are more effective than placebo.[14,15] In this report, we address major depressive episodes alone, which we refer to as major depression, as well as a broader grouping of major or minor depression, which we refer to as such or by the more general terms "depression" or "depressive illness." We necessarily rely on the specific definitions of minor depression used by the different authors of the reviewed studies.

Another mental disorder that can occur in the perinatal period is postpartum psychosis. Unlike postpartum depression, postpartum psychosis is a relatively rare event with an estimated incidence of 1.1 to 4.0 cases per 1,000 deliveries.[16] The onset of postpartum psychosis is usually acute, within the first 2 weeks of delivery, and appears to be more common in women with a strong family history of bipolar or schizoaffective disorder.[17] Postpartum psychosis is an important disorder in its own right, but it is not addressed specifically in this report.

Perinatal depression, major or minor, often goes unrecognized because many of the discomforts of pregnancy and the puerperium are similar to symptoms of depression.[18,19] The onset of major depression is believed to be impressively common in the postpartum period; researchers have found a 3-fold increase in the onset of major or minor depression in the first 5 weeks postpartum compared to women of similar age, marital status, and parity at nonchildbearing times.[20] However, the precise levels of the prevalence and incidence of perinatal depression are uncertain. Published estimates of the rate of major or minor depression in the postpartum period range widely—from 5 percent to more than 25 percent of new mothers—depending on the assessment method, the timing of the assessment, and population characteristics.[21-23]

Note: Appendixes and Evidence Tables cited in this report are provided electronically at http://www.ahrq.gov/clinic/epcindex.htm.

Although many screening instruments have been developed or modified to detect major or minor depression in pregnant and newly delivered women, the evidence on their screening accuracy relative to a reference standard has yet to be systematically reviewed and assessed.[24] Evidence on the effectiveness of screening all pregnant women and providing a preventive intervention to those scoring at high risk has also not been systematically investigated and evaluated.[24]

To address these gaps, the Agency for Healthcare Research and Quality (AHRQ) in collaboration with the Safe Motherhood Group (SMG) commissioned the RTI International-University of North Carolina (RTI-UNC) Evidence-based Practice Center (EPC) to conduct a systematic evidence review on three questions related to perinatal depression. These questions (provided in Table 1) address the prevalence and incidence of perinatal depression, the accuracy of screening instruments for perinatal depression, and the effectiveness of interventions for women who are found to be at high risk for developing perinatal depression.

Table 1. Key questions for the evidence report on perinatal depression

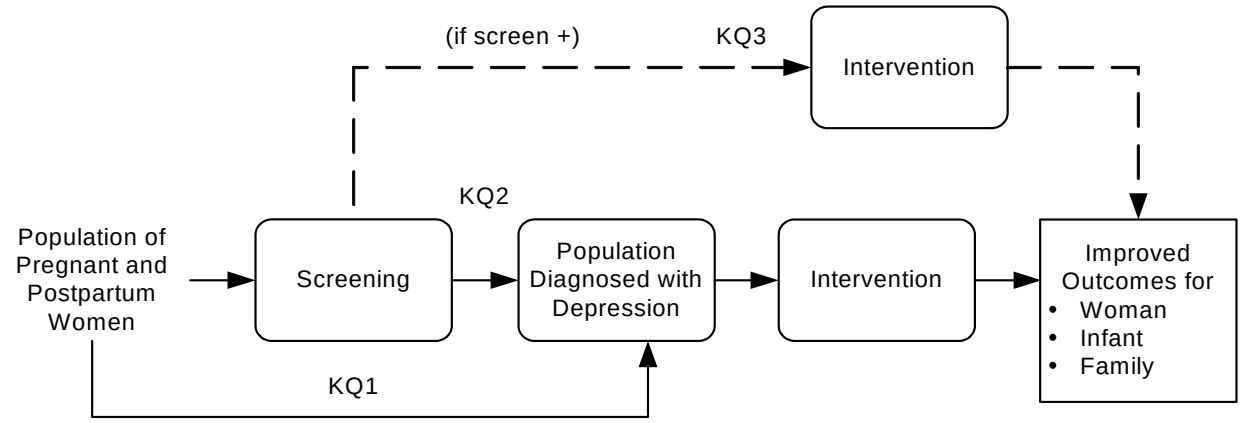
Key Question	
1	What is the incidence and prevalence of depression (major or minor) during pregnancy and during the postpartum period? Is it increased during pregnancy and the postpartum period compared to nonchildbearing periods?
2	What is the accuracy of different screening tools for detecting depression during pregnancy and the postpartum period?
3	Does prenatal or early postnatal screening for depressive symptoms with subsequent intervention lead to improved outcomes?

We show a simple schematic of the causal pathway for the screening and treatment of perinatal depression and the links addressed by the three study questions in Figure 1. For all three questions, we begin with a general population of pregnant or postpartum women. The first key question addresses the percentage of this population diagnosed with depression at various points and periods of time throughout pregnancy and the first postpartum year—that is, the prevalence and incidence of the disorder. Prevalence and incidence can be measured in different ways and may vary by population characteristics. We synthesize available evidence on prevalence and incidence measured in a similar manner at or over the same general period of time and analyze the impact of selected population and study characteristics. Studies with a comparison group of women of similar age during nonchildbearing periods are also reviewed to determine whether the prevalence or incidence of depression increases during pregnancy and the first postpartum year.

The second key question addresses the accuracy of different screening instruments for postpartum depression—that is, how well different instruments detect pregnant or postpartum women who have depression (sensitivity) and pregnant and postpartum women who do not have depression (specificity). We identify and abstract English-language and non-English-language studies of various cutoff scores for a variety of commonly used instruments but review only the English-language studies.

Finally, we review studies that provide evidence on whether interventions can reduce the prevalence and incidence of perinatal depression for women who are screened and found to be at high risk for the disorder. We also summarize evidence in these studies on the effect of

Figure 1. Causal pathway for the screening and treatment of perinatal depression



screening with subsequent intervention on other health outcomes for the woman and her infant. This third question addresses whether the screening process itself ultimately leads to improved outcomes for perinatal depression. Studies had to use some form of screening to identify women for testing interventions involving a technique to address psychological status in the woman and had to have an outcome measured related to depression severity.

In this report, we provide the results of our systematic search and review of the published literature for evidence addressing these questions. In conducting this study, our intent was to answer the questions using the most reliable evidence available, obtain a sense of the strength of the available evidence, and identify gaps in the knowledge base that require further research. We follow a discussion of our general approach and methods in Chapter 2 with discussions of each of the question-specific methods and findings (Chapters 3, 4, and 5). In Chapter 6, we discuss our main conclusions, comment on the state of the evidence, and offer an agenda for future research studies. Appendix A presents the exact search strings for the electronic database searches. Appendix B contains copies of our quality rating forms. Appendix C presents the evidence tables, Appendix D provides a list of excluded articles, and Appendix E provides acknowledgments.

Chapter 2. Methods

In conducting this systematic review, we followed standardized procedures developed by the Agency for Healthcare Research and Quality (AHRQ) in collaboration with all its Evidence-based Practice Centers (EPCs) for such reviews. This chapter documents how we implemented those procedures to answer the three key questions on perinatal depression. We first discuss the role of the Technical Expert Advisory Group (TEAG). We then describe our inclusion/exclusion criteria, our strategy for identifying articles relevant for addressing the key questions, and our process for abstracting relevant information from the eligible articles and generating evidence tables. We also discuss our criteria for grading the quality of individual articles and the strength of the evidence as a whole. Finally, we explain the peer review process.

Role of the Technical Expert Advisory Group

Throughout the project, we enlisted the assistance of a TEAG to react to work in progress and advise us on substantive issues or possibly overlooked areas of research. The TEAG included four individuals with collective expertise in obstetrics, psychiatry, psychology, and research methods and both clinical and research experience in perinatal depression (see Appendix E, Acknowledgments). As in all such systematic reviews, the TEAG contributed to AHRQ's broader goals of (1) creating and maintaining science partnerships as well as public-private partnerships and (2) meeting the needs of an array of potential customers and users of its products. Thus, the TEAG was both an additional resource and a sounding board during the project.

To ensure robust, scientifically relevant work, we called on the TEAG to participate in conference calls and discussions through e-mail to

- refine the analytic framework and key questions at the beginning of the project;
- discuss the preliminary assessment of the literature, including inclusion/exclusion criteria;
- identify relevant literature not revealed through our literature searches;
- provide input on the information and categories included in evidence tables;
- review proposed methods for data synthesis; and
- help interpret preliminary findings.

Because of their extensive knowledge of this topic, we also asked TEAG members to participate in the external peer review of the draft report.

Note: Appendixes and Evidence Tables cited in this report are provided electronically at http://www.ahrq.gov/clinic/epcindex.htm.

Literature Search Strategy

To ensure a comprehensive and reproducible literature search and appraisal, we identified relevant research studies using an explicit search strategy and uniformly applied a set of inclusion and exclusion criteria to the identified studies. We describe our criteria and approach in this section.

Inclusion and Exclusion Criteria

To identify relevant studies, we generated a list of inclusion and exclusion criteria for each key question. We made the criteria fairly restrictive to ensure that our conclusions would be based on the highest quality data available with the lowest risk of bias. Some criteria were common across the three key questions; others were specific to the question. Table 2 summarizes the criteria.

Table 2. Inclusion/exclusion criteria by key question

Category	Inclusion	Exclusion
All Key Questions		
Publication date	1980 through March 2004	
Setting	Developed countries only Any clinical setting or homes	Less-developed countries
Populations	Humans only Depressive illness assessed during pregnancy or first postpartum year	Animal studies Trials addressing exclusively bipolar disorder, a primary psychotic disorder, or maternity blues
Study design	Original data	Case reports, case series, letters, editorials, and non-systematic reviews that have no original data
Prevalence and Incidence (Key Question 1)		
Study design	Prevalence or incidence study Epidemiologic cohort or weighted to be representative	
Study population	Diagnosis of major depressive episode or postpartum depressive episode using criterion standard (see text)	Depressive disorder identified only by screen
Screening Accuracy (Key Question 2)		
Study design	Must have criterion standard (see text) Studies must be prospective	Case-control studies
Outcomes of interest	Sensitivity and specificity	
Study population	Patients who are screened for depression during pregnancy or during 12 months postpartum	Patients with known current depressive episode
Screening Interventions Criteria (Key Question 3)		
Study design	Randomized controlled trial or prospective cohort study	Case-control studies
Outcomes of interest	Clinical status and functioning	
Study population	Patients identified by a screen during pregnancy or during 12 months postpartum as being at high risk of having depression	Patients with known current depressive episode

For all key questions, studies had to report on original data, be in English, and be published from January 1980 through March 2004. This time frame ensured that the applied reference standards were consistent with the *Diagnostic and Statistical Manual for Mental Disorders, Third Edition* (DSM-III), or later criteria for the diagnosis of depression. The study could be conducted in any clinical setting or home but had to be from a developed country to increase the likelihood of being generalizable to the US population. In our original criteria submitted in the research proposal, we proposed including only studies done in the United States, the United Kingdom and other Commonwealth/English-speaking countries, Europe, and Scandinavia. However, we determined after abstract review that such limitations would leave out a large number of relevant studies. Therefore, we modified our inclusion criteria to accept any study conducted in developed countries where the population could be generalized to pregnant and postpartum women in the United States, regardless of the language spoken. We excluded studies published before 1980 or in a language other than English and those on women in less developed countries. We also excluded studies of women with major or minor depression in which the outcomes of interest were not distinguishable from those for women with bipolar disorder, primary psychotic disorders, or maternity blues.

In addition, studies for all key questions had to assess women for major depression either alone or together with minor depression during pregnancy or the first year postpartum by means of a clinical assessment or structured clinical interview. For Key Question (KQ) 1, we excluded studies of the prevalence and incidence of perinatal depression that relied solely on self-report screens to identify depression. For KQs 2 and 3, we excluded studies that included women with known depressive disorders at the outset. In KQ 2, study investigators used the clinical assessment or structured clinical interview as the criterion or gold standard with which to assess the properties of the screening instrument. In many KQ 3 studies, investigators used the clinical assessment to measure the depression outcomes from screening with subsequent intervention among women found to be at elevated risk of depression. Studies that measured women's mood using self-report measures only were also included in KQ 3.

For KQ 1, we included both prospective and retrospective studies of the prevalence and incidence of perinatal depression and studies that were conducted for purposes other than determining the prevalence and incidence of perinatal depression but nevertheless included a population-based estimate meeting the other inclusion criteria (e.g., studies of the properties of screening instruments). Furthermore, to answer the second part of KQ 1, we included both clinical trials and case-control studies comparing the incidence or prevalence of depression among pregnant women and newly delivered mothers to prevalence among women of similar age during other nonchildbearing periods of their lives. We included only prospective studies in those reviewed for KQs 2 and 3.

Literature Search and Retrieval Process

We used three strategies to identify studies providing evidence related to the key questions: systematic searches of electronic databases using both search terms and author names, hand searches of reference lists of included articles, and consultation with the TEAG. First, we generated a list of Medical Subject Heading (MeSH) search terms for each key question in the feasibility study. We used these terms to search standard electronic databases: MEDLINE, Cumulative Index to Nursing & Allied Health Literature (CINAHL), PsycINFO, Sociofile, and the Cochrane Library.

We conducted the electronic database searches twice. We initially did them in April 2003 for the feasibility study.[24] That study included three additional key questions, including questions on natural history, risk factors, and treatment effectiveness for perinatal depression. We found relevant articles for the three key questions of the current study under the natural history and treatment effectiveness searches. We therefore conducted these and the incidence or prevalence and mass screening searches again in March 2004 to capture any studies published and posted in the interim.

The subject headings used and the total yield from each source are shown in Table 3 by key question. We found a total of 837 unduplicated citations in the electronic searches and picked up an additional 9 citations through the hand searches and discussion with the TEAG, for a total of 846 citations. We also searched the Cochrane Collaboration database for prior systematic reviews using the keywords "perinatal" and "depression." This search yielded 38 reviews.

Table 3. Literature search strategies and yield

Key Question	Search Terms	Yield
All	MEDLINE and& CINAHL: ('Puerperal Disorders' and (Depression or 'Depressive Disorder')) or 'Depression, Postpartum/ or perinatal depression.mp' PsycINFO: "Depression, Postpartum" Sociofile: "Postpartum Depression"	
KQ 1	. . . and "Natural History" or "Cohort Studies" or "Longitudinal Studies" or . . . and Incidence or Prevalence	MEDLINE = 165 CINAHL = 42 PsycINFO = 88 Sociofile = 21 Total unduplicated = 256
KQ 2	. . . and "Mass Screening"	MEDLINE = 67 CINAHL = 25 PsycINFO = 28 Sociofile = 1 Total = unduplicated 96
KQ 3	. . . and treatment.mp or Therapeutics or "treatment failure" or "treatment outcomes" or "treatment duration" or treatment errors" or "treatment delay" or "treatment complications"	MEDLINE = 513 CINAHL = 90 PsycINFO = 91 Sociofile = 5 Total unduplicated = 485

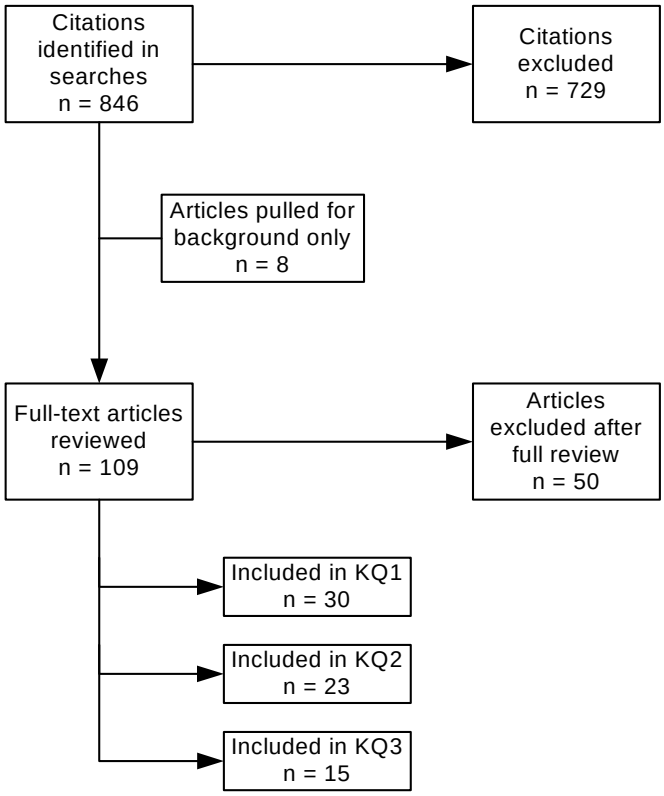
Three senior reviewers with clinical expertise in perinatal depression reviewed the abstracts of articles identified during the literature search. Two clinicians evaluated each abstract against the inclusion criteria and resolved any differences in inclusion by consensus. In several instances, the abstracts did not provide enough information to make an inclusion decision; we pulled full articles to review for those studies. Of the 846 articles identified, 729 did not meet the inclusion criteria for any of the key questions and were therefore excluded, 8 studies were pulled for background only, and the remaining 109 articles were pulled for a full review.

Among the 109 studies pulled for full review, 50 did not meet our inclusion/exclusion criteria for any of the three key questions. The most common reason for exclusion was the absence of a gold standard (i.e., either a clinical assessment or structured clinical interview) for assessing

depression, which eliminated 26 studies. Ten of the studies pulled for the evaluation of the properties of screening instruments were excluded because they did not report sensitivity and specificity or data from which these statistics could be computed. Other reasons for exclusion were depression assessed after the first year postpartum, no depression outcome measure, a retrospective study design, and restriction of the study sample to specific population subgroups (e.g., teens, patients of psychiatric hospitals). We based the last exclusion on two lines of reasoning. First, although groups such as adolescents are a key subgroup, our charge was to ensure that our results were generalizable to the broader US population. Second, these specific subpopulations are different enough from the remainder of the population that they warrant separate consideration. We excluded only one study because it was limited to an adolescent population.

We included the remaining 59 studies in our review, and some met the inclusion criteria for more than one key question. We abstracted 30 studies for KQ 1, 23 for KQ 2, and 15 for KQ 3. We provide a graphical presentation of the disposition of the citations in Figure 2.

Figure 2. Perinatal depression article disposition



Data Collection and Assessment

The data collection process involved abstracting relevant information from the eligible articles and generating evidence tables that present the key details of the study design and the major findings from the articles. A trained member of the study team read and abstracted each article; a second member checked the table entries for accuracy against the original article.

Appendix C contains the final evidence tables in their entirety. They provide the study design details and major findings. The dimensions of each study design abstracted vary by key question, but they contain some common elements, such as author, year of publication, study location (e.g., country, state), population description, and sample size. We also collected information on the clinical interview instrument and diagnostic criteria used to diagnose depression and the age and racial and ethnic distribution of study subjects in each study.

The study results are recorded in the form reported in the article. However, for assessing consistency of results across the studies and for combining study results in a meta-analysis (see below), we also transformed the study results when necessary into consistent outcome measures using the appropriate statistical formulas. These computed data elements are shown in bold in the evidence tables (Appendix C).

We conducted data abstraction electronically in a word processing program and in such a way that study identifiers and results were easily transferred from the forms to electronic files for input into programs for meta-analysis.

Meta-analysis

We conducted a meta-analysis of the different prevalence and incidence estimates from studies abstracted for KQ 1 to arrive at single prevalence and incidence estimates for particular periods and points in time. We elaborate on these methods in Chapter 3. We also conducted meta-analyses of the different estimates of the receiver operating characteristics (ROC) curves for screening instruments evaluated for KQ 2, as described in Chapter 4. Because of the diversity of screening instruments and prevention interventions in the studies found for KQ 3, we did not conduct a meta-analysis for this key question.

Quality of Individual Articles

At the same time that we abstracted information on the study designs and findings in the included articles, we rated the quality of the studies. We developed a quality rating form for the screening accuracy (KQ 2) articles from criteria identified by the Cochrane Methods Working Group on Systematic Review of Screening and Diagnostic Tests.[25] For studies addressing KQ1 and KQ 3, we modified the quality rating forms developed by Downs and Black for RCTs and observational studies.[26] These forms are provided in Appendix B.

The quality rating forms rated the reporting completeness and clarity, external validity, internal validity, and the power or precision of each study for the relevant key questions. Hence, the ratings refer to the usefulness or quality of the article for our purposes and not necessarily for the original purpose of the research or article. Studies that were included in more than one key

question were rated separately for each key question. The specific quality items rated are described in more detail in Chapters 3, 4, and 5 for KQs 1, 2, and 3, respectively.

The senior abstractor completed the quality rating form for each article; another project team member then reviewed the completed form for accuracy and completeness. The overall quality scores of these articles are recorded in the evidence tables (Appendix C); scores on each of the domains are provided in Chapters 3, 4, and 5. All graded studies were included in the analysis regardless of their quality score. However, evidence from studies graded as poor were given less weight in the qualitative and quantitative syntheses and discussion.

Strength of Overall Evidence

In addition to the individual studies, we also rated the strength of the collective evidence on each key question. We applied four separate criteria: (1) number of studies, (2) aggregate sample sizes over the studies, (3) quality of the individual studies, and (4) representativeness of the study populations in the studies.

External Peer Review

As is customary for all evidence reports and systematic reviews done for AHRQ, the RTI-UNC EPC requested review of this report from a wide array of outside experts in the field and from relevant professional societies and public organizations. AHRQ has also requested review from its own staff and appropriate federal agencies. We provide a list of the external peer reviewers in Appendix E. This report reflects substantive and editorial comments from this external peer review.

Chapter 3. Prevalence and Incidence of Perinatal Depression

Introduction

Perinatal depression is generally recognized to be a common affliction among women during pregnancy and the first postpartum year. However, estimates of the prevalence and incidence of the condition vary widely—from 5 percent to more than 25 percent of pregnant women and new mothers—depending on the assessment method, the timing of the assessment, and population characteristics.[21,22,27] To estimate disease burden more accurately and thereby better target and prioritize health care expenditures, we need more precise estimates of the prevalence and incidence of perinatal depression.

Two prior systematic reviews of the prevalence of perinatal depression—one for the early postpartum months and the other for pregnancy—are notable. O'Hara and Swain conducted the first meta-analysis of the prevalence of postpartum depression and investigated sources of variability in the prevalence estimates across studies.[21] The authors combined estimates from 59 studies in which depression had been assessed at least 2 weeks postpartum using either a clinical interview or a validated self-report measure with an established cutoff (i.e., Beck Depression Inventory [BDI] ≥ 10; Edinburgh Postnatal Depression Scale [EPDS] ≥ 13; Zung Depression Scale ≥ 48; Center for Epidemiological Studies—Depression [CES-D] scale ≥ 16). Based on a total sample of 12,810 postpartum women, they estimated the average prevalence of postpartum depression to be 13.0 percent, with a 95% confidence interval (CI) of 12.3 percent to 13.4 percent. They found that self-report measures yielded significantly higher estimates of postpartum depression than interview-based methods and that longer evaluation periods resulted in higher estimates. The number of days postpartum when the depression assessment was made and the country in which the study was conducted did not significantly affect the prevalence estimates in their analysis.

More recently, Bennett et al. conducted a meta-analysis of prevalence estimates for depression during pregnancy.[27] The authors combined estimates from 21 studies meeting predetermined inclusion criteria, including the assessment of depression by a structured clinical interview, the BDI, or the EPDS. Based on a total sample of 19,284 pregnant women, they estimated the prevalence of depression to be 7.4 percent (95% CI, 2.2 percent to 12.6 percent) during the first trimester, 12.8 percent (95% CI, 10.7 percent to 14.8 percent) during the second trimester, and 12.0 percent (95% CI, 7.4 percent to 16.7 percent) during the third trimester. The 95% CIs of these estimates overlap substantially, indicating that, given available evidence, the prevalence of depression during pregnancy cannot be said to differ significantly by trimester. The authors also found that, compared with structured clinical interviews, the self-report BDI produced significantly higher prevalence estimates, whereas the self-report EPDS produced statistically equivalent estimates.

Several factors point to the need for a reassessment of the prevalence of depression during pregnancy and the postpartum period at this time. First, the clinical definition of major depression has changed over time, becoming more precise. Definitions of major depression prior

Note: Appendixes and Evidence Tables cited in this report are provided electronically at http://www.ahrq.gov/clinic/epcindex.htm.

to the 1987 revision of the *Diagnostic and Statistical Manual of Mental Disorders, Third Edition* (DSM-III-R), were broader than subsequent definitions and likely included some minor depression and dysthymia. Minor depression is a proposed diagnosis for further study for which the 1994 DSM, fourth edition (DSM-IV), has defined research criteria;[28] however, it has not yet been added to the DSM-IV. Furthermore, DSM-IV has an even more precise definition of major depression, requiring a minimum number of depressive symptoms and functional impairment, whereas DSM-III-R required only counts of depressive symptoms. Most of the literature reviewed in the O'Hara and Swain study[21] (published in 1996) was published before 1994. Determining whether more recent studies affect the combined prevalence estimates and CIs is crucial to improving understanding of this disorder.

Most of the studies in Bennett et al.[27] (done in 2004) were published after 1994. Whether the combined prevalence estimates refer to major and minor depression together or major depression alone is not clear. The text of the article discusses major depression, but the tables clearly indicate the inclusion of minor depression.

Second, neither review distinguished between measures of the point prevalence, the percentage of the population with depression at a given point in time (e.g., at 24 weeks gestational age or 9 weeks postpartum), and measures of period prevalence, the percentage of the population with depression over a period of time (e.g., during pregnancy or from delivery to the end of the first 3 months postpartum). Both types of estimates are used for the single combined prevalence estimates, although O'Hara and Swain did test the effect of differing time points and durations for the depression assessment in a meta-regression.

Third, neither of the reviews presented evidence of the incidence of perinatal depression—the percentage of the population with depressive episodes that begin within a given period of time.

Fourth, overall prevalence estimates from both reviews are confounded by false positives because they included prevalence estimates from studies that assessed depression with self-report instruments. As mentioned above, both systematic reviews found that self-report instruments produce significantly higher prevalence estimates than do clinical interviews.

Finally, although both systematic reviews discussed prevalence estimates for women who were not pregnant and had not recently delivered a child, neither study rigorously reviewed the evidence that compares depression rates for women during pregnancy and the first postpartum year to the rates for women of a similar age during nonchildbearing times.

This chapter reviews the literature addressing Key Question (KQ) 1: What is the prevalence and incidence of depression (major and minor) during pregnancy and during the first year postpartum? Is the prevalence or incidence increased during pregnancy and the first postpartum year compared to nonchildbearing periods?

Methods

We abstracted study features and all estimates of the prevalence and incidence of major and minor depression together and of major depression alone from the 30 included studies found through our literature searches described in Chapter 2. During the abstraction process, we graded the quality of the study based on selected study features. We then analyzed the estimates using a variety of meta-analytic methods described in this section.

Evaluation of the Quality and Strength of the Evidence

Appendix B presents the quality rating form used for articles considered for KQ 1. The total possible score for these studies was 20 for studies without a comparison group and 25 for studies with a comparison group. For both types of studies, we considered those articles with a score of 16 or greater to be good, those with scores between 10 and 15 to be fair, and those with scores of 9 and below to be poor. The domains and maximum points possible for each domain are as follows:

- Reporting (domain score of 9): Eight items covering study aims, measures, patient populations, findings, and statistical presentation; each scored yes or no (1 or 0), except for an item concerning principal confounders that was scored yes, partially, or no (2, 1, or 0, respectively).

- External validity (domain score of 3): Three items relating to the representativeness of populations from which people were recruited and of settings and clinicians that treat such patients; each scored yes, no, or unable to determine (1, 0, or 0, respectively).

- Internal validity–bias (domain score of 3): Three items relating to issues such as validation of the depression diagnosis through clinical interview, follow-up periods, and appropriate statistical tests; each scored yes, no, or unable to determine (1, 0, or 0, respectively).

- Internal validity–confounding (domain score of 2 for studies without a comparison group and 4 for studies with a comparison group): Two items relating to sources of comparison groups, one for the adequacy of adjustments for confounding, and one for the handling of loss to follow-up; each scored yes, no, or unable to determine (1, 0, or 0, respectively).

- Precision (domain score of 3 for studies without a comparison group and 6 with a comparison group): One item relating to the number of pregnant or postpartum women assessed for depression, with scores of 3 for more than 1,000 women, 2 for 250 to 1,000 women, 1 for 30 to 250 women, and 0 for fewer than 30 women. For studies with a comparison group, a second item gave points based on the size of the smallest comparison group: a score of 3 for more than 2,000 women, 2 for 1,000 to 2,000 women, 1 for 500 to 1,000 women, and 0 for fewer than 500 women.

Best Estimates of Prevalence and Incidence

We abstracted all estimates of the prevalence and incidence of major and minor depression together and major depression alone. We distinguished prevalence estimates by whether they were point or period estimates and both prevalence and incidence estimates by the time period covered. Time periods for point prevalence estimates were defined as trimesters during pregnancy and months during the first postpartum year. Estimates taken at different weeks of gestation but within the same trimester of pregnancy were considered as being conducted in the same time period (e.g., estimates taken week 14 through week 27 of gestation were considered the second trimester). Similarly, estimates taken at different weeks postpartum but within the

same month postpartum were considered within the same time period (e.g., estimates taken during week 1 through week 4 postpartum would be considered month 1; week 5 through week 9 postpartum, month 2). Where we found two or more estimates within the same trimester of pregnancy or month postpartum, we used meta-analysis to obtain a combined estimate for that trimester or month. We then graphed the resulting estimates to determine how they changed throughout pregnancy and the first postpartum year.

We conducted similar procedures for period prevalence and incidence estimates. The relevant time periods were either single trimesters and months or multiple trimesters and months. Because we found fewer estimates of these types, however, we graphed period prevalence and incidence estimates for only the first 3 months postpartum.

We combined all estimates with the same diagnosis, estimate type, and time period using the *meta* command in Stata. This procedure uses the inverse-variance weighting method to calculate random effects summary estimates. It also produces (1) Q tests of the homogeneity of the estimates and (2) forest plots of the individual study and combined estimates and their CIs. To satisfy the normalcy assumptions of these methods, we first transformed the prevalence estimates into log odds estimates.

We reviewed the forest plots of the studies in each summary estimate to determine whether we could identify the source of any heterogeneity between studies. We then reran the meta-analyses excluding studies that were obvious outliers and for which we could identify the source of the bias. The new summary estimates are considered our best estimates of the prevalence and incidence of perinatal depression for the general female population in the United States.

Analysis of Confounders

To analyze associations between the prevalence of depression and study characteristics, we conducted cumulative meta-analysis and a series of meta-regressions on the point prevalence estimates for major and minor depression together and major depression alone. In the cumulative meta-analysis, we added studies one by one, based on publication year, to produce a new combined estimate with the cumulative evidence for each year. This procedure allowed us to see trends in the estimate over time. We conducted cumulative meta-analysis on the 2-month point prevalence estimates using the *metacum* command in Stata.

We then used the Stata *metareg* command to estimate several different meta-regression models. For all models, we used the log odds as the dependent variable and included the time point at which depression was assessed and indicators for whether the study enrolled only low-risk women and only women of low socioeconomic status (SES) as explanatory variables. The time point was represented by a categorical variable with included values for the first, second, and third trimesters and the first, second, and third months postpartum. The reference category for this variable was 4 to 12 months postpartum.

We estimated seven different models. Each had a different set of additional explanatory variables:

1. No additional explanatory variables;

2. Publication year;

3. Study country, categorized as the United States (the reference category), other western countries, and Asian countries;

4. Interview type, categorized as the Schedule for Affective Disorders and Schizophrenia (SADS) (the reference category), the Structured Clinical Interview for DSM Diagnoses (SCID), and other interview types;

5. Diagnostic criteria, categorized as Research Diagnostic Criteria (RDC) (the reference category), DSM III-R, DSM IV, and other criteria;

6. Whether depression was assessed only for women who were designated as at risk based on a screening instrument; and

7. The quality rating score.

Comparison with Other Women

To answer the second part of KQ 1, whether the prevalence and incidence of depression is higher during pregnancy and the first year postpartum compared to nonchildbearing periods, we computed odds ratios for studies with a comparison group of women of similar age during nonchildbearing times. Because the types and timing of prevalence and incidence estimates did not overlap in these studies, except for one time point, we did not conduct meta-analyses of the log odds ratios.

Results

We found 28 prospective studies and two retrospective studies that met our inclusion criteria. Only three of the prospective studies included a comparison group of nonpregnant women of similar age.[a] In this section, we first describe the study characteristics and then present our analysis of the study results.

Study Characteristics

The major characteristics of the 30 studies are summarized in Table 4 by study type and alphabetically within type. The 25 prospective studies without a comparison group are shown first,[19,23,29-51] followed by the three prospective studies with a comparison group,[20,52,53] and, finally, the two retrospective studies.[54,55] Important study characteristics include the precision or size of the studies, the representativeness of the study populations, the methods and timing used to assess the mother's mood, and the quality rating of study design. Each of these characteristics is addressed in turn below.

[a] Two studies assessed the mood of fathers in addition to the mothers.[30,45] We do not address the comparison of mothers and fathers in this chapter because it is beyond the scope of this study.

Precision. The study sample sizes ranged from 54 to 4,964 women; the median sample size was 202 women. Although all the studies had an adequate sample size to provide a prevalence estimate of 10 percent with 80 percent power at a 95% confidence level, most were not large enough to allow subgroup analyses.

The three studies with comparison groups included 313, 232, and 179 women in the comparison groups.[20,52,53] These sample sizes are inadequate to detect a difference as large as 5 percentage points in incidence or prevalence at 80 percent power and a 95% confidence level; a minimum sample size of more than 500 per group is required.

Table 4. Major characteristics of studies of prevalence and incidence of perinatal depression

Author, Year	Country	Sample Size	Who Interviewed	When Interviewed	Interview Type	Diagnostic Criteria
Prospective Cohort Studies without Comparison Groups						
Affonso et al., 1990[29]	US	202	All	Pregnancy & PP	SADS-PPG	RDC
Areias et al., 1996[30]	Portugal	54	All	Pregnancy & PP	SADS	RDC
Berle et al., 2003[31]	Norway	411	All EPDS ≥ 8 & some < 8	PP	MINI-V4.4/ MADRS	DSM-IV
Campbell and Cohn, 1991[32]	US	1,033	All	PP	SADS	RDC
Cooper et al., 1996[33]	England	4,964	EPDS ≥ 8	PP	SCID	DSM-III-R
Cox et al., 1982[34]	Scotland	105	All	PP	SPI	Pitt's
Garcia-Esteve et al., 2003[35]	Spain	1,123	All EPDS ≥ 9 & some < 9	PP	SCID-NP	DSM-IV
Gotlib et al., 1989[36]	Canada	295	All BDI ≥ 10 & some < 10	Pregnancy & PP	SADS	RDC
Hobfoll et al., 1995[37]	US	192	All	Pregnancy & PP	SADS	RDC
Kent et al., 1999[38]	Australia	710	GHQ28 > 4	PP	CIDI-A	DSM-III-R
Kitamura et al., 1993[39]	Japan	120	All	Pregnancy	SADS/ SADS-C	RDC
Kitamura et al., 1999[40]	Japan	111	All	Pregnancy & PP	SADS	RDC
Kumar and Robson, 1984[41]	England	196	All	Pregnancy & PP	SPI	RDC
Lee et al., 2001[42]	Hong Kong	781	All GHQ > 4 & some ≤ 4	PP	Modified SCID	Modified DSM-III-R
Lee et al., 2001[43]	Hong Kong	145	All	PP	Modified SCID	Modified DSM-III-R
Lucas et al., 2001[44]	Spain	641	BDI > 21	PP	Not specified	DSM-III-R
Matthey et al., 2003[45]	Australia	408	All	PP	DIS	DSM-IV
Murray and Cox, 1990[46]	England	100	All	Pregnancy	SPI	RDC
O'Hara et al., 1984[19]	US	99	All	Pregnancy & PP	SADS	RDC
Pop et al., 1993[47]	Netherlands	293	All	Pregnancy & PP	Not specified	RDC
Watson et al., 1984[48]	England	128	All	Pregnancy & PP	SPI	ICD-9

Table 4. Major characteristics of studies of prevalence and incidence of perinatal depression (continued)

Author, Year	Country	Sample Size	Who Interviewed	When Interviewed	Interview Type	Diagnostic Criteria
Whiffen, 1988[49]	Canada	115	All	PP	SADS	RDC
Yamashita et al., 2000[50]	Japan	88	All	PP	SADS	RDC
Yonkers et al., 2001[23]	US	802	All IDS ≥ 18 or EPDS ≥ 12 & some < 12	PP	SCID	DSM-IV
Yoshida et al., 1997[51]	England	98	All	PP	SADS	RDC
Prospective Studies with Comparison Groups						
Cooper et al., 1988[52]	England	483 cases 313 controls	All GHQ ≥ 12 & some < 12	PP	PSE/ MADRS	PSE ID/ Catego Class
Cox et al., 1993[20]	England	232 cases 232 controls	All EPDS ≥ 9 & some < 9	PP	SPI	RDC
O'Hara et al., 1990[53]	US	182 cases 179 controls	All	Pregnancy & PP	SADS	RDC
Retrospective Studies						
Bryan et al., 1999[54]	US	403	—	PP	Medical records	Diagnosis of 2 or more symptoms
Georgiopoulos et al., 2001[55]	US	342	—	PP	Medical records	Diagnosis

BDI, Beck Depression Inventory; CIDI-A, Composite International Diagnostic Interview; DIS, Diagnostic Inventory Schedule; DSM-III-R, Diagnostic and Statistical Manual of Mental Disorders, Third Edition, Revised; DSM-IV, Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition; EPDS, Edinburgh Postnatal Depression Scale; GHQ, General Health Questionnaire; ICD-9, International Classification of Diseases, Ninth Edition; MADRS, Montgomery-Asburg Depression Rating Scale; MINI-V4.4, Mini International Neuropsychiatric Interview, Version 4.4; PP, postpartum; PSE, Present State Examination; PSE ID, PSE Index of Definition; RDC, Research Diagnostic Criteria; SADS, Schedule for Affective Disorders and Schizophrenia; SADS-C, SADS Change Version; SADS-PPG, SADS-Pregnancy and Postpartum Guidelines; SCID, Structured Clinical Interview for DSM-III-R; SCID-NP, Structured Clinical Interview for DSM-III-R, Non-Patient Version; SPI, Standardized Psychiatric Interview.

Representativeness. Included studies represented a wide array of developed nations, but the study subjects were not a good representation of the racial and ethnic mix of the US population (Table 4). Seven of the prospective studies were located in England; six in the United States; three in Japan; two each in Canada, Australia, and Hong Kong; and one each in the Netherlands, Norway, Portugal, and Scotland. The two retrospective studies investigated depression diagnoses documented in the Olmsted County, Minnesota, population-based databases during two different 12-month periods—the 12 months following deliveries occurring in 1993 and the 12 months following deliveries among women visiting the Olmsted County and Mayo clinics in 1997 and 1998.[54,55]

None of the studies was designed to compare rates of depression among women of different racial and ethnic groups. Sixteen of the 30 studies did not even specify the racial and ethnic composition of the study subjects. Among the other 14 studies, 5 included only white non-Hispanic women;[19,32,38,44,47] two studies included only Chinese women;[42,43] and two others included only Japanese women.[50,51] The remaining five studies noted a racially mixed population, but all had a predominant race or ethnicity. In four of these studies, 73 percent to 90

percent of the women were white non-Hispanic,[29,36,37,48] and, in the fourth, 75 percent were Hispanic.[23]

Depression Assessment. Our inclusion criteria required that the study use a clinical interview or assessment to validate depression diagnoses. The prospective studies differed in who received a clinical interview, the interview instrument, the diagnostic criteria used to identify a depressive episode from the interview responses, and when the interview was conducted. These differences can affect the resulting estimates of prevalence and incidence.

Eighteen of the 28 prospective studies conducted a clinical interview on all study women. The remaining 10 studies first had study subjects complete a self-report depression screening instrument, such as the EPDS, the BDI, or the General Health Questionnaire (GHQ), a broader measure designed to assess the presence of psychiatric distress related to general medical illness. These studies then administered a clinical interview to women scoring over a predetermined cutoff on the screening instrument. Seven of the 10 studies also interviewed a small sample (e.g., 10 percent) of the women scoring below the cutoff, but few of the studies used the results from these interviews to adjust the final prevalence estimates for false negatives. Most studies used low enough cutoff scores that the resulting downward bias in the estimates was minimal. The one exception was the Lucas et al. study, which used a high cutoff of 21 on the BDI and did not interview any women scoring below the cutoff or adjust the resulting prevalence rates in any way, thereby introducing a significant, uncorrected downward bias.[44]

Different interview instruments have been developed for identifying depression diagnoses. These different instruments use different criteria for diagnosing depression. Little is known about how these different instruments and diagnostic criteria affect the prevalence and incidence estimates.

The most frequently used instrument among our studies was the SADS. This semistructured interview is widely used in clinical research and has well-established reliability and validity.[56] O'Hara et al. adapted the SADS for use with pregnant and postpartum women.[19] Twelve of the 28 prospective studies used this interview instrument.

Five of the studies used the section of the SCID that covers depressive disorders.[57,58] The SCID allows the interviewer to use additional questions to inquire about idioms of distress that are specific to the local context. Lee et al. used this feature of the SCID to incorporate questions about traditional Chinese customs used during the puerperium that may affect the clinical presentation of postpartum depression.[42,43] They also modified the instrument to identify cases of minor depression.

Five other studies used the Standardized Psychiatric Interview (SPI) of Goldberg et al.[59] The SPI includes 10 five-point scales that rate the severity of neurotic symptoms in the 7 days preceding the interview and a rating of 12 abnormalities observed during the interview.

Other interview instruments used include the Composite International Diagnostic Interview (CIDI-A),[60] the Diagnostic Interview Schedule,[61] the Mini International Neuropsychiatric Interview (MINI-V4.4),[62] the Present State Examination (PSE),[63] and the Montgomery and Asberg Depression Rating Scale (MADRS).[64]

All studies that used the SADS and three of the studies that used the SPI based depression diagnosis on the RDC.[65] To be diagnosed with depression, women had to have reported that they felt sad, tearful, or blue for at least 2 weeks. The 2-week criterion serves to rule out women who were experiencing postpartum blues only. In addition, for a diagnosis of major depression, the women had to have reported at least three or four additional symptoms, such as sleeping

disturbances, loss of appetite, fatigue, loss of interest in usual activities or the ability to concentrate, psychomotor retardation, and suicidal thoughts. Women with only two to four of these symptoms were classified as having minor depression. The RDC attempts to differentiate between normal physical effects of pregnancy and the puerperium and actual symptoms of depression.

Five of the prospective studies based diagnoses of depression on DSM-III-R criteria and four based diagnoses on DSM-IV criteria. A diagnosis of major depression based on the DSM-III-R criteria is comparable with the RDC for definite major depression.[66] However, the RDC includes criteria for minor depression, which, as mentioned above, received its first DSM mention in the fourth edition (DSM-IV)[28] as a proposed category for further study. Other criteria used for diagnoses of depression included Pitt's criteria;[67] the International Classification of Diseases, Ninth Edition (ICD-9); and PSE Index of Definition (PSE ID) and Catego Class.[63]

Finally, because the prevalence and incidence of depression may not be constant throughout pregnancy and the first postpartum year, the timing of the clinical interview is also very important. Most of the studies we reviewed administered the clinical interview at multiple points in time throughout pregnancy and the first postpartum year, allowing for multiple estimates of prevalence and incidence. The 28 prospective studies provided 80 estimates of the prevalence and incidence of major and minor depression and 70 estimates of the prevalence and incidence of major depression alone. Clinical assessments of depression were taken at different points in time throughout pregnancy and the first postpartum year. Graphical presentations of the timing of each of the estimates by diagnosis and estimate type are shown in Figures 3 through 8 as follows:

- Point prevalence estimates
  - 42 for major and minor depression (Figure 3)
  - 46 for major depression alone (Figure 4)

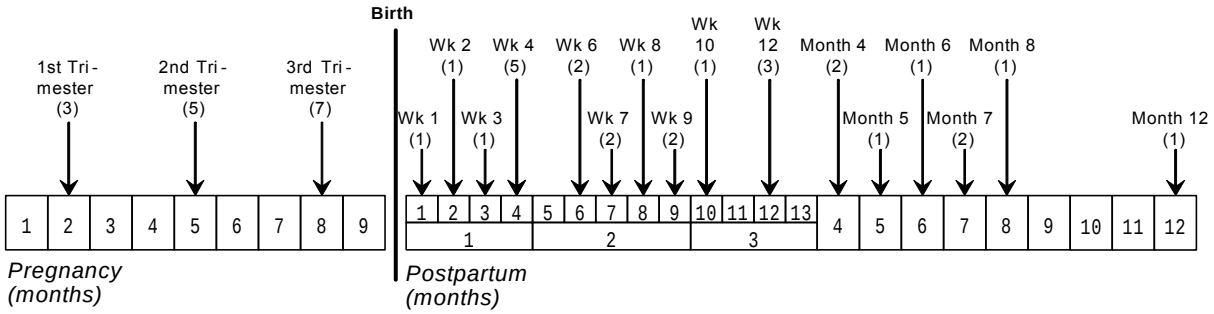
- Period prevalence estimates
  - 17 for major and minor depression (Figure 5)
  - 12 for major depression alone (Figure 6)

- Incidence estimates
  - 21 for major and minor depression (Figure 7)
  - 12 for major depression alone (Figure 8).

The numbers in parentheses in these figures are the number of estimates found in the 28 studies for that point or period of time.

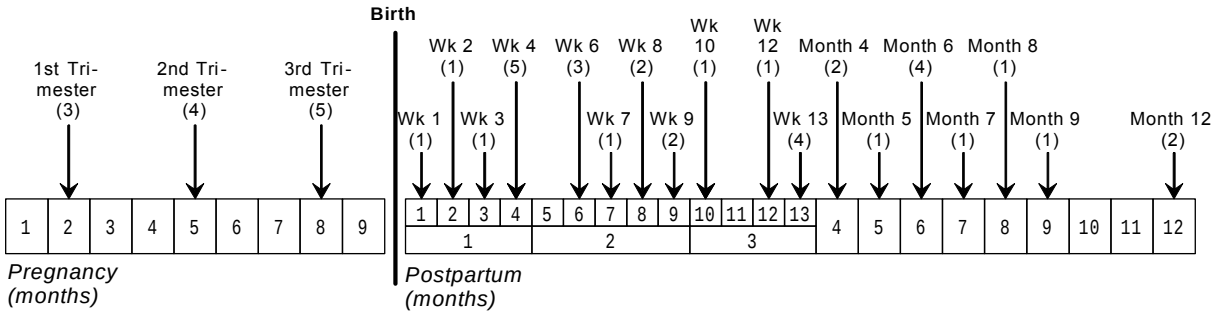
For the two retrospective studies, the investigators had abstracted information on symptoms and diagnoses of depression from medical records beginning at delivery and extending to 1 year postpartum. Both studies provided only estimates of 1-year period prevalence. Bryan et al.[54] provided estimates of the prevalence for both major and minor depression and major depression alone, whereas Georgiopoulos et al.[55] provided only the prevalence of major depression alone. Bryan et al. identified a woman as having postpartum depression if any of the following criteria were found in her medical records:[54] (1) two notations at least 2 weeks apart of symptoms of depression; (2) a documented diagnosis of depression by a physician, psychologist, nurse practitioner, or midwife; (3) a new prescription for an antidepressant with no evidence that it was for chronic pain or for any indication other than depression; and (4) documentation of symptoms

Figure 3. Estimates of point prevalence of major and minor depression by time of assessment



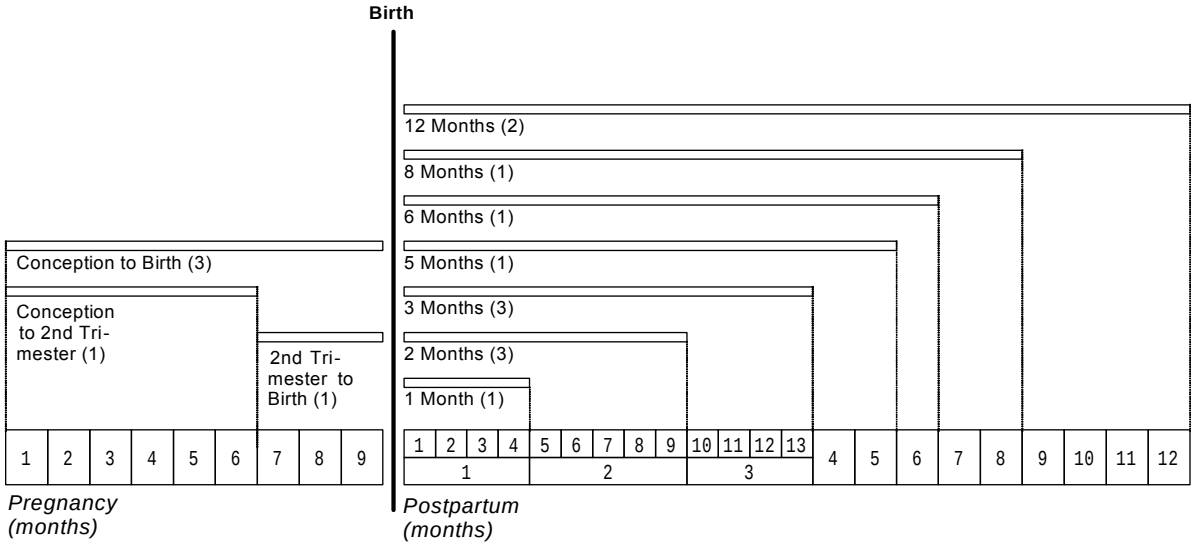
Note: The number of estimates at each point in time is shown in parentheses.

Figure 4. Estimates of point prevalence of major depression by time of assessment



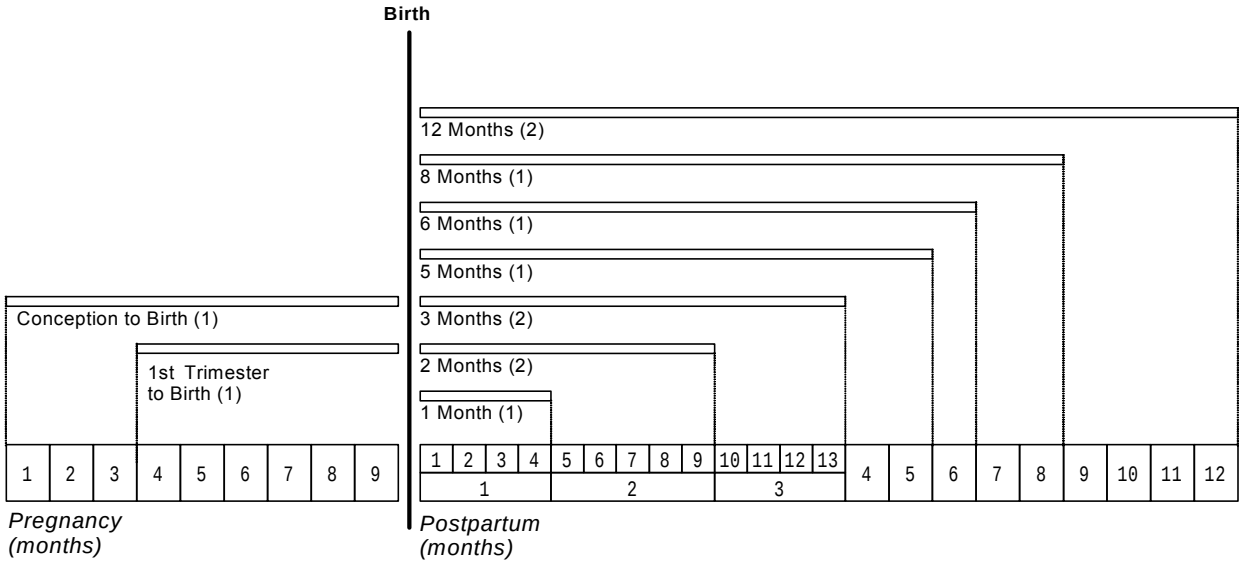
Note: The number of estimates at each point in time is shown in parentheses.

