Comparative Effectiveness Research Review Disposition of Comments Report

Research Review Title: *Efficacy and Safety of Screening for Postpartum Depression*

Draft review available for public comment from July 31 to August 28, 2012.

Suggested citation: Myers ER, Aubuchon-Endsley N, Bastian LA, Gierisch JM, Kemper AR, Swamy GK, Wald MF, McBroom AJ, Lallinger KR, Gray RN, Green C, Sanders GD. Efficacy and Safety of Screening for Postpartum Depression. Comparative Effectiveness Review 106. (Prepared by the Duke Evidence-based Practice Center under Contract No. 290-2007-10066-I.) AHRQ Publication No. 13-EHC064-EF. Rockville, MD: Agency for Healthcare Research and Quality; April 2013. www.effectivehealthcare.ahrq.gov/reports/final.cfm.

Comments to Research Review

The Effective Health Care (EHC) Program encourages the public to participate in the development of its research projects. Each comparative effectiveness research review is posted to the EHC Program Web site in draft form for public comment for a 4-week period. Comments can be submitted via the EHC Program Web site, mail or email. At the conclusion of the public comment period, authors use the commentators' submissions and comments to revise the draft comparative effectiveness research review.

Comments on draft reviews and the authors' responses to the comments are posted for public viewing on the EHC Program Web site approximately 3 months after the final research review is published. Comments are not edited for spelling, grammar, or other content errors. Each comment is listed with the name and affiliation of the commentator, if this information is provided. Commentators are not required to provide their names or affiliations in order to submit suggestions or comments.

The tables below include the responses by the authors of the review to each comment that was submitted for this draft review. The responses to comments in this disposition report are those of the authors, who are responsible for its contents, and do not necessarily represent the views of the Agency for Healthcare Research and Quality.

Source: http://effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=1437
Published Online: April 9, 2013

Commentator & Affiliation	Section	Comment	Response
Coyne, James	Abstract	The document that follows the structured abstract does not deliver on the promise of the objectives and the bulk of the document mainly focuses on the performance characteristics of screening instruments, without acknowledging or making use of what is known from the evaluation of depression screening instruments in other general and specialty medical settings. Overall, this document adds little new to the literature concerning depression screening, could serve to perpetuate problems in the existing literature, and fails to establish the discontinuity between screening for postpartum depression and screening for depression and other medical contexts. If a greater continuity had been recognized, some flaws in the approach taken by these investigators could have been avoided.	We have added additional discussion of the findings of the USPSTF review of screening for depression in adults, which overall are quite similar to those of our review in terms of performance characteristics of screening tests and the importance of appropriate management of positive screening tests.
Jesse, D. Elizabeth	Abstract	Was your summary of BDI or BDI-II? Your conclusions are based on only 15 studies. Since many nurses have published about PPD, could you expand your search to Medline and Cinahl? Clarify that EPDS, PDSS and BDI-II are screening instruments that would detect PP depressive symptoms, but not PPD. Please clarify your terms. Your conclusion is overreaching. Just as a screening tes tfor diabetes is sensitive, the person would still be ill without a plan to address the illness/symptoms. Screening for symptoms for RISK of PPD is a great first step but there needs to be further evaluation for PPD and a plan, ie therapy, support network, follow up for it to be of value.	We have clarified use of BDI and BDI-II throughout the report. The draft report conclusions were based on results from a total of 36 unique studies represented by 40 publications. The conclusions in the final report are based on results from 41 unique studies (46 publications). As described in the Methods section, the MEDLINE database (PubMed) was searched to identify Peer Reviewer-reviewed published literature, as were Embase, PsycINFO, and the Cochrane Database of Systematic Reviews. Additional published evidence was identified through manual review of the reference lists of relevant key articles/reviews. The search strategies (including specific databases to be searched) were discussed and confirmed with the TEP and AHRQ during development of the review protocol. Possible use of CINAHL was discussed at that time; the TEP agreed that the included databases were appropriate and sufficient to identify the relevant literature.

