Review | **EVIDENCE REPORT FOR THE USPSTF**

Primary Care Screening for and Treatment of Depression in Pregnant and Postpartum Women
Evidence Report and Systematic Review for the US Preventive Services Task Force

Elizabeth O'Connor, PhD; Rebecca C. Rossom, MD, MSCR; Michelle Henninger, PhD; Holly C. Groom, MPH; Brittany U. Burda, MPH

IMPORTANCE Depression is a source of substantial burden for individuals and their families, including women during the pregnant and postpartum period.

OBJECTIVE To systematically review the benefits and harms of depression screening and treatment, and accuracy of selected screening instruments, for pregnant and postpartum women. Evidence for depression screening in adults in general is available in the full report.

DATA SOURCES MEDLINE, PubMed, PsycINFO, and the Cochrane Collaboration Registry of Controlled Trials through January 20, 2015; references; and government websites.

STUDY SELECTION English-language trials of benefits and harms of depression screening, depression treatment in pregnant and postpartum women with screen-detected depression, and diagnostic accuracy studies of depression screening instruments in pregnant and postpartum women.

DATA EXTRACTION AND SYNTHESIS Two investigators independently reviewed abstracts and full-text articles and extracted data from fair- and good-quality studies. Random-effects meta-analysis was used to estimate the benefit of cognitive behavioral therapy (CBT) in pregnant and postpartum women.

MAIN OUTCOMES AND MEASURES Depression remission, prevalence, symptoms, and related measures of depression recovery or response; sensitivity and specificity of selected screening measures to detect depression; and serious adverse effects of antidepressant treatment.

RESULTS Among pregnant and postpartum women 18 years and older, 6 trials (n = 11 869) showed 18% to 59% relative reductions with screening programs, or 2.1% to 9.1% absolute reductions, in the risk of depression at follow-up (3-5 months) after participation in programs involving depression screening, with or without additional treatment components, compared with usual care. Based on 23 studies (n = 5398), a cutoff of 13 on the English-language Edinburgh Postnatal Depression Scale demonstrated sensitivity ranging from 0.67 (95% CI, 0.18-0.96) to 1.00 (95% CI, 0.67-1.00) and specificity consistently 0.87 or higher. Data were sparse for Patient Health Questionnaire instruments. Pooled results for the benefit of CBT for pregnant and postpartum women with screen-detected depression showed an increase in the likelihood of remission (pooled relative risk, 1.34 [95% CI, 1.19-1.50]; No. of studies [K] = 10, I^2 = 7.9%) compared with usual care, with absolute increases ranging from 6.2% to 34.6%. Observational evidence showed that second-generation antidepressant use during pregnancy may be associated with small increases in the risks of potentially serious harms.

CONCLUSIONS AND RELEVANCE Direct and indirect evidence suggested that screening pregnant and postpartum women for depression may reduce depressive symptoms in women with depression and reduce the prevalence of depression in a given population. Evidence for pregnant women was sparser but was consistent with the evidence for postpartum women regarding the benefits of screening, the benefits of treatment, and screening instrument accuracy.

Supplemental content at jama.com

CME Quiz at jamanetworkcme.com and CME Questions page 412

Author Affiliations: Kaiser Permanente Research Affiliates Evidence-based Practice Center, Center for Health Research, Portland, Oregon (O'Connor, Henninger, Groom, Burda); Kaiser Permanente Research Affiliates Evidence-based Practice Center, HealthPartners Institute for Education and Research, Bloomington, Minnesota (Rossom).

Corresponding Author: Elizabeth O'Connor, PhD, Kaiser Permanente Research Affiliates Evidence-based Practice Center, Center for Health Research, Kaiser Permanente Northwest, 3800 N Interstate Ave, Portland, OR 97227 (elizabeth.oconnor@kpchr.org).

Major depressive disorder (MDD) is the leading cause of disease-related disability in women around the world.[1] In a study of US women assessed in 2005, 9.1% of pregnant women and 10.2% of postpartum women met criteria for a major depressive episode.[2] Maternal depression can affect offspring as well, leading to lower-quality interactions with the mother,[3] higher rates of emotional and behavioral problems, worse social competence with peers, and poorer adjustment to school.[4-6] In 2009, the US Preventive Services Task Force (USPSTF) recommended screening adults for depression when staff-assisted depression care supports are in place to ensure accurate diagnosis, effective treatment, and follow-up (B recommendation).[7] The USPSTF recommended against routinely screening adults for depression when such support is not in place but acknowledged there may be considerations that support screening for depression in an individual patient (C recommendation).[7] These recommendations were based on a combination of results from the 2002 USPSTF review,[8] which included very little evidence related to pregnant and postpartum women, and a targeted update published in 2009, which excluded studies limited to pregnant and postpartum women.[9] We undertook the current review to help the USPSTF update its recommendation on depression screening and expand it to include evidence related to pregnant and postpartum women.

Methods

Scope of Review
Detailed methods are available in the full evidence report at http://www.uspreventiveservicestaskforce.org/Page/Document/final-evidence-review144/depression-in-adults-screening1.[10] Evidence related to general and older adults was only minimally changed from the previous review and are also presented in the full report. In this article, the focus is on the direct and indirect evidence for depression screening of pregnant and postpartum women, where most new evidence was found. The analytic framework and key questions (KQs) to guide the portion of our review related to pregnant and postpartum women are shown in Figure 1.

Data Sources and Searches
An initial search was conducted for existing synthesized literature and guidelines related to depression screening and treatment in MEDLINE/PubMed, the Database of Abstracts of Reviews of Effects, Cochrane Database of Systematic Reviews, BMJ Clinical Evidence, Institute of Medicine, the National Institute for Health and Clinical Excellence, PsycINFO, the Agency for Healthcare Research and Quality, the American Psychiatric Association, the American Psychological Association, the Campbell Collaboration, the Canadian Agency for Drugs and Technologies in Health, the National Health Services' Health Technology Assessment Programme, and the Centre for Reviews and Dissemination, from 2008 through October 3, 2013. The search strategies are listed in the eMethods in the Supplement.

For pregnant and postpartum women, abstracts and full-text articles were systematically evaluated to identify existing systematic reviews to incorporate into the review, based on an approach outlined by Whitlock et al.[11] Three good-quality reviews were identified that served as foundational reviews for 1 or more KQs. These

reviews were chosen based on relevance (ie, inclusion and exclusion criteria that were at least as inclusive as our review), having conducted a good-quality search, having reported good-quality article evaluation methods, and recency.[12-14] For the question of harms of antidepressants (KQ5), 1 of the foundational reviews was of sufficient quality, and the evidence base was so extensive, that this review was used directly as evidence in the report and individual studies included in this review were not reevaluted.[14] The other 2 foundational reviews were used for study identification, and then a search was conducted for additional original research published after the search windows of these foundational reviews.[12,13] All studies included in each of these 2 foundational reviews were evaluated against our a priori inclusion/exclusion criteria.

We searched for newly published literature in the following databases: MEDLINE/PubMed, PsycINFO, and the Cochrane Central Register of Controlled Trials through January 20, 2015. The bridge search started from January 1, 2012, because there was at least 1 foundational review with a search period for each KQ that extended into 2012. Reference lists of other relevant publications were reviewed to identify additional potentially relevant studies that were not identified by the literature searches or foundational reviews.

Since January 2015, we continued to conduct ongoing surveillance through article alerts and targeted searches of high-impact journals to identify major studies published in the interim that may affect the conclusions or understanding of the evidence and therefore the related USPSTF recommendation. The last surveillance was conducted on December 9, 2015, and identified no new studies.

Study Selection
Two investigators independently reviewed 6536 titles and abstracts and 478 full-text articles against prespecified inclusion criteria (Figure 2). Disagreements were resolved through discussion or consultation with other investigators. We included English language fair- and good-quality studies involving women who were 18 years and older and pregnant or postpartum (within 1 year of birth at enrollment) and living in "very high-developed" countries according to the World Health Organization.[15] Studies limited to persons with other medical or mental health conditions were excluded; however, studies that included some persons with such conditions were not excluded, as long as it was not a requirement of participation.

For benefits and harms of depression screening (KQ1, KQ3), we included randomized or nonrandomized clinical trials conducted in primary care settings, including obstetrics/gynecology or, for postpartum women, pediatrics. To allow determination of the full population effect of screening programs, studies that included some participants who already had a medical record diagnosis of depression or were being treated for depression were not excluded. Studies of depression screening could also include additional treatment elements, as long as the screening test results were given to the primary care clinician. A requirement was that the control group either was not screened (KQ1) or did not have screening test results sent to their clinician (KQ1a). Outcomes had to be reported at a minimum of 6 weeks after randomization.

For diagnostic accuracy (KQ2), we examined studies of the Patient Health Questionnaire (PHQ) or Edinburgh Postnatal Depression Scale (EPDS) compared with a valid reference standard, which was defined as a structured or semistructured diagnostic interview with a trained interviewer or a nonbrief (>5 minutes) unstructured

Figure 1. Analytic Framework
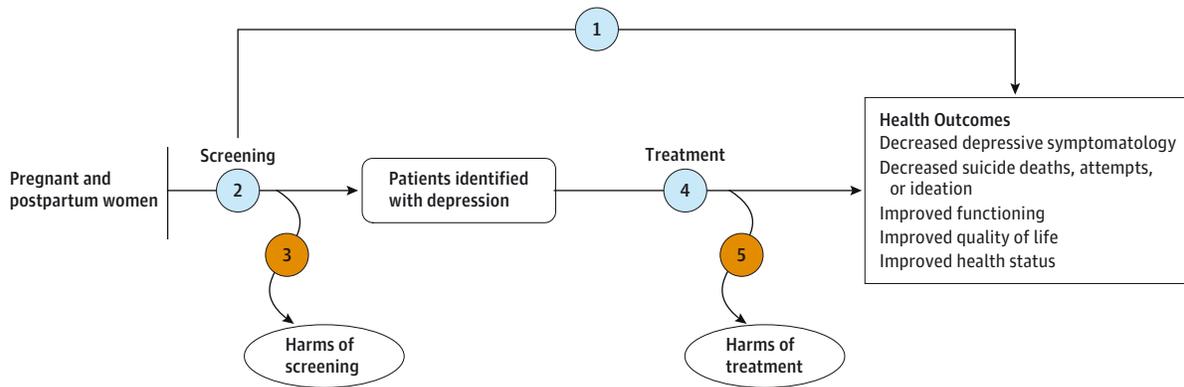
Key questions

1   Do primary care depression screening programs in pregnant and postpartum women result in improved health outcomes (decreased depressive symptomatology; decreased suicide deaths, attempts, or ideation; improved functioning; improved quality of life; or improved health status)?

   a. Does sending depression screening test results to clinicians (with or without additional care management supports) result in improved health outcomes?

   b. Does the effect of screening vary by population characteristics?[a]

2   What is the test performance of the most commonly used depression screening instruments in pregnant and postpartum women in primary care?

   a. Do the test performance characteristics of the screening instruments vary by population characteristics?[a]

3   What are the harms associated with primary care depression screening programs in pregnant and postpartum women?

   a. Do the harms vary by population characteristics?[a]

4   Does treatment (psychotherapy, antidepressants, or collaborative care) result in improved health outcomes (decreased depressive symptomatology; decreased suicide deaths, attempts, or ideation; improved functioning; improved quality of life; or improved health status) in pregnant and postpartum women who screen positive for depression in primary care?

   a. Do the effects of the interventions vary by population characteristics?

5   What are the harms of treatment in pregnant and postpartum women who screen positive for depression in primary care?

   a. Do the harms vary by population characteristics?[a]

   b. What is the prevalence of other selected serious harms of treatment with antidepressants in the general (ie, not limited to primary care) population of pregnant and postpartum women?

[a] Population characteristics include sex, age, race/ethnicity, comorbid conditions, and new-onset depression vs recurrent depression.
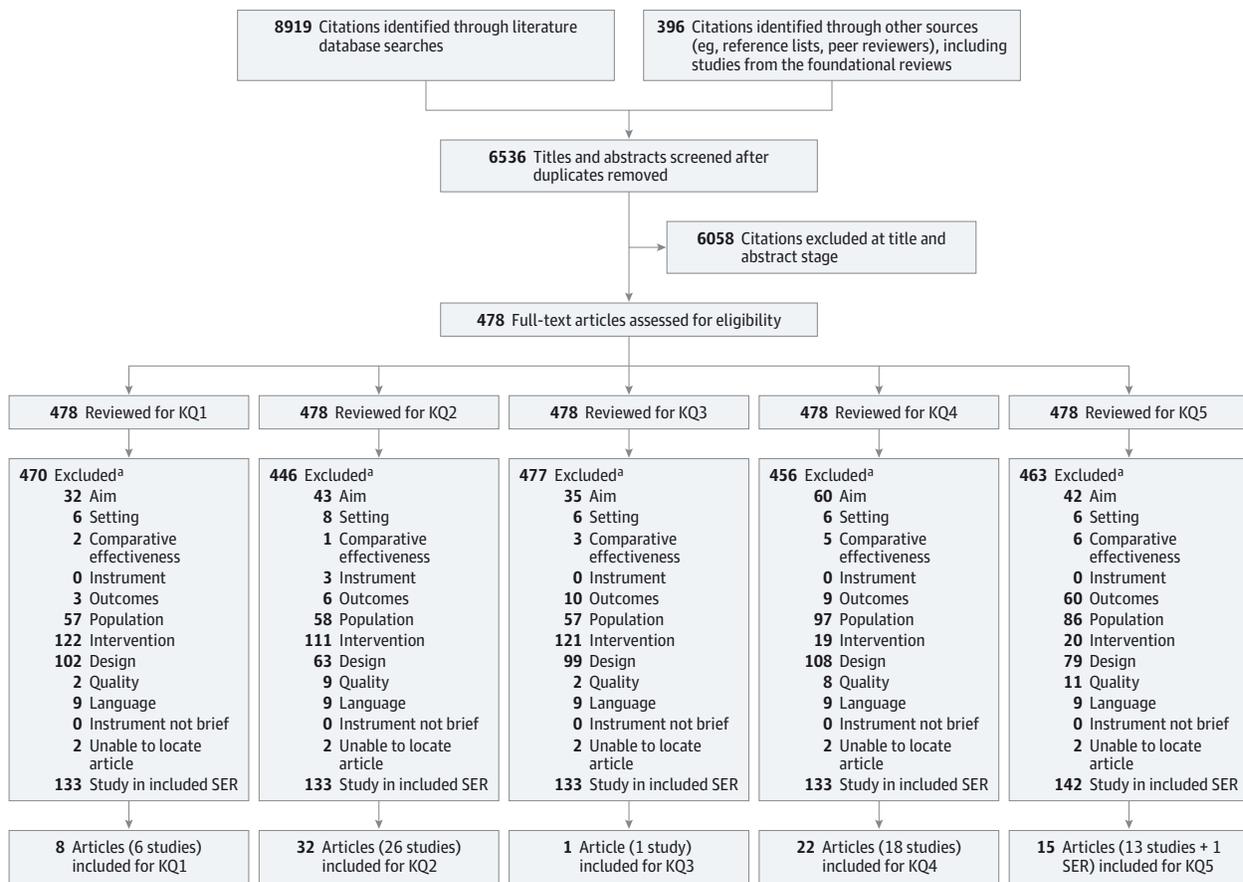
interview with a mental health clinician. Studies that gave the reference test only to a subset of participants had to make appropriate adjustments to their analysis or provide sufficient data to allow statistically adjusted analysis. Studies had to report sensitivity or specificity or the raw data to allow their calculation. The time between the index and reference tests could not exceed 2 weeks on average. In addition, these studies had to include patients comprising a wide spectrum of symptom severity, comparable with what would occur in typical primary care settings, including those without symptoms, those with subclinical symptomatology, and those with diagnostic-level symptomatology (ie, case-control designs were excluded). Studies of non-English versions of the instruments were included as long as the study was published in English.