Figure 5. Estimates of period prevalence of major and minor depression by time period of assessment



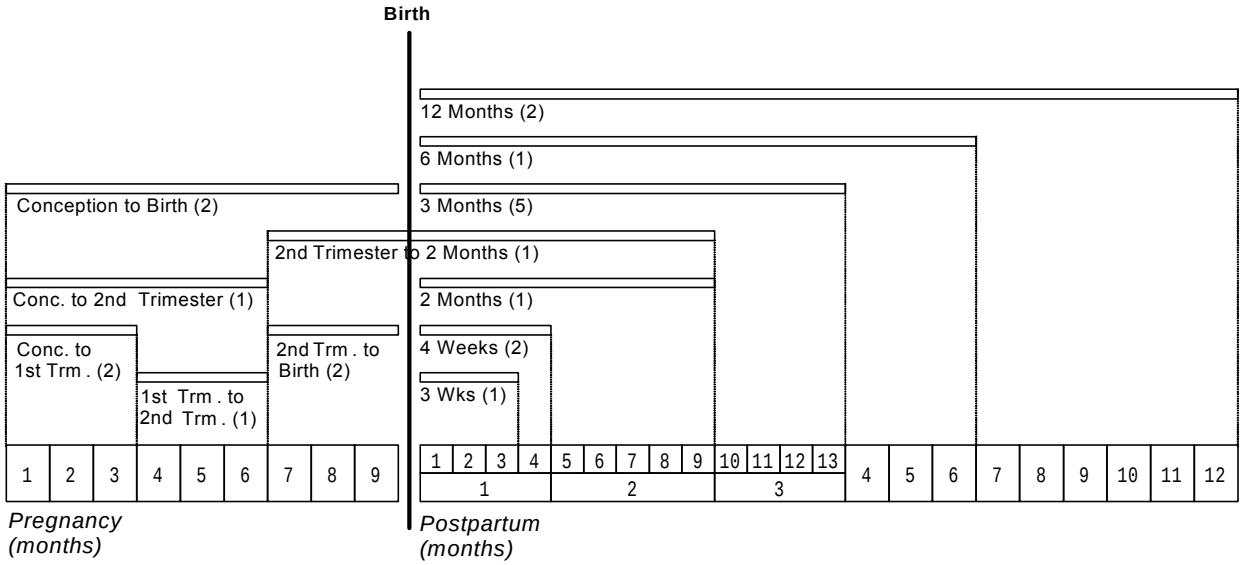
Note: The number of estimates for each period in time is shown in parentheses.

Figure 6. Estimates of period prevalence of major depression by time period of assessment



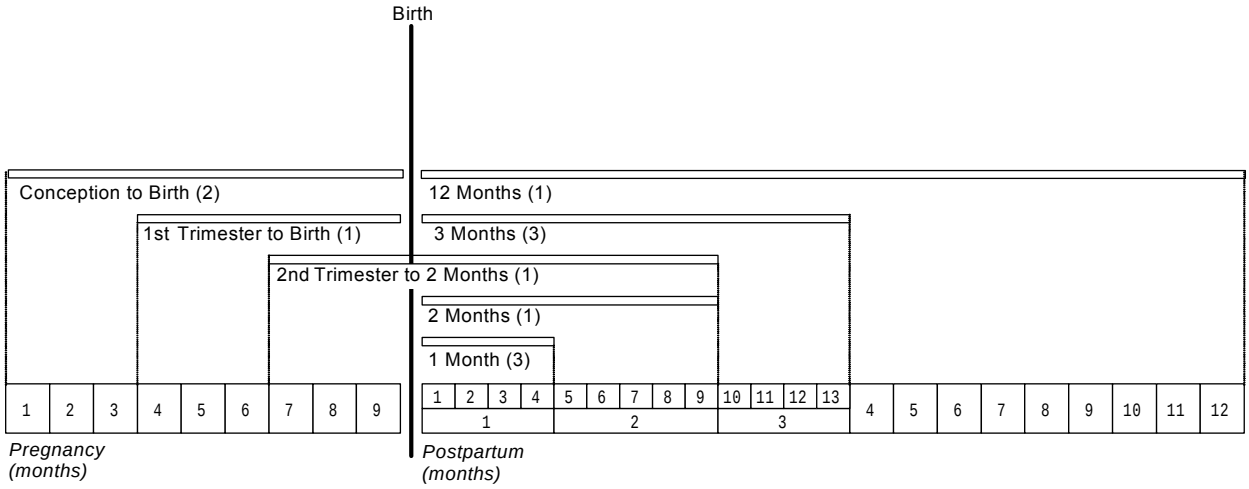
Note: The number of estimates for each period in time is shown in parentheses.

Figure 7. Estimates of incidence of major and minor depression by time period of assessment



Note: The number of estimates for each period in time is shown in parentheses.

Figure 8. Estimates of incidence of major depression by time period of assessment



Note: The number of estimates for each period in time is shown in parentheses.

sufficient to meet the DSM-IV criteria of major depression. Georgiopoulos et al.[55] based their prevalence estimate solely on a documented diagnosis of postpartum depression.

Quality Rating. We show the results of the quality rating of the included articles in Table 5 by study type. Studies were rated on reporting completeness, external validity, internal validity, and precision. The average overall quality rating score, out of a possible 20 points, was 11.1 for prospective studies without comparison groups and 12.0 for the retrospective studies. For prospective studies with comparison groups, which had 25 possible points, the average overall quality score was 11.7. Thus, we would rate the overall body of evidence for the prevalence and incidence of perinatal depression as fair at best.

In general, studies ranked good on reporting. The 28 prospective studies, both those with and those without comparison groups, scored an average of 6.0 out of 9 possible points for reporting. The retrospective studies scored 5.0 on average. Most studies clearly described the purpose of the study, the method of assessing depression, the characteristics of the patients in the study, and the study findings. Most studies also provided adequate information to estimate the random variability in the estimates and reported actual probability values for the statistical significance of the main outcomes. Fewer studies provided the distribution of the major principal confounders and described the characteristics of patients lost to follow-up. In particular, studies often did not discuss whether the women had prior depressive episodes or obstetrical complications and frequently did not report the women's socioeconomic status or race and ethnicity. Most studies also did not specifically exclude cases of bipolar disorder or psychosis.

Virtually all prospective studies rated poor on external validity. Prospective studies without a comparison group averaged 0.8 points out of 3 possible points; those with a comparison group averaged 0.3 points. These studies seldom supplied adequate information to determine whether study subjects were representative of the patient population of the facilities from which they were recruited and whether the recruitment facilities were representative of the facilities frequented by the general population in the geographic area. In contrast, the two retrospective studies, which were conducted using the Olmsted County Health Department and Mayo Clinic databases, included the majority of all newly delivered women in the county and therefore scored an average of 2.5 points on external validity.

Table 5. Quality rating of studies of the prevalence and incidence of perinatal depression

Author, Year	Reporting (9)	External Validity (3)	Internal Validity–Bias (3)	Internal Validity–Confounding (2)	Precision (3)	Total Score (20)
Prospective Cohort Studies without Comparison Groups						
Affonso et al., 1990[29]	4	0	3	0	1	8
Areias et al., 1996[30]	8	0	2	1	1	12
Berle et al., 2003[31]	5	0	2	0	2	9
Campbell and Cohn, 1991[32]	6	0	3	0	3	12
Cooper et al., 1996[33]	7	0	2	0	3	12
Cox et al., 1982[34]	5	1	3	1	1	11
Garcia-Esteve et al., 2003[35]	7	0	2	1	3	13
Gotlib et al., 1989[36]	5	2	1	1	2	11
Hobfoll et al., 1995[37]	6	3	2	0	1	12

Author, Year	Reporting (9)	External Validity (3)	Internal Validity–Bias (3)	Internal Validity–Confounding (2)	Precision (3)	Total Score (20)
Prospective Cohort Studies without Comparison Groups						
Kent et al., 1999[38]	7	1	2	0	2	12
Kitamura et al., 1993[39]	8	0	3	1	1	13
Kitamura et al., 1999[40]	4	1	3	1	1	10
Kumar and Robson, 1984[41]	7	0	3	0	1	11
Lee et al., 2001[42]	6	2	2	1	1	12
Lee et al., 2001[43]	5	2	0	0	1	8
Lucas et al., 2001[44]	5	0	2	0	2	9
Matthey et al., 2003[45]	6	0	3	0	2	11
Murray and Cox, 1990[46]	6	0	3	0	1	10
O'Hara et al., 1984[19]	6	0	3	0	1	10
Pop et al., 1993[47]	7	1	3	0	2	13
Watson et al., 1984[48]	7	2	3	0	1	13
Whiffen, 1988[49]	6	0	3	0	1	10
Yamashita et al., 2000[50]	6	0	3	0	1	10
Yonkers et al., 2001[23]	8	0	2	2	2	14
Yoshida et al., 1997[51]	7	0	3	0	1	11
Average	6.0	0.6	2.4	0.4	1.5	11.1
Retrospective Studies						
Bryan et al., 1999[54]	8	3	2	1	2	16
Georgiopoulos et al., 2001[55]	2	2	1	1	2	8
Average	5.0	2.5	1.5	1.0	2.0	12.0
Prospective Studies with Comparison Groups						
Cooper et al., 1988[52]	6	0	2	0	2	10
Cox et al., 1993[20]	5	1	2	3	1	12
O'Hara et al., 1990[53]	7	0	3	2	1	13
Average	6.0	0.3	2.3	1.7	1.3	11.7

Note: Numbers in parentheses are total possible points.

We separated scores for internal validity into two sets of study design characteristics: those that may bias the prevalence estimates and those that reflect possible confounding factors, which relate to the comparability of the comparison groups and whether losses of patients to follow-up were taken into consideration. The prospective studies scored high on the first measure of internal validity; the studies without a comparison group averaged 2.4 of 3 points and the studies with a comparison group averaged 2.3 points. Virtually all prospective studies assessed the mood of study women within 2 weeks of designated times during pregnancy and postpartum and applied appropriate statistical tests for measuring incidence or prevalence. However, as noted above, 10 studies introduced potential bias by not administering the clinical interview to all study women.

The retrospective studies averaged a lower 1.5 points. Diagnoses were not validated through clinical interview for all women, and Georgiopoulos et al. did not provide adequate information

to determine whether they used appropriate statistical techniques to compute the prevalence estimate.[55]

Studies with comparison groups could get 4 possible points for the internal validity confounding score. We awarded 2 additional points if the cases and controls were recruited from the same population and over the same period of time. Only two of the three prospective studies with comparison groups met these criteria. The comparison group in the Cooper et al. study comprised women interviewed by another researcher over a different time period in a different city. Study women were recruited from the appointments diary of the prenatal clinic and the delivery booking diary of the general practitioner unit of the John Radcliffe Hospital in Oxford; the comparison group was derived from a community sample of Edinburgh women of similar age but who were not pregnant and had not delivered in the previous 12 months.[52]

By contrast, in the Cox et al. study, both cases and controls resided in the North Staffordshire Health District.[20] Cases were recruited from the prenatal clinic lists of the North Staffordshire Maternity Hospital; controls matching cases on marital status, number of children, and age (within 5 years) were recruited from four general practice registers. The O'Hara et al. study recruited cases from a public obstetrics and gynecology clinic and two private practices at the University of Iowa Hospitals and Clinics.[53] Each subject was asked to provide the names of five acquaintances similar in age, marital status, work status, and number of children. The acquaintance most similar to the subject was selected as a control.

We also gave points for the internal validity confounding measure if the investigators made adjustments or discussed the possible direction and magnitude of any biases from confounding factors and if they took the loss of patients to follow-up into account in their prevalence or incidence estimate. A minority of studies met either of these criteria, resulting in an average score on this measure of 0.4 out of 4 possible points for prospective studies without comparison groups, 1.7 for prospective studies with comparison groups, and 1.0 for the retrospective studies.

Finally, we gave 17 studies with 30 to 250 pregnant or recently delivered women a precision score of 1, 10 studies with 250 to 1,000 women a precision score of 2, and 3 studies with more than 1,000 women a precision score of 3. None of the studies had a comparison group of at least 500 women; therefore, we awarded no additional points for precision. The average precision score was 1.5 for prospective studies without comparison groups, 1.3 for prospective studies with comparison groups, and 2.0 for the retrospective studies.

In summary, the included studies generally were rated as good on reporting and internal validity for bias, poor on external validity and internal validity for confounding, and only fair on precision.

Results from Prospective Studies

Our original estimates of point prevalence, period prevalence, and incidence rates computed from all of the estimates in the included studies are shown by time period in Table 6 for major and minor depression together and in Table 7 for major depression alone. For time periods for which we had more than one estimate, we show the combined estimate from the meta-analysis and the *P*-value for the Q test of homogeneity. This statistic tests the null hypothesis that the estimates come from the same distribution—that is, whether or not the studies appear statistically to measure the same phenomenon. A P-value < 0.05 suggests that they do not.

The results of these tests indicate that considerable heterogeneity exists across the studies included in many of the pooled estimates, particularly among the point prevalence estimates.

Therefore, we first discuss the results of our analysis of outliers and then discuss the results of the revised meta-analyses. We finish this section by presenting the findings from the studies with comparison groups of nonchildbearing women.

Outliers. In a review of the forest plots of the meta-analyses of the prevalence and incidence estimates, we found estimates from several studies consistently to be outliers for all time periods at which they assessed the women's mood. Two studies included only women at low risk of depression.[29,32] Affonso et al.[29] included only primigravida women with a viable fetus who were married or living with the infant's father and who had no recent depression episodes. Campbell and Cohn[32] included only primiparous women who delivered full-term, single infants without major complications and who were Caucasian, married, over 17 years of age, and had at least a high school education. The estimates from these studies were consistently lower than the estimates from the other studies.

Two additional studies included only women of lower socioeconomic status.[23,37] These studies generally provided higher estimates of depression prevalence and incidence than the other studies.

The Lucas et al. study included only women who screened positive for depression on the BDI.[44] The cutoff used (> 21) was so high that the bias from false negatives produced consistently lower prevalence estimates compared to the other studies.

Finally, because of its size, the Cooper et al. study dominated the combined 2-month point prevalence estimate for major depression alone.[33] However, the 15.3 percent estimated point prevalence from this study is outside the 95% CI of the combined estimate for major and minor depression. The purpose of the study was not to produce a prevalence estimate but rather to develop a predictive index for postpartum depression. Furthermore, many of the clinical

Table 6. Original estimates of prevalence and incidence of major and minor depression

Start Date	End Date	Studies	Estimate	95% Confidence Interval	*P*-Value for Test of Homogeneity
		Point Prevalence			
	1st trimester	29,40,41	6.4%	2.3%-16.2%	0.002
	2nd trimester	19,36,37,41,53	11.0%	5.7%-20.4%	0.000
	3rd trimester	29,36,37,40,41,46,47	8.7%	4.9%-15.0%	0.000
	1 week PP	40	5.5%	1.8%-12.4%	
	1 month PP	23,29,36,40,42,47,50	8.8%	6.4%-11.9%	0.002
	2 months PP	31,32,35,37,43,49,53	11.3%	7.7%-16.2%	0.000
	3 months PP	41,42,47,50	12.9%	10.6%-15.8%	0.707
	4 months PP	29,47	4.3%	0.6%-25.4%	0.001
	5 months PP	47	10.6%	7.3%-14.7%	
	6 months PP	20	9.9%	6.4%-14.5%	
	7 months PP	41,47	10.6%	7.1%-15.6%	0.180
	8 months PP	47	6.5%	4.0%-9.9%	
	12 months PP	41	6.5%	2.7%-12.9%	
		Period Prevalence			
Conception	2nd trimester	30	9.3%	3.1%-20.3%	
Conception	Birth	30,39,41	18.4%	14.3%-23.3%	0.931
2nd trimester	3rd trimester	36	10.2%	7.0%-14.2%	
Birth	1 month PP	50	13.6%	7.3%-22.6%	
Birth	2 months PP	19,32,45	8.9%	6.8%-11.7%	0.135
Birth	3 months PP	30,50,51	19.2%	10.7%-31.9%	0.016
Birth	5 months PP	34	29.1%	20.6%-38.9%	
Birth	6 months PP	20	13.8%	9.6%-18.9%	
Birth	8 months PP	47	20.8%	16.3%-25.9%	
Birth	12 months PP	30	53.7%	39.6%-67.4%	
		Incidence			
Conception	1st trimester	39,41	11.3%	7.8%-16.3%	0.757
Conception	2nd trimester	30	5.8%	1.2%-16.0%	
Conception	Birth	30,39	14.5%	8.1%-24.4%	0.192
1st trimester	2nd trimester	41	2.7%	0.6%-7.6%	
2nd trimester	3rd trimester	36,41	2.2%	1.1%-4.1%	0.627
2nd trimester	2 months PP	37	12.5%	7.9%-18.5%	
Birth	1 month PP	36,42,50	7.8%	3.6%-16.1%	0.003
Birth	2 months PP	19	10.3%	5.1%-18.1%	
Birth	3 months PP	30,41,42,50,51	14.5%	10.9%-19.2%	0.142
Birth	6 months PP	20	11.1%	7.3%-16.0%	
Birth	12 months PP	30	49.0%	34.4%-63.7%	

Table 7. Original estimates of prevalence and incidence of major depression

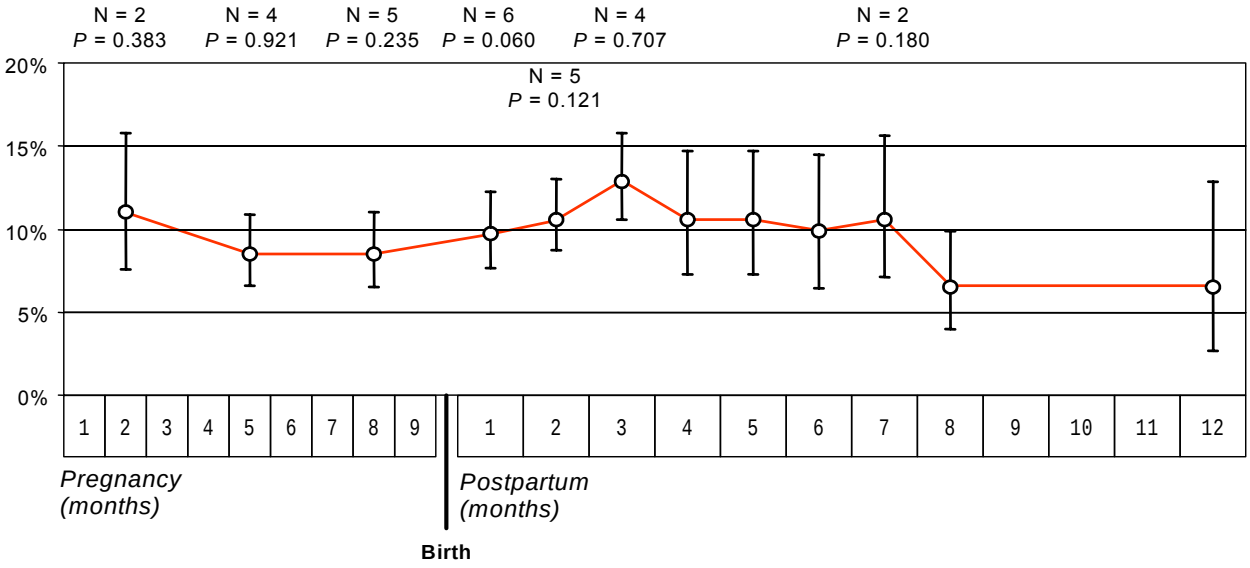
Start Date	End Date	Studies	Estimate	95% Confidence Interval	*P*-Value for Test of Homogeneity
		Point Prevalence			
	1st trimester	29,40,41	2.4%	0.7%-8.2%	0.032
	2nd trimester	19,37,48,53	6.4%	3.7%-11.0%	0.029
	3rd trimester	29,37,40,46,47	3.4%	1.8%-6.4%	0.116
	1 week PP	40	0.0%	0.0%-3.2%	
	1 month PP	23,42,44,50	2.8%	1.5%-5.5%	0.000
	2 months PP	31,33,35,37,42,48,49	6.8%	3.8%-11.9%	0.000
	3 months PP	42,44,50	3.8%	2.4%-6.1%	0.010
	4 months PP	29,47	2.3%	1.1%-4.9%	0.435
	5 months PP	47	2.1%	0.8%-4.4%	
	6 months PP	20,38,44,52	4.2%	2.1%-8.7%	0.000
	7 months PP	47	3.1%	1.4%-5.8%	
	8 months PP	47	1.0%	0.2%-3.0%	
	9 months PP	44	0.0%	0.0%-0.7%	
	12 months PP	44,52	1.3%	0.0%-56.6%	0.206
		Period Prevalence			
Conception	Birth	39	12.7%	7.1%-20.4%	
1st trimester	Birth	48	9.4%	4.9%-15.8%	
Birth	1 month PP	50	5.7%	1.9%-12.8%	
Birth	2 months PP	19,32	6.5%	5.2%-8.2%	0.516
Birth	3 months PP	50,51	7.1%	4.1%-11.7%	0.626
Birth	5 months PP	34	12.6%	6.9%-20.6%	
Birth	6 months PP	20	6.5%	3.7%-10.4%	
Birth	8 months PP	47	6.8%	4.2%-10.4%	
Birth	12 months PP	44,48	6.6%	0.5%-51.7%	0.000
		Incidence			
Conception	Birth	30,39,48	7.5%	3.8%-14.2%	0.116
2nd trimester	2 months PP	37	3.0%	1.0%-6.8%	
Birth	1 month PP	23,42,50	3.9%	2.9%-5.4%	0.429
Birth	2 months PP	48	8.1%	4.0%-14.4%	
Birth	3 months PP	42,50,51	6.5%	4.2%-9.6%	0.767
Birth	12 months PP	30	30.6%	18.3%-45.4%	

interviews were conducted by telephone and the article did not state whether a clinician or lay person conducted the interview. Thus, the procedures for assessing depression in this study may have introduced significant bias in the prevalence estimate.

We reran the meta-analyses excluding these six studies to produce "best estimates" of the prevalence of perinatal depression. The final best estimates are shown in Table 8 for major and minor depression together and in Table 9 for major depression alone.

Point Prevalence. We show the best estimates for the point prevalence of major and minor depression graphically in Figure 9. This figure graphs the mean estimate and corresponding 95% CI for each trimester of pregnancy and month postpartum in the first year following delivery. The number of studies that we used to compute the estimate and the *P*-value for the Q test of homogeneity among the studies are shown above each estimate. For points in time for which no numbers are shown, we found only a single estimate.

Figure 9. Best estimates of point prevalence of major and minor depression



Note: For times with an estimate from a single study, no N or *P*-value is shown.
N = number of studies on which the combined estimate is based.
P = *P*-value for the Q test of homogeneity.

As shown in Figure 9, prevalence in the first trimester is 11.0 percent but drops to 8.5 percent in the second and third trimesters. Following delivery, prevalence of major and minor depression begins to rise and is highest in the third month at 12.9 percent. In the fourth through seventh month postpartum, prevalence declines slightly, staying in the range of 9.9 percent to 10.6 percent, after which it declines to 6.5 percent. However, all of these estimates have broad 95% CIs, suggesting that a considerable amount of uncertainty remains in the precise values of the estimates and that the differences in the estimates over time may be attributed to chance or to uncontrolled factors. We cannot say with certainty from these data that perinatal depression is higher at any particular trimester during pregnancy or month in the first postpartum year.

Table 8. Best estimates of prevalence and incidence of major and minor depression

Start Date	End Date	Studies	Estimate	95% Confidence Interval	*P*-Value for Test of Homogeneity
		Point Prevalence			
	1st trimester	40,41	11.0%	7.6%-15.8%	0.383
	2nd trimester	19,36,41,53	8.5%	6.6%-10.9%	0.921
	3rd trimester	36,40,41,46,47	8.5%	6.5%-11.0%	0.235
	1 week PP	40	5.5%	1.8%-12.4%	
	1 month PP	23,36,40,42,47,50	9.7%	7.7%-12.3%	0.060
	2 months PP	31,35,43,49,53	10.6%	8.7%-13.0%	0.121
	3 months PP	41,42,47,50	12.9%	10.6%-15.8%	0.707
	4 months PP	47	10.6%	7.3%-14.7%	
	5 months PP	47	10.6%	7.3%-14.7%	
	6 months PP	20	9.9%	6.4%-14.5%	
	7 months PP	41,47	10.6%	7.1%-15.6%	0.180
	8 months PP	47	6.5%	4.0%-9.9%	
	12 months PP	41	6.5%	2.7%-12.9%	
		Period Prevalence			
Conception	2nd trimester	30	9.3%	3.1%-20.3%	
Conception	Birth	30,39,41	18.4%	14.3%-23.3%	0.931
2nd trimester	3rd trimester	36	10.2%	7.0%-14.2%	
Birth	1 month PP	50	13.6%	7.3%-22.6%	
Birth	2 months PP	19,45	9.6%	8.0%-11.4%	0.362
Birth	3 months PP	30,50,51	19.2%	10.7%-31.9%	0.016
Birth	5 months PP	34	29.1%	20.6%-38.9%	
Birth	6 months PP	20	13.8%	9.6%-18.9%	
Birth	8 months PP	47	20.8%	16.3%-25.9%	
Birth	12 months PP	30	53.7%	39.6%-67.4%	
		Incidence			
Conception	1st trimester	39,41	11.3%	7.8%-16.3%	0.757
Conception	2nd trimester	30	5.8%	1.2%-16.0%	
Conception	Birth	30,39	14.5%	8.1%-24.4%	0.192
1st trimester	2nd trimester	41	2.7%	0.6%-7.6%	
2nd trimester	3rd trimester	36,41	2.2%	1.1%-4.1%	0.627
Birth	1 month PP	36,42,50	7.8%	3.6%-16.1%	0.003
Birth	2 months PP	19	10.3%	5.1%-18.1%	
Birth	3 months PP	30,41,42,50,51	14.5%	10.9%-19.2%	0.142
Birth	6 months PP	20	11.1%	7.3%-16.0%	
Birth	12 months PP	30	49.0%	34.4%-63.7%	

NOTE: Best estimates reflect the single or combined estimate at each point or period of time remaining after estimates with obvious, identifiable biases have been dropped.
PP, postpartum.

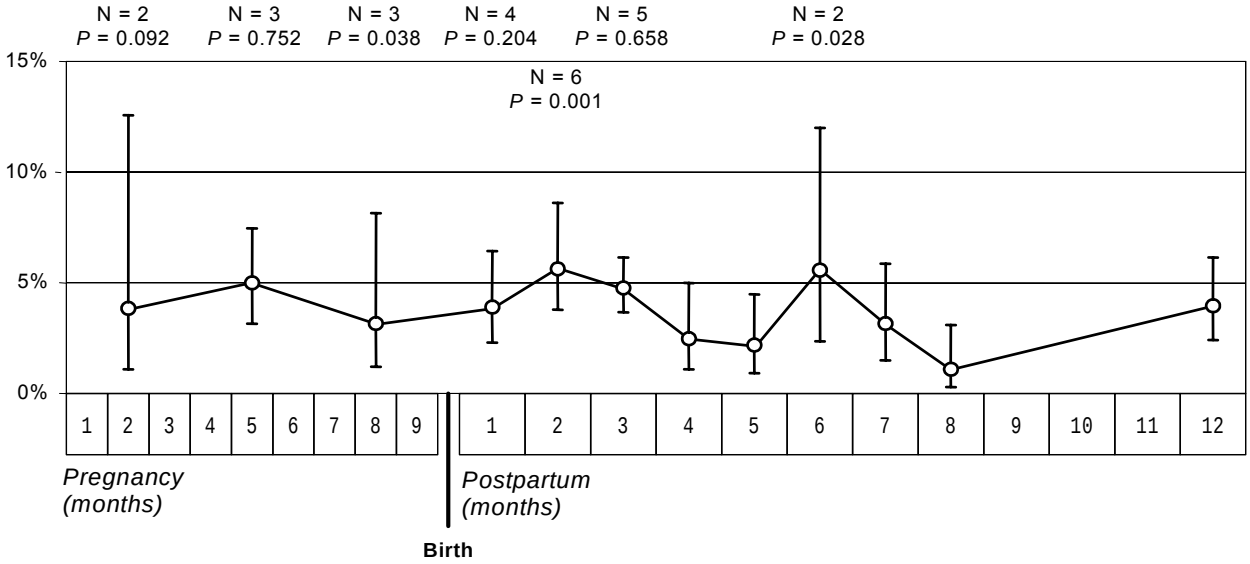
Table 9. Best estimates of prevalence and incidence of major depression

Start Date	End Date	Studies	Estimate	95% Confidence Interval	*P*-Value for Test of Homogeneity
		Point Prevalence			
	1st trimester	40,41	3.8%	1.0%-12.6%	0.092
	2nd trimester	19,48,53	4.9%	3.1%-7.4%	0.752
	3rd trimester	40,46,47	3.1%	1.1%-8.1%	0.038
	1 week PP	40	0.0%	0.0%-3.2%	
	1 month PP	40,42,47,50	3.8%	2.2%-6.4%	0.204
	2 months PP	31,35,43,48,49,53	5.7%	3.8%-8.7%	0.001
	3 months PP	41,42,47,50,52	4.7%	3.6%-6.1%	0.658
	4 months PP	47	2.4%	1.0%-4.9%	
	5 months PP	47	2.1%	0.8%-4.4%	
	6 months PP	20,52	5.6%	2.4%-12.1%	0.028
	7 months PP	47	3.1%	1.4%-5.8%	
	8 months PP	47	1.0%	0.2%-3.0%	
	12 months PP	52	3.9%	2.3%-6.1%	
		Period Prevalence			
Conception	Birth	39	12.7%	7.1%-20.4%	
1st trimester	Birth	48	9.4%	4.9%-15.8%	
Birth	1 month PP	50	5.7%	1.9%-12.8%	
Birth	2 months PP	19	8.1%	3.6%-15.3%	
Birth	3 months PP	50,51	7.1%	4.1%-11.7%	0.626
Birth	5 months PP	34	12.6%	6.9%-20.6%	
Birth	6 months PP	20	6.5%	3.7%-10.4%	
Birth	8 months PP	47	6.8%	4.2%-10.4%	
Birth	12 months PP	48	21.9%	15.1%-30.0%	
		Incidence			
Conception	Birth	30,39,48	7.5%	3.8%-14.2%	0.116
Birth	1 month PP	42,50	5.2%	3.1%-8.9%	0.819
Birth	2 months PP	48	8.1%	4.0%-14.4%	
Birth	3 months PP	42,50,51	6.5%	4.2%-9.6%	0.767
Birth	12 months PP	30	30.6%	18.3%-45.4%	

NOTE: Best estimates reflect the single or combined estimate at each point or period of time remaining after estimates with obvious, identifiable biases have been dropped.
PP, postpartum.

The best estimates for the point prevalence of major depression alone (Figure 10) are more variable and no more precise than those for major and minor depression together. Episodes of major depression comprise less than half of all cases of depression in the perinatal period, except during three seemingly peak times. As shown in Figure 10, the prevalence of major depression is highest in the second trimester (4.9 percent), 2 months postpartum (5.7 percent), and 6 months postpartum (5.6 percent). However, the 95% CIs for these estimates are very wide and overlap those at other times. Thus, we cannot say with certainty that major depression peaks at these points in time. Furthermore, the tests for homogeneity show that considerable heterogeneity persists among studies in the combined estimates.

Figure 10. Best estimates of point prevalence of major depression



ote: For times with an estimate from a single study, no N or *P*-value is shown.
N = number of studies on which the combined estimate is based.
P = *P*-value for the Q test of homogeneity.N

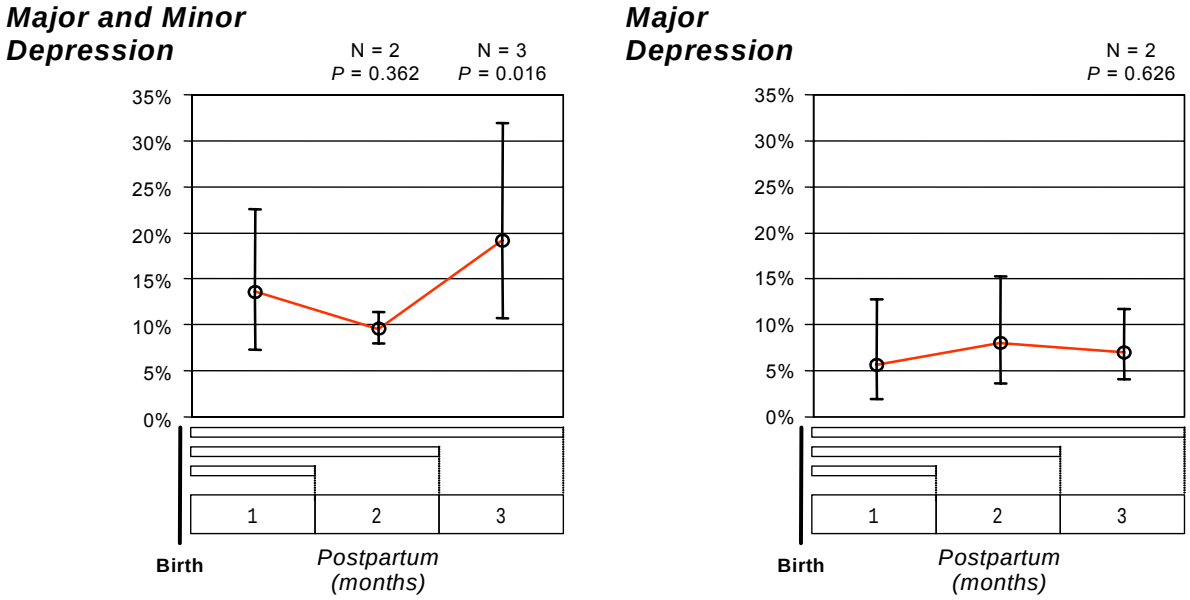
Period Prevalence. The many fewer estimates of period prevalence allow us to say little about the period prevalence for major and minor depression. As shown in Tables 8 and 9, the best estimates suggest that as many as 18.4 percent of pregnant women are depressed during their pregnancy (i.e., from conception to birth), with as many as 12.7 percent having an episode of major depression. Furthermore, as many as 19.2 percent of new mothers may have major or minor depression in the first 3 months following delivery (Table 8), with as many as 7.1 percent having major depression (Table 9).

However, all estimates have wide 95% CIs. Moreover, as shown in Figure 11, the best estimates of different durations are not consistent over longer periods of time. We would expect the period prevalence for major and minor depression from birth to 2 months postpartum to be higher than the period prevalence from birth to 1 month postpartum and the period prevalence

for major depression from birth to 3 months postpartum to be higher than the period prevalence from birth to 2 months postpartum, but we do not see these patterns.

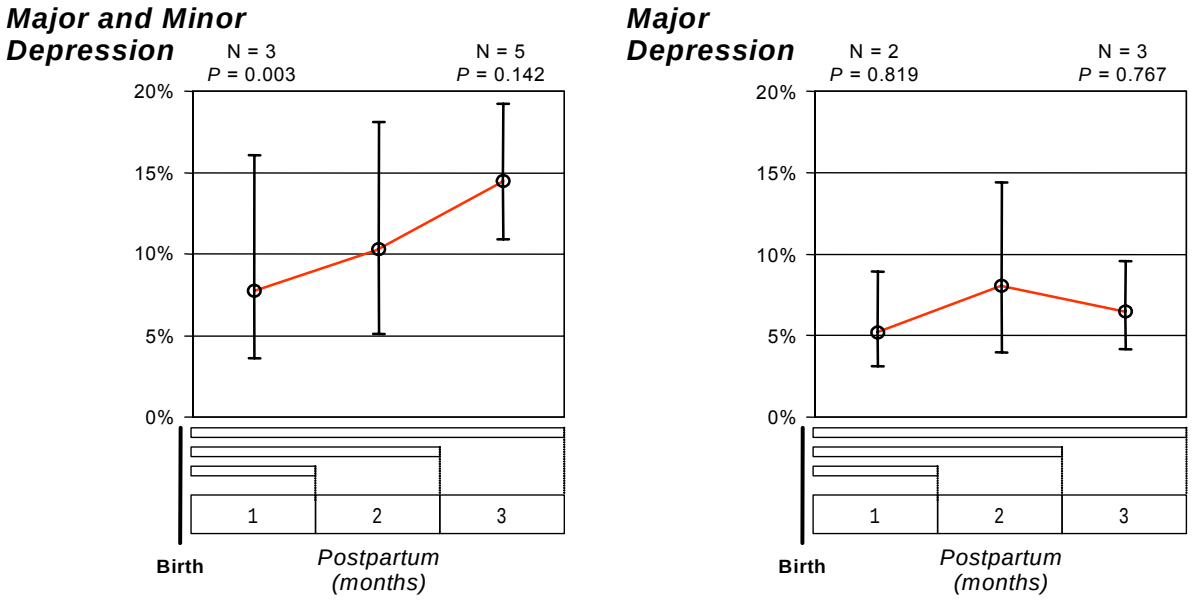
Incidence. We also found few estimates of the incidence of depression—the percentage of women with depressive episodes that begin during pregnancy or the first year postpartum. The studies we found suggest that as many as 14.5 percent of pregnant women have a new episode of major or minor depression during pregnancy, and 14.5 percent have a new episode during the first 3 months postpartum (Table 8). Considering major depression alone, 7.5 percent of women may have a new episode during pregnancy and 6.5 percent during the first 3 months after delivery (Table 9). Figure 12 shows that, although the incidence estimates for major and minor depression in the first 3 months postpartum follow the expected upward trend, the incidence estimates of major depression alone do not.

Figure 11. Best estimates of period prevalence of depression



Note: For times with an estimate from a single study, no N or *P*-value is shown.
N = number of studies on which the combined estimate is based.
P = *P*-value for the Q test of homogeneity.

Figure 12. Best estimates of incidence of major and minor depression

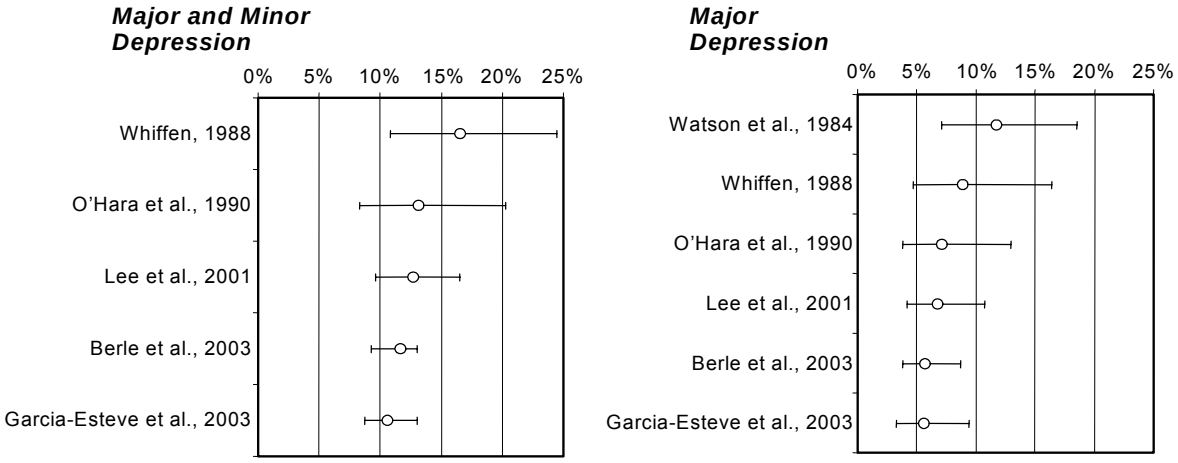


Note: For times with an estimate from a single study, no N or *P*-value is shown.
N = number of studies on which the combined estimate is based.
P = *P*-value for the Q test of homogeneity.

Analysis of Confounders

The results of the cumulative meta-analysis are graphed in Figure 13. They clearly show the impact of the more precise diagnostic criteria in more recent studies. For both major and minor depression together (left panel) and major depression alone (right panel), the cumulative combined 2-month point prevalence estimate drifts downward as more recent studies are added. Thus, the more precise criteria in the more recent studies identify fewer women as depressed. However, we did not find a statistically significant effect of the year of publication in our meta-regression.

Figure 13. Cumulative meta-analysis for point prevalence of depression at 2 months postpartum



We provide the estimated coefficients, their standard errors, and *P*-values from the different meta-regression models in Table 10 for major and minor depression together and in Table 11 for major depression alone. We have bolded coefficients significantly different from zero at the $\alpha = 0.05$ level. The results for major and minor depression show large, positive coefficients for the 2-month postpartum and 3-month postpartum time periods compared to the 4- to 12-month postpartum period (Table 10). These findings suggest a higher prevalence of depression during these 2 months. However, the coefficients are both significant only in the equation that includes diagnostic criteria (Model 5, Table 10). The 2-month postpartum time period is also large and positive for major depression alone, but significant only in the equations including diagnostic criteria (Model 5, Table 11) and whether only women who screened positive for depression were interviewed (Model 6, Table 11). None of the coefficients for the trimesters of pregnancy is statistically significant, suggesting that the prevalence of depression during pregnancy is similar to that during the last three quarters of the first postpartum year.

The low-risk indicator has a statistically significant, negative coefficient for both sets of diagnoses, as expected (Tables 10 and 11). Low SES has a statistically significant, positive coefficient only for major and minor depression together (Table 10). The latter result suggests that the prevalence of major depression is similar among SES groups but that minor depression may be more prevalent among lower SES groups.

The meta-regression results also suggest that prevalence can vary by the clinical instrument and diagnostic criteria used to assess depression. The SCID instrument defined fewer women with major and minor depression than did the SADS interview (Table 10), but the coefficient for this variable is not significant in the equation for major depression alone (Table 11), suggesting that the difference is in the identification of women with minor depression. DSM-IV and other diagnostic criteria (e.g., Pitt, ICD-9) defined fewer women as depressed than did the RDC in the equation for major and minor depression (Table 10), and DSM-III-R and other criteria defined significantly more women as suffering from major depression than did the RDC (Table 11).

Finally, studies with higher quality rating scores have lower log odds, suggesting lower prevalence of depression, but the coefficient of this variable is only marginally significant ($P = 0.072$) in the equation for major depression alone (Model 7, Table 11) and is not significant in the equation for major and minor depression together (Model 7, Table 10). No statistically

significant results were found for study country or whether the study interviewed only women who screened positive for depression, although the signs of the coefficients for these variables are as predicted.

Comparison with Other Women

The three prospective studies with comparison groups of women of similar age in nonchildbearing periods had adequate data to compute 13 estimates of the relative prevalence and incidence of depression. The estimated odds ratios and corresponding 95% CIs are shown in Table 12.

None of the odds ratios for prevalence, which covered different time periods in the first postpartum year, indicated a statistically significant difference. In addition, the National Comorbidity Survey fielded in 1990-1992 found a 5.9 percent current 30-day prevalence of major depressive episodes among women ages 15 to 54 years using the CIDI instrument and DSM-III-R criteria.[68] This finding is approximately equivalent to our best 1-month postpartum period prevalence of 5.7 percent and to the point prevalence at 2 months and 6 months postpartum (5.7 percent and 5.6 percent, respectively) shown in Table 9. Thus, the evidence indicates no difference in the prevalence of postpartum depression among pregnant or newly delivered women and women at other times in their childbearing years.

The single estimate of the incidence of major and minor depression (Table 12) shows a significant 3-fold difference in the odds of having a new episode of major or minor depression among women in their first 5 weeks postpartum compared to women who were not pregnant and had not recently given birth.[20] However, by 6 months postpartum, the difference in the incidence had narrowed and was no longer significant (Table 12).

Results from Retrospective Studies

The prevalence estimates from the retrospective studies measure something different than the prospective studies. In the prospective studies, all study women recruited from prenatal clinics or maternity wards were screened and interviewed for depression. Thus, all (or nearly all) women with depression in the populations so defined are identified. In the retrospective studies, only those women with depression detected through the course of medical contacts during the year were identified.

Table 10. Meta–regression results for log odds of a diagnosis of major and minor depression

Explanatory Variables	Model						
	1	2	3	4	5	6	7
Constant	–2.291 (0.149) ***P* = 0.000**	–2.159 (0.174) ***P* = 0.000**	–2.564 (0.339) ***P* = 0.000**	–2.189 (0.218) ***P* = 0.000**	–2.273 (0.098) ***P* = 0.000**	–2.276 (0.151) ***P* = 0.000**	–2.125 (0.626) ***P* = 0.001**
1st trimester vs. 4 to 12 months PP	0.065 (0.322) *P* = 0.840	0.068 (0.316) *P* = 0.830	0.064 (0.337) *P* = 0.850	0.032 (0.310) *P* = 0.917	0.064 (0.238) *P* = 0.788	0.056 (0.324) *P* = 0.863	0.042 (0.336) *P* = 0.901
2nd trimester vs. 4 to 12 months PP	0.080 (0.244) *P* = 0.744	0.029 (0.242) *P* = 0.903	0.182 (0.274) *P* = 0.508	–0.004 (0.260) *P* = 0.989	0.012 (0.166) *P* = 0.943	0.091 (0.246) *P* = 0.711	0.065 (0.252) *P* = 0.798
3rd trimester vs. 4 to 12 months PP	–0.014 (0.229) *P* = 0.953	–0.011 (0.224) *P* = 0.960	–0.009 (0.235) *P* = 0.971	–0.065 (0.226) *P* = 0.775	–0.075 (0.156) *P* = 0.630)	–0.007 (0.230) *P* = 0.976	–0.028 (0.237) *P* = 0.904
1 month PP vs. 4 to 12 months PP	–0.115 (0.222) *P* = 0.606	–0.033 (0.226) *P* = 0.883	–0.109 (0.242) *P* = 0.652	–0.029 (0.237) *P* = 0.902	0.147 (0.160) *P* = 0.357	–0.054 (0.240) *P* = 0.822	–0.120 (0.226) *P* = 0.594
2 months PP vs. 4 to 12 months PP	0.336 (0.211) *P* = 0.110	0.426 (0.216) ***P* = 0.049**	0.379 (0.223) *P* = 0.089	0.404 (0.226) *P* = 0.073	0.377 (0.167) ***P* = 0.024**	0.361 (0.214) *P* = 0.092	0.323 (0.219) *P* = 0.139
3 months PP vs. 4 to 12 months PP	0.346 (0.255) *P* = 0.175	0.400 (0.252) *P* = 0.113	0.339 (0.273) *P* = 0.214	0.425 (0.245) *P* = 0.082	0.377 (0.175) ***P* = 0.031**	0.354 (0.256) *P* =00.167	0.342 (0.258) *P* = 0.185
Low risk	–1.436 (0.271) ***P* = 0.000**	–1.494 (0.271) ***P* = 0.000**	–1.195 (0.389) ***P* = 0.002**	–1.529 (0.269) ***P* = 0.000**	–1.230 (0.195) ***P* = 0.000**	–1.474 (0.277) ***P* = 0.000**	–1.457 (0.278) ***P* = 0.000**
Low SES	0.753 (0.204) ***P* = 0.000**	0.818 (0.204) ***P* = 0.000**	0.988 (0.331) ***P* = 0.003**	0.772 (0.192) ***P* = 0.000**	1.083 (0.149) ***P* = 0.000**	0.737 (0.206) ***P* = 0.000**	0.774 (0.219) ***P* = 0.000**
Publication year	——	–0.018 (0.013) *P* = 0.170	——	——	——	——	——
Other western countries vs. US	——	——	0.273 (0.305) *P* = 0.371	——	——	——	——
Asian countries vs. US	——	——	0.285 (0.349) *P* = 0.414	——	——	——	——
SCID vs. SADS	——	——	——	–0.489 (0.202) ***P* = 0.015**	——	——	——
Other interview type vs. SADS	——	——	——	–0.098 (0.174) *P* = 0.574	——	——	——

Table 10. Meta–regression results for log odds of a diagnosis of major and minor depression (continued)

Explanatory Variables	Model						
	1	2	3	4	5	6	7
DSM III–R vs. RDC	——	——	——	——	–0.113 (0.188) *P* = 0.548	——	——
DSM IV vs. RDC	——	——	——	——	–0.381 (0.190) ***P* = 0.045**	——	——
Other diagnostic criteria vs. RDC	——	——	——	——	–1.487 (0.268) ***P* = 0.000**	——	——
Interviewed women with positive screens only vs. all	——	——	——	——	——	–0.096 (0.140) *P* = 0.490	——
Quality score	——	——	——	——	——	——	–0.014 (0.051) *P* = 0.783

Notes: Estimated coefficients are shown along with their standard errors in parentheses and the *P* -value for a test of statistically significant differences from zero. *P*-values shown in bold type are significant at the < 0.05 level.
DSM III-R, Diagnostic and Statistical Manual of Mental Disorders, Third Edition; DSM IV, Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition; PP, postpartum; RDC, Research Diagnostic Criteria; SADS, Schedule for Affective Disorders and Schizophrenia; SCID, Structured Clinical Interview for DSM-IV; SES, socioeconomic status.

Table 11. Meta–regression results for log odds of a diagnosis of major depression

Explanatory Variables	Model						
	1	2	3	4	5	6	7
Constant	–3.206 (0.209) ***P* = 0.000**	–3.299 (0.272) ***P* = 0.000**	–3.447 (0.515) ***P* = 0.000**	–3.010 (0.373) ***P* = 0.000**	–3.419 (0.191) ***P* = 0.000**	–3.454 (0.239) ***P* = 0.000**	–1.677 (0.871) ***P* = 0.054**
1st trimester vs. 4 to 12 mos PP	0.052 (0.516) *P* = 0.920	0.033 (0.523) *P* = 0.950	–0.180 (0.558) *P* = 0.747)	–0.086 (0.558) *P* = 0.877	0.271 (0.447) *P* = 0.545	0.278 (0.512) *P* = 0.587	–0.085 (0.514) *P* = 0.868
2nd trimester vs. 4 to 12 mos PP	0.375 (0.384) *P* = 0.329	0.424 (0.399) 0.288	0.471 (0.466) *P* = 0.312	0.245 (0.444) *P* = 0.582	0.517 (0.315) *P* = 0.101	0.634 (0.392) *P* = 0.105	0.339 (0.376) *P* = 0.368
3rd trimester vs. 4 to 12 mos PP	–0.272 (0.410) *P* = 0.507	–0.277 (0.415) *P* = 0.503	–0.372 (0.426) *P* = 0.382	–0.351 (0.436) *P* = 0.421	–0.052 (0.354) *P* = 0.883	–0.012 (0.417) *P* = 0.976	–0.379 (0.406) *P* = 0.350
1 month PP vs. 4 to 12 mos PP	0.021 (0.356) *P* = 0.954	–0.033 (0.377) *P* = 0.929	–0.185 (0.409) *P* = 0.651	–0.168 (0.415) *P* = 0.686	–0.073 (0.290) *P* = 0.800	0.060 (0.342) *P* = 0.861	0.035 (0.349) *P* = 0.920
2 mos PP vs. 4 to 12 mos PP	0.557 (0.291) *P* = 0.056	0.533 (0.300) *P* = 0.076	0.538 (0.304) *P* = 0.077	0.377 (0.356) *P* = 0.290	0.573 (0.257) ***P* = 0.026**	0.613 (0.279) ***P* = 0.028**	0.465 (0.289) *P* = 0.107
3 mos PP vs. 4 to 12 mos PP	0.231 (0.342) *P* = 0.499	0.228 (0.347) 0.510	0.097 (0.364) *P* = 0.791	0.139 (0.364) *P* = 0.703	0.159 (0.273) *P* = 0.561	0.279 (0.328) *P* = 0.395	0.180 (0.336) *P* = 0.592
Low risk	–1.501 (0.671) ***P* = 0.025**	–1.436 (0.687) ***P* = 0.036**	–1.045 (0.822) *P* = 0.204	–1.537 (0.695) ***P* = 0.027**	–1.340 (0.609) ***P* = 0.028**	–1.384 (0.658) ***P* = 0.036**	–1.915 (0.702) ***P* = 0.006**
Low SES	0.459 (0.323) *P* = 0.155	0.428 (0.333) *P* = 0.199	0.759 (0.497) *P* = 0.126	0.379 (0.345) *P* = 0.273	0.498 (0.262) *P* = 0.057	0.432 (0.308) *P* = 0.161	0.636 (0.331) *P* = 0.054
Publication year	—	0.010 (0.020) *P* = 0.604	—	—	—	—	—
Other western countries vs. US	—	—	0.238 (0.470) *P* = 0.612	—	—	—	—
Asian countries vs. US	—	—	0.602 (0.533) *P* = 0.258	—	—	—	—
SCID vs. SADS	—	—	—	0.112 (0.332) *P* = 0.736	—	—	—
Other interview type vs. SADS	—	—	—	–0.201 (0.306) *P* = 0.511	—	—	—

Table 11. Meta–regression results for log odds of a diagnosis of major depression (continued)

Explanatory Variables	Model						
	1	2	3	4	5	6	7
DSM III–R vs. RDC	—	—	—	—	0.815 (0.241) ***P* = 0.001**	—	—
DSM IV vs. RDC	—	—	—	—	–0.198 (0.356) *P* = 0.578	—	—
Other diagnostic criteria vs. RDC	—	—	—	—	0.414 (0.218) *P* = 0.058	—	—
Interviewed women with positive screens only vs. all	—	—	—	—	—	0.441 (0.222) ***P* = 0.047**	—
Quality score	—	—	—	—	—	—	–0.132 (0.073) *P* = 0.072

Notes: *P*-values shown in bold type are significant at the < 0.05 level.
DSM III-R, Diagnostic and Statistical Manual of Mental Disorders, Third Edition; DSM IV, Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition; PP, postpartum; RDC, Research Diagnostic Criteria; SADS, Schedule for Affective Disorders and Schizophrenia; SCID, Structured Clinical Interview for DSM-IV; SES, socioeconomic status.