Commentator & Affiliation	Section	Comment	Response
Jesse, D. Elizabeth	Abstract	(Continued)	(Continued from previous cell) We have revised the abstract to reflect revisions throughout the report. To the reviewer's point, the screening instruments measure symptoms, but we very explicitly used a diagnosis of PPD based on DSM-IV criteria as the reference standard for sensitivity and specificity, so we believe our initial statement is correct. We believe our conclusion that there is insufficient evidence to draw any conclusions about the net balance of benefits and harms of screening for postpartum depression, or about whether specific tools or strategies would result in a more favorable balance, is appropriately justified based on the limitations of the identified evidence. We agree that questions regarding the effectiveness and harms of downstream therapy for PPD following screening are of significant interest (and importance). Such questions are currently being pursued in a second AHRQ comparative effectiveness review. We have added clarification in the report that a second review will address questions regarding treatment.
Peer Reviewer 2	Executive Summary	Does not include references for the included studies which we have recently been asked to have in ES. Fine if not required.	The number of citations in the Executive Summary is limited per the AHRQ format to make the ES document more concise. Complete lists of references for both included and excluded articles are provided in the full report.
Peer Reviewer 2	Executive Summary	[Pg. ES-2, l. 48] Suggest delete "consistent" or use persistent, long-standing	We have replaced the term "consistent" with "persistent" and restructured the sentence for clarity.
Peer Reviewer 2	Executive Summary	[Pg. ES-2, l. 54-55] Clarify last sentence: "…, or between [among] strategies."	We have revised this sentence for clarity.
Peer Reviewer 2	Executive Summary	[Pg. ES-3, l. 10-15] Suggest reword/divide sentence	We have restructured this statement for clarity in both the ES and main report.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 2	Executive Summary	[Pg. ES-3, l. 16-18] Change screening "of" to screening "for". Semantically "By summarizing data that support improved screening" implies that you left out non-supportive data. Know this is not what you intended. Also suggest rewording to "we hope to improve outcomes for women, their partners, and…" (Just my preference - big fan of standard English with minimal jargon.)	We have corrected this statement. Thank you—we agree this statement was unclear, and we have revised it to the following: "By summarizing the available evidence on the accuracy and effectiveness of screening for postpartum depression, we hope to provide a resource to organizations developing recommendations to enhance patient-centered outcomes for women, their partners, and children, ideally with efficient use of clinical resources."
Peer Reviewer 2	Executive Summary	[ES- Table A p. ES-11] Number of subjects info missing in rows 21 and 23	We have added the subject numbers for the remaining rows.
Peer Reviewer 2	Executive Summary	[Pg. ES-11] Two question screen questionnaire lists one study but number of subjects is missing.	We have added the subject numbers for the remaining rows.
Peer Reviewer 2	Executive Summary	[Pg. ES-11, l. 31-34 and Results p. 32 l. 51-53] Study numbers by country US=3, Europe=7, Asia=2, 1 each from UK, Australia, and Israel totals 15 but text says only 14 unique studies. Was the UK study also included in the "Europe" count? Any reason UK is considered separate from Europe?	Thank you for noticing this error; the numbers have been corrected to reflect the totals per the final set of included studies. The UK was considered separately from Europe primarily because screening instruments were administered in English, enhancing comparability to a US non-immigrant setting.
Peer Reviewer 2	Executive Summary	[ES- Table B] Number of subjects missing for certain studies.	We have added the subject numbers for the remaining rows.
Peer Reviewer 2	Executive Summary	[Pg. ES-13, l. 54 and Results p. 40, l. 41-42] Is funding relevant information here? Suggest delete sentence "Of the two KQ3a…" and change to "A prospective investigation of maternal mental illness conducted at a single U.S academic center enrolled women prior to 28 weeks…"	We have restructured this sentence in the Results and the ES to remove discussion of funding sources.
Peer Reviewer 2	Executive Summary	[Pg. ES-14, l.7-9 and Results p. 40, l. 49-50] Is funding information relevant? (Agree there are cases in which it is.) If you wish to highlight might do consistently throughout.	We have restructured this sentence in the Results and the ES to remove discussion of funding sources.
Peer Reviewer 2	Executive Summary	[Pg. ES-14, l. 41-42] Suggest reword to "The single U.S. study included only women receiving…"	We have reworded this sentence for clarity.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 2	Executive Summary	[Pg. ES-14, l. 43-44] Suggest reword “The other studies were conducted in… “; please specify which study provided harms data.	We have reworded this sentence for clarity and specified which study provided evidence regarding harms.
Peer Reviewer 2	Executive Summary	[Pg. ES-14, l. 47-55] Is there a consistent order throughout in which studies are presented? Not completely clear, might state. Often discuss good quality studies first, followed by fair, then poor.	There is not a specified order for discussion of individual studies; in some cases, the authors of the individual chapters used chronological order, while in other cases they are listed by study quality, or by subtopic (for example, risk factors).
Peer Reviewer 2	Executive Summary	[Pg. ES-15, l. 35-36] For the “the cross-sectional study was rated as good quality” what quality scoring method was used? Apologies if I missed this or misunderstood methods.	The methodological quality (or risk of bias) of individual studies was assessed according to processes detailed in the AHRQ Methods Guide and Medical Test Guide. Citations for these sources, details on the specific elements considered, and definitions of overall quality ratings are provided in the Methods section of the full report.
Peer Reviewer 2	Executive Summary	[Pg. ES-16, l. 44-45 and Results p. 22 l. 55-57 and p. 57 l] 13 States 48% studies conducted in UK or Europe. This does not agree with % presented on pg. 22 (53%)	Thank you for noticing this error; the numbers have been corrected to reflect the totals per the final set of included studies.
Peer Reviewer 2	Executive Summary	[Pg. ES-16, l. 55-56] Delete duplicate word “that”.	This has been corrected.
Peer Reviewer 2	Executive Summary	[Pg. ES-19, l. 43] Insert “of” after “low strength”.	This has been corrected.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 2	Executive Summary	Tables in Executive Summary and Report text: "Magnitude of effect" as a column header in SOE tables could use some definition. I would advocate for a modification of the standard SOE table in order to have the table say exactly what you are reporting - for example in Table A "Diagnostic Test Performance" or "Diagnostic Test Characterisitic". I tend to think of magnitude of effect as an outcome of intervention or change measure - doesn't seem to fit properties as well. Some of the tables like SOE Table C/E, include headers that have elements like "and provider" that are not part of the SOE, indeed in some cases where there was no evidence. Could the tables be more clear if the table either showed the element "provider type" and no literature? Or what we more often do, could you drop the item not found in the literature from the table title and state as you do in these instances in text that there was no literature identified?	We have made the suggested change to the SOE table for KQ 1 in the ES/main text. We have removed elements that are not included from the headers.
TEP Member 1	Executive Summary	One page 9 [top of page=ES-2 bottom] line 28 it could be more clearly delineated how screening can improve outcomes through the treatment of depression.	We have added a sentence about how screening can potentially improve outcomes by identifying undiagnosed depression which would otherwise go untreated, or get treated at a more severe stage.
TEP Member 1	Executive Summary	Methods section is very well done. On page 12 [top of page=ES-5 bottom] line 21 should include rational for limiting to economically developed countries.	We have added the rationale for focusing on economically developed countries to the Methods section of the ES.
TEP Member 3	Executive Summary	[Results section]: References for the cited studies were not included in the executive summary and this would be helpful	The number of citations in the Executive Summary is limited per the AHRQ format to make the ES document more concise. Complete lists of references for both included and excluded articles are provided in the full report.

Commentator & Affiliation	Section	Comment	Response
TEP Member 8	Executive Summary	In the methods section of the executive summary, it should be explicitly stated that only studies with diagnostic standard follow-up of a positive screen were included.	We have added a statement to the Methods section of the ES to clarify this point, as follows: “Studies reporting depression outcomes were required to include confirmation of depression with a reference standard.”
TEP Member 10	Executive Summary	In the Executive Summary, the Two-Question Screen is mentioned in Table A. However, the Two-Question screen is not mentioned/described immediately before or after Table A. Provide a brief description of the Two-Question screen in the Exec Summary; perhaps the explanation should go somewhere in lines 51-56 on page ES-8.	We have expanded the description of the two-question screen.
TEP Member 10	Executive Summary	[ES-2]Pg 9 of 189, line 38 Include "culturally-appropriate screening tools.	We have made this revision.
TEP Member 10	Executive Summary	[ES-3]Pg 10 of 189, line 12: Include "ethnicity" after "race".	We have added ethnicity.
TEP Member 10	Executive Summary	[ES-3]General comment: Pg 10 of 189, lines 16-26 describe a laudable scope, however, the results and conclusion indicate that none of this was achieved. Perhaps instead of stating "will benefit" throughout the paragraph, it should state "potentially will benefit" or "will increase understanding of the benefit".	We have revised this section to read: “By summarizing the available evidence on the accuracy and effectiveness of screening for postpartum depression, we hope to provide a resource to organizations developing recommendations to enhance patient-centered outcomes for women, their partners, and children, ideally with efficient use of clinical resources.”
TEP Member 10	Executive Summary	[ES-7]Pg 14 of 189, line 19: Briefly define 'sufficient studies' in parentheses.	We have added a definition of sufficient studies (three or more) to this sentence.
TEP Member 10	Executive Summary	[Pg ES-18, line 6] take the 's' off of "outcomes."	We have made this revision.
TEP Member 10	Executive Summary	[ES19, line 37 to 38] I would recommend adding a focus on parental or paternal health outcomes.	We have added a sentence to this effect.
TEP Member 10	Executive Summary	[Pg ES-16, line 42-43] The sentence does not make sense. What are community patients? Do you mean perinatal women seen at perinatal public health clinics?	We have clarified this sentence as “Many included studies recruited populations whose demographics differed considerably from patients in the broader community.”
TEP Member 10	Executive Summary	[Pg ES-19, line 25] replace "unwanted" with "adverse."	We have made this revision.

