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

AHRQ
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