Table 12. Odds ratios for studies with comparison groups of women during nonchildbearing periods

Diagnosis Estimate Type Author, Year	Time Period	Odds Ratio	95% Confidence Interval
Major and Minor Depression			
Point			
O'Hara et al., 1990[53]	2nd trimester	1.41	0.61-3.26
O'Hara et al., 1990[53]	9 weeks PP	1.37	0.67-2.83
Cox et al., 1993[20]	6 months PP	1.00	0.54-1.84
Period			
Cox et al., 1993[20]	Birth to 6 months PP	1.04	0.61-1.76
Incidence			
Cox et al., 1993[20]	Birth to 5 weeks PP	3.26*	1.17-9.06
Cox et al., 1993[20]	Birth to 6 months PP	1.48	0.77-2.82
Major Depression			
Point			
O'Hara et al., 1990[53]	2nd trimester	1.28	0.47-3.51
O'Hara et al., 1990[53]	9 weeks PP	1.33	0.45-3.90
Cooper et al., 1988[52]	3 months PP	0.85	0.33-2.17
Cox et al., 1993[20]	6 months PP	1.00	0.37-2.71
Cooper et al., 1996[33]	6 months PP	1.53	0.65-3.58
Cooper et al., 1996[33]	12 months PP	0.50	0.17-1.46
Period			
Cox et al., 1993[20]	6 months PP	1.16	0.54-2.51

* Statistically significant at $P < 0.05$.
PP, postpartum.

The two retrospective studies that met our inclusion criteria provided estimates of the period prevalence of major depression in the first postpartum year. The first study conducted in 1993 found that 1.2 percent of postpartum women in Olmsted County, Minnesota, had a major depressive episode during their first postpartum year and that 2.5 percent had a major or minor episode.[54] These rates are significantly below the 3-month period prevalence of 7.1 percent for major depression alone and the 19.2 percent for major and minor depression reported in Tables 8 and 9.

In 1997-1998, universal screening for depression with the EPDS at the 6-week postpartum visit was implemented in Olmsted County. As a result, the prevalence of a diagnosis of major depression among postpartum women rose to 10.7 percent, suggesting that the screening score posted in medical charts led clinicians to become more aware of their patients' mental state.[55]

Discussion

We found 30 studies providing estimates of the prevalence of perinatal depression but only 13 providing estimates of the incidence of the disorder. The studies were generally of moderate

size—too small for reliable subgroup analyses. Furthermore, the study populations were typically restricted to a local community or geographic region served by one provider or a small number of providers of obstetrical services and were not representative of the racial and ethnic mix of the countries in which the studies were conducted. Other confounders included the risk status of women at study entry, their socioeconomic status, the interview methods, and the diagnostic criteria used to identify cases.

Combining point prevalence estimates of depression assessed at the same point in time and distinguishing whether they included minor depression, we found that the best estimates of the point prevalence of major and minor depression ranged from 8.5 percent to 11.0 percent at different times during pregnancy and from 6.5 percent to 12.9 percent at different times during the first year postpartum. Including only major depression, the best point prevalence estimates ranged from 3.1 percent to 4.9 percent at different times during pregnancy and from 1.0 percent to 5.9 percent at different times during the first postpartum year.

Period prevalence estimates show that as many as 19.2 percent of women have a depressive episode during the first 3 months postpartum, with as many as 7.1 percent having a major depressive episode during this time. Most of these episodes began following delivery. Incidence estimates show that, during the same 3-month period, 14.5 percent of women had a new depressive episode with as many as 6.5 percent having a major depressive episode. However, all of these estimates have wide 95% CIs, indicating that the amount of uncertainty in their precise values is considerable.

Our best estimates of prevalence and incidence were somewhat lower than those found in prior systematic reviews because we excluded studies that assessed depression based on self-report screens alone, which tend to overestimate prevalence. In addition, we separate out estimates of major and minor depression from estimates of major depression alone and estimates of point prevalence from estimates of period prevalence. Finally, we include more recent studies that use more precise criteria to identify major depression.

We found that the available evidence does not support the hypothesis that the prevalence of depression is higher during pregnancy or in the first year postpartum compared to nonchildbearing times. A single study suggested that the incidence of new depressive episodes (major and minor) is greater in the first 5 weeks postpartum than at other times.[20]

Nevertheless, pregnancy and the early postpartum period provide opportunities to screen for depression through regular prenatal and postpartum physician contacts. Because the poor outcomes of suffering from depression during the perinatal period can be farther reaching—affecting not only the woman but her newborn child and other family members—it behooves us to investigate the efficacy of screening and treatment programs for these women.

Chapter 4. Screening Accuracy

Introduction

Screening for perinatal depression is an important first step in identifying women who are at risk of having perinatal depression. It is only an initial step—after a positive screen, a depressive illness must be confirmed by a follow-up diagnostic examination and determination by a clinician.

To be useful screening tools, instruments must be able to identify accurately and reliably the illness in the population of interest; they also need to rule out, accurately, persons in the population who do not have the illness. Assessment of a screening test's accuracy depends on knowing whether a disease is truly present, i.e., comparison to a reference standard. This section addresses the second Key Question (KQ) from the Safe Motherhood Group (SMG) and the Agency for Healthcare Research and Quality (AHRQ): "What is the accuracy of different screening tools for detecting depression during pregnancy and during the postpartum period?"

The two most commonly used measures of accuracy are sensitivity and specificity. Sensitivity refers to the proportion of patients with a disease who test positive ("true positives") using a screening tool. A sensitive test is one that is usually positive in the presence of disease. In general, a highly sensitive test should be selected when the consequence of missing a disease would be a clearly bad outcome. Screens with high sensitivity are most useful to clinicians when the result is negative; negative results can help rule out a disease.

Specificity refers to the proportion of patients without a disease who test negative ("true negatives") using the screening tool. A specific test is one that is usually negative in the absence of disease. A highly specific test, then, should be selected when false-positive results can substantially harm the patient in some way. Screens with high specificity are most useful to the clinician when the result is positive; the positive result can rule in the disease.

Screening tools have varying sensitivities and specificities as a function of which cutoff point, or threshold, clinicians and others use. The optimal cutoff depends on prevalence of disease (as explored in Chapter 3), benefits and harm of therapy, and risks and costs of administering the screening test.

Methods

Chapter 2 provides the detailed methods we used to search and review the literature on screening instrument accuracy. In this discussion, we elaborate on some of these methods.

Note: Appendixes and Evidence Tables cited in this report are provided electronically at http://www.ahrq.gov/clinic/epcindex.htm.

Inclusion and Exclusion Criteria

Studies to be retained had to report directly or to provide data allowing us to calculate our primary outcomes of interests—sensitivity and specificity. We required that the screening instrument be compared to a reference standard for a diagnosis of depression. Reference standards could be one of two types. The first includes a clinical assessment by a mental health professional based on criteria from the Diagnostic and Statistical Manual of Mental Disorders (DSM), the Research Diagnostic Criteria (RDC), the Bedford College Checklist,[69] or the International Classification of Diseases (ICD). The second involves a research-based diagnosis obtained by structured or semistructured clinical interview, such as the Structured Clinical Interview for Depression (SCID), the Diagnostic Interview Schedule (DIS), the Schedule for Affective Disorders and Schizophrenia (SADS), or Goldberg's Standardized Psychiatric Interview (SPI); each of these confirms a diagnosis based on one of the above systems of criteria.

Depressive illness can be either a major depressive disorder or a minor depression. The latter is understood to be an impairing, episodic depression with clear symptoms exceeding a normal state but without severity reaching the diagnostic criteria for major depressive disorder. For this chapter, we are concerned with the ability of screening tools to detect either major depression *or* minor depression in a given individual (because an individual can have only one or the other of these diagnoses), so the terminology intentionally differs from that used in Chapter 3.

We excluded studies that included patients with a known current depressive illness (for whom a screen would not provide new information). Furthermore, we excluded studies on women with bipolar disorder or a primary psychotic disorder and studies in which women with diagnosed depression could not be distinguished from women with maternity blues, a transient, subthreshold cluster of depressive symptoms commonly described in up to 50 percent of postpartum women.

Data Analysis

Our main outcomes of interest were sensitivity and specificity of the screening approaches or instruments as described in the selected articles. When calculating outcomes ourselves or doing other analyses, we used Stata, version 8. For each reported instrument and associated cutoff, we calculated sensitivity and specificity from the published data. We constructed 95% confidence intervals (CIs) using exact methods. For instruments with three or more outcome values reported, we created plots of the sensitivity or specificity with associated 95% CIs to provide a graphic description of the degree of consistency of results. In addition, where possible we estimated pooled sensitivity and specificity values using meta-analytic methods for fixed effects. We evaluated heterogeneity using the Q statistic test for homogeneity. In several circumstances, pooled estimates were impossible to calculate because of perfect estimates of sensitivity (i.e., 100 percent) with associated variance estimates equal to 0.

Evaluation of Quality and Strength of Evidence

We developed a quality rating form for these articles on screening accuracy from criteria identified by the Cochrane Methods Working Group on Systematic Review of Screening and Diagnostic Tests.[25] The quality rating forms, provided in Appendix B, rated reporting, external

validity, and internal validity. The senior abstractor completed the quality rating form for each article; another project team member reviewed a sample of the completed forms for accuracy and completeness.

We rated retained studies on three separate categories of quality then summed the individual category scores for a total score. The domains and maximum points possible for each domain are as follows:

- Reporting (domain score of 10): Nine items covering study aims, description of depression assessment, potential confounders described, and instrument procedures described, each scored yes or no (1 or 0), except for an item concerning principal confounders that was scored yes, partially, or no (2, 1, or 0). We considered 0 to 3 as poor, 4 to 7 as fair, and 8 to 10 as good.

- External validity (domain score of 3): Three items relating to representativeness of populations from which people were recruited and of settings and clinicians that treat such patients, each scored yes or no (1 or 0). We considered 0 or 1 as poor, 2 as fair, and 3 as good.

- Internal validity (domain score of 8): Six items relating to both bias and confounding in the use of the screen and reference standard, each scored yes or no (1 or 0), except for an item assessing whether all screens were done independently on each person, all tests done on each person but not independently, or different tests done on different persons and not randomly allocated (2, 1, or 0, respectively). We considered 0 to 2 as poor, 3 to 5 as fair, and 6 to 8 as good.

The maximum total quality score was 21. We considered 0 to 7 as poor, 8 to 14 as fair, and 15 to 21 as good.

Results

Study Characteristics

Our literature review of screening tools for detecting depression during pregnancy and the postpartum period identified no relevant systematic reviews. We did find 23 studies meeting our inclusion criteria. Of these, 10 were studies involving screening instruments in English;[32,46,70-77] 13 involved non-English screening instruments.[31,35,42,43,50,51,78-84]

The major characteristics of these studies are summarized in Table 13 and detailed in Evidence Table 3 (Appendix C). The studies represent a wide variety of countries. Of the 10 studies using an English-language screening instrument, two were conducted in US populations,[32,71] six were performed in the United Kingdom,[46,70,73-76] and one each was conducted in Canada[77] and Australia.[72] Of the 13 studies using a non-English screening instrument, four were conducted in Chinese,[42,43,81,82] three were in Japanese,[50,51,80] and one each was in German (Austria),[83] Swedish,[84] French,[78] Spanish (Spain),[35] Norwegian,[31] and

English/Africaans.[79] We will focus on the screening instruments used in the English language, given their greater relevance to our population of interest.

Unfortunately, the racial and ethnic mix of the study populations for the studies using English language screening instruments was poorly representative of the US population (our target of interest). Of the 10 studies, only the two studies conducted in the United States reported race and ethnicity.[32,71] These populations were overwhelmingly Caucasian; in by far the largest study,[32] 100 percent of the 1,007 women enrolled were white, and, in the other, 87 percent of the women were white.[71]

When reported, the mean age of women in these studies ranged from approximately 24 to 31 years. Of these 10 studies, only one was conducted during pregnancy.[46] The remaining nine studies were conducted postpartum between 2 weeks and 6 months after delivery, with most

Table 13. Major characteristics of studies of screening for perinatal depression

Author, Year	Place/ Sample Size	Depression Type and Prevalence	Screening Method(s) and Cutoffs Used	Timing of Screenings	Criterion Standard
Prenatal Period					
Murray and Cox, 1990[46]	UK 100	Major depression: 6% major or minor depression: 14%	EPDS: cutoffs vary from ≥ 11 to ≥ 15	28 to 34 weeks GA	SPI to obtain RDC diagnosis
Postpartum Period					
Ballard et al., 1994[70]	UK 200	Major depression alone: 12%	EPDS: cutoff 13	6 months PP	PAS to obtain RDC diagnosis
Beck and Gable, 2001[71]	US 150	Major depression alone: 12% Major or minor depression: 19%	PDSS ≥ 81 EPDS ≥ 13 BDI-II ≥ 21	Between 2nd and 12th week PP	SCID-DSM-IV for DSM-IV diagnosis
Boyce et al., 1993[72]	Australia 103	Major depression alone: 9%	EPDS ≥ 13 GHQ: NR Pitt Scales: NR	≤ 6 months PP	DIS to obtain DSM-III-R diagnosis
Campbell and Cohn, 1991[32]	US 1,007	Major or minor depression: 9%	CES-D	6 to 8 weeks after delivery	Modified SADS to obtain RDC diagnosis
Cox et al., 1996[73]	UK 128	Major depression alone: 6% Major or minor depression: 16%	EPDS ≥ 13 (primarily) but also ≥ 10, ≥ 11, ≥ 12, ≥ 14, ≥ 15	Not reported in relationship to time of birth	SPI to obtain RDC diagnosis
Harris et al., 1989[74]	Wales 147	Major depression alone: 15%	BDI: ≥ 11 EPDS: ≥ 13	6 to 8 weeks PP	Clinical examination for DSM-III criterion
Leverton and Elliott, 2000[75]	England 199	Major or minor depression: Catego: 5%; Bedford: 8%	EPDS ≥ 13	3 months PP	PSE with 2 standards used: Bedford College and Catego diagnosis
Murray and Carothers, 1990[76]	England 646	Not provided, but data suggest major depression alone: 6% Major or minor depression: 15%	EPDS ≥ 13	6 weeks PP	SPI to obtain RDC diagnosis
Whiffen, 1988[77]	Canada 120	Major or minor depression: 18%	BDI ≥ 10	6 to 8 weeks PP	SADS to obtain RDC diagnosis

BDI, Beck Depression Inventory; CES-D, Center for Epidemiological Studies-Depression scale; DIS, Diagnostic Inventory Schedule; DSM-III-R, Diagnostic and Statistical Manual of Mental Disorders, third edition, revised; DSM-IV, Diagnostic and Statistical Manual of Mental Disorders, fourth edition; EPDS, Edinburgh Postnatal Depression Scale; GA, gestational age; GHQ, General Health Questionnaire; PAS, Psychiatric Assessment Schedule; PDSS, Postpartum Depression Screening Scale; PP, postpartum; PSE, Present State Examination; RDC, Research Diagnostic Criteria; SADS, Schedule for Affective Disorders and Schizophrenia; SCID, Structured Clinical Interview for DSM-III-R; SPI, Standardized Psychiatric Interview.

occurring between weeks 8 and 12. Individual study sizes ranged from 103 to 1,007, with an aggregate sample size of 2,800.

Studies might use one or more screening tools; the selected articles evaluated four different screening tools.

Screening Instruments Used

The key features of the four different types of screening instruments used are summarized in Table 14. Eight studies assessed the Edinburgh Postnatal Depression Scale (EPDS); seven of these used the 10-item version,[46,71-76] and one used a 13-item version.[70] The EPDS had been developed specifically for assessing postpartum depression and relies much less than standard depression screens on somatic, or physical, questions. In its most common form, it is a 10-item self-report screening scale for postpartum depression that is specifically aimed at exploring mood symptoms in the postpartum period.[85] Questions on the EPDS scale are framed within the "past seven days" and the response format is frequency-based. Each item is scored on a 4-point scale (0 to 3); the minimum and maximum scores are 0 and 30, respectively. It takes less than 5 minutes to administer. The responses to the 10 items are summed to obtain a score.

Table 14. Key features of screening instruments for perinatal depression

Screening Tool	Method of Administration	Number of Items	Score Ranges	Time to Complete	Time Frame Covered
EPDS	Self-administered	10-item* 13-item	0-30 0-39	< 5 minutes	In the past 7 days
BDI†	Interviewer- or self-administered	21-item	0-63	5-10 minutes	Last week including today
BDI-II†	Interviewer- or self-administered	21-item	0-63	5-10 minutes	During the past 2 weeks
PDSS	Self-administered	35-item	35-175	5-10 minutes	Over the past 2 weeks
CES-D	Self-administered	20-item	0-80	1-2 minutes	Past 7 days

*The 10-item EPDS is more commonly administered than the 13-item version.
†BDI and BDI-II were originally designed to be administered by an interviewer but are most often self-reported.

Three studies assessed the Beck Depression Inventory (BDI).[71,74,77] The BDI is a list of 21 symptoms and attitudes that are each rated on intensity.[86] Versions include the BDI, which uses "last week, including today" as the time frame for symptoms;[86] the BDI-II, which uses 2 weeks as the time frame for symptoms;[87] and the BDI-PC, which also has a 2-week time frame.[88] The versions used most often (BDI or BDI-II) are scored by summing the ratings that respondents give to the 21 items. Although originally designed to be administered by trained interviewers, it is most often self-administered and takes 5 to 10 minutes to complete. This instrument has been used to measure severity of depression in depressed samples and also to assess depression in general population samples. Because of its reliance on somatic symptoms, some experts worry that it may produce higher scores and more false-positive results in pregnant women than in other respondents.

One study used the Postpartum Depression Screening Scale (PDSS).[71] The PDSS is a 35-item Likert-type self-report instrument created specifically for new mothers that can be administered in 5 to 10 minutes. Written at a third-grade reading level, PDSS items are brief and easy to understand. Mothers respond using a 5-point scale ranging from "strongly disagree" to "strongly agree." The test yields an overall severity score falling into one of three ranges: normal adjustment, significant symptoms of postpartum depression, and positive screen for major postpartum depression. The PDSS also provides scores for seven symptom areas: Sleeping/Eating Disturbances, Anxiety/Insecurity, Emotional Lability, Mental Confusion, Loss of Self, Guilt/Shame, and Suicidal Thoughts.

Another study used the Center for Epidemiological Studies Depression Scale (CES-D).[32] The CES-D was designed to measure current level of depressive symptomatology and especially depressive affect.[89] The 20 items were chosen from five previously used depression scales to represent all major components of depressive symptomatology, and it was designed to apply to a general population. Each item is rated on 4-point scales indicating the degree of its occurrence during the past week. The scales range from "rarely or none of the time" to "most all of the time." The scale can distinguish between clinical groups and general community groups. It takes approximately 5 to 10 minutes to complete; scoring takes about 1 to 2 minutes. Although it is usually scored continuously, various cutoff scores for clinical depression have reasonable associations with a clinical diagnosis. A cutoff score of 16 or higher has been suggested as a positive screen for depression.[89]

Reference Standards Used

Investigators used a variety of strategies to confirm the diagnosis of depression. Six studies used the RDC[65] for depressive illness as the reference standard but employed different instruments to identify patients meeting this standard. Three studies used the Standardized Psychiatric Interview,[46,73,76] two studies used a version of the Schedule for Affective Disorders and Schizophrenia,[32,77] and one study used the Psychiatric Assessment Schedule (an adaptation of the Present State Examination).[70]

Other reference standards were also employed. Beck and Gable used the Structured Clinical Interview for DSM IV to confirm the diagnosis of depressive illness per DSM IV criteria;[71] Boyce et al. used the Diagnostic Interview Schedule, based on DSM III-R criteria, as the reference standard to confirm depressive illness;[72] Harris et al. used a clinical assessment of whether a patient's presentation met DSM III criteria for depressive illness;[74] and Leverton and Elliott used the Present State Examination to identify whether patients met depressive illness criteria by either the Bedford College Criteria or the Catego criteria (based on ICD-8 criteria).[75]

Classifications of Depressive Illness

Investigators classified depressive illness into one of two categories that reflected how perinatal depression is described in the scientific literature: major depression alone or major or minor depression. Patients identified as major depression alone met criteria for an episode of severe depressive illness according to the standardized criteria. In this report, we refer to major depressive episodes as major depression. For major depressive disorders, clearly effective interventions have been identified in clinical trials. Seven studies provided this classification.[46,70-74,76]

The point prevalence for major depression alone was 6 percent in the single prenatal study,[46] somewhat higher than the 3.1 percent "best estimate" that we discussed in Chapter 3. For the postpartum studies, the point prevalence for the six studies reporting on major depression alone ranged from 6 percent to 15.5 percent;[70-74,76] this frequency is somewhat higher than the postpartum results from KQ 1 showing a best estimate prevalence between 1 and 3 months postpartum of 3.8 percent and 4.7 percent, respectively.

The major or minor depression category of depressive illness requires that patients meet diagnostic criteria for either a major depressive episode or a minor depressive episode. Minor depression is an impairing yet less severe constellation of depressive symptoms[13] for which controlled trials have not consistently indicated that particular interventions are more effective than placebo.[14,15] In this report, we refer to this grouping as major or minor depression, or by the more general terms of "depression" or "depressive illness." Seven studies classified depression in this way.[32,46,71,73,75-77]

In the single prenatal screening study, the point prevalence of major or minor depression in the third trimester (14 percent) was greater than our best estimate from KQ 1 for this time period (8.5 percent).[73] For the postpartum studies, prevalence rates ranged from 5 percent to 19 percent; these figures are somewhat higher than our best estimate range for point prevalence of 9.7 percent to 12.9 percent in the first 3 months postpartum. Given that this distinction substantially affects screening accuracy at a particular cutoff, we sort the results below by these two case definitions.

Quality Rating

Table 15 documents the details of our grading of individual studies. For reporting completeness, we rated studies as fair; they averaged 6.1 of a possible 10 points. Three studies scored in the good range (8 or above).[32,70,72] For external validity, studies ranged from poor to good, averaging a poor-to-fair rating of 1.6 overall (of a possible 3 points), suggesting that at best they were a fair representation of each individual study's target population. Given that only two studies (Campbell and Cohn with an external validity score of 3[32] and Beck and Gable with an external validity score of 1[71]) were conducted on US populations, the generalizability of these results to our target population appears limited. For internal validity, studies scored better, ranging from 4 to 8, with an overall average in the good range (7.0). Total scores for the three categories ranged from fair to good, with an overall average of 14.7 of a possible 21 points; of these 10 studies, six scored in the good overall quality range (15 or higher).

Prenatal Screening Results

One English study of 100 subjects used the10-item EPDS to screen women in their third trimester of pregnancy (Table 13).[46] For major depressive disorder alone (n = 6 depressed patients), sensitivity and specificity point estimates at all cutoff points (12, 13, 14, 15) were quite good, although the sensitivity estimates were imprecise (as demonstrated by the wide CIs). At all cutoff points used, sensitivity was 100 percent, and each cutoff had a wide CI from 0.54 to 1.0 (Figure 14a). Specificity varied among the different cutoff points, with means varying from 0.79 (at a cutoff of 12) up to 0.96 (at a cutoff of 15), and all CIs were more precise, reflecting the larger number of subjects (n = 94) without major depression alone (Figure 14b). At the traditional postpartum cutoff of ≥ 13, sensitivity was 100 percent and specificity was 87 percent.

As a screening instrument for major or minor depression (n = 14 depressed patients), overall test performance was worse. Sensitivity was much lower, ranging from 0.71 (at a cutoff of 11 or greater) to 0.57 (at a cutoff of 14 or greater), and CIs remained wide (Figure 15a). Specificity remained relatively good, varying from 0.72 (cutoff of 11 or greater) to 0.95 (cutoff of 14 or greater), with reasonable precision (Figure 15b). At the same ≥ 13 cutoff, sensitivity was 64 percent and specificity was 90 percent.

In summary, one prenatal screening study is of good quality. However, the inclusion of only six women with major depression substantially limits conclusions about the accuracy of prenatal depression screens. Indeed, the sensitivity results at 100 percent for each cutoff dramatically underscore the small number of depressed patients involved.

Results for major or minor depression from this one study are similarly limited. Only 14 depression cases are involved. Sensitivity and specificity estimates appeared to be lower than those for major depression alone. In particular, sensitivity estimates appeared worse than those for major depression alone, but again CIs are wide.

Table 15. Quality rating of studies of screening for perinatal depression

Author, Year	Reporting (10)	External Validity (3)	Internal Validity (8)	Total Score (21)
Studies with Screener in English				
Prenatal Period				
Murray and Cox, 1990[46]	5	3	8	16
Postpartum Period				
Ballard et al., 1994[70]	9	1	8	18
Beck and Gable, 2001[71]	6	1	8	15
Boyce et al., 1993[72]	8	3	5	16
Campbell and Cox, 1991[32]	8	3	8	19
Cox et al., 1996[73]	5	0	8	13
Harris et al., 1989[74]	5	2	6	13
Leverton and Elliott, 2000[75]	5	0	7	12
Murray and Carothers, 1990[76]	4	3	8	15
Whiffen, 1988[77]	6	0	4	10
Average	6.1	1.6	7.0	14.7

Note: Maximum possible score is shown in parentheses.

Figure 14a. Sensitivity of screening by Edinburgh Postnatal Depression Scale: prenatal period, major depression alone

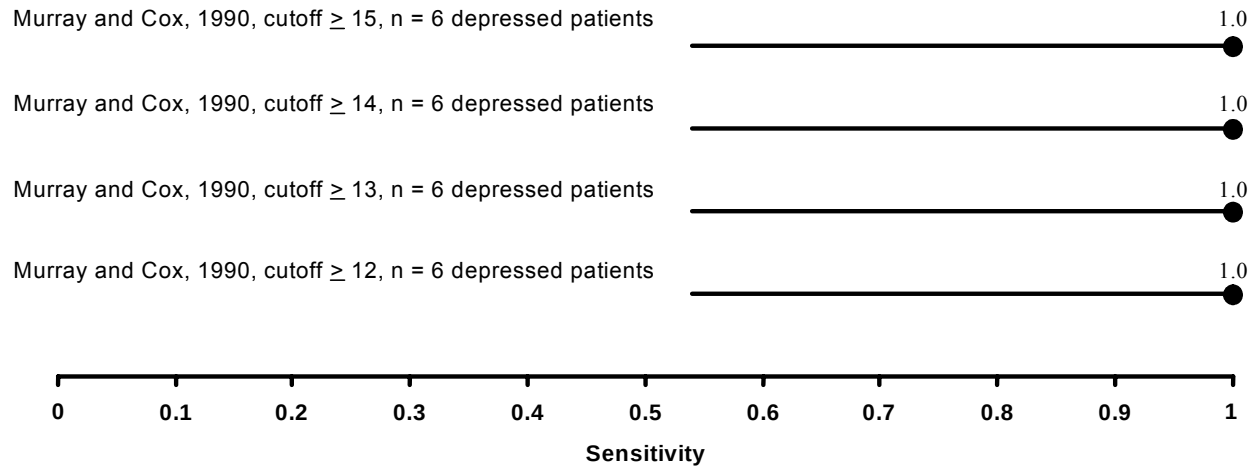


Figure 14b. Specificity of Edinburgh Postnatal Depression Scale: prenatal period, major depression alone

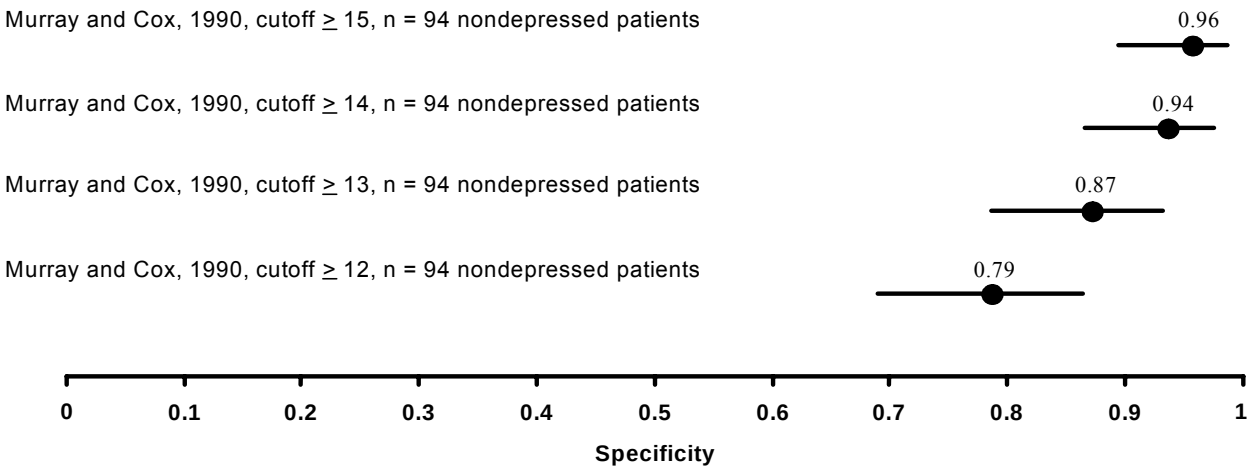


Figure 15a. Sensitivity of screening by Edinburgh Postnatal Depression Scale: prenatal period, major or minor depression

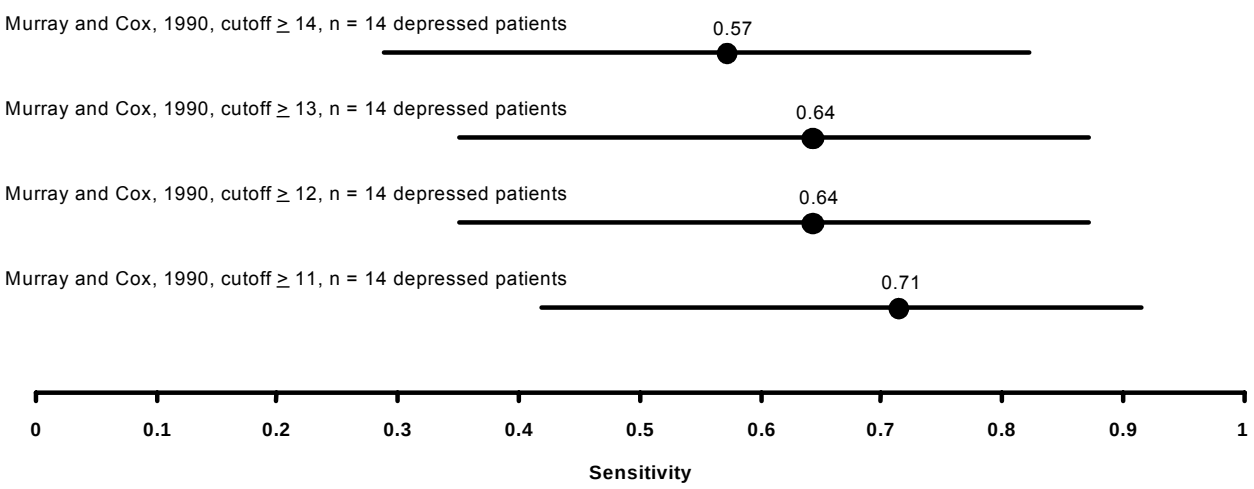
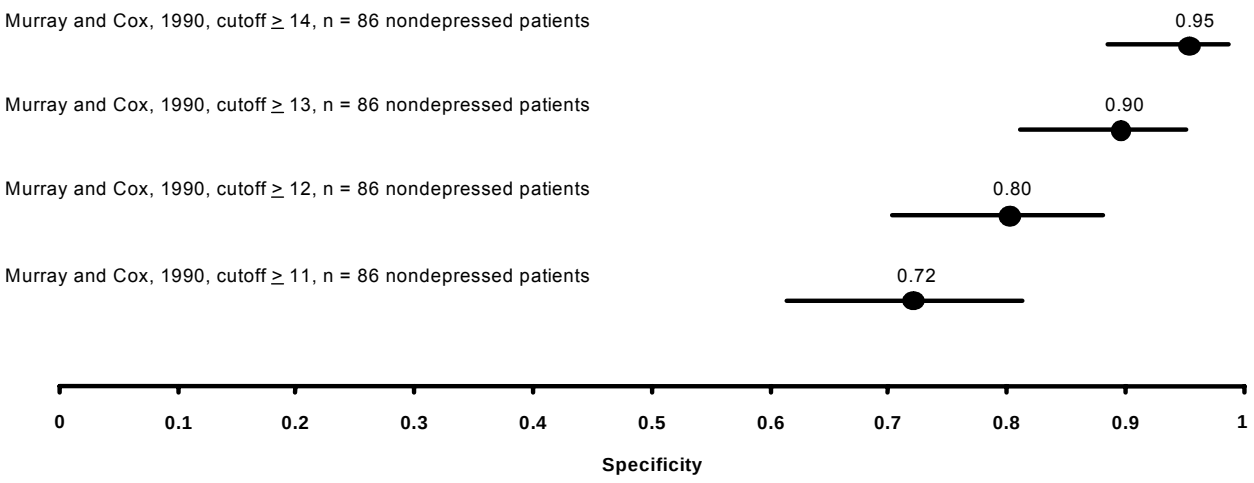


Figure 15b. Specificity of screening by Edinburgh Postnatal Depression Scale: prenatal period, major or minor depression

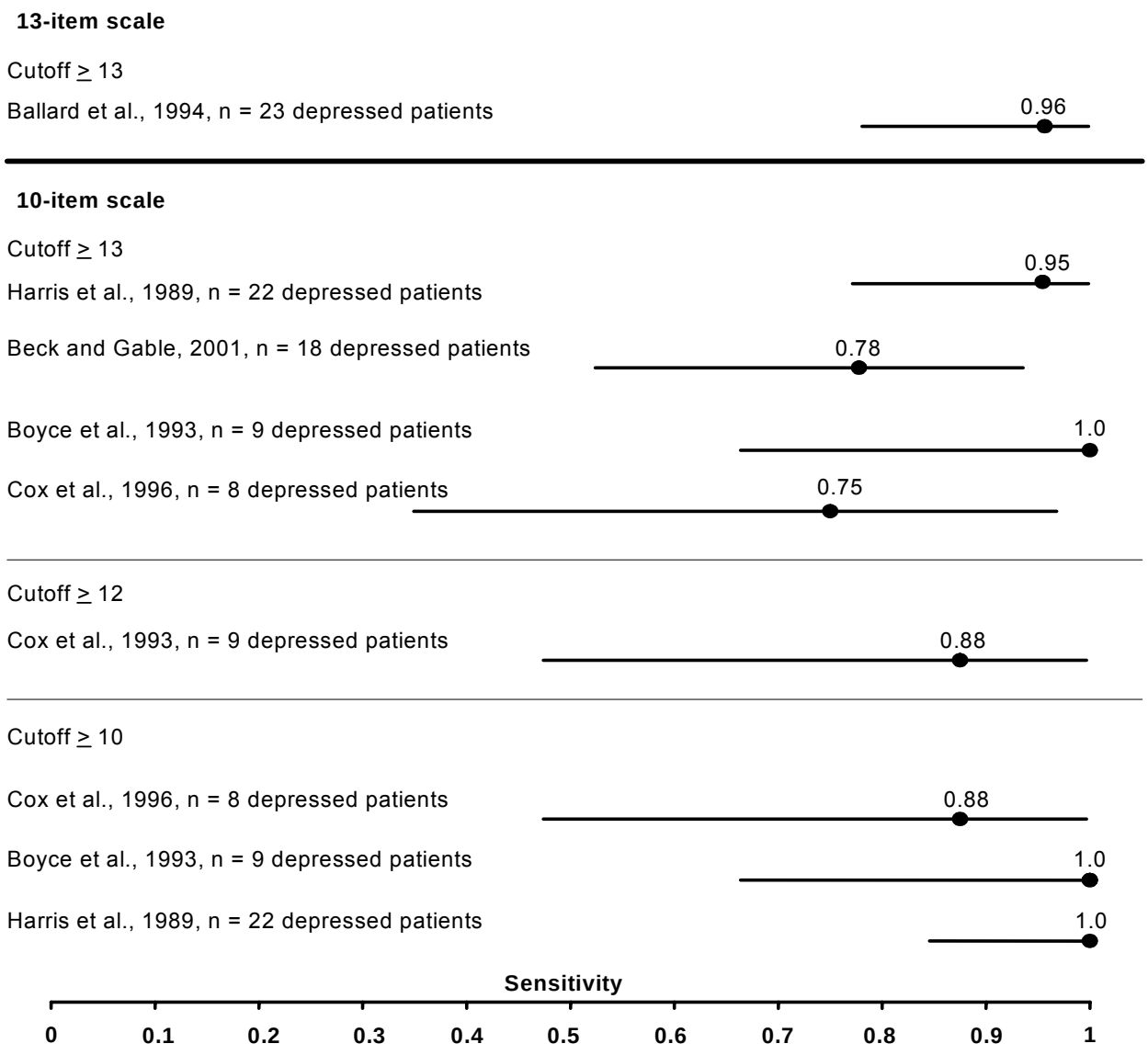


Postpartum Screening Results

Nine studies provided sensitivity and specificity estimates in the postpartum period (Table 16). These studies used one of four screening instruments—EPDS, BDI, PDSS, and CES-D—at a variety of cutoff points. We review the results separately for each scale for major depression alone and for major or minor depression.

Edinburgh Postnatal Depression Scale. The EPDS was the most common tool reported, involving 1,573 patients from eight studies;[46,70-76] 80 patients had major depression alone, and 83 patients had major or minor depression. Murray and Carothers reported test characteristics for both major depression alone and major and minor depression together (not in Table 16), but they did not give information allowing us to calculate CIs for the results;[76] we address their work separately.

Figure 16a. Sensitivity of screening by Edinburgh Postnatal Depression Scale: postpartum period, major depression alone

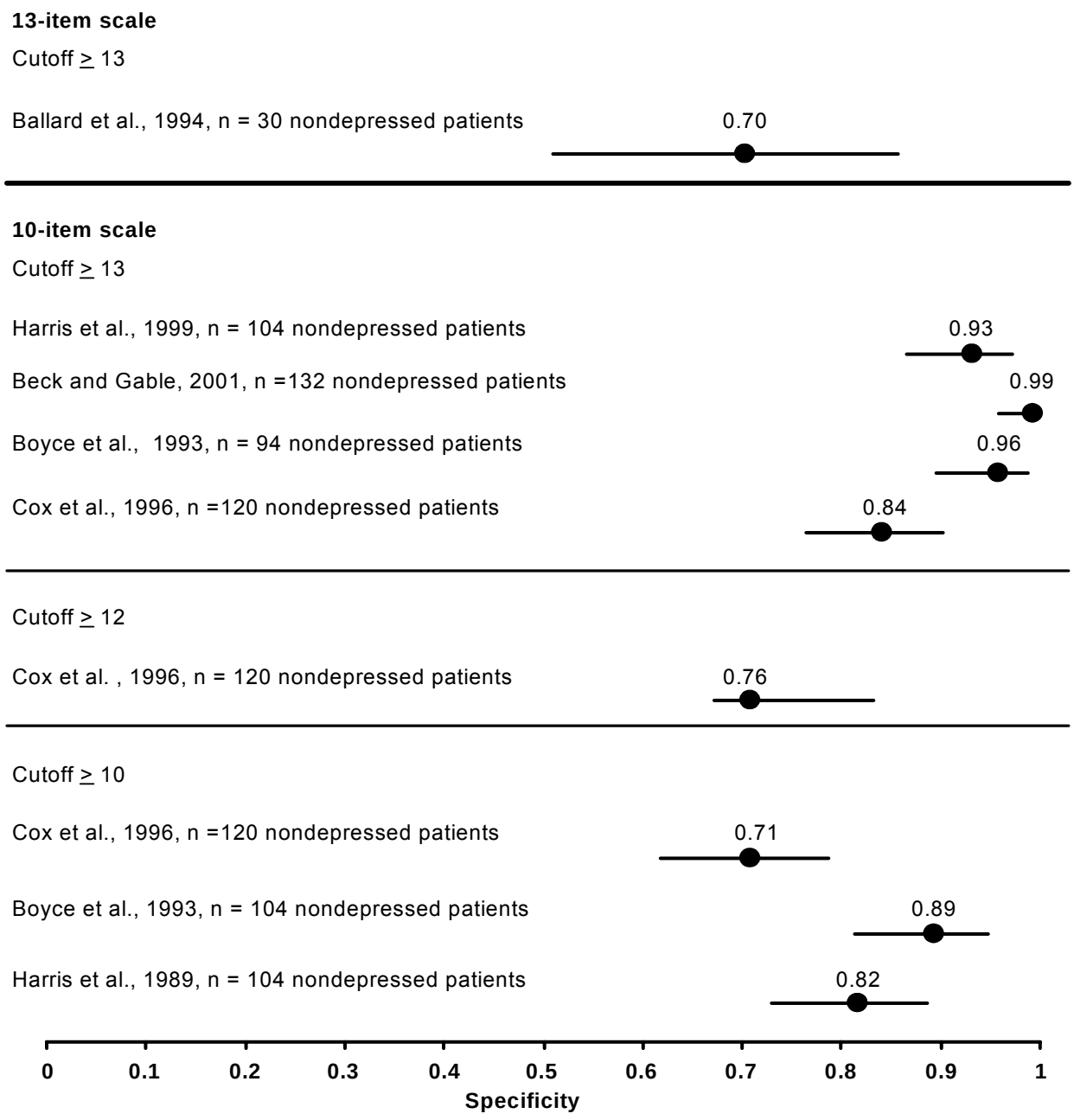


Major Depression Alone. Figures 16a and 16b present the sensitivity and specificity estimates for the five studies reporting on major depression alone (a total of 927 patients).[70-74] We show data according to the version and cutoff point used. The sensitivity graphs show the number of depressed patients; the specificity graphs use the number of nondepressed patients.

For the Ballard et al. study employing the 13-item version (n = 23 depressed women),[70] we used only the cutoff of ≥ 13. Mean sensitivity was 0.96 and mean specificity was 0.70, with relatively wide CIs for both point estimates.

Results for the remaining four major depression alone studies are listed below the solid line in Figures 16a and 16b.[71-74] All used a cutoff point of 13. Sensitivities in these studies range from 0.75 to 1.0, with very wide CIs.

Figure 16b. Specificity of screening by Edinburgh Postnatal Depression Scale: postpartum period, major depression alone



Specificities ranged from 0.84 to 0.99 and appeared to be more precise than sensitivities, as indicated by the much narrower CIs. Of note, results at this threshold from these individual studies of the 10-item screen indicated that sensitivities were similar to the value reported in the one 13-item screen study, but specificities were higher with the 10-item version.

We attempted to conduct a meta-analysis of the sensitivity results from the four studies using the cutoff point of 13 or greater. The Boyce et al. study[72] reported a sensitivity point estimate of 1.0, thus we were unable to generate a meaningful standard error; consequently, we could not include this result in the sensitivity meta-analysis. Leaving this study out, our meta-analysis produced a sensitivity point estimate of 0.91 (95% CI, 0.84 to 0.99); the test for heterogeneity was not significant ($P = 0.141$). We were able to include all four studies in our meta-analysis of specificity, but heterogeneity was significant ($P < 0.001$), precluding a pooled specificity estimate.

One study assessed a cutoff point of ≥ 12.[73] It reported a sensitivity of 0.88 (with a wide CI) and a specificity of 0.76 (with a narrow CI).

Three studies reported a cutoff of ≥ 10, all producing estimates with imprecise sensitivities yet relatively precise specificities.[72-74] Point estimates for sensitivity ranged from 0.88[73] to 1.0.[72,74] Because two studies reported a perfect sensitivity of 1.0, we could not determine a pooled sensitivity estimate. Specificity ranged from 0.71 to 0.89, but heterogeneity was significant ($P = 0.002$), precluding a pooled estimate.

Major or Minor Depression. For major or minor depression (1,343 patients), four studies[71,73,75,76] reported test characteristics for the 10-item EPDS (Table 16). All but Murray and Carothers[76] allowed a calculation of confidence intervals and are presented in Figures 17a and 17b.

Two studies report a cutoff score of ≥ 13.[73,75] Sensitivities were low (0.62[73] and 0.44[75]) and imprecise (wide CIs). Specificities were high (0.89 and 0.92, respectively) and quite precise. A meta-analysis at this cutoff produced a pooled sensitivity estimate of 0.54 (95% CI, 0.39 to 0.70) without significant heterogeneity ($P = 0.266$) and a pooled specificity estimate of 0.91 (95% CI, 0.88 to 0.94) without significant heterogeneity ($P = 0.410$).

One study reported a cutoff score of 12 or greater.[73] Relative to a threshold of 13 or more, this score appeared to improve sensitivity and decrease specificity, with the precision remaining unchanged.

Three studies reported results with a cutoff score of ≥ 10.[71,73,75] Reported sensitivities ranged from 0.59 to 0.81, and specificities ranged from 0.77 to 0.88. Again, sensitivity estimates were quite imprecise, whereas specificity estimates were quite precise. A meta-analysis of these results produced a pooled sensitivity estimate of 0.68 (95% CI, 0.58 to 0.78) without significant heterogeneity ($P = 0.140$). Specificities could not be pooled because of significant heterogeneity ($P = 0.068$).

Murray and Carothers reported sensitivities and specificities at various cutoff points as estimated by logistic regression analyses on results from 646 subjects;[76] they addressed both major depression alone and major or minor depression. Because we could not calculate CIs from their reported results, we do not show them in Table 16; their reported test characteristics are listed in Evidence Table 4 (Appendix C). For major depression alone, their sensitivity and specificity results mirrored the other studies' point estimates in Figures 16a and 16b. For major or minor depression, although specificities were similar to those of other studies, sensitivities were slightly higher than those reported in Figures 17a and 17b. For example, at a cutoff of ≥ 10, sensitivity was reported as 0.89; at a cutoff of ≥ 13, sensitivity was 0.68.

Beck Depression Index. Three studies involving the BDI (417 patients in all) reported lower sensitivity and slightly higher specificity than did the EPDS studies.[71,74,77] Two studies[71,74] reported results for major depression only using the BDI-II and BDI, respectively (Figures 18a and 18b). Using a cutoff of ≥ 21, Beck and Gable reported a sensitivity of 0.56 and a specificity of 1.0,[71] Harris et al. reported a sensitivity of 0.32 and a specificity of 0.99 at this cutoff.[74] Harris et al. also reported test characteristics for cutoff points of ≥ 13 and ≥ 11;[74] these cutoff

Table 16. Sensitivity and specificity of perinatal depression screens

Author, Year	Cutoff (≥)	Point Estimate for Sensitivity 95% CI	Point Estimate for Specificity 95% CI
Prenatal period			
EPDS, Major depression			
Murray and Cox, 1990[46]	15	1.0 0.54-1.0	0.96 0.89-0.99
	14	1.0 0.54-1.0	0.94 0.87-0.98
	13	1.0 0.54-1.0	0.87 0.79-0.93
	12	1.0 0.54-1.0	0.79 0.69-0.86
EPDS, Major or minor depression			
Murray and Cox, 1990[46]	14	0.57 0.29-0.82	0.95 0.89-0.99
	13	0.64 0.35-0.87	0.90 0.81-0.95
	12	0.64 0.35-0.87	0.80 0.70-0.88
	11	0.71 0.42-0.92	0.72 0.61-0.81
Postpartum Period			
EPDS, Major depression			
Ballard et al., 1994 (13-item version)[70]	13	0.96 0.78-1.0	0.70 0.51-0.85
Harris et al., 1989[74]	13	0.95 0.77-1.0	0.93 0.87-0.97
	10	1.0 0.85-1.0	0.82 0.73-0.89
Beck and Gable, 2001[71]	13	0.78 0.52-0.94	0.99 0.96-1.0
Boyce et al., 1993[72]	13	1.0 0.67-1.0	0.96 0.89-0.99
	10	1.0 0.66-1.0	0.89 0.81-0.95
Cox et al., 1996[73]	13	0.75 0.35-0.97	0.84 0.76-0.90
	12	0.88 0.47-1.0	0.76 0.67-0.83

Author, Year	Cutoff (≥)	Point Estimate for Sensitivity 95% CI	Point Estimate for Specificity 95% CI
	10	0.88 0.47-1.0	0.71 062-0.79
EPDS, Major or minor depression			
Cox et al., 1996[73]	13	0.62 0.38-0.82	0.89 0.81-0.94
Cox et al., 1996[73]	12	0.76 0.53-0.92	0.81 0.73-0.88
Cox et al., 1996[73]	10	0.81 0.58-0.95	0.77 0.67-0.84
Beck and Gable, 2001[71]	10	0.59 0.43-0.73	0.86 0.78-0.92
Leverton and Elliott, 2000[75] (Bedford Criteria)	13	0.44 0.38-0.82	0.92 0.87-0.95
Leverton and Elliott, 2000[75]	10	0.69 0.41-.89	0.85 0.79-.0.90
BDI, Major depression			
Beck and Gable, 2001[71] (BDI-II)	21	0.56 0.31-0.78	1.0 0.97-1.0
Harris et al., 1989[74] (BDI)	21	0.32 0.13-0.57	0.99 0.95-1.0
	13	0.63 0.38-0.84	0.92 0.85-0.96
	11	0.68 0.43-0.87	0.88 0.82-0.94
BDI, Major or minor depression			
Beck and Gable, 2001[71] (BDI-II)	15	0.57 0.41-0.71	0.97 0.92-1.0
Whiffen, 1988[77] (BDI)	10	0.48 0.26-0.70	0.86 0.78-0.92
PDSS, Major depression			
Beck and Gable, 2001[71]	81	0.94 0.73-1.0	0.98 0.94-1.0
PDSS, Major or minor depression			
Beck and Gable, 2001[71]	61	0.91 0.79-0.98	0.72 0.62-0.80
CES-D, Major or minor depression			
Campbell and Cohn, 1991[32]	16	0.60 0.50-0.70	0.92 0.90-0.93
	21	0.43 0.33-0.54	0.97 0.95-0.98

BDI, Beck Depression Inventory; CES-D, Center for Epidemiological Studies – Depression Scale; CI, confidence interval; EPDS, Edinburgh Postnatal Depression Scale; PDSS, Postpartum Depression Screening Scale.

scores each produced slightly higher sensitivity and slightly lower specificity than did a cutoff of 21. Relative to the EPDS for major depression alone, Beck and Gable results showed sensitivities that remained substantially lower and specificities that appeared to be slightly higher, although wide CIs preclude a confident comparison.