Commentator & Affiliation	Section	Comment	Response
TEP Member 10	Executive Summary	[ES19, line 44 -48] I do not entirely agree with the statement that the outcomes relevant to timing, setting, or provider are more related to aspects of the screening/referral/diagnosis process other than the test characteristics of the specific screening instrument used in the study. If this were true, then why was the Two-Question screener used in one study? I do feel that the screening process has a substantial influence on outcomes based on the Federal Healthy Start program experiences.	Based on additional evidence identified in the updated search, we have emphasized that the evidence suggests that setting is very important for screening effectiveness.
TEP Member 10	Executive Summary	[ES19, line 53 – 56] Revise the sentence to read [see revisions in all caps] "...postpartum depression, CONSISTENT WITH THE 2005 REVIEW, there CONTINUES TO BE VERY LIMITED AND insufficient EMPIRICAL evidence to draw any conclusions about the net balance of benefits and harms of screening for postpartum depression, or about whether specific tools or strategies would result in a more favorable balance" I feel that it is very important to send the clear message that the science is lacking, not that screening/referral has not been proven to be beneficial. When the USPSTF states the latter, screening and referral practices and reimbursement opportunities decrease; this severely disadvantages the infant, mother, and family.	We have made the suggested change, and as other reviewers have suggested, have emphasized that the lack of direct evidence of benefit is not synonymous with evidence of a lack of benefit.

Commentator & Affiliation	Section	Comment	Response
Jesse, D. Elizabeth	Executive Summary	For the most part, this is a well-done report. Here are some issues to be addressed, some major, some minor. p. ES-2 -- what does the term "consistent uncertainty" mean?	We have replaced the term “consistent” with “persistent” and restructured this sentence for clarity.
		A limitation of the CER should be mentioned. If so, would mention that it is difficult to find direct studies on KQ4. That is why the USPSTF uses a logic model that permits the use of separate studies on the effectiveness of treatment. I have mentioned this several times during the process, to no avail (except that there will be a separate SER on PPD treatment) but would like it mentioned now. Thus, it may not be correct to say (p. ES-17), the evidence is insufficient.	We have revised this section considering evidence published subsequent to the initial draft (the Yawn 2012 paper).
		p. ES-14: heading says KQ4 AND KQ5 but the paragraph doesn't address comparative harms of screening. There is some language previously in the ES.	Under the ES heading that refers to KQ 4 and KQ 5, we discuss the evidence provided by the one study that reported data for a potential harm of increased unscheduled doctor visits for infants of screened women.
		p. ES-18 and elsewhere in the document: cost is mentioned. 2 issues: it is unclear whether it's the cost of the research or the cost of the screening and associated care that is the subject. Are AHRQ CERs allowed to mention cost?	Costs were not formally considered. The revised report places a stronger emphasis (partly in response to additional evidence published subsequent to the first draft) that screening effectiveness appears to be highly dependent on the availability of systematic resources for ensuring appropriate diagnosis and treatment, and that choices about screening strategies (including choice of test) have significant impact on these resources.
		p. ES_19 re risk factors: are the risk factors being mentioned valid? these are all correlational, so rather than put them in a model to stratigy care, would suggest that they be validated.	We have clarified that our purpose in reviewing the risk factor literature was primarily in the context of improving test performance, rather than as potential targets for interventions themselves. An association does not have to be causal in order to be useful in improving posttest probabilities.

Commentator & Affiliation	Section	Comment	Response
Yogman, Michael	Executive Summary	While this review is comprehensive in scientific rigor based on the scarce high quality published data, the conclusions based on this inadequate data are not supported by this review, which at best is a major critique of the scarcity of data on an important public health issue. This review neglected to cite the recent IOM report on this subject which is a serious ommission. First, the evidence review provides stronf support for the sensitivity and specificity of the EPDS, probably better than most widely used screening instruments. Second, the absence of strong support for positive screens leading to improved maternal outcomes is more an indictment of inaccessible mental health services rather than a critique of screening. Postpartum women may not want/or be able to leave their infant to visit a mental health provider .Very few mental health providers see mothers and infants together even though limited studies suggest that the dyad is the more appropriate target for intervention. Most mental health providers are better reimbursed by insurers for treating depression with medications and postpartum nursing mothers may not be candidates for this treatment. The most important target of effective screening is an improvement in parent infant interaction but this review does not highlight the need for well validated measures of this outcome and research to assess it. Third, the attribution that an increased number of unscheduled infant clinic visits in positively screened mothers is a harm is likely to be wrong;in fact it is likely to be a benefit which could be documented with larger sample sizes and longer follow up to result in fewer inflicted injuries and accidents if mothers are getting extra support from their pediatricians.	We have included discussion of the IOM report findings and recommendations, which are not inconsistent with ours, in both the introduction and discussion. In regards to specific points: 1) As noted in responses to other reviewers, we have not identified anywhere in the literature an explicit discussion of the tradeoffs in false-negative and false-positive probabilities for both individual patients and the health system in general that would allow us, or any group, to identify an optimal screening test and threshold. We have reemphasized this in the revised report—clinical considerations should drive the choice of test and threshold, rather than an arbitrary cutoff based on optimizing the area under the ROC curve. 2) We have reemphasized the point about appropriate followup, with additional evidence published subsequent to the draft review. We have also emphasized the lack of data on maternal–infant dyad outcomes, and the need for additional data. 3) We have included further discussion of the point about whether increased utilization of health services among infants is a harm rather than a benefit. The IOM report explicitly refers to “maladaptive” utilization, including both underutilization and overutilization. Although we agree that the evidence of an association between maternal depression and infant (and longer term outcomes) is strong, as noted by the reviewer, there is a lack of evidence that screening, even with effective treatment, results in improvement in these outcomes.