For studies of the benefits of antidepressants and behavioral-based treatments (KQ4), trials were included that had a minimum of 6 weeks' follow-up after randomization that took place in primary or specialty care settings or online. Trials had to use population-

based screening to identify eligible patients. Studies were considered to include population-based screening if they attempted to recruit all or a consecutive or a random subset of women in a specific setting or population during the study's recruitment window, with individual outreach to potential participants for depression screening as part of determination of study eligibility. Thus, studies were excluded in which recruitment was based on referral, recruitment was from populations of known or likely depressed patients (eg, persons identified as depressed in their medical records), or volunteers were recruited through media or other advertising. Control groups could include usual care, no intervention, waitlist, attention control, or a minimal intervention (eg, ≤15 minutes of information, not intended to be a therapeutic dose).

These same studies were also examined for harms of treatment (KQ5). For serious harms of antidepressants in general populations of pregnant and postpartum women (not limited to screen-detected, KQ5b), systematic reviews, randomized or nonrandomized

Figure 2. Literature Flow Diagram


```
┌─────────────────────────────────┐   ┌─────────────────────────────────┐
│ 8919 Citations identified       │   │ 396 Citations identified through│
│ through literature database     │   │ other sources (eg, reference    │
│ searches                        │   │ lists, peer reviewers),         │
│                                 │   │ including studies from the      │
│                                 │   │ foundational reviews            │
└─────────────────────────────────┘   └─────────────────────────────────┘
                    │
┌─────────────────────────────────┐
│ 6536 Titles and abstracts       │
│ screened after duplicates       │
│ removed                         │
└─────────────────────────────────┘
                    │       ┌─────────────────────────────────┐
                    │──────▶│ 6058 Citations excluded at title│
                    │       │ and abstract stage              │
                    │       └─────────────────────────────────┘
┌─────────────────────────────────┐
│ 478 Full-text articles assessed │
│ for eligibility                 │
└─────────────────────────────────┘
```

478 Reviewed for KQ1	478 Reviewed for KQ2	478 Reviewed for KQ3	478 Reviewed for KQ4	478 Reviewed for KQ5
470 Excluded[a]	446 Excluded[a]	477 Excluded[a]	456 Excluded[a]	463 Excluded[a]
32 Aim	43 Aim	35 Aim	60 Aim	42 Aim
6 Setting	8 Setting	6 Setting	6 Setting	6 Setting
2 Comparative effectiveness	1 Comparative effectiveness	3 Comparative effectiveness	5 Comparative effectiveness	6 Comparative effectiveness
0 Instrument	3 Instrument	0 Instrument	0 Instrument	0 Instrument
3 Outcomes	6 Outcomes	10 Outcomes	9 Outcomes	60 Outcomes
57 Population	58 Population	57 Population	97 Population	86 Population
122 Intervention	111 Intervention	121 Intervention	19 Intervention	20 Intervention
102 Design	63 Design	99 Design	108 Design	79 Design
2 Quality	9 Quality	2 Quality	8 Quality	11 Quality
9 Language	9 Language	9 Language	9 Language	9 Language
0 Instrument not brief	0 Instrument not brief	0 Instrument not brief	0 Instrument not brief	0 Instrument not brief
2 Unable to locate article	2 Unable to locate article	2 Unable to locate article	2 Unable to locate article	2 Unable to locate article
133 Study in included SER	133 Study in included SER	133 Study in included SER	133 Study in included SER	142 Study in included SER
8 Articles (6 studies) included for KQ1	32 Articles (26 studies) included for KQ2	1 Article (1 study) included for KQ3	22 Articles (18 studies) included for KQ4	15 Articles (13 studies + 1 SER) included for KQ5

[a] Details about reasons for exclusion are as follows. Aim: study aim not relevant. Setting: study was not conducted in a setting or country relevant to US primary care. Comparative effectiveness: study did not have a control group. Instrument: study did not use an included screening instrument. Outcomes: study did not have relevant outcomes or had incomplete outcomes. Population: study was not conducted in a pregnant or postpartum population or was limited to a narrow population not broadly representative of primary care. Intervention: study used an excluded intervention or screening approach. Design: study did not use an included design. For review for key question 2 (KQ2), design included >2 weeks between screening and reference test, or reference test was not applied to full range of screening results or could not adjust for partial verification. Quality: study did not meet criteria for fair or good quality (ie, it was poor quality) using study design–specific criteria developed by the US Preventive Services Task Force for randomized clinical trials,[16] the Quality Assessment of Diagnostic Accuracy Studies 2 for diagnostic accuracy studies,[17] the Newcastle-Ottawa Scale[18] for observational studies, or A Measurement Tool to Assess Systematic Reviews (AMSTAR) for systematic reviews.[19] The criteria and definitions of good, fair, poor are provided in eTable 1 in the Supplement. Language: study was published in a non-English language. Instrument not brief: study included a screening instrument that was not brief (ie, exceeded 15 minutes to complete). Study included in systematic evidence review (SER): study was included in existing SER that was included as evidence.

clinical trials, and large comparative observational studies were included. Maternal harms included suicidality, serotonin syndrome, cardiac effects, seizures, bleeding, cardiometabolic effects, miscarriage, and preeclampsia. Infant harms included neonatal death, major malformations, small for gestational age and low birth weight, seizures, serotonin withdrawal syndrome, neonatal respiratory distress, cardiopulmonary effects, and other serious events requiring medical attention. Comparative cohort studies had to have a minimum of 10 cases in each exposure group and include appropriate controls who were not taking antidepressants.

Data Extraction and Quality Assessment

Two investigators independently assessed the quality of the included studies by using criteria defined by the USPSTF[16] and supplemented with criteria from the Quality Assessment of Diagnostic Accuracy 2 (QUADAS-2)[17] for diagnostic accuracy studies, the Newcastle-Ottawa Scale (NOS)[18] for observational studies, and A Measurement Tool to Assess Systematic Reviews (AMSTAR) for systematic reviews (eTable 1 in the Supplement).[19] Each study was assigned a final quality rating of good, fair, or poor; disagreements between the investigators were resolved through discussion. We rated and excluded studies as poor quality if there was a major "fatal flaw" (eg, attrition was >40%, differential attrition >20%) or multiple important limitations that could invalidate the results.

One investigator abstracted data from the included studies, and a second investigator checked the data for accuracy. We abstracted study design characteristics, population demographics, baseline history of depression and other mental health conditions, screening and intervention details, depression outcomes, other health outcomes (eg, suicidality, mortality, quality of life, function-

Table 1. Study Characteristics of Trials of Benefits of Screening (Key Questions 1 and 1a)

Source	Quality[a]	No. of Patients	Study Design	Intervention	Planned Follow-up, mo	Country	Age, Mean (Range), y	Race/Ethnicity, No. (%)	Weeks Postpartum or Gestational Week at Baseline
Leung et al,[27] 2011	Good	462	RCT	EPDS screening	4, 16	Hong Kong	NR (NR)	NR	8 (postpartum)
Wickberg et al,[29] 2005	Fair	669	Cluster RCT	EPDS screening results feedback to clinician, brief depression training	2.75	Sweden	NR (NR)	NR	25 (gestation)
Yawn et al,[30] 2012	Fair	2343	Cluster RCT	EPDS and PHQ-9 screening results feedback to clinician, clinician training and supports	6, 12	United States	26.4 (≥18)	Black: 421 (18) Hispanic: 282 (12) White: NR	8 (postpartum)
MacArthur et al,[28] 2002	Fair	2064	Cluster RCT	EPDS screening, midwife training and supports	3	United Kingdom	NR (NR)	NR	4 (postpartum)
Morrell et al,[22] 2009	Fair	4084	Cluster RCT	EPDS screening, CBT or person-centered counseling	5	United Kingdom	NR (≥18)	Black: NR Hispanic: NR White: 3892 (95.3)	6 (postpartum)
Glavin et al,[26] 2010	Fair	2247	CCT[b]	EPDS screening, redesigned follow-up care	1.5, 4.5	Norway	32.5 (≥18)	NR	6 (postpartum)

Abbreviations: CBT, cognitive behavioral therapy; CCT, controlled clinical trial; EPDS, Edinburgh Postnatal Depression Scale; NR, not reported; PHQ, Patient Health Questionnaire; RCT, randomized clinical trial.

[a] Quality assessed using criteria developed by the US Preventive Services Task Force.[16]

[b] Group assignment was nonrandom.

ing, health status, infant outcomes, emergency department visits, inpatient stays), adverse events, and diagnostic accuracy statistics.

Data Synthesis and Analysis

We created summary tables of study characteristics, population characteristics, intervention characteristics, and outcomes separately for each KQ. These tables and forest plots of the results were used to examine the consistency, precision, and relationship of effect size with key potential modifiers. We had a sufficient number of trials with acceptable comparability to conduct a meta-analysis of trials examining the benefits of cognitive behavioral therapy (CBT) and related approaches. Because this analysis included 10 studies with low statistical heterogeneity, as assessed by the I^2 statistic, and fairly comparable sample sizes, a random-effects model was used (DerSimonian and Laird),[20] with a sensitivity analysis using a restricted maximum likelihood model with the Knapp-Hartung modification for small samples.[21] Funnel plots and the Egger test were used to examine the risk of small study effects. For the studies of instrument accuracy (KQ2), sensitivity and specificity with Jeffrey confidence intervals were calculated, using data from 2 × 2 tables that included true positives, false positives, false negatives, and true negatives. Several studies only verified a negative screening result in a random sample of participants below a predetermined threshold (which was lower than the typical cutoffs for a positive screener in all cases).[22-24] For these studies, the proportion with a depressive disorder according to the reference standard was applied to the full sample of those below the threshold, and sensitivity and specificity were calculated based on these extrapolated results.[25] In all cases, there were no false negatives, so sensitivity did not change, but specificity increased with extrapolation, although we were unable to accurately determine the number of noncases for 1 study and so did not calculate specificity.[22] Side-by-side plots of sensitivity and specificity were created in R version 3.2.2 (R Foundation); all other analyses were conducted in Stata

version 13.1 (StataCorp). All significance testing was 2-sided, and results were considered statistically significant if the P value was .05 or less.

Results

This article focused only on the evidence related to pregnant and peripartum women, which covers most of the new evidence since the previous review and omits coverage of some sub-KQs that had no or minimal evidence, specifically key questions related to variation in results by population characteristics (KQs 1b, 2a, 3a, 4a, and 5a).
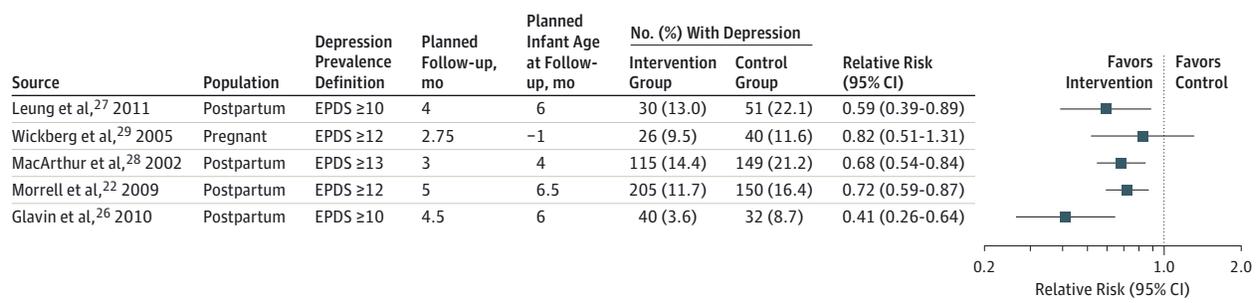
Benefits of Screening

Key Question 1. Do primary care depression screening programs in pregnant and postpartum women result in improved health outcomes (decreased depressive symptomatology; decreased suicide deaths, attempts, or ideation; improved functioning; improved quality of life; or improved health status)?

Key Question 1a. Does sending depression screening test results to clinicians (with or without additional care management supports) result in improved health outcomes?
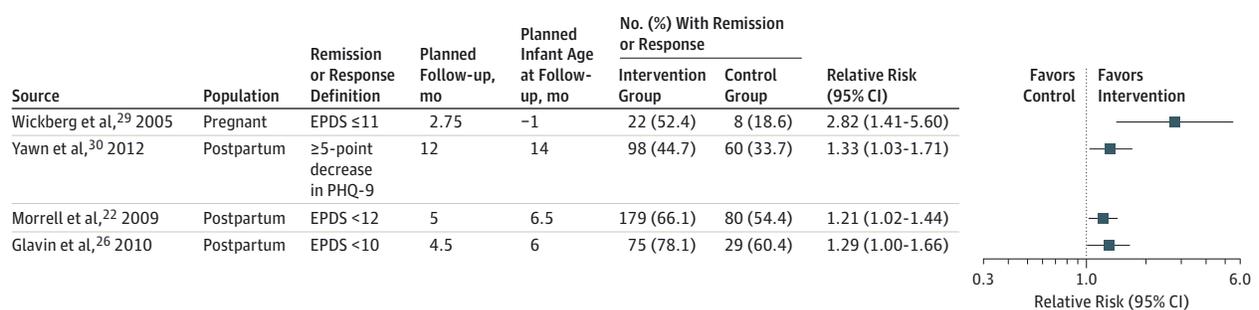
One good-quality and 5 fair-quality trials were included that examined the benefits of screening for pregnant and postpartum depression (n = 11 869)[22,26-30] with or without additional clinician training or treatment components (Table 1; trials are arranged in increasing order of the extensiveness of the treatment components in addition to screening). Five trials focused on postpartum women,[22,26-28,30] and the sixth focused on pregnant women.[29] All trials studied women identified in health care settings and included all study-eligible women regardless of screening test results.[22,26-30] Two trials included unscreened control groups[27,28] (KQ1), and 4 screened all participants but sent results to only the intervention group's clinicians (KQ1a).[22,26,29,30] Trials screened women

Figure 3. Benefits of Screening Programs: Depression Prevalence Reported Within 1 Year (Key Questions 1 and 1a)

Source	Population	Depression Prevalence Definition	Planned Follow-up, mo	Planned Infant Age at Follow-up, mo	No. (%) With Depression		Relative Risk (95% CI)
					Intervention Group	Control Group	
Leung et al,[27] 2011	Postpartum	EPDS ≥10	4	6	30 (13.0)	51 (22.1)	0.59 (0.39-0.89)
Wickberg et al,[29] 2005	Pregnant	EPDS ≥12	2.75	−1	26 (9.5)	40 (11.6)	0.82 (0.51-1.31)
MacArthur et al,[28] 2002	Postpartum	EPDS ≥13	3	4	115 (14.4)	149 (21.2)	0.68 (0.54-0.84)
Morrell et al,[22] 2009	Postpartum	EPDS ≥12	5	6.5	205 (11.7)	150 (16.4)	0.72 (0.59-0.87)
Glavin et al,[26] 2010	Postpartum	EPDS ≥10	4.5	6	40 (3.6)	32 (8.7)	0.41 (0.26-0.64)

EPDS indicates Edinburgh Postnatal Depression Scale. Error bars indicate 95% confidence interval.