For major or minor depression, two articles reported BDI test characteristics using different thresholds.[71,77] Beck and Gable,[71] using a cutoff of ≥ 15 on the BDI-II, reported a sensitivity of 0.57 and a specificity of 0.97. The BDI study by Whiffen employed a cutoff of ≥ 10 and reported a sensitivity of 0.48 and a specificity of 0.86.[77]

Figure 17a. Sensitivity of screening by Edinburgh Postnatal Depression Scale: postpartum period, major or minor depression

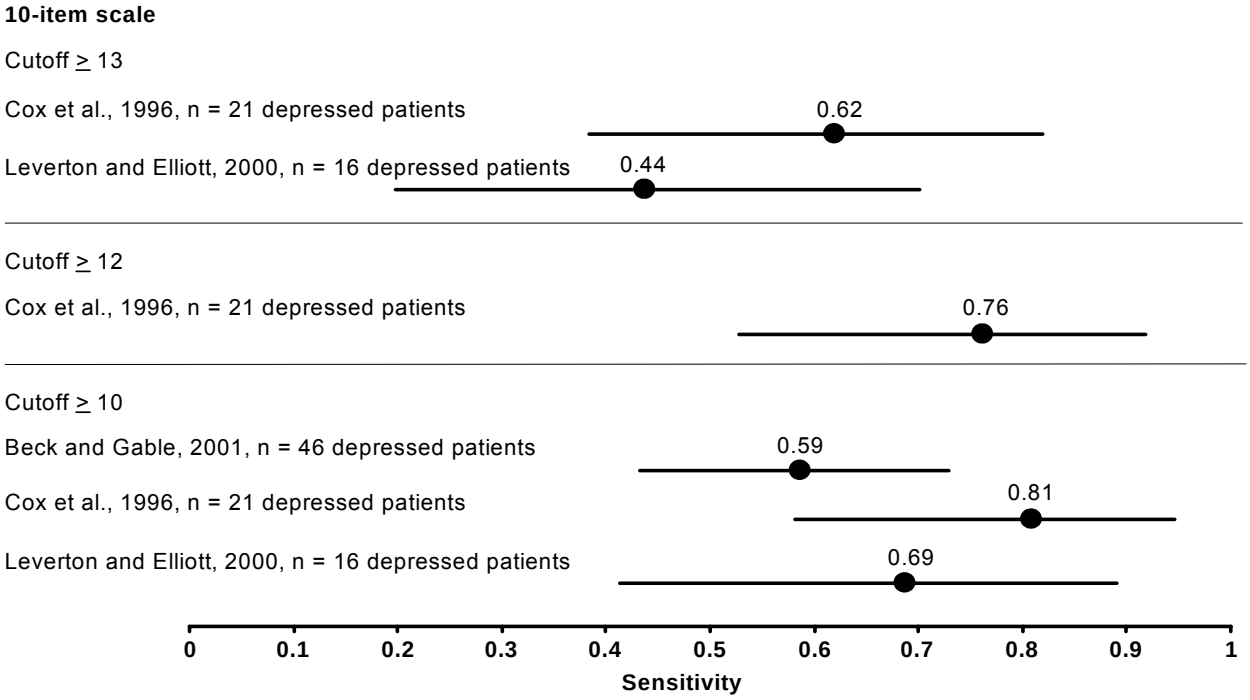


Figure 17b. Specificity of screening by Edinburgh Postnatal Depression Scale: postpartum period, major or minor depression

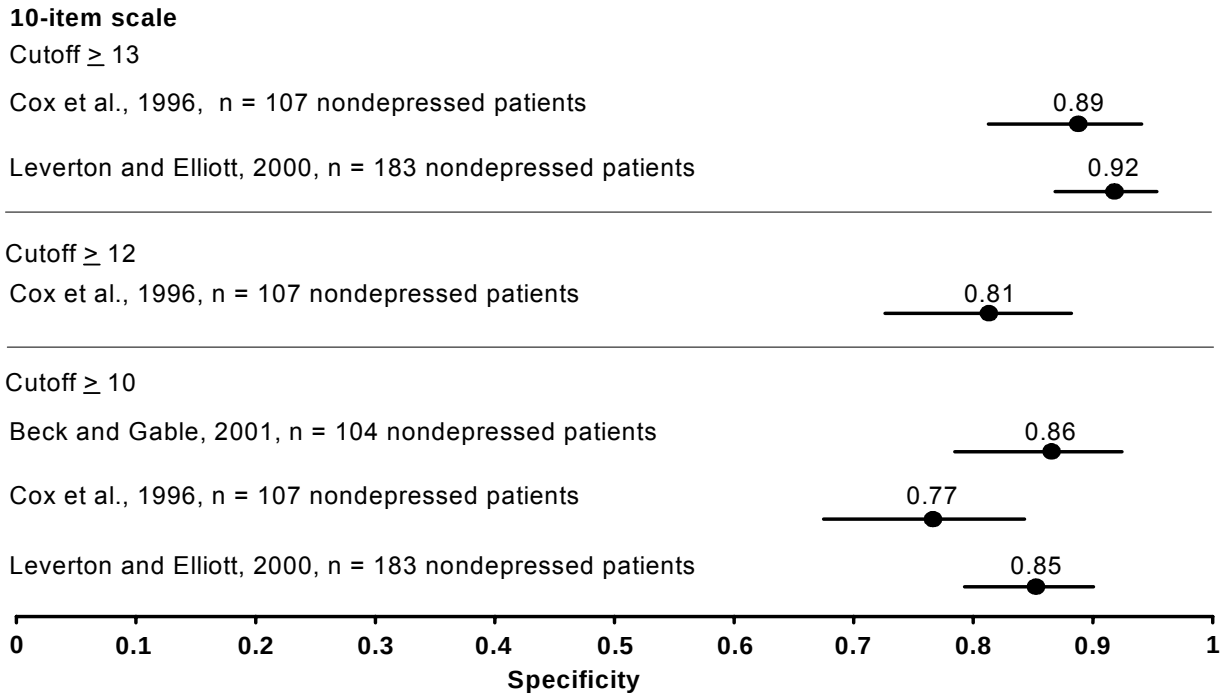


Figure 18a. Sensitivity of screening by BDI: postpartum period, major depression

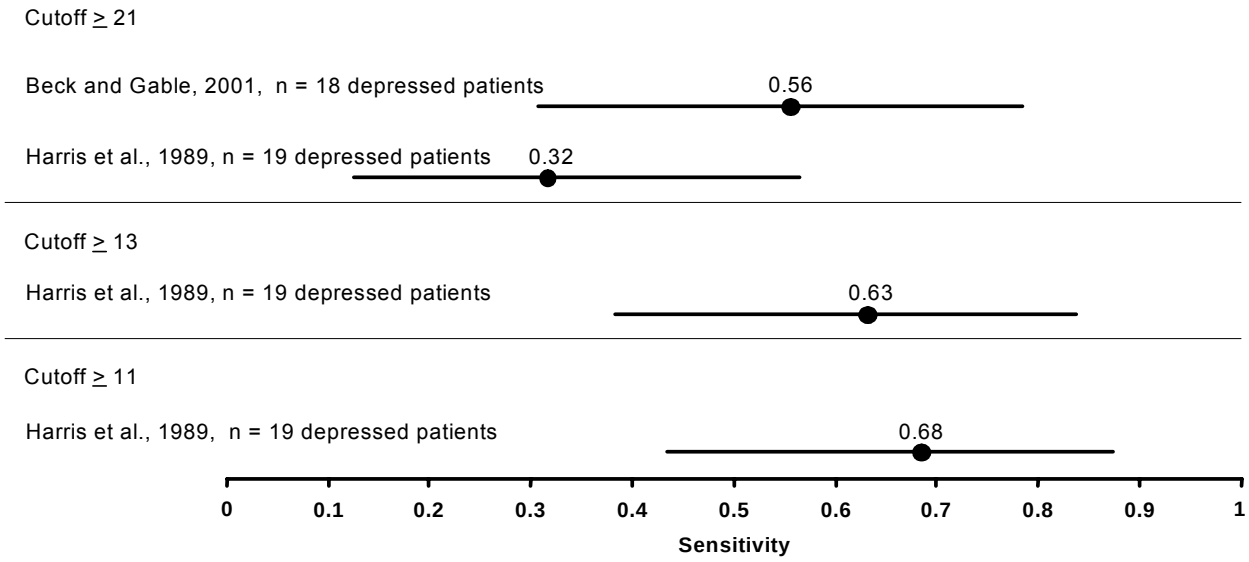
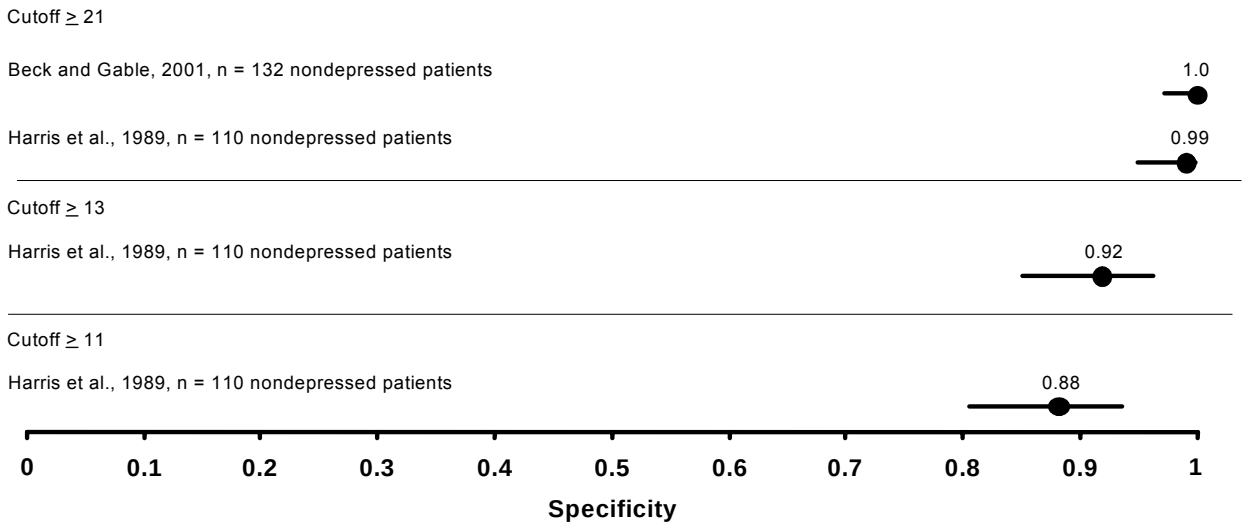


Figure 18b. Specificity of screening by BDI: postpartum period, major depression alone



Postpartum Depression Screening Scale. One study of the PDSS (150 patients) reported high sensitivity (0.94) and high specificity (0.98) for major depressive disorder alone at a cutoff of ≥ 80.[71] The investigators also reported lower sensitivity (0.91) and lower specificity (0.72) for major or minor depression using a cutoff of ≥ 60.

Center for Epidemiological Studies – Depression Scale. One study of the CES-D (1,007 patients) used two cutoff points (≥ 21 and ≥ 16).[32] It reported low sensitivity (0.60 and 0.43, respectively) and high specificity (0.92 and 0.97, respectively) for major or minor depression.

Summary of Results of Screening Instrument Review

The available evidence for both major depression alone and major or minor depression together is characterized by studies including markedly low numbers of depressed patients, a narrow racial and ethnic mix, varying cutoff points, and varying reference standards. These factors combine to preclude definitive conclusions or recommendations about screening instruments or thresholds.

Screening Instruments. For major depression alone, all screening instruments investigated (EPDS, BDI, PDSS) provided similarly high degrees of specificity at various cutoffs. Because of wide CIs, however, conclusions about sensitivity are more restricted. Heterogeneity among the studies limited our ability to synthesize these results quantitatively. In most instances, we could not obtain a more precise estimate. For an EPDS cutoff of ≥ 13 for patients with major depression alone, sensitivity estimates were combined in a meta-analysis to produce a point estimate of 0.91; however, heterogeneity precluded a meta-analysis for a specificity point estimate.

The EPDS and PDSS (with point estimates ranging from 0.75 to 1.0 at various cutoffs) appeared to be more sensitive than the BDI instruments (0.32 to 0.68 at various cutoffs), but the

wide CIs overlapped nearly completely. A recent meta-analysis of prevalence estimates found that, compared with structured clinical interviews, the EPDS produced statistically equivalent prevalence estimates whereas the BDI produced significantly higher estimates.[27] Together, these findings suggest that a positive screen with EPDS may be more clinically useful than screens with the other instruments.

For major or minor depression, sensitivity point estimates for each tool at each cutoff were consistently lower than those for major depression alone, although specificities were quite similar to those for major depression alone. We were able to synthesize EPDS results quantitatively at a cutoff of ≥ 13, producing a sensitivity point estimate of 0.54 with a wide CI (95% CI, 0.39 to 0.70) and a specificity estimate of 0.91 with a narrow CI (95% CI, 0.88 to 0.94). At an EPDS cutoff of ≥ 10, we were able to produce a pooled sensitivity estimate of 0.68 (95% CI, 0.58 to 0.78), but heterogeneity precluded a pooled analysis for specificity.

In short, estimates of specificity are relatively precise, but estimates of sensitivity are imprecise. This pattern of results prevents any substantive conclusions about the accuracy of these tools for identifying true positives. This imprecision can be attributed to the consistently low number of patients with a depression diagnosis, a fact reflected by a number of studies reporting 100 percent sensitivity, and it is a major limitation of the currently available data. Because of this imprecision, we cannot meaningfully compare sensitivities of screening instruments.

Cutoff Points. For an individual screening instrument, we cannot make any substantive conclusions about the use of a particular cutoff point. As noted above, the wide CIs for sensitivity prevent one from confidently distinguishing one sensitivity result from another. However, two further guides that bear directly on the choice of a threshold need to be considered before a particular threshold could be suggested.

First, the relative cost, or value, of errors in screening tests (false-negative compared to false-positive results) needs to be clarified. False-negative results (miss true depression) can lead to bad outcomes such as continued morbidity, costs of unnecessary tests, and similar effects. By contrast, false-positive results (identifying depression when it is not there), can lead to unnecessary time, effort, and financial cost for diagnostic workup as well as potential side effects of a treatment that is not indicated. If false-negative and false-positive results are equally bad, then a screening test should try to minimize both equally to identify the most effective cutoff.

If missing depression in a patient is worse than falsely identifying depression in a patient (i.e., a false-negative classification is worse than a false-positive one), then one would want a test that maximizes sensitivity and has the highest negative predictive value. Said another way, the preferred test would be one in which the greatest proportion of those screening negative do not have the disease. By contrast, if falsely identifying a patient as having depression (a false positive) is worse, then one would want a test that maximizes specificity and has the highest positive predictive value. Clinical intuition suggests that missing a diagnosis is worse than making an incorrect diagnosis. We could find no literature addressing the trade-off of false-positive versus false-negative diagnoses in this clinical situation.

A second important guide in choosing a cutoff is the prevalence of a disease in a particular population. Regardless of test characteristics, in populations in which the prevalence of depression is relatively high, the number of false-negative results is higher; in populations in which the prevalence is relatively low, the number of false-positive results is higher. Therefore, the choice of a test and cutoff may differ depending on whether the population has a higher

prevalence of depression (e.g., a high-risk postpartum clinic) or a lower prevalence (e.g., a healthy baby clinic). As a result, these three variables—sensitivity and specificity, the predictive value of screening errors (false positives versus false negatives), and the prevalence of the disease—must be clarified before clinicians or researchers can choose a specific test and related cutoff.

The above limitations notwithstanding, the tools we have reviewed above appear to be able to identify depressive illness in pregnant and postpartum women with a degree of accuracy similar to that for depression screen results in other nonpsychiatric settings. Screening results in primary care for a combined major or minor depression group are not available, but the results in primary care settings for major depression alone are similar to those reported for perinatal depression. For example, in a synthesis of depression case-identifying instruments in primary care settings using selection criteria similar to ours, Williams et al. reported a median sensitivity for major depression of 85 percent (range, 50 percent to 97 percent), and a median specificity of 74 percent (range, 51 percent to 98 percent).[90] This review included both women and men, which might explain the lower measures of accuracy; female gender appears to improve the accuracy of depression screens in primary care settings.[91]

Interpretation of Results

The small numbers of relevant articles limits our interpretation of the results. Given that most of the articles address the EPDS, we will use this instrument as an example. Because of the reports of 100 percent sensitivity in the prenatal tests of the EPDS (underscoring the very small number of prenatal depressed patients involved), we consider application of our results only to the postpartum population, and we draw on the prevalence data reported in Chapter 3 for KQ 1. We caution that, given the low numbers of depressed patients in the postpartum studies, the sensitivity estimates are likely to be inaccurate. Also, the majority of postpartum screens were performed 6 to 8 weeks after delivery, so the examples below apply only to that time period.

For major depression alone, the estimated point prevalence for the 6- to 8-week postpartum period is 6.8 percent, although the confidence interval around this estimate is wide. EPDS screens using the most commonly cited cutoff of 13 have a sensitivity of 91 percent and a specificity of approximately 95 percent. To illustrate this scenario, consider using this tool and cutoff for 1,000 patients. This EPDS screen would produce 62 true-positive cases and 6 false-negative cases, and 47 false-positive cases and 885 true-negative cases. The positive predictive value is 57 percent, meaning that the probability that a woman with a positive screen truly has major depressive disorder is slightly more than half. The negative predictive value (i.e., the probability that a woman with a negative screen would not have depressive illness) is 99 percent.

For major or minor depression, the estimated point prevalence from KQ 1 is 11.3 percent. EPDS screens tested for this population most commonly reported a cutoff of 10. This threshold at 6 to 8 weeks postpartum has a sensitivity of 68 percent and a specificity of approximately 80 percent. For 1,000 patients, the screen would produce 77 true-positive cases and 36 false-negative results, and 177 false-positive cases and 710 true-negative cases. The positive predictive value is 30 percent, and the negative predictive value is 95 percent.

Discussion

Conclusions

Very little is known about the accuracy of depression screening tests in pregnant and postpartum women. The available evidence is limited in several ways. It has a very narrow racial and ethnic mix. Study samples have prevalence rates of depression that are, by design, somewhat higher than our best estimate prevalence rates from KQ 1 (which would produce a higher positive predictive value). Most important, the available data involve small numbers of depressed patients. We could not address the limits of the small numbers of depressed patients using meta-analytic procedures. Case definitions, reference standards, screening tools, and screening thresholds all varied across the studies, and the heterogeneity of study methods constrained our ability to synthesize the data and obtain pooled estimates.

Despite these limitations, the available evidence does indicate that depression screens are feasible to administer in perinatal settings. It also suggests that the estimates of sensitivity and specificity, although limited, appear equivalent to those that have been reported in primary care settings. In particular, specificity is relatively good, suggesting a relatively good positive predictive value.

Future Research

Further studies in this area need to standardize the above parameters we have examined in this chapter (instruments and, in particular, cutoff points), involve a more representative mix of racial and ethnic groups, test the screening tools in populations with a frequency of depression more reflective of the actual prevalence, and include a larger number of depressed patients to clarify the accuracy of depression screening tools and make them more relevant to the population of interest. Given the currently available evidence, we offer six future research recommendations.

First, subsequent studies on the test characteristics of screeners must be designed with sample size estimates that take into account prevalence and that project a reasonable width of sensitivity confidence intervals for the particular illness. For example, studies would need to screen 1,000 women to identify 34 with major depression or 110 with major or minor depression. This sample size might be enough for precise estimates for women with major or minor depression as a group, but it may not be enough for precise estimates for major depression alone.

Second, the sample should represent the target population. Specifically, subsequent studies need to provide a more representative racial and ethnic mix. In addition, studies should incorporate a range of other demographic variables that could influence screening performance, such as socioeconomic status measures, and assess the screening tools in these subpopulations.

Third, as in the Beck and Gable study,[71] subsequent studies should assess and directly compare multiple screening instruments. This design provides a head-to-head comparison that allows researchers and clinicians to understand which screening instruments are more accurate than others in different settings.

Fourth, studies evaluating both the risks and benefits of screening, specifically assessing the relative cost of false-negatives and false-positive results, will provide insights on how to consider target sensitivity and specificity when attempting to maximize cost-effectiveness.

Fifth, subsequent depression screening studies should carefully consider whether to target major depression alone, for which beneficial treatments clearly exist, or the traditional combined category of major or minor depression, a heterogeneous group for which treatment benefit is unclear. Our results suggest that the sensitivity of screening instruments is generally greater for the major depression alone group.

Sixth, the bulk of the screening studies we reviewed were conducted in the first 3 months postpartum. Subsequent studies should examine screening not just in the first 3 months postpartum but also at 6 weeks, 6 months, and 12 months postpartum. If peak prevalence and incidence occur within the first 6 weeks, the obstetrics clinic is a prime place to target resources for such a program. If, however, peaks occur after this time, most postpartum women will have completed follow-up care with an obstetrician, so programs in an obstetrics clinic may be less helpful. In this case, it is possible that programs targeting new mothers in family medicine, internal medicine, or pediatric clinics might be more effective.

Use in Clinical Settings

In the interim, what is a clinician to do? The best available evidence supports the conclusion that screening instruments with reasonable test characteristics appear feasible to use in a perinatal population with a depression prevalence between 5 percent and 10 percent. Given that use of the tools likely carries low risk, and that they all have reasonable specificity (and, thus, a reasonable positive predictive value), the selection of a tool would be guided by an interest in maximizing sensitivity. For the category of major or minor depression, sensitivity estimates were quite similar for all instruments. However, for major depression alone, sensitivity estimates for the EPDS and PDSS appear to be higher than those for the BDI. The standard cutoffs of ≥ 13 for the EPDS and ≥ 81 for the PDSS appear to be reasonable thresholds.

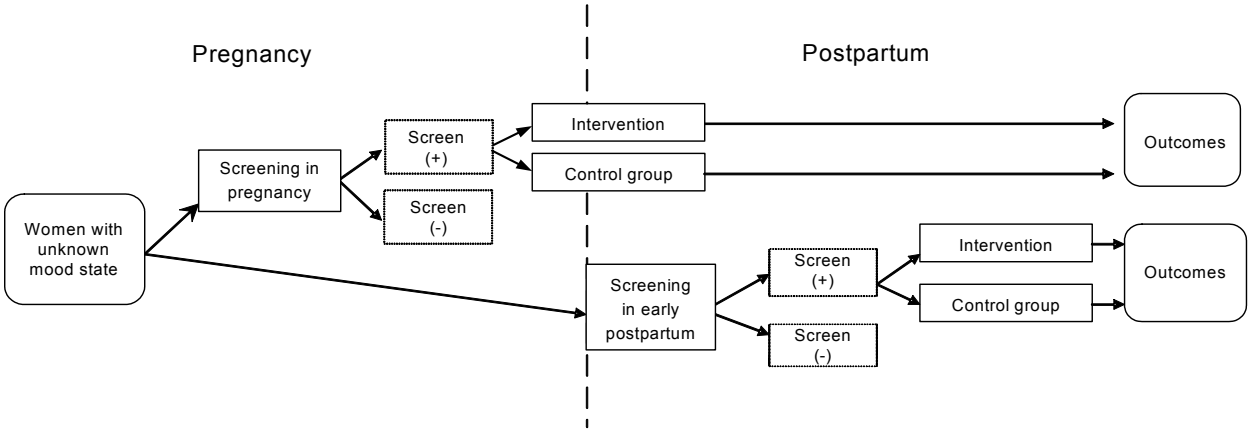
Having an instrument that can accurately identify women at risk of having perinatal depression is an important and necessary link in improving the clinical outcomes of women with perinatal depression: women who may benefit from a depression intervention first need to be recognized. Nonetheless, it remains merely an initial step. A more important question is whether screening pregnant or postpartum women to identify those at risk of having depression, and subsequently providing an intervention, ultimately leads to improved outcome. We address this key question in our next chapter.

Chapter 5. Impact of Depression Screening and Interventions on Patient Outcomes

Introduction

In agreement with the Safe Motherhood Group and the Agency for Healthcare Research and Quality (AHRQ), we directed part of our work to Key Question (KQ) 3: Does prenatal or early postnatal screening for depressive symptoms with subsequent intervention lead to improved outcomes? That is, does screening for depression during pregnancy or the postpartum period and implementing an intervention improve outcomes related to maternal depressive symptoms? To address KQ 3, we developed an analytic framework (Figure 19), which begins (left side) by identifying a cohort of women with unknown mood state, continues through implementation of a formal screening of the cohort either during pregnancy or in the postpartum period, and (right side) ends with studies of an intervention to assess how it may affect outcome measures of postpartum depression.

Figure 19. Causal Pathway for Key Question 3 on Screening and Treatment Outcomes



As described in this chapter, screening can be done in various settings and with various instruments (as discussed for KQ 2). Interventions are both nonpharmacologic (e.g., counseling and behavioral intervention programs aimed at mothers or, in some cases, both parents or mother-infant dyads) and pharmacologic (e.g., antidepressants). These interventions can be

Note: Appendixes and Evidence Tables cited in this report are provided electronically at http://www.ahrq.gov/clinic/epcindex.htm.

implemented in various outpatient settings (e.g., clinics, homes) and delivered by various types of health professionals, and they may be group efforts or one-on-one activities.

Methods

Chapter 2 documents the methods we used to conduct literature searches and title and abstract or full article reviews. We did not identify any studies that specifically examined the cascade of screening-treatment-outcomes. Thus, we do not have any direct evidence pertaining to KQ3.

All the trials included for KQ3 are treatment studies that had a screening component (either a formal depression screening instrument or other type of screen that identified women at risk of a depressive illness). We included studies conducted worldwide in developed countries where the population could be generalized to pregnant and postpartum women in the United States, regardless of the language spoken. We also included both randomized controlled trials (RCTs) and prospective cohort studies. Additionally, for inclusion in KQ 3, patients were identified by a screen done either during pregnancy or during 12 months postpartum and considered to be "at risk" of having a depressive illness.

We excluded all case-control studies and studies in which patients had had a documented current depressive episode before the initial screening. Furthermore, we excluded two studies that had originally been reviewed for the feasibility study, one because it did not use any screening[92] and one because it had no depression severity outcome.[93]

We attempted to synthesize the results of the included studies quantitatively, but the study methods (screening instruments, type of intervention, intensity of intervention, outcomes measured) were so heterogeneous that a combined result would have little meaning. We also attempted to compare effect sizes in an exploratory analysis of the various studies, but the data necessary to compute these were not available.

Appendix B presents the quality rating form used for articles considered for KQ 3. The total possible score for these studies was 29. We characterized studies with scores of 20 or greater as good, those with scores between 15 and 19 as fair, and those with scores of 14 and below as poor. The domains and maximum points possible for each domain are as follows:

- Reporting (domain score of 11): 10 items covering study aims, measures, patient populations, findings, and statistical presentation; each scored yes or no (1 or 0), except for an item concerning principal confounders that was scored yes, partially, or no (2, 1, or 0, respectively).

- External validity (domain score of 3): Three items relating to representativeness of populations from which people were recruited and of settings and clinicians that treat such patients; each scored yes or no (1 or 0).

- Internal validity–bias (domain score of 7): Seven items relating to issues such as blinding subjects and outcomes assessors, follow-up periods, appropriate statistical tests, and use of reliable and valid outcome measures; each scored yes, no, or unable to determine (1, 0, or 0, respectively).

- Internal validity–confounding (domain score of 6): Six items relating to sources of intervention and control groups, randomization of study subjects and concealment of allocation, adequacy of adjustments for confounding, and loss to follow-up; each scored yes, no, or unable to determine (1, 0, or 0, respectively).

- Power (domain score of 2): One item about use of power analysis to determine sample size; scored no, yes for one measure, or yes for two or more measures (0, 1, and 2, respectively).

Results

We reviewed a total of 60 titles and abstracts or full articles for KQ 3 drawn from several searches done for the feasibility study and later for this update. Ultimately, we retained 15 studies that met our inclusion criteria. Table 17 summarizes the major characteristics of the 15 included studies, Table 18 presents the results of our quality ratings, and Table 19 shows the results of the various depression assessments made among cases and controls.

The types and frequency of screening measures and the types of interventions applied varied appreciably among the studies we reviewed. Of the 15 studies retained for the full study, 4 examined intervention efforts for which screening had been done in the prenatal period and 11 studies examined screening and interventions in the postpartum period. The remainder of this section reports on the studies in these two main categories.

Prenatal Studies

Of the four studies examining screening, interventions, and outcomes in the prenatal period,[94-97] three were RCTs[94,95,97] and one was a nonrandomized controlled trial.[96] All four studies (published between 1995 and 2001) were set in prenatal clinics. Sample sizes for screening ranged from 37 to 209, for a total population of 473 women. The types of screening instruments used to identify patients with depressive symptoms differed among these studies; similarly, the outcome measures differed, although three studies used the Edinburgh Postnatal Depression Scale (EPDS) as one measure. All four studies implemented some type of psychological intervention, generally characterized as group classes or sessions relating to prenatal preparation, skills, and perinatal support. One study was considered fair; the other three were poor.

Brugha et al. screened 209 women with a modified General Health Questionnaire Depression Score (GHQ-D) to study the effect of six weekly prenatal group therapy classes called "Preparing for Parenthood" compared to routine prenatal care.[95] In this study, which we graded as fair, the program aimed to increase social support and problem-solving skills. Outcome measures that assessed maternal mood and depressive symptoms at 3 months postpartum included the GHQ-D (cutoff ≥ 2), the EPDS (cutoff score of > 11), and a Schedule for Clinical Assessment in Neuropsychiatry (SCAN, related to the International Classification of Diseases [ICD], version 10). Assignment to the intervention group did not significantly improve postpartum depression. On the GHQ-D, 26 percent of the intervention group and 22 percent of the control group scored at or above 2, with an adjusted odds ratio of 1.19 (95% confidence

interval [CI], 0.59 to 2.37). On the EPDS, 16 percent of the intervention group and 19 percent of the control group scored above 11 (an adjusted odds ratio of 0.83 (95% CI, 0.39 to 1.79).

In the earliest study we included in this group (1995), Stamp et al. used a study-specific, modified prenatal questionnaire as a screening instrument and assigned 129 patients to either two prenatal group classes plus one postpartum class (at 6 weeks) or routine care.[94] Outcome measures were the EPDS at 6 weeks and 6 months postpartum, using a cutoff point of > 12 for major depression and > 9 for major or minor depression. The intervention did not significantly reduce rates of postpartum depression on either measure. For example, at 6 weeks, 13 percent of the intervention group and 17 percent of the control group had EPDS scores greater than 12; at 6 months, the figures were 15 percent and 10 percent, respectively.

Table 17. Major characteristics of studies of screening with interventions for prenatal or postpartum depression

Author, Year	Country	Study Design	Sample Size	Setting	Type of Screening	Type of Intervention
Screening during Pregnancy						
Brugha et al., 2000[95]	UK	RCT	209	Prenatal clinic	Modified GHQ-D	Structured group prenatal preparation classes
Elliott et al., 2000[96]	UK	Nonran-domized controlled trial	98	Prenatal clinic	Leverton Questionnaire Crown Crisp	Structured group prenatal preparation classes
Stamp et al., 1995[94]	Australia	RCT	129	Prenatal clinic	Modified prenatal questionnaire	Perinatal support group
Zlotnick et al., 2001[97]	US	RCT	37	Prenatal clinic	Screening survey	Four prenatal therapy/skills groups
Screening during Postpartum Period						
Armstrong et al., 1999[98]	Australia	RCT	181	PP hospital ward	Adverse family risk factors from Brisbane Evaluation of Needs Questionnaire	Regular home visits by child-health nurses
Chabrol et al., 2002[99]	France	RCT	859 screened 258 randomized	PP hospital ward	EPDS	One CBT prevention group during the delivery hospital stay followed by an at-home CBT- based program in women with major depression
Chen et al., 2000[100]	Taiwan	RCT	414 screened 115 randomized	PP hospital ward	Taiwanese BDI Measures of support	Four weekly PP support group sessions
Dennis 2003[101]	Canada	RCT	501 screened 44 randomized	Child immuniza-tion clinics	EPDS	Telephone-based peer support
Fleming et al., 1992[102]	Canada	Nonran-domized controlled trial	781 screened 152 enrolled	PP hospital ward	EPDS CES MAACL	PP social support group
Hiscock and Wake, 2002[103]	Australia	RCT	155 screened 99 randomized	Child-health center	EPDS	Infant sleep intervention group
Honey et al., 2002[104]	UK	RCT	45 randomized	Mother/baby clinic	EPDS	Psycho-educational group

Table 17. Major characteristics of studies of screening with interventions for prenatal or postpartum depression (continued)

Author, Year	Country	Study Design	Sample Size	Setting	Type of Screening	Type of Intervention
Horowitz et al., 2001[105]	US	RCT	1,215 screened 122 randomized	Community sample of PP women	EPDS	Coached behavioral intervention to promote maternal-baby interaction
Onozawa et al., 2001[106]	UK	RCT	59	PP hospital ward	EPDS	Infant massage plus support group
Wisner and Wheeler, 1994[107]	US	Open trial	23	PP hospital ward	Prior history of PPD	Antidepressant medication
Wisner et al., 2001[108]	US	RCT	581 screened 56 randomized	PP hospital ward	Prior history of PPD	Antidepressant medication

BDI, Beck Depression Inventory; CBT, cognitive behavioral therapy; CES, Current Experience Scale; EPDS, Edinburgh Postnatal Depression Scale; GHQ-D, General Health Questionnaire Depression Score; MAACL, Multiple Affect Adjective Checklist; PP, postpartum; PPD, postpartum depression; RCT, randomized controlled trial.

Table 18. Quality rating of studies of screening with interventions for prenatal or postpartum depression

Author, Year	Reporting (11)	External Validity (3)	Internal Validity–Bias (7)	Internal Validity–Confounding (6)	Power (2)	Total Score (29)
Screening during Pregnancy						
Brugha et al., 2000[95]	7	0	3	6	1	17
Elliott et al., 2000[96]	5	0	2	4	0	11
Stamp et al., 1995[94]	6	0	1	5	1	13
Zlotnick et al., 2001[97]	5	0	4	3	0	12
Screening during Postpartum Period						
Armstrong et al., 1999[98]	9	1	5	3	1	19
Brisco et al., 1989[93]	6	1	4	3	0	14
Chabrol et al., 2002[99]	8	0	2	5	0	15
Chen et al., 2000[100]	8	0	3	3	0	14
Dennis, 2003[101]	9	2	5	6	0	22
Fleming et al., 1992[102]	6	0	3	2	0	11
Hiscock and Wake, 2002[103]	7	1	5	5	1	19
Honey et al., 2002[104]	9	0	3	4	0	16
Horowitz et al., 2001[105]	7	0	4	3	1	15
Onozawa et al., 2001[106]	10	0	3	3	0	16
Wisner and Wheeler, 1994[107]	8	0	3	1	0	12
Wisner et al., 2001[108]	7	0	6	4	1	18

Note: Maximum possible score in parentheses.

Table 19. Major outcomes of studies of screening and interventions for perinatal depression

Author, Year	Type of Intervention	Outcome Measures	Significant Differences between Intervention and Control Group
Screening during Pregnancy			
Brugha et al., 2000[95]	Structured group prenatal preparation classes	GHQ-D EPDS SCAN	No significant differences on any measure
Elliott et al., 2000[96]	Structured group prenatal preparation classes	EPDS PSE	Intervention group had significantly lower EPDS scores in first time mothers; no significant difference on PSE for diagnosis of major depression
Stamp et al., 1995[94]	Perinatal support group	EPDS	No significant differences on this measure
Zlotnick et al., 2001[97]	Four prenatal therapy/skills groups	BDI SCID	Intervention group had a significantly greater change over time; at follow-up, intervention group had a significantly lower level of maternal depression
Screening during Postpartum Period			
Armstrong et al., 1999[98]	Regular home visits by child health nurses	EPDS PSI Child health HOME	For secondary outcomes, intervention group had significantly lower depression scores and a positive effect on parent-infant interaction
Chabrol et al., 2002[99]	One CBT-based prevention group during the PP hospitalization, followed by an at-home CBT-based program in women with major depression	EPDS HAM-D BDI	Intervention group had significant reductions in frequency of depressive symptoms
Chen et al., 2000[100]	Four weekly PP support group sessions	BDI PSS ISEL	Intervention group had significant lower rates of depression and rates of perceived stress and more interpersonal support
Dennis, 2003[101]	Telephone-based peer support	EPDS	Intervention group had significantly lower EPDS scores
Fleming et al., 1992[102]	PP social support group	EPDS CES	No significant differences on any measure
Hiscock et al., 2002[103]	Infant sleep intervention (controlled crying) group	EPDS Maternal and infant sleep quality Maternal stress	Intervention group members with higher depression scores at baseline had significantly greater improvement in EPDS scores and reported improvements in sleep quality
Honey et al., 2002[104]	Psycho-educational group	EPDS	Intervention group had significant reductions in depressive symptoms

Table 19. Major outcomes of studies of screening and interventions for perinatal depression (continued)

Author, Year	Type of Intervention	Outcome Measures	Significant Differences between Intervention and Control Group
Horowitz et al., 2001[105]	Coached behavioral intervention to promote maternal-baby responsiveness	BDI-II DMC	No significant differences for maternal depression; intervention group showed significantly better mother-infant responsiveness
Onazawa et al., 2001[106]	Infant massage classes plus support group	EPDS Videotape of mother-infant interaction	No significant differences for maternal depression; intervention group showed significant improvements in mother-infant interaction
Wisner and Wheeler, 1994[107]	Antidepressant (nortriptyline)	Clinical interview IDD	Intervention group had significantly lower proportion of new episodes of major depression
Wisner et al., 2001[108]	Antidepressant (nortriptyline)	RDC HAM-D	No significant differences in the rate of recurrence

BDI, Beck Depression Inventory; CBT, cognitive behavioral therapy; CES, Current Experience Scale; DMC, Dyadic Mutuality Code; EPDS, Edinburgh Perinatal Depression Scale; GHQ-D, General Health Questionnaire–Depression Subscale; HAM-D, Hamilton Depression Rating Scale; HOME, Home Observation for Measurement of the Environment; IDD, Inventory to Diagnose Depression; ISEL, Interpersonal Support Evaluation List; PSI, Parenting Stress Index; PSE, Present State Examination; PSS, Perceived Stress Scale; RDC, Research Diagnostic Criteria; SCAN, Schedule for Clinical Assessment in Neuropsychiatry; SCID, Structured Clinical Interview for Diagnosis.

Elliott et al. screened 98 women with the Leverton Questionnaire and the depression, anxiety, and somatic subscales of the Crown Crisp Experiential Index.[96] The authors studied a preventive group of psychosocial intervention versus routine care; they also looked at differences between first- and second-time mothers. The structured group intervention was conducted once per month for 5 months during the prenatal period (starting at 24 weeks) and for 6 months postpartum. Outcome measures included the EPDS and the Present State Examination (PSE), as well as a self-rating questionnaire, at 3 and 12 months postpartum. For first-time mothers, the median EPDS score was significantly lower in the intervention group (Mann-Whitney one-tailed test, $P = 0.005$); for second-time mothers, the median EPDS did not differ significantly between the two groups. The PSE served as a formal diagnosis of depression, and the investigators reported no significant differences in diagnosis of major depression. When the authors included cases of borderline depression or "minor depression" in the analysis, first- time mothers in the intervention group were significantly less likely to have a diagnosis of depression than controls (19 percent and 39 percent, respectively, Chi-square = 2.64, one-tailed test, $P < 0.05$). PSE scores did not differ significantly between groups in second-time mothers.

In the only study in this category done in the United States, Zlotnick et al. used the Beck Depression Inventory (BDI) and a Diagnostic and Statistical Manual (DSM-IV) Structured Clinical Interview for Depression (SCID) as a positive screen among women of low socioeconomic status (SES).[97] They excluded patients who met criteria for a current episode of major depression based on the SCID. A total of 37 patients with a positive screen were assigned to either a four-session Interpersonal-Therapy-Oriented Group (given weekly) or to a usual-care group. Outcome measures included the BDI before and after the intervention and the SCID at 3 months postpartum. Women in the intervention group had a significantly greater change in their BDI scores from baseline than did those in the control group ("pre" versus "post" intervention Beck scores were 13.0 and 8.4, respectively, for the treatment group). In contrast, the control

group "pre" versus "post" intervention scores were 9.2 and 11.3, respectively, suggesting that they got worse over time. The change between the intervention and the control group was significant (t-test = 3.50; df = 33; $P = 0.001$). In addition, women in the intervention group had a significantly decreased rate of major depression during the postpartum period as measured by the SCID at 3 months postpartum; no women in the intervention group and 33 percent of women in the usual-care group developed postpartum depression ($P > 0.02$).

These four small studies of programs for women identified by screening prenatally did not, collectively, produce many positive results from the various psychosocial interventions as compared with usual care. All of these studies scored poor on external validity (0 of 3 points), and two of the four had 0 of 2 points for power. The four studies did, at best, only a fair job of reporting data (from 5 to 7 of 11 points). For bias, the study scores ranged from 1 to 4 of 7 possible points; for confounding, they ranged from 3 to 5 of 6 points. Given the heterogeneity in populations, the screening instruments and cutoff points for defining "at-risk" individuals, the interventions themselves, and the outcome measures used, we cannot draw any overall conclusions about the utility of such programs.

Postpartum Studies

Of the 11 studies examining screening and intervention outcomes only in the postpartum period,[98-108] eight were RCTs published between 1992 and 2003,[98,100,101,103-106,108] and three were controlled trials published between 1992 and 2002.[99,102,107] Sample sizes ranged from 23 to 1,215, for a total population of 4,289 women.

As with the prenatal screening studies, the screening instruments used to identify patients with depressive symptoms differed among the postnatal studies, although the EPDS was used in the majority of studies and two studies by the same investigator team used "prior history of postpartum depression." The treatment interventions also differed considerably. Nine of these studies involved various behavioral and psycho-educational programs or other innovative activities (e.g., infant massage or infant sleep interventions); two involved tests of antidepressants. Unlike the prenatal studies, the settings varied from postpartum hospital wards to child-health and immunization clinics. Finally, the outcome measures also varied across these studies, but the EPDS was most commonly used (in seven studies). We graded one study good, seven fair, and three poor.

Behavioral and Psychosocial Interventions. Of the nine studies in this subgroup, one was conducted in the United States; the remainder were in Australia (two studies), Canada (two studies), the United Kingdom (not otherwise specified, two studies), and France and Taiwan (one study each). Using "number randomized or enrolled" as the metric, the sample sizes ranged from 45 to 859. We describe the studies below according to quality grade and sample size.

In a recent study rated good that randomized participants for the intervention (not screening), Dennis et al. screened 501 women recruited from child immunization clinics between 8 and 12 weeks postpartum.[101] Inclusion criteria included the mother's being at least 18 years of age, having a singleton birth, and delivering a full-term infant. Women were screened using the EPDS (cutoff score > 9). The 44 women with a positive screen were randomized to a "mother-to-mother" peer support telephone intervention or to routine care. The outcome measures were the EPDS at 4 and 8 weeks after randomization. The women in the intervention group had

significantly lower EPDS scores than those in the control group: 15 percent of the intervention group and 52.4 percent of the control group had an EPDS > 12 at 8 weeks ($P = 0.02$).

In the largest of the seven studies rated fair, Chabrol et al. screened 859 women and identified 258 who were at risk based on an EPDS > 9 on day 2 or 3 postpartum.[99] They assigned these 258 women randomly to receive a cognitive behavioral therapy (CBT) (n = 130) intervention or to routine care (n = 128) during the postpartum hospitalization. CBT is a form of psychotherapy that actively examines how cognitions influence emotions or affect and involves active exploration, clarification, and testing of the patient's perceptions and beliefs.[109] Outcome measures for the prevention intervention included the EPDS (cutoff ≥ 11) taken at 4 to 6 weeks postpartum.

Women in the CBT group who continued to have positive screens on the EPDS (defined as EPDS score ≥ 11) at 4 to 6 weeks were assessed for major depression in a clinical interview using the Mini-Neuropsychiatric Interview (MINI) and DSM-IV criteria. Those with major depression were offered an at-home CBT program for five to eight additional sessions. These women were then compared with women with probable major depression in the control group at 10 to 12 weeks using the EPDS, Hamilton Depression Rating Scale (HAM-D), and the BDI. Women in the control group received one initial home visit assessment but then received only weekly telephone checks.

The study results demonstrated that women in the prevention intervention group had significant reductions in the frequency of depressive symptoms. At 4 to 6 weeks postpartum, 30.2 percent of those in the CBT group versus 48.2 percent in the control group ($P = 0.0067$) were still depressed (based on an EPDS score of ≥ 11). Additionally, the intensity of depressive symptoms measured by the mean score on the EPDS was significantly lower in the prevention group than in the control group: mean EPDS scores, respectively, of 8.5 (standard deviation [SD] 4) and 10.3 (SD 4.4) (t-test = 3.06, df = 209, $P = 0.0024$); the analyses indicate a medium effect size (ES, 0.42). At 10 to 12 weeks postpartum after completion of the home-based CBT intervention, women in the intervention group had significantly lower scores on all measures of depressive symptoms (HAM-D, BDI, EPDS) than did those in the control group. Specifically, the intervention and the control group mean scores were as follows: HAM-D, 5.7 versus 16.2 (t-test = 8.4, $P < 0.0001$); BDI, 4.7 versus 15.7 (t-test = 9, $P < 0.0001$); and EPDS, 5.9 versus 13.7 (t-test = 7.7; $P < 0.0001$).

Armstrong et al. screened 181 women with good literacy skills in the immediate postpartum period by asking about a history of trauma or abuse or a positive screen for adverse family characteristics on the Brisbane Evaluation of Needs Questionnaire.[98] Women were randomized to receive 6 months of home visits by a child-health nurse or routine primary care. Primary outcome measures involved measures of child health, parental and family functioning (measured by the EPDS [cutoff > 12] and the Parenting Stress Index [PSI]), quality of the home environment (HOME assessment), and satisfaction with community services. All assessments were administered immediately postpartum.

If we focus primarily on scores of maternal depression and functioning at 6 weeks postpartum, women in the intervention group had significantly lower depression scores than the control group: 5.8 percent in the intervention group and 20.7 percent in the control group ($P = 0.003$) with EPDS > 12. Additionally, women in the intervention group had significantly lower (better) PSI scores at 6 weeks than controls (15.3 versus 38.4, $P < 0.001$). The investigators also reported that the total HOME score differed significantly between groups: 28.34 for the intervention group versus 25.51 for the control group ($P < 0.001$), providing

evidence for the positive effect the intervention had on influencing parent-infant interaction and the home environment for the child.

Hiscock and Wake recruited 155 women from a child-health center at 7 to 8 months postpartum and screened with the EPDS (cutoff > 12).[103] Other inclusion criteria included reported child sleep problems. Of these women, 99 were considered depressed (baseline EPDS ≥ 10) and randomly assigned to either an infant sleep intervention group or a usual-care group. The infant sleep intervention comprised three private sessions (one session every 2 weeks) held at the local child-health center where sleep management plans were discussed, including an emphasis on controlled crying (where parents responded to their infants' crying at increasing time intervals, allowing the infant to fall asleep unaided). At the 10- to 12-month follow-up assessment, outcome measures included the EPDS, measures of sleep quality, and measures of maternal stress and coping.

The results of this study were mixed. Women who began with higher (worse) scores of depression at baseline had a significantly greater improvement in their EPDS scores than did those in the control group. At the 10-month follow-up, women in the intervention group had a 6.0 point decrease (95% CI, 7.5 to 4.0) in EPDS score compared to a 3.7 point decrease (95% CI, 4.9 to 2.6) in the control group ($P = 0.01$). At the 12-month follow-up visit, the intervention group had a 6.5 point decrease (95% CI, 7.9 to 5.1) in EPDS score compared to a 4.2 point decrease (95% CI, 5.9 to 2.5) in the control group ($P = 0.04$). Also, at the 10-month follow-up, women in the intervention group reported improvements in their own sleep quality, including being more likely than control mothers to rate their own sleep quality as "very good" and less likely to rate it as "very bad" (Chi square = 9.93; $P = 0.02$). They also reported having "enough sleep" and were less likely to have "not enough" sleep (Chi square = 8.11, $P = 0.04$).

Horowitz et al. screened 1,215 women at 2 to 4 weeks postpartum with the EPDS (cutoff > 10). Women with positive screens (n = 122) were randomly assigned to either an interactive coaching intervention or a control group. The coached behavioral intervention was designed to promote maternal-infant responsiveness. All women in the study (both intervention and control groups) received three home visits when their infants were 4 to 8 weeks, 10 to 14 weeks, and 14 to 18 weeks of age; the women in the intervention group practiced the coaching intervention during these visits. Outcome measures included the BDI-II for maternal depression and, secondarily, the Dyadic Mutuality Code (DMC), a measure of the level of responsiveness in the maternal-infant relationship. Responsiveness was defined as "the mother's ability to accommodate to her infant's behavior and to give it meaning through regulation of her own behavioral responses" (p. 326). The intervention and control groups did not differ significantly in terms of maternal depression scores (BDI-II) at any time period. The DMC showed a significantly better outcome for mother-infant responsiveness for the treatment group ($P = 0.06$).

Onozawa et al. screened 581 primiparous women with the EPDS (cutoff ≥ 13) at 4 weeks postpartum.[106] Of 91 women who had a positive screen, 59 agreed to participate in the study. Participants were randomized to either a 5-week infant massage class with a support group or the support group only. The 1-hour infant massage class (approved by the International Association of Infant Massage) taught parents the techniques of infant massage by encouraging parents to observe and respond to their infants' body language and cues and to adjust their touch accordingly. Outcome measures included maternal depression on the EPDS (cutoff ≥ 13) at 4 weeks and 2 months postpartum and a videotaped mother-infant interaction that assessed the mother's attitude toward the infant, the infant's response to the mother, and the overall quality of the interaction. At 14 weeks postpartum, EPDS scores had fallen for both groups (reported as a

change in median EPDS score from baseline to final visit), but the intervention group demonstrated a significantly greater change in scores than did the control group (intervention group baseline of 15.0 and final visit score of 5.0, versus the control group score of 16.0 at baseline to 10.0 at the final visit; $P = 0.03$). Additionally, significant improvements in all aspects of mother-infant interaction as measured by the videotape were seen only in the massage group ($P = 0.0004$).

Honey et al. used the EPDS (cutoff > 12) to screen postpartum women recruited through mother-baby clinics but assessed at home.[104] The 45 women with a positive screen on the EPDS were randomly assigned to either an 8-week psycho-educational group (PEG) or to a routine care group. Outcome measures included the EPDS (cutoff > 12) after completion of the PEG and at a 6-month follow-up. At the end of the 8-week assessment interval, the women in the PEG did not differ significantly from those in the routine-care group. By contrast, at the 6-month follow-up assessment, the percentage of women scoring below the EPDS cutoff for a probable major depressive episode was significantly higher in the PEG group than in the routine-care group (65 percent versus 36 percent, Chi square = 3.75; $P \leq 0.05$). An additional analysis demonstrated that the use of antidepressant medication during the study had no impact on the improvement in mood observed at the 6-month follow-up assessment.