Commentator & Affiliation	Section	Comment	Response
Yogman, Michael	Executive Summary	(Continued from previous cell) Untreated parental depression influences a host of regulatory problems in the parent-infant dyad influencing sleep, feeding ,crying, which may in turn lead to a lifelong cycle of toxic stress injury to the brain,later developmental problems and long lasting health problems as adults, none of which were even acknowledged in the discussion of the reports findings.Interestingly, the report failed to emphasize the strong finding of environmental risk factors (KQ3) strongly predicitive of postpartum depression--single parents , low income , premorbid depression-a very important public health finding to target resources. The report continually raises the specter of hypothetical harm from stigmatization but ignores the real harm we see everyday of parents failing to acknowledge their need for and seek help for fear of being stigmatized because this problem is not widely acknowledged and screened for . The normative aspects of postpartum adjustment and need for universal post partum support need more emphasis. This report would do better to acknowledge the seriousness of untreated postpartum depression, encourage open discussion of the issue , referral and availabilty of appropriate services rather than worrying about the hypothetical harm of stigma associated with screening. Finally I give the report credit for mentioning the lack of data on paternal depression for this is a common underdiagnosed disorder, presents differently than in women and needs better screening tools and treatment modalities.	(Continued from previous cell) 4) We have expanded our discussion of risk factors; we also note, as have others, the methodological difficulties in disentangling the potential effects of risk factors for depression, many of which are also associated with adverse developmental outcomes, from depression itself on the risk of longer term development and health problems. 5) Although we are not sure we would characterize 5 references to stigmatization in an over 6000 word report as “continual,” we have revised our discussion to reflect the reviewer’s excellent point about potential barriers to receiving appropriate services created by it. We note that this issue was raised specifically by the TEP and has been raised in other reviews of screening for depression, including the USPSTF review of screening in adults. 6) Thank you.
Peer Reviewer 1	Introduction	The Introduction provides the background needed to explain the scope of the report. It is clear and concise. My stakeholder group is consumers. They are confused and frustrated by lack of consistent recommendations regarding postpartum depression screening. This report speaks to this issue.	Thank you.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 2	Introduction	Appreciated note about DSM-V. "There is high-quality evidence for effective treatment of patients who meet criteria for major depression in other settings; evidence is inconsistent for postpartum depression." I'm being dense - what is the evidence for screening and treatment minor depression? Is postpartum generally only major or do some of the other diagnostic standards also include minor depression? Like the scope of intro, clear but not long winded.	We have extensively revised this discussion to clarify the state of the evidence.
Peer Reviewer 3	Introduction	The Background section is concise and appropriately describes the evolving definition of pp depression.	Thank you.
Peer Reviewer 3	Introduction	[Table 2, p 3] is very helpful in establishing context by illustrating the varying guidelines	Thank you.
Peer Reviewer 3	Introduction	[p. 4] The descriptions of sensitivity, specificity, PPV and NPV are basic, but clear and likely helpful for some readers.	Thank you.
Peer Reviewer 3	Introduction	The key questions, analytic framework, and search strategy are well formulated and important conceptually.	Thank you.
Peer Reviewer 4	Introduction	The introduction is fair, but you can tell from the beginning that the authors are not clinically familiar with this topic area.	One of the members of the review team is a psychiatrist whose practice largely consists of perinatal and postpartum patients; we also received significant input from a psychiatric social worker whose practice is limited to pregnant and postpartum patients. As noted in both the report and other responses, the inclusion/exclusion criteria were developed with significant input from the TEP, which included multiple members with experience in perinatal psychiatry. We would also argue that the basic considerations involved in evaluating screening are independent of the condition for which screening is proposed, and that many of the clinical issues involved with evaluating screening strategies, which is the focus of this report, are more relevant to the clinicians who will be performing the screening.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 5	Introduction	More discussion of how postpartum depression differs clinically from depression at other life stages. Are the same treatments known to be effective in general population similarly effective among women with postpartum depression?	We have included a discussion of the comparative epidemiology and clinical outcomes between postpartum and other women of childbearing age.
Peer Reviewer 6	Introduction	see above . The scope of the report is very well framed in the Introduction	Thank you.
Peer Reviewer 7	Introduction	[Page 2 Line 27]. References for effect of maternal depression on infant and child development seem off (10-13). Should reference IOM report, Depression in Parents, Parenting, and Children as well as more relevant studies.	We have revised these references and included a discussion of the IOM report.
Peer Reviewer 7	Introduction	[Page 4 line 6] reference to efforts at state level to require offering screening for PPD is from NJ. I would recommend citing efforts in Illinois that provide compensation and systems-levels supports for PPD screening. Would be good to check if there are any published articles about Illinois experience.	We have included a reference to a citation regarding the Illinois experience.
TEP Member 2	Introduction	Outlines the issues clearly. Teh prevalence rates are low compared to many published rates.	We have clarified that these are point prevalences, which is the measure most appropriate for estimating positive and negative predictive value of screening, and included a reference to the estimated period prevalence, which is the more commonly reported measure.
TEP Member 3	Introduction	The following statement is not accurate [p. 1; also ES-2]: A new set of diagnostic criteria for psychiatric illness, the Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (DSM-V), is currently scheduled for release in May 2013; preliminary discussions suggest that the overall diagnostic framework for postpartum depression (i.e., major depression with a specification of postpartum onset) will remain unchanged, although the window for diagnosis may be extended to 6 months after delivery. ---DSM V is now called “DSM 5” ---it is not likely that the risk period will be extended to 6 months ---it is likely that the risk period will include some time period of pregnancy	We have corrected the name and eliminated any discussion of potential revisions.