Figure 4. Benefits of Screening Programs: Depression Remission or Response Reported Within 1 Year (Key Questions 1 and 1a)

Source	Population	Remission or Response Definition	Planned Follow-up, mo	Planned Infant Age at Follow-up, mo	No. (%) With Remission or Response		Relative Risk (95% CI)
					Intervention Group	Control Group	
Wickberg et al,[29] 2005	Pregnant	EPDS ≤11	2.75	−1	22 (52.4)	8 (18.6)	2.82 (1.41-5.60)
Yawn et al,[30] 2012	Postpartum	≥5-point decrease in PHQ-9	12	14	98 (44.7)	60 (33.7)	1.33 (1.03-1.71)
Morrell et al,[22] 2009	Postpartum	EPDS <12	5	6.5	179 (66.1)	80 (54.4)	1.21 (1.02-1.44)
Glavin et al,[26] 2010	Postpartum	EPDS <10	4.5	6	75 (78.1)	29 (60.4)	1.29 (1.00-1.66)

EPDS indicates Edinburgh Postnatal Depression Scale; PHQ-9, 9-item Patient Health Questionnaire. Error bars indicate 95% confidence interval.

at week 25 of gestation[29] or 4 to 8 weeks postpartum.[22,26-28,30] Only 1 trial was conducted in the United States.[26] Both of the individually randomized trials excluded women who were currently being treated for depression[26,27]; however, the trials that randomized at the level of a midwife or medical practice all had very broad inclusion criteria and did not exclude women being treated for depression.[22,28-30] All studies used the EPDS for screening; cutoffs for screening positive ranged from 10 to 13. While 1 trial focused narrowly on the benefit of adding the EPDS to the usual clinical evaluation,[22] others provided a wide range of components in addition to the screening intervention, such as clinician training and support, person-centered counseling, or redesigned follow-up care.

At follow-up, which ranged from 1.5 to 16 months, 5 of 6 trials reported the proportion of women scoring above a specified cutoff on the EPDS, which we refer to as depression prevalence (**Figure 3**). In pregnant and postpartum women, there were relative reductions of 18% to 59% in the risk of depression at follow-up compared with usual care, which translated to 2.1% to 9.1% absolute reductions in depression prevalence, according to a variety of EPDS cutoff definitions. For example, depression prevalence (defined as an EPDS score ≥10) was 13% in the screened group in the Hong Kong–based screening-only intervention in the near term (4 months) but 22.1% in the nonscreened group.[27] However, this effect was not sustained at 16 months.[27] In the study of pregnant women that included feedback of screening results and a 1-afternoon depression training session for midwives, the effect size was smaller and not statistically significant, with 9.5% of women in the intervention group reporting EPDS scores of 12 or more at follow-up,

compared with 11.6% of women in usual care.[29] In the 3 studies that reported outcomes similar to remission (ie, no longer screened positive) or treatment response (ie, showed a predetermined level of improvement on a scale score) in postpartum women, there was a 21% to 33% increase in the likelihood of remission or response at 4.5 to 12 months (6-14 months postpartum), ranging from 10.0% to 33.8% absolute increases in the likelihood or remission or response (**Figure 4**).[22,26,30] The effect was even larger in the trial of pregnant women, but last follow-up was only at 2.75 months.[29]

The results most applicable to US primary care come from a fair-quality US trial of screening plus clinician supports.[30] Forty-five percent of intervention participants reported a 5-point or greater drop in their PHQ-9 scores, the a priori definition of clinical meaningful benefit, whereas 34% of those receiving usual care reported such a drop (odds ratio [OR], 1.74 [95% CI, 1.05-5.86], adjusted for depression history, marital status, income, education, age, and degree of parenting stress). This trial was rated as fair primarily because attrition was greater than 25% in both groups, which was a common problem in the studies on the benefits of screening for depression.

Performance Characteristics of the EPDS and PHQ

Key Question 2. What is the test performance of the most commonly used depression screening instruments in pregnant and postpartum women in primary care?

We identified 23 studies[22-24,31-50] (n = 5398) that examined the accuracy of the EPDS and 3 studies that examined the PHQ (n = 777)[51-53] relative to a diagnostic interview (**Table 2**, EPDS studies

Table 2. Study Characteristics of Diagnostic Accuracy Studies of the EPDS and PHQ (Key Question 2)

Source	Quality[a]	No. of Patients	Reference Standard	Country (Language)	Age, Mean (Range), y	Race/Ethnicity, No. (%)	Weeks Postpartum or Gestational Week at Baseline
English EPDS							
Tandon et al,[42] 2012	Fair	95	SCID-I/NP diagnosis of (1) MDD and (2) major or minor depression	United States	24.4 (NR)	Black: 95 (100) Hispanic: NR White: NR	Pregnant or 26 weeks postpartum
Harris et al,[39] 1989	Fair	126	*DSM-II* criteria for (1) MDD and (2) major or minor depression	United Kingdom	24.6 (17-40)	NR	6 (postpartum)
Clarke,[49] 2008	Fair	103	SCID for MDD	Canada	23.8 (18-42)	NR	4-52 (postpartum)
Morrell et al,[22] 2009	Fair	860	SCAN interview diagnosis of depression	United Kingdom	NR (≥18)	Black: NR Hispanic: NR White: 3892 (95.3)	6 (postpartum)
Beck and Gable,[32] 2001	Fair	150	*DSM-IV* diagnosis of (1) MDD and (2) any depressive disorder	United States	31 (18-46)	Black: 12 (8) Hispanic: 5 (3.3) White: 130 (86.7)	2-12 (postpartum)
Cox et al,[23] 1996	Fair	272	SPI interview criteria for (1) MDD and (2) major or minor depression	United Kingdom	25.4 (NR)	NR	24 (postpartum)
Murray and Cox,[48] 1990	Fair	100	SPI using RDC for (1) MDD and (2) major or minor depression	United Kingdom	24.6 (NR)	NR	28-34 (gestation)
Leverton and Elliott,[41] 2000	Fair	199	PSE interview and Bedford College diagnosis of (1) case depression and (2) borderline or case depression	United Kingdom	NR (NR)	NR	12 (postpartum)
Non-English EPDS							
Alvarado et al,[47] 2015	Fair	111	*DSM-IV* or *ICD-9* diagnosis of MDD based on MINI interview	Chile (Spanish)	25 (18-43)	NR	28 (gestation)
Adouard et al,[31] 2005	Fair	60	MINI *DSM-IV* criteria for MDD	France (French)	31.5 (23-46)	NR	28-34 (gestation)
Benvenuti et al,[33] 1999	Fair	113	MINI *DSM-III-R* criteria for any depressive disorder	Italy (Italian)	31.9 (NR)	NR	0.5 (postpartum)
Carpiniello et al,[36] 1997	Fair	61	Clinically depressed by the PSE interview	Italy (Italian)	31.6 (22-43)	NR	4-6 (postpartum)
Felice et al,[50] 2006	Fair	223	*ICD-9* based on CIS-R interview for severe, moderate, or mild depression episode	Malta (Maltese)	27.1 (15-34)	NR	Average 18.6 (gestation)
Bunevicius et al,[35] 2009b	Fair	230	SCID-NP diagnosis of (1) MDD and (2) any depressive disorder during first trimester	Lithuania (Lithuanian)	29 (18-43)	NR	First trimester (gestation)
Garcia-Esteve et al,[24] 2003	Fair	1123	SCID diagnosis of (1) MDD and (2) any depressive disorder	Spain (Spanish)	30.2 (NR)	NR	6 (postpartum)
Töreki et al,[44] 2013	Good	219	SCID *DSM-IV* criteria for (1) MDD and (2) any depressive disorder	Hungary (Hungarian)	30.0 (17-42)	NR	12 (gestation)
Töreki et al,[45] 2014	Fair	266	SCID diagnosis of (1) MDD and (2) any depressive disorder	Hungary (Hungarian)	30.5 (18-42)	NR	6 (postpartum)
Guedeney and Fermanian,[38] 1998	Fair	87	RDC diagnosis of major or minor depressive disorder	France (French)	30.4 (20-42)	NR	16 (postpartum)
Yamashita et al,[46] 2000	Fair	75	SADS diagnostic interview for major or minor depression	Japan (Japanese)	31 (19-41)	NR	4 (postpartum)
Bunevicius et al,[34] 2009a	Fair	94	CIDI (short form) diagnosis of any depressive disorder	Lithuania (Lithuanian)	29 (20-43)	NR	2 (postpartum)
Lee et al,[40] 2001	Fair	145	SCID-NP diagnosis of major or minor depression	Hong Kong (Chinese)	29 (16-42)	Black: 0 Hispanic: 0 White: 0	6 (postpartum)
Chen et al,[37] 2013	Fair	487	*DSM-IV-TR* clinical interview diagnosis of any depressive disorder	Singapore (Chinese)	30.4 (19-43)	Black: 0 Hispanic: 0 White: 0	1-22 (postpartum)
Teng et al,[43] 2005	Fair	199	MINI *DSM-IV* diagnosis of any depressive disorder	Taiwan (Taiwanese)	29 (16-41)	NR	6 (postpartum)
English PHQ							
Smith et al,[53] 2010	Fair	213	CIDI for MDD	United States	28.9 (≥17)	Black: 43 (20.1) Hispanic: 21 (9.8) White: 135 (63.1)	<17 (gestation)
Gjerdingen et al,[51] 2009b	Fair	438	SCID for MDD	United States	29.1 (≥12)	Black: 89 (17.6) Hispanic: 14 (2.8) White: 339 (67)	4 (postpartum)
Mann et al,[52] 2012	Fair	126	*DSM-IV* interview using guidance from the SCID for major or minor depression	United Kingdom	27.4 (≥18)	Black: 6 (3.9) Hispanic: NR White: 86 (56.6)	26-28 (gestation)

Abbreviations: CIDI, Composite International Diagnostic Interview; CIS-R, Clinical Interview Schedule-Revised; *DSM, Diagnostic and Statistical Manual of Mental Disorders*; EPDS, Edinburgh Postnatal Depression Scale; *ICD-9, International Classification of Diseases, Ninth Revision*; MDD, major depressive disorder; MINI, Mini International Neuropsychiatric Interview; NP, nonpatient; NR, not reported; PHQ, Patient Health Questionnaire; PSE, Present State Examination; RDC, Research

Diagnostic Criteria; SADS, Schedule for Affective Disorders and Schizophrenia; SCAN, Schedules for Clinical Assessment in Neuropsychiatry; SCID, Structured Clinical Interview for *DSM* Disorders; SPI, Standardized Psychiatric Interview.

[a] Quality assessed using criteria from the Quality Assessment of Diagnostic Accuracy Studies 2.[17]

are arranged in the order of decreasing proportion meeting the reference standard diagnosis, separately for English-version EPDS and non–English-version EPDS; PHQ studies are ordered by the PHQ versions reported). Eight of the included studies used the English-language version of the EPDS (n = 1905).[22,23,32,39,41,42,48,49] Six of the English-language EPDS studies assessed postpartum women, usually between 6 and 12 weeks postpartum,[24,26,28,34,37,39] 1 assessed pregnant women,[48] and 1 assessed women at any point during pregnancy and up to 26 weeks postpartum.[42] We focused on the English-language EPDS and standard cutoff scores of 10 (indicating moderate-level symptoms) and 13 (indicating probable depressive disorder) for the EPDS.

At a cutoff score of 13 for identifying MDD, the sensitivity of the English-language EPDS ranged from 0.67 (95% CI, 0.18-0.96) to 1.00 (95% CI, 0.67-1.00), with most of the results between 0.75 and 0.82 (Figure 5 and eFigure 1 in the Supplement). The largest of these studies,[22] from the United Kingdom, reported a sensitivity of 0.79 (95% CI, 0.72-0.85), which was very similar to that seen in a relatively recent US-based study with low-income African American women with a high rate of depression (0.81 [95% CI, 0.64-0.93]).[42] The specificity of the English-language EPDS was 0.87 or greater in all studies. Sensitivity for detecting depressive disorders, including both major and minor depression, using the cutoff of 10 or greater ranged from 0.63 (95% CI, 0.44-0.79)[23] to 0.84.[42,49] At a cutoff score of 10, the study of low-income African American women reported[42] sensitivity of 0.84 (95% CI, 0.69-0.94) and specificity of 0.81 (95% CI, 0.70-0.89) for identifying major or minor depression in pregnant and postpartum women combined. The estimates were very similar for pregnant and postpartum women.[42]

The PHQ studies covered 3 different versions of the PHQ (PHQ-2,[51-53] PHQ-8,[53] PHQ-9[51]) and 3 different scoring methods for the PHQ-2 (Figure 6 and eFigure 2 in the Supplement). Sensitivities and specificities were fairly wide-ranging across different versions, scoring methods, diagnostic comparators, and cutoffs, and no single method was reported in more than 1 study.

Harms of Screening

Key Question 3. What are the harms associated with primary care depression screening programs in pregnant and postpartum women?

Among the trials addressing the benefits of screening (KQ1), 1 trial reported that there were no adverse effects of depression screening in postpartum women (n = 462; Table 1)[27]; the remaining 5 trials did not report on harms.