Fleming et al. screened 781 primiparous women with full-term deliveries and no psychiatric history during their first 2 weeks postpartum using the EPDS (cutoff ≥ 13), the Current Experience Scale (CES, cutoff ≥ 35), and the Multiple Affect Adjective Checklist (MAACL, cutoff ≥ 21).[102] Women with a positive screen (n = 142) were assigned (not randomly) to either a postpartum social support group that included both depressed and nondepressed women or a usual-care group.

Outcome measures included the EPDS and the CES at the same cutoff scores used for screening. At the 6-week and 5-month follow-up assessments, the groups did not differ significantly with respect to rates of maternal depression, and the support groups had no apparent effect on the mothers' general affective mood. However, women in the social support group had a statistically significant increase in the number of maternal-infant interactions and noted decreased infant crying compared to women in the routine-care group.

Chen et al. screened 414 women at 3 weeks postpartum using the Taiwanese BDI (cutoff ≥ 10).[100] Of these, 115 women with positive screens were randomized to weekly support groups or to a routine-care group; 60 patients were available for analysis. Outcome measures included the BDI (cutoff ≥ 10), Perceived Stress Scale (PSS), and measures of interpersonal support. At the 15-week follow-up assessment, women in the intervention group had significantly lower rates of depression: 33.3 percent of the intervention group and 60.0 percent of the control group had BDI values equal to or greater than 10 ($P < 0.05$). The rate of perceived stress was also significantly lower in the intervention group than the control group (t-test = 3.75, $P < 0.01$). Finally, women receiving the intervention reported significantly more interpersonal support as measured by the Interpersonal Support Evaluation List than those in the control group (t-test = 2.81, $P < 0.01$).

Pharmacologic Studies. Two of the studies were psychotropic medication trials to prevent the occurrence of postpartum depression. The women were not directly screened with any instrument, but rather were included if they had a previous history of postpartum depression. The same research team conducted both of these pharmacologic trials. In the first trial, Wisner and Wheeler studied the efficacy of antidepressant treatment in women with a previous history of

postpartum depression (i.e., at high risk of maternal depression but no history of psychosis or bipolar disorder).[107] At-risk postpartum women (n = 23), who had had at least one episode of postpartum depression were treated in an open clinical trial with the tricyclic antidepressant nortriptyline and postpartum monitoring or with postpartum monitoring only. Outcome measures included a clinical assessment of major depression and the Inventory to Diagnose Depression scale. After 3 months, study results demonstrated a significantly greater proportion of new-episode major depression in those patients who received monitoring alone than in those in the medication group (62.5 percent of those in the monitoring group; 6.7 percent in the medication group; $P = 0.0086$).

In a later Wisner et al. RCT, 56 women with a prior history of postpartum depression within the past 5 years but no depressive episode upon enrollment, as diagnosed by standardized research diagnostic criteria, were randomized to either a nortriptyline group or a placebo group immediately postpartum.[108] Outcome measures of recurrence of perinatal depression included the HAM-D and Research Diagnostic Criteria (RDC). In contrast to the earlier open-label trial, the investigators reported no difference in the rate of recurrence of depression (one-fourth) between women treated with nortriptyline and those receiving placebo (23 percent versus 24 percent, respectively).

None of the studies used treatment interventions that are recognized as the gold standard treatment for major depressive illness according to current American Psychiatric Association guidelines. These guidelines specify that the gold standard include antidepressant medication plus psychotherapy.[12]

The overall quality of these 11 postpartum studies was fair; one study was rated good; two, poor. Of a possible quality score of 29, one study scored 22, seven scored between 15 and 20, and two at or below 14 (Table 18). External validity was generally poor; the majority of studies scored 0 of 3 points. The bias measure of internal validity ranged between scores of 1 and 6 (of a possible 7); 6 of 11 studies scored 4 or better. For confounding, scores ranged from 1 to 6 (of a possible 6); 6 of 11 studies scored 4 or better. Power was generally poor, reflecting the small sample sizes; 9 of 11 studies scored 0 (of a possible 2).

Only three studies had quality scores of 18 or higher: Wisner et al.,[108] Armstrong et al.,[98] and Dennis et al.[101] All three enrolled women in the postpartum period and had a fairly intensive treatment approach consisting of weekly interventions (Wisner et al. for 20 weeks, Dennis et al. for 8 weeks, and Armstrong et al. for 6 weeks). The Wisner et al. study had a pharmacologic intervention with weekly assessments of efficacy but no psychotherapeutic intervention; by contrast, the Dennis et al. and Armstrong et al. studies had weekly psychotherapeutic interventions but no pharmacologic intervention. Interestingly, although Wisner et al. treated patients for 20 weeks postpartum with antidepressant medication, their study did not have a significant result. This finding may suggest that psychosocial support and psychotherapeutic intervention are both critical as part of a treatment plan for women with postpartum depression.

Discussion

Conclusions

The 15 studies examined a variety of screening and treatment interventions for women identified as being at risk (sometimes at high risk) for postpartum depression. The majority of these studies focused on intervention strategies in the postpartum period; all but two dealt with a wide array of psychosocial, education, skill-building, and other mother-child behavioral activities. Generally, the more successful efforts occurred in the studies in which screening and interventions were carried out in the postpartum, not the prenatal, period. Once again, none of the studies had a treatment intervention with both psychotherapeutic and pharmacologic components that would be considered "gold standard therapy" for the treatment of major depression.

Overall, many of the studies suggest that providing the mother with some form of psychosocial program to increase maternal support or improve maternal-child interaction may decrease the rate of postpartum depression. Across the nine nonpharmacologic studies, about 20 outcomes were assessed; of these, 12 showed significant effects for the intervention group. Taking only the outcomes dealing specifically with depression, nine significant effects were reported. The two small pharmacologic trials from the same group yielded conflicting results about the impact of nortriptyline in reducing recurrence of maternal depression.

Only one study[97] specifically studied low SES women—a matter of some interest to the Safe Motherhood Group. Low SES women with at least one risk factor for postpartum depression who participated in weekly prenatal survival skills classes were less likely to develop postpartum depression compared to controls. This small study suggested that increasing support and parenting skills may help to decrease postpartum depression in this particular population.

Study Limitations

This set of studies, however, has several limitations, and it can be regarded as offering, at best, only fair evidence about the utility of screening plus prevention or treatment programs or even interventions alone. Although a variety of interventions may be helpful in treating women with or at risk of perinatal depression, the available evidence does not directly address whether *screening* with subsequent intervention improves outcomes. Screening, in the classic sense, implies "examination of a group *of usually asymptomatic individuals* to detect those with a high probability of having a given disease" (italics added; http://dictionary.reference.com/search?q=screening); this meaning can be extended to using appropriate screening or diagnostic tools within populations with known risk factors. These studies provide little guidance in answering the practical question of whether clinicians should screen all women in the perinatal period (i.e., essentially an asymptomatic population with respect to depression) for risk factors or latent depression, or whether they should screen only women who have known prior histories or risk factors for depression.

The studies are generally small, with poor generalizability (especially to the heterogeneous childbearing population of the United States). We contemplated and rejected the idea of any quantitative analyses: populations, settings, and screening and outcome measures—let alone interventions—were simply too disparate for anything but qualitative synthesis.

Future Research

To overcome some of these problems in understanding the impact of programs designed to prevent the problems of perinatal depression or to mitigate the considerable deleterious effects of this disorder on mothers, infants, and families, considerably more and better research needs to be conducted. Possibly the most important issue is for future studies to enroll adequate samples of women and, if screening is the question, to screen quite large numbers of women to produce sample sizes with adequate power to detect relevant differences between treatment and control groups in later phases of these studies. Virtually all studies appeared to be underpowered to start with, and some lost participants along the way. This deficiency hampers investigators and policymakers in making sense of, or decisions based on, much of this work.

Moreover, a greater effort should be made, at least in US-based studies, to focus on ethnic and disadvantaged populations, such as low-income women. Even if the incidence and prevalence of perinatal depression were "evenly" spread over population groups in this country, the underfunding of health care for many (e.g., lack of insurance, poor coverage of mental health benefits in insurance plans, unavailability of publicly funded services) and the more precarious economic resources and family support for some populations means that additional attention needs to be paid to them. For example, programs may need to be designed to take lack of transportation, child care, or telephone access into account.

In addition, researchers might direct attention to several other variables that appear to be important. They include first-time versus second-time mothers, maternal comorbidities and lifestyle behaviors, family structure (make-up) and available support, and status of infant at birth (e.g., full term or not, healthy or not). Another gap may be programs intended to assist the mother-father dyad or, indeed, to assist fathers in providing the emotional or physical support needed to forestall depression in new mothers.

Ideally, researchers would employ similar screening measures with similar cutoff points so that some elements of separate studies could be compared more readily. Not all of the screening instruments used appear to be sufficiently well-targeted to perinatal depression (i.e., even if they are reliable, their validity for this purpose may be called into question). Moreover, some instruments may be relatively infeasible for use in certain populations (e.g., immigrants) or in cases in which patient self-report is important and literacy may be low. For these situations, some work to calibrate well-known instruments that have been specifically designed for this disorder and that have acceptable test properties against each other might be useful. Calibrating less well-known or well-proven instruments against some agreed-upon reference ("gold") standard instrument in this area might also be valuable. Testing these in different settings, trying to use shorter instruments, attempting to take literacy levels into account, and in other ways improving the screening armamentarium are also important steps. In that way, investigators and clinicians can have a better selection of proven screening tools for future research or clinical practice applications.

Another element warranting more clarity is the purpose of the screening-cum-intervention effort. All appear to relate to populations of women at risk of perinatal depression (particularly postpartum depression that goes beyond "maternal blues"), but the severity level of being at risk differed in these studies. Moreover, women could have had no prior history of depression (or perinatal depression) and be at risk; alternatively, they could have had some history, especially of postpartum depression, and be, essentially, at "high" risk. These distinctions did not seem to be well or consistently described across these studies. They also have implications for the goals of

the interventions themselves: for example, preventing any "first episode" of depression, mitigating the effects of a first episode that is not wholly prevented, or preventing a recurrence.

Interventions tested in the future would, ideally, be those shown to have some promise so far (e.g., as reflected in some of the studies reported here). The components of the programs should be of appropriate length and intensity, and published articles should describe them thoroughly. In addition, interventions should be consistent with current evidence-based practice standards for the treatment of major depression. Multiple studies of the same interventions, perhaps at different time periods or different settings and populations, might be helpful in completing the picture of the impact of screening and interventions on occurrence or reoccurrence of perinatal depression. Finally, outcome measures should be appropriate to the research questions and preferably selected from among the more reliable, valid, and widely used instruments. These steps might help fill the gaps in this knowledge base and permit those performing systematic reviews to compare and synthesize studies more readily.

Chapter 6. Conclusions and Recommendations

In an effort to identify the evidence base addressing important questions on the epidemiology, screening and diagnosis, and management of perinatal depression, the Safe Motherhood Group (SMG) and the Agency for Healthcare Research and Quality (AHRQ) initially requested a feasibility study to determine whether enough high-quality evidence existed on six separate issues to support a full evidence report. After reviewing our feasibility study,[24] SMG and AHRQ requested an evidence report focusing on the three key questions (KQs) covered in this review.

We applied rigorous selection criteria and assessed the quality of each study, bringing a public health perspective to an area of research that traditionally has not had this focus. Our report was limited to depressive illness without psychotic symptoms, the latter complication being much less common and much more challenging to identify and manage. We made a distinction between results involving major depression alone, a discrete clinical syndrome for which treatment is clearly indicated, and results referring to patients with either major or minor depression, for which management is less clear.

This evidence report comprises a comprehensive review of all the available research. In this final chapter, we first review the major findings pertaining to each question and the strength of overall evidence about these issues; we then present some observations and recommendations about future research.

Conclusions

Key Question 1: Prevalence and Incidence of Perinatal Depression

For KQ 1, we identified 30 studies of generally moderate size that provide estimates of the prevalence of perinatal depression; 13 of these inquiries provide estimates of incidence. Studies were generally of good quality for reporting completeness and internal validity for bias; by contrast, they were of fair quality for precision and only poor quality for external validity and internal validity for confounding. In particular, the study populations were not representative of the racial and ethnic mix of the countries in which the studies were performed and especially not of the United States.

Our final best estimates of prevalence and incidence were somewhat lower than those reported in prior systematic reviews because we excluded studies that assessed depression based on self-report screens alone, which have been found to overestimate prevalence. Also, we separated out estimates of major and minor depression from estimates of major depression alone. Finally, we included more recent studies that use more precise criteria to identify major depression.

For major depression alone, our final combined point prevalence estimates ranged from 3.1 percent to 4.9 percent at different times during pregnancy and from 1.0 percent to 5.9 percent at different times during the first postpartum year. For major and minor depression, our final

Note: Appendixes and Evidence Tables cited in this report are provided electronically at http://www.ahrq.gov/clinic/epcindex.htm.

combined estimates of point prevalence ranged from 8.5 percent to 11.0 percent at different times during pregnancy and between 6.5 percent and 12.9 percent at different times during the first year postpartum. This nearly 2-fold higher rate suggests that approximately half of the women experience a major depressive episode and half a minor depressive episode at any given time. Confidence intervals surrounding all these estimates remain wide, suggesting that a fair amount of uncertainty remains in the combined estimates.

Fewer estimates were available for the incidence of depression. These limited data suggest that as many as 14.5 percent of pregnant women have a new episode of either major or minor depression during pregnancy, and 14.5 percent have a new episode during the first 3 months postpartum. Considering only major depression, 7.5 percent may have a new episode during pregnancy, with 6.5 percent having a new episode in the first 3 months postpartum.

Are the prevalence and incidence of depression during the perinatal period higher than the rates during nonchildbearing periods? We found three studies that measured the prevalence of major or minor depression and major depression alone for women at different times during these two periods. None of these estimates shows a statistically significant difference. Only one study[20] directly compared the incidence (new onset) of perinatal depression to that of nonchildbearing women of similar age; women at 5 weeks postpartum were more than three times as likely as the comparison group to have a new episode of major or minor depression. By 6 months postpartum, this difference had disappeared. An incidence for major depression alone was not reported.

That these estimates did not appear significantly different from those of nonchildbearing women of the same age does not reduce the dramatic burden experienced by women postpartum. Indeed, these estimates, based on the best available evidence, suggest that perinatal depression, whether major or minor depression, is a very common complication of pregnancy. Furthermore, and arguably more important, after labor and delivery this dramatically common complication, rather than primarily affecting one individual, now directly affects two: mother and child.

Key Question 2: Accuracy of Perinatal Depression Screening Tools

For our analysis of the accuracy of screening tools (KQ 2), we identified 10 studies reporting test characteristics for English-language screeners. In general, studies were of fair to good quality, although external validity was only poor to fair. In particular, the study populations were nearly entirely white, so the accuracy of these screeners in nonwhite perinatal populations is not clear. A major limitation in the available evidence is the very small number of depressed patients involved, a fact that results in substantial imprecision in the point estimate of sensitivity and prevents one from reasonably determining an ideal cutoff point.

For depression during pregnancy, we found only one study reporting on screening accuracy in a population with 6 patients with major depression and 14 patients with either major or minor depression. For major depression, sensitivities for the Edinburgh Postnatal Depression Scale (EPDS) at all evaluated thresholds (12, 13, 14, 15) were 1.0, underscoring the markedly small number of depressed patients involved; specificities ranged from 0.79 (at EPDS $\geq$ 12) to 0.96 (at EPDS $\geq$ 15). For major or minor depression, sensitivity was much poorer (0.57 to 0.71); specificity remained fairly high (0.72 to 0.95).

For postpartum depression also, the small number of depressed patients involved in the studies precluded identifying an optimal screener or an optimal threshold for screening. Our ability to conduct a meta-analysis of the results of different studies was limited by the use of

multiple cutoffs and other differences across studies that precluded a meaningful interpretation of the results. Where we were able to combine the results, the pooled estimates did not add to what one could conclude from individual studies.

For women with major depression alone, specificity for all screeners (the Beck Depression Inventory [BDI], the Postpartum Depression Screening Scale [PDSS], and the EPDS) was relatively high. This finding suggests that a positive screen was accurate in ruling major depression in; that is, the risk that a screen with one of these instruments would be falsely positive was low. By contrast, sensitivities varied much more. The EPDS and the PDSS appeared to be more sensitive (with estimates ranging from 0.75 to 1.0 at different thresholds) than the BDI instruments (with estimates from 0.32 to 0.68), but the wide confidence intervals (CIs) overlapped nearly completely. This means that we could not say with confidence that the specificity estimates using the different tools were different.

The point estimates are consistent with what is reported for depression screeners in primary care settings.[90] Still, the imprecision is important to clarify. If falsely missing depression (a false negative) is worse than falsely identifying it, as may be the case with this disorder, clinicians must be able to feel confident that the screen is usually positive if the disease is there and that a negative result can help rule out the illness.

For patients with major or minor depression, results were reported for EPDS, BDI, PDSS, and the Center for Epidemiologic Studies–Depression (CES-D). Specificity estimates remained relatively high, but sensitivity results were much lower (ranging from 0.43 to 0.71) than for major depression alone. This means that the ability of the screening instrument to score women as positive for this condition when the disease is present was poorer than for major depression alone. Again, neither any particular cutoff nor any particular screening instrument performed differently from the others. No available comparators were found for primary care populations.

Our results suggest that various screening instruments can identify perinatal depression, most accurately major depression, but clinicians need to know more about the precision of individual instruments. If one assumes that the risk of a false-negative depression screen is worse than the risk of a false-positive screen, perinatal depression is a condition in which sensitivity is likely to be more important than specificity. Whether as a screen for major depression alone or for major or minor depression, specificities appear high and relatively precise. By contrast, sensitivity for identifying either category is imprecise and differs by diagnostic category. For major depression alone, point estimates are equivalent to those in primary care medical settings. For major or minor depression, however, sensitivity is quite low. At this time, these screens do not appear to be useful for identifying patients in this latter category of illness.

Key Question 3: Screening and Treatment Outcomes

KQ 3 concerned issues of whether screening ultimately leads to improved patient outcomes. Although it is the most vital question from the public health perspective, it is the one with the most limited evidence. Indeed, the studies that we identified were not designed to test whether *screening* for depression (versus not screening) improved patient outcomes. Such a design would randomize patients to be screened or not to be screened and then compare subsequent outcomes. We found no studies designed in this way.

Instead, we made use of studies in which women were screened by formal depression screening or the presence of a risk factor associated with perinatal depression to identify those at risk of having a depressive illness; then, for those screening positive, the investigators compared

the outcomes of women receiving a treatment intervention to those in a control group. This design tests whether, among women identified as at risk of depression by a screen, an intervention improves outcomes compared to the outcomes in a control group. This is an important intermediary step, but it does not directly test whether screening itself improves outcome compared to not screening. All the trials included are treatment studies that had a screening component (either a formal depression screening instrument, or other type of screen that identified women at risk of a depressive illness) but did not have diagnostic confirmation of depression.

We attempted to synthesize the results of the included studies quantitatively, but the study methods (screening instruments, type of intervention, intensity of intervention, outcomes measured) were so heterogeneous that a meta-analytic synthesis would not be meaningful. We also attempted to compare effect sizes to attempt an exploratory analysis of the various studies, but the data necessary to compute these were not available.

For patients whose screening results identified them as at risk of perinatal depression and for whom a subsequent intervention was provided, we identified 15 studies. Four small prenatal studies involved various psychosocial interventions. Quality was poor for three of these studies and fair for one. Overall, the effects of the interventions in these studies were not consistently superior to those in the control groups.

The 11 postpartum studies were of overall fair quality and had larger sample sizes than the prenatal trials. Study populations reflected only a limited racial and ethnic mix, and both external validity and the power to demonstrate statistically significant differences were generally poor. Again, screening tools and interventions varied considerably; the latter involved both psychosocial and pharmaceutical interventions.

Results were mixed. Of the nine trials that employed a psychosocial intervention, six studies[98-101,103,104] reported significant benefit for depression outcomes in the experimental group compared to those in the control group. The one RCT involving pharmacologic intervention did not show benefit relative to the control group.[108] Overall, the evidence available is not sufficient to draw conclusions about this key question. These results, although limited, do suggest that providing some form of psychosocial support to pregnant women at risk of having a depressive illness may decrease depressive symptoms.

Recommendations for Future Research

The available research suggests that depression is one of the more common complications of the prenatal and postpartum periods and that fairly accurate and feasible screening measures are available. The prenatal or postpartum periods are clearly not times for nonpsychiatric clinicians to ignore depression screening, which is routinely recommended for patients seen in primary care settings.[110,111]

Specifics of the course of a depressive illness with onset during the perinatal period, including the severe physiologic and psychological challenges unique to this period that complicate the identification and management of perinatal depression, seem to suggest that this topic would have a substantial degree of high-quality research. We were surprised by the paucity of such evidence in this area. If one assumes that perinatal depression is a significant mental health and public health problem, then larger scale studies are needed involving each of these

domains. The small number and small size of relevant studies are not adequate to guide national policy.

Reflecting on the three key questions addressed in this report, we have concluded generally that the level of research warrants both improvement and expansion. The three results chapters discuss the limitations and gaps in these areas in more detail. We summarize here our suggestions for additional research efforts for the future.

For KQ 1, prevalence studies need to account better for the racial and ethnic mix of perinatal depression in the US population. We do not have good evidence about whether and, if so, how perinatal depression rates differ among various ethnic groups. The absence of information on nonwhite populations was dramatic. Better understanding any racial and ethnic variations could help clinicians know where to target screening programs and researchers know where to target studies on screening tools, and it could help researchers clarify the need for more nationally representative perinatal depression samples. Furthermore, researchers need to clarify whether the incidence of perinatal depression is greater than the incidence of depression in nonchildbearing women of similar ages.

For KQ 2, the quality grades point to several areas in which improvements in study design and conduct are needed. In particular, future studies on the test characteristics of screeners must be designed with sample size estimates that take prevalence into account and that project a reasonably precise estimate of sensitivity for the particular illness. Moreover, samples should more closely mirror the target population; specifically, subsequent studies need to provide a more representative racial and ethnic mix. In addition, studies should incorporate a range of other demographic variables that could influence screening performance, such as socioeconomic status measures, and assess the screening tools in these subpopulations.

Furthermore, as Beck and Gable did,[71] future research should continue to assess and directly compare multiple screening instruments. This design would provide a head-to-head comparison to allow an evaluation of which screening instrument is more accurate in the setting in which the investigations are carried out. Moreover, studies evaluating the cost-effectiveness of screening, specifically assessing the relative costs of false-negative and false-positive designation, the degree of provider burden, and patient acceptability, are needed to provide insights on how to consider target sensitivity and specificity when attempting to maximize cost-effectiveness.

Diagnosis is another area of concern. Subsequent studies should carefully consider whether to target major depression alone, for which beneficial treatments clearly exist, or the combined category that includes minor depression, a heterogeneous group for which treatment benefit is unclear. Given that the results suggest that available screening tools identify major depression alone more accurately, and noting that the general benefit of interventions is more apparent for major depression alone, we believe that an evidence-based public health perspective recommends targeting major depression alone.

Timing is another factor of future studies deserving more thought. The issue here involves both the need for more epidemiology to confirm prevalence rates at different times as well as the need to confirm what time point(s) would identify the greatest number of depressed women. The bulk of the few screening studies we identified had been conducted in the first 3 months postpartum. Our best estimates of prevalence suggest that depression may remain high for several more months.

More studies are needed to better delineate periods of peak prevalence and incidence, to include not just 3 months but also 6 weeks, 6 months, and 12 months, and subsequent screening studies need to consider testing properties of screening at these later time periods. The very

small number of adequate studies currently available hampers plans for screening and intervention programs because the best time for screening, and hence the best clinic location, is not clear. If peak prevalence and incidence occur within the first 6 weeks, the obstetrics clinic is a prime place to target resources for such a program. If, however, peaks occur after this time, most postpartum women will have completed follow-up care with an obstetrician, so programs in an obstetrics clinic may be less helpful. In this case, programs targeting new mothers in family medicine, internal medicine, or pediatric clinics might be more effective.

For KQ 3, several similar or related issues emerged as well. First, studies addressing the relationship between screening and outcome need to recruit and retain sample sizes that are large enough to yield adequate power to detect relevant differences. Second, screening and outcome studies must include populations with a racial and ethnic mix that is more representative of the US populations than the work we have seen to date. Third, interventions involved should be more consistent with what we know to be evidence-based treatments for depression,[12] i.e., antidepressant medications[112] and/or psychotherapies such as cognitive behavioral therapy[113] or interpersonal psychotherapy.[114]

Type of screening measures used henceforth is another major issue. Of the three KQ 3 studies rated as good,[98,101,108] only Dennis and colleagues used a depression screener (EPDS).[101] Researchers should consider developing and using standardized screening measures, and similar cutoff points, so that some elements of separate studies could be compared more readily. Screening tools with the best supporting evidence would seem to be the best candidates. While the evidence base remains quite limited and any conclusions preliminary, at this time those instruments would appear to be either the EPDS or the PDSS. For major depression alone, an EPDS cutoff of ≥ 13 or a PDSS cutoff of ≥ 81 are reasonably supported by the evidence. For major or minor depression, we found the results too inconclusive to make even a preliminary recommendation.

Finally, studies should be designed to address whether the screening process itself leads to better access to proven treatment and improved outcome relative to usual care. We support additional research on interventions per se, but we conclude that important questions remain about the impact of the screening element. Reviewing studies that used screening as a means of identifying women potentially at high risk and enrolling them in interventional studies is not a sufficient approach to answering issues about the effectiveness of screening.

References and Included Studies

1. Kessler RC, McGonagle KA, Zhao S, et al. Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. Results from the National Comorbidity Survey. Arch Gen Psychiatry 1994; 51(1):8-19.

2. Robins L, Regier D. Psychiatric Disorders in America. New York: Free Press, 1991.

3. Depression Guideline Panel. Depression in Primary Care: Volume 1. Detection and Diagnosis. Clinical Practice Guideline, Number 5. Rockville, Md: Agency for Health Care Policy and Research, 1993; AHCPR No. 93-0550.

4. Shaw J, Kennedy SH, Joffe RT. Gender differences in mood disorders: A clinical focus. In. Gender Psychopathol. Washington, DC: American Psychiatric Press, 1996: 89-111.

5. Burke KC, Burke JD Jr, Rae DS, et al. Comparing age at onset of major depression and other psychiatric disorders by birth cohorts in five US community populations. Arch Gen Psychiatry 1991; 48(9):789-95.

6. Kessler RC. Epidemiology of women and depression. J Affect Disord 2003; 74(1):5-13.

7. Murray L, Stein A. The effects of postnatal depression on the infant. Baillieres Clin Obstet Gynaecol 1989; 3(4):921-33.

8. Marmorstein NR, Malone SM, Iacono WG. Psychiatric disorders among offspring of depressed mothers: associations with paternal psychopathology. Am J Psychiatry 2004; 161(9):1588-94.

9. Burke L. The impact of maternal depression on familial relationships. Int Rev Psychiatry 2003; 15(3):243-55.

10. Stein A, Gath DH, Bucher J, et al. The relationship between post-natal depression and mother-child interaction. Br J Psychiatry 1991; 158:46-52.

11. Flynn HA, Davis M, Marcus SM, et al. Rates of maternal depression in pediatric emergency department and relationship to child service utilization. Gen Hosp Psychiatry 2004; 26(4):316-22.

12. American Psychiatric Association (APA). Practice guideline for the treatment of patients with major depression (revision). Am J Psychiatry 2000; 157(4).

13. Wagner HR, Burns BJ, Broadhead WE, et al. Minor depression in family practice: Functional morbidity, comorbidity, service utilization, and outcomes. Psychol Med 2000; 30(2):1377-90.

14. Oxman TE, Sengupta A. Treatment of minor depression. Am J Geriatr Psychiatry 2002; 10(3):256-64.

15. Judd LL, Rapaport MH, Yonkers KA, et al. Randomized, placebo-controlled trial of fluoxetine for acute treatment of minor depressive disorder. Am J Psychiatry 2004; 161(10):1864-71.

16. Bloch M, Daly RC, Rubinow DR. Endocrine factors in the etiology of postpartum depression. Compr Psychiatry 2003; 44(3):234-46.

17. Jones I, Craddock N. Familiality of the puerperal trigger in bipolar disorder: results of a family study. Am J Psychiatry 2001; 158(6):913-7.

18. Klein M, Essex MJ. Pregnant or depressed? The effect of overlap between symptoms of depression and somatic complaints of pregnancy on rates of major depression during the second trimester. Depression 1994; 2:1994-5.

19. O'Hara MW, Neunaber DJ, Zekoski EM. Prospective study of postpartum depression: Prevalence, course, and predictive factors. J Abnorm Psychol 1984; 93(2):158-71.

20. Cox JL, Murray D, Chapman G. A controlled study of the onset, duration and prevalence of postnatal depression. Br J Psychiatry 1993; 163:27-31.

21. O'Hara MW, Swain AM. Rates and risk of postpartum depression -- A meta-analysis. Int Rev Psychiatry 1996; 8:37-54.

22. Llewellyn AM, Stowe ZN, Nemeroff CB. Depression during pregnancy and the puerperium. J Clin Psychiatry 1997; 58 Suppl 15:26-32.

23. Yonkers KA, Ramin SM, Rush AJ, et al. Onset and persistence of postpartum depression in an inner-city maternal health clinic system. Am J Psychiatry 2001; 158(11):1856-63.

24. Gaynes B, Gavin N, Meltzer-Brody S, Sleath B, Sutton S. Perinatal Depression: Feasibility Study. Rockville, Md.: Agency for Healthcare Quality and Research (AHRQ), 2003.

25. Cochrane Methods Working Group. Based on the Cochrane Methods Working Group on Systematic Review of Screening and Diagnostic Tests: Recommended Methods, updated June 6, 1996. 1996.

26. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. J Epidemiol Community Health 1998; 52(6):377-84.

27. Bennett HA, Einarson A, Taddio A, et al. Prevalence of depression during pregnancy: Systematic review. Obstet Gynecol 2004; 103(4):698-709.

28. American Psychiatric Association (APA). Diagnostic and statistical manual of mental disorders. Fourth edition. Washington, DC: American Psychiatric Association, 1994.

29. Affonso DD, Lovett S, Paul SM, et al. A standardized interview that differentiates pregnancy and postpartum symptoms from perinatal clinical depression. Birth 1990; 17(3):121-30.

30. Areias ME, Kumar R, Barros H, et al. 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(1):30-5.

31. Berle J, Aarre T, Mykletun A, et al. Screening for postnatal depression. Validation of the Norwegian version of the Edinburgh Postnatal Depression Scale, and assessment of risk factors for postnatal depression. J Affect Disord 2003; 76(1-3):151-6.

32. Campbell SB, Cohn JF. Prevalence and correlates of postpartum depression in first-time mothers. J Abnorm Psychol 1991; 100(4):594-9.

33. Cooper PJ, Murray L, Hooper R, et al. The development and validation of a predictive index for postpartum depression. Psychol Med 1996; 26(3):627-34.

34. Cox JL, Connor Y, Kendell RE. Prospective study of the psychiatric disorders of childbirth. Br J Psychiatry 1982; 140:111-7.

35. Garcia-Esteve L, Ascaso C, Ojuel J, et al. Validation of the Edinburgh Postnatal Depression Scale (EPDS) in Spanish mothers. J Affect Disord 2003; 75(1):71-6.

36. Gotlib IH, Whiffen VE, Mount JH, et al. Prevalence rates and demographic characteristics associated with depression in pregnancy and the postpartum. J Consult Clin Psychol 1989; 57(2):269-74.

37. 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(3):445-53.

38. Kent GN, Stuckey BG, Allen JR, et al. Postpartum thyroid dysfunction: Clinical assessment and relationship to psychiatric affective morbidity. Clin Endocrinol 1999; 51(4):429-38.

39. Kitamura T, Shima S, Sugawara M, et al. Psychological and social correlates of the onset of affective disorders among pregnant women. Psychol Med 1993; 23:967-75.

40. Kitamura T, Sugawara M, Shima S, et al. Temporal variation of validity of self-rating questionnaires: Improved validity of repeated use of Zung's Self-Rating Depression Scale among women during the perinatal period. J Psychosom Obstet Gynecol 1999; 20(2):112-7.

41. Kumar R, Robson KM. A prospective study of emotional disorders in childbearing women. Br J Psychiatry 1984; 144:35-47.

42. Lee D, Yip A, Chiu H, et al. A psychiatric epidemiological study of postpartum Chinese women. Am J Psychiatry 2001; 158(2):220-6.

43. Lee D, Yip A, Chiu H, et al. Screening for postnatal depression: Are specific instruments mandatory? J Affect Disord 2001; 63(1-3):233-8.

44. Lucas A, Pizarro E, Granada ML, et al. Postpartum thyroid dysfunction and postpartum depression: Are they two linked disorders? Clin Endocrinol 2001; 55(6):809-14.

45. Matthey S, Barnett B, Howie P, et al. Diagnosing postpartum depression in mothers and fathers: Whatever happened to anxiety? J Affect Disord 2003; 74(2):139-47.

46. Murray D, Cox JL. Screening for depression during pregnancy with the Edinburgh Depression Scale (EPDS). J Reprod Infant Psychol 1990; 8(2):99-107.

47. Pop VJ, Essed GG, de Geus CA, et al. Prevalence of post partum depression--or is it post-puerperium

depression? Acta Obstet Gynecol Scand 1993; 72(5):354-8.

48. Watson JP, Elliott SA, Rugg AJ, et al. Psychiatric disorder in pregnancy and the first postnatal year. Br J Psychiatry 1984; 144:453-62.

49. Whiffen V. Vulnerability of postpartum depression: A prospective multivariate study. J Abnorm Psychol 1988; 97(4):467-74.

50. Yamashita H, Yoshida K, Nakano H, et al. Postnatal depression in Japanese women. Detecting the early onset of postnatal depression by closely monitoring the postpartum mood. J Affect Disord 2000; 58(2):145-54.

51. Yoshida K, Marks M, Kibe N, et al. Postnatal depression in Japanese women who have given birth in England. J Affect Disord 1997; 43(1):69-77.

52. Cooper PJ, Campbell EA, Day A, et al. Non-psychotic psychiatric disorder after childbirth. A prospective study of prevalence, incidence, course and nature. Br J Psychiatry 1988; 152:799-806.

53. O'Hara MW, Zekoski EM, Philipps LH, et al. Controlled prospective study of postpartum mood disorders: Comparison of childbearing and nonchildbearing women. J Abnorm Psychol 1990; 99(1):3-15.

54. 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(4):351-8.

55. Georgiopoulos AM, Bryan TL, Wollan P, et al. Routine screening for postpartum depression. J Fam Pract 2001; 50(2):117-22.

56. Endicott J, Spitzer RL. A diagnostic interview: The schedule for affective disorders and schizophrenia. Arch Gen Psychiatry 1978; 35(7):837-44.

57. Spitzer RL, Williams JBW, Gibbon M, First MB. Structured Clinical Interview for SSM-III-R. Washington, DC: American Psychiatric Press, 1990.

58. Frist MB, Spitzer RL, Gibbon M, Williams JBW. Structured Clinical Interview for DSM-IV Axis I Disorders (SCID), Clinical Version. Washington, DC: American Psychiatric Press, 1996.

59. Goldberg DP, Cooper B, Eastwood MR, et al. A standardized psychiatric interview for use in community surveys. Br J Prev Soc Med 1970; 24(1):18-23.

60. Janca A, Ustun TB, Sartorius N. New versions of World Health Organization instruments for the assessment of mental disorders. Acta Psychiatr Scand 1994; 90(2):73-83.

61. Robins LN, Helzer JE, Croughan J, et al. National Institute of Mental Health Diagnostic Interview Schedule. Its history, characteristics, and validity. Arch Gen Psychiatry 1981; 38(4):381-9.

62. Lecrubier Y, Sheehan D, Weiller E, et al. The Mini International Neuropsychiatric Interview (M.I.N.I.) a short diagnostic structured interview: Reliability and validity according to the CIDI. Euro Psychiatry 1997; 12:224-31.

63. Wing JK, Cooper JE, Sartorius N. The Measurement and Classification of Psychiatric Symptoms. Cambridge: Cambridge University Press, 1974.

64. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. Br J Psychiatry 1979; 134:382-9.

65. Spitzer RL, Endicott J, Robins E. Research diagnostic criteria: Rationale and reliability. Arch Gen Psychiatry 1978; 35(6):773-82.

66. American Psychiatric Association (APA). Diagnostic and statistical manual of mental disorders. Third edition. Washington, DC: American Psychiatric Association, 1987.

67. Pitt B. "Atypical" depression following childbirth. Br J Psychiatry 1968; 114(516):1325-35.

68. Blazer DG, Kessler RC, McGonagle KA, et al. The prevalence and distribution of major depression in a national community sample: The National Comorbidity Survey. Am J Psychiatry 1994; 151(7):979-86.

69. Finlay-Jones R, Brown GW, Duncan-Jones P, et al. Depression and anxiety in the community: Replicating the diagnosis of a case. Psychol Med 1980; 10(3):445-54.

70. Ballard CG, Davis R, Cullen PC, et al. Prevalence of postnatal psychiatric morbidity in mothers and fathers. Br J Psychiatry 1994; 164(6):782-8.

71. Beck CT, Gable RK. Comparative analysis of the performance of the Postpartum Depression Screening Scale with two other depression instruments. Nurs Res 2001; 50(4):242-50.

72. Boyce P, Stubbs J, Todd A. The Edinburgh Postnatal Depression Scale: Validation for an Australian sample. Aust N Z J Psychiatry 1993; 27(3):472-6.

73. Cox J, Chapman G, Murray D, et al. Validation of the Edinburgh Postnatal Depression Scale (EPDS) in non-postnatal women. J Affect Disord 1996; 39(3):185-9.

74. Harris B, Huckle P, Thomas R, et al. The use of rating scales to identify post-natal depression. Br J Psychiatry 1989; 154:813-7.

75. Leverton TJ, Elliott SA. Is the EPDS a magic wand? 1. A comparison of the Edinburgh Postnatal Depression Scale and health visitor report as predictors of diagnosis on the Present State Examination. J Reprod Infant Psychol 2000; 18(4):279-96.

76. Murray L, Carothers A. The validation of the Edinburgh Post-natal Depression Scale on a community sample. Br J Psychiatry 1990; 157:288-90.

77. Whiffen VE. Screening for postpartum depression: A methodological note. J Clin Psychol 1988; 44(3):367-71.

78. Guedeney N, Fermanian J. Validation study of the French version of the Edinburgh Postnatal Depression Scale (EPDS): New results about use and psychometric properties. Eur Psychiatry 1998; 13:83-9.

79. Lawrie T, Hofmeyr G, de Jager M, et al. Validation of the Edinburgh Postnatal Depression Scale on a cohort of South African women. S Afr Med J 1998; 88(10):1340-4.

80. Kitamura T, Shima S, Sugawara M, et al. Temporal variation of validity of self-rating questionnaires: Repeated use of the General Health Questionnaire and Zung's Self-rating Depression Scale among women during antenatal and postnatal periods. Acta Psychiatr Scand 1994; 90(6):446-50.

81. Lee D, Yip A, Chiu H, et al. Screening for postnatal depression using the double-test strategy. Psychosom Med 2000; 62(2):258-63.

82. Lee D, Yip S, Chiu H, et al. Detecting postnatal depression in Chinese women. Validation of the Chinese version of the Edinburgh Postnatal Depression Scale. Br J Psychiatry 1998; 172:433-7.

83. Muzik M, Klier C, Rosenblum K, et al. Are commonly used self-report inventories suitable for screening postpartum depression and anxiety disorders? Acta Psychiatr Scand 2000; 102(1):71-3.

84. Wickberg B, Hwang C. Counselling of postnatal depression: A controlled study on a population based Swedish sample. J Affect Disord 1996; 39(3):209-16.

85. 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.

86. Beck AT, Ward CH, Mendelson M, et al. An inventory for measuring depression. Arch Gen Psych 1961; 4:561-71.

87. Beck AT, Steer RA, Brown GK. Manual for Beck Depression Inventory II (BDI-II). San Antonio, Tx: Psychology Corporation, 1996.

88. Beck AT, Guth D, Steer RA, et al. Screening for major depression disorders in medical inpatients with the Beck depression inventory for primary care. Behav Res Ther 1997; 35:785-91.

89. Radloff LS. The CES-D Scale: A self-report depression scale for research in the general population. Appl Psychol Measure 1977; 1:385-401.

90. Williams JJr , Pignone M, Ramirez G, et al. Identifying depression in primary care: A literature synthesis of case-finding instruments. Gen Hosp Psych 2002; 24:225-37.

91. Henkel V, Mergl R, Kohnen R, et al. Use of brief depression screening tools in primary care: Consideration of heterogeneity in performance in different patient groups. Gen Hosp Psychiatry 2004; 26(3):190-8.

92. Meager I, Milgrom J. Group treatment for postpartum depression: A pilot study. Aust N Z J Psychiatry 1996; 30(6):852-60.

93. Brisco M. The detection of emotional disorders in the post natal period by health visitors. Health Visit 1989; 62(11):336-8.

94. Stamp GE, Williams AS, Crowther CA. Evaluation of antenatal and postnatal support to overcome postnatal depression: A randomized, controlled trial. Birth 1995; 22(3):138-43.

95. Brugha TS, Wheatley S, Taub NA, et al. Pragmatic randomized trial of antenatal intervention to prevent post-natal depression by reducing psychosocial risk factors. Psychol Med 2000; 30(6):1273-81.

96. Elliott SA, Leverton TJ, Sanjack M, et al. Promoting mental health after childbirth: A controlled trial of primary prevention of postnatal depression. Br J Clin Psychol 2000; 39 (Pt 3):223-41.

97. Zlotnick C, Johnson SL, Miller IW, et al. Postpartum depression in women receiving public assistance: Pilot study of an interpersonal-therapy-oriented group intervention. Am J Psychiatry 2001; 158(4):638-40.

98. Armstrong K, Fraser J, Dadds M, et al. A randomized, controlled trial of nurse home visiting to vulnerable families with newborns. J Paediatr Child Health 1999; 35(3):237-44.

99. Chabrol H, Teissedre F, Saint-Jean M, et al. Prevention and treatment of post-partum depression: A controlled randomized study on women at risk. Psychol Med 2002; 32(6):1039-47.

100. Chen CH, Tseng YF, Chou FH, et al. Effects of support group intervention in postnatally distressed women. A controlled study in Taiwan. J Psychosom Res 2000; 49(6):395-9.

101. Dennis CL. The effect of peer support on postpartum depression: A pilot randomized controlled trial. Can J Psychiatry 2003; 48(2):115-24.

102. Fleming AS, Klein E, Corter C. The effects of a social support group on depression, maternal attitudes and behavior in new mothers. J Child Psychol Psychiatry 1992; 33(4):685-98.

103. Hiscock H, Wake M. Randomised controlled trial of behavioural infant sleep intervention to improve infant sleep and maternal mood. Br Med J 2002; 324(7345):1062-5.

104. Honey KL, Bennett P, Morgan M. A brief psycho-educational group intervention for postnatal depression. Br J Clin Psychol 2002; 41(Pt 4):405-9.

105. Horowitz JA , Bell M, Trybulski J, et al. Promoting responsiveness between mothers with depressive symptoms and their infants. J Nurs Scholarsh 2001; 33(4):323-9.

106. Onozawa K, Glover V, Adams D, et al. Infant massage improves mother-infant interaction for mothers with postnatal depression. J Affect Disord 2001; 63(1-3):201-7.

107. Wisner KL, Wheeler SB. Prevention of recurrent postpartum major depression. Hosp Community Psychiatry 1994; 45(12):1191-6.

108. Wisner KL, Perel JM, Peindl KS, et al. Prevention of recurrent postpartum depression: A randomized clinical trial. J Clin Psychiatry 2001; 62 (2):82-6.

109. Turk DC, Meichenbaum D, Genest M. Pain and Behavioral Medicine. New York: Guilford, 1983.

110. US Preventive Services Task Force. Screening for depression: recommendations and rationale. Ann Intern Med 2002; 136(10):760-4.

111. Pignone MP, Gaynes BN, Rushton JL, et al. Screening for depression in adults: a summary of the evidence for the US Preventive Services Task Force. Ann Intern Med 2002; 136(10):765-76.

112. Hoffbrand S, Howard L, Crawley H. Antidepressant treatment for post-natal depression. Nurs Times 2001; 97(45):35.

113. Cooper P, Murray L. The impact of psychological treatments of postpartum depression on maternal mood and infant development. In: Cooper P, Murray L, editor(s). Postpartum depression and child development. New York: Guilford, 1997: 201-20.

114. O'Hara MW, Stuart S, Gorman LL, et al. Efficacy of interpersonal psychotherapy for postpartum depression. Arch Gen Psychiatry 2000; 57(11):1039-45.

Glossary

Bipolar disorder – a type of mood disorder characterized by both (1) one or more major depressive episodes and (2) either one or more manic or mixed episodes (Bipolar 1) or hypomanic episodes (Bipolar II). The disorder may or may not be accompanied by psychotic symptoms. In community samples, the prevalence of bipolar disorder (approximately 1 percent) is lower than the prevalence of major depressive disorder (at least 6 percent). Given that management of bipolar disorder is notably different from that of major depressive disorder, making such a diagnostic distinction is critical.

External validity – the extent to which a study's conclusions can be applied to populations and settings outside those of the study itself.

Incidence – the percentage of the population with an illness episode that begins within a given period of time (e.g., during pregnancy or within the first 3 months following delivery).

Internal validity – the extent to which a study is appropriately designed and conducted to measure what it is intended to measure.

Major depressive disorder – a type of mood disorder characterized by one or more major depressive episodes. The Diagnostic and Statistical Manual, version III (DSM-III) defines a major depressive episode as a period of at least 2 weeks during which an individual experiences daily disturbance in mood (intense feelings of sadness or loss of interest in activities that are usually pleasurable) and at least four of eight symptoms: (1) too much or too little sleep, (2) appetite or weight disturbance, (3) psychomotor agitation or retardation, (4) loss of energy, (5) feelings of worthlessness or excessive guilt, (6) problems with concentration or indecisiveness, (7) loss of interest in sex, and (8) recurrent suicidal thoughts or attempts. DSM-IV changed these criteria to the following: (1) symptoms must be present most of the day and nearly every day during the episode, (2) clinically significant distress or impairment in functioning must be present, (3) the syndrome must not be the result of the direct physiologic effects of a substance or a general medical condition, (4) major depressive disorder is still diagnosed after an acute grief reaction if the syndrome lasts for more than 2 months.

Major depressive disorder is not diagnosed if the syndrome is attributable to an acute grief reaction or a nonaffective psychotic condition such as schizophrenia. In addition, major depressive disorder is not diagnosed if there is a history of a manic, hypomanic, or mixed episode.

Maternity blues – a subthreshold cluster of depressive symptoms commonly described in up to 50 percent of postpartum women. This transient condition does not require an intervention.

Meta-analysis – a quantitative approach for systematically combining evidence from multiple previous research studies on a particular parameter or association to arrive at a conclusion about the body of research on that parameter or association.

Note: Appendixes and Evidence Tables cited in this report are provided electronically at http://www.ahrq.gov/clinic/epcindex.htm.

Meta-regression – a statistical analysis of the association between one or more study characteristics and the observed magnitude of effect.

Minor depressive disorder (also known as minor depression) – a subthreshold diagnosis with a variety of definitions, but in general seen as one or more episodes of depression lasting 2 weeks or more but with fewer symptoms than required for a diagnosis of major depressive disorder.

Period prevalence – the percentage of the population with depression over a period of time (e.g., during pregnancy or from delivery to the end of the first 3 months postpartum).

Perinatal depression – a condition encompassing major and minor depressive episodes that occur during pregnancy (prenatal) or within the first 12 months following delivery (postpartum).

Point prevalence – the percentage of the population with a condition at a given point in time (e.g., at 24 weeks gestation or 9 weeks postpartum).

Postpartum – for the purposes of this review, the period from parturition to 12 months after delivery.

.

Postpartum depression – according to DSM-IV, a specific type of major depressive disorder with onset of a major depressive episode within 4 weeks postpartum.

Postpartum psychosis – also known as puerperal psychosis, this condition is a severe and rare postpartum disorder, affecting 1 to 2 per 1,000 births. Women with postpartum psychosis present with new onset of delusions or prominent hallucinations. More than half of these episodes meet the criteria for major depressive disorder, and many women ultimately prove to have bipolar illness. Management of postpartum psychosis substantially differs from the much more common presentation of major depressive disorder with postpartum onset.

Power (statistical power) – the probability of detecting as "statistically significant" a postulated level of effect.

Precision – a measure of how close an estimator is expected to be to the true value of a parameter. Precision is related to the standard error of the estimator; less precision is reflected by a larger standard error.

Prenatal– the period of pregnancy from conception to parturition.

Puerperium – the 6-week period following delivery.

Reference standard (also known as gold standard) – the diagnostic assessment against which the screening test is compared to gauge the accuracy of the screening test. The reference standard determines the actual presence of disease. For psychiatric illness, the reference standard is often a clinical assessment by a mental health professional or a structured or semi-structured diagnostic interview.

Screen (also screening) – the use of a measure or test, often a formal instrument or tool, to classify an individual with respect to her likelihood of having a particular disorder. A screen itself does not diagnose the illness—those screening positive require subsequent diagnostic confirmation to confirm the presence of the disease.

Sensitivity – the ability of a test to identify correctly those who have a condition, computed as the percentage of true positive values correctly predicted by the test. A sensitive test identifies few false-negative cases.

Specificity – the ability of a test to identify correctly those who do not have a condition, computed as the percentage of true negative values correctly predicted by the test. A specific test identifies few false-positive cases.

U.S. Department of Health and Human Services
Mike Leavitt, *Secretary*

Office of Public Health and Science
Richard H. Carmona, M.D., M.P.H., F.A.C.S., *Surgeon General of the United States*

Agency for Healthcare Research and Quality
Carolyn M. Clancy, M.D., *Director*

Appendixes

Exact Search Strings

Database: MEDLINE <1966 to March Week 3 2004>
Search Strategy:
---

1 exp Puerperal Disorders/ (16527)
2 exp Depression/ (32747)
3 exp Depressive Disorder/ (42005)
4 2 or 3 (73267)
5 1 and 4 (1452)
6 exp Depression, Postpartum/ or perinatal depression.mp. (753)
7 5 or 6 (1467)
13 limit 7 to (human and english language) (1299)

CINAHL used these terms as well.

PsycINFO has "Depression, Postpartum" as a Major Descriptor that yields 379.

Sociofile indexes 105 records to "Postpartum Depression".

For Key Question 1, the following terms were used:

20 exp Natural History/ (8432)
21 8 and 20 (0)

When "Natural History" yielded no results, the following terms were used:

22 exp Cohort Studies/ (466831)
23 8 and 22 (112)
24 exp Longitudinal Studies/ (438062)
25 8 and 24 (101)
26 23 or 25 (134)

CINAHL (using similar terms) = 35

PsycINFO (natural history, cohort, longitudinal) = 65

Sociofile (natural history, cohort, longitudinal) = 20

Total from all databases for Key Question 1 = 254

After duplicates, book chapters, foreign language articles and dissertations were removed, the total unduplicated count for KQ1 = 210.

For Key Question 2, Incidence, the following terms were used:

MEDLINE
19 exp INCIDENCE/ (76679)
20 8 and 19 (31)

CINAHL (Incidence) = 7

PsycINFO (Incidence) = 23

Sociofile (Incidence) = 1

Total file = 62, minus duplications, dissertations, etc = 46

For Key Question 3, Risk, the following terms were used:

16 exp Risk Factors/ (221767)
17 8 and 16 (153)

CINAHL (Risk Factors) = 32

PsycINFO (risk) = 59

Sociofile (risk) = 11

Total from all databases for Key Question 3 = 255

After duplicates, book chapters, foreign language articles and dissertations were removed, the total unduplicated count for KQ3 = 204.