Commentator & Affiliation	Section	Comment	Response
TEP Member 4	Introduction	Clear, concise	Thank you.
TEP Member 5	Introduction	To start with a small point, it is DSM-5 not DSM-V. The references (3-5) at the end of paragraph were inadequate, especially as they addressed psychotherapy. In fact, the authors did a relatively poor job in discussing treatment. Again this issue comes up in Potential Harms of Screening in the discussion of treatment for minor depression. The issues that seem to plague pharmacotherapy are somewhat different than those that plague psychotherapy. Again, the authors need to be much more careful in their language. As an example, they cite the Gaynes paper (3) as evidence that treatment does not improve outcomes for minor depression. This report barely addresses the issue and should not be viewed as any kind of a primary source. One of my major concerns about this review is the fact that it did not include studies that were included in the prior review. This is a significant weakness and it is manifest throughout the entire report. The authors undertook a quantitative review yet ignored a good bit of the relevant literature. This is inexplicable and it unnecessarily limited the power of their analyses.	We have corrected the reference to DSM-5. During the initial scoping of the report with AHRQ, the key informants, and the TEP, it was clear that resources to include an evaluation of therapy were not available. Since that time, a separate followup review of treatment effectiveness has been funded and is currently ongoing. During these discussions, there was also agreement to focus on the literature published subsequent to the earlier AHRQ report. It is certainly possible that more expansive inclusion criteria might have led to different conclusions; however, given that almost all of the literature limitations that precluded extensive quantitative synthesis in the 2005 report were also present in the subsequently published literature, we believe the likelihood that inclusion of previously published studies would have allowed meaningful quantitative synthesis is relatively low.
TEP Member 6	Introduction	Introduction is clear with a well defined population and goal.	Thank you.
TEP Member 7	Introduction	The introduction provides the framework for the this review, and why it is am important topic.	Thank you.
TEP Member 8	Introduction	Clear overview and background are presented.	Thank you.

Commentator & Affiliation	Section	Comment	Response
TEP Member 8	Introduction	On p. 7 of the Intro, paragraph in lines 39-50 has an unobjective tone, as if the authors have already concluded that screening for PPD is a useful health quality measure. For example, the first sentence "By summarizing data that support improved screening of postpartum depression, we hope to enhance patient-centered outcomes for women, their partners, and children..." What is "improved screening?"	Thank you—we agree this statement was unclear, and we have revised it to read: "By summarizing the available evidence on the accuracy and effectiveness of screening for postpartum depression, we hope to provide a resource to organizations developing recommendations to enhance patient-centered outcomes for women, their partners, and children, ideally with efficient use of clinical resources."
TEP Member 10	Introduction	General comment: It would be useful to include in-text citations in the Background section.	We thank the reviewer for this suggestion. For consistency across reports, referencing in the Executive Summary and main report follow the standard citation format specified by AHRQ.
Coyne, James	Introduction	[ES-1 and pg. 1] Background The statement "The impact of depression and postpartum women is at least as great as that for depression in other populations" is noncontroversial, but seems to avoid dealing with the common myth that postpartum depression is more common than depression among women of childbearing age who have not just had a baby. Similarly, there is a bit of evasion in subsequently stating the prevalence of major depression for postpartum without making any comparisons to other women of childbearing age.	These are excellent points, and we have revised this section to clearly state that the prevalence of major depression is similar in postpartum and nonpregnant/nonpostpartum women of the same age.
Coyne, James	Introduction	[ES-3 and pg. 6]Scope and Key Questions "By summarizing data that support improved screening of postpartum depression" assumes that the benefits of screening have been established, when decidedly they have not been.	We agree that that this wording inadvertently gives the impression that the benefits have been established. We have revised it to read, "By summarizing the available evidence on the accuracy and effectiveness of screening for postpartum depression, we hope to provide a resource to organizations developing recommendations to enhance patient-centered outcomes for women, their partners, and children, ideally with efficient use of clinical resources."
Coyne, James	Introduction	[pg. 4] Potential Harms of Screening The focuses too much on harm to the individual patient, rather than to systems of care that are already inadequate and underresourced.	This is an excellent point, which we have added to the discussion.