Benefits of Treatment

Key Question 4. Does treatment (psychotherapy, antidepressants, or collaborative care) result in improved health outcomes (decreased depressive symptomatology; decreased suicide deaths, attempts, or ideation; improved functioning; improved quality of life; or improved health status) in pregnant and postpartum women who screen positive for depression in primary care?
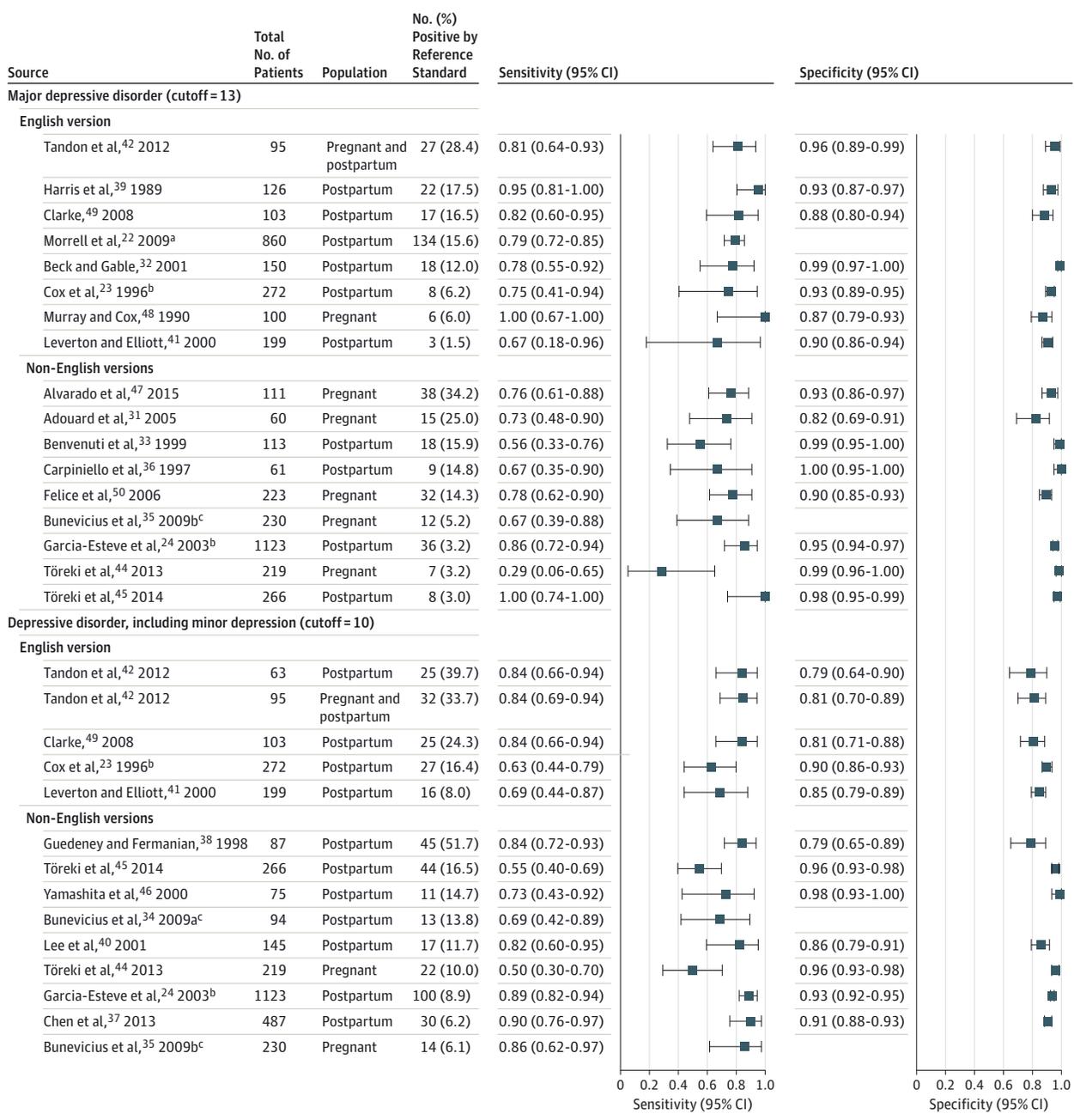
We identified 2 good-quality and 16 fair-quality trials (n = 1638) that examined the benefits of interventions in pregnant and postpartum women who had screened positive for depression in a primary care or community setting, generally compared with usual care[54-71] (Table 3, trials are arranged in increasing order of estimated contact hours with the intervention). One trial combined

treatment in depressed women and prevention in women who were not depressed, but we only included results related to the depressed subgroup (n = 324).[63] Fifteen of 18 trials recruited women during the postpartum period (≤22 weeks) and 3 during their pregnancy.[63,64,71] All 18 trials reported outcomes during the postpartum period. Time to follow-up varied widely, from 6 weeks[59,69] to 18 months.[56] Furthermore, trials varied in time between end of treatment and follow-up assessment, with 7 trials conducting follow-up assessment within 2 weeks of when treatment ended,[55,57,62,65,66,69,71] while the remaining had a lag of 1 to 7 months between end of treatment and follow-up assessment. The most common behavioral interventions were CBT or related interventions that included traditional CBT components, such as stress management, goal setting, and problem solving, including 2 trials conducted with pregnant women.[63,64] Other intervention approaches included fluoxetine,[55] a health care system–level stepped-care intervention,[57] nondirective counseling,[56,60,69] psychodynamic therapy,[58] an information-only intervention,[59] and 2 different approaches to improving the mother-infant relationship.[58,62]

Of 18 trials, 15 reported an outcome similar to depression remission (usually the proportion below a specified cut point on a depression symptom scale) at follow-up ranging from 1.5 to 18 months (Figure 7, only outcomes within 1 year shown).[54,56-61,63-67,69-71] All 10 trials that used CBT or related interventions showed an increased likelihood of remission with treatment in the short-term, although not all results were statistically significant.[54,56,61,63-67,70,71] Effect sizes were similar for pregnant and postpartum women for CBT. Pooled results that used only the longest follow-up (within 1 year) showed an increase in the likelihood of remission with CBT (DerSimonian and Laird pooled relative risk [RR], 1.34 [95% CI, 1.19-1.50]; No. of studies (K) = 10; I^2 = 7.9%) compared with usual care, with absolute increases ranging from 6.2% to 34.6%. Results were almost identical in sensitivity analyses using a more conservative pooling method, with even lower statistical heterogeneity (restricted maximum likelihood pooled RR, 1.34 [95% CI, 1.17-1.53]; K = 10, I^2 = 0%). Increased hours of contact might be associated with larger effect sizes, but because contact hours, sample size, control group remission rates, and time to follow-up were all confounded with each other, conclusions could not be drawn about their relative importance. The funnel plot (eFigure 3 in the Supplement) suggested an increased risk of small studies bias, consistent with increased risk of publication bias; the Egger test did not identify a statistically significant small studies bias, but power was limited. The possibility of correlation between sample size and effect size raises the concern that the pooled effect may overestimate the true effect.

Results for the outcome of continuous symptom score showed a similar pattern (Figure 8 and eFigure 4 in the Supplement), although only 7 of the trials were available for pooling.[54,61,64-67,71] All of the trials showed greater symptom reduction in the intervention groups. Results were not statistically significant in 3 trials[64,66,67]; however, unadjusted mean differences were statistically significant in 1 of these.[67] With usual care, EPDS scores declined by an average of 2 to 6 points, compared with 5 to 10 points in intervention groups. The pooled standardized mean difference in change between groups was −0.82 (95% CI, −1.10 to −0.54; K = 7, I^2 = 35.4%), consistent with a medium to large effect size according to Cohen's

Figure 5. Diagnostic Accuracy of the Edinburgh Postnatal Depression Scale Relative to a Diagnostic Interview (Key Question 2)

Source	Total No. of Patients	Population	No. (%) Positive by Reference Standard	Sensitivity (95% CI)	Specificity (95% CI)
Major depressive disorder (cutoff = 13)					
English version					
Tandon et al,[42] 2012	95	Pregnant and postpartum	27 (28.4)	0.81 (0.64-0.93)	0.96 (0.89-0.99)
Harris et al,[39] 1989	126	Postpartum	22 (17.5)	0.95 (0.81-1.00)	0.93 (0.87-0.97)
Clarke,[49] 2008	103	Postpartum	17 (16.5)	0.82 (0.60-0.95)	0.88 (0.80-0.94)
Morrell et al,[22] 2009[a]	860	Postpartum	134 (15.6)	0.79 (0.72-0.85)	
Beck and Gable,[32] 2001	150	Postpartum	18 (12.0)	0.78 (0.55-0.92)	0.99 (0.97-1.00)
Cox et al,[23] 1996[b]	272	Postpartum	8 (6.2)	0.75 (0.41-0.94)	0.93 (0.89-0.95)
Murray and Cox,[48] 1990	100	Pregnant	6 (6.0)	1.00 (0.67-1.00)	0.87 (0.79-0.93)
Leverton and Elliott,[41] 2000	199	Postpartum	3 (1.5)	0.67 (0.18-0.96)	0.90 (0.86-0.94)
Non-English versions					
Alvarado et al,[47] 2015	111	Pregnant	38 (34.2)	0.76 (0.61-0.88)	0.93 (0.86-0.97)
Adouard et al,[31] 2005	60	Pregnant	15 (25.0)	0.73 (0.48-0.90)	0.82 (0.69-0.91)
Benvenuti et al,[33] 1999	113	Postpartum	18 (15.9)	0.56 (0.33-0.76)	0.99 (0.95-1.00)
Carpiniello et al,[36] 1997	61	Postpartum	9 (14.8)	0.67 (0.35-0.90)	1.00 (0.95-1.00)
Felice et al,[50] 2006	223	Pregnant	32 (14.3)	0.78 (0.62-0.90)	0.90 (0.85-0.93)
Bunevicius et al,[35] 2009b[c]	230	Pregnant	12 (5.2)	0.67 (0.39-0.88)	
Garcia-Esteve et al,[24] 2003[b]	1123	Postpartum	36 (3.2)	0.86 (0.72-0.94)	0.95 (0.94-0.97)
Töreki et al,[44] 2013	219	Pregnant	7 (3.2)	0.29 (0.06-0.65)	0.99 (0.96-1.00)
Töreki et al,[45] 2014	266	Postpartum	8 (3.0)	1.00 (0.74-1.00)	0.98 (0.95-0.99)
Depressive disorder, including minor depression (cutoff = 10)					
English version					
Tandon et al,[42] 2012	63	Postpartum	25 (39.7)	0.84 (0.66-0.94)	0.79 (0.64-0.90)
Tandon et al,[42] 2012	95	Pregnant and postpartum	32 (33.7)	0.84 (0.69-0.94)	0.81 (0.70-0.89)
Clarke,[49] 2008	103	Postpartum	25 (24.3)	0.84 (0.66-0.94)	0.81 (0.71-0.88)
Cox et al,[23] 1996[b]	272	Postpartum	27 (16.4)	0.63 (0.44-0.79)	0.90 (0.86-0.93)
Leverton and Elliott,[41] 2000	199	Postpartum	16 (8.0)	0.69 (0.44-0.87)	0.85 (0.79-0.89)
Non-English versions					
Guedeney and Fermanian,[38] 1998	87	Postpartum	45 (51.7)	0.84 (0.72-0.93)	0.79 (0.65-0.89)
Töreki et al,[45] 2014	266	Postpartum	44 (16.5)	0.55 (0.40-0.69)	0.96 (0.93-0.98)
Yamashita et al,[46] 2000	75	Postpartum	11 (14.7)	0.73 (0.43-0.92)	0.98 (0.93-1.00)
Bunevicius et al,[34] 2009a[c]	94	Postpartum	13 (13.8)	0.69 (0.42-0.89)	
Lee et al,[40] 2001	145	Postpartum	17 (11.7)	0.82 (0.60-0.95)	0.86 (0.79-0.91)
Töreki et al,[44] 2013	219	Pregnant	22 (10.0)	0.50 (0.30-0.70)	0.96 (0.93-0.98)
Garcia-Esteve et al,[24] 2003[b]	1123	Postpartum	100 (8.9)	0.89 (0.82-0.94)	0.93 (0.92-0.95)
Chen et al,[37] 2013	487	Postpartum	30 (6.2)	0.90 (0.76-0.97)	0.91 (0.88-0.93)
Bunevicius et al,[35] 2009b[c]	230	Pregnant	14 (6.1)	0.86 (0.62-0.97)	

Error bars indicate 95% confidence interval.

[a] Morrell et al (2009)[22] did not report sufficient data to extrapolate the number of false positives and true negatives; therefore, specificity could not be calculated.

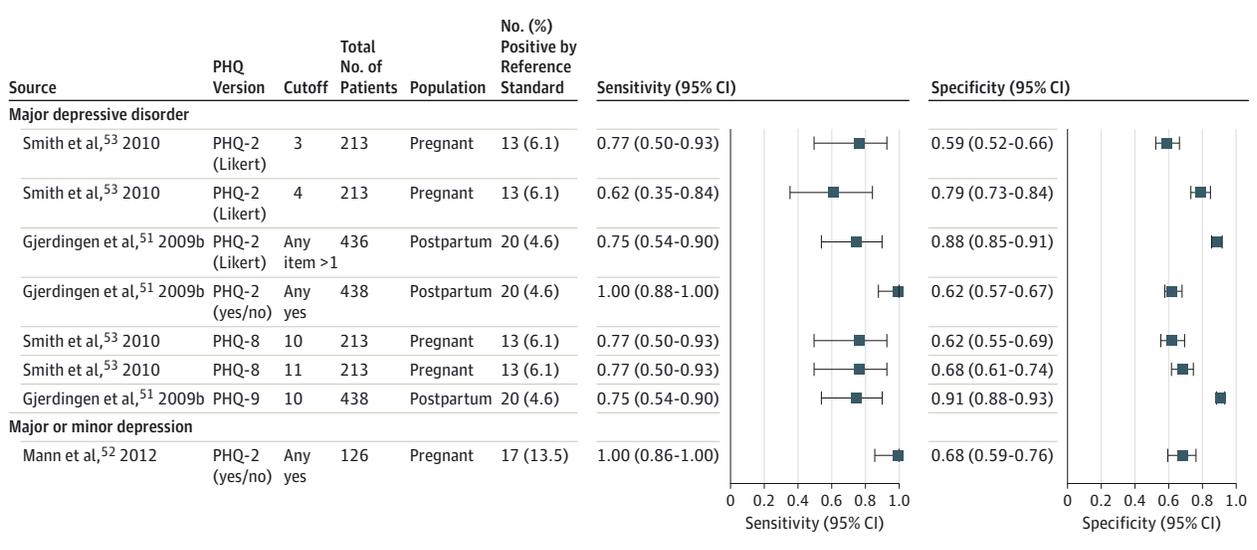[b] Data were extrapolated from partial verification.

[c] Bunevicius et al (2009a)[34] and Bunevicius et al (2009b)[35] did not report the number of false positives or true negatives; therefore, specificity could not be calculated.

suggested convention.[72] Average baseline EPDS scores were generally at or above the cutoff of 13 (cutoff for identifying MDD), and at follow-up most CBT group averages were below 10 (cutoff for identifying minor or major depressive disorder), which put them in the mild depressive symptom range, on average. Some studies showed average EPDS scores below 10 at follow-up in both the intervention and usual care groups[64,67]; in other trials, the usual care groups remained above 10 while the intervention groups were below 10[54,70]

or showed mixed results over time.[56] Other instruments showed comparable results.

