Evidence-based Practice Center Systematic Review Protocol

Project Title: Efficacy and Safety of Screening for Postpartum Depression

I. Background and Objectives for the Systematic Review

Postpartum Depression

Depression is a potentially life-threatening condition with a substantial impact on quality of life. The impact of depression in postpartum women is at least as great as that for depression in other populations. Postpartum depression is defined in the *Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, Text Revision* (hereafter, *DSM-IV-TR*) as a major depressive disorder according to the diagnostic criteria listed in Table 1, with a secondary criterion of onset of symptoms within 4 weeks of delivery.[1] A new set of diagnostic criteria for psychiatric illness, the *Diagnostic and Statistical Manual of Mental Disorders, 5th Edition* (*DSM-V*), is currently scheduled for release in May 2013; preliminary discussions suggest that the overall diagnostic framework for postpartum depression (i.e., major depression with a specification of postpartum onset) will remain unchanged, although the window for diagnosis may be extended to 6 months after delivery.[2]

Other diagnostic standards allow the definition of onset to extend beyond 4 weeks and up to 12 months after delivery and/or add a "minor depression" subcategory (2 to 4 of the symptoms listed in Table 1). There is high-quality evidence for effective treatment of patients who meet criteria for major depression in other settings; evidence is inconsistent for postpartum depression.[3-5]

Table 1. DSM-IV-TR diagnostic criteria for major depressive disorder[1]

Criterion	Description
A.	Five (or more) of the symptoms below have been present during the same 2-week period and represent a change from previous functioning; at least one of the symptoms is either 1) depressed mood or 2) loss of interest or pleasure. (Note: Do not include symptoms that are clearly due to a general medical condition, or mood-incongruent delusions or hallucinations.) • Depressed mood most of the day, nearly every day, as indicated by either subjective report (e.g., feels sad or empty) or observation made by others (e.g., appears tearful) • Markedly diminished interest in pleasure in all, or almost all, activities most of the day, nearly every day (as indicated by either subjective account or observation made by others) • Significant weight loss when not dieting or weight gain (e.g., change of more than 5% body weight in a month), or decrease or increase in appetite nearly every day • Insomnia or hypersomnia nearly every day • Psychomotor agitation or retardation nearly every day (observable by others, not merely subjective feelings of restlessness or being slowed down) • Fatigue or loss of energy nearly every day • Feelings of worthlessness or excessive or inappropriate guilt (which may be delusional) nearly every day (not merely self-reproach or guilt about being sick) • Diminished ability to think or concentrate, or indecisiveness, nearly every day (either subjective account or as observed by others) • Recurrent thoughts of death (not just fear of dying), recurrent suicidal ideation without a specific plan,

Criterion	Description
	or a suicide attempt or a specific plan for committing suicide
B.	The symptoms do not meet the criteria for mixed episode (*DSM-IV-TR*, p. 365)
C.	The symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning
D.	The symptoms are not due to the direct physiological effects of a substance (e.g., a drug of abuse, medication) or a general condition (e.g., hypothyroidism)
E.	The symptoms are not better accounted for by bereavement, that is, after the loss of a loved one, the symptoms persist for longer than 2 months or are characterized by marked functional impairment, morbid preoccupation with worthlessness, suicidal ideation, psychotic symptoms, or psychomotor retardation

Abbreviations: *DSM-IV-TR = Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, Text Revision*

The most recent United States-based synthesis of the evidence, performed for the Agency for Healthcare Research and Quality (AHRQ) in 2005,[3,4] estimated that the prevalence of major depression alone during the first postpartum year is 1.0 to 5.9 percent and that the prevalence of major and minor depression combined is 6.5 to 12.9 percent. Incidence estimates for the first 3 postpartum months were up to 6.5 percent for major depression alone and 14.5 percent for major and minor depression. At the time of the AHRQ review, consistent limitations in the literature included small sample size (precluding subgroup analyses) and lack of generalizability.

Although the risk of suicide in women may be lower during pregnancy and the postpartum period,[6] a review of maternal mortality in the United Kingdom during the 1990s found that suicide was the leading cause of maternal mortality, accounting for 29 percent of maternal deaths.[7,8] Postpartum depression may increase the risk of infant mortality through neglect, abuse, or homicide.[9] Maternal depression clearly affects maternal-infant interactions and some measures of infant development.[10-13] Health care resource utilization is greater for women with postpartum depression than for postpartum women who are not depressed;[14] data on resource use for their infants are inconsistent.[15,16] Outcomes in the studies included in the two most recent systematic reviews were primarily scores on measures of depression, which are often used as end points in clinical trials of depression therapy; other clinical outcomes, such as measures of infant health, were not included.[3,4,17]

Potential Benefits of Screening for Postpartum Depression

Given the potential impact of postpartum depression on maternal and infant health, there has been considerable interest in strategies aimed at identifying women who are at risk for postpartum depression or who have postpartum depression, with the ultimate goal being the application of effective preventive or therapeutic interventions. Key components of any particular screening strategy for postpartum depression include 1) which screening test or instrument to use, 2) when to screen, 3) who should screen, and 4) how to use the results of the screening test. However, there is considerable uncertainty about all of these components, as seen in existing recommendations. All major organizations providing care to pregnant and postpartum women and infants recognize the risk of postpartum depression and the potential benefit of screening, but the strength of recommendations is variable. For example, none of the United States-based organizations recommend use of a specific instrument (Table 2). Factors limiting the strength of recommendations include the lack of sufficient data on the most appropriate

screening instrument, the optimal time(s) for screening,[18] issues concerning reimbursement and the scope of practice,[10,18] and the need for adequate systems to ensure appropriate care for women identified through screening.[10,11,19]

Table 2. Guidelines/recommendations for screening for postpartum depression

Organization	Statement	Date
U.S. Preventive Services Task Force (USPSTF)[19]	No specific recommendations for postpartum depression. Grade B recommendation *for* screening "when staff-assisted depression care supports are in place to assure accurate diagnosis, effective treatment, and follow-up;" Grade C recommendation *against* screening when such supports are not in place.	2009
American College of Obstetricians and Gynecologists, Committee on Obstetric Practice[18]	At this time there is insufficient evidence to support a firm recommendation for universal antepartum or postpartum screening. There are also insufficient data to recommend how often screening should be done. However, screening for depression has the potential to benefit a woman and her family and should be strongly considered. Medical practices should have a referral process for identified cases. Women with current depression or a history of major depression warrant particularly close monitoring and evaluation.	February 2010
American Academy of Pediatrics, Committee on Psychosocial Aspects of Child and Family Health[20]	Screening can be integrated, as recommended by *Bright Futures* and the American Academy of Pediatrics Mental Health Task Force, into the well-child care schedule and included in the prenatal visit. This screening has proven successful in practice in several initiatives and locations and is a best practice for primary care pediatricians caring for infants and their families. Intervention and referral are optimized by collaborative relationships with community resources and/or by colocated/integrated primary care and mental health practices.	November 2010
American Academy of Family Physicians	No specific recommendations for postpartum depression; general recommendations for screening follow those of the USPSTF.[19]	2010
American College of Nurse Midwives[21]	The American College of Nurse Midwives supports universal screening, treatment, and/or referral for depression in women as a part of routine primary health care.	2003
United Kingdom National Institute for Health and Clinical Excellence[22]	At a woman's first contact with a primary care provider, at her booking visit, and postnatally (usually at 4 to 6 weeks and 3 to 4 months), health care professionals (including midwives, obstetricians, health visitors, and general practitioners) should ask two questions to identify possible depression: • During the past month, have you often been bothered by feeling down, depressed, or hopeless? • During the past month, have you often been bothered by having little interest or pleasure in doing things? A third question should be considered if the woman answers "yes" to either of the initial questions: • Is this something you feel you need or want help with? Health care professionals may consider the use of self-report measures such as the Edinburgh Postnatal Depression Scale (EPDS), the Hospital Anxiety and Depression Scale (HADS), or the Patient Health Questionnaire-9 (PHQ-9) as part of a subsequent assessment or for the routine monitoring of outcomes.	April 2007

Potential Harms of Screening for Postpartum Depression

In their 2009 recommendations on screening for depression in adults, the U.S. Preventive Services Task Force (USPSTF) identified "false-positive results, the inconvenience of additional diagnostic workup, the costs and adverse effects of treatment of patients who are incorrectly identified as being depressed, and potential adverse effects of labeling" as potential harms but found no evidence for any of these harms.[19] Whether any of these harms is more likely when screening for postpartum depression is unclear. However, it is possible that pregnant and postpartum women may be at increased risk of harm from screening, given that many of the signs and symptoms included in the diagnostic criteria for depression (Table 1) are common and normal responses to pregnancy, childbirth, and caring for infants. Furthermore, many studies of postpartum depression include "minor depression" as a diagnostic category. Previous reviews have concluded that there is a lack of evidence that treatment of symptoms not meeting criteria for major depression improves outcomes.[3,4,23] If a diagnosis of "minor depression" does not lead to effective treatment, then patients are exposed to the potential side effects of therapy (particularly medical therapy) in addition to being labeled as depressed without a concomitant improvement in health for themselves or their child. Finally, when comparing different strategies for screening women, differences in both false-positive and false-negative results are important, especially for women who might have been helped by earlier identification of depression through screening.