Commentator & Affiliation	Section	Comment	Response
Coyne, James	Introduction	[pg. 4Accuracy of Screening Instruments This section fails to state the screening instruments are to be evaluated in terms of their ability to detect otherwise on detected and presumably on treated depression.	We have clarified this section; in addition to their ability to detect otherwise undetected and untreated depression, screening instruments also need to be evaluated based on their "ability" to falsely diagnose a condition, and, unfortunately, these two abilities are inextricably linked.
Jesse, D. Elizabeth	Introduction	I suggest that you add physician/provider factors for factors limiting screening, such as inexperience, lack of time or desire to know. In clinical setting, a dominant philosophy has been, "Don't ask, don't tell." Yes, it is good your report discussed adequate systems. It can be considered unethical to screen women without a safety net and protocol in place. Since EPDS is specific for pregnancy, a good clinician can know if the S&S are different than postpartum changes. The MINI could be used with EPDs for clinical dx. Do you mean to imply that women should not be screened until there are more RCTs? Dr. Caron Zlotnick has completed several RCTs with women who suffer from PPDS. I was surprised by your statement that there is paucity of evidence for balance of screening or not. There is such a problem of up to 50% of women in pregnancy with PPDS yet there are rarely found. Now your report is suggesting that we not find them postpartum? Cheryl Beck's research shows that half of women with antepartum depressive symptoms go on to have PPD. There is such a need to find these women so treatment can be initiated. Question two needs to include factor: current antepartum depression (APD) or antepartum depressive symptoms (APDS). Question 4 "no screening" does not seem ethical to me Question 6: include affected by physician or other health professionals level of comfort, ability, and knowledge Stake holders: Did you reach out to nurses and nurse-midwives?	The Key Questions were developed in consultation with AHRQ, the topic nominator, and a wide range of Key Informants and Technical Expert Panel members who are acknowledged in the report. Once finalized at this stage of the review, these questions cannot be changed, although the reviewer makes excellent suggestions. We do not suggest that women should not be screened until more RCTs are conducted. We have attempted to clarify to the extent possible that the state of the evidence is insufficient to draw a conclusion about overall net benefits, rather than that the evidence points to no benefit, and to emphasize the need for additional research to clarify these points. To the extent that policy and practice are (or should be) evidence-based, our conclusion that more specific evidence is needed to inform that policy and practice is quite different from a conclusion that identifying and providing care for perinatally depressed women is not worthwhile (in other words, "absence of evidence is not evidence of absence"). We have emphasized this in the revised discussion section and executive summary. We explicitly included midwives and nurse-midwives in our list of stakeholders.
Peer Reviewer 1	Methods	The methods are thorough, explained clearly and logically.	Thank you.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 2	Methods	Yes. Personally would have liked to see likelihood ratios but that likely falls in the zone of too much detail. Since they are readily calculated by those with interest it is not a problem but might have made for interesting figure graphing pretest probability by post-test probability and showing what range of priors might actually be informed by findings of screening. Imagine when our clinical priors are already elevated that screening adds little to need to evaluation more and that a positive screening with a very low prior might also not activate care teams. Maybe for a later paper?	We agree that the issue of how priors affect posttest probability is an important one. We had originally planned on more extensive modeling of the effect of varying priors based on timing, predictive algorithms, or other factors, but, given the level of evidence, the degree of uncertainty would be too high. Such an exercise would be quite valuable for value-of-information analysis or other quantitative methods for research prioritization, and we are planning subsequent development of the model.
Peer Reviewer 2	Methods	Methods/Results: This section (pg. ES-7 and pgs. 17-18) includes a detailed description of meta-analysis and other quantitative models but it is not clear if the methods described were, in fact, used for this report. In at least two places the team states that meta-analysis was not possible. (pg 24, l. 33 and pg. 33 l. 11). Might be good to clarify if the analytic methods were used, and if yes, specify for which KQ. If they were not used these sections can be abbreviated. Not clear if the diagnostic test properties which were used for the simulation were meta-estimates since text states: "The values for sensitivity and specificity (along with CIs) were derived from the literature review." Like the inclusion of simulation and the subsequent related text.	Thank you; we have revised the Methods section in the main report and ES to distinguish between the methods that were proposed (and would have been used if the evidence had allowed) and those that we were able to use. Thank you.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 2	Methods	Were other quality scoring instruments used in addition to QUADAS-2	Yes. To assess quality for studies presenting information on patient-centered intermediate, final, and adverse effect outcomes, we followed the process outlined in the AHRQ Methods Guide to (1) classify the study design, (2) apply predefined criteria for quality and critical appraisal, and (3) arrive at a summary judgment of the study's quality. The citation for the Methods Guide, details on the specific elements considered in rating quality, and definitions of overall quality ratings are provided in the Methods section of the full report. An abbreviated description of the quality rating process is provided in the ES.
Peer Reviewer 3	Methods	[P 12], Second bullet point under Exclusion criteria for Populations: Please clarify why you excluded studies that attempted to detect (ie, screen for) depression during pregnancy. On page 14, under Timing, it states that prenatal screening was included.	We have revised to clarify that we excluded studies where the outcome was depression during pregnancy, but included studies that assessed women prenatally for risk of postpartum depression.
Peer Reviewer 3	Methods	[P 22] articles excluded: I note that 445 articles were excluded because they had "No outcomes of interest". Were the exclusion criteria for outcomes (p 13) too strict? One could question the decision to exclude studies because the outcome was measured only with the screening instrument and not confirmed with a reference standard. Especially when tests such as the EPDS have a reasonably high specificity.	The outcomes were initially suggested by AHRQ in the request for the review, and refined through discussion s with key informants and the TEP. As discussed in the report, the definition of "reasonably high specificity" is not clear. In the draft report, and in the revisions, we have provided some numerical examples of the number of women with false positive results with tests of specificity of 80-90%. In the range of prevalences reported in the literature, at least half of all positive results would be false positives. Drawing inferences about either association or effectiveness when at least 50% of those with the "outcome" truly do not have it seems problematic.
Peer Reviewer 3	Methods	[P 17 Lines 50-58] Redundant with page 4. You could edit out everything after the first sentence in line 50.	We have eliminated the text in this section discussing likelihood ratios.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 4	Methods	Too restrictive--- to the point that very few articles are included and the authors are therefore unable to address the identified key questions.	As noted in a prior response to a similar comment, we would certainly be willing to discuss any limitations resulting from these criteria, but, without a specific recommendation from the reviewer about alternative inclusion/exclusion criteria or what clinically meaningful aspects of perinatal depression were overlooked on the basis of the inclusion/exclusion criteria, we cannot address the reviewer's general concerns.
Peer Reviewer 5	Methods	Generally the methods are appropriate and clearly explained. How many studies were excluded becuase there was information about postpartum depression and other outcomes without linking to screening; is there strong evidence for effective treatments for postpartum depression (regardless of whether cases are identified through screening)?	Thank you. 44 articles were excluded for providing information about postpartum depression and other outcomes without linking to screening. These excluded articles are counted within the literature flow diagram as part of the "No outcomes of interest" group. While treatments for postpartum depression are beyond the scope of this review, we agree that questions regarding the effectiveness and harms of downstream therapy/interventions are of significant interest. Such questions are currently being pursued in a second AHRQ comparative effectiveness review on the effectiveness and safety of treatment options.
Peer Reviewer 6	Methods	Rigor of the Methods is both explained and clearly noted in the numerous tables	Thank you.
Peer Reviewer 7	Methods	Inclusion and exclusion criteria were justifiable; search strategies were clear; statistical methods appropriate.	Thank you.
TEP Member 1	Methods	Methods section is very well done. On page 12 line 21 should include rational for limiting to economically developed countries.	Thank you; we have added the rationale for focusing on economically developed countries to the Methods section of the ES as suggested.
TEP Member 2	Methods	1. Inclusion adn exlcusion criteria---clearly stated and appropriate.	Thank you.
TEP Member 2	Methods	2. Search strategies---clearly stated and appear to be adequately broad adn logical. The day may soon come that we need to add Google.	Thank you.

Commentator & Affiliation	Section	Comment	Response
TEP Member 2	Methods	3. The quriement of SCID or SCAN for the diagnosis of PPD for use in the sensitivity and specificity studies is appropriate. I am very glad that was not required in the intervention RCTs since none of the studies requiring that in the US have sufficient numbers of women who completed that referral.	Acknowledged.
TEP Member 2	Methods	4. The statiscs appeared appropriate. The modeling section might benefit from the information from the Yawn 2012 paper which did use two step screening with senstive EPDS and follow up with more specific PHQ-9.	Thank you; the Yawn 2012 article was identified during the search update and has been included in the final report. We have added a reference to the use of serial testing in the Yawn paper, but cannot make direct comparisons.
TEP Member 3	Methods	the authors need to describe why they started their review in 2004 (i.e. related to prior AHRQ report)	We have clarified our rationale in the Methods section: “Given the findings of the 2005 review regarding the level of evidence, we chose these dates after consultation with AHRQ, Key Informants, and the TEP in order to maximize efficiency. The primary impediment to formal data synthesis in the 2005 review was study heterogeneity. Therefore, it was unlikely that we would be able to combine literature identified in that report with newer data in any subsequent meta-analyses. This led us to conclude that qualitative comparison of our findings to those of prior reviews would be a more useful approach.
TEP Member 4	Methods	All done according to protocol, clear and appropriate. Some of the outcomes (eg breastfeeding) seemed a little far from the primary issue, but I can see why it was there.	Thank you.