For Key Question 4, Therapies, the following terms were used:

MEDLINE
12 treatment.mp. or exp Therapeutics/ (2537613)
14 8 and 12 (513)

CINAHL (Treatment) = 90

PsycINFO (Treatment) = 91

Sociofile (Treatment) = 5

Total file = 699, minus duplications, dissertations, etc = 485

For Key Questions 5 and 6, Screening Accuracy and Screening Barriers, searches focused on "screening" and will give the total pool to investigators for finer sorting between questions.

MEDLINE

9 exp mass screening/ (62902)
10 8 and 9 (67)

CINAHL (screening) = 25

PsycINFO (screening) = 28

Sociofile (screening) = 1

Total from all databases for Key Questions 5 & 6 = 121

After duplicates, book chapters, foreign language articles and dissertations were removed, the total unduplicated count for KQ 5&6 = 96.

Quality Checklist for RCTs and Observational Studies of Prevalence and Incidence Studies

Reviewer's initials _____ Article # (from reference manager) _____

First Author _____ Journal:_____

Year published _____

Reporting	**Yes**	**No**
1. Is the hypothesis/aim/objective of the study clearly described?	1	0
2. Is the method of assessing depression clearly described in the Introduction or Methods section?	1	0
3. Are the characteristics of the patients included in the study clearly described?	1	0

	Yes	**P***	**No**
4. Are the distributions of principal confounders in each group of subjects to be compared clearly described? (i.e., bipolar, . . .)	2	1	0

	Yes	**No**
5. Are the main findings of the study clearly described?	1	0
6. Does the study provide estimates of or adequate information to estimate the random variability in the prevalence/ incidence rate?	1	0
7. Have the characteristics of patients lost to follow-up been described?	1	0
8. Have actual probability values been reported (e.g., 0.035 rather than <0.05) for the main outcomes except where the probability value is less than 0.001?	1	0

Total Reporting Score: _____

*P = Partially
U/D = Unable to Determine

External Validity	**Yes**	**No**	**U/D**
9. Were the subjects asked to participate in the study representative of the entire population from which they were recruited?	1	0	0
10. Were those subjects who were prepared to participate representative of the entire population from which they were recruited?	1	0	0
11. Were the staff, places, and facilities where the patients were treated representative of the treatment the majority of patients receive?	1	0	0

Total External Validity Score: _____

Internal Validity - Bias	**Yes**	**No**	**U/D**
12. Was the depression diagnosis or absence thereof verified through clinical interview for all study subjects?	1	0	0
13. In trials and cohort studies, do the analyses adjust for different lengths of follow-up of patients, or in case-control studies, is the time period over which mood is assessed the same for cases and control?	1	0	0
14. Were the statistical tests used to assess the main outcomes appropriate?	1	0	0

Total Bias Score: _____

*P = Partially
U/D = Unable to Determine

Internal Validity - Confounding	**Yes**	**No**	**U/D**	**NA**
15. Were the patients in different intervention groups (trials and cohort studies) or were the cases and controls (case-control studies) recruited from the same population?	1	0	0	0
16. Were study subjects in different intervention groups (trials and cohort studies) or were the cases and controls (case-control studies) recruited over the same period of time?	1	0	0	0
17. Was there adequate adjustment for confounding in the analyses from which the main findings were drawn?	1	0	0	
18. Were losses of patients to follow-up taken into account?	1	0	0	

Total Confounding Score: _____

Precision

19. Did the study have sufficient precision to provide a prevalence estimate where the probability value for the estimate being greater than zero is less than 5%?

Sample size	
< 30	0
30-250	1
250-1000	2
1000+	3

20. Did the study have sufficient precision to detect a clinically important effect where the probability value for a difference being due to chance is less than 5%?

Size of *smallest* group	
No comparison group	0
< 500	0
500-1000	1
1000-2000	2
2000+	3

Total Precision Score: _____

Total Quality Score: _____

*P = Partially
U/D = Unable to Determine

Instructions

2. If the main outcomes are first mentioned in the Results section, the question should be answered no.

3. In cohort studies and trials, inclusion and/or exclusion criteria should be given. In case-control studies, a case-definition and the source for controls should be given.

4. Principal confounders is include bipolar disorders, psychoses, substance abuse, and major medical problems. Give one point if some confounders are described and two only if most of these principal confounders are described.

5. Simple outcome data (including denominators and numerators) should be reported for all major findings so that the reader can check the major analyses and conclusions. (This question does not cover statistical tests which are considered below).

6. In non normally distributed data the inter-quartile range of results should be reported. In normally distributed data the standard error, standard deviation or confidence intervals should be reported. If the distribution of the data is not described, it must be assumed that the estimates used were appropriate and the question should be answered yes.

7. This should be answered yes where there were no losses to follow-up or where losses to follow-up were so small that findings would be unaffected by their inclusion. This should be answered no where a study does not report the number of patients lost to follow-up.

9. The study must identify the source population for patients and describe how the patients were selected. Patients would be representative if they comprised the entire source population, an unselected sample of consecutive patients, or a random sample. Random sampling is only feasible where a list of all members of the relevant population exists. Where a study does not report the proportion of the source population from which the patients are derived, the question should be answered as unable to determine.

10. The proportion of those asked who agreed should be stated. Validation that the sample was representative would include demonstrating that the distribution of the main confounding factors was the same in the study sample and the source population.

11. The question should be answered no if, for example, the intervention was undertaken in a specialist center unrepresentative of the hospitals most of the source population would attend.

13. Where follow-up was the same for all study patients the answer should be yes. If different lengths of follow-up were adjusted for by, for example, survival analysis the answer should be yes. Studies where differences in follow-up are ignored should be answered no.

*P = Partially
U/D = Unable to Determine

14. The statistical techniques used must be appropriate to the data. For example, non-parametric methods should be used for small sample sizes. Where little statistical analysis has been undertaken but where there is no evidence of bias, the question should be answered yes. If the distribution of the data (normal or not) is not described it must be assumed that the estimates used were appropriate and the question should be answered yes.

15. For example, patients for all comparison groups should be selected from the same hospital. The question should be answered unable to determine for cohort and case-control studies where there is no information concerning the source of patients included in the study.

16. For a study which does not specify the time period over which patients were recruited, the question should be answered as unable to determine.

17. This question should be answered no for trials if: the main conclusions of the study were based on analyses of treatment rather than intention to treat; the distribution of known confounders in the different treatment groups was not described; or the distribution of known confounders differed between the treatment groups but was not taken into account in the analyses. In non-randomised studies if the effect of the main confounders was not investigated or confounding was demonstrated but no adjustment was made in the final analyses the question should be answered as no.

18. If the numbers of patients lost to follow-up are not reported, the question should be answered as unable to determine. If the proportion lost to follow-up was too small to affect the main findings, the question should be answered yes.

SOURCE: Based on a modified version of the form found in Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomized and non-randomised studies of health care interventions. J Epidemiol Community Health, 1998;52:377-87.

*P = Partially
U/D = Unable to Determine

Quality Checklist for Studies of Screening Instruments/Procedures

Reviewer's initials _____ Article # (from reference manager) _____

First Author _____ Journal:_____

Year published _____

Reporting		**Yes**	**No**
1.	Is the hypothesis/aim/objective of the study clearly described?	1	0
2.	Are the performance measures to be assessed clearly described in the Introduction or Methods sections, including explicit threshold values?	1	0
3.	Are the characteristics of the patients included in the study clearly described?	1	0
4.	Are the instruments/procedures under study clearly described?	1	0

		Yes	**P***	**No**
5.	Are the distributions of principal confounders in each group of subjects to be compared clearly described?	2	1	0

		Yes	**No**
6.	Are the main findings of the study clearly described?	1	0
7.	Does the study provide estimates of the random variability in the data for the main performance measures?	1	0
8.	Have the characteristics of patients excluded because the test was infeasible or result was indeterminate been described?	1	0
9.	Have the actual probability values been reported for the main performance measures except when the probability value is less than 0.001?	1	0

Total Reporting Score: _____

*P = Partially
U/D = Unable to Determine

External Validity		**Yes**	**No**
10.	Were the subjects asked to participate in the study representative of the entire population from which they were recruited?	1	0
11.	Were those subjects who were prepared to participate representative of the entire population from which they were recruited?	1	0
12.	Were the staff, places, and facilities where the patients were screened representative of the settings in which the instrument/procedures will be used?	1	0

Total External Validity Score: _____

Internal Validity		**Yes**	**No**
13.	Was the test compared with a valid reference standard?	1	0
14.	Were the test and reference standard measured independently (blind) of each other?		
	2 Test measured independently of reference standard and reference standard independently of test (MOST VALID)		
	1 Test measured independently of reference standard but not vice versa		
	1 Reference standard measured independently of test but not vice versa		
	0 Test and reference standard not measured independently of each other (LEAST VALID)		
15.	Was the choice of patients who were assessed by the reference standard independent of the test's results? (Avoidance of verification bias)	1	0
16.	Was the test measured independently of all other clinical information?	1	0
17.	Was the reference standard measured before any interventions were stated with knowledge of test results? (Avoidance of treatment paradox)	1	0
18.	Were tests compared in a valid design? Categories are:		
	2 all tests done independently (i.e., blind to the results of the other tests) on each person (MOST VALID)		
	2 different tests done on randomly allocated individuals		

1 all tests done on each person but not assessed independently

0 different tests done on different individuals, not randomly allocated (LEAST VALID)

Total Internal Validity Score: _____

Power

19. Did the study mention having conducted a power analysis to determine the sample size needed to analyze one or more cut-off levels?

No	0
Yes, one cut-off level	1
Yes, two or more cut-off levels	2

Total Score: _____

Source: Based on the Cochrane Methods Working Group on Systematic Review of Screening and Diagnostic Tests: Recommended Methods, updated June 6, 1996. [http://som/flinders.edu.au/cochrane/]

Quality Checklist for RCTs and Observational Studies of Treatment Studies

Reviewer's initials _____ Article # (from reference manager) _____

First Author _____ Journal:_____

Year published _____

Reporting		**Yes**	**No**
1.	Is the hypothesis/aim/objective of the study clearly described?	1	0
2.	Are the main outcomes to be measured clearly described in the Introduction or Methods section?	1	0
3.	Are the characteristics of the patients included in the study clearly described?	1	0
4.	Are the interventions under study clearly described?	1	0

		Yes	**P***	**No**
5.	Are the distributions of principal confounders in each group of subjects to be compared clearly described?	2	1	0

		Yes	**No**
6.	Are the main findings of the study clearly described?	1	0
7.	Does the study provide estimates of the random variability in the data for the main outcomes?	1	0
8.	Have all important adverse events that may be a consequence of the intervention been reported?	1	0
9.	Have the characteristics of patients lost to follow-up been described?	1	0
10.	Have actual probability values been reported (e.g., 0.035 rather than <0.05) for the main outcomes except where the probability value is less than 0.001?	1	0

Total Reporting Score: _____

*P = Partially

U/D = Unable to Determine

External Validity	**Yes**	**No**	**U/D**
11. Were the subjects asked to participate in the study representative of the entire population from which they were recruited?	1	0	0
12. Were those subjects who were prepared to participate representative of the entire population from which they were recruited?	1	0	0
13. Were the staff, places, and facilities where the patients were treated representative of the treatment the majority of patients receive?	1	0	0

Total External Validity Score: _____

Internal Validity - Bias	**Yes**	**No**	**U/D**
14. Was an attempt made to blind study subjects to the intervention they have received?	1	0	0
15. Was an attempt made to blind those measuring the main outcomes of the intervention?	1	0	0
16. If any of the results of the study were based on "data dredging", was this made clear?	1	0	0
17. In trials and cohort studies, do the analyses adjust for different lengths of follow-up of patients, or in case-control studies, is the time period between the intervention and outcome the same for cases and control?	1	0	0
18. Were the statistical tests used to assess the main outcomes appropriate?	1	0	0
19. Was compliance with the intervention/s reliable?	1	0	0
20. Were the main outcome measures used accurate (valid and reliable)?	1	0	0

Total Bias Score: _____

Internal Validity - Confounding	**Yes**	**No**	**U/D**
21. Were the patients in different intervention groups (trials and cohort studies) or were the cases and controls (case-control studies) recruited from the same population?	1	0	0
22. Were study subjects in different intervention groups (trials and cohort studies) or were the cases and controls (case-control studies) recruited over the same period of time?	1	0	0

*P = Partially

U/D = Unable to Determine

23.	Were study subjects randomized to intervention groups?	1	0	0	
24.	Was the randomized intervention assignment concealed from both patients and health care staff until recruitment was complete and irrecoverable?	1	0	0	
25.	Was there adequate adjustment for confounding in the analyses from which the main findings were drawn?	1	0	0	
26.	Were losses of patients to follow-up taken into account?	1	0	0	

Total Confounding Score: _____

Power

27. Did the study mention having conducted a power analysis to determine the sample size needed to detect a significant difference in effect size for one or more outcome measures?

No	0
Yes, one measure	1
Yes, two or more measures	2

Total Score: _____

*P = Partially

U/D = Unable to Determine

Instructions

2. If the main outcomes are first mentioned in the Results section, the question should be answered no.

3. In cohort studies and trials, inclusion and/or exclusion criteria should be given. In case-control studies, a case-definition and the source for controls should be given.

4. Treatments and placebo (where relevant) that are to be compared should be clearly described.

5. Principal confounders is include bipolar disorders, psychoses, substance abuse, and major medical problems. Give one point if some confounders are described and two only if most of these principal confounders are described.

6. Simple outcome data (including denominators and numerators) should be reported for all major findings so that the reader can check the major analyses and conclusions. (This question does not cover statistical tests which are considered below).

7. In non normally distributed data the inter-quartile range of results should be reported. In normally distributed data the standard error, standard deviation or confidence intervals should be reported. If the distribution of the data is not described, it must be assumed that the estimates used were appropriate and the question should be answered yes.

8. This should be answered yes if the study demonstrates that there was a comprehensive attempt to measure adverse events. (A list of possible adverse events is provided.)

9. This should be answered yes where there were no losses to follow-up or where losses to follow-up were so small that findings would be unaffected by their inclusion. This should be answered no where a study does not report the number of patients lost to follow-up.

11. The study must identify the source population for patients and describe how the patients were selected. Patients would be representative if they comprised the entire source population, an unselected sample of consecutive patients, or a random sample. Random sampling is only feasible where a list of all members of the relevant population exists. Where a study does not report the proportion of the source population from which the patients are derived, the question should be answered as unable to determine.

12. The proportion of those asked who agreed should be stated. Validation that the sample was representative would include demonstrating that the distribution of the main confounding factors was the same in the study sample and the source population.

13. For the question to be answered yes, the study should demonstrate that the intervention was representative of that in use in the source population. The question should be

answered no if, for example, the intervention was undertaken in a specialist center unrepresentative of the hospitals most of the source population would attend.

14. For studies where the patients would have no way of knowing which intervention they received, this should be answered yes.

16. Any analyses that had not been planned at the outset of the study should be clearly indicated. If no retrospective unplanned subgroup analyses were reported, then answer yes.

17. Where follow-up was the same for all study patients the answer should be yes. If different lengths of follow-up were adjusted for by, for example, survival analysis the answer should be yes. Studies where differences in follow-up are ignored should be answered no.

18. The statistical techniques used must be appropriate to the data. For example, non-parametric methods should be used for small sample sizes. Where little statistical analysis has been undertaken but where there is no evidence of bias, the question should be answered yes. If the distribution of the data (normal or not) is not described it must be assumed that the estimates used were appropriate and the question should be answered yes.

19. Where there was non compliance with the allocated treatment or where there was contamination of one group, the question should be answered no. For studies where the effect of any misclassification was likely to bias any association to the null, the question should be answered yes.

20. For studies where the outcome measures are clearly described, the question should be answered yes. For studies which refer to other work or that demonstrates the outcome measures are accurate, the question should be answered as yes.

21. For example, patients for all comparison groups should be selected from the same hospital. The question should be answered unable to determine for cohort and case-control studies where there is no information concerning the source of patients included in the study.

22. For a study which does not specify the time period over which patients were recruited, the question should be answered as unable to determine.

23. Studies which state that subjects were randomized should be answered as yes except where method of randomization would not ensure random allocation. For example, alternate allocation would score no because it is predictable.

24. All non-randomised studies should be answered no. If assignment was concealed from patients but not from staff, it should be answered no.

25. This question should be answered no for trials if: the main conclusions of the study were based on analyses of treatment rather than intention to treat; the distribution of known

confounders in the different treatment groups was not described; or the distribution of known confounders differed between the treatment groups but was not taken into account in the analyses. In non-randomised studies if the effect of the main confounders was not investigated or confounding was demonstrated but no adjustment was made in the final analyses the question should be answered as no.

26. If the numbers of patients lost to follow-up are not reported, the question should be answered as unable to determine. If the proportion lost to follow-up was too small to affect the main findings, the question should be answered yes.

SOURCE: Based on a modified version of the form from Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomized and non-randomised studies of health care interventions. J Epidemiol Community Health, 1998;52:377-87.

Perinatal Depression List of Acronyms

Adj	adjusted
B	Bedford
BDI	Beck Depression Inventory
BDI-II	Beck Depression Inventory - II
C	Catego
CCEI	Crown-Crisp Experiental Index
CES	Current Experience Scale
CES-D	Center for Epidemiological Studies – Depression Scale
CI	confidence interval
CIDI-A	Composite International Diagnostic Interview - Auto
Dept	department
DIS	diagnostic interview schedule
DMC	Dyadic Mutality Code
DSM-III	Diagnostic and Statistical Manual for Mental Disorders, Third Edition
DSM-III-R	Diagnostic and Statistical Manual for Mental Disorders, Third Edition - Revised
DSM-IV	Diagnostic and Statistical Manual for Mental Disorders, Fourth Edition
dx	diagnosis
EPDS	Edinburgh Postnatal Depression Scale
GA	gestational age
GHQ	General Health Questionnaire
GHQ-D	General Health Questionnaire - Depression
GP	general practitioner
GP/psych	general practitioner/psychiatrist
HDRS	Hamilton Depression Rating Scale
HMO	health maintenance organization
HOME	Home Observation for Measurement of Environment
hr(s)	hour(s)
HS	high school
ICD-9	International Classification of Diseases, Ninth Edition
IDD-10	International Classification of Disease, Tenth Edition

IDS	Inventory of Depressive Symptomology
LQ	Leverton Questionnaire
MAACL	Multiple Affect Adjective Check List
MADRS	Montgomery and Asberg Depression Rating Scale
MDE	major depressive episode
MINI	Mini International Neuropsychiatric Interview
MINI-V4.4	Mini International Neuropsychiatric Interview, Version 4.4
mo(s)	month(s)
NA	not applicable
NICU	Neonatal Intensive Care Unit
No.	number
NPV	negative predictive value
NR	not reported
NS	not significant
Ob-Gyn	obstretrics and gynecology
OR	odds ratio
PAS	Psychiatric Assessment Schedule
PDSS	Postpartum Depression Screening Scale
PEG	Psycho Educational Group
PP	Postpartum
PPG	Postpartum Guidelines
PSE	Present State Examination
PSE-ID	Present State Examination – Index of Definition
RCT	randomized controlled trials
RDC	research diagnostic criteria
SADS	Schedule for Affective Disorders and Schizophrenia
SADS-C	Schedule for Affective Disorders and Schizophrenia – Change version
SADS-L	Schedule for Affective Disorders and Schizophrenia – Long
SCAN	Schedules for Clinical Assessment in Neuropsychiatry
SCID	Structured Clinical Interview for DSM-IV
SCID-German	Structured Clinical Interview for DSM-IV – German
SCIP-NP	Structured Clinical Interview for DSM-III-R – non-patient
SCLR-90	Symptom checklist Revised - 1990

SD	standard deviation
Sensi	sensitivity
Speci	specificity
SIDS	Sudden Infant Death Syndrome
SPI	Standardized Psychiatric Interview
SRQ	self-reported questionnaire
TSH	thyroid stimulating hormone
UK	United Kingdom
Univ.	University
USA	United States of America
vs.	versus
wk(s)	week(s)
yr(s)	year(s)

Evidence Table 1. Key Question 1: Study design for studies of the prevalence and incidence of perinatal depression

Author, Year Quality Rating	Study Characteristics	Population Description	Depression Measure
Prospective Studies without Comparison Groups			
Author Affonso et al., 1990 **Quality rating** 8	**Study design** Prospective cohort **Sample size** 202 women **Place** California	**Population** Women recruited from 3 HMO clinics at their 1st prenatal care visit **Age** Mean: 30 yrs Range: 20 to 40 yrs **Race/ethnicity** White: 76% Black: 8% Asian: 6% Hispanic: 6% Other: 4% **Inclusion/exclusion criteria** *Included:* Primigravida, viable fetus, married or living with the infant's father, and no depression episode 12 mos prior to pregnancy. *Excluded:* Those with a depression episode within the past 2 yrs if younger than 20 or within the past 5 yrs if over 20 and those undergoing therapy for 3 continuous mos	**Diagnoses** Major and minor **Interview subjects** All study women **Clinical instrument** SADS modified for PPG **Diagnostic criteria** RDC **Interview times** 10 to 14 wks GA, 30 to 32 wks GA, 1 to 2 wks PP, and 14 wks PP

Evidence Table 1. Key Question 1: Study design for studies of the prevalence and incidence of perinatal depression (continued)

Type of Estimate	Time Period of Estimate	Diagnoses Included	Confounders	No. Depressed / Total No. (%)
Prospective Cohort Studies without Comparison Groups				
Point	10 to 14 wks GA	Major and minor	Low risk	3/202 (**1.5%**)
	10 to 14 wks GA	Major	Low risk	2/202 (**1.0%**)
	30 to 32 wks GA	Major and minor	Low risk	1/202 (**0.5%**)
	30 to 32 wks GA	Major	Low risk	0/202 (**0.0%**)
	1 to 2 wks PP	Major and minor	Low risk	3/202 (**1.5%**)
	1 to 2 wks PP	Major	Low risk	2/202 (**1.0%**)
	14 wks PP	Major and minor	Low risk	3/202 (**1.5%**)
	14 wks PP	Major	Low risk	0/202 (**0.0%**)

Evidence Table 1. Key Question 1: Study design for studies of the prevalence and incidence of perinatal depression (continued)

Author, Year Quality Rating	Study Characteristics	Population Description	Depression Measure
Author Areias et al., 1996 **Quality rating** 12	**Study design** Prospective cohort **Sample size** 54 women and 42 husbands/partners **Place** Portugal	**Population** Pregnant women recruited from 2 prenatal clinics **Age** Mean: 25 yrs Range: 17 to 38 yrs **Race/ethnicity** NR **Inclusion/exclusion criteria** *Included:* Primiparous women ≤ 24 wks GA at entry *Excluded:* Women with inadequate education to complete the questionnaire	**Diagnoses** Major, minor, and intermittent **Interview subjects** All study women **Clinical instrument** SADS or SADS-L **Diagnostic criteria** RDC **Interview times** 6 mos GA and 12 mos PP. A subset of 24 women were also interviewed at 3 mos PP
Author Berle et al., 2003 **Quality rating** 9	**Study design** Prospective cohort **Sample size** 411 women **Place** Norway	**Population** Women attending routine PP visits 6 to 12 wks PP **Age** Mean depressed: 30 yrs Mean nondepressed: 29.8 yrs **Race/ethnicity** NR **Inclusion/exclusion criteria** None	**Diagnoses** Major and minor **Interview subjects** All women scoring ≥ 8 on the EPDS and every 10th woman scoring < 8 **Clinical instrument** MINI-V4.4 and the MADRS **Diagnostic criteria** DSM-IV **Interview times** 6 to 12 wks PP

Evidence Table 1. Key Question 1: Study design for studies of the prevalence and incidence of perinatal depression (continued)

Type of Estimate	Time Period of Estimate	Diagnoses Included	Confounders	No. Depressed / Total No. (%)
New episode	During pregnancy	Major and minor	NR	**5**/52 (**9.6%**)
	During pregnancy	Major	NR	**2**/52 (**3.8%**)
	0 to 6 mos GA	Major and minor	Intermittent dx	3/52 (**5.8%**)
	0 to 3 mos PP	Major and minor	NR	12/49 (**24.5%**)
	4 to 12 mos PP	Major and minor	NR	12/46 (**26.0%**)
	0 to 12 mos PP	Major and minor	NR	**24**/49 (**49.0%**)
	0 to 12 mos PP	Major	NR	**15**/49 (**30.6%**)
Period	During pregnancy	Major and minor	Intermittent dx	9/54 (16.7%)
	0 to 6 mos GA	Major and minor	Intermittent dx	5/54 (**9.3%**)
	0 to 3 mos PP	Major and minor	Intermittent dx	17/54 (**31.5%**)
	4 to 12 mos PP	Major and minor	Intermittent dx	20/54 (**37.0%**)
	0 to 12 mos PP	Major and minor	Intermittent dx	29/54 (**53.7%**)
Point	6 to 12 wks PP	Major and minor	EPDS ≥ 9	41/411 (10.0%)
	6 to 12 wks PP	Major	EPDS ≥ 9	27/411 (6.6%)

Evidence Table 1. Key Question 1: Study design for studies of the prevalence and incidence of perinatal depression (continued)

Author, Year Quality Rating	Study Characteristics	Population Description	Depression Measure
Author Campbell and Cohn, 1991 **Quality rating** 12	**Study design** Prospective cohort **Sample size** 1033 women **Place** Pennsylvania	**Population** Women who delivered at Magee-Women's Hospital from 7/1986 to 7/1990 **Age** Mean depressed: 28.5 (SD 3.6) yrs Mean nondepressed: 28.8 (SD 3.6) yrs **Race/ethnicity** White: 100% **Inclusion/exclusion criteria** *Included:* Primiparous women who delivered full-term, single infants without major complications and who were Caucasian, married, 18+ yrs of age, and had at least a high school education. *Excluded:* Women with adopted or stepchildren at home	**Diagnoses** Major and minor **Interview subjects** All study women **Clinical instrument** Modified version of SADS via telephone **Diagnostic criteria** RDC plus had to have depressed mood for ≥ 2 wks and ≥ 3 other symptoms **Interview time** 6 to 8 wks PP

Evidence Table 1. Key Question 1: Study design for studies of the prevalence and incidence of perinatal depression (continued)

Type of Estimate	Time Period of Estimate	Diagnoses Included	Confounders	No. Depressed / Total No. (%)
Point	6 to 8 wks PP	Major and minor	Low risk	57/1,033 (**5.5**%)
Period	0 to 6 to 8 wks PP	Major and minor	Low risk	96/1,033 (**9.3**%)
	0 to 6 to 8 wks PP	Major	Low risk	36/1,033 (**3.5%**)

Evidence Table 1. Key Question 1: Study design for studies of the prevalence and incidence of perinatal depression (continued)

Author, Year Quality Rating	Study Characteristics	Population Description	Depression Measure
Author Cooper et al., 1996 **Quality rating** 12	**Study design** Prospective cohort **Sample size** 4,954 women **Place** England	**Population** Women attending the prenatal clinic at the Rosie Maternity Hospital over a 3-yr period **Age** Mean: 28.2 (SD 4.9) yrs **Race/ethnicity** NR **Inclusion/exclusion criteria** *Included*: All women in 1st 24 mos of the study and only primiparous women in the last 12 mos of the study	**Diagnoses** Major **Interview subjects** EPDS mailed at 5 wks PP; women scoring above the cutoff on the EPDS were contacted by telephone. Initially, the cutoff was set at 7 but was changed to 8 and later to 9. **Clinical instrument** Multiparous women were assessed by telephone and primiparous women were assessed in person with the SCID **Diagnostic criteria** DSM-III-R **Interview time** 6 to 10 wks PP
Author Cox et al., 1982 **Quality rating** 11	**Study design** Prospective cohort **Sample size** 105 women **Place** Scotland	**Population** Women who attended prenatal clinics at the Simpson Memorial Maternity Pavilion from 1/1978 to 11/1979 **Age** Mean: 26 yrs **Race/ethnicity** NR **Inclusion/exclusion criteria** *Included:* Women who resided in Edinburgh, were < 20 wks gestation at entry, and delivered a live infant *Excluded:* Women with language difficulties	**Diagnoses** Major and minor **Interview subjects** All study women **Clinical instrument** SPI **Diagnostic criteria** Pitt's criteria **Interview times** At 1st visit to clinic; at or about 35 wks GA; within 10 days of delivery; within 3 to 5 mos of delivery

Evidence Table 1. Key Question 1: Study design for studies of the prevalence and incidence of perinatal depression (continued)

Type of Estimate	Time Period of Estimate	Diagnoses Included	Confounders	No. Depressed / Total No. (%)
Point	6 to 10 wks PP	Major	EPDS ≥ 8	756/4,954 (15.3%)
Period	1 wk to 3 to 5 mos PP	Major and minor	NR	30/103 (**29.1%**)
	1 wk to 3 to 5 mos PP	Major	NR	13/103 (**12.6%**)

Evidence Table 1. Key Question 1: Study design for studies of the prevalence and incidence of perinatal depression (continued)

Author, Year Quality Rating	Study Characteristics	Population Description	Depression Measure
Author Garcia-Esteve et al., 2003 **Quality rating** 13	**Study design** Prospective cohort **Sample size** 1123 women **Place** Barcelona, Spain	**Population** Women attending routine PP checkups at 6 wks PP at the public Maternity Hospital of Barcelona **Age** Mean depressed: 29.8 yrs Mean nondepressed: 30.2 yrs **Race/ethnicity** NR **Inclusion/exclusion criteria** *Included:* Spanish-speaking women *Excluded:* Those suffering from mourning or organic depression	**Diagnoses** Major and minor **Interview subjects** Women scoring ≥ 9 on the EPDS and a random 10% sample of those scoring < 9 **Diagnostic instrument** Modified SCID-NP **Diagnostic criteria** DSM-IV **Interview time** 6 wks PP
Author Gotlib et al., 1989 **Quality rating** 11	**Study design** Prospective cohort **Sample size** 295 women **Place** Canada	**Population** Participants were recruited from a consecutive series of pregnant patients through the obstetrics department of a large, urban hospital and from the private practices of more than 15 physicians **Age** Mean: 27.8 (SD 3.4) yrs Range: 18 to 40 yrs **Race/ethnicity** White: 90% **Inclusion/exclusion criteria** None	**Diagnoses** Major and minor **Interview subjects** All women with BDI ≥ 10 and 34 women with BDI < 10 **Clinical instrument** Shortened version of SADS via telephone by a clinical psychologist **Diagnostic criteria** RDC **Interview times** 24 and 36 wks GA and 4 wks PP

Evidence Table 1. Key Question 1: Study design for studies of the prevalence and incidence of perinatal depression (continued)

Type of Estimate	Time Period of Estimate	Diagnoses Included	Confounders	No. Depressed / Total No. (%)
Point	6 wks	Major and minor	EPDS ≥ 9	100/1123(8.9**%**)
	6 wks	Major	EPDS ≥ 9	36/1123 (**3.2%**)
New episode	24 to 36 wks GA	Major and minor	BDI ≥ 10	5/270 (**1.9%**)
	36 wks GA to 4 wks PP	Major and minor	BDI ≥ 10	10/285 (**3.5**%)
Point	24 wks GA	Major and minor	BDI ≥ 10	**27**/295 (9.2%)
	36 wks GA	Major and minor	BDI ≥ 10	**24**/295 (8.0%)
	4 wks PP	Major and minor	BDI ≥ 10	20/295 (6.8%)
	4 wks PP	Major and minor	Adj for BDI	**38**/295 (13%)
Period	24 to 36 wks GA	Major and minor	BDI ≥ 10	30/295 (10.2%)

Evidence Table 1. Key Question 1: Study design for studies of the prevalence and incidence of perinatal depression (continued)

Author, Year Quality Rating	Study Characteristics	Population Description	Depression Measure
Author Hobfoll et al., 1995 **Quality rating** 12	**Study design** Prospective cohort **Sample size** 192 women **Place** Ohio	**Population** Women were recruited over 2.5 yrs from a randomly selected sample of the patient population meeting study criteria at 3 obstetrics clinics for low-income women in a mid-sized Midwestern city **Age** Mean: 24.5 (SD 5.1) yrs Range: 17 to 40 yrs **Race/ethnicity** African American: 27% European American: 73% **Inclusion/exclusion criteria** *Included:* 17 to 40 yrs of age, 16 to 24 wks GA at entry, free of serious medical complications, and of either African American or European American descent	**Diagnoses** Major and minor **Interview subjects** All study women **Clinical instrument** Modified version of the SADS **Diagnostic criteria** RDC **Interview times** During the 2nd and 3rd trimesters and at 7 to 9 wks PP
Author Kent et al., 1999 **Quality rating** 12	**Study design** Prospective cohort **Sample size** 710 women **Place** Australia	**Population** Women were randomly selected from the Western Australian Midwives' Notification System database **Age** Range: 20 to 45 yrs **Race/ethnicity** White: 100% **Inclusion/exclusion criteria** *Included:* Caucasian women aged 20 to 45 yrs at 4.5 to 5.5 mos PP who were residents of Perth and had no traumatic birth events *Excluded:* Women with pre-existing thyroid disease, PP thyroid dysfunction on thyroid medication prior to entry, and women with normal TSH but low free T_4	**Diagnoses** Major **Interview subjects** All women with GHQ28 > 4 **Clinical instrument** CIDI-A **Diagnostic criteria** DSM-III-R **Interview time** 6 mos PP

Evidence Table 1. Key Question 1: Study design for studies of the prevalence and incidence of perinatal depression (continued)

Type of Estimate	Time Period of Estimate	Diagnoses Included	Confounders	No. Depressed / Total No. (%)
New episode	2nd or 3rd trimester to 7 to 9 wks PP	Major and minor	Low income	21/168 (**12.5%**)
	2nd or 3rd trimester to 7 to 9 wks PP	Major	Low income	5/168 (**3.0%)**
Point	2nd trimester	Major and minor	Low income	53/192 (27.6%)
	2nd trimester	Major	Low income	22/192 (11.5%)
	3rd trimester	Major and minor	Low income	47/192 (24.5%)
	3rd trimester	Major	Low income	8/192 (4.2%)
	7 to 9 wks PP	Major and minor	Low income	45/192 (23.4%)
	7 to 9 wks PP	Major	Low income	14/192 (7.3%)
Point	6 mos PP	Major	GHQ28 > 4 only	67/710 (9.4%)

Evidence Table 1. Key Question 1: Study design for studies of the prevalence and incidence of perinatal depression (continued)

Author, Year Quality Rating	Study Characteristics	Population Description	Depression Measure
Author Kitamura et al., 1993 **Quality rating** 13	**Study design** Prospective cohort **Sample size** 120 women **Place** Japan	**Population** Women were recruited from patients at a prenatal clinic in the obstetrics department of a general hospital on 2 given days per wk **Age** Mean: 27.9 (SD 4.6) yrs Range: 17 to 42 yrs **Race/ethnicity** NR **Inclusion/exclusion criteria** *Excluded:* Women > 12 wks GA	**Diagnoses** Major and minor **Interview subjects** All study women **Clinical instrument** SADS and SADS-C administered by a psychiatrist **Diagnostic criteria** RDC **Interview times** Early (when fetal heart beat was 1st confirmed) and late (about 34 wks GA) pregnancy
Author Kitamura et al., 1999 **Quality rating** 10	**Study design** Prospective cohort **Sample size** 111 women **Place** Japan	**Population** Women attending a prenatal clinic at a general hospital in an industrial city between 8/1984 and 2/1986 **Age** Mean: 28 (SD 5) yrs Range: 17 to 42 yrs **Race/ethnicity** NR **Inclusion/exclusion criteria** *Included:* > 12 wks GA and planning to give birth in another hospital	**Diagnoses** Major and minor **Interview subjects** All study women **Clinical instrument** SADS administered by a psychiatrist **Diagnostic criteria** RDC **Interview times** 1st and 3rd trimesters, on the 5th day PP, and 1 mo PP

Evidence Table 1. Key Question 1: Study design for studies of the prevalence and incidence of perinatal depression (continued)

Type of Estimate	Time Period of Estimate	Diagnoses Included	Confounders	No. Depressed / Total No. (%)
New episode	During pregnancy	Major and minor	NR	19/108 (**17.6%**)
	During pregnancy	Major	NR	13/108 (**12.0%**)
	1st trimester	Major and minor	NR	13/108 (**12.0%**)
Period	During pregnancy	Major and minor	NR	21/110 (**19.1%**)
	During pregnancy	Major	NR	14/110 (**12.7%**)
Point	1st trimester	Major and minor	NR	10/111 (**9.0%**)
	1st trimester	Major	NR	7/111 (**6.3%)**
	3rd trimester	Major and minor	NR	9/102 (**8.8%**)
	3rd trimester	Major	NR	4/102 (**3.9%**)
	5 days PP	Major and minor	NR	5/91 (**5.5%**)
	5 days PP	Major	NR	0/91(**0.0%**)
	1 mo PP	Major and minor	NR	7/101 (**6.9%**)
	1 mo PP	Major	NR	3/101 (**3.0%**)

Evidence Table 1. Key Question 1: Study design for studies of the prevalence and incidence of perinatal depression (continued)

Author, Year Quality Rating	Study Characteristics	Population Description	Depression Measure
Author Kumar and Robson, 1984 **Quality rating** 11	**Study design** Prospective cohort **Sample size** 119 entering during pregnancy and 77 entering PP **Place** England	**Population** Primiparous women booked at a prenatal clinic. Additional primiparous and multiparous women meeting other criteria above were enrolled from PP wards after delivery **Age** Mean: 28 yrs Range: 19 to 40 yrs **Race/ethnicity** NR **Inclusion/exclusion criteria** *Included:* Primiparous women who were < 12 to 14 wks GA, married or had stable common-law partners, had spent 5 of previous 10 yrs in Britain, and resided within a reasonable distance of central London	**Diagnoses** Major and minor **Interview subjects** All study women **Clinical instrument** SPI **Diagnostic criteria** RDC **Interview times** Within a 4-wk base period at entry, 12 wks, and 1 yr PP; subjects rated as "cases" during 1st trimester were also followed up at 24 or 36 wks GA, and those rated as cases at 12 wks PP were interviewed at 1 yr
Author Lee et al., 2001 **Quality rating** 12	**Study design** Prospective cohort **Sample size** 781 women **Place** Hong Kong	**Population** Women recruited from consecutive patients admitted to the prenatal booking clinic of the Prince of Wales Hospital for their 1st prenatal visit **Age** Mean: 29.0 (SD 4.9) yrs for women assessed at 3 mos Mean: 29.2 (SD 4.9) yrs for women not assessed at 3 mos **Race/ethnicity** Chinese: 100% **Inclusion/exclusion criteria** *Excluded:* Women who were not of Chinese ethnicity or were not permanent residents of Hong Kong	**Diagnoses** Major and minor **Interview subjects** Women with GHQ ≥ 5 and a random sample of women with GHQ < 5 **Clinical instrument** SCID **Diagnostic criteria** DSM-III-R **Interview time** 3 mos PP

Evidence Table 1. Key Question 1: Study design for studies of the prevalence and incidence of perinatal depression (continued)

Type of Estimate	Time Period of Estimate	Diagnoses Included	Confounders	No. Depressed / Total No. (%)
New episode	Preconception to 12 wks GA	Major and minor	NR	12/112 (**10.7%**)
	12 wks GA to 24 wks GA	Major and minor	NR	3/113 (**2.7%**)
	24 wks GA to 36 wks GA	Major and minor	NR	3/114 (**2.6%)**
	3rd trimester to12 wks PP	Major and minor	NR	15/114 (**13.2%**)
Point	12 wks GA	Major and minor	NR	15/119 (**12.5%**)
	12 wks GA	Major	NR	1/119 (**0.8%**)
	24 wks GA	Major and minor	NR	9/119 (**7.6%**)
	36 wks GA	Major and minor	NR	7/119 (**5.9%**)
	12 wks PP	Major and minor	NR	16/114 (**14.0%**)
	12 wks PP	Major	NR	3/114 (**2.6%**)
	28 wks PP	Major and minor	NR	15/112 (**13.4%**)
	1 yr PP	Major and minor	NR	7/108 (**6.5%**)
Period	During pregnancy	Major and minor	NR	22/119 (**18.5%**)
New episode	0 to 1 mo PP	Major and minor	GHQ > 4 Chinese only	9.4% (SE 2.0)
	0 to 1 mo PP	Major	GHQ > 4 Chinese only	5.0% (SE 2.0)
	0 to 3 mos PP	Major and minor	GHQ > 4 Chinese only	10.4% (SE 2.8)
	0 to 3 mos PP	Major	GHQ > 4 Chinese only	5.6% (SE 2.0)
Point	1 mo PP	Major and minor	GHQ > 4 Chinese only	10.3% (SE 2.8)
	1 mo PP	Major	GHQ > 4 Chinese only	5.5% (SE 2.0)
	3 mos PP	Major and minor	GHQ > 4 Chinese only	11.2% (SE 2.8)
	3 mos PP	Major	GHQ > 4 Chinese only	6.1% (SE 2.0)

Evidence Table 1. Key Question 1: Study design for studies of the prevalence and incidence of perinatal depression (continued)

Author, Year Quality Rating	Study Characteristics	Population Description	Depression Measure
Author Lee et al., 2001 **Quality rating** 12	**Study design** Prospective cohort **Sample size** 145 women **Place** Hong Kong	**Population** Women admitted to the PP wards of the Prince of Wales Hospital over a 3-mo period **Age** Mean: 29 yrs Range: 16 to 42 yrs **Race/ethnicity** Chinese: 100% **Inclusion/exclusion criteria** *Excluded:* Non-Chinese women and those without permanent residency rights in Hong Kong	**Diagnoses** Major and minor **Interview subjects** All study women **Clinical instrument** Chinese version of SCID **Diagnostic criteria** DSM-III-R **Interview time** 6 wks PP
Author Lucas et al., 2001 **Quality rating** 9	**Study design** Prospective cohort **Sample size** 641 women **Place** Spain	**Population** Consecutive series of healthy Caucasian women recruited between their 36th wk of pregnancy and 4th day PP at a university hospital from 3/1993 to 6/1997 **Age** Mean: 28 (SD 4.6) yrs Range: 17 to 42 yrs **Race/ethnicity** White: 100% **Inclusion/exclusion criteria** *Excluded:* Women with concurrent autoimmune disease and previous thyroid disease	**Diagnoses** Major **Interview subjects** Women with BDI > 21 **Clinical instrument** Clinical evaluation by psychiatrist with unspecified instrument **Diagnostic criteria** DSM-III-R **Interview times** 1, 3, 6, 9, and 12 mos PP

Evidence Table 1. Key Question 1: Study design for studies of the prevalence and incidence of perinatal depression (continued)

Type of Estimate	Time Period of Estimate	Diagnoses Included	Confounders	No. Depressed / Total No. (%)
Point	6 wks PP	Major and minor	Chinese only	17/145 (11.7%)
	6 wks PP	Major	Chinese only	8/145 (5.5%)
Point	1 mo PP	Major	BDI > 21	3/605 (**0.5%**)
	3 mos PP	Major	BDI > 21	5/552 (**0.9%**)
	6 mos PP	Major	BDI > 21	3/574 (**0.5%**)
	9 mos PP	Major	BDI > 21	0/431 (**0.0%**)
	12 mos PP	Major	BDI > 21	0/444 (**0.0%**)
Period	0 to 1 yr PP	Major	BDI > 21	11/641 (1.7%)

Evidence Table 1. Key Question 1: Study design for studies of the prevalence and incidence of perinatal depression (continued)

Author, Year Quality Rating	Study Characteristics	Population Description	Depression Measure
Author Matthey et al., 2003 **Quality rating** 11	**Study design** Prospective cohort **Sample size** Sample 1: 216 Sample *2:* 192 **Place** Australia	**Population** 1st-time mothers were recruited from prenatal classes in a public hospital **Age** *Sample 1:* Mean: 27.2 (SD 4.2) yrs Range: 18 to 41 yrs *Sample 2:* Mean: 27.5 (SD 3.5) yrs Range: 19 to 38 yrs **Race/ethnicity** NR **Inclusion/exclusion criteria** None	**Diagnoses** Major and minor **Interview subjects** All study women **Clinical instrument** DIS **Diagnostic criteria** DSM-IV **Interview time** 6 wks PP
Author Murray and Cox, 1990 **Quality rating** 10	**Study design** Prospective cohort **Sample size** 100 women **Place** England	**Population:** Convenience sample of women attending the prenatal clinic of the North Staffordshire Maternity Hospital **Age** Mean: 24.6 yrs **Race/ethnicity** NR **Inclusion/exclusion criteria** *Included:* Women 28 to 34 wks GA	**Diagnoses** Major and minor **Interview subjects** All study women **Clinical instrument** Modified SPI **Diagnostic criteria** RDC **Interview time** 3rd trimester

Evidence Table 1. Key Question 1: Study design for studies of the prevalence and incidence of perinatal depression (continued)

Type of Estimate	Time Period of Estimate	Diagnoses Included	Confounders	No. Depressed / Total No. (%)
Period Sample 1	0 to 6 wks PP	Major and minor	NR	21/216 (9.7%)
Period Sample 2	0 to 6 wks PP	Major and minor	NR	9/192 (4.7%)
Point	3rd trimester	Major and minor	NR	14/100 (14.0%)
	3rd trimester	Major	NR	6/100 (6.0%)

Evidence Table 1. Key Question 1: Study design for studies of the prevalence and incidence of perinatal depression (continued)

Author, Year Quality Rating	Study Characteristics	Population Description	Depression Measure
Author O'Hara et al., 1984 **Quality rating** 10	**Study design** Prospective cohort **Sample size** 99 women **Place** Iowa	**Population** Women recruited in the 2nd trimester of pregnancy from a public ob-gyn clinic and 2 private practices at the Univ. of Iowa Hospital and Clinics **Age** Mean: 26.5 (SD 4.2) yrs **Race/ethnicity** White: 98% **Inclusion/exclusion criteria** *Included:* Married and 18+ yrs of age	**Diagnoses** Major and minor **Interview subjects** All study women **Clinical instrument** Modified SADS **Diagnostic criteria** RDC **Interview times** 2nd trimester and 9 wks PP
Author Pop et al., 1993 **Quality rating** 13	**Study design** Prospective cohort **Sample size** 293 women **Place** The Netherlands	**Population** Caucasian women registered for prenatal care between 11/1988 and 4/1989 in a semi-urban, semi-rural area **Age** Multipara mean: 30.5 yrs Primipara mean: 27.4 yrs **Race/ethnicity** White: 100% **Inclusion/exclusion criteria** None	**Diagnoses** Major and minor **Interview subjects** All study women **Clinical interview** NR **Diagnostic criteria** RDC **Interview times** 32 wks GA and 4, 10, 16, 22, 28, and 34 wks PP

Evidence Table 1. Key Question 1: Study design for studies of the prevalence and incidence of perinatal depression (continued)

Type of Estimate	Time Period of Estimate	Diagnoses Included	Confounders	No. Depressed / Total No. (%)
New episode	0 to 9 wks PP	Major and minor	NR	10/97 (**10.3%**)
Point	2nd trimester	Major and minor	NR	9/99 (**9.1%**)
	2nd trimester	Major	NR	6/99 (**6.1%**)
Period	0 to 9 wks PP	Major and minor	NR	12/99 (**12.1%**)
	0 to 9 wks PP	Major	NR	8/99 (**8.1%**)
Point	32 wks GA	Major and minor	NR	21/293 (**7.2%**)
	4 wks PP	Major and minor	NR	27/293 (**9.2%**)
	10 wks PP	Major and minor	NR	41/293 (**14.0%**)
	16 wks PP	Major and minor	NR	31/293 (**10.6%**)
	22 wks PP	Major and minor	NR	31/293 (**10.6%**)
	28 wks PP	Major and minor	NR	26/293 (**8.9%**)
	34 wks PP	Major and minor	NR	19/293 (**6.5%**)
	32 wks GA	Major	NR	**3**/293 (1.0%)
	4 wks PP	Major	NR	**6**/293 (2.0%)
	10 wks PP	Major	NR	**12**/293 (4.1%)
	16 wks PP	Major	NR	**7**/293 (2.4%)
	22 wks PP	Major	NR	**6**/293 (2.0%)
	28 wks PP	Major	NR	**9**/293 (3.1%)
	34 wks PP	Major	NR	**3**/293 (1.0%)
Period	0 to 34 wks PP	Major and minor	NR	61/293 (20.8%)
	0 to 34 wks PP	Major	NR	20/293 (6.8%)

Evidence Table 1. Key Question 1: Study design for studies of the prevalence and incidence of perinatal depression (continued)

Author, Year Quality Rating	Study Characteristics	Population Description	Depression Measure
Author Watson et al., 1984 **Quality rating** 13	**Study design** Prospective cohort **Sample size** 128 women **Place** England	**Population** Women who attended a prenatal clinic in South London inner-city area for the 1st time between 9/1977 and 9/1978 **Age** Multipara mean: 28 yrs Primipara mean: 24 yrs **Race/ethnicity** UK origin: 94 (73%) West Indian origin: 17 (13%) Other: 18 (14%) **Inclusion/exclusion criteria** *Excluded:* Those who spoke insufficient English, were past the 24th wk of pregnancy, or were planning to move out of the area in the near future	**Diagnoses** Major **Interview subjects** All study women **Clinical instrument** SPI **Diagnostic criteria** ICD-9 **Interview times** 16 wks of GA and 6 wks PP
Author Whiffen, 1988 **Quality rating** 10	**Study design** Prospective cohort **Sample size** 115 women **Place** Canada	**Population** Women were recruited during their 3rd trimester through public health prenatal classes **Age** Mean: 28.1 (SD 3.9) yrs **Race/ethnicity** NR **Inclusion/exclusion criteria** *Included:* Primiparous, 18+ yrs old, intending to deliver in a hospital, involved in a marital or common-law relationship, and at low risk for complications *Excluded:* Women with inadequate English language skills	**Diagnostic criteria** Major and minor **Interview subjects** All study women **Clinical instrument** Modified version of the SADS **Diagnostic criteria** RDC **Interview times** 35 wks GA and 6 to 8 wks PP

Evidence Table 1. Key Question 1: Study design for studies of the prevalence and incidence of perinatal depression (continued)

Type of Estimate	Time Period of Estimate	Diagnoses Included	Confounders	No. Depressed / Total No. (%)
New episode	1st trimester to birth	Major	NR	7/123 (**5.7%**)
	0 to 6 wks PP	Major	NR	10/123 (**8.1%**)
Point	16 wks GA	Major	NR	5/128 (**3.9%**)
	6 wks PP	Major	NR	15/128 (**11.7%**)
Period	1st trimester to birth	Major	NR	12/128 (**9.4%**)
	0 to 1 yr PP	Major	NR	28/128 (**21.9%**)
Point	6 to 8 wks PP	Major and minor	NR	**19**/115 (16.5%)
	6 to 8 wks PP	Major	NR	**7**/115 (6.1%)

Evidence Table 1. Key Question 1: Study design for studies of the prevalence and incidence of perinatal depression (continued)

Author, Year Quality Rating	Study Characteristics	Population Description	Depression Measure
Author Yamashita et al., 2000 **Quality rating** 10	**Study design** Prospective cohort **Sample size** 88 women **Place** Japan	**Population** Consecutive patients admitted for delivery to the maternity ward of Kyushu Univ. Hospital from 12/1994 to 12/1996 **Age** Mean: 31 yrs Range: 19 to 41 yrs **Race/ethnicity** Asian: Assumed 100% **Inclusion/exclusion criteria** None	**Diagnoses** Major and minor **Interview subjects** All study women **Clinical instrument** SADS via telephone by psychiatrists **Diagnostic criteria** RDC **Interview times** 3 wks and 3 mos PP
Author Yonkers et al., 2001 **Quality rating** 14	**Study design** Prospective cohort **Sample size** 802 women **Place** Texas	**Population** Consecutive patients who came for their initial PP appointments on selected days at 4 publicly funded inner-city community maternal health clinics **Age** Mean: 24.2 (SD 5.6) yrs **Race/ethnicity** White non-Hispanic: 20 (2%) African American: 162 (20%) Hispanic: 604 (75%) Asian: 5 (<1%) Other: 11(1%) **Inclusion/exclusion criteria** *Included:* Had completed their pregnancies or miscarried	**Diagnoses** Major and minor **Interview subjects** Women with IDS ≥ 18 or EPDS ≥ 12 and the 1st 42 participants who screened negative for depressive symptoms. Also 50 randomly selected patients who did not go to the clinics at the appointed times were screened by telephone to investigate whether depressed women are less likely to keep their clinic appointments **Clinical instrument** SCID **Diagnostic criteria** DSM-IV **Interview times** 4 to 5 wks PP