Accuracy of Screening Instruments for Postpartum Depression

In evaluating strategies involving screening for postpartum depression, patients, providers, and policymakers must consider the tradeoffs between the likely benefits and harms of screening. Although direct evidence from appropriately designed trials is ideal, such data are often lacking (and are lacking for screening for postpartum depression). In such cases, inferences must be drawn from data on how well the screening test or strategy distinguishes between patients who truly have the condition of interest and those who do not, which is usually reported as the strategy's sensitivity (the likelihood that people with the condition will have a positive test) and specificity (the likelihood that people without the condition will have a negative test). The sensitivity and specificity of a test are characteristics that are independent of the population being tested. Higher sensitivity means fewer people with the condition are missed, while higher specificity means fewer people without the condition will be falsely identified; importantly, sensitivity and specificity are indirectly correlated—increasing sensitivity decreases specificity and vice versa. One advantage of sensitivity and specificity is that, because they are characteristics of the tests themselves, sensitivity and specificity estimates of a given test can be compared and pooled across different studies.

Sensitivity and specificity are not, however, directly useful clinically; the more relevant test characteristics are the positive predictive value (PPV; the likelihood that a person with a positive test has the condition of interest) and the negative predictive value (NPV; the likelihood that a person with a negative test does not have the condition of interest). These characteristics are functions of test sensitivity and specificity *and* the underlying likelihood of the condition of interest (prevalence). Because of this dependence on prevalence, the PPV and NPV of a specific

test can vary across studies, depending on the population. The PPV and NPV of a test or strategy can be directly estimated from a study in a specific population or can be indirectly estimated from given estimates of the test sensitivity and specificity and the population prevalence. A test with a certain sensitivity and specificity might have a quite different PPV and NPV when used in different settings or at different times. Greater certainty about how the PPV and NPV vary across populations, settings, and timing would help in developing specific recommendations about when, whom, and how often to screen.

One of the consistent uncertainties identified in current recommendations is how well currently available tests and strategies for identifying women with, or at risk for, postpartum depression perform. For example, the committee opinion on screening for depression during and after pregnancy developed by the American Congress of Obstetricians and Gynecologists[18] lists seven different tests—the Edinburgh Postnatal Depression Scale (EPDS), the Postpartum Depression Screening Scale (PDSS), the Patient Health Questionnaire-9 (PHQ-9), the Beck Depression Inventory (BDI), the Beck Depression Inventory II (BDI-II), the Center for Epidemiologic Studies Depression Scale (CES-D), and the Zung Self-Rating Depression Scale (Zung SDS)—with wide ranges for the reported sensitivity and specificity, but it does not provide specific guidance on which test might be most appropriate in a particular setting.

Another issue is that sensitivity and specificity may also vary based on the definition of "disease." For example, the 2005 AHRQ Evidence Report on postpartum depression[3,4] found that the sensitivity of all instruments reviewed was greater for "mild" depression when compared with "major" depression." As noted above, if treatment of mild depression does not lead to improved outcomes, then this greater "sensitivity" does not translate into a better test.

Clinical and Socioeconomic Factors Affecting Risk for Postpartum Depression

Consistent risk factors for postpartum depression identified in the literature include a history of depression before pregnancy, depression or anxiety during pregnancy, experiencing stressful life events during pregnancy or the early postpartum period, and low levels of social support; maternal age, income, and parity may also affect risk.[24-28] Because the outcomes of screening for any condition are dependent on the likelihood of that condition at the time of screening, selective use of specific tools to screen women at higher risk for postpartum depression when one or more risk factors are present may be a viable strategy.

Other Factors Affecting Performance of Screening for Postpartum Depression

Timing

Many of the signs and symptoms that make up the diagnostic criteria for depression are also common physiological or emotional responses to pregnancy and caring for an infant, and their prevalence can vary depending on when the measurement is performed. The presence of similar signs/symptoms in women who have and do not have depression could affect the specificity, and thus the false-positive rate, of a given screening test. In addition, testing during the prenatal period is seeking either to identify current depression (which by definition would not be postpartum depression), or to identify women *at risk* for postpartum depression; the performance

of a test designed to identify patients at higher risk before they develop a condition is often quite different than the performance of a test designed to detect the condition itself.

Setting

Setting is inevitably related to timing; however, setting may have other effects on test performance. For example, the willingness of a woman to admit to symptoms of depression might vary depending on the setting—that is, her comfort level and familiarity with a provider or her concerns about being judged as a parent. Setting may also play a crucial role in determining whether women with a positive screening test result receive appropriate diagnostic and treatment services.

Provider

As with setting, the provider and the nature of his/her relationship with the patient may affect the willingness of the patient to admit to symptoms of depression. The provider's ability to appropriately administer a given screening tool may be affected by his/her training or the nature of his/her usual practice. Finally, as with setting, even if the sensitivity/specificity/predictive values of the test are unchanged, the ability of the provider to provide appropriate diagnosis and treatment to a patient with a positive test may vary based on available resources, skill and training of provider, or the context of the visit.

Effective Management of Positive Screening Tests for Postpartum Depression

Screening is often focused during pregnancy or the first 3 postpartum months in settings where care is provided to pregnant or postpartum women by providers such as obstetricians, family practitioners, or nurse-midwives. All of the existing recommendations for screening emphasize the need for systems or procedures to ensure that women identified as being at risk for postpartum depression receive appropriate diagnostic services, and, if a diagnosis of depression is confirmed, appropriate treatment (Table 2). Because the risk of postpartum depression extends throughout the first 12 months after delivery, maternal depression may affect outcomes for the infant, and settings where care is provided to the infant provide an opportunity for postpartum depression screening. Clinicians who provide care for infants have proposed the possibility of including screening for maternal depression as part of routine infant care,[11,20] but issues regarding scope of practice, legal liability, and appropriate referral remain challenges.[10]

Rationale for Evidence Review

Despite recognition that a) postpartum depression is common, b) it may have serious effects on both mothers and infants, and c) screening instruments are available, uncertainty about whether, when, and how to screen for postpartum depression remains, as seen in the various recommendations summarized in Table 2. Sources for this uncertainty include:

- Imprecision in the published sensitivity and specificity estimates for the various instruments at the time the recommendations were drafted. Incorporating additional data

published subsequently should add greater precision to these estimates by increasing the overall sample size and may make any differences between specific tests more apparent.

- Uncertainty about the ability of screening strategies to consistently identify the women most likely to benefit from available treatments and followup. For example, in populations at very low risk for postpartum depression, lower specificity would result in a low NPV and could result in a high absolute number of women referred for additional diagnostic evaluation.
- Lack of direct evidence of benefits from screening. For screening to be of benefit, the test has to be able to accurately distinguish between women likely to benefit from further evaluation and treatment and those at low risk for the condition of interest; women identified as being at higher risk of the condition have to be able to receive appropriate diagnostic services; and, for those definitively identified with the condition, effective treatment needs to be available. Our review will focus on the first two aspects of screening benefits. If we assume that women identified through screening whose symptoms meet the diagnostic criteria for depression are given effective treatments, then a study that randomized women to no screening versus screening, or to screening with two different instruments, would address the question of screening benefit, especially if the treatments were standardized. Addressing the question of which treatments are most effective would require a different design.
- Issues related to management of women with a positive screening result. Although all recommendations related to screening commented on the need for appropriate systems or mechanisms for managing women with a positive screening test, there is no mention of the possible harms, such as anxiety created by a positive screening test result or the potential stigma associated with a diagnosis of depression.