Commentator & Affiliation	Section	Comment	Response
TEP Member 5	Methods	Overall, I was very impressed with the methodology of the study. Everything was described in a very clear manner and in a way that probably would allow another investigator to replicate the study findings. The major problem was in the decision to only review the most recent literature. Also, it was unclear whether the authors systematically surveyed major investigators for unpublished work. The report did make reference to the "grey" literature but was rather ambiguous as to what that included. The authors did a reasonably good job of defining terms, but I thought could have done a better job of defining "bias." In the section on Data Synthesis the authors discuss the value of using likelihood ratios but I could not find any further reference to them. Did the authors mean RR or OR? It is difficult to say. The authors should go through manuscript and make sure that they actually follow through on what they say they are doing.	Thank you. (1) We have clarified our rationale in the Methods section: "Given the findings of the 2005 review regarding the level of evidence, we chose these dates after consultation with AHRQ, Key Informants, and the TEP in order to maximize efficiency. The primary impediment to formal data synthesis in the 2005 review was study heterogeneity. Therefore, it was unlikely that we would be able to combine literature identified in that report with newer data in any subsequent meta-analyses. This led us to conclude that qualitative comparison of our findings to those of prior reviews would be a more useful approach. (2) Grey literature databases searched included ClinicalTrials.gov; the World Health Organization (WHO) International Clinical Trials Registry Platform (ICTRP) search portal; and the ProQuest COS Conference Papers Index. Publishers of proprietary depression screening tools were also contacted (directly by the AHRQ Scientific Resource Center) to provide unpublished data or other additional information to be considered. These sources are specified in the Methods section of the main report and in the Methods section of the ES. A full listing of the search dates and terms used in searching these sources is provided in Appendix A.

Commentator & Affiliation	Section	Comment	Response
TEP Member 5	Methods	(Continued from previous cell)	(Continued from previous cell) (3) Although we believe the definition of "bias" in the setting of discussion of research findings is generally understood, we have added the additional statement: In this context, "bias" refers to the degree to which a study's results are due to aspects of the study design (choice of population, allocation of treatment, uneven distribution of risk factors, etc.) rather than the specific factor (risk factor or exposure, screening test, treatment, etc.) of interest. (4) We have eliminated the text in the Methods section referring to likelihood ratios. (5) We agree the language in the Methods section may have been confusing. We have revised the text to clearly distinguish between the proposed methodology established prior to performing the review, and the statistical analyses we were able to perform in practice with the final set of available evidence.
TEP Member 6	Methods	Methods are well defined and appropriate. inclusion and exclusion criteria makes sense.	Thank you.
TEP Member 7	Methods	There was a great deal of variability in the 15 studies that were selected, so the authors were hampered to a degree.	Acknowledged; the available studies for KQ 1 (15 in the draft report and 18 in the final) do present considerable variability.
TEP Member 8	Methods	It struck me as unusual (and perhaps not appropriate) that the authors conducted ad hoc analyses/simulations--is it customary for EPC reports to generate and report "new" findings generated by the team, ie not appearing in the published literature?	The use of a simulation model was part of the protocol from the beginning of the project, and discussed with the TEP. Decision/simulation models are an alternative method for quantitatively synthesizing the literature, analogous to a meta-analysis—the results of both are "new" in the sense that they have not previously been published. It is not uncommon for EPC reports to include such models, and they can often provide additional insight that is not available with other methods.

Commentator & Affiliation	Section	Comment	Response
TEP Member 10	Methods	[pg. 12]Table 3 indicates that findings related to KQ 3 are viewed as interim outcomes measured by diagnosis of depression from DSM-IV-TR. I agree with this. What is confusing is that several intermediate outcomes for KQ 4 and 5 (e.g., receipt of appropriate diagnostic and treatment services) as well as the final outcome for KQ 4 (e.g., scores; service utilization) all relate to KQ 3 as well. Should KQ 3 be listed with the appropriate intermediate (KQ 4 & 5) and final outcomes (KQ 4) too?	KQ 3 focuses specifically on the outcomes of test performance (sensitivity, specificity, predictive values), so that intermediate outcomes such as receipt of services do not apply. There is certainly an indirect association between test performance and some of these outcomes (for example, a test with a low specificity might generate too many positive results and subsequently affect the proportion of women receiving followup services), but are not directly relevant to KQ 3.
TEP Member 10	Methods	[pg. 12] In Table 3, setting and timing are listed in the far left column. Shouldn't there be a note in the middle column that these variables relate to KQ 3?	Timing of screening and setting in which screening is conducted were indeed specific factors under consideration in KQ 3. In contrast, the Timing and Setting rows in Table 3 refer to elements of the PICOTS (Populations, Interventions, Comparators, Outcomes, Timings, and Settings of interest) criteria used to establish a framework for the review as a whole. In this table, the Timing and Setting rows refer to the overall inclusion and exclusion parameters related to timing and setting that were applied to all of the articles considered during the review (across all KQs).
TEP Member 10	Methods	[pg. 12]In Table 3, shouldn't "Outcomes Measured pre-delivery" also be an exclusion criterion for the variable, Setting?	In this table, the Setting row is an element of the PICOTS framework and refers to the overall inclusion and exclusion parameters related to setting that were applied to all of the articles considered during the review (across all KQs). We are comfortable indicating the exclusion criterion of “outcomes measured pre-delivery” in both the Timing and the Outcomes rows, but do not believe it would be helpful to also list it in the row for Setting where we describe the inclusion/exclusion parameters relating to the physical locations in which studies were conducted.
TEP Member 10	Methods	Under the Study Selection section and else where as appropriate, mention the qualifications of the reviewers and arbitrator.	We have added a description of the reviewers’ expertise in the Methods section on Study Selection.

Commentator & Affiliation	Section	Comment	Response
TEP Member 10	Methods	[ES-8]pg 15 of 189 line 10, should patient income be rewritten as "family income"?	We have made the suggested revision.
Coyne, James	Methods	[pg. 16] Quality Assessment of Individual Studies The quality assessment is decidedly primitive in terms of the evaluation of screening instruments. First, studies not rated in terms of whether cutpoints were allowed to freely vary so that the "optimal" cut point that is identified capitalizes on chance or idiosyncratic characteristics of the sample. Second, it is not noted whether studies excluded patients who had already been identified and were being treated. Think of it: what if evaluation of mammography failed to exclude women already diagnosed with breast cancer or currently receiving treatment for breast cancer?	As stated in the Methods section, we used the QUADAS system for grading studies of screening instruments. While we agree that reporting across a range of cutpoints is preferred, almost all of the studies did report values for sensitivity and specificity across a range of cutpoints. We agree that studies of screening tests that do not exclude subjects with known preexisting disease are biased, and any studies that did so would be judged as high risk of bias. We have also expanded our discussion of what we feel is the most relevant overlooked point—the tradeoffs between false negatives and false positives have not been explicitly considered, and, without knowing what sensitivity and specificity should be, it is impossible to evaluate the clinical or public health value of any test, no matter how methodologically rigorous the evaluation.