Evidence Table 1. Key Question 1: Study design for studies of the prevalence and incidence of perinatal depression (continued)

Type of Estimate	Time Period of Estimate	Diagnoses Included	Confounders	No. Depressed / Total No. (%)
New episode	0 to 3 wks PP	Major and minor	NR	12/88 (**13.6%**)
	0 to 3 wks PP	Major	NR	5/88 (**5.7%**)
	0 to 3 mos PP	Major and minor	NR	15/88 (**17.0%**)
	0 to 3 mos PP	Major	NR	7/88 (**8.0%**)
Point	3 wks PP	Major and minor	NR	12/88 (**13.6%**)
	3 wks PP	Major	NR	5/88 (**5.7%**)
	3 mos PP	Major and minor	NR	9/88 (**10.2%**)
	3 mos PP	Major	NR	6/88 (**6.8%**)
Period	0 to 3 wks PP	Major and minor	NR	12/88 (**13.6%**)
	0 to 3 wks PP	Major	NR	5/88 (**5.7%**)
	0 to 3 mos PP	Major and minor	NR	15/88 (**17.0%**)
	0 to 3 mos PP	Major	NR	7/88 (**8.0%**)
New episode	0 to 1 mo PP	Major	Multiethnic Low income IDS ≥ 18 or EPDS ≥ 12	26/776 (3.3%) With all lost to followup *not* depressed
	0 to 1 mo PP	Major	Multiethnic Low income IDS ≥ 18 or EPDS ≥ 12	**42**/776 (**5.4**%) With all lost to followup depressed
Point	1 mo PP	Major and minor	Multiethnic Low income IDS ≥ 18 or EPDS ≥ 12	**58**/802 (**7.2%**) With all lost to followup *not* depressed
	1 mo PP	Major	Multiethnic Low income IDS ≥ 18 or EPDS ≥ 12	52/802 (6.5%) With all lost to followup *not* depressed
	1 mo PP	Major	Multiethnic Low income IDS ≥ 18 or EPDS ≥ 12	**68**/802 (8.5%) With all lost to followup depressed
	1 mo PP	Major	African American Low income IDS ≥ 18 or EPDS ≥ 12	11/162 (6.8%) With all lost to followup *not* depressed
	1 mo PP	Major	African American Low income IDS ≥ 18 or EPDS ≥ 12	20/162 (12.3%) With all lost to followup depressed
	1 mo PP	Major	Hispanic Low income IDS ≥ 18 or EPDS ≥ 12	29/604 (4.8%) With all lost to followup *not* depressed
	1 mo PP	Major	Hispanic Low income IDS ≥ 18 or EPDS ≥ 12	**45**/604 (7.4%) With all lost to followup depressed

Evidence Table 1. Key Question 1: Study design for studies of the prevalence and incidence of perinatal depression (continued)

Author, Year Quality Rating	Study Characteristics	Population Description	Depression Measure
Author Yoshida et al., 1997 **Quality rating** 11	**Study design** Prospective cohort **Sample size** 98 women **Place** England	**Population** Recruited from prenatal classes for pregnant Japanese women and an advertisement inserted in a maternity guidebook for pregnant Japanese women **Age** Mean: 30.0 (SD 2.7) yrs **Race/ethnicity** Japanese: 100% **Inclusion/exclusion criteria** *Included:* Pregnant Japanese women living in London	**Diagnoses** Major and minor **Interview subjects** All study women **Clinical instrument** Japanese translation of the SADS administered by a Japanese psychiatrist **Diagnostic criteria** RDC **Interview times** 3 mos PP
Prospective Studies with Comparison Groups			
Author Cooper et al., 1988 **Quality rating** 10	**Study design** Prospective cohort study with comparison group **Sample size** 483 cases and 313 controls **Place** England	**Population** Cases were recruited from the appointments diary of the prenatal clinic and the delivery booking diary of the General Practitioner Unit at the John Radcliffe Hospital; every 2nd woman identified was approached. Comparison sample was derived from a community sample of Edinburgh women **Age** Cases mean: 27.2 yrs Comparison sample mean: 28.8 yrs **Race/ethnicity** NR **Inclusion/exclusion criteria** *Included:* Cases had to reside in Oxford City and have an expected delivery date in 9-mo recruitment window; comparison women had to be 16 to 40 yrs of age, not currently pregnant, and not pregnant in previous 12 mos	**Diagnoses** Major **Interview subjects** At about 34 wks GA, all cases were assessed using the GHQ. At 3 mos PP, psychiatric diagnoses were validated for a random sample of cases, all cases with GHQ ≥12, and sample of cases with GHQ < 12. At 6 mos PP, all GHQ ≥ 12 cases and PSE cases from 3-mo interview were reinterviewed. At 12 mos, a random sample of women, all GHQ ≥ 12 cases from prenatal, 3- and 6-mo PP periods, a subsample of GHQ < 12 and all 3- and 6-mo PSE cases were reinterviewed **Clinical instrument** PSE/MADRS **Diagnostic criteria** PSE ID/Cartego class **Interview times** 3, 6, and 12 mos PP

Evidence Table 1. Key Question 1: Study design for studies of the prevalence and incidence of perinatal depression (continued)

Type of Estimate	Time Period of Estimate	Diagnoses Included	Confounders	No. Depressed / Total No. (%)
New episode	0 to 3 mos PP	Major and minor	NR	12/98 (12.2%)
	0 to 3 mos PP	Major	NR	6/98 (**6.1%**)
Period	0 to 3 mos PP	Major and minor	NR	12/98 (12.2%)
	0 to 3 mos PP	Major	NR	6/98 (**6.1%**)
Prospective Studies with Comparison Groups				
Point *cases*	3 mos PP	Major	(1)	**10**/460 (**2.2%**)
	6 mos PP	Major	(1)	**17**/442 (**3.8%**)
	12 mos PP	Major	(1)	**6**/462 (**1.3%**)
Point *controls*	3 mos PP	Major	(1)	**8**/313 (**2.6%**)
	6 mos PP	Major	(1)	**8**/313 (**2.6%**)
	12 mos PP	Major	(1)	**8**/313 (**2.6%**)

Evidence Table 1. Key Question 1: Study design for studies of the prevalence and incidence of perinatal depression (continued)

Author, Year Quality Rating	Study Characteristics	Population Description	Depression Measure
Author Cox et al., 1993 **Quality rating** 12	**Study design** Prospective case/control **Sample size** 232 cases and 232 controls **Place** England	**Population** The index group comprised random samples of women from the prenatal clinic lists of the North Staffordshire Hospital and women who had not been seen in a prenatal clinic identified from the birth register. The control-group women were recruited from 4 general practice age/sex registers and were 16 to 45 yrs of age, not currently pregnant, and had NR birth in the previous 12 mos **Age** Case mean: 25.4 (SD 5.2) yrs Controls mean: 27.2 (SD 4.7) yrs **Race/ethnicity** NR **Inclusion/exclusion criteria** None	**Diagnoses** Major and minor **Interview subjects** Women with EPDS ≥ 9 and a sample of women with EPDS < 9 **Clinical instrument** *S*PI during home visit **Diagnostic criteria** RDC **Interview time** 6 mos PP
Author O'Hara, 1990 **Quality rating** 13	**Study design** Prospective case/control **Sample size** 182 cases and 179 controls **Place** Iowa	**Population** Women recruited from a public ob-gyn clinic and 2 private practices at the Univ. of Iowa Hospitals and Clinics. Each subject was asked to provide the names of 5 acquaintances who were similar in age, marital status, work status, and had a similar number of children. The acquaintance most similar to the subjects was selected as a control **Age** Case mean: 27.02 (SD 4.71) yrs Controls mean: 27.51 (SD 5.04) yrs **Race/ethnicity** NR **Inclusion/exclusion criteria** *Included:* 18+ yrs of age	**Diagnoses** Major and minor **Interview subjects** All study women **Clinical instrument** SADS and SADS-L **Diagnostic criteria** RDC **Interview times** 2nd trimester and at 34 wks GA, 3, 6, and 9 wks PP

Evidence Table 1. Key Question 1: Study design for studies of the prevalence and incidence of perinatal depression (continued)

Type of Estimate	Time Period of Estimate	Diagnoses Included	Confounders	No. Depressed / Total No. (%)
New *cases*	0 to 5 wks PP	Major and minor	EPDS ≥ 9	16/225 (**7.1**%)
	0 to 6 mos PP	Major and minor	EPDS ≥ 9	25/225 (**11.1%**)
Point *cases*	6 mos PP	Major and minor	EPDS ≥ 9	21/232 (9.1%)
	6 mos PP	Major and minor	(2)	23/232 (9.9%)
	6 mos PP	Major	EPDS ≥ 9	8/232 (3.5%)
Period *cases*	0 to 6 mos PP	Major and minor	EPDS ≥ 9	32/232 (13.8%)
	0 to 6 mos PP	Major	EPDS ≥ 9	15/232 (6.5%)
New *controls*	0 to 5 wks PP	Major and minor	EPDS ≥ 9	5/218 (**2.3**%)
	0 to 6 mos PP	Major and minor	EPDS ≥ 9	17/218 (**7.8%**)
Point *controls*	6 mos PP	Major and minor	EPDS ≥ 9	19/232 (8.2%)
	6 mos PP	Major and minor	(2)	23/232 (9.9%)
	6 mos PP	Major	EPDS ≥ 9	8/232 ((3.5%)
Period *controls*	0 to 6 mos PP	Major and minor	EPDS ≥ 9	31/232 (13.4%)
	0 to 6 mos PP	Major	EPDS ≥ 9	13/232 (5.6%)
Point *cases*	2nd trimester	Major and minor	NR	14/182 (7.7%)
	2nd trimester	Major	NR	9/182 (**4.9%**)
	9 wks PP	Major and minor	NR	19/182 (10.4%)
	9 wks PP	Major	NR	8/182 (**4.4%**)
Point *controls*	2nd trimester	Major and minor	NR	10/179 (5.6%)
	2nd trimester	Major	NR	7/179 (**3.9%**)
	9 wks PP	Major and minor	NR	14/179 (7.8%)
	9 wks PP	Major	NR	6/179 (**3.4%**)

Evidence Table 1. Key Question 1: Study design for studies of the prevalence and incidence of perinatal depression (continued)

Author, Year Quality Rating	Study Characteristics	Population Description	Depression Measure
Retrospective Studies			
Author Bryan et al., 1999 **Quality rating** 16	**Study design** Retrospective chart review **Sample size** 403 women **Place** Minnesota	**Population** Random sample of Olmsted County residents who gave birth in the county between 1/1/1993 and 12/31/1993 **Age** < 19 yrs: 8 (2%) 19 to 32 yrs: 265 (65.8%) 33+ yrs: 130 (32.2%) **Race/ethnicity** NR **Inclusion/exclusion criteria** *Excluded:* Women whose pregnancies terminated prior to 24 wks GA and non-Olmsted County residents at the time of delivery; women with preexisting depression without substantial remission prior to delivery	**Diagnoses** Major and minor **Clinical assessment** Information on symptoms and diagnoses of depression were abstracted from medical records for 1 yr PP. Depression defined by • 2 notations at least 2 wks apart of symptoms of depression • documented diagnosis of depression by a medical provider • a new prescription for antidepressant with no evidence that it was for an indication other than depression • documentation of symptoms sufficient to meet the DSM-IV criteria for major depression
Author Georgiopoulos et al., 2001 **Quality rating** 8	**Study design** Population-based prospective study with retrospective record review **Sample size** 342 women **Place** Minnesota	**Population** Residents of Olmstead County, MN, visiting Olmstead Medical Center or Mayo Clinic in 1997 to 1998 who scored ≥ 10 in the routine EPDS screening project and a sample of women scoring < 10 with an indication of suicidal ideation **Age** Mean: 29 yrs Range: 16 to 46 yrs **Race/ethnicity** NR **Inclusion/exclusion criteria** None	**Diagnoses** Major **Clinical assessment** Documented diagnosis of PP depression in medical records during 1 yr PP

Evidence Table 1. Key Question 1: Study design for studies of the prevalence and incidence of perinatal depression (continued)

Type of Estimate	Time Period of Estimate	Diagnoses Included	Confounders	No. Depressed / Total No. (%)
Retrospective Studies				
New episode	0 to 1 yr PP	Major and minor	NR	10/398 (2.5%)
Period	0 to 1 yr PP	Major and minor	NR	15/403 (3.7%)
	0 to 1 yr PP	Major	NR	5/403 (**1.2%**)
Period	0 to1 yr PP	Major	NR	10.7% (weighted for the entire population)

Notes: Bolded numbers were computed from reported numbers.
(1) Different subsamples of women (random, GHQ ≥12, GHQ < 12) interviewed at different time periods; rates were based on cases with a Catego class of retarded depression and were adjusted for the full sample.
(2) Corrected for loss to followup by applying the positive predictive value to the EPDS high scorers who could not be interviewed and the negative predictive value to the EPDS low scorers not interviewed.

Evidence Table 2. Key Question 2: Studies of screening tools for detecting depression during pregnancy and first year postpartum: Abstract form results (English)

Author, Year Quality Rating	Study Characteristics	Population Description	Criterion Standard
Author Ballard et al., 1994 **Quality rating** 18	**Sample size** 200 **Place** UK **Recruitment setting** Maternity hospital	**Age** Mean: 28.8 yrs **Race/ethnicity** NR **Inclusion/exclusion criteria** *Included:* Married or cohabitating mothers *Excluded:* Lack of English language skills	**Instrument** PAS (adaptation of PSE) **Diagnostic criteria** RDC **Timing** Within 2 wks of date that completed EPDS questionnaire was received – approximately 6 mos PP
Author Beck and Gable, 2001 **Quality rating:** 15	**Sample size** 150 **Place** US **Recruitment setting** • Childbirth classes • Newspaper advertisements	**Age** Mean: 31 yrs Range: 18 to 46 yrs **Race/ethnicity** Caucasian: 87% Black: 8% Hispanic: 4% Asian: 1% **Inclusion/exclusion criteria** *Included:* Women 2 to 12 wks PP, 18+ yrs of age, able to read English, delivered live, healthy infant	**Instrument** SCID conducted by nurse psychotherapist **Diagnostic criteria** DSM-IV **Timing** Immediately following completion of screening instruments between 2 and 12 wks PP

Evidence Table 2. Key Question 2: Studies of screening tools for detecting depression during pregnancy and first year postpartum: Abstract form results (English) (continued)

Diagnosis	Screening Method	Prevalence No. Screened & Diagnosed	Sensit. %	Specific. %	Likelihood Ratio +	Likelihood Ratio –
Major	**Instrument** EPDS **Cutoff scores** ≥ 13	15.5% **Screened** + 31, – **22**; **Diagnosed** + 23, – **30**	95.7	71	3.3	0.06
Major	**Instrument** PDSS **Cutoff scores** ≥ 81	12% **Screened** + 20, – 130; **Diagnosed** + 18, – 132	94	98	47	0.061
Major	**Instrument** EPDS **Cutoff scores** ≥ 13	12% **Screened** + 15, – 135; **Diagnosed** + 18, – 132	78	99	78	0.222
Major	**Instrument** BDI-II **Cutoff scores** ≥ 21	12% **Screened** + 10, – 140; **Diagnosed** + 18, – 132	56	100	0	0.44
Major or minor	**Instrument** BDI-II **Cutoff scores** ≥ 10	19% **Screened** + 42, – 108; **Diagnosed** + 46, – 104	59	88	4.9	0.47
Major or minor	**Instrument** BDI-II **Cutoff scores** ≥ 15	19% **Screened** + 29, – 121; **Diagnosed** + 46, – 104	57	97	19	0.44
Major or minor	**Instrument** PDSS **Cutoff scores** ≥ 61	19% **Screened** + NR, – NR; **Diagnosed** + 46, – 104	91	72	3.25	0.13

Evidence Table 2. Key Question 2: Studies of screening tools for detecting depression during pregnancy and first year postpartum: Abstract form results (English) (continued)

Author, Year Quality Rating	Study Characteristics	Population Description	Criterion Standard
Author Boyce et al., 1993 **Quality rating** 16	**Sample size** 103 **Place** Australia **Recruitment setting** Mother's Advisory Clinics	**Age** Mean: 28.4 yrs **Race/ethnicity** NR **Inclusion/exclusion criteria** *Included:* ≤ 6 mos PP *Excluded:* Puerperal psychosis	**Instrument** DIS administered by a psychologist **Diagnostic criteria** DSM-III-R **Timing** Mean: 12 wks PP (SD 6.8 wks) Median: 10 wks PP Range: 2 to 29 wks Time lag between screening and gold standard NR
Author Campbell and Cohn, 1991 **Quality rating** 19	**Sample size** 1007 **Place** USA **Recruitment setting** Women who delivered in urban tertiary women's hospital	**Age** NR **Race/ethnicity** Caucasian: 100% **Inclusion/exclusion criteria** *Included:* Primiparous; full-term, singleton delivery; Caucasian; married; at least 18 yrs of age; at least HS education *Excluded:* Major delivery complications, adoptive or stepchildren in the home	**Instrument** Modified SADS **Diagnostic criteria** RDC **Timing** 6 to 8 wks PP (same time as screening)

Appendix C. Evidence Tables (continued)

Evidence Table 2. Key Question 2: Studies of screening tools for detecting depression during pregnancy and first year postpartum: Abstract form results (English) (continued)

Diagnosis	Screening Method	Prevalence No. Screened & Diagnosed				Sensit. %	Specific. %	Likelihood Ratio +	Likelihood Ratio –
Major	**Instrument** EPDS **Cutoff scores** ≥ 13	8.7%				100	96	2.50	NA
		Screened		**Diagnosed**					
		+	–	+	–				
		13	90	9	94				
Major	**Instrument** EPDS **Cutoff scores** ≥ 13	8.7%				100	89	9.09	NA
		Screened		**Diagnosed**					
		+	–	+	–				
		19	84	9	94				
Major	**Instrument** CES-D **Cutoff scores** ≥ 21	9%				43.5	96.7	14.3	0.59
		Screened		**Diagnosed**					
		+	–	+	–				
		70	937	92	915				
Major	**Instrument** CES-D **Cutoff scores** ≥ 16	9%				59.8	91.6	7.11	0.44
		Screened		**Diagnosed**					
		+	–	+	–				
		132	875	92	915				

Evidence Table 2. Key Question 2: Studies of screening tools for detecting depression during pregnancy and first year postpartum: Abstract form results (English) (continued)

Author, Year Quality Rating	Study Characteristics	Population Description	Criterion Standard
Author Cox et al., 1996	**Sample size** 128	**Age** Mean: 27.2 yrs	**Instrument** SPI
Quality rating 13	**Place** UK	**Race/ethnicity** NR	**Diagnostic criteria** RDC
	Recruitment setting PP women recruited from GP age/sex registries	**Inclusion/exclusion criteria** None	**Timing** Occurred subsequent to initial EPDS but exact timing NR

Evidence Table 2. Key Question 2: Studies of screening tools for detecting depression during pregnancy and first year postpartum: Abstract form results (English) (continued)

Diagnosis	Screening Method	Prevalence No. Screened & Diagnosed				Sensit. %	Specific. %	Likelihood Ratio +	Likelihood Ratio –
Major	**Instrument** EPDS **Cutoff scores** ≥ 13	6.25%				75	84	4.7	0.30
		Screened		**Diagnosed**					
		+	–	+	–				
		25	103	8	120				
Major	**Instrument** EPDS **Cutoff scores** ≥ 12	6.25%				88	76	3.7	0.16
		Screened		**Diagnosed**					
		+	–	+	–				
		36	92	8	120				
Major	**Instrument** EPDS **Cutoff scores** ≥ 10	6.25%				88	77	3.0	0.17
		Screened		**Diagnosed**					
		+	–	+	–				
		42	86	8	120				
Major or minor	**Instrument** EPDS **Cutoff scores** ≥ 13	16.41%				62	89	5.6	0.43
		Screened		**Diagnosed**					
		+	–	+	–				
		25	103	21	107				
Major or minor	**Instrument** EPDS **Cutoff scores** ≥ 12	16.41%				76	81	4.0	0.30
		Screened		**Diagnosed**					
		+	–	+	–				
		36	92	21	107				
Major or minor	**Instrument** EPDS **Cutoff scores** ≥ 10	16.41%				81	77	3.5	0.25
		Screened		**Diagnosed**					
		+	–	+	–				
		42	86	21	107				

Evidence Table 2. Key Question 2: Studies of screening tools for detecting depression during pregnancy and first year postpartum: Abstract form results (English) (continued)

Author, Year Quality Rating	Study Characteristics	Population Description	Criterion Standard
Author Harris et al., 1989 **Quality rating** 13	**Sample size** 147 **Place** Wales **Recruitment setting** Prior study of women delivering at Caerphilly Miners' Hospital	**Age** Mean: 24.6 yrs **Race/ethnicity** NR **Inclusion/exclusion criteria** *Included:* Women with (n = 65) and without (n = 82) autoimmune thyroid disorder *Excluded:* All other thyroid disorders	**Instrument** Raskin 3 Area Scale for Depression and MADRS conducted by experienced psychiatrist **Diagnostic criteria** DSM-III **Timing** 6 wks PP the clinical interview was conducted at the PP followup visit; screening instruments were completed subsequently at home and mailed back within 2 wks
Author Leverton and Elliott, 2000 **Quality rating** 13	**Sample size** 199 **Place** England **Recruitment setting** Prenatal clinic	**Age** NR **Race/ethnicity** NR **Inclusion/exclusion criteria** *Included:* Women expecting their 1st or 2nd child and designated as vulnerable on the Leverton Questionnaire for vulnerability	**Instrument** PSE conducted by a psychiatrist **Diagnostic criteria** One classification determined by Bedford (B) College criteria (gives diagnosis consistent with current MDE criteria and a "borderline depression" diagnosis consistent with minor depression) and another by Catego (C) diagnoses **Timing** 3 mos PP

Appendix C. Evidence Tables (continued)

Evidence Table 2. Key Question 2: Studies of screening tools for detecting depression during pregnancy and first year postpartum: Abstract form results (English) (continued)

Diagnosis	Screening Method	Prevalence No. Screened & Diagnosed				Sensit. %	Specific. %	Likelihood Ratio +	Likelihood Ratio –
Major	**Instrument** EPDS **Cutoff scores** ≥ 13	15% **Screened** +	**Screened** –	**Diagnosed** +	**Diagnosed** –	95	93	13.57	00.54
		21	105	22	104				
Major	**Instrument** EPDS **Cutoff scores** ≥ 10	15% **Screened** +	**Screened** –	**Diagnosed** +	**Diagnosed** –	100	82	3.6	0.0
		41	85	22	104				
Major	**Instrument** BDI **Cutoff scores** ≥ 21	15% **Screened** +	**Screened** –	**Diagnosed** +	**Diagnosed** –	32	99	32	0.69
		NR	NR	19	110				
Major	**Instrument** BDI **Cutoff scores** ≥ 13	15% **Screened** +	**Screened** –	**Diagnosed** +	**Diagnosed** –	63	92	7.9	0.40
		NR	NR	19	110				
Major	**Instrument** BDI **Cutoff scores** ≥ 11	15% **Screened** +	**Screened** –	**Diagnosed** +	**Diagnosed** –	68	88	5.67	0.364
		26	103	19	110				
Major or minor	**Instrument** EPDS **Cutoff scores** ≥ 13	C: 5% B: 8% **Screened** +	**Screened** –	**Diagnosed** +	**Diagnosed** –	C: 70 B: 44	C: 93 B: 92	C: 9.45 B: 5.50	C: 0.32 B: 0.61
		C: 21 B: 21	C: 178 B: 178	C: 10 B: 16	C: 189 B: 183				
Major or minor	**Instrument** EPDS **Cutoff scores** ≥ 10	C: 5% B: 8% **Screened** +	**Screened** –	**Diagnosed** +	**Diagnosed** –	C: 90 B: 69	C: 84 B: 85	C: 5.62 B: 4.60	C: 0.12 B: 0.36
		C: 39 B: 39	C: 160 B: 160	C: 10 B: 16	C: 189 B: 183				

Evidence Table 2. Key Question 2: Studies of screening tools for detecting depression during pregnancy and first year postpartum: Abstract form results (English) (continued)

Author, Year Quality Rating	Study Characteristics	Population Description	Criterion Standard
Author Murray and Carothers, 1990 **Quality rating** 15	**Sample size** 646 **Place** England **Recruitment setting** PP wards	**Age** NR **Race/ethnicity** NR **Inclusion/exclusion criteria** *Included:* 20 to 40 yrs of age, married or cohabitating, primiparous, healthy infant	**Instrument** SPI **Diagnostic criteria** RDC **Timing** Interview conducted presumably a couple of wks later than EPDS, which was conducted at 6 wks PP

Evidence Table 2. Key Question 2: Studies of screening tools for detecting depression during pregnancy and first year postpartum: Abstract form results (English) (continued)

Diagnosis	Screening Method	Prevalence No. Screened & Diagnosed				Sensit. %	Specific. %	Likelihood Ratio +	Likelihood Ratio –
Major and minor	**Instrument** EPDS **Cutoff scores** ≥ 14	NR				73.1	97.5	29.2	0.28
		Screened		**Diagnosed**					
		+	–	+	–				
		NR	NR	NR	NR				
Major and minor	**Instrument** EPDS **Cutoff scores** ≥ 13	NR				81.1	95.7	18.9	0.20
		Screened		**Diagnosed**					
		+	–	+	–				
		NR	NR	NR	NR				
Major and minor	**Instrument** EPDS **Cutoff scores** ≥ 12	NR				88.0	92.5	13.5	0.24
		Screened		**Diagnosed**					
		+	–	+	–				
		NR	NR	NR	NR				

Evidence Table 2. Key Question 2: Studies of screening tools for detecting depression during pregnancy and first year postpartum: Abstract form results (English) (continued)

Author, Year Quality Rating	Study Characteristics	Population Description	Criterion Standard
Author Murray and Cox, 1990	**Sample size** 100	**Age** Mean: 24.6 yrs	**Instrument** SPI
Quality rating 16	**Place** UK	**Race/ethnicity** NR	**Diagnostic criteria** RDC
	Recruitment setting Prenatal clinic of a large maternity hospital	**Inclusion/exclusion criteria** *Included:* All women between 28 and 34 wks GA	**Timing** 28 to 34 wks GA
Author Whiffen 1988	**Sample size** 120	**Age** Mean: 28 yrs	**Instrument** SADS
Quality rating 10	**Place** Ontario, Canada	**Race/ethnicity** NR	**Diagnostic criteria** RDC
	Recruitment setting Public health prenatal classes	**Inclusion/exclusion criteria** *Included:* 18 yrs of age or older, married or common-law, low risk for pregnancy complications, primiparous	**Timing** 7.6 days (± 5.9) after screening, which was conducted 6 to 8 wks PP

Evidence Table 2. Key Question 2: Studies of screening tools for detecting depression during pregnancy and first year postpartum: Abstract form results (English) (continued)

Diagnosis	Screening Method	Prevalence No. Screened & Diagnosed				Sensit. %	Specific. %	Likelihood Ratio +	Likelihood Ratio –
Major	**Instrument** EPDS **Cutoff scores** ≥ 15	6%				100	96	25.0	0.0
		Screened		**Diagnosed**					
		+	–	+	–				
		10	90	6	94				
Major	**Instrument** EPDS **Cutoff scores** ≥ 14	6%				100	94	16.6	0.0
		Screened		**Diagnosed**					
		+	–	+	–				
		12	88	6	94				
Major	**Instrument** EPDS **Cutoff scores** ≥ 13	6%				100	87	7.7	0.0
		Screened		**Diagnosed**					
		+	–	+	–				
		18	82	6	94				
Major	**Instrument** EPDS **Cutoff scores** ≥ 12	6%				100	79	4.8	0.0
		Screened		**Diagnosed**					
		+	–	+	–				
		26	74	6	94				
Major or minor	**Instrument** EPDS **Cutoff scores** ≥ 14	14%				57	95	11.4	0.45
		Screened		**Diagnosed**					
		+	–	+	–				
		21	79	14	86				
Major or minor	**Instrument** EPDS **Cutoff scores** ≥ 13	14%				64	90	6.4	0.4
		Screened		**Diagnosed**					
		+	–	+	–				
		28	72	14	86				
Major or minor	**Instrument** EPDS **Cutoff scores** ≥ 12	14%				64	80	3.2	0.45
		Screened		**Diagnosed**					
		+	–	+	–				
		40	60	14	86				
Major or minor	**Instrument** EPDS **Cutoff scores** ≥ 11	14%				71	72	2.5	0.40
		Screened		**Diagnosed**					
		+	–	+	–				
		48	52	14	86				
Major or minor	**Instrument** BDI **Cutoff scores** ≥ 9	17.5%				47.6	85.9	3.38	0.61
		Screened		**Diagnosed**					
		+	–	+	–				
		24	96	21	99				

Evidence Table 3. Key Question 2: Studies of screening tools for detecting depression during pregnancy and first year postpartum: Abstract form results (non-English)

Author, Year Quality Rating	Study Characteristics	Population Description	Criterion Standard
Author Berle et al., 2003	**Sample size** 411	**Age** Mean: 30.2 yrs	**Instrument** MINI-V4.4 and MADRS
Quality rating 13	**Place** Norway	**Race/ethnicity** NR	**Diagnostic criteria** DSM-IV
	Recruitment setting PP clinics	**Inclusion/exclusion criteria** None	**Timing** 6 to 12 wks PP

Evidence Table 3. Key Question 2: Studies of screening tools for detecting depression during pregnancy and first year postpartum: Abstract form results (non-English) (continued)

Diagnosis	Screening Method	Prevalence No. Screened & Diagnosed	Sensit. %	Specific. %	Likelihood Ratio +	Likelihood Ratio –
Major	**Instrument** EPDS (Norwegian version) **Cutoff scores** ≥ 13	6.6% **Screened** + 57, – 354; **Diagnosed** + 27, – 384	56	89	5.09	0.49
Major	**Instrument** EPDS (Norwegian version) **Cutoff scores** ≥ 12	6.6% **Screened** + 90, – 321; **Diagnosed** + 27, – 384	78	82	4.33	0.27
Major	**Instrument** EPDS (Norwegian version) **Cutoff scores** ≥ 11	6.6% **Screened** + 122, – 289; **Diagnosed** + 27, – 384	96	75	3.84	5.33
Major or minor	**Instrument** EPDS (Norwegian version) **Cutoff scores** ≥ 11	10% **Screened** + 53, – 44; **Diagnosed** + 41, – 59	0.83	0.83	3.1	0.32
Major or minor	**Instrument** EPDS (Norwegian version) **Cutoff scores** ≥ 13	10% **Screened** + 59, – 39; **Diagnosed** + 41, – 59	0.49	0.95	18	0.14
Major or minor	**Instrument** EPDS (Norwegian version) **Cutoff scores** ≥ 9	10% **Screened** + 72, – 26; **Diagnosed** + 41, – 59	0.95	0.51	19.3	0.10

Evidence Table 3. Key Question 2: Studies of screening tools for detecting depression during pregnancy and first year postpartum: Abstract form results (non-English) (continued)

Author, Year Quality Rating	Study Characteristics	Population Description	Criterion Standard
Author Garcia-Esteve et al., 2003 **Quality rating** 18	**Sample size** 1123 **Place** Spain **Recruitment setting** Public maternity hospital	**Age** NR **Race/ethnicity** NR **Inclusion/exclusion criteria** *Included:* Women attending routine PP checkup *Excluded:* Not fluent in Spanish, mourning, organic depression	**Instrument** SCIP-NP **Diagnostic criteria** DSM-III-R **Timing** 6 wks PP

Appendix C. Evidence Tables (continued)

Evidence Table 3. Key Question 2: Studies of screening tools for detecting depression during pregnancy and first year postpartum: Abstract form results (non-English) (continued)

Diagnosis	Screening Method	Prevalence No. Screened & Diagnosed	Sensit. %	Specific. %	Likelihood Ratio +	Likelihood Ratio –
Major	**Instrument** EPDS (Spanish version): 10/11 **Cutoff scores** ≥ 13	3.5% Screened: + 44, – 280 Diagnosed: + 36, – 288	86.1	95.4	14.6	0.15
Major	**Instrument** EPDS (Spanish version): 10/11 **Cutoff scores** ≥ 12	3.5% Screened: + 50, – 274 Diagnosed: + 36, – 288	91.7	94.1	15.5	0.09
Major	**Instrument** EPDS (Spanish version): 10/11 **Cutoff scores** ≥ 11	3.5% Screened: + 60, – 264 Diagnosed: + 36, – 288	100	91.8	12.2	NA
Major	**Instrument** EPDS (Spanish version): 10/11 **Cutoff scores** ≥ 10	3.5% Screened: + 68, – 256 Diagnosed: + 36, – 288	100	88.8	4.7	NA
Major or minor	**Instrument** EPDS (Spanish version): 10/11 **Cutoff scores** ≥ 13	5.7% Screened: + 66, – 258 Diagnosed: + 100, – 224	62.0	98.1	32.7	0.39
Major or minor	**Instrument** EPDS (Spanish version): 10/11 **Cutoff scores** ≥ 12	5.7% Screened: + 76, – 248 Diagnosed: + 100, – 224	70.0	97.3	25.9	0.31
Major or minor	**Instrument** EPDS (Spanish version): ≥ 10 **Cutoff scores** ≥ 11	5.7% Screened: + 89, – 245 Diagnosed: + 100, – 224	79.0	95.5	17.6	0.22
Major or minor	**Instrument** EPDS (Spanish version): ≥ 9 **Cutoff scores** ≥ 10	5.7% Screened: + 104, – 220 Diagnosed: + 100, – 224	89.0	93.3	14.6	0.02

Evidence Table 3. Key Question 2: Studies of screening tools for detecting depression during pregnancy and first year postpartum: Abstract form results (non-English) (continued)

Author, Year Quality Rating	Study Characteristics	Population Description	Criterion Standard
Author Guedeney and Fermanian, 1998 **Quality rating** 18	**Sample size** 87 **Place** France **Recruitment setting** Infant health clinics	**Age** Mean: 30.4 yrs **Race/ethnicity** NR **Inclusion/exclusion criteria** *Included:* Living in Paris, fluent in French, 1st 4 mos PP *Excluded:* History of psychotic illness or PP psychosis	**Instrument** PSE **Diagnostic criteria** RDC **Timing** Up to 4 mos PP
Author Kitamura et al., 1994 **Quality rating** 11	**Sample size** 120 **Place** Japan **Recruitment setting** Dept. of obstetrics of a general hospital	**Age** Mean: 28 yrs **Race/ethnicity** Japanese **Inclusion/exclusion criteria** *Included:* Pregnant women attending prenatal clinic *Excluded:* At or over 12 wks GA	**Instrument** SADS conducted by two psychiatrists blinded to results of the other screening methods **Diagnostic criteria** RDC **Timing** Directly following each screen conducted: • early pregnancy • late pregnancy • 5 days PP • 1 mo PP

Appendix C. Evidence Tables (continued)

Evidence Table 3. Key Question 2: Studies of screening tools for detecting depression during pregnancy and first year postpartum: Abstract form results (non-English) (continued)

Diagnosis	Screening Method	Prevalence No. Screened & Diagnosed	Sensit. %	Specific. %	Likelihood Ratio +	Likelihood Ratio –
Major or minor	**Instrument** EPDS (French version) **Cutoff scores** ≥ 13	51.7% **Screened** + 28, – 59; **Diagnosed** + 45, – 42	60	97	20	0.41
Major or minor	**Instrument** EPDS (French version) **Cutoff scores** ≥ 12	51.7% **Screened** + 33, – 54; **Diagnosed** + 45, – 42	73	95	14.6	0.28
Major or minor	**Instrument** EPDS (French version) **Cutoff scores** ≥ 11	51.7% **Screened** + 39, – 48; **Diagnosed** + 45, – 42	80	92	10	0.22
Major or minor	**Instrument** SDS **Cutoff scores** ≥ 22	1st trim: 12.5% **Screened** + 108, – 98; **Diagnosed** + 12, – 96	90.9	70.0	3.03	0.13
Major or minor	**Instrument** SDS **Cutoff scores** ≥ 22	3rd trim: 10% **Screened** + 98, – 88; **Diagnosed** + 10, – 88	70.0	76.1	2.93	0.39
Major or minor	**Instrument** SDS **Cutoff scores** ≥ 22	PP day 5: 7.1% **Screened** + 11, – 80; **Diagnosed** + 6, – 85	27.3	85.0	1.82	0.85
Major or minor	**Instrument** SDS **Cutoff scores** ≥ 22	PP mo 1: 7.4% **Screened** + 9, – 92; **Diagnosed** + 7, – 94	44.9	88.0	3.74	0.63

Evidence Table 3. Key Question 2: Studies of screening tools for detecting depression during pregnancy and first year postpartum: Abstract form results (non-English) (continued)

Author, Year Quality Rating	Study Characteristics	Population Description	Criterion Standard
Author Lawrie et al., 1998 **Quality rating** 20	**Sample size** 103 **Place** South Africa **Recruitment setting** PP clinic	**Age** 28.1 yrs **Race/ethnicity** Black **Inclusion/exclusion criteria** *Included:* Women who had obstetrical complications, cesareans, or requested sterilization	**Instrument** Structured interview and MADRS **Diagnostic criteria** DSM-IV **Timing** Starting at 6 wks PP over 3-mo period
Author Lee et al., 2001 **Quality rating** 14	**Sample size** 781 **Place** Hong Kong **Recruitment setting** Prenatal clinic of a univ. hospital	**Age** Mean: 29 yrs **Race/ethnicity** Chinese **Inclusion/exclusion criteria** *Excluded:* Not of Chinese ethnicity, no long-term residential rights	**Instrument** SCID (all women who had ≥ 5 on GHQ and 10% of women with < 5) **Diagnostic criteria** DSM-III-R **Timing** 3 mos PP

Evidence Table 3. Key Question 2: Studies of screening tools for detecting depression during pregnancy and first year postpartum: Abstract form results (non-English) (continued)

Diagnosis	Screening Method	Prevalence No. Screened & Diagnosed	Sensit. %	Specific. %	Likelihood Ratio +	Likelihood Ratio –
Major or minor	**Instrument** EPDS **Cutoff scores** ≥ 13	NR% **Screened** + 7, – 96; **Diagnosed** + NR, – NR	87.5	72.3	3.16	0.17
Major or minor	**Instrument** EPDS **Cutoff scores** ≥ 10	NR% **Screened** + NR, – NR; **Diagnosed** + NR, – NR	100.0	58.5	2.41	NA
Major or minor	**Instrument** EPDS **Cutoff scores** ≥ 9	NR% **Screened** + NR, – NR; **Diagnosed** + NR, – NR	100.0	51.1	2.04	NA
Major	**Instrument** GHQ **Cutoff scores** ≥ 5	6.1% **Screened** + NR, – NR; **Diagnosed** + 26, – 127	NR	NR	NR	NR
Major or minor	**Instrument** GHQ **Cutoff scores** ≥ 5	11.2% **Screened** + 1, – 62; **Diagnosed** + 45, – 82	95	68	2.97	0.07

Evidence Table 3. Key Question 2: Studies of screening tools for detecting depression during pregnancy and first year postpartum: Abstract form results (non-English) (continued)

Author, Year Quality Rating	Study Characteristics	Population Description	Criterion Standard
Author Lee et al., 2000 **Quality rating** 16	**Sample size** 145 **Place** Hong Kong **Recruitment setting** PP ward of a university hospital	**Age** Mean: 29 yrs Range: 16 to 42 yrs **Race/ethnicity** Chinese **Inclusion/exclusion criteria** *Included:* All Chinese women admitted to PP ward of Dept. of ob/gyn from 11/1996 to 1/1997 *Excluded:* Non-Chinese women and those who did not have permanent residency in Hong Kong	**Instrument** SCID-NP (nonpatient version; Chinese) **Diagnostic criteria** DSM-III-R **Timing** 6 wks PP
Author Lee et al., 1998 **Quality Rating** 20	**Sample size** 330 eligible, 220 (67%) agreed to participate, 145 (66% participants) completed 6-wk followup assessment **Place** Hong Kong **Recruitment setting** PP wards of the Prince of Wales Hospital in Hong Kong from 11/1996 to 1/1997	**Age** Mean: 29 yrs Range: 16 to 42 yrs **Race/ethnicity** Asian **Inclusion/exclusion criteria** *Included:* Chinese women admitted to PP wards *Excluded:* Non-Chinese, nonpermanent Hong Kong residents (i.e., illegal immigrants)	**Instrument** Chinese nonpatient version of SCID modified to allow diagnosis of DSM IV-minor depressive disorder and to make “6-wk” rather than “1-wk” diagnoses **Diagnostic criteria** DSM-IV **Timing** Interviewed at 6 wks PP Screened 2 days PP and 6 wks PP

Evidence Table 3. Key Question 2: Studies of screening tools for detecting depression during pregnancy and first year postpartum: Abstract form results (non-English) (continued)

Diagnosis	Screening Method	Prevalence No. Screened & Diagnosed	Sensit. %	Specific. %	Likelihood Ratio +	Likelihood Ratio –
Major or minor	**Instrument** EPDS **Cutoff scores** ≥ 10	12% **Screened** + 32, – 113; **Diagnosed** + 17, – 128	82	86	5.86	0.21
Major or minor	**Instrument** GHQ **Cutoff scores** ≥ 5	12% **Screened** + 29, – 16; **Diagnosed** + 17, – 128	88	89	8.00	0.13
Major or minor	**Instrument** BDI-II **Cutoff scores** ≥ 10	12% **Screened** + 34, – 111; **Diagnosed** + 17, – 128	0.94	0.86	6.71	0.07
Major or minor	**Instrument** Chinese version of the EPDS **Cutoff scores** ≥ 10	12% **Screened** + 32, – 113; **Diagnosed** + 17, – 128	82	86	5.86	1.16
Major or minor	**Instrument** GHQ **Cutoff scores** ≥ 5	12% **Screened** + 31, – 116; **Diagnosed** + 19, – 128	88	89	8.0	0.13

Evidence Table 3. Key Question 2: Studies of screening tools for detecting depression during pregnancy and first year postpartum: Abstract form results (non-English) (continued)

Author, Year Quality Rating	Study Characteristics	Population Description	Criterion Standard
Author Muzik et al., 2000 **Quality rating** 10	**Sample size** 50 **Place** Austria **Recruitment setting** Prior study participants	**Age** Mean: 28 yrs Range: 21 to 40 yrs **Race/ethnicity** NR **Inclusion/exclusion criteria** *Included:* Women in a larger study with EPDS ≥ 7 at either 3 or 6 mos PP	**Instrument** SCID-German version by psychiatrist **Diagnostic criteria** DSM-III-R **Timing** 3 or 6 mos PP (did not clarify, but implies near time of screen)
Author Wickberg and Hwang, 1996 **Quality rating** 18	**Sample size** 128 women used in validation study drawn from a sample of 1655 women screened **Place** Sweden **Recruitment setting** 17 child health clinics near Göteberg, Sweden, and Möindal	**Age** Mean: 28.1 yrs Range: 18 to 42 yrs **Race/ethnicity** NR **Inclusion/exclusion criteria** *Included:* Swedish speaking, completed both screenings, score > 11.5 on both or 2nd screening or else randomly chosen to be one of 37 who scored less than 12 *Excluded:* Already seeing a GP/psych for depression, refusal to participate	**Instrument** MADRS **Diagnostic criteria** DSM-III-R **Timing** 1 to 2 wks after completing the EPDS screen, which was conducted 2 and 3 mos PP
Author Yamashita et al., 2000 **Quality rating** 8	**Sample size** 88 **Place** Japan **Recruitment setting** Maternity ward of a university hospital	**Age** Mean: 31 yrs Range: 19 to 41 yrs **Race/ethnicity** Asian **Inclusion/exclusion criteria** *Included:* Pregnant, Japanese	**Instrument** SADS conducted by psychiatrists **Diagnostic criteria** RDC **Timing** Screening was conducted at 1 mo PP (T1) visit and by mail 3 mos PP (T2)

Evidence Table 3. Key Question 2: Studies of screening tools for detecting depression during pregnancy and first year postpartum: Abstract form results (non-English) (continued)

Diagnosis	Screening Method	Prevalence No. Screened & Diagnosed				Sensit. %	Specific. %	Likelihood Ratio +	Likelihood Ratio –
		12%							
	Instrument	**Screened**		**Diagnosed**					
Major	EPDS	+	–	+	–	66	92	8.25	0.37
	Cutoff scores ≥ 11	10	40	10	40				
		12%							
	Instrument	**Screened**		**Diagnosed**					
Major	SDS	+	–	+	–	89	77	3.87	0.14
	Cutoff scores ≥ 50	18	32	10	40				
		12%							
	Instrument	**Screened**		**Diagnosed**					
Major	SCLR-90	+	–	+	–	78	87	6.00	0.25
	Cutoff scores ≥ 50	13	37	10	40				
		44%							
	Instrument	**Screened**		**Diagnosed**					
Major	EPDS	+	–	+	–	85	63	2.27	0.25
	Cutoff scores ≥ 13	47	45	56	72				
		44%							
	Instrument	**Screened**		**Diagnosed**					
Major	EPDS	+	–	+	–	96	49	1.89	0.08
	Cutoff scores ≥ 12	128	37	56	72				
		NR%							
	Instrument	**Screened**		**Diagnosed**					
Major or minor	EPDS	+	–	+	–	T1: 82 T2: NR	T1: 95 T2: NR	T1: 16.4 T2: NR	T1: 0.19 T2: NR
	Cutoff scores ≥ 9	T1: 12 T2: NR	T1: 64 T2: 71	T1: 11 T2: 15	T1: 64 T2: 78				

Evidence Table 3. Key Question 2: Studies of screening tools for detecting depression during pregnancy and first year postpartum: Abstract form results (non-English) (continued)

Author, Year Quality Rating	Study Characteristics	Population Description	Criterion Standard
Author Yoshida et al., 1997	**Sample size** 98	**Age** Mean: 30 yrs	**Instrument** SADS
Quality rating 12	**Place** England	**Race/ethnicity** Asian	**Diagnostic criteria** RDC
	Recruitment setting • Prenatal classes for pregnant women • Advertisement in a maternity guidebook for pregnant Japanese women in UK	**Inclusion/exclusion criteria** *Included:* Pregnant, Japanese	**Timing** 3 mos PP (covered retrospective diagnosis at 1 mo as well as current diagnosis at 3 mos)

Evidence Table 3. Key Question 2: Studies of screening tools for detecting depression during pregnancy and first year postpartum: Abstract form results (non-English) (continued)

Diagnosis	Screening Method	Prevalence No. Screened & Diagnosed				Sensit. %	Specific. %	Likelihood Ratio +	Likelihood Ratio –
Major or minor	**Instrument** EPDS **Cutoff scores** ≥9	12%				25	96	6.25	0.78
		Screened		**Diagnosed**					
		+	–	+	–				
		6	91	8	89				
Major or minor	**Instrument** EPDS **Cutoff scores** ≥7	12%				50	80	2.50	0.63
		Screened		**Diagnosed**					
		+	–	+	–				
		22	75	8	89				

Evidence Table 4. Key Question 3: Studies of screening interventions for perinatal depression: Screening during pregnancy

Author, Year Quality Rating	Study Characteristics	Population Description	Treatment
Author Brugha et al., 2000 **Quality rating** 17	**Place** UK **Recruitment setting** Prenatal clinic of a UK general hospital **Sample size** *No. randomized:* Intervention: 103 Control: 106 *No. analyzed:* Intervention: 94 Control: 96 **Design** RCT	**Age** Median: 19 yrs Range: 16 to 38 yrs **Race/ethnicity** European: 73% Asian and other: 27% **Depression risk criteria** *Instrument:* Modified GHQ-D *Cutoff scores:* Presence of one of six depression items *Screen timing:* 12 to 20 wks GA **Other inclusion criteria** *Included:* 16+ yrs of age, 1st pregnancy, English speaking, living within travel distance of hospital	**Intervention group** Preparing for Parenthood Intervention—six structured, 2-hour-long prenatal classes and a PP reunion class at 8 wks PP designed to increase social support and present problem-solving skills **Control group** Routine prenatal care

Evidence Table 4. Key Question 3: Studies of screening interventions for perinatal depression: Screening during pregnancy (continued)

Assessment Timing	Depression Outcomes	Other Outcomes
3 mos PP	**GHQ-D ≥ 2** Intervention: 26% Control: 22% Adjusted OR: 1.19 (95% CI, 0.59 to 2.37) **EPDS ≥ 11** Intervention: 16% Control: 19% Adjusted OR: 0.83 (95% CI, 0.39 to 1.79) **Diagnosis of depression using SCAN and ICD-10 criteria** Intervention: 3% Control: 6% Adjusted OR: 0.47 (95% CI, 0.11 to 1.96)	**Social support** Intervention had no statistically significant impact on social support

Evidence Table 4. Key Question 3: Studies of screening interventions for perinatal depression: Screening during pregnancy (continued)

Author, Year Quality Rating	Study Characteristics	Population Description	Treatment
Author Elliott et al., 2000 **Quality rating** 11	**Place** UK **Recruitment setting** Prenatal clinic **Sample size** Intervention: 47 Control: 51 **Design** Controlled trial	**Age** NR **Race/ethnicity** NR **Depression risk criteria** *Instrument:* LQ and the depression, anxiety, and somatic subscales of the CCEI *Cutoff scores:* • Score of 2 on any one vulnerability question on the LQ or score of 1 on more than one vulnerability question -OR- • Score of ≥ 10 on CCEI anxiety subscale -OR- • 2nd-time mothers who felt more tense or depressed than usual after birth of 1st child *Screen timing:* 1st prenatal clinic appointment **Other inclusion criteria** *Included:* 1st- and 2nd-time mothers in ongoing relationships attending the same clinic in the same 6-mo window	**Intervention group** "Preparing for Parenthood" for 1st-time mothers and "Surviving Parenthood" for 2nd-time mothers, which included meetings run by a psychologist and health visitor held monthly for 5 mos during pregnancy, starting at 24 wks GA and continuing for 6 mos PP and included a mid-pregnancy health visitor visit **Control group** Routine prenatal and PP care

Evidence Table 4. Key Question 3: Studies of screening interventions for perinatal depression: Screening during pregnancy (continued)

Assessment Timing	Depression Outcomes	Other Outcomes
3 mos PP and 1 yr PP	**EPDS Score** *1st-time mothers at 3 mos:* Intervention: median 3.0, range 0 to 17, SD 4.48 Control: median 8.0, range 1 to 16, SD 4.53 Mann-Whitney: 141, one-tailed $P = 0.005$ *2nd-time mothers at 3 mos:* Intervention: median 6.5, range 1 to 247, SD 6.10 Control: median 9.0, range 1 to 23, SD 6.60 Mann-Whitney: 319, one-tailed P = NS *1st- and 2nd-time mothers at 12 mos:* No significant differences **CCEI & SRQ** Results were similar to those for EPDS **Diagnosis of depression using the PSE** *1st- and 2nd-time mothers at 3 mos and 1 yr:* No significant differences **Diagnosis of borderline or case depressions using Bedford College Criteria** *1st-time mothers at 3 mos:* Intervention: 19% borderline or case depressions Control: 39% borderline or case depressions $X^2(1)$: 2.64, one-tailed $P < 0.05$ *2nd-time mothers at 3 mos:* No significant differences	NR