A preliminary search of the literature indexed in PubMed® from 2005 forward using search terms similar to those used in the 2005 AHRQ Evidence Report[3,4] identified between 1,000 and 1,500 articles, suggesting that there may be sufficient additional data available to refine estimates of sensitivity and specificity for greater precision.

II. The Key Questions

The draft Key Questions (KQs) developed during Topic Refinement were available for public comment from November 8 to December 6, 2011. The comments received reinforced the uncertainties about screening tools as discussed with the Key Informants and reflected in the draft KQs. The comments did not lead to significant changes but were helpful in identifying additional factors of interest in KQ 2 and for clarifying the wording of the questions.

Based on the public comments and subsequent discussions with AHRQ, the following changes of note were made to the KQs:

- KQ 2: Change a reference to "*patient* factors" to "*individual* factors." Include cultural factors and history of intimate partner violence as potential factors that may affect baseline risk of postpartum depression.
- KQ 3a: Explicitly indicate that frequency of screening is a factor under consideration.

- KQ 3c: Explicitly indicate that family practitioners are included among the types of providers under consideration.

Additional comments received and considered, but not incorporated into the project plan, included recommendations to:

- consider comparative effectiveness of treatment strategies;
- expand the population to include screening for postpartum depression in fathers/partners;
- expand the population to include depression in pregnant women; and
- include an assessment of cost-effectiveness.

Although of interest, these suggestions are beyond the scope of this review.

The revised KQs are as follows:

Question 1

This question has two parts:

a. What are the sensitivity and specificity of currently available screening instruments for detecting postpartum depression, and how do these translate into the likelihood of false-negative and false-positive results in different populations and settings?
b. Are there clinically relevant differences in the ability of currently available screening instruments to correctly identify specific signs or symptoms of depression (e.g., suicidal ideation)?

Question 2

This question has two parts:

a. Are there individual factors (age, race, parity, history of mood disorders, history of intimate partner violence, perinatal outcomes, cultural factors) that affect the baseline risk of postpartum depression and, therefore, the subsequent positive and negative predictive values of screening instruments?
b. Are there validated predictive models or algorithms based on such factors that would improve the performance of screening instruments?

Question 3

Are the performance characteristics (sensitivity, specificity, predictive values) of screening instruments affected by:

a. Timing (prenatal, peripartum, or at various times in the first postpartum year) and frequency of screening?

b. Setting (prenatal visit, hospital/birthing center/home, postpartum maternal visit, or well-child visit)?

c. Provider (obstetrician, midwife, pediatrician, family practitioner, other health provider)?

Question 4

What are the comparative benefits of screening for postpartum depression when compared to no screening, or between different screening strategies (based on choice of screening instrument, timing, setting, etc.)?

Question 5

What are the comparative harms of screening for postpartum depression when compared to no screening, or between different screening strategies (based on choice of screening instrument, timing, setting, etc.)?

Question 6

Is the likelihood of an appropriate action (referral, diagnosis, treatment, etc.) after a positive screening result affected by timing, setting, patient characteristics, or other factors?

PICOTS (Population, Intervention, Comparator, Outcome, Timing, and Setting)

- **Population:**

  Pregnant women (although outcomes are focused on mothers and infants/children after delivery, some screening strategies may be applied during the prenatal period) and women during the first 12 months after delivery. Subgroups of potential interest include pregnant and postpartum women who differ by race/ethnicity, income, parity, cultural norms, history of mood disorders, perinatal outcomes, and history of intimate partner violence.

- **Interventions:**

  o Screening for depression through:

  ▪ Identification of risk factors for women at increased risk for postpartum depression (e.g., age, previous history of mental illness, history of intimate partner violence, or adverse perinatal outcome), followed by screening with a validated instrument (KQs 4–6). Validation is defined as documentation of standard psychometric properties of reliability and validity, with the reference standard for validity including, but not necessarily limited to, either a clinical assessment by a mental health professional based on criteria from the *DSM-IV-TR*, the Research Diagnostic Criteria (RDC), the Bedford College checklist, or the International Classification of Diseases (ICD); or a research-based diagnosis obtained by a 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).[3] This identification of risk factors can occur at various times throughout pregnancy and the first 12 postpartum months, followed by a defined action based on results of the screening test:

- – If the screening test results are positive, referral for diagnosis and treatment (if diagnosis confirms depression) or a diagnostic evaluation and treatment of confirmed depression by the same professional who conducted the screening
- – If the screening test results are negative, usual prenatal/postpartum care

- ■ Screening all pregnant/postpartum women using a validated instrument (as described above) at various times throughout pregnancy and the first 12 postpartum months in settings related to prenatal, peripartum, and pediatric care, followed by a defined action based on the results of the screening test (KQs 1–3):

  - – If the screening test results are positive, referral for diagnosis and treatment (if diagnosis confirms depression) or a diagnostic evaluation and treatment of confirmed depression by the same professional who conducted the screening
  - – If the screening test results are negative, usual prenatal/postpartum care

- • **Comparator:**

  - o No formal protocol for screening at any time during pregnancy or the first 12 postpartum months, screening with another validated instrument, or screening with the same instrument under different conditions (e.g., different settings or different timing)

- • **Outcomes:**

  - o Performance characteristics, using standard diagnostic (not screening) instruments for depression as a reference standard. Potential reference standards include, but are not necessarily limited to, either a clinical assessment by a mental health professional based on criteria from the DSM-IV-TR, the Research Diagnostic Criteria (RDC), the Bedford College Checklist, or the International Classification of Diseases (ICD); or a research-based diagnosis obtained by a 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)[3]:

    - ■ KQs 1–3:

      - – Sensitivity
      - – Specificity
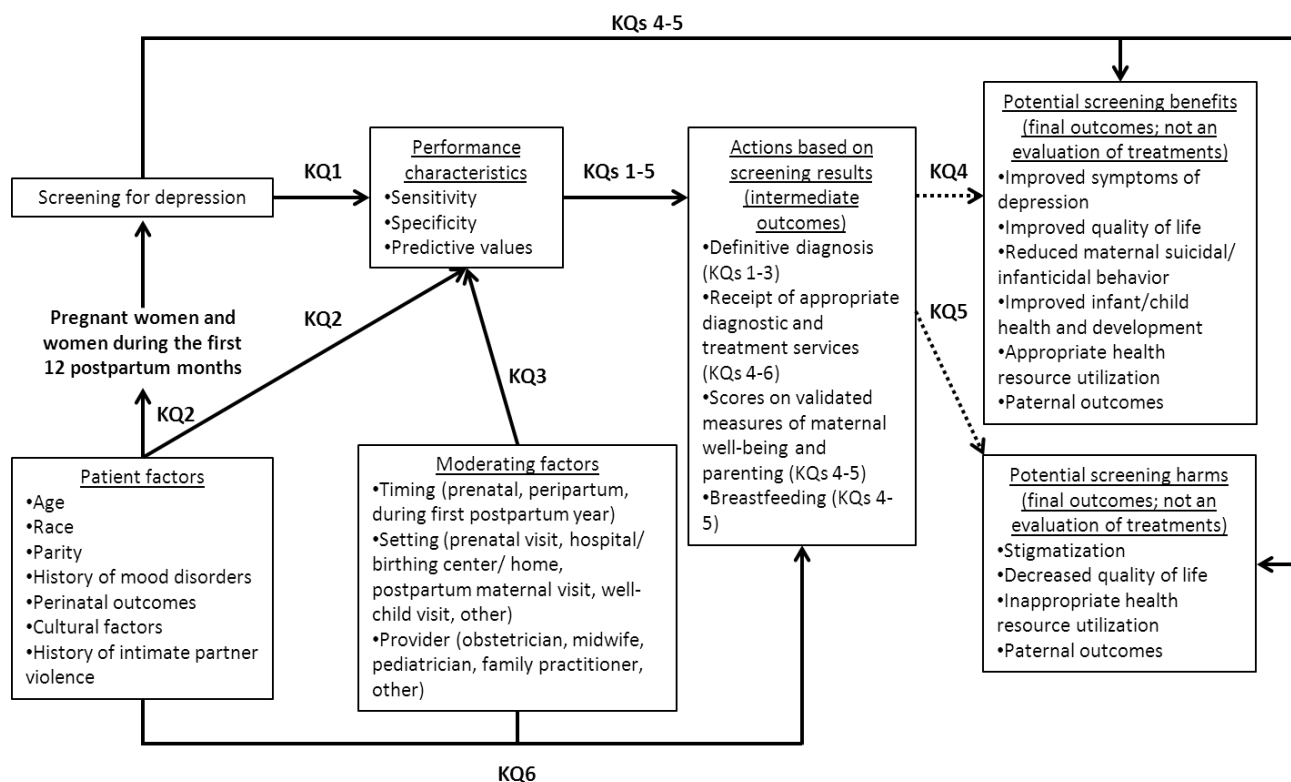   - – Predictive values