The 1 trial that examined pharmacotherapy (n = 87) reported a 10-point reduction in the EPDS with fluoxetine after 12 weeks, compared with a 7-point reduction in those taking a placebo (P < .05). Results were similar for 2 other continuous measures of depression symptom severity, but this trial did not report a dichotomous remission-related outcome.

Figure 6. Diagnostic Accuracy of the Patient Health Questionnaire Relative to a Diagnostic Interview (Key Question 2)

Source	PHQ Version	Cutoff	Total No. of Patients	Population	No. (%) Positive by Reference Standard	Sensitivity (95% CI)	Specificity (95% CI)
Major depressive disorder							
Smith et al,[53] 2010	PHQ-2 (Likert)	3	213	Pregnant	13 (6.1)	0.77 (0.50-0.93)	0.59 (0.52-0.66)
Smith et al,[53] 2010	PHQ-2 (Likert)	4	213	Pregnant	13 (6.1)	0.62 (0.35-0.84)	0.79 (0.73-0.84)
Gjerdingen et al,[51] 2009b	PHQ-2 (Likert)	Any item >1	436	Postpartum	20 (4.6)	0.75 (0.54-0.90)	0.88 (0.85-0.91)
Gjerdingen et al,[51] 2009b	PHQ-2 (yes/no)	Any yes	438	Postpartum	20 (4.6)	1.00 (0.88-1.00)	0.62 (0.57-0.67)
Smith et al,[53] 2010	PHQ-8	10	213	Pregnant	13 (6.1)	0.77 (0.50-0.93)	0.62 (0.55-0.69)
Smith et al,[53] 2010	PHQ-8	11	213	Pregnant	13 (6.1)	0.77 (0.50-0.93)	0.68 (0.61-0.74)
Gjerdingen et al,[51] 2009b	PHQ-9	10	438	Postpartum	20 (4.6)	0.75 (0.54-0.90)	0.91 (0.88-0.93)
Major or minor depression							
Mann et al,[52] 2012	PHQ-2 (yes/no)	Any yes	126	Pregnant	17 (13.5)	1.00 (0.86-1.00)	0.68 (0.59-0.76)

PHQ indicates Patient Health Questionnaire. Error bars indicate 95% confidence interval.

Harms of Treatment

Key Question 5. What are the harms of treatment in pregnant and postpartum women who screen positive for depression in primary care?

Key Question 5b. What is the prevalence of other selected serious harms of treatment with antidepressants in the general (ie, not limited to primary care) population of pregnant and postpartum women?

The examination of harms of antidepressants was limited to second-generation agents: selective serotonin reuptake inhibitors (SSRIs), selective norepinephrine reuptake inhibitors, bupropion, nefazodone, trazodone, and mirtazapine. Ten of the included studies on harms of treatment for depression were of good quality, and 4 were of fair quality (**Table 4**, studies are ordered by study design, then by primary reported outcome). Of the trials that addressed benefits of treatment, which all involved screen-identified patients, only the trial of fluoxetine also reported on harms of treatment.[55] At 12 weeks of follow-up, 1 of 43 women (2.3%) taking fluoxetine and 3 of 44 women (6.8%) taking the placebo discontinued it due to adverse effects.

Considering studies not limited to women with screen-detected depression, a good-quality systematic review published in 2013[14] identified 15 observational studies providing evidence of the harms of antidepressants at unknown dosages in pregnant depressed women. The review included an additional 109 observational studies that provided evidence of the harms of antidepressants in pregnant women in whom depression status in either or both treatment groups was unknown. When available, data limited to depressed women were our focus.

An additional 12 fair- or good-quality large observational studies were identified that were published between 2012 and 2014 and that examined the harms of antidepressants in pregnant or postpartum women (n = 4 759 735).[73-84] Three were case-control studies[82-84]; the remaining were cohort studies that used national registers or administrative health data to examine exposures and outcomes retrospectively in women who had been pregnant. Five studies provided evidence of outcomes in pregnant women with known depression who were or were not exposed to antidepressants.[74-76,78,84] The remaining 7 studies compared outcomes in exposed vs unexposed pregnant women with unknown depression status, although most of these analyses adjusted for presence of depression[79] or conducted some analyses that were restricted to depressed women.[77,80,81]

Detailed results of the harms of treatment are shown in eTable 2 in the Supplement. There was evidence that use of some antidepressants during pregnancy, particularly SSRIs and venlafaxine, are associated with increased risk of preeclampsia, postpartum hemorrhage, and miscarriage as well as a number of adverse infant outcomes, including neonatal or postneonatal death, preterm birth, small for gestational age, neonatal seizures, serotonin withdrawal syndrome, neonatal respiratory distress, pulmonary hypertension, or major congenital malformations. The absolute increase in risk for most infant outcomes was very small, given the rarity of the events, and sometimes occurred only with higher levels of exposure. For example, a large retrospective cohort study reported a more than doubling of seizure occurrence in infants of depressed women who had been provided 3 or more prescription fills for antidepressants of any kind (but primarily SSRIs). However, the absolute risk remained quite small (0.66% among exposed infants vs 0.28% in unexposed infants; unadjusted OR, 2.39 [95% CI, 1.57-3.64]).[75] In that study, there was no similar association among women with 1 or 2 prescription fills for antidepressants.

More common outcomes showed potentially important absolute increases. One study in the 2013 review[14] reported neonatal respiratory distress among 7.8% of infants not exposed to SSRIs in utero, compared with 13.9% of exposed infants, and a pooled estimate combining 3 studies showed an increased odds of respiratory

Because non-CBT approaches, including fluoxetine, were highly variable in their effects and were limited by lack of replication, firm conclusions about those approaches could not be drawn.

Table 3. Study Characteristics of Trials of Benefits of Treatment (Key Question 4)

Source	Quality[a]	No. of Patients	Study Design	Intervention	Planned Follow-up, mo	Country	Age, Mean (Range), y	Race/Ethnicity, No. (%)	Weeks Postpartum or Gestational Week at Baseline
McGregor et al,[64] 2014	Fair	42	CCT[b]	CBT	4, 6	Canada	NR (≥16)	NR	22 (gestation)
Milgrom et al,[66] 2011	Fair	68	RCT	CBT	2	Australia	31.5 (NR)	NR	16 (postpartum)
Cooper et al,[56] 2003	Good	193	RCT	CBT (G1), psychodynamic (G2), or nondirective counseling (G3)	4, 9, 18	United Kingdom	27.7 (17-42)	NR	0 (postpartum)
Prendergast and Austin,[67] 2001	Fair	37	RCT	CBT	1.5, 8	Australia	32.2 (NR)	NR	10 (postpartum)
O'Mahen et al,[71] 2013	Fair	55	RCT	CBT	4	United States	27.0 (18-43)	Black: 32 (58.2) Hispanic: NR White: 17 (30.9)	31 (gestation)
Kozinszky et al,[63] 2012	Good	324	RCT	CBT-related	4.75	Hungary	27.3 (NR)	NR	27 (gestation)
Ammerman et al,[54] 2013	Fair	93	RCT	CBT-related	4.75, 7.75	United States	21.9 (16-37)	Black: 30 (32.2) Hispanic: 7 (7.5) White: 58 (62.4)	12 (postpartum)
Honey et al,[61] 2002	Fair	45	RCT	CBT-related	2, 8	United Kingdom	27.9 (NR)	NR	22 (postpartum)
Milgrom et al,[65] 2005	Fair	192	RCT	CBT (Coping with Depression course) (G1) or CBT-related (G2)	12	Australia	29.7 (NR)	NR	12 (postpartum)
Wiklund et al,[70] 2010	Fair	67	RCT	CBT	2.75	Sweden	NR (NR)	NR	0 (postpartum)
Holden et al,[60] 1989	Fair	55	RCT	Nondirective counseling	3.25	United Kingdom	26.2 (NR)	NR	10 (postpartum)
Wickberg and Hwang,[69] 1996	Fair	41	RCT	Nondirective counseling	1.5	Sweden	28.4 (NR)	NR	12 (postpartum)
Segre et al,[68] 2015	Fair	66	RCT	Nondirective counseling	2	United States	26.31 (≥14)	Black: 22 (33.3) Hispanic: 27 (40.9) White: 22 (33.3)	NR
Goodman et al,[58] 2015	Fair	42	RCT	Perinatal dyadic psychotherapy	3, 6	United States	30.7 (NR)	Black: NR Hispanic: 10 (23.8) White: 25 (59.5)	5 (postpartum)
Heh and Fu,[59] 2003	Fair	70	RCT	Information support	1.5	Taiwan	27.1 (20-35)	NR	6 (postpartum)
Horowitz et al,[62] 2001	Fair	122	RCT	Interaction coaching	1.5, 2.5	United States	31 (17-41)	Black: 9 (7.4) Hispanic: 9 (7.4) White: 84 (68.9)	6 (postpartum)
Gjerdingen et al,[57] 2009	Fair	39	RCT	Stepped care	9	United States	27.6 (≥16)	NR	0 (postpartum)
Appleby et al,[55] 1997	Fair	87	RCT	Fluoxetine and CBT	3	United Kingdom	25.3 (NR)	NR	7 (postpartum)

Abbreviations: CBT, cognitive behavioral therapy; CCT, controlled clinical trial; G1, G2, etc, group 1, group 2, etc; NR, not reported; RCT, randomized clinical trial.

[a] Quality assessed using criteria developed by the US Preventive Services Task Force.[16]

[b] Group assignment was nonrandom.

distress with SSRI exposure (pooled OR, 1.91; 95% CI, 1.63-2.24; I^2 = 0%).[14] As another example, a large US-based cohort study found development of preeclampsia among 8.9% of depressed women exposed to venlafaxine compared with 5.4% of unexposed women.[81] However, because these are observational studies, causality cannot be determined; it is not possible to control for all possible confounders related to depression, particularly the fact that women with more severe depression may be more likely to take antidepressants during pregnancy.
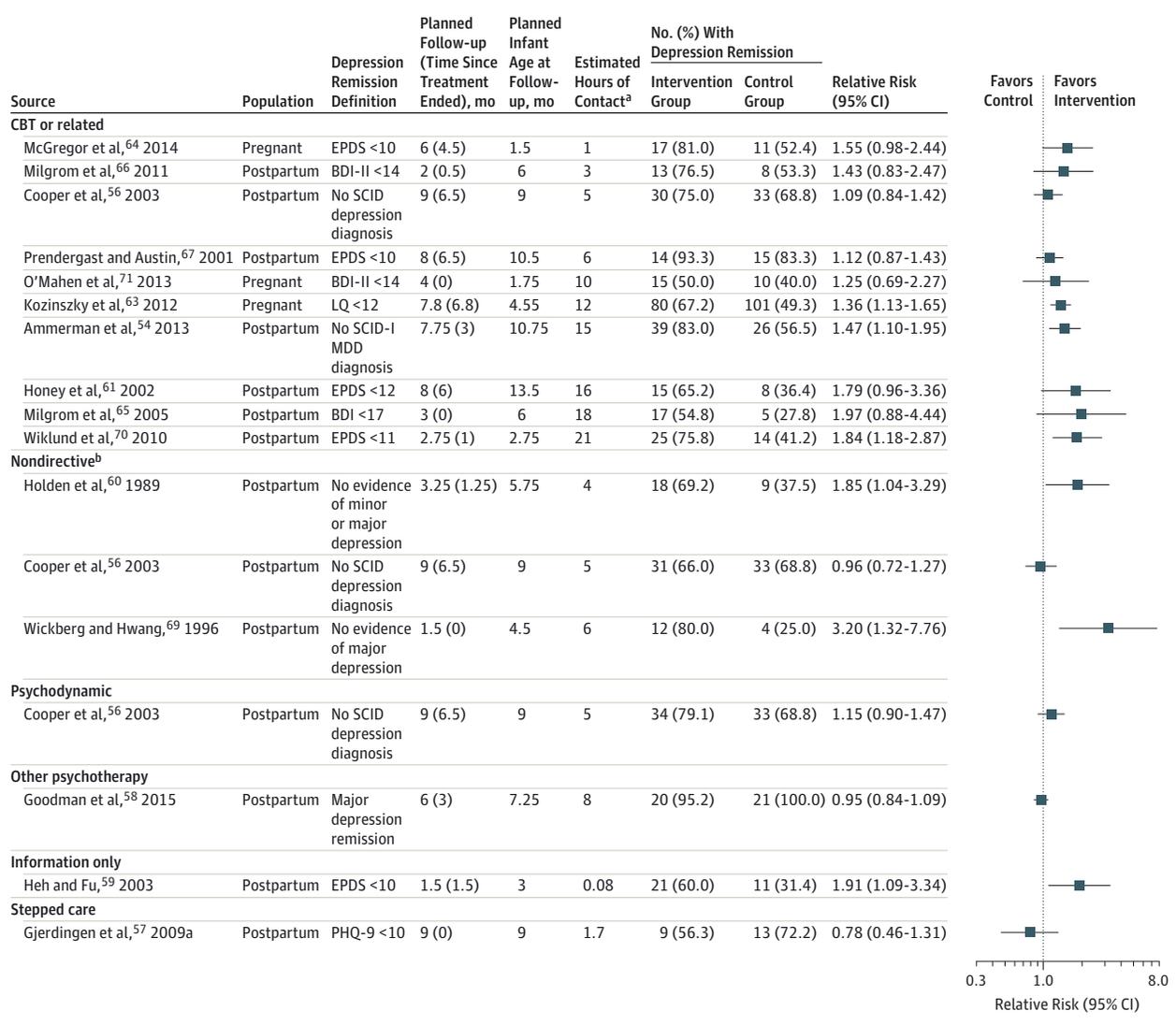
Discussion

We examined recent information on the benefits and harms of depression screening and treatment and the accuracy of selected screening instruments for pregnant and postpartum women to support the USPSTF updated recommendation on these topics.

Evidence suggested that programs to screen pregnant and postpartum women, with or without additional treatment-related supports, reduced the prevalence of depression and increased remission or treatment response (Table 5). Most of the screening trials included in this review provided treatment elements beyond screening, such as clinician training and supports, treatment protocols, or counseling with specially trained clinicians. Sensitivity of the English-language version of the EPDS was estimated to be approximately 0.80 and specificity approximately 0.90, using a cutoff of 13 to detect postpartum MDD. Further, evidence suggested that CBT improved depression in women with postpartum depression. In addition, the use of second-generation antidepressants during pregnancy may be associated with increased risk of some serious harms.