Evidence Table 4. Key Question 3: Studies of screening interventions for perinatal depression: Screening during pregnancy (continued)

Author, Year Quality Rating	Study Characteristics	Population Description	Treatment
Author Stamp et al., 1995 **Quality rating** 13	**Place** Australia **Recruitment setting** Prenatal clinics of a tertiary referral Women's and Children's Hospital (no privately insured women attend these clinics) **Sample size** Intervention: 64 Control: 65 **Design** RCT	**Age** Intervention: mean 25.6 (SD 4.4) yrs Control: mean 27.5 (SD 5.2) yrs **Race/ethnicity** NR **Depression risk criteria** *Instrument:* Modified prenatal screening questionnaire *Cutoff scores:* ≥ 2 *Screen timing:* ≤ 24 wks GA **Other inclusion criteria** *Included:* English speaking, singleton pregnancy, live in the metropolitan area	**Intervention group** 2 prenatal groups at 32 and 26 wks GA plus 1 PP group held at 6 wks. Groups were a combination of education and social and psychological support **Control group** Routine prenatal care
Author Zlotnick et al., 2001 **Quality rating** 12	**Place** USA **Recruitment setting** Prenatal clinic at a general hospital in the Northeast **Sample size** Intervention: 17 Control: 18 **Design** RCT	**Age** Mean: 23.4 (SD 4.41) yrs Range: 18 to 38 yrs **Race/ethnicity** Caucasian: 46% Other: 54% **Depression risk criteria** *Instrument*: Survey assessing risk factors for perinatal depression *Cutoff score*: At least 1 risk factor *Screen timing*: 20 to 32 wks GA **Other inclusion criteria** *Included:* Receipt of public assistance, not depressed at screening according to SCID for DSM-IV	**Intervention group** 4 weekly hour-long interpersonal-therapy-oriented survival skills group sessions **Control group** Routine care

Evidence Table 4. Key Question 3: Studies of screening Interventions for perinatal depression: Screening during pregnancy (continued)

Assessment Timing	Depression Outcomes	Other Outcomes
6 wks, 12 wks, and 6 mos	**EPDS > 9 (6 wks)** Intervention: 34% Control: 34% OR: 1.00 (95% CI, 0.45 to 2.21) **EPDS > 9 (12 wks)** Intervention: 22% Control: 26% OR: NR **EPDS > 9 (6 mos)** Intervention: 23% Control: 16% OR: 1.55 (95% CI, 0.58 to 4.22) **EPDS > 12 (6 wks)** Intervention: 13% Control: 17% OR: 0.69 (95% CI, 0.23 to 2.03) **EPDS > 12 (12 wks)** Intervention: 11% Control: 15% OR: 0.69 (95% CI, 0.22 to 2.14) **EPDS > 12 (6 mos)** Intervention: 15% Control: 10% OR: 1.62 (95% CI, 0.47 to 5.91)	NR
3 mos	**Pre vs. post BDI scores** Intervention: 13.0 (SD 6.9) vs. 8.4 (SD 7.8) Control: 9.2 (SD 6.5) vs. 11.3 (SD 4.8) t-test(33): 3.50, $P = 0.001$ **Diagnosis of Depression using SCID for DSM-IV** Intervention: 0% Control: 33% $X^2(1)$: $P = 0.02$	NR

Evidence Table 5. Key Question 3: Studies of screening interventions for perinatal depression: Screening during postpartum

Author, Year Quality Rating	Study Characteristics	Population Characteristics Inclusion Criteria	Treatment
Author Armstrong et al., 1999 **Quality rating** 19	**Place** Australia **Recruitment setting** Hospital maternity ward **Sample size** Intervention: 90 Control: 91 **Design** RCT	**Age** < 18 yrs: 6.6% **Race/ethnicity** Aboriginal or Torres Strait Islander: 5.6% intervention and 9.0% control **Depression risk criteria** *Instrument*: Adverse family characteristics identified with the Brisbane Evaluation of Needs Questionnaire *Cutoff score:* • 1 or more of 1st tier risk factors -OR- • 3 or more of 2nd tier risk factors *Screen timing:* Immediately following birth **Other inclusion criteria** *Included:* Good English literacy skills	**Intervention group** Home visits by child health nurse – weekly for 1st 6 wks, fortnightly for wks 7 to 13, and monthly for mos 4 to 6 **Control group** Routine primary care

Evidence Table 5. Key Question 3: Studies of screening interventions for perinatal depression: Screening during postpartum (continued)

Assessment Timing	Depression Outcomes	Other Outcomes
6 wks PP	**Mean EPDS scores** Intervention: 5.67 (SD 4.14) Control: 7.90 (SD 5.89) *F*(1, 169): 7.35, *P* = 0.004	**Child health (30-item self-report)** No significant differences in rate of breastfeeding, knowledge or practice of SIDS risk minimizing, or use of health services
	EPDS > 12 Intervention: 5.8% Control: 20.7% X^2(1): 8.30, *P* = 0.003 Impact of change in EPDS scores over time was concentrated among the primiparous women	**Parenting Stress Index (child reinforces parent subscale)** Intervention: 9.59 (SD 2.92) Control: 11.12 (SD 3.78) *F*(1, 169): 8.72, *P* = 0.004 High scores indicate that parent-child interactions are damaging to parental perception of their competence
		HOME Inventory (45 items) Intervention: 28.34 (SD 2.90) Control: 25.51 (SD 0.59) *F*(1, 169): *P* < 0.001 Intervention group had significantly better home environment than control group
		Patient satisfaction (10 items on health care) Greater satisfaction was found for the home-based program compared with standard services

Evidence Table 5. Key Question 3: Studies of screening interventions for perinatal depression: Screening during postpartum (continued)

Author, Year Quality Rating	Study Characteristics	Population Characteristics Inclusion Criteria	Treatment
Author Chabrol et al., 2002 **Quality rating** 15	**Place** France **Recruitment setting** Hospital maternity ward **Sample size** *1st intervention*: Intervention: 97 Control: 114 *2nd intervention*: Intervention: 18 Control: 30 **Design** Quasi-randomized controlled trial (no randomization for screening, only for treatment)	**Age** Intervention: mean 30.4 (SD 4) yrs Control: mean 29.6 (SD 5) yrs **Race/ethnicity** Caucasian: 100% **Depression risk criteria** *Instrument:* EPDS *Cutoff score:* ≥ 9 *Screen timing:* Majority at the 2nd and 3rd day PP **Other inclusion criteria** *Included:* No current treatment with psychiatrists or psychologists, good French language skills	**Intervention group** *1st intervention:* 1-hr educational, supportive and cognitive behavioral session during hospitalization *2nd intervention:* Patients in the intervention group with EPDS ≥ 11 and depression determined by DSM-IV with MINI after 4 to 6 wks received additional 5 to 8 sessions of an at-home cognitive behavioral therapy program **Control group** Routine primary care

Evidence Table 5. Key Question 3: Studies of screening interventions for perinatal depression: Screening during postpartum (continued)

Assessment Timing	Depression Outcomes	Other Outcomes
1st intervention: 4 to 6 wks PP *2nd intervention*: 10 to 12 wks PP	**EPDC ≥ 11 (4 to 6 wks PP)** Intervention: 30.2% depressed Control: 48.2% depressed X^2(1): 7.36, P = 0.0067 **Mean EPDS scores (4 to 6 wks PP)** Intervention: 8.5 (SD 4.0) Control: 10.3 (SD 4.4) *t*-test(209): 3.06, P = 0.0024 **Mean HDRS scores (10 to 12 wks)** Intervention: 5.7 (SD 3.3) Control: 16.2 (SD 4.5) *t*-test(49): 8.4, P < 0.0001 **Mean BDI scores (10 to 12 wks)** Intervention: 4.7 (SD 3.0) Control: 15.7 (SD 4.4) *t*-test(49): 9.0, P < 0.0001 **Mean EPDS scores (10 to 12 wks)** Intervention: 5.9 (SD 2.7) Control: 13.7 (SD 3.6) *t*-test(49): 7.7, P < 0.0001 **Recovery rate: HDRS < 7 (10 to 12 wks)** Intervention: 66.6% Control: 6.6% X^2(1): 16.8, P < 0.0001 **Recovery rate: BDI < 4 (10 to 12 wks)** Intervention: 61.1% Control: 3.3% X^2(1): 17.7, P < 0.0001	NR

Evidence Table 5. Key Question 3: Studies of screening interventions for perinatal depression: Screening during postpartum (continued)

Author, Year Quality Rating	Study Characteristics	Population Characteristics Inclusion Criteria	Treatment
Author Chen et al., 2000 **Quality rating** 14	**Place** Taiwan **Recruitment setting** Hospital maternity wards **Sample size** Intervention: 30 Control: 30 **Design** RCT	**Age** Mean: 29.1 (SD 4.2) yrs Range: 19 to 40 yrs **Race/ethnicity** NR **Depression risk criteria** *Instrument:* Taiwanese BDI *Cutoff score:* ≥ 10 *Screen timing:* 3 wks PP **Other inclusion criteria** *Included:* > 18 yrs of age, survival of infant, at least HS education	**Intervention group** Support groups of 5 to 6 mothers with their infants and a registered nurse met for 4 weekly sessions of 1.5 to 2 hrs duration **Control group** No support group
Author Dennis, 2003 **Quality rating** 22	**Place** Canada **Recruitment setting** Child immunization clinics **Sample size** Intervention: 20 Control: 22 **Design** RCT (for intervention, not for screening)	**Age** 18 to 24 yrs: 14% 25 to 34 yrs: 76% 35 yrs or over: 10% **Race/ethnicity** NR **Depression risk criteria** *Instrument:* EPDS *Cutoff score:* > 9 *Screen timing:* 8 to12 wks PP **Other inclusion criteria** *Included:* At least 18 yrs of age, singleton birth, English-speaking, able to access phone calls, birth at 37 wks GA or more, resided in surrounding region	**Intervention group** Lay peer support in the form of a paired peer volunteer (mother who previously had PP depression and attended a 4-hr training) who provided telephone-based peer support **Control group** Standard community PP services only

Evidence Table 5. Key Question 3: Studies of screening interventions for perinatal depression: Screening during postpartum (continued)

Assessment Timing	Depression Outcomes	Other Outcomes
15 wks PP	**Mean change in BDI score** Intervention: -6.60 (SD 5.89) Control: -1.40 (SD 8.33) **BDI ≥ 10** Intervention: 33.3% Control: 60.0% Significance tests for within-group changes over time were conducted but not for between-group differences in changes over time. (Both groups started out with 100% BDI ≥ 10)	**Mean change in Perceived Stress Scale** Intervention: -3.75 (SD 4.53) Control: -1.30 (SD 4.26) **Mean change in Interpersonal Support Evaluation List** Intervention: 2.60 (SD 5.08) Control: 0.00 (SD 5.14) **Mean change in Self-Esteem Inventory** Intervention: 1.03 (SD 4.57) Control: 1.03 (SD 2.97)
4 and 8 wks post-randomization, which occurred 8 to 12 wks PP	**EPDS > 9 (after 4 wks of support)** Intervention: 45.0% Control: 72.7% OR: 3.26 (95% CI, 0.90 to 11.81) **EPDS > 9 (after 8 wks of support)** Intervention: 35.0% Control: 76.2% OR: 5.94 (95% CI, 1.52 to 23.18) **EPDS > 12 (after 4 wks of support)** Intervention: 10.0% Control: 40.9% OR: 6.23 (95% CI, 1.15 to 33.77) $X^2(1)$: 5.18, $P = 0.02$ **EPDS > 12 (after 8 wks of support)** Intervention: 15.0% Control: 52.4% OR: 6.23 (95% CI, 1.40 to 27.84) Adjusted OR: 4.7 (95% CI, 0.91 to 25.46) $X^2(1)$: 6.37, $P < 0.01$ **Mean EPDS score (after 4 wks of support)** Intervention: 8.5 (SD 3.7) Control: 12.1 (SD 4.6) *t*-test(40): 2.8, $P = 0.008$ **Mean EPDS score (after 8 wks of support)** *t*-test(39): 2.9, $P = 0.006$	**Mean maternal self-esteem score (after 8 wks of support)** Intervention: 30.00 (SD 4.21) Control: 28.57 (SD 3.83) Difference not statistically significant **Mean child care stress score (after 8 wks of support)** Intervention: 4.95 (SD 2.68) Control: 6.48 (SD 3.63) Difference not statistically significant **Mean maternal loneliness score (after 8 wks of support)** Intervention: 20.37 (SD 5.23) Control: 23.91 (SD 6.07) Difference not statistically significant Note that baseline measures are also provided but the significance of the difference in changes over time is not computed

Evidence Table 5. Key Question 3: Studies of screening interventions for perinatal depression: Screening during postpartum (continued)

Author, Year Quality Rating	Study Characteristics	Population Characteristics Inclusion Criteria	Treatment
Author Fleming et al., 1992 **Quality rating** 11	**Place** Canada **Recruitment setting** Hospital maternity ward **Sample size** Intervention: 44 Control 1: 15 Control 2: 83 **Design** Controlled trial, no randomization	**Age** Range: 22 to 36 yrs **Race/ethnicity** NR **Depression risk criteria** *Instruments:* CES, EPDS, MAACL *Cutoff scores:* CES ≥ 35 and either EPDS ≥ 13 or MAACL ≥ 21 *Screen timing:* 2 wks PP **Other inclusion criteria** *Included:* Primiparous, married or cohabiting, full-term vaginal deliveries, no known past or current serious psychiatric or gynecologic histories, English speaking	**Intervention group** PP social support group meeting for 2 hrs weekly for 8 wks **Control group 1** Group-by-mail sent weekly information on PP depression on the same topics and schedule as the support group meetings **Control group 2** No intervention

Evidence Table 5. Key Question 3: Studies of screening interventions for perinatal depression: Screening during postpartum (continued)

Assessment Timing	Depression Outcomes	Other Outcomes
6 wks and 5 mos	**CES mood score** No statistically significant effects of the social support intervention compared to either the mail or no intervention groups	**Negative self image** Social support intervention had a statistically significant negative impact **Attachment to infant** No statistically significant effects of the social support intervention **Maternal-infant interaction** Social support intervention had a statistically significant increase in number of approaches to the infant **Non-cry vocalizations** Babies of social support women decreased crying from 6 wks to 5 mos whereas those of the other groups either had no change or increased crying

Evidence Table 5. Key Question 3: Studies of screening interventions for perinatal depression: Screening during postpartum (continued)

Author, Year Quality Rating	Study Characteristics	Population Characteristics Inclusion Criteria	Treatment
Author Hiscock and Wake, 2002 **Quality rating** 19	**Place** Australia **Recruitment setting** Maternal and child health center **Sample size** Intervention: 33 Control: 33 **Design** RCT	**Age** Intervention: 34.1 (SD 3.6) yrs Control: 33.3 (SD 5.6) yrs **Race/ethnicity** NR **Depression risk criteria** *Instrument:* EPDS *Cutoff score:* ≥ 10 *Screen timing:* 7 to 9 mos PP **Other inclusion criteria** *Included:* Reported infant sleep problems, adequate English skills to complete questionnaires, not receiving treatment for PP depression, reported no thoughts of self harm, infants had no major medical or developmental problem, not receiving help for infant sleep problems	**Intervention group** Infant Sleep Intervention, comprised three private sessions (one session every 2 wks) held at the local maternal and child health center where sleep management plans included an emphasis on "controlled crying"; (parents were encouraged to respond to their infant's cry at increasing time intervals, allowing the infant to fall asleep by itself) **Control group** Mailed single sheet describing normal sleep patterns

Evidence Table 5. Key Question 3: Studies of screening interventions for perinatal depression: Screening during postpartum (continued)

Assessment Timing	Depression Outcomes	Other Outcomes
2 mos and 4 mos after randomization, which occurred 7 to 9 mos PP	**Mean change in EPDS (baseline to 2 mos)** Intervention: -6.0 (95% CI, -7.5 to -4.0) Control: -3.7 (95% CI, -4.9 to -2.6) *t*-test(64): *P* = 0.01 **Mean change in EPDS (baseline to 4 mos)** Intervention: -6.5 (95% CI, -7.9 to -5.1) Control: -4.2 (95% CI, -5.9 to -2.5) *t*-test(64): *P* = 0.04 No significant differences between groups for higher cutoff points	**Infant's sleep problems resolved (baseline to 2 mos)** Intervention: 26/33 Control: 13/33 X^2(1): *P* = 0.001 **Infant's sleep problems resolved (baseline to 4 mos)** Intervention: 21/32 Control: 14/30 X^2(1): *P* = 0.13 **Mother's sleep quality (baseline to 2 mos)** Intervention mothers more likely to rate their own sleep quality as "very good" compared with controls X^2(1): 7.58, *P* = 0.06 **Mother's sleep quality (baseline to 4 mos)** No significant differences between intervention and control women **Mothers report having enough sleep (baseline to 2 mos)** Intervention mothers more likely to report having enough sleep compared with controls X^2(1): 5.00, *P* = 0.09 **Mothers report having enough sleep (baseline to 4 mos)** No significant differences between intervention and control women **Mothers report "no stress" (baseline to 2 mos and baseline to 4 mos)** No significant differences between intervention and control women

Evidence Table 5. Key Question 3: Studies of screening interventions for perinatal depression: Screening during postpartum (continued)

Author, Year Quality Rating	Study Characteristics	Population Characteristics Inclusion Criteria	Treatment
Author Honey et al., 2002 **Quality rating** 16	**Place** UK **Recruitment setting** Referred by health visitor **Sample size** Intervention: 23 Control: 22 **Design** RCT	**Age** Intervention: 29.30 (SD 5.36) yrs Control: 26.48 (SD 5.68) yrs **Race/ethnicity** NR **Depression risk criteria** *Instrument:* EPDS *Cutoff score:* > 12 *Screen timing:* < 12 mos PP, mean 5.98 mos PP (SD 2.34 mos) **Other inclusion criteria** *Included:* Attending mother and baby clinics, not exhibiting psychotic symptoms, most recent child under 12 mos	**Intervention group** PEG consisting of 8 weekly, 2-hr meetings run by health visitors **Control group** Routine primary care
Author Horowitz, et al., 2001 **Quality rating** 15	**Place** Boston, USA **Recuitment setting** Hospital maternity wards **Sample size** Intervention: 60 Control: 57 **Design** RCT	**Age** Mean: 31 yrs Range: 17 to 41 yrs **Race/ethnicity** White: 68.9% Black: 7.4% Hispanic: 7.4% Mixed: 7.4% Other: 4% Asian: 3.3% Native American: 1.6% **Depression risk criteria** *Instrument:* EPDS *Cutoff score:* ≥ 10 *Screen timing:* 2 to 4 wks PP **Other inclusion criteria** None	**Intervention group** Interactive coaching, designed to promote maternal-infant responsiveness; composed of three home visits when infants were 4 to 8 wks, 10 to 14 wks, and 14 to18 wks **Control group** Standard community services

Evidence Table 5. Key Question 3: Studies of screening interventions for perinatal depression: Screening during postpartum (continued)

Assessment Timing	Depression Outcomes	Other Outcomes
At end of 8-wk PEG and 6 mos following end of PEG	**Mean EPDS score (baseline)** Intervention: 19.35 (SD 4.39) Control: 17.95 (SD 3.95) **Mean EPDS score (end of 8-wk PEG)** Intervention: 14.87 (SD 5.97) Control: 16.95 (SD 5.44) **Mean EPDS score (6 mos after PEG)** Intervention: 12.55 (SD 4.62) Control: 15.63 (SD 7.28) **Mean change in EPDS scores (baseline to 6 mos after PEG)** Intervention women had a significant greater decrease in means scores compared to control women F (2, 43): 3.16, $P < 0.05$ **EPDS ≤ 12 (end of 8-wk PEG)** Intervention: 35% Control: 27% X^2(1): 0.30, $P > 0.1$ **EPDS ≤ 12 (6 mos after PEG)** Intervention: 65% Control: 36% X^2(1): 3.75, $P \leq 0.05$	**Social support** No statistically significant effect of intervention **Marital relationship** No statistically significant effect of intervention **Coping scales** No statistically significant effect of intervention
4 to 8 wks PP; 10 to 14 wks PP; and 14 to 18 wks PP	**Mean BDI-II score (4 to 8 wks PP)** Intervention: 15.5 (SD 1.17) Control: 13.24 (SD 0.92) **Mean BDI-II score (10 to 14 wks PP)** Intervention: 10.99 (SD 0.96) Control: 10.10 (SD 0.84) **Mean BDI-II score (14 to 18 wks PP)** Intervention: 10.27 (SD 0.99) Control: 9.51 (SD 0.77) **Difference in changes in mean BDI-II scores over time** F (2, 115): 0.36, $P = 0.67$	**Mother-infant responsiveness measured by mean DMC score (4 to 8 wks PP)** Intervention: 8.83 (SD 1.76) Control: 8.67 (SD 1.64) **Mean DMC score (10 to 14 wks PP)** Intervention: 9.73 (SD 1.65) Control: 8.77 (SD 1.72) **Mean DMC score (14 to 18 wks PP)** Intervention: 9.55 (SD 1.77) Control: 8.80 (SD 1.86) **Difference in changes in mean DMC scores over time** F (2, 115): 2.14, $P = 0.121$

Evidence Table 5. Key Question 3: Studies of screening interventions for perinatal depression: Screening during postpartum (continued)

Author, Year Quality Rating	Study Characteristics	Population Characteristics Inclusion Criteria	Treatment
Author Onazawa et al., 2001 **Quality rating** 16	**Place** UK **Recruitment setting** Hospital maternity ward **Sample size** Intervention: 19 Control: 15 **Design** RCT	**Age** Intervention: mean 32 yrs (95% CI, 29.6 to 34.5) Control: mean 33 yrs (95% CI, 31.2 to 35.9) **Race/ethnicity** White: 88% Other: 12% **Depression risk criteria** *Instrument:* EPDS *Cutoff score:* ≥ 13 *Screen timing:* 4 wks PP **Other inclusion criteria** *Included:* Primiparous, 18 to 45 yrs of age, singleton birth from 37 to 42 wks GA with no congenital abnormalities and not requiring NICU care	**Intervention group** Infant massage class plus support group (1-hr classes weekly for 5 wks). Infant massage class designed to teach parents the techniques of infant massage by encouraging parents to observe and respond to their infant's body language and cues and to adjust touch accordingly **Control group** Support group only for 5 wks
Author Wisner and Wheeler, 1994 **Quality rating** 12	**Place** USA **Recruitment setting** University-based, outpatient program for pregnant and PP women **Sample size** Intervention: 15 Control: 8 **Design** Open controlled trial	**Age** NR **Race/ethnicity** NR **Depression risk criteria** At least one episode of PP depression in prior pregnancy **Other inclusion criteria** *Included:* No history of psychosis or bipolar disorder, married or in stable relationship for at least 6 mos, no major obstetrical problems	**Intervention group** PP antidepressant medication and monitoring **Control group** PP monitoring only

Evidence Table 5. Key Question 3: Studies of screening interventions for perinatal depression: Screening during postpartum (continued)

Assessment Timing	Depression Outcomes	Other Outcomes
14 wks PP	**Median EPDS score (baseline)** Intervention: 15.0 (95% CI, 14.0 to 18.1) Control: 16.0 (95% CI, 14.7 to 18.7) **Median EPDS score (final session)** Intervention: 5.0 (95% CI, 2.2 to 7.8) Control: 10.0 (95% CI, 7.7 to 11.8) **Change in median EPDS score** Intervention: 12.0 (94% CI, 8.0 to 14.2) Control: 6.0 (95% CI, 4.6 to 9.0) Z: -2.2, *P* = 0.03	All measures of mother-infant interaction showed significantly greater improvement in intervention group compared with control group
3 mos PP	**Recurrence of DSM-III-R major depression** Intervention: 6.7% Control: 62.5% Fisher's exact test, two-tailed: *P* = 0.0086 OR: 19.2 (95% CI, 1.5 to 1,179)	NR

Evidence Table 5. Key Question 3: Studies of screening interventions for perinatal depression: Screening during postpartum (continued)

Author, Year Quality Rating	Study Characteristics	Population Characteristics Inclusion Criteria	Treatment
Author Wisner et al., 2001	**Place** USA	**Age** NR	**Intervention group** Nortriptyline
Quality rating 18	**Recruitment setting** NR	**Race/ethnicity** NR	**Control group** Placebo
	Sample size Intervention: 26 Control: 25	**Depression risk criteria** At least one past episode of PP-onset major depression within past 5 yrs	
	Design RCT	**Other inclusion criteria** *Included:* ≤ 35 weeks GA, 21 to 45 yrs of age, no depressive episode since the conception of index pregnancy, not exposed to antidepressants after 1st trimester, no other Axis I diagnosis except generalized anxiety or panic disorder	

Evidence Table 5. Key Question 3: Studies of screening interventions for perinatal depression: Screening during postpartum (continued)

Assessment Timing	Depression Outcomes	Other Outcomes
20 wks PP	**Recurrence of RDC major depression** Intervention: 23% Control: 24% Fisher's exact test, two-tailed: *P* = 1.00 **Time to recurrence** No statistically significant difference in time to recurrence between nortriptyline and placebo. Exact log-rank: < 0.00, *P* = 0.83	NR

References

Affonso DD, Lovett S, Paul SM, et al. A standardized interview that differentiates pregnancy and postpartum symptoms from perinatal clinical depression. Birth 1990; 17(3):121-30.

Areias ME, Kumar R, Barros H, et al. 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(1):30-5.

Armstrong KL, Fraser JA, Dadds MR, et al. A randomized, controlled trial of nurse home visiting to vulnerable families with newborns. J Paediatr Child Health 1999; 35(3):237-44.

Ballard CG, Davis R, Cullen PC, et al. Prevalence of postnatal psychiatric morbidity in mothers and fathers. Br J Psychiatry 1994; 164(6):782-8.

Beck CT, Gable RK. Comparative analysis of the performance of the Postpartum Depression Screening Scale with two other depression instruments. Nurs Res 2001; 50(4):242-50.

Berle J, Aarre T, Mykletun A, et al. Screening for postnatal depression. Validation of the Norwegian version of the Edinburgh Postnatal Depression Scale, and assessment of risk factors for postnatal depression. J Affect Disord 2003; 76(1-3):151-6.

Boyce P, Stubbs J, Todd A. The Edinburgh Postnatal Depression Scale: Validation for an Australian sample. Aust N Z J Psychiatry 1993; 27(3):472-6.

Brugha TS, Wheatley S, Taub NA, et al. Pragmatic randomized trial of antenatal intervention to prevent post-natal depression by reducing psychosocial risk factors. Psychol Med 2000; 30(6):1273-81.

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(4):351-8.

Campbell SB, Cohn JF. Prevalence and correlates of postpartum depression in first-time mothers. J Abnorm Psychol 1991; 100(4):594-9.

Chabrol H, Teissedre F, Saint-Jean M, et al. Prevention and treatment of post-partum depression: A controlled randomized study on women at risk. Psychol Med 2002; 32(6):1039-47.

Chen CH, Tseng YF, Chou FH, et al. Effects of support group intervention in postnatally distressed women. A controlled study in Taiwan. J Psychosom Res 2000; 49(6):395-9.

Cooper PJ, Campbell EA, Day A, et al. Non-psychotic psychiatric disorder after childbirth. A prospective study of prevalence, incidence, course and nature. Br J Psychiatry 1988; 152:799-806.

Cooper PJ, Murray L, Hooper R, et al. The development and validation of a predictive index for postpartum depression. Psychol Med 1996; 26(3):627-34.

Cox JL, Connor Y, Kendell RE. Prospective study of the psychiatric disorders of childbirth. Br J Psychiatry 1982; 140:111-7.

Cox JL, Murray D, Chapman G. A controlled study of the onset, duration and prevalence of postnatal depression. Br J Psychiatry 1993; 163:27-31.

Cox J, Chapman G, Murray D, et al. Validation of the Edinburgh Postnatal Depression Scale (EPDS) in non-postnatal women. J Affect Disord 1996; 39(3):185-9.

Dennis CL. The effect of peer support on postpartum depression: A pilot randomized controlled trial. Can J Psychiatry 2003; 48(2):115-24.

Elliott SA, Leverton TJ, Sanjack M, et al. Promoting mental health after childbirth: A controlled trial of primary prevention of postnatal depression. Br J Clin Psychol 2000; 39 (Pt 3):223-41.

Fleming AS, Klein E, Corter C. The effects of a social support group on depression, maternal attitudes and behavior in new mothers. J Child Psychol Psychiatry 1992; 33(4):685-98.

Garcia-Esteve L, Ascaso C, Ojuel J, et al. Validation of the Edinburgh Postnatal Depression Scale (EPDS) in Spanish mothers. J Affect Disord 2003; 75(1):71-6.

Georgiopoulos AM, Bryan TL, Wollan P, et al. Routine screening for postpartum depression. J Fam Pract 2001; 50(2):117-22.

Gotlib IH, Whiffen VE, Mount JH, et al. Prevalence rates and demographic characteristics associated with depression in pregnancy and the postpartum. J Consult Clin Psychol 1989; 57(2):269-74.

Guedeney N, Fermanian J. Validation study of the French version of the Edinburgh Postnatal Depression Scale (EPDS): New results about use and psychometric properties. Eur Psychiatry 1998; 13:83-9.

Harris B, Huckle P, Thomas R, et al. The use of rating scales to identify post-natal depression . Br J Psychiatry 1989; 154:813-7.

Hiscock H, Wake M. Randomised controlled trial of behavioural infant sleep intervention to improve infant sleep and maternal mood. Br Med J 2002; 324(7345):1062-5.

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(3):445-53.

Honey KL, Bennett P, Morgan M. A brief psycho-educational group intervention for postnatal depression. Br J Clin Psychol 2002; 41(Pt 4):405-9.

Horowitz JA, Bell M, Trybulski J, et al. Promoting responsiveness between mothers with depressive symptoms and their infants. J Nurs Scholarsh 2001; 33(4):323-9.

Kent GN, Stuckey BG, Allen JR, et al. Postpartum thyroid dysfunction: Clinical assessment and relationship to psychiatric affective morbidity. Clin Endocrinol 1999; 51(4):429-38.

Kitamura T, Shima S, Sugawara M, et al. Psychological and social correlates of the onset of affective disorders among pregnant women. Psychol Med 1993; 23:967-75.

Kitamura T, Shima S, Sugawara M, et al. Temporal variation of validity of self-rating questionnaires: Repeated use of the General Health Questionnaire and Zung's Self-rating Depression Scale among women during antenatal and postnatal periods. Acta Psychiatr Scand 1994; 90(6):446-50.

Kitamura T, Sugawara M, Shima S, et al. Temporal variation of validity of self-rating questionnaires: Improved validity of repeated use of Zung's Self-Rating Depression Scale among women during the perinatal period. J Psychosom Obstet Gynecol 1999; 20(2):112-7.

Kumar R, Robson KM. A prospective study of emotional disorders in childbearing women. Br J Psychiatry 1984; 144:35-47.

Lawrie T, Hofmeyr G, de Jager M, et al. Validation of the Edinburgh Postnatal Depression Scale on a cohort of South African women. S Afr Med J 1998; 88(10):1340-4.

Lee D, Yip A, Chiu H, et al. A psychiatric epidemiological study of postpartum Chinese women. Am J Psychiatry 2001; 158(2):220-6.

Lee D, Yip A, Chiu H, et al. Screening for postnatal depression using the double-test strategy. Psychosom Med 2000; 62(2):258-63.

Lee D, Yip A, Chiu H, et al. Screening for postnatal depression: Are specific instruments mandatory? J Affect Disord 2001; 63(1-3):233-8.

Lee D, Yip S, Chiu H, et al. Detecting postnatal depression in Chinese women. Validation of the Chinese version of the Edinburgh Postnatal Depression Scale. Br J Psychiatry 1998; 172:433-7.

Leverton TJ, Elliott SA. Is the EPDS a magic wand? 1. A comparison of the Edinburgh Postnatal Depression Scale and health visitor report as predictors of diagnosis on the Present State Examination. J Reprod Infant Psychol 2000; 18(4):279-96.

Lucas A, Pizarro E, Granada ML, et al. Postpartum thyroid dysfunction and postpartum depression: Are they two linked disorders? Clin Endocrinol 2001; 55(6):809-14.

Matthey S, Barnett B, Howie P, et al. Diagnosing postpartum depression in mothers and fathers: Whatever happened to anxiety? J Affect Disord 2003; 74(2):139-47.

Murray D, Cox JL. Screening for depression during pregnancy with the Edinburgh Depression Scale (EPDS). J Reprod Infant Psychol 1990; 8(2):99-107.

Murray L, Carothers A. The validation of the Edinburgh Post-natal Depression Scale on a community sample. Br J Psychiatry 1990; 157:288-90.

Muzik M, Klier C, Rosenblum K, et al. Are commonly used self-report inventories suitable for screening postpartum depression and anxiety disorders? Acta Psychiatr Scand 2000; 102(1):71-3.

O'Hara MW, Neunaber DJ, Zekoski EM. Prospective study of postpartum depression: Prevalence, course, and predictive factors. J Abnorm Psychol 1984; 93(2):158-71.

O'Hara MW, Zekoski EM, Philipps LH, et al. Controlled prospective study of postpartum mood disorders: Comparison of childbearing and nonchildbearing women. J Abnorm Psychol 1990; 99(1):3-15.

Onozawa K, Glover V, Adams D, et al. Infant massage improves mother-infant interaction for mothers with postnatal depression. J Affect Disord 2001; 63(1-3):201-7.

Pop VJ, Essed GG, de Geus CA, et al. Prevalence of post partum depression--or is it post-puerperium depression? Acta Obstet Gynecol Scand 1993; 72(5):354-8.

Stamp GE, Williams AS, Crowther CA. Evaluation of antenatal and postnatal support to overcome postnatal depression: A randomized, controlled trial. Birth 1995; 22(3):138-43.

Watson JP, Elliott SA, Rugg AJ, et al. Psychiatric disorder in pregnancy and the first postnatal year. Br J Psychiatry 1984; 144:453-62.

Whiffen VE. Screening for postpartum depression: A methodological note. J Clin Psychol 1988; 44(3):367-71.

Whiffen V. Vulnerability of postpartum depression: A prospective multivariate study. J Abnorm Psychol 1988; 97(4):467-74.

Wickberg B, Hwang C. The Edinburgh Postnatal Depression Scale: Validation on a Swedish community sample. Acta Psychiatr Scand 1996; 94(3):181-4.

Wisner KL, Perel JM, Peindl KS, et al. Prevention of recurrent postpartum depression: A randomized clinical trial. J Clin Psychiatry 2001; 62(2):82-6.

Wisner KL, Wheeler SB. Prevention of recurrent postpartum major depression. Hosp Community Psychiatry 1994; 45(12):1191-6.

Yamashita H, Yoshida K, Nakano H, et al. Postnatal depression in Japanese women. Detecting the early onset of postnatal depression by closely monitoring the postpartum mood. J Affect Disord 2000; 58(2):145-54.

Yonkers KA, Ramin SM, Rush AJ, et al. Onset and persistence of postpartum depression in an inner-city maternal health clinic system. Am J Psychiatry 2001; 158(11):1856-63.

Yoshida K, Marks M, Kibe N, et al. Postnatal depression in Japanese women who have given birth in England. J Affect Disord 1997; 43(1):69-77.

Zlotnick C, Johnson SL, Miller IW, et al. Postpartum depression in women receiving public assistance: Pilot study of an interpersonal-therapy-oriented group intervention. Am J Psychiatry 2001; 158(4):638-40.

Excluded Articles

Abramowitz JS, Schwartz SA, Moore KM, et al. Obsessive-compulsive symptoms in pregnancy and the puerperium: A review of the literature. J Anxiety Disord 2003; 17(4):461-78.

Affonso DD, Mayberry LJ, Lovett S, et al. Pregnancy and postpartum depressive symptoms. J Womens Health 1993; 2(2):157-64.

Affonso DD, Arizmendi TG. Disturbances in post-partum adaptation and depressive symptomatology. J Psychosom Obstet Gynaecol 1986; 5(1):15-32.

Areias ME, Kumar R, Barros H, et al. 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(1):30-5.

Areias ME, Kumar R, Barros H, et al. Correlates of postnatal depression in mothers and fathers. Br J Psychiatry 1996; 169(1):36-41.

Beck CT. Postpartum depression predictors inventory--revised. Adv Neonatal Care 2003; 3(1):47-8.

Beck CT. Recognizing and screening for postpartum depression in mothers of NICU infants. Adv Neonatal Care 2003; 3(1):37-46.

Beck CT, Gable RK. Comparative analysis of the performance of the Postpartum Depression Screening Scale with two other depression instruments. Nurs Res 2001; 50(4):242-50.

Beck C, Gable R. Postpartum depression screening scale: Spanish version. Nurs Res 2003; 52(5):296-306.

Beeghly M, Olson KL, Weinberg MK, et al. Prevalence, stability, and socio-demographic correlates of depressive symptoms in Black mothers during the first 18 months postpartum. Matern Child Health J 2003; 7(3):157-68.

Bijl RV, van Zessen G, Ravelli A. Psychiatric morbidity among adults in The Netherlands: The NEMESIS-Study. II. Prevalence of psychiatric disorders. Netherlands Mental Health Survey and Incidence Study. Ned Tijdschr Geneeskd 1997; 141(50):2453-60.

Brugha TS, Wheatley S, Taub NA, et al. Pragmatic randomized trial of antenatal intervention to prevent post-natal depression by reducing psychosocial risk factors. Psychol Med 2000; 30(6):1273-81.

Caravale B, Allemand F, Libenson MH. Factors predictive of seizures and neurologic outcome in perinatal depression. Pediatr Neurol 2003; 29(1):18-25.

Carro MG, Grant KE, Gotlib IH, et al. Postpartum depression and child development: An investigation of mothers. Dev Psychopathol 1993; 5(4):567-79.

Chan S, Levy V. Postnatal depression: A qualitative study of the experiences of a group of Hong Kong Chinese women. J Clin Nurs 2004; 13(1):120-3.

Cooper PJ, Campbell EA, Day A, et al. Non-psychotic psychiatric disorder after childbirth. A prospective study of prevalence, incidence, course and nature. Br J Psychiatry 1988; 152:799-806.

Cooper PJ, Murray L, Hooper R, et al. The development and validation of a predictive index for postpartum depression. Psychol Med 1996; 26(3):627-34.

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.

Cutrona CE. Causal attributions and perinatal depression. J Abnorm Psychol 1983; 92(2):161-72.

Demyttenaere K, Lenaerts H, Nijs P, et al. Individual coping style and psychological attitudes during pregnancy and predict depression levels during pregnancy and during postpartum. Acta Psychiatr Scand 1995; 91(2):95-102.

Dennis C. Can we identify mothers at risk for postpartum depression in the immediate postpartum period using the Edinburgh Postnatal Depression Scale? J Affect Disord 2004; 78(2):163-9.

Eberhard-Gran M, Tambs K, Opjordsmoen S, et al. A comparison of anxiety and depressive symptomatology in postpartum and non-postpartum mothers. Soc Psychiatry Psychiatr Epidemiol 2003; 38(10):551-6.

Edwards M, Waldorf M. Reclaiming Birth: History and heroines of American Childbirth Reform. 1984: The Crossing Press, 1984: 4-6.

Ellison M, Hall J. Social stigma and compounded losses: Quality-of-life issues for multiple-birth families. Fertil Steril 2003; 80(2):405-14.

Evins GG, Theofrastous JP, Galvin SL. Postpartum depression: A comparison of screening and routine clinical evaluation. Am J Obstet Gynecol 2000; 182(5):1080-2.

Feggetter G, Cooper P, Gath D. Non-psychotic psychiatric disorders in women one year after childbirth. J Psychosom Res 1981; 25(5):369-72.

Fergusson DM, Horwood LJ, Thorpe K. Changes in depression during and following pregnancy. ALSPAC Study Team. Study of Pregnancy and Children. Paediatr Perinat Epidemiol 1996; 10(3):279-93.

Ghubash R, Abou-Saleh MT, Daradkeh TK. The validity of the Arabic Edinburgh Postnatal Depression Scale. Soc Psychiatry Psychiatr Epidemiol 1997; 32(8):474-6.

Gilman SE, Kawachi I, Fitzmaurice GM, et al. Family disruption in childhood and risk of adult depression. Am J Psychiatry 2003; 160(5):939-46.

Greene SM, Nugent JK, Wieczorek Deering DE, et al. The patterning of depressive symptoms in a sample of first-time mothers. Ir J Psychol 1991; 12(2):263-75.

Guedeney A, Doleans MC, Huot-Marchand M. Early screening of withdrawal reaction in the Maternal and Infant Welfare Protection program. Arch Pediatr 2003; 10(Suppl 1):131s-3s.

Guedeney N, Fermanian J, Guelfi JD, et al. The Edinburgh Postnatal Depression Scale (EPDS) and the detection of major depressive disorders in early postpartum: Some concerns about false negatives. J Affect Disord 2000; 61(1-2):107-12.

Harris B, Fung H, Johns S, et al. Transient postpartum thyroid dysfunction and postnatal depression. J Affect Disord 1989; 17(3):243-9.

Harris B, Johns S, Fung H, et al. The hormonal environment of post-natal depression . Br J Psychiatry 1989; 154:660-7.

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(3):445-53.

Holt WJ. The detection of postnatal depression in general practice using the Edinburgh postnatal depression scale. N Z Med J 1995; 108(994):57-9.

Honey KL, Bennett P, Morgan M. Predicting postnatal depression. J Affect Disord 2003; 76(1-3):201-10.

Kessler RC, McGonagle KA, Zhao S, et al. Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. Results from the National Comorbidity Survey. Arch Gen Psychiatry 1994; 51(1):8-19.

Kessler RC, Zhao S, Blazer DG, et al. Prevalence, correlates, and course of minor depression and major depression in the National Comorbidity Survey. J Affect Disord 1997; 45(1-2):19-30.

Kitamura T, Shima S, Sugawara M, et al. Psychological and social correlates of the onset of affective disorders among pregnant women. Psychol Med 1993; 23:967-75.

Kurki T HVRRMHYO. Depression and anxiety in early pregnancy and risk for preeclampsia. Obstet Gynecol 2000; 95(4):487-90.

Leavitt JW. Brought to Bed: Childbearing in America 1750 to 1950. New York: Oxford University Press, 179-86.

Lee D, Yip A, Chan S, et al. Postdelivery screening for postpartum depression. Psychosom Med 2003; 65(3):357-3361.

Leopold K, Zoschnick L. Postpartum depression. Female Pat 1997; 22:40-9.

Leverton TJ, Elliott SA. Transition to parenthood groups: A preventive intervention for perinatal depression? In: van Hall EV, Evereard W, eds. The free woman: woman's health in the 1990s. Invited papers of the 9th international conference of psychosomatic obstetrics and gynecology. Vol. 479-486. Lancaster, Pa: Parthenon Press, 1989.

McGill H, Burrows VL, Holland LA, et al. Postnatal depression: A Christchurch study. N Z Med J 1995; 108(999):162-5.

McKenry PC, Browne DH, Kotch JB, et al. Mediators of depression among low-income, adolescent mothers of infants: A longitudinal perspective. J Youth Adolescence 1990; 19(4):327-47.

McMahon C, Barnett B, Kowalenko N, et al. Postnatal depression, anxiety and unsettled infant behaviour. Aust N Z J Psychiatry 2001; 35(5):581-8.

Meager I, Milgrom J. Group treatment for postpartum depression: A pilot study. Aust N Z J Psychiatry 1996; 30(6):852-60.

Morris-Rush JK, Freda MC, Bernstein PS. Screening for postpartum depression in an inner-city population. Am J Obstet Gynecol 2003; 188(5):1217-9.

Muzik M, Klier C, Rosenblum K, et al. Are commonly used self-report inventories suitable for screening postpartum depression and anxiety disorders? Acta Psychiatr Scand 2000; 102(1):71-3.

Neugebauer R. Rate of depression in the puerperium. Br J Psychiatry 1983; 143:421-2.

Nott PN. Extent, timing and persistence of emotional disorders following childbirth. Br J Psychiatry 1987; 151:523-7.

O'Hara ME, Zekoski EM. Postpartum depression, a comprehensive review . Motherhood and Mental Illness 2. Butterworth & Co. Ltd, 1988: 17-63.

Okano T, Nomura J, Kumar R, et al. An epidemiological and clinical investigation of postpartum psychiatric illness in Japanese mothers. J Affect Disord 1998; 48(2-3):233-40.

Owen PJ, Lazarus JH. The treatment of post-partum thyroid disease. J Endocrinol Invest 2003; 26(4):290-1.

Peindl KS, Wisner KL. Successful recruitment strategies for women in postpartum mental health trials. J Psychiatr Res 2003; 37(2):117-25.

Pfost KS, Lum CU, Stevens MJ. Femininity and work plans protect women against postpartum dysphoria. Sex Roles 1989; 21(5-6):423-31.

Philipps LH, O'Hara MW. Prospective study of postpartum depression: 4 1/2-year follow-up of women and children. J Abnorm Psychol 1991; 100(2):151-5.

Posmontier B, Horowitz JA. Postpartum practices and depression prevalences: Technocentric and ethnokinship cultural perspectives . J Transcult Nurs 2004; 15(1):34-43.

Posner NA, Unterman RR, Williams KN, et al. Screening for postpartum depression: An antepartum questionnaire. J Reprod Med 1997; 42:207-15.

Priest SR, Henderson J, Evans SF, et al. Stress debriefing after childbirth: A randomised controlled trial. Med J Aust 2003; 178(11):542-5.

Rahman A, Iqbal Z, Harrington R. Life events, social support and depression in childbirth: Perspectives from a rural community in the developing world. Psychol Med 2003; 33(7):1161-7.

Rees WD. Parental depression before and after childbirth. An assessment with the Beck Depression Inventory. J R Coll Gen Pract 1971; 21(102):26-31.

Schaper A, Rooney B, Kay N, et al. Use of the Edinburgh Postnatal Depression Scale to identify

postpartum depression in a clinical setting. J Reprod Med 1994; 39(8):620-4.

Shakespeare J, Blake F, Garcia J. A qualitative study of the acceptability of routine screening of postnatal women using the Edinburgh Postnatal Depression Scale. Br J Gen Pract 2003; 53(493):614-9.

Sharp DJ. Validation of the 30-item General Health Questionnaire in early pregnancy. Psychol Med 1988; 18(2):503-7.

Shimizu Y, Kaplan B. Postpartum depression in the United States and Japan. J Cross-Cultural Psychol 1987; 18(1):15-30.

Silver L. Postnatal depression: An overview. J Fam Health Care 2003; 13(6):144-5.

Teissedre F, Chabrol H. Detecting women at risk for postnatal depression using the Edinburgh Postnatal Depression Scale at 2 to 3 days postpartum. Can J Psychiatry 2004; 49(1):51-4.

Thorpe K, Dragonas T, Golding J. The effects of psychosocial factors on the mother's emotional well-being. J Reprod Infant Psychol 1992; 10(4):205.

Troutman BR, Cutrona CE. Nonspychotic post partum depression among adolescent mothers. J Abnorm Psychol 1990; 99:69-78.

Verkerk GJ, Pop VJ, Van Son MJ, et al. Prediction of depression in the postpartum period: A longitudinal follow-up study in high-risk and low-risk women. J Affect Disord 2003; 77(2):159-66.

Wang SY, Jiang XY, Jan WC, et al. A comparative study of postnatal depression and its predictors in Taiwan and mainland China. Am J Obstet Gynecol 2003; 189(5):1407-12.

Webster J, Pritchard M, Creedy D, et al. A simplified predictive index for the detection of women at risk for postnatal depression. Birth 2003; 30(2):101-8.

Whiffen VE. Screening for postpartum depression: A methodological note. J Clin Psychol 1988; 44(3):367-71.

Whiffen V. Vulnerability of postpartum depression: A prospective multivariate study. J Abnorm Psychol 1988; 97(4):467-74.

Wickberg B, Hwang C. Counselling of postnatal depression: A controlled study on a population based Swedish sample. J Affect Disord 1996; 39(3):209-16.

Yonkers KA, Ramin SM, Rush AJ, et al. Onset and persistence of postpartum depression in an inner-city maternal health clinic system. Am J Psychiatry 2001; 158(11):1856-63.

Acknowledgments

This study was supported by Contract 290-02-0016 from the Agency for Healthcare Research and Quality (AHRQ), Task Order No. 4. We acknowledge the continuing support of Kenneth Fink, MD, MGA, MPH, Director of the AHRQ Evidence-Based Practice Center (EPC) Program, and Marian James, PhD, the AHRQ Task Order Officer for this project.

The investigators deeply appreciate the considerable support, commitment, and contributions of the EPC team staff at RTI International and the University of North Carolina (UNC). From UNC, we thank EPC Co-Director, Timothy S. Carey, MD, MPH; EPC Literature Search Specialist, B. Lynn Whitener, PhD; and Research Assistant Leah Randolph, MA. We also express our gratitude to Loraine Monroe, EPC word processing specialist, Debra Bost, EPC editor, and Kathleen Mohar, Manager, Publications Specialist Group at RTI International.

Technical Expert Advisory Group

We extend our appreciation to the members of our Technical Expert Advisory Group (TEAG), who provided advice and input during our research process. The RTI-UNC EPC team solicited the views of TEAG members from the beginning of the project. TEAG members also provided insights into and reactions to work in progress and advice on substantive issues and overlooked areas of research. TEAG members participated in refining the analytic framework and key questions and discussing the preliminary assessment of the literature, including inclusion/exclusion criteria and methods for data synthesis. The TEAG was both a substantive resource and a "sounding board" throughout the study. It was also the body from which expertise was formally sought at several junctions. TEAG members are listed below:

Jeffrey Kuller, MD
Associate Professor
Division of Maternal-Fetal Medicine
Duke University Medical Center

Michael W. O'Hara, PhD
Professor
University of Iowa

Susan F. Meikle, MD, MSPH
Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Katherine L. Wisner, MD, MS
Director
Women's Behavioral HealthCARE
Professor of Psychiatry, Obstetrics and Gynecology and Reproductive Sciences and Epidemiology
University of Pittsburgh Medical Center

Peer Reviewers

We gratefully acknowledge the following individuals who reviewed the initial draft of this report and provided us with constructive feedback. External reviewers comprised clinicians, researchers, representatives of professional societies, and potential users of the report. We would also like to extend our appreciation to David Atkins, MD, MPH from AHRQ for contributing peer review comments. Our peer review panel also includes all members of the TEAG. Peer review was a separate duty for these individuals and not part of their commitment as TEAG members. All are active professionals in the field. The peer reviewers were asked to provide comments on the content, structure, and format of the evidence report and to complete a checklist. The peer reviewers' comments and suggestions formed the basis of our revisions to the evidence report. Acknowledgments are made with the explicit statement that this does not constitute endorsement of the report.

Individuals

Cheryl Beck, DNSc, CNM, FAAN
Professor
School of Nursing
University of Connecticut

Diana Dell, MD, FACOG
Women's Behavioral Health Program
Department of Obstetrics and Gynecology
Duke University Medical Center

Judith Lumley, MD
Centre for Mother's and Children's Health
La Trobe University
Carlton, Victoria, Australia

Organizations

Shoshana Bennett, PhD
Postpartum Support International

Janet Chapin, RN, MPH
American College of Obstetricians and Gynecologists

Ruth Johnson, CNM
American College of Nurse-Midwives

Marlene Freeman, MD
American Psychiatric Association

Gwen Gjerdingen, MD, MS
American Association of Family Physicians

Sheila M. Marcus, MD
American Psychiatric Association

Laura J. Miller, MD
American Psychiatric Association

Darrel A. Regier, MD, MPH
American Psychiatric Association

Kimberly A. Yonkers, MD
American Psychiatric Association