- o Intermediate outcomes

  - ▪ KQs 1–3:

    - – Confirmed diagnosis of depression based on the *DSM-IV-TR* criteria using a validated instrument, such as a clinical assessment by a mental health professional based on criteria from the *DSM-IV-TR*, the Research Diagnostic Criteria (RDC), the Bedford College Checklist, or the International Classification of Diseases (ICD); or a research-based diagnosis obtained by a 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).

  - ▪ KQs 4 and 5:

    - – Receipt of appropriate diagnostic and treatment services for symptoms of depression
    - – Scores on validated measures of maternal well-being and parenting
    - – Breastfeeding

  - ▪ KQ 6:

    - – Receipt of appropriate diagnostic and treatment services for symptoms of depression

- o Final outcomes

  - ▪ KQ 4:

    - – Scores on validated diagnostic instruments for depression
    - – Health-related quality of life, based on validated measures such as the Medical Outcomes Study 36-Item Short-Form Health Survey (SF-36)
    - – Maternal suicidal/infanticidal behavior
    - – Scores on validated instruments of infant health and development, including, but not necessarily limited to, the Bayley Scales of Infant Development (BSID)
    - – Maternal health-system resource utilization, including number of visits and estimates of total and attributable costs
    - – Infant health-system resource utilization, including number of visits and estimates of total and attributable costs
    - – Paternal outcomes, including scores on validated mental health instruments, health-related quality of life, and health-system resource utilization (measured as described above for maternal outcomes)

- o Adverse effects (harms) of intervention(s)

    - KQ 5:

        – Scores on validated measures of stigmatization, including but not necessarily limited to, the Internalized Stigma of Mental Illness (ISMI) scale[29]
        – Health-related quality of life, based on validated measures such as the SF-36
        – Maternal health-system resource utilization, including number of visits and estimates of total and attributable costs
        – Infant health-system resource utilization, including number of visits and estimates of total and attributable costs
        – Paternal outcomes, including scores on validated mental health instruments, health-related quality of life, and health-system resource utilization (measured as described above for maternal outcomes)

- **Timing:**

    - o Intervention:

        - Prenatal period
        - Immediate postpartum period (up to 6 weeks after delivery)
        - Up to 12 months after delivery

    - o Outcomes:

        - First 12 months after delivery (as listed above under Outcomes)

            – Mother
            – Infant
            – Father

        - Longer term (as listed above under Outcomes)

            – Mother
            – Child
            – Father

- **Setting:**

    - o Settings:

        - Prenatal care
        - Hospital

- Birthing center
- Home delivery
- Short-term postpartum followup
- Well-child visit
- Other

o Other providers:

- Obstetricians
- Family practitioners
- Nurse-midwives
- Mental health professionals
- Other health care providers (e.g., lactation consultants, social workers, behavioral health specialists)

III. Analytic Framework

Figure: Provisional analytic framework for efficacy and safety of screening for postpartum depression

Abbreviation: KQ = key question

IV. Methods

In developing this comprehensive review, we will apply the rules of evidence and evaluation of strength of evidence recommended by the AHRQ Evidence-based Practice Center (EPC) Program in its *Methods Guide for Effectiveness and Comparative Effectiveness Reviews*[30] and draft *Methods Guide for Medical Test Reviews*[31] (hereafter referred to as the *Methods Guides*). We will solicit feedback about conduct of the work (such as development of search strategies) from the Task Order Officer (TOO) and the Technical Expert Panel (TEP). We will follow the recommendations in the *Methods Guides* for literature search strategies, inclusion/exclusion of studies in our review, abstract screening, data abstraction and management, assessment of methodological quality of individual studies, data synthesis, and grading of evidence for each KQ.

A. Criteria for Inclusion/Exclusion of Studies in the Review

We will use the following inclusion/exclusion criteria for studies in our systematic review.

Table 3. Inclusion and exclusion criteria

Study Characteristic	Inclusion Criteria	Exclusion Criteria
Populations	- Pregnant women and women up to 12 months postpartum - Subgroups of potential interest include: - Race/ethnicity - Income - Parity - Cultural norms - History of mood disorders - Perinatal outcomes - History of intimate partner violence	- Women currently undergoing treatment for depression - Studies where the primary objective is to detect depression during pregnancy (rather than identify risk factors for postpartum depression) - Studies exclusively addressing bipolar disorder, a primary psychotic disorder, or maternity blues; or studies that include these populations and do not report results for subjects not fitting these subgroups separately

Study Characteristic	Inclusion Criteria	Exclusion Criteria

Interventions	• Screening using a validated screening instrument for depression, including, *but not necessarily limited to*:
	  o Bromley Postnatal Depression Scale (BPDS)
	  o Edinburgh Postnatal Depression Scale (EPDS)
	  o Postpartum Depression Screening Scale (PDSS)
	  o Leverton Questionnaire (LQ)
	  o Center for Epidemiologic Studies Depression Scale (CES-D)
	  o Hospital Anxiety and Depression Scale (HADS)
	  o Patient Health Questionnaire-9 (PHQ-9)
	  o Beck Depression Inventory (BDI IA, II)
	  o Zung Self-Rating Depression Scale (Zung SDS)
	  o Hamilton Rating Scale for Depression (HRSD)
	  o Postpartum Depression Predictors Inventory–Revised (PDPI-R)
	  o General Health Questionnaire (GHQ-D)
	  o Montgomery Asberg Depression Rating Scale (MADRS)
	  o Generalized Contentment Scale
	  o Patient Health Questionnaire-2 (PHQ-2)
	  o Primary Care Evaluation of Mental Disorders Patient Health Questionnaire (PRIME-MD PHQ)
		• Validation studies, or screening conducted using a nonvalidated instrument

Comparators	• No formal protocol for screening, screening with another validated instrument, or screening with the same instrument under different conditions (e.g., different settings or different timing)	• Comparison to screening with a nonvalidated instrument

Outcomes	• Performance characteristics (KQs 1–3):
	  o Sensitivity
	  o Specificity
	  o Predictive values
	• Intermediate outcomes
	  o KQs 1–3:
	    ▪ Diagnosis of depression based on the *DSM-IV-TR* criteria using a validated instrument
	  o KQs 4 and 5:
	    ▪ Receipt of appropriate diagnostic and treatment services for symptoms of depression
	    ▪ Scores on validated measures of maternal well-being and parenting
	    ▪ Breastfeeding
	  o KQ 6:
	    ▪ Receipt of appropriate diagnostic and treatment services for symptoms of depression
	• Final outcomes (KQ 4):
	  o Scores on validated diagnostic instruments for depression
		None