Evidence primarily focused on postpartum women, except for harms of antidepressants, but the little evidence among pregnant women suggested comparable effect with postpartum women. Im-

Figure 7. Benefits of Depression Treatment: Depression Remission or Response (Key Question 4)

Source	Population	Depression Remission Definition	Planned Follow-up (Time Since Treatment Ended), mo	Planned Infant Age at Follow-up, mo	Estimated Hours of Contact[a]	No. (%) With Depression Remission		Relative Risk (95% CI)
						Intervention Group	Control Group	
CBT or related								
McGregor et al,[64] 2014	Pregnant	EPDS <10	6 (4.5)	1.5	1	17 (81.0)	11 (52.4)	1.55 (0.98-2.44)
Milgrom et al,[66] 2011	Postpartum	BDI-II <14	2 (0.5)	6	3	13 (76.5)	8 (53.3)	1.43 (0.83-2.47)
Cooper et al,[56] 2003	Postpartum	No SCID depression diagnosis	9 (6.5)	9	5	30 (75.0)	33 (68.8)	1.09 (0.84-1.42)
Prendergast and Austin,[67] 2001	Postpartum	EPDS <10	8 (6.5)	10.5	6	14 (93.3)	15 (83.3)	1.12 (0.87-1.43)
O'Mahen et al,[71] 2013	Pregnant	BDI-II <14	4 (0)	1.75	10	15 (50.0)	10 (40.0)	1.25 (0.69-2.27)
Kozinszky et al,[63] 2012	Pregnant	LQ <12	7.8 (6.8)	4.55	12	80 (67.2)	101 (49.3)	1.36 (1.13-1.65)
Ammerman et al,[54] 2013	Postpartum	No SCID-I MDD diagnosis	7.75 (3)	10.75	15	39 (83.0)	26 (56.5)	1.47 (1.10-1.95)
Honey et al,[61] 2002	Postpartum	EPDS <12	8 (6)	13.5	16	15 (65.2)	8 (36.4)	1.79 (0.96-3.36)
Milgrom et al,[65] 2005	Postpartum	BDI <17	3 (0)	6	18	17 (54.8)	5 (27.8)	1.97 (0.88-4.44)
Wiklund et al,[70] 2010	Postpartum	EPDS <11	2.75 (1)	2.75	21	25 (75.8)	14 (41.2)	1.84 (1.18-2.87)
Nondirective[b]								
Holden et al,[60] 1989	Postpartum	No evidence of minor or major depression	3.25 (1.25)	5.75	4	18 (69.2)	9 (37.5)	1.85 (1.04-3.29)
Cooper et al,[56] 2003	Postpartum	No SCID depression diagnosis	9 (6.5)	9	5	31 (66.0)	33 (68.8)	0.96 (0.72-1.27)
Wickberg and Hwang,[69] 1996	Postpartum	No evidence of major depression	1.5 (0)	4.5	6	12 (80.0)	4 (25.0)	3.20 (1.32-7.76)
Psychodynamic								
Cooper et al,[56] 2003	Postpartum	No SCID depression diagnosis	9 (6.5)	9	5	34 (79.1)	33 (68.8)	1.15 (0.90-1.47)
Other psychotherapy								
Goodman et al,[58] 2015	Postpartum	Major depression remission	6 (3)	7.25	8	20 (95.2)	21 (100.0)	0.95 (0.84-1.09)
Information only								
Heh and Fu,[59] 2003	Postpartum	EPDS <10	1.5 (1.5)	3	0.08	21 (60.0)	11 (31.4)	1.91 (1.09-3.34)
Stepped care								
Gjerdingen et al,[57] 2009a	Postpartum	PHQ-9 <10	9 (0)	9	1.7	9 (56.3)	13 (72.2)	0.78 (0.46-1.31)

Favors Control ← → Favors Intervention

Relative Risk (95% CI): 0.3   1.0   8.0

BDI indicates Beck Depression Inventory; CBT, cognitive behavioral therapy; EPDS, Edinburgh Postnatal Depression Scale; LQ, Leverton Questionnaire; MDD, major depressive disorder; PHQ, Patient Health Questionnaire; RR, relative risk; SCID, Structured Clinical Interview for Depression. Error bars indicate 95% confidence interval.

[a] Hours of contact were estimated based on planned number and length of sessions.

[b] Nondirective therapy involves empathic, reflective listening rather than advice or direction in behavior change.

portant limitations to the evidence were noted for all bodies of evidence, including relatively small number of studies, few trials with good applicability to primary care in the United States, and many studies with very small study sizes.

The direct evidence of effects of screening for depression suggested that programs that include screening reduce the overall prevalence of depression and increase the likelihood of remission or treatment response in postpartum women. Results in pregnant women were consistent with postpartum women, although they came from only a single, smaller study. The direct (KQ1 and KQ1a) evidence base is relatively small (6 trials, most with fairly short follow-up) but included almost 12 000 women. Only 1 of these trials was conducted in the United States. Two trials provided minimal additional components beyond screening: one demonstrated reduced preva-

lence of depression[27] and the other increased response to treatment.[29] The results of this evidence report are consistent with 2 recent comprehensive reviews of depression identification in pregnant and postpartum women, which included overlapping (but not identical) evidence bases.[12,13] One review concluded that their included studies showed that using the EPDS had beneficial effects, but the authors could not disentangle the effects of using an identification strategy from the effects of subsequent interventions provided.[13] The other review concluded that screening was associated with modest improvement in depression across a variety of low-intensity interventions.[12]

One concern about the trials of screening programs is that 4 of the 6 studies did not exclude women who were previously known to be depressed. Because depression is often inadequately treated,[7,8]

Figure 8. Benefits of Depression Treatment: Depression Symptoms (Key Question 4)

| Source | Outcome | Intervention Group | | Control Group | | Mean Difference in Change From Baseline (95% CI) | Favors Intervention | Favors Control |
		No. of Patients	Mean Change From Baseline (95% CI)	No. of Patients	Mean Change From Baseline (95% CI)			
CBT or related								
McGregor et al,[64] 2014	EPDS	21	−6.22 (−8.13 to −4.31)	21	−3.76 (−5.67 to −1.85)	−2.46 (−4.98 to 0.06)		
Milgrom et al,[66] 2011	BDI-II	23	−20.5 (−25.18 to −15.82)	23	−16.9 (−21.58 to −12.22)	−3.60 (−9.69 to 2.49)		
Cooper et al,[56] 2003	EPDS	40	−5.1 (NA)	48	−3.2 (NA)	−1.90 (NA)		
Prendergast and Austin,[67] 2001	EPDS	15	−9.7 (−11.74 to −7.66)	18	−6.0 (−8.01 to −3.99)	−3.70 (−6.14 to −1.26)		
O'Mahen et al,[71] 2013	BDI-II	21	−14.74 (−18.73 to −10.75)	21	−3.17 (−7.16 to 0.82)	−11.57 (−16.06 to −7.08)		
Ammerman et al,[54] 2013	EPDS	47	−10.18 (−12.02 to −8.34)	46	−5.98 (−7.82 to −4.14)	−4.20 (−6.92 to −1.48)		
Honey et al,[61] 2002	EPDS	23	−6.8 (−8.75 to −4.85)	22	−2.32 (−4.27 to −0.37)	−4.48 (−7.70 to −1.26)		
Milgrom et al,[65] 2005	BDI	31	−8.5 (−10.97 to −6.03)	18	−1.1 (−3.65 to 1.45)	−7.40 (−11.41 to −3.39)		
Wiklund et al,[70] 2010	EPDS	33	−9.3 (NA)	34	-3.8 (NA)	−5.50 (NA)		
Nondirective[a]								
Segre et al,[68] 2015	EPDS	39	−6.85 (−8.57 to −5.13)	21	−3.96 (−5.73 to −2.19)	−2.89 (−5.74 to −0.04)		
Cooper et al,[56] 2003	EPDS	46	−4.1 (NA)	48	−3.2 (NA)	−0.90 (NA)		
Wickberg and Hwang,[69] 1996	MADRS	20	−8.7 (NA)	21	−2.4 (NA)	−6.30 (NA)		
Psychodynamic								
Cooper et al,[56] 2003	EPDS	43	−3.1 (NA)	48	−3.2 (NA)	0.10 (NA)		
Other psychotherapy								
Horowitz et al,[62] 2001	BDI-II	60	−5.23 (−7.41 to −3.05)	57	−3.73 (−5.92 to −1.54)	−1.50 (−4.22 to 1.22)		
Goodman et al,[58] 2015	EPDS	21	−7.62 (−9.15 to −6.09)	21	−6.09 (−7.62 to −4.56)	−1.53 (−3.74 to 0.68)		
Information only								
Heh and Fu,[59] 2003	EPDS	35	−5.7 (−7.04 to −4.36)	35	−4.2 (−5.54 to −2.86)	−1.50 (−3.10 to 0.10)		
Stepped care								
Gjerdingen et al,[57] 2009a	PHQ-9	16	−1.5 (−5.72 to 2.72)	18	−4.1 (−8.29 to 0.09)	2.60 (−2.43 to 7.63)		
Fluoxetine and CBT								
Appleby et al,[55] 1997	EPDS	43	−9.9 (−11.73 to −8.07)	44	−7 (−8.83 to −5.17)	−2.90 (−5.25 to −0.55)		

-20 -15 -10 -5 0 5 10
Mean Difference in Change From Baseline (95% CI)

Some studies did not provide sufficient data to calculate the 95% confidence interval; these are indicated by a data marker with no error bars on the forest plot and NA (not available) in the data columns. BDI indicates Beck Depression Inventory; CBT, cognitive behavioral therapy; EPDS, Edinburgh Postnatal Depression Scale; MADRS, Montgomery-Asberg Depression Rating Scale; PHQ, Patient Health Questionnaire. Error bars indicate 95% confidence interval.

[a] Nondirective therapy involves empathic, reflective listening rather than advice or direction in behavior change.

however, it may also be important for persons who are still depressed despite previous treatment efforts to be identified so their clinician can continue to help them until they are able to find a successful treatment. While this falls outside the traditional definition of screening, it is nevertheless a potentially important side benefit of depression screening programs. Further, depression screening presents an opportunity to query suicidal ideation among those who screen positive. While the USPSTF has not recommended routine screening for suicide risk, they did note that "primary care clinicians should be aware of psychiatric problems in their patients and should consider asking these patients about suicidal ideation and referring them" for treatment.[85] Thus, pragmatically, identifying incompletely treated patients could be considered an added benefit of routine depression screening, although this falls more in the realm of depression management than prevention through early detection, which is the traditional definition of screening.

In addition to the direct evidence, we also considered indirect evidence on screening accuracy and the benefits and harms of treatment for depression in pregnant and postpartum women. While the range of sensitivities and specificities were quite wide for the English-language version of the EPDS, the largest studies and the study most

applicable to the US health care system reported sensitivities around 0.80 and specificities of 0.87 and higher at a cutoff of 13 to detect MDD, primarily in postnatal women. This body of evidence was fairly large (K = 23), but only 8 studies addressed the English-language version of the EPDS and only 2 of these were conducted in the United States. Furthermore, the literature on the English-language version was limited by small study sizes. However, the broad use of the EPDS and the relatively acceptable results despite the various languages and country populations can be seen as reassuring for its applicability to a diverse US pregnant and postpartum population. Evidence on the accuracy of the PHQ for pregnant and postpartum women was very limited. Other reviews drew similar conclusions and included additional screening instruments.[12,13]

Cognitive behavioral therapy and related behaviorally based approaches reduced the symptoms of postpartum depression and increased the likelihood of remission compared with usual care among depressed pregnant and postpartum women identified through screening. There were insufficient data to determine whether the use of other treatment modalities was beneficial in either pregnant or postpartum women, including only a single small trial of pharmacotherapy. Results were mixed in the studies conducted in the

Table 4. Study Characteristics of Studies of Harms of Treatment (Key Question 5)

Source	Quality[a]	No. of Patients	Study Design	Exposure	Planned Follow-up, mo	Country	Age, Mean (Range), y	Race/Ethnicity, No. (%)	Weeks Postpartum or Gestational Week at Baseline
Appleby et al,[55] 1997	Fair	87	RCT	Fluoxetine and CBT	3	United Kingdom	25.3 (NR)	NR	7 (postpartum)
Palmsten et al,[81] 2013a	Good	85 326	Cohort	Second-generation antidepressants	NR	United States	23.7 (12-55)	Black: 19 220 (22.5) Hispanic: 10 045 (11.8) White: 50 224 (58.9)	NR
Palmsten et al,[80] 2013b	Good	102 722	Cohort	Second-generation antidepressants	NR	United States	23.5 (12-55)	Black: 19 719 (19.2) Hispanic: 10 624 (10.3) White: 65 611 (63.9)	NR
Lupattelli et al,[79] 2014	Fair	57 279	Cohort	Second-generation antidepressants	NR	Norway	NR (NR)	NR	NR
Andersen et al,[73] 2014	Good	1 279 840	Cohort	SSRIs	NR	Denmark	NR (NR)	NR	NR
Kjaersgaard et al,[78] 2013	Good	1 005 319	Cohort	Second-generation antidepressants	NR	Denmark	30.2 (NR)	NR	NR
Hayes et al,[75] 2012	Good	228 876	Cohort	Second-generation antidepressants	NA	United States	23.2 (15-44)	Black: 95 503 (41.7) Hispanic: NR White: 127 592 (55.7)	NR
Jensen et al,[77] 2013	Good	673 853	Cohort	Second-generation antidepressants	NR	Denmark	29 (NR)	NR	NR
Ban et al,[74] 2014	Good	349 127	Cohort	SSRIs	NA	United Kingdom	30 (14-45)	NR	NR
Polen et al,[82] 2013	Fair	27 045	Case-control	Venlafaxine	NR	United States	NR (NR)	Black: NR Hispanic: NR White: 15 861 (58.6)	NR
Yazdy et al,[84] 2014	Fair	2624	Case-control	SSRIs	12	United States	NR (NR)	Black: 414 (15.8) Hispanic: 311 (11.9) White: 1757 (67)	NR
Louik et al,[83] 2014	Good	16 524	Case-control	SSRIs	6	United States	NR (NR)	NR	NR
Huybrechts et al,[76] 2014	Good	931 259	Cohort	Second-generation antidepressants	NR	United States	24.0 (NR)	Black: 318 807 (34.2) Hispanic: 168 462 (18.1) White: 373 242 (40.1)	NR
McDonagh et al,[14] 2013	Good	NR	Systematic review, included 124 studies reporting harms of second-generation antidepressants	Antidepressants	No minimum follow-up	Economically advanced	No restrictions	No restrictions	Pregnancy through 52 weeks postpartum

Abbreviations: CBT, cognitive behavioral therapy; NA, not applicable; NR, not reported; RCT, randomized clinical trial; SSRIs, selective serotonin reuptake inhibitors.