Study Characteristic	Inclusion Criteria	Exclusion Criteria
	o Health-related quality of life, based on validated measures o Maternal suicidal/infanticidal behaviors o Scores on validated instruments of infant health and development o Maternal health-system resource utilization, including number of visits and estimates of total and attributable costs o Infant health-system resource utilization, including number of visits and estimates of total and attributable costs o Paternal outcomes, including scores on validated mental health instruments, health-related quality of life, and health-system resource utilization (measured as described above for maternal outcomes) • Adverse effects (KQ 5): o Scores on validated measures of stigmatization o Health-related quality of life, based on validated measures o Maternal health-system resource utilization, including number of visits and estimates of total and attributable costs o Infant health-system resource utilization, including number of visits and estimates of total and attributable costs o Paternal outcomes, including scores on validated mental health instruments, health-related quality of life, and health-system resource utilization (measured as described above for maternal outcomes)	
Timing	• Intervention o Prenatal period o Immediate postpartum period (up to 6 weeks after delivery) o Up to 12 months after delivery • Followup o Begins at delivery; timing of followup will not be limited[a]	• Predelivery outcomes
Setting	• Any clinical provider setting, home • Studies conducted in a high-income economy as defined by the World Bank.[32] We restrict the study to economically developed countries—countries that have greater cultural and health care system similarities to the United States—to improve applicability of the study results to U.S. populations.	None

Study Characteristic	Inclusion Criteria	Exclusion Criteria
Study design	• Original data • Randomized trials, prospective and retrospective observational studies with comparator; for test characteristics, cross-sectional studies are acceptable if they include patients with diagnostic uncertainty and direct comparison of test results with an appropriate reference standard • Randomized controlled trials: All sample sizes • Observational studies: sample size ≥100 subjects	• Editorials, nonsystematic reviews, letters, case series, case reports
Publications	• English-language only • Peer-reviewed articles • Relevant systematic review, meta-analysis, or methods article (to be used for background only)	Given the high volume of literature available in English-language publications, the focus of our review on applicability to populations in the United States, and the scope of our current KQs, non–English-language articles will be excluded.[b]

[a]For all included studies, we will indicate the total number of participants enrolled and longest length (weeks or months) of followup, if relevant.

[b]It is the opinion of the investigators that the resources required to translate non–English-language articles would not be justified by the low potential likelihood of identifying relevant data unavailable from English-language sources. We will monitor the number of articles excluded at the abstract stage for English language and determine whether this exclusion criterion should be revisited.[33]

Abbreviations: BDI-IA = Beck Depression Inventory-IA; BDI-II = Beck Depression Inventory-II; BPDS = Bromley Postnatal Depression Scale; CES-D = Center for Epidemiologic Studies Depression Scale; DSM-IV-TR = Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, Text Revision; EPDS = Edinburgh Postnatal Depression Scale; GHQ-D = General Health Questionnaire; HADS = Hospital Anxiety and Depression Scale; HRSD = Hamilton Rating Scale for Depression; KQ = key question; LQ = Leverton Questionnaire; MADRS = Montgomery Asberg Depression Rating Scale; PDPI-R = Postpartum Depression Predictors Inventory-Revised; PDSS = Postpartum Depression Screening Scale; PHQ-2 = Patient Health Questionnaire-2; PHQ-9 = Patient Health Questionnaire-9; PRIME-MD PHQ = Primary Care Evaluation of Mental Disorders Patient Health Questionnaire; Zung SDS = Zung Self-Rating Depression Scale

B. Searching for the Evidence: Literature Search Strategies for Identification of Relevant Studies To Answer the Key Questions

To identify the relevant published literature, we will search PubMed®, EMBASE®, PsycINFO®, and the Cochrane Database of Systematic Reviews, starting with articles published subsequent to the March 2004 search end date of the 2005 AHRQ Evidence Report on postpartum depression. Where possible, we will use existing validated search filters (such as the Clinical Queries Filters in PubMed®). An experienced search librarian will guide all searches. Our proposed search strategy for PubMed is included in Appendix A; this strategy will be adapted as necessary for use in the other databases. We will supplement the electronic searches with a manual search of citations from a set of key primary and review articles. The reference list for identified pivotal articles will be manually hand-searched and cross-referenced against our library, and additional manuscripts will be retrieved. All citations will be imported into an electronic database (EndNote® Version X4; Thomson Reuters, Philadelphia, PA). As a mechanism to ascertain publication bias, we will search ClinicalTrials.gov to identify completed but unpublished studies. While the draft report is under peer review, we will update the search

and include any eligible studies determined either during that search or identified by peer or public reviewers in the final report.

We will use several approaches to identify relevant gray literature, including requesting Scientific Information Packets from identified publishers of proprietary depression screening tools among those listed in Appendix B. We will also search the gray literature of study registries and conference abstracts. Results of this search will be used to identify additional relevant publications for completed studies of interest. Gray literature databases will include ClinicalTrials.gov, the World Health Organization (WHO) International Clinical Trials Registry Platform Search Portal, and the ProQuest COS Conference Papers Index.

For searches conducted in PubMed, EMBASE, PsycINFO, and the Cochrane Database of Systematic Reviews, two reviewers will screen the titles and abstracts of the results for potential relevance to the research questions using prespecified inclusion/exclusion criteria. For the gray literature searches, either one or two reviewers (depending on literature volume) will perform an initial screen to identify studies of interest. Full publications associated with these studies of interest will be screened by two reviewers in the same fashion as described above. Articles included by either reviewer will undergo full-text screening. At the full-text screening stage, two independent reviewers must agree on a final inclusion/exclusion decision. Articles meeting eligibility criteria will be included for data abstraction. All results will be tracked in the DistillerSR data synthesis software program (Evidence Partners Inc., Manotick, ON, Canada).

C. Data Abstraction and Data Management

The research team will create data abstraction forms for the KQs that will be programmed in the DistillerSR software. Based on clinical and methodological expertise, a pair of researchers will be assigned to abstract data from each of the eligible articles. One researcher will abstract the data, and the second will over-read the article and the accompanying abstraction to check for accuracy and completeness. Disagreements will be resolved by consensus or by obtaining a third reviewer's opinion if consensus cannot be reached. Guidance documents will be drafted and provided to the researchers to aid both reproducibility and standardization of data collection.

Data abstraction forms will be designed to collect the data required to evaluate the specified eligibility criteria for inclusion in this review, as well as demographic and other data needed for determining outcomes (performance characteristics as well as intermediate, final, and adverse event outcomes). We will pay particular attention to describing the details of the screening intervention including setting, provider, and timing and frequency of screening; patient characteristics (e.g., age, parity); and study design (e.g., randomized controlled trial [RCT] versus observational study) that may be related to outcomes. In addition, we will describe comparators carefully as treatment standards may have changed during the study period. Harms outcomes will be framed to help identify adverse events (e.g., stigmatization, decreased quality of life). Data necessary for assessing quality and applicability, as described in the *Methods Guides*, will also be abstracted. Before they are used, abstraction form templates will be pilot-tested with a sample of included articles to ensure that all relevant data elements are captured and that there is consistency/reproducibility between abstractors. Forms will be revised as necessary before full abstraction of all included articles.

D. Assessment of Methodological Quality of Individual Studies

We will assess the methodological quality, or risk of bias, for each individual study by using the assessment instruments detailed by the AHRQ EPC Program's *Methods Guides*.[30,31] Briefly, we will rate each study as being of good, fair, or poor quality based on its adherence to well-accepted standard methodologies (e.g., QUADAS-2[34] for studies of diagnostic accuracy). For all studies, the overall study quality will be assessed as follows:

- **Good** (low risk of bias). These studies had the least bias, and the results were considered valid. These studies adhered to the commonly held concepts of high quality, including the following: a clear description of the population, setting, approaches, and comparison groups; appropriate measurement of outcomes; appropriate statistical and analytic methods and reporting; no reporting errors; a low dropout rate; and clear reporting of dropouts.
- **Fair**. These studies were susceptible to some bias, but not enough to invalidate the results. They did not meet all the criteria required for a rating of good quality because they had some deficiencies, but no flaw was likely to cause major bias. The study may have been missing information, making it difficult to assess limitations and potential problems.
- **Poor** (high risk of bias). These studies had significant flaws that might have invalidated the results. They had serious errors in design, analysis, or reporting; large amounts of missing information; or discrepancies in reporting.

The grading will be outcome-specific such that a given study that analyzes its primary outcome well but did an incomplete analysis of a secondary outcome would be assigned a different quality grade for each of the two outcomes. Studies of different designs will be graded within the context of their respective designs. Thus, RCTs will be graded good, fair, or poor, and observational studies will separately be graded good, fair, or poor.

E. Data Synthesis

We will begin by summarizing key features of the included studies for each KQ. To the degree that data are available, we will abstract information on study design; participant characteristics; clinical settings; interventions; and intermediate, final, and adverse event outcomes.

We will then determine the feasibility of completing a quantitative synthesis (i.e., meta-analysis). Feasibility depends on the volume of relevant literature, conceptual homogeneity of the studies, and completeness of results reporting. When a meta-analysis is appropriate, we will use random-effects models to quantitatively synthesize the available evidence. We will test for heterogeneity using graphical displays and test statistics (Q and I^2 statistics), while recognizing that the ability of statistical methods to detect heterogeneity may be limited. For comparison, we will also perform fixed-effect meta-analyses. We will present summary estimates, standard errors, and confidence intervals. We anticipate that intervention effects may be heterogeneous. We hypothesize that the methodological quality of individual studies, study type, characteristics of the screening population (e.g., age, parity), and characteristics of the screening intervention (e.g., setting, provider) will impact intervention effects. If there are sufficient studies, we will

perform subgroup analyses and/or meta-regression analyses to examine these hypotheses. An example of such a subgroup analysis would be a comparison of effectiveness estimates for RCTs versus observational studies or a comparison of estimates of the association between a history of intimate partner violence and postpartum depression for cohort versus case-control studies.

We will also use an existing simulation model of pregnancy and neonatal outcomes[35] to estimate the balance of benefits and harms of different strategies based on the literature review, using the benefits and harms listed above. Because there are numerous unresolved issues about the use of quality adjusted life years in the setting of maternal-child health,[36] we will use the estimated likelihood of specific outcomes as the model output. Specific benefits include estimates of treated depression, maternal and infant health care visits, and other outcomes for which our review identifies evidence. Based on our preliminary review and discussions with the Key Informants (KIs) and the TEP, data on harms, in particular, are likely to be sparse. We can, however, readily estimate the number of false-positive screening test results, or total referrals for further evaluation, under different scenarios. This allows an approach that compares total tests or false-positive results as a measure of "cost" or "harm" to a measure of benefit, such as "cases of depression detected." Such an approach has been used by modelers supporting the USPSTF in making recommendations—for example, in colorectal cancer screening, where the metric was colonoscopies per cancer death prevented, or in cervical cancer screening, where the metric was colposcopies per cancer death prevented.

The model simulates pregnancy from conception through delivery and can subsequently simulate both maternal and child outcomes. Child outcomes are conditioned on gestational age at delivery and maternal race/ethnicity; both maternal and child outcomes can also easily be conditioned on maternal exposures at any point in gestation. In this context, using this model, estimates of benefits and harms can be generated for specific screening tests, at different times during and after pregnancy, for mothers and infants (and for fathers, if data are available). For example, the model could compare estimated maternal and infant outcomes from screening with a test of sensitivity X% and specificity Y% at 36 weeks gestation and 6 weeks postpartum, versus screening with a test of sensitivity A% and specificity B% at each well-child visit. The values for sensitivity and specificity (along with confidence intervals) will be derived from the literature review. The model will also incorporate variability in followup and appropriate treatment after a positive screening test result. We will use probabilistic sensitivity analysis to assess overall uncertainty based on the available literature, and use a modified value-of-information approach to help prioritize future research needs.[37] Outcomes included in the model, in addition to those discussed above (such as false-positive results, or number of health care encounters), will be those for which there is sufficient evidence identified in the literature review, and which can be meaningfully incorporated into a model; for example, although there may be valid evidence on health-related quality of life, these data may not be readily translatable into quality-adjusted life expectancy.

F. Grading the Evidence for Each Key Question

We will grade the strength of evidence for each outcome assessed across studies. The strength of evidence will be assessed by using the approach described in the *Methods Guides*.[30,31,38] In brief, the approach requires assessment of four domains: risk of bias, consistency, directness, and precision. Additional domains are to be used when appropriate:

coherence, dose-response association, impact of plausible residual confounders, strength of association (magnitude of effect), and publication bias. These domains will be considered qualitatively, and a summary rating of high, moderate, or low strength of evidence will be assigned after discussion between two reviewers. In some cases, high, moderate, or low ratings will be impossible or imprudent to make, for example, when no evidence is available or when evidence on the outcome is too weak, sparse, or inconsistent to permit any conclusion to be drawn. In these situations, a grade of insufficient will be assigned. This four-level rating scale is defined as follows:

- High—High confidence that the evidence reflects the true effect. Further research is very unlikely to change our confidence in the estimate of effect.
- Moderate—Moderate confidence that the evidence reflects the true effect. Further research may change our confidence in the estimate of effect and may change the estimate.
- Low—Low confidence that the evidence reflects the true effect. Further research is likely to change the confidence in the estimate of effect and is likely to change the estimate.
- Insufficient—Evidence either is unavailable or does not permit estimation of an effect.

G. Assessing Applicability

We will assess applicability across our KQs using the method described in the *Methods Guides*.[30,31,39] In brief, this latter method uses the PICOTS (Population, Intervention, Comparator, Outcome, Timing, and Setting) format as a way to organize information relevant to applicability. Items of particular interest that may contribute to heterogeneity and impact applicability include setting (e.g., country, provider), comparator, spectrum of disease (e.g., screening population or preselected group), patient income, race, ethnicity, parity, and partner support. We will use a checklist to guide the assessment of applicability. We will use these data to evaluate the applicability to clinical practice, paying special attention to study eligibility criteria, demographic features of the enrolled population in comparison to the target population, characteristics of the intervention used in comparison with care models currently in use, and clinical relevance and timing of the outcome measures. We will summarize issues of applicability qualitatively.

V. References

1. American Psychiatric Association. Diagnostic and statistical manual of mental Ddsorders, 4th Edition, Text Revision (DSM-IV-TR). Arlington, VA: American Psychiatric Association; 2000.
2. Jones I. DSMV: the perinatal onset specifier for mood disorders. Arlington, VA: American Psychiatric Association; 2010. Available at: http://www.dsm5.org/Documents/Mood%20Disorders%20Work%20Group/Ian%20Jones%20memo-post-partum.pdf. Accessed August 10, 2011.
3. Gaynes BN, Gavin N, Meltzer-Brody S, et al. Perinatal Depression: Prevalence, Screening Accuracy, and Screening Outcomes. Evidence Report/Technology Assessment No. 119 (Prepared by RTI–University of North Carolina Evidence-based Practice Center under

Contract No. 290-02-0016). Rockville, MD: Agency for Healthcare Research and Quality; February 2005. AHRQ Publication No. 05-E006-2. Available at: http://www.ahrq.gov/downloads/pub/evidence/pdf/peridepr/peridep.pdf. Accessed August 10, 2011.

4. Gaynes BN, Gavin N, Meltzer-Brody S, et al. Perinatal depression: prevalence, screening accuracy, and screening outcomes. Evid Rep Technol Assess (Summ) 2005 Feb;(119):1-8. PMID: 15760246.

5. Ng RC, Hirata CK, Yeung W, et al. Pharmacologic treatment for postpartum depression: a systematic review. Pharmacotherapy 2010;30(9):928-41. PMID: 20795848.

6. Appleby L. Suicide during pregnancy and in the first postnatal year. BMJ. 1991;302(6769):137-40. PMID: 1995132.

7. Lindahl V, Pearson JL, Colpe L. Prevalence of suicidality during pregnancy and the postpartum. Arch Women Ment Health 2005;8(2):77-87. PMID: 15883651.

8. Oates M. Perinatal psychiatric disorders: a leading cause of maternal morbidity and mortality. Br Med Bull 2003;67:219-29. PMID: 14711766.

9. Spinelli MG. Maternal infanticide associated with mental illness: prevention and the promise of saved lives. Am J Psychiatry 2004;161(9):1548-57. PMID: 15337641.

10. Chaudron LH, Szilagyi PG, Campbell AT, et al. Legal and ethical considerations: risks and benefits of postpartum depression screening at well-child visits. Pediatrics 2007;119(1):123-8. PMID: 17200279.