[a] Quality assessed using criteria developed by the US Preventive Services Task Force[16] for randomized clinical trials, the Newcastle-Ottawa Scale[18] for observational studies, or A Measurement Tool to Assess Systematic Reviews (AMSTAR) for systematic reviews.[19]

United States: 1 found benefit at both the 4.5- and 7.5-month follow-ups,[54] but the other did not find statistically significant group differences.[71] Effect sizes in CBT trials were very similar between the 2 trials of pregnant women and the trials of postpartum women. Although not limited to studies of women with screen-detected depression, other reviews have also concluded that behaviorally based treatment of depression is beneficial during the postpartum period and that data are lacking on the use of antidepressants.[86,87]

The generalizability of clinical trial treatment results may be reduced by restrictive inclusion and exclusion criteria. For example,

excluding persons with greater disease severity and comorbidities may overestimate the effects of treatment.[88,89] The treatment studies in our review generally excluded women with the greatest disease severity (such as history of psychosis, current suicidal ideation, and need for crisis management). Furthermore, bias related to small sample sizes has been reported in the psychotherapy literature[90,91] and was a possible issue in our included studies, although one of those reports suggested that the statistical significance of pooled results was only minimally affected by this bias.[91] Limiting trials to those that used screening for case-finding, rather

Table 5. Summary of Evidence: Screening and Treatment for Depression in Pregnant and Postpartum Women

Key Question Topic	No. of Studies	No. of Participants	Study Design	Summary of Findings (Including Consistency and Precision)	Applicability	Limitations (Includes Reporting Bias)	Overall Study Quality
Key question 1, 1a: Benefits of screening	6	11 869	5 RCTs, 1 CCT[a]	Trials reported reasonably consistent (yet imprecise) relative decreases in prevalence of depression (18%-59%) with depression screening (± additional components) and increases in remission/treatment response (21%-33%) in women with depressive symptoms at baseline. Two interventions focused on screening with minimal additional supports or counseling reduced depression in the near-term (≤4 mo). Four interventions with additional clinician supports or counseling consistently improved depression outcomes.	All studies were conducted in maternal health or other primary care settings; however, only 1 was conducted in the United States, and 3 involved home visits, which are rarely used in the United States.	A small number of studies were included; wide range of intervention approaches with no replication of any interventions were included that reported minimal descriptions of the samples (eg, age, race/ethnicity, previous depression); minimal information on the role of screening in the beneficial results was available. Reporting bias not detected. Quality: 1 good-quality trial, 5 fair-quality trials	Fair
Key question 2: Performance characteristics of the EPDS	23 (8 English-language version)	5398 (1905 English-language version)	Diagnostic accuracy	For detecting MDD in the first 3 mo postpartum, sensitivity of the English-language EPDS was estimated to be approximately 0.80 and specificity approximately 0.90, with a cutoff of 13. Evidence was reasonably consistent and reasonably precise. In a population with 10% MDD prevalence, PPV was estimated at 47% for detecting MDD. Using a cutoff of 10 for detecting depressive disorders, including minor depression, evidence was somewhat inconsistent and imprecise: sensitivity was estimated between 0.63 to 0.84, specificity between 0.80 and 0.90. PPVs were 43% and 50% at these sensitivity levels and specificity of 0.85 in a population with 15% prevalence of depressive disorders.	Only 2 studies of English-language version were conducted in the United States, but study with best applicability reported relatively good performance characteristics.	Limited data on the English-language version, much of it collected 15-25 y ago; studies generally had small No. of observations; training and fidelity associated with the reference standard were rarely reported; the English-language version of 2 studies did not report interval between EPDS and reference test. Reporting bias is possible, based on use of optimal cutoffs, but most English-version EPDS studies reported on commonly used cutoffs of 10 and 13. Quality: 1 good-quality study, 22 fair-quality studies	Fair
Key question 2: Performance characteristics of the PHQ	3	777	Diagnostic accuracy	Sensitivity and specificity were fairly wide-ranging over different versions of the PHQ, different scoring methods, different cutoffs, and different comparators (MDD vs major or minor depression). Sensitivities ranged from 0.62 to 1.00 and specificities ranged from 0.59 to 0.91. Evidence was inconsistent and imprecise.	Two of 3 studies were conducted in the United States, within past 5 y, including 18%-20% black participants, but other racial/ethnic minority groups were not represented.	A small number of studies were included, with no replication for any specific version, scoring method, cutoff, and comparator; studies had small samples, resulting in ≤5 false negatives. Reporting bias was possible, as 1 of the 3 studies reported optimal cut points based on receiver operating curve. Quality: 3 fair-quality studies	Fair
Key question 3: Harms of screening	1	462	1 RCT	One of the included studies reported no adverse effects. No additional data on harms of screening beyond trials of screening's benefit was identified. No evidence of paradoxical deleterious effects was observed.	Only a single study conducted in Hong Kong.	No evidence directly examined harms. Quality: 1 good-quality study	Good
Key question 4: Benefits of treatment	18	1638	17 RCTs, 1 CCT[a]	CBT and related therapeutic approaches showed reasonably consistent and somewhat precise relative increases in likelihood of remission: RR, 1.34 (95% CI, 1.19-1.50) in the short term (<8 mo) and reduced symptom severity. Larger effects were generally associated with greater contact hours; however, contact time was confounded with other important sources of heterogeneity. Data were insufficient to evaluate other treatment approaches, including stepped care (K = 1) and fluoxetine (K = 1).	Limited to studies of screen-detected depression conducted in or recruited from primary care, but only 3 were conducted in the United States; little information was available about population characteristics, particularly racial/ethnic background.	Most trials were small and with ≥1 methodological limitations. Reporting bias was possible, as a variety of definitions were used for remission, and it is possible that the definition with largest effect was presented in some studies. Quality: 2 good-quality studies, 16 fair-quality studies	Fair
Key question 5: Harms of treatment (behaviorally based)	0	NA	NA	None of the studies reported on adverse events or other specific harms. No additional data addressing harms of treatment beyond trials of treatment's benefit were found. No evidence of paradoxical deleterious effects was observed.	NA	No evidence directly examined harms.	NA

(continued)

Table 5. Summary of Evidence: Screening and Treatment for Depression in Pregnant and Postpartum Women (continued)

Key Question Topic	No. of Studies	No. of Participants	Study Design	Summary of Findings (Including Consistency and Precision)	Applicability	Limitations (Includes Reporting Bias)	Overall Study Quality
Key question 5: Harms of treatment (antidepressants)	14	4 759 822 (excluding studies in the SER)	1 SER, 1 RCT, 9 large cohort studies, 3 large case-control studies	Second-generation antidepressants were associated with a higher risk of some serious adverse effects with a consistent and reasonably precise effect for most outcomes. Positive associations between antidepressants and harms for preeclampsia (venlafaxine), postpartum hemorrhage (SSRI for ≥60 d exposure, SNRI), miscarriage (SSRI first trimester, SNRI), neonatal or postneonatal death (SSRI), preterm birth (SSRI in first and second trimesters, SNRI), small for gestational age (SSRI), infant seizures (SSRI), serotonin withdrawal syndrome (SSRI, SNRI), neonatal respiratory distress (SSRI), pulmonary hypertension (SSRI, particularly late in pregnancy), major malformations (SSRI), and cardiac malformations (paroxetine, venlafaxine, bupropion). Negative studies are not summarized here: for most studies showing positive associations, other studies revealed no associations.	Approximately one-third of studies were conducted in the United States and the majority of the others were conducted in Europe.	Almost all evidence was in observational studies rather than trials, so causality could not be clearly determined; many studies compared harms in groups of women with unknown depression status, exaggerating the potential for confounding by indication; no data were available to examine harms by dose, although some did examine harms by length of exposure; most studies used pharmacy fills to examine exposure but did not verify whether women were taking antidepressants as prescribed. Reporting bias was unlikely as most studies included a limited number of outcomes and used medical records to ascertain exposure and outcomes. Quality: 1 good-quality SER, 1 fair-quality randomized trial, 9 good-quality cohort and case-control studies, 3 fair-quality cohort and case-control studies	Good

Abbreviations: CBT, cognitive behavioral therapy; EPDS, Edinburgh Postnatal Depression Scale; MDD, major depressive disorder; K, No. of studies; NA, not applicable; PPV, positive predictive value; RCT, randomized clinical trial; RR, relative risk; SER, systematic evidence review; SNRI, selective norepinephrine reuptake inhibitor; SSRI, selective serotonin reuptake inhibitor.

a In the controlled clinical trial (CCT), group assignment was nonrandom.

than including trials with referral-based and self-selected entry, likely limited the degree of overestimation in this review. Trials that recruit through screening generally have smaller effect sizes than those enrolling self-selected volunteers from broad-based community recruitment through media ads and other means.[92]

There was very little evidence related to the harms of behaviorally based treatment in pregnant and postpartum women and no evidence that these treatments could be harmful. Data on the harms of antidepressant use in postpartum women were insufficient, with only a single small 12-week trial of fluoxetine. Evidence on harms of antidepressants was almost entirely limited to pregnant women, in contrast to the other bodies of literature in this review. The imbalance of evidence of benefits and harms on antidepressants is likely due to the difficulty of conducting randomized clinical trials in pregnant and breastfeeding women, yet observational studies are feasible and have the best chance of identifying rare harms, for which studies with very large sample sizes are needed. Results did suggest possible risk of harm. While these data were limited to observational studies, many were very large, population-based studies that controlled for depression status in some way. Nevertheless, causality could not be definitively determined from these studies. Pragmatically, CBT is not an option for every depressed woman because some will not want such therapy, some will not have access to trained CBT clinicians, and some may not respond fully to CBT treatment. For women with more severe depression who are not interested in or able to participate in CBT, further research is needed on the risks vs benefits of antidepressant therapy in order to guide shared decision making.

The evidence we included in this analysis targeted primarily postpartum women (except for harms of antidepressants, which pertained to prenatal women only). However, the little evidence found regarding pregnant women suggested comparable effects with postpartum women for benefits of screening, accuracy of the EPDS, and benefits of CBT.

Important limitations to the evidence reviewed were noted for all bodies of evidence, including the number of studies, study size, inconsistency in the specific outcomes reported, and applicability of trials to primary care in the United States. In addition, the scope of this review excluded areas of research that may be pertinent to depression screening in pregnant and postpartum women. For example, examination of screening instrument accuracy was limited to only 2 instruments, the PHQ and the EPDS. Nontrial evidence related to harms of screening or behaviorally based treatment was excluded, although the risks of these interventions are likely to be minimal. Furthermore, evidence of using antidepressants was limited to a prespecified list of serious harms; we did not examine other harms that, even if not life-threatening, might be clinically important, such as developmental outcomes (eg, autism) and behavioral outcomes (eg, crying or sleeping issues) in infants. Also, we did not review the effectiveness in pregnant and postpartum women of interventions that are widely available but generally offered outside of the health care setting (eg, yoga, exercise, or light therapy). As the scope of this review was limited to adults, studies focused on pregnant or postpartum females younger than 18 years were not included. In addition, a potential methodological limitation is reliance on other reviews to identify evidence for some years and, for harms of antidepressants, reliance on the synthesized work of previous reviewers. Although we assessed the pertinent sections of these

reviews' methods as being of good quality, it is nonetheless possible that they missed or incorrectly interpreted evidence.

Conclusions

The direct evidence suggested that screening pregnant and postpartum women for depression may reduce depressive symptoms in women with depression and reduce the prevalence of depression in a given population, particularly in the presence of additional treatment supports (eg, treatment protocols, care management, and availability of specially trained depression care clinicians). The indirect evidence showed that screening instruments can identify pregnant and postpartum women who need further evaluation and may need treatment. The only identified harm of treatment was the use of antidepressants during pregnancy, although the absolute risk of harm appeared to be small and CBT appeared to be an effective alternative treatment approach.

REFERENCES

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

2. Hoertel N, López S, Peyre H, et al. Are symptom features of depression during pregnancy, the postpartum period and outside the peripartum period distinct? results from a nationally representative sample using item response theory (IRT). *Depress Anxiety*. 2015;32(2):129-140.

3. Stein A, Gath DH, Bucher J, Bond A, Day A, Cooper PJ. The relationship between post-natal depression and mother-child interaction. *Br J Psychiatry*. 1991;158:46-52.

4. van Wijngaarden B, Schene AH, Koeter MW. Family caregiving in depression: impact on caregivers' daily life, distress, and help seeking. *J Affect Disord*. 2004;81(3):211-222.

5. Kersten-Alvarez LE, Hosman CM, Riksen-Walraven JM, van Doesum KT, Smeekens S, Hoefnagels C. Early school outcomes for children of postpartum depressed mothers: comparison with a community sample. *Child Psychiatry Hum Dev*. 2012;43(2):201-218.

6. Beardslee WR, Versage EM, Gladstone TR. Children of affectively ill parents: a review of the past 10 years. *J Am Acad Child Adolesc Psychiatry*. 1998;37(11):1134-1141.

7. US Preventive Services Task Force. Screening for depression in adults: US Preventive Services Task Force Recommendation Statement. *Ann Intern Med*. 2009;151(11):784-792.

8. 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):765-776.

9. O'Connor EA, Whitlock EP, Gaynes BN, Beil TL. *Screening for Depression in Adults and Older Adults in Primary Care: An Updated Systematic Review: Evidence Synthesis No. 75 [AHRQ Publication No. 10-05143-EF-1]*. Rockville, MD: Agency for Healthcare Research and Quality; 2009.

10. O'Connor E, Rossom RC, Henninger M, et al. *Screening for Depression in Adults: An Updated Systematic Evidence Review for the US Preventive Services Task Force: Evidence Synthesis No. 128 [AHRQ Publication No. 14-05208-EF-1]*. Rockville, MD: Agency for Healthcare Research and Quality; 2016.

11. Whitlock EP, Lin JS, Chou R, Shekelle P, Robinson KA. Using existing systematic reviews in complex systematic reviews. *Ann Intern Med*. 2008;148(10):776-782.

12. Myers ER, Aubuchon-Endsley N, Bastian LA, et al. *Efficacy and Safety of Screening for Postpartum Depression [Comparative Effectiveness Review No. 106]*. Rockville, MD: Agency for Healthcare Research and Quality; 2013.

13. 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; 13(36):1-145,147-230.

14. McDonagh M, Matthews A, Phillipi C, et al. *Treatment of Depression During Pregnancy and the Postpartum Period*. Rockville, MD: Agency for Healthcare Research and Quality; 2013.

15. The Rise of the South: Human Progress in a Diverse World [Human Development Report 2013]. United Nations Development Programme. http://hdr.undp.org/sites/default/files/reports/14 /hdr2013_en_complete.pdf. Accessed December 28, 2015.

16. Harris RP, Helfand M, Woolf SH, et al; Methods Work Group, Third US Preventive Services Task Force. Current methods of the US Preventive Services Task Force: a review of the process. *Am J Prev Med*. 2001;20(3)(suppl):21-35.

17. Whiting PF, Rutjes AW, 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):529-536.

18. Wells GA, Shea B, O'Connell D, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. http://www.ohri.ca/programs/clinical _epidemiology/oxford.asp. Accessed January 21, 2014.

19. Shea BJ, Grimshaw JM, Wells GA, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol*. 2007;7:10.

20. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;7(3):177-188.