11. Gjerdingen DK, Yawn BP. Postpartum depression screening: importance, methods, barriers, and recommendations for practice. J Am Board Fam Med 2007;20(3):280-8. PMID: 17478661.

12. McLearn KT, Minkovitz CS, Strobino DM, et al. Maternal depressive symptoms at 2 to 4 months post partum and early parenting practices. Arch Pediatr Adolesc Med 2006;160(3):279-84. PMID: 16520447.

13. Paris R, Bolton RE, Weinberg MK. Postpartum depression, suicidality, and mother-infant interactions. Arch Womens Ment Health 2009;12(5):309-21. PMID: 19728036.

14. Petrou S, Cooper P, Murray L, et al. Economic costs of post-natal depression in a high-risk British cohort. Br J Psychiatry 2002;181:505-12. PMID: 12456521.

15. Sills MR, Shetterly S, Xu S, et al. Association between parental depression and children's health care use. Pediatrics. 2007;119(4):e829-36. PMID: 17403826.

16. Anderson LN, Campbell MK, daSilva O, et al. Effect of maternal depression and anxiety on use of health services for infants. Can Fam Physician 2008;54(12):1718-9.e5. PMID: 19074718.

17. Hewitt C, Gilbody S, Brealey S, et al. Methods to identify postnatal depression in primary care: an integrated evidence synthesis and value of information analysis. Health Technol Assess 2009 Jul;13(36):1-145, 147-230. PMID: 19624978.

18. American College of Obstetricians and Gynecologists, Committee on Obstetric Practice. Committee opinion no. 453: screening for depression during and after pregnancy. Obstet Gynecol 2010;115(2 Pt 1):394-5. PMID: 20093921.

19. U.S. Preventive Services Task Force. Screening for depression in adults: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med 2009;151(11):784-92. PMID: 19949144.

20. Earls MF; Committee on Psychosocial Aspects of Child and Family Health, American Academy of Pediatrics. Incorporating recognition and management of perinatal and postpartum depression into pediatric practice. Pediatrics 2010;126(5):1032-9. PMID: 20974776.

21. American College of Nurse Midwives, Division of Women's Health Policy and Leadership. Position statement: depression in women. Approved March 2002. Reviewed December 2003. Available at: www.midwife.org/siteFiles/position/Depression_in_Women_05.pdf. Accessed August 10, 2011.

22. National Institute for Health and Clinical Excellence. Antenatal and Postnatal Mental Health: Clinical Management and Service Guidance. NICE Clinical Guideline 45. London: National Institute for Health and Clinical Excellence; February 2007. Available at: http://guidance.nice.org.uk/CG45/NICEGuidance/pdf/English. Accessed August 10, 2011.

23. Barbui C, Cipriani A, Patel V, et al. Efficacy of antidepressants and benzodiazepines in minor depression: systematic review and meta-analysis. Br J Psychiatry 2011;198(1):11-6. PMID: 21200071.

24. Eberhard-Gran M, Eskild A, Tambs K, et al. Depression in postpartum and non-postpartum women: prevalence and risk factors. Acta Psychiatr Scand 2002;106(6):426-33. PMID: 12392485.

25. Josefsson A, Angelsiöö L, Berg G, et al. Obstetric, somatic, and demographic risk factors for postpartum depressive symptoms. Obstet Gynecol 2002;99(2):223-8. PMID: 11814501.

26. McCoy SJB, Beal JM, Shipman SBM, et al. Risk factors for postpartum depression: a retrospective investigation at 4-weeks postnatal and a review of the literature. J Am Osteopath Assoc 2006;106(4):193-8. PMID: 16627773.

27. Robertson E, Grace S, Wallington T, et al. Antenatal risk factors for postpartum depression: a synthesis of recent literature. Gen Hosp Psychiatry 2004;26(4):289-95. PMID: 15234824.

28. Vesga-Lopez O, Blanco C, Keyes K, et al. Psychiatric disorders in pregnant and postpartum women in the United States. Arch Gen Psychiatry 2008;65(7):805-15. PMID: 18606953.

29. Van Brakel WH. Measuring health-related stigma—a literature review. Psychol Health Med 2006;11(3):307-34. PMID: 17130068.

30. Agency for Healthcare Research and Quality. Methods Guide for Effectiveness and Comparative Effectiveness Reviews. Rockville, MD: Agency for Healthcare Research and Quality; August 2011. AHRQ Publication No. 10(11)-EHC063-EF. Available at: http://www.effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productid=318. Accessed January 3, 2012.

31. Agency for Healthcare Research and Quality. Methods Guide for Medical Test Reviews. Rockville, MD: Agency for Healthcare Research and Quality; November 2010. Available at: http://effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productid=558. Accessed January 3, 2012.

32. The World Bank. Country and Lending Groups. Available at: http://data.worldbank.org/about/country-classifications/country-and-lending-groups. Accessed January 3, 2012.

33. Moher D, Fortin P, Jadad AR, et al. Completeness of reporting of trials published in languages other than English: implications for conduct and reporting of systematic reviews. Lancet 1996;347(8998):363-6. PMID: 8598702.

34. Whiting PF, Rutjes AWS, Westwood ME, et al; QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;155(8):529-36. PMID: 22007046.

35. Myers ER, Misurski DA, Swamy GK. Influence of timing of seasonal influenza vaccination on effectiveness and cost-effectiveness in pregnancy. Am J Obstet Gynecol 2011;204(6 Suppl 1):S128-40. PMID: 21640230.

36. Ungar WJ, ed. Economic evaluation in child health. New York: Oxford University Press; 2010.

37. Myers E, Sanders GD, Ravi D, et al. Evaluating the Potential Use of Modeling and Value-of-Information Analysis for Future Research Prioritization Within the Evidence-based Practice Center Program Methods Future Research Needs Report No. 5 (Prepared by the Duke Evidence-based Practice Center under Contract No. 290-2007-10066-I). Rockville, MD: Agency for Healthcare Research and Quality; June 2011. AHRQ Publication No. 11-EHC030-EF. Available at:http://www.effectivehealthcare.ahrq.gov/ehc/products/220/700/MFRN5_20111213.pdf. Accessed January 3, 2012.

38. Owens DK, Lohr KN, Atkins D, et al. AHRQ series paper 5: grading the strength of a body of evidence when comparing medical interventions—Agency for Healthcare Research and Quality and the Effective Health-Care Program. J Clin Epidemiol 2010;63(5):513-23. PMID: 19595577.

39. Atkins D, Chang SM, Gartlehner G, et al. Assessing applicability when comparing medical interventions: AHRQ and the Effective Health Care Program. J Clin Epidemiol 2011;64(11):1198-207. PMID: 21463926.

VI. Definition of Terms

AHRQ	Agency for Healthcare Research and Quality
BDI-IA	Beck Depression Inventory-IA
BDI-II	Beck Depression Inventory-II
BPDS	Bromley Postnatal Depression Scale
BSID	Bayley Scales of Infant Development
CER	Comparative Effectiveness Review
CES-D	Center for Epidemiologic Studies Depression Scale
DIS	Diagnostic Interview Schedule
DSM-IV-TR	*Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, Text Revision*
DSM-V	*Diagnostic and Statistical Manual of Mental Disorders, 5th Edition*
EPC	Evidence-based Practice Center
EPDS	Edinburgh Postnatal Depression Scale
GHQ-D	General Health Questionnaire
HADS	Hospital Anxiety and Depression Scale
HRSD	Hamilton Rating Scale for Depression
ICD	International Classification of Diseases
ISMI	Internalized Stigma of Mental Illness
KQ	key question

LQ	Leverton Questionnaire
MADRS	Montgomery Asberg Depression Rating Scale
NPV	negative predictive value
PDPI-R	Postpartum Depression Predictors Inventory-Revised
PDSS	Postpartum Depression Screening Scale
PHQ-2	Patient Health Questionnaire-2
PHQ-9	Patient Health Questionnaire-9
PICOTS	Population, Intervention, Comparator, Outcome, Timing, and Setting
PPV	positive predictive value
PRIME-MD PHQ	Primary Care Evaluation of Mental Disorders Patient Health Questionnaire
QUADAS	Quality Assessment of Diagnostic Accuracy Studies
RCT	randomized controlled trial
RDC	Research Diagnostic Criteria
SADS	Schedule for Affective Disorders and Schizophrenia
SCID	Structured Clinical Interview for Depression
SF-36	Medical Outcomes Study 36-Item Short-Form Health Survey
SPI	Goldberg's Standardized Psychiatric Interview
TEP	Technical Expert Panel
TOO	Task Order Officer
Zung SDS	Zung Self-Rating Depression Scale

VII. Summary of Protocol Amendments

In the event of protocol amendments, the date of each amendment will be accompanied by a description of the change and the rationale.