21. Knapp G, Hartung J. Improved tests for a random effects meta-regression with a single covariate. *Stat Med*. 2003;22(17):2693-2710.

22. Morrell CJ, Warner R, Slade P, et al. Psychological interventions for postnatal

depression: cluster randomised trial and economic evaluation: the PoNDER trial. *Health Technol Assess*. 2009;13(30):iii-iv,xi-xiii,1-153.

23. Cox JL, Chapman G, Murray D, Jones P. Validation of the Edinburgh Postnatal Depression Scale (EPDS) in non-postnatal women. *J Affect Disord*. 1996;39(3):185-189.

24. Garcia-Esteve L, Ascaso C, Ojuel J, Navarro P. Validation of the Edinburgh Postnatal Depression Scale (EPDS) in Spanish mothers. *J Affect Disord*. 2003;75(1):71-76.

25. Begg CB, Greenes RA. Assessment of diagnostic tests when disease verification is subject to selection bias. *Biometrics*. 1983;39(1):207-215.

26. Glavin K, Smith L, Sørum R, Ellefsen B. Redesigned community postpartum care to prevent and treat postpartum depression in women: a one-year follow-up study. *J Clin Nurs*. 2010;19(21-22):3051-3062.

27. Leung SS, Leung C, Lam TH, et al. Outcome of a postnatal depression screening programme using the Edinburgh Postnatal Depression Scale: a randomized controlled trial. *J Public Health (Oxf)*. 2011;33(2):292-301.

28. MacArthur C, Winter HR, Bick DE, et al. Effects of redesigned community postnatal care on women's health 4 months after birth: a cluster randomised controlled trial. *Lancet*. 2002;359 (9304):378-385.

29. Wickberg B, Tjus T, Hwang P. Using the EPDS in routine antenatal care in Sweden: a naturalistic study. *J Reprod Infant Psychol*. 2005;23(1):33-41.

30. Yawn BP, Dietrich AJ, Wollan P, et al; TRIPPD practices. TRIPPD: a practice-based network effectiveness study of postpartum depression screening and management. *Ann Fam Med*. 2012;10 (4):320-329.

31. Adouard F, Glangeaud-Freudenthal NM, Golse B. Validation of the Edinburgh Postnatal Depression Scale (EPDS) in a sample of women with high-risk pregnancies in France. *Arch Womens Ment Health*. 2005;8(2):89-95.

32. 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):242-250.

33. Benvenuti P, Ferrara M, Niccolai C, Valoriani V, Cox JL. The Edinburgh Postnatal Depression Scale: validation for an Italian sample. *J Affect Disord*. 1999;53(2):137-141.

34. Bunevicius A, Kusminskas L, Bunevicius R. Validation of the Lithuanian version of the Edinburgh Postnatal Depression Scale. *Medicina (Kaunas)*. 2009;45(7):544-548.

35. Bunevicius A, Kusminskas L, Pop VJ, Pedersen CA, Bunevicius R. Screening for antenatal depression with the Edinburgh Depression Scale. *J Psychosom Obstet Gynaecol*. 2009;30(4):238-243.

36. Carpiniello B, Pariante CM, Serri F, Costa G, Carta MG. Validation of the Edinburgh Postnatal Depression Scale in Italy. *J Psychosom Obstet Gynaecol*. 1997;18(4):280-285.

37. Chen H, Bautista D, Ch'ng YC, Li W, Chan E, Rush AJ. Screening for postnatal depression in Chinese-speaking women using the Hong Kong translated version of the Edinburgh Postnatal Depression Scale. *Asia Pac Psychiatry*. 2013;5(2): E64-E72.

38. Guedeney N, Fermanian J. Validation study of the French version of the Edinburgh Postnatal Depression Scale (EPDS): new results about use and psychometric properties. *Eur Psychiatry*. 1998; 13(2):83-89.

39. Harris B, Huckle P, Thomas R, Johns S, Fung H. The use of rating scales to identify post-natal depression. *Br J Psychiatry*. 1989;154:813-817.

40. Lee DT, Yip AS, Chiu HF, Leung TY, Chung TK. Screening for postnatal depression: are specific instruments mandatory? *J Affect Disord*. 2001; 63(1-3):233-238.

41. 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): 279-296.

42. Tandon SD, Cluxton-Keller F, Leis J, Le HN, Perry DF. A comparison of three screening tools to identify perinatal depression among low-income African American women. *J Affect Disord*. 2012;136 (1-2):155-162.

43. Teng HW, Hsu CS, Shih SM, Lu ML, Pan JJ, Shen WW. Screening postpartum depression with the Taiwanese version of the Edinburgh Postnatal Depression scale. *Compr Psychiatry*. 2005;46(4): 261-265.

44. Töreki A, Andó B, Keresztúri A, et al. The Edinburgh Postnatal Depression Scale: translation and antepartum validation for a Hungarian sample. *Midwifery*. 2013;29(4):308-315.

45. Töreki A, Andó B, Dudas RB, et al. Validation of the Edinburgh Postnatal Depression Scale as a screening tool for postpartum depression in a clinical sample in Hungary. *Midwifery*. 2014;30(8): 911-918.

46. Yamashita H, Yoshida K, Nakano H, Tashiro N. Postnatal depression in Japanese women: detecting the early onset of postnatal depression by closely monitoring the postpartum mood. *J Affect Disord*. 2000;58(2):145-154.

47. Alvarado R, Jadresic E, Guajardo V, Rojas G. First validation of a Spanish-translated version of the Edinburgh Postnatal Depression Scale (EPDS) for use in pregnant women: a Chilean study. *Arch Womens Ment Health*. 2015;18(4):607-612.

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

49. Clarke PJ. Validation of two postpartum depression screening scales with a sample of First Nations and Métis women. *Can J Nurs Res*. 2008; 40(1):113-125.

50. Felice E, Saliba J, Grech V, Cox J. Validation of the Maltese version of the Edinburgh Postnatal Depression Scale. *Arch Womens Ment Health*. 2006;9(2):75-80.

51. Gjerdingen D, Crow S, McGovern P, Miner M, Center B. Postpartum depression screening at well-child visits: validity of a 2-question screen and the PHQ-9. *Ann Fam Med*. 2009;7(1):63-70.

52. Mann R, Adamson J, Gilbody SM. Diagnostic accuracy of case-finding questions to identify perinatal depression. *CMAJ*. 2012;184(8):E424-E430.

53. Smith MV, Gotman N, Lin H, Yonkers KA. Do the PHQ-8 and the PHQ-2 accurately screen for

depressive disorders in a sample of pregnant women? *Gen Hosp Psychiatry*. 2010;32(5):544-548.

54. Ammerman RT, Putnam FW, Altaye M, Stevens J, Teeters AR, Van Ginkel JB. A clinical trial of in-home CBT for depressed mothers in home visitation. *Behav Ther*. 2013;44(3):359-372.

55. Appleby L, Warner R, Whitton A, Faragher B. A controlled study of fluoxetine and cognitive-behavioural counselling in the treatment of postnatal depression. *BMJ*. 1997;314(7085):932-936.

56. Cooper PJ, Murray L, Wilson A, Romaniuk H. Controlled trial of the short- and long-term effect of psychological treatment of post-partum depression: I, Impact on maternal mood. *Br J Psychiatry*. 2003;182:412-419.

57. Gjerdingen D, Crow S, McGovern P, Miner M, Center B. Stepped care treatment of postpartum depression: impact on treatment, health, and work outcomes. *J Am Board Fam Med*. 2009;22(5):473-482.

58. Goodman JH, Prager J, Goldstein R, Freeman M. Perinatal Dyadic Psychotherapy for postpartum depression: a randomized controlled pilot trial. *Arch Womens Ment Health*. 2015;18(3):493-506.

59. Heh SS, Fu YY. Effectiveness of informational support in reducing the severity of postnatal depression in Taiwan. *J Adv Nurs*. 2003;42(1):30-36.

60. Holden JM, Sagovsky R, Cox JL. Counselling in a general practice setting: controlled study of health visitor intervention in treatment of postnatal depression. *BMJ*. 1989;298(6668):223-226.

61. Honey KL, Bennett P, Morgan M. A brief psycho-educational group intervention for postnatal depression. *Br J Clin Psychol*. 2002; 41(pt 4):405-409.

62. Horowitz JA, Bell M, Trybulski J, et al. Promoting responsiveness between mothers with depressive symptoms and their infants. *J Nurs Scholarsh*. 2001;33(4):323-329.

63. Kozinszky Z, Dudas RB, Devosa I, et al. Can a brief antepartum preventive group intervention help reduce postpartum depressive symptomatology? *Psychother Psychosom*. 2012;81 (2):98-107.

64. McGregor M, Coghlan M, Dennis CL. The effect of physician-based cognitive behavioural therapy among pregnant women with depressive symptomatology: a pilot quasi-experimental trial. *Early Interv Psychiatry*. 2014;8(4):348-357.

65. Milgrom J, Negri LM, Gemmill AW, McNeil M, Martin PR. A randomized controlled trial of psychological interventions for postnatal depression. *Br J Clin Psychol*. 2005;44(pt 4):529-542.

66. Milgrom J, Holt CJ, Gemmill AW, et al. Treating postnatal depressive symptoms in primary care: a randomised controlled trial of GP management, with and without adjunctive counselling. *BMC Psychiatry*. 2011;11:95.

67. Prendergast J, Austin MP. Early childhood nurse-delivered cognitive behavioural counselling for post-natal depression. *Australas Psychiatry*. 2001;9(3):255-259.

68. Segre LS, Brock RL, O'Hara MW. Depression treatment for impoverished mothers by point-of-care providers: a randomized controlled trial. *J Consult Clin Psychol*. 2015;83(2):314-324.

69. Wickberg B, Hwang CP. Counselling of postnatal depression: a controlled study on a population based Swedish sample. *J Affect Disord*. 1996;39(3):209-216.

70. Wiklund I, Mohlkert P, Edman G. Evaluation of a brief cognitive intervention in patients with signs of postnatal depression: a randomized controlled trial. *Acta Obstet Gynecol Scand*. 2010;89(8):1100-1104.

71. O'Mahen H, Himle JA, Fedock G, Henshaw E, Flynn H. A pilot randomized controlled trial of cognitive behavioral therapy for perinatal depression adapted for women with low incomes. *Depress Anxiety*. 2013;30(7):679-687.

72. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Earlbaum Assoc; 1988.

73. Andersen JT, Andersen NL, Horwitz H, Poulsen HE, Jimenez-Solem E. Exposure to selective serotonin reuptake inhibitors in early pregnancy and the risk of miscarriage. *Obstet Gynecol*. 2014; 124(4):655-661.

74. Ban L, Gibson JE, West J, et al. Maternal depression, antidepressant prescriptions, and congenital anomaly risk in offspring: a population-based cohort study. *BJOG*. 2014;121 (12):1471-1481.

75. Hayes RM, Wu P, Shelton RC, et al. Maternal antidepressant use and adverse outcomes: a cohort study of 228,876 pregnancies [published correction appears in *Am J Obstet Gynecol*. 2013;208(4):326]. *Am J Obstet Gynecol*. 2012;207(1):49.e1-49.e9.

76. Huybrechts KF, Palmsten K, Avorn J, et al. Antidepressant use in pregnancy and the risk of cardiac defects. *N Engl J Med*. 2014;370(25):2397-2407.

77. Jensen HM, Grøn R, Lidegaard O, Pedersen LH, Andersen PK, Kessing LV. The effects of maternal depression and use of antidepressants during pregnancy on risk of a child small for gestational age. *Psychopharmacology (Berl)*. 2013;228(2):199-205.

78. Kjaersgaard MI, Parner ET, Vestergaard M, et al. Prenatal antidepressant exposure and risk of spontaneous abortion: a population-based study. *PLoS One*. 2013;8(8):e72095.

79. Lupattelli A, Spigset O, Koren G, Nordeng H. Risk of vaginal bleeding and postpartum hemorrhage after use of antidepressants in pregnancy: a study from the Norwegian Mother and Child Cohort Study. *J Clin Psychopharmacol*. 2014; 34(1):143-148.

80. Palmsten K, Hernández-Díaz S, Huybrechts KF, et al. Use of antidepressants near delivery and risk of postpartum hemorrhage: cohort study of low income women in the United States. *BMJ*. 2013; 347:f4877.

81. Palmsten K, Huybrechts KF, Michels KB, et al. Antidepressant use and risk for preeclampsia. *Epidemiology*. 2013;24(5):682-691.

82. Polen KN, Rasmussen SA, Riehle-Colarusso T, Reefhuis J; National Birth Defects Prevention Study. Association between reported venlafaxine use in early pregnancy and birth defects: National Birth Defects Prevention Study, 1997-2007. *Birth Defects Res A Clin Mol Teratol*. 2013;97(1):28-35.

83. Louik C, Kerr S, Mitchell AA. First-trimester exposure to bupropion and risk of cardiac malformations. *Pharmacoepidemiol Drug Saf*. 2014; 23(10):1066-1075.

84. Yazdy MM, Mitchell AA, Louik C, Werler MM. Use of selective serotonin-reuptake inhibitors during pregnancy and the risk of clubfoot. *Epidemiology*. 2014;25(6):859-865.

85. LeFevre ML; US Preventive Services Task Force. Screening for suicide risk in adolescents, adults, and older adults in primary care: US Preventive Services Task Force recommendation statement. *Ann Intern Med*. 2014;160(10):719-726.

86. Sockol LE, Epperson CN, Barber JP. A meta-analysis of treatments for perinatal depression. *Clin Psychol Rev*. 2011;31(5):839-849.

87. Howard LM, Molyneaux E, Dennis CL, Rochat T, Stein A, Milgrom J. Non-psychotic mental disorders in the perinatal period. *Lancet*. 2014;384(9956): 1775-1788.

88. Blanco C, Olfson M, Goodwin RD, et al. Generalizability of clinical trial results for major depression to community samples: results from the National Epidemiologic Survey on Alcohol and Related Conditions. *J Clin Psychiatry*. 2008;69(8): 1276-1280.

89. Wisniewski SR, Rush AJ, Nierenberg AA, et al. Can phase III trial results of antidepressant medications be generalized to clinical practice? a STAR*D report. *Am J Psychiatry*. 2009;166(5): 599-607.

90. Cuijpers P, Smit F, Bohlmeijer E, Hollon SD, Andersson G. Efficacy of cognitive-behavioural therapy and other psychological treatments for adult depression: meta-analytic study of publication bias. *Br J Psychiatry*. 2010;196(3):173-178.

91. Niemeyer H, Musch J, Pietrowsky R. Publication bias in meta-analyses of the efficacy of psychotherapeutic interventions for depression. *J Consult Clin Psychol*. 2013;81(1):58-74.

92. Cuijpers P, Van Straten A, Warmerdam L, Smits N. Characteristics of effective psychological treatments of depression: a metaregression analysis. *Psychother Res*. 2008;18(2):225-236.