VIII. Review of Key Questions

For all EPC reviews, key questions were reviewed and refined as needed by the EPC with input from Key Informants and the Technical Expert Panel (TEP) to assure that the questions are specific and explicit about what information is being reviewed. In addition, for Comparative Effectiveness reviews, the key questions were posted for public comment and finalized by the EPC after review of the comments.

IX. Key Informants

Key Informants are the end-users of research, including patients and caregivers, practicing clinicians, relevant professional and consumer organizations, purchasers of health care, and others with experience in making health care decisions. Within the EPC program, the Key Informant role is to provide input into identifying the Key Questions for research that will inform healthcare decisions. The EPC solicits input from Key Informants when developing questions for systematic review or when identifying high-priority research gaps and needed new research. Key Informants are not involved in analyzing the evidence or writing the report and have not reviewed the report, except as given the opportunity to do so through the peer or public review mechanism.

Key Informants must disclose any financial conflicts of interest greater than $10,000 and any other relevant business or professional conflicts of interest. Because of their role as end-users, individuals are invited to serve as Key Informants and those who present with potential conflicts may be retained. The TOO and the EPC work to balance, manage, or mitigate any potential conflicts of interest identified.

X. Technical Experts

Technical Experts comprise a multidisciplinary group of clinical, content, and methodological experts who provide input in defining populations, interventions, comparisons, or outcomes as well as identifying particular studies or databases to search. They are selected to provide broad expertise and perspectives specific to the topic under development. Divergent and conflicted opinions are common and perceived as healthy scientific discourse that results in a thoughtful, relevant systematic review. Therefore study questions, design, and/or methodological approaches do not necessarily represent the views of individual technical and content experts. Technical Experts provide information to the EPC to identify literature search strategies and recommend approaches to specific issues as requested by the EPC. Technical Experts do not do analysis of any kind nor contribute to the writing of the report and have not reviewed the report, except as given the opportunity to do so through the public review mechanism.

Technical Experts must disclose any financial conflicts of interest greater than $10,000 and any other relevant business or professional conflicts of interest. Because of their unique clinical or content expertise, individuals are invited to serve as Technical Experts and those who present with potential conflicts may be retained. The TOO and the EPC work to balance, manage, or mitigate any potential conflicts of interest identified.

XI. Peer Reviewers

Peer reviewers are invited to provide written comments on the draft report based on their clinical, content, or methodological expertise. Peer review comments on the preliminary draft of the report are considered by the EPC in preparation of the final draft of the report. Peer reviewers do not participate in writing or editing of the final report or other products. The synthesis of the scientific literature presented in the final report does not necessarily represent the views of individual reviewers. The dispositions of the peer review comments are documented and will, for Comparative Effectiveness Reviews (CERs) and Technical briefs, be published 3 months after the publication of the Evidence report.

Potential Reviewers must disclose any financial conflicts of interest greater than $10,000 and any other relevant business or professional conflicts of interest. Invited Peer Reviewers may not have any financial conflict of interest greater than $10,000. Peer reviewers who disclose potential business or professional conflicts of interest may submit comments on draft reports through the public comment mechanism.

XII. EPC Team Disclosures

The EPC project team has no conflicts of interest to report.

XIII. Role of the Funder

This project was funded under Contract No. HHSA 290-2007-1066-I from the Agency for Healthcare Research and Quality, U.S. Department of Health and Human Services. The Task Order Officer reviewed contract deliverables for adherence to contract requirements and quality. The authors of this report are responsible for its content. Statements in the report should not be construed as endorsement by the Agency for Healthcare Research and Quality or the U.S. Department of Health and Human Services.

Appendix A. Proposed PubMed Literature Search Strategy

Set #	Terms	Results
#1	"Maternal Health Services"[Mesh] OR "Pregnancy"[Mesh] OR "Pregnant Women"[Mesh] OR Puerperal Disorders[Mesh] OR prenatal[tiab] OR perinatal[tiab] OR postnatal[tiab] OR pregnancy[tiab] OR pregnant[tiab] OR postpartum[tiab] OR post-partum[tiab]	810687
#2	Depression[Mesh] OR Depressive Disorder[Mesh] OR depression[tiab]	234185
#3	#1 AND #2	10492
#4	postpartum period/psychology[mesh] OR depression, postpartum[mesh]	3509
#5	#3 OR #4	11289
#6	postpartum depression/diagnosis[mesh] OR mass screening[mesh] OR questionnaires[mesh] OR Interviews as Topic[Mesh] OR Psychometrics[Mesh] OR Psychiatric Status Rating Scales[Mesh] OR questionnaire[tiab] OR questionnaires[tiab] OR screening[tiab] OR screen[tiab] OR scale[tiab] OR instrument[tiab] OR instruments[tiab] OR EPDS[tiab] OR "Edinburgh postnatal depression"[tiab] OR BDI[tiab] OR "beck depression inventory"[tiab] OR PDSS[tiab] OR "Postpartum Depression Screening Scale"[tiab] OR BPDS[tiab] OR "Bromley Postnatal Depression Scale"[tiab] OR LQ[tiab] OR "Leverton Questionnaire"[tiab] OR CES-D[tiab] OR "Center for Epidemiologic Studies Depression Scale"[tiab] OR HADS[tiab] OR "Hospital Anxiety and Depression Scale"[tiab] OR PHQ-9[tiab] OR "Patient Health Questionnaire-9"[tiab] OR "Zung SDS"[tiab] OR "Zung Self-Rating Depression Scale"[tiab] OR HRSD[tiab] OR "Hamilton Rating Scale for Depression"[tiab] OR PDPI-R[tiab] OR "Postpartum Depression Predictors Inventory-Revised"[tiab] OR GHQ-D[tiab] OR "General Health Questionnaire"[tiab] OR MADRS[tiab] OR "Montgomery Asberg Depression Rating Scale"[tiab] OR "generalized contentment scale"[tiab] OR "patient health questionnaire-2"[tiab] OR "phq-2"[tiab] OR "primary care evaluation of mental disorders patient health questionnaire"[tiab] OR "prime-md phq"[tiab]	1113231
#7	#5 AND #6	4018
#8	#7 NOT (Editorial[ptyp] OR Letter[ptyp] OR Case Reports[ptyp] OR Comment[ptyp]) NOT (animals[mh] NOT humans[mh])	3831
#9	#8 Limits: English, 2004 - present	2340

Appendix B. Screening Instruments

Screening instruments considered in this review will include, but are not limited to, the instruments named in the following list. Any additional screening instruments identified during the review will be assessed for validity against the criteria in the protocol and will be included if they meet those criteria.

- Bromley Postnatal Depression Scale (BPDS)
- Edinburgh Postnatal Depression Scale (EPDS)
- Postpartum Depression Screening Scale (PDSS)
- Leverton Questionnaire (LQ)
- Center for Epidemiologic Studies Depression Scale (CES-D)
- Hospital Anxiety and Depression Scale (HADS)
- Patient Health Questionnaire-9 (PHQ-9)
- Beck Depression Inventory (IA, II)
- Zung Self-Rating Depression Scale (Zung SDS)
- Hamilton Rating Scale for Depression (HRSD)
- Postpartum Depression Predictors Inventory-Revised (PDPI-R)
- General Health Questionnaire (GHQ-D)
- Montgomery Asberg Depression Rating Scale (MADRS)
- Generalized Contentment Scale
- Patient Health Questionnaire-2 (PHQ-2)
- Primary Care Evaluation of Mental Disorders Patient Health Questionnaire (PRIME-MD PHQ)