Commentator & Affiliation	Section	Comment	Response
Jesse, D. Elizabeth	Methods	Literature review strategy: Since many nurses have published about PPD, you could expand your search to Medline and Cinahl. Clarify if you were only seeking studies that used EPDS and other screening tools and that your reviewers confirmed that there was a PPD clinical diagnosis from clinical interview or instrument such as the MINI to confirm a diagnosis.	As described in the Methods, the MEDLINE database (PubMed) was searched to identify Peer Reviewer-reviewed published literature, as were Embase, PsycINFO, and the Cochrane Database of Systematic Reviews. Additional published evidence was identified through manual review of the reference lists of relevant key articles/reviews. The search strategies (including specific databases to be searched) were discussed and confirmed with the TEP and AHRQ during development of the review protocol. Possible use of CINAHL was discussed at that time; the TEP agreed that the included databases were appropriate and sufficient to identify the relevant literature. The searches were broadly designed to identify any relevant validated screening instrument for depression, including, but not limited to the list provided in Methods section Table 3. Specific search terms for each source are provided in Appendix A. Studies reporting depression outcomes were required to include confirmation of depression with a reference standard. This is specified in Table 3. We have added a statement to the Methods section of the ES to emphasize this point.
Peer Reviewer 1	Results	Yes, the studies are described with clarity and enough detail. Tables were clear.	Thank you.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 2	Results	Yes, detail level is good in both ES and full report. Key questions are detailed and nuanced - ambitious and well done. Appreciate the often missing emphasis on diagnostics test characteristics. I am not aware of any missing literature that would have be eligible for review. Specific ideas in the attachment. None concerning, most small edit ideas.	Thank you. Thank you. Acknowledged. Thank you; these items are addressed individually within this table.
Peer Reviewer 2	Results	[pages 33-39] Results Tables 7, 8, 10, 11, 12: Might be useful to include N and quality score in the first column after author name. There is room for this info and it is cumbersome for the reader to have to search for it in Appendix E.	We have added this information in the indicated tables.
Peer Reviewer 2	Results	Results [pg. 42-43] Detailed synthesis for KQ4: might discuss good quality study first.	We have revised the text to discuss the good-quality study first.
Peer Reviewer 2	Results	Results [pg. 45, l. 13] Insert word “to” between “compared children”	This sentence has been corrected.
Peer Reviewer 2	Results	Results [p. 45, l. 55] What screening tool did 5th study use?	The sentence has been revised to specify the screening tools used in all of the studies relevant to KQ 6.
Peer Reviewer 2	Results	Why is UK listed separately from Europe? In a couple of places (see detailed comments below) the counts don’t add up- staff caught this and wondered if you may have double-counted UK/Europe studies?	Thank you for noticing this error; the numbers have been corrected to reflect the totals per the final set of included studies. The UK was considered separately from Europe primarily because screening instruments were administered in English, enhancing comparability to a US non-immigrant setting.
Peer Reviewer 2	Results	Overview tables for each KQ might be helpful. You could take the information from Appendix E and present it by KQ in each section	Given the small number of studies, which facilitates somewhat more detailed descriptions, we have opted to streamline the main text and leave the detailed study characteristics in the Appendix. We would be happy to revise after additional consultation with AHRQ.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 3	Results	In general, the Figures are clear and helpful. You could perhaps make it more obvious that Figures 5 and 6 show not only varying thresholds, but also two different test versions. In general, the key points for each question seem to be a fair summary of the studies.	We have clarified the figures.
Peer Reviewer 3	Results	[P 23, lines 53-57] "…it is possible to use an initial simple step for selecting patients for more specific screening instruments." This finding deserves greater emphasis in abstract and conclusions.	We have emphasized this further, especially since the recent study by Yawn 2012 shows further support for a two-stage screening process.
Peer Reviewer 3	Results	Pages 46 and 47 are very dense. Edit if possible.	We have revised for clarity.
Peer Reviewer 3	Results	[P 37, Table 10] It is confusing that the Andersson study is divided out into two analyses and that the variable chronic disease is bolded in one and not the other. (Similar situation for 1st trimester BMI). Are you trying to make a point about what variables were controlled for?	We have clarified that the bolding refers to the specific variable under consideration, while the nonbolded variables are the others included in the relevant multivariate analysis.
Peer Reviewer 3	Results	[P 44, line 52] Are more doctor visits really a "harm"? There are many situations in pediatrics where a visit to the physician results in no improvement of the child's "baseline health," but reassures the parents, which is very beneficial. What's the plausible connection between being screened and taking your child to the doctor anyway? In general, I suspect the "harms" of screening, including stigmatization and anxiety, while they need to be considered, may be overblown.	We have added a statement emphasizing that whether extra visits are a "harm" is unclear, and expanded the discussion around this point. The point here is that pre-visit "baseline health" was controlled for, not post-visit. We would argue that, if baseline child health measures are equivalent, then a difference in visits between screened and unscreened parents may be overutilization by screened parents, or underutilization by unscreened parents. Without better outcome measures for children, this remains an area of uncertainty.
Peer Reviewer 3	Results	[P 45, lines 36-43] These differences outcomes based on timing deserve greater emphasis in abstract and conclusions.	An additional study included in the updated search was inconsistent with these previous findings. The additional evidence identified in the updated search also more strongly makes the case for systematic factors more than timing, and we have emphasized this in the revised report.
Peer Reviewer 4	Results	The results section is appropriate--- given that very few articles are included in the report, the results are not particularly long.	Acknowledged.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 5	Results	I liked the presentation of results on the screening instrucments.	Thank you.
Peer Reviewer 5	Results	for KQ5, it is really unclear whether the single RCT from Hong Kong is relevant to US since it is unknown what patterns of utilization are the context. I"m not sure it makes sense to feature this finding in the abstract and some discussion of context of care patterns should be considered.	We have added a statement emphasizing that whether extra visits are a “harm” is unclear, and expanded the discussion around this point in the discussion section, as well as acknowledging the issues regarding applicability given the context.
Peer Reviewer 5	Results	The summary of key points at the top of p42 seems more optimistic than the statement in the abstract.	We have revised the abstract to reflect the revised report.
Peer Reviewer 6	Results	Results are well outlined although the inability to make conclusions despite the 189 pages of this report is not adequately described .	We have revised the discussion section.
Peer Reviewer 7	Results	Overall the amount of detail was appropriate. However, in the detailed synthesis for KQ4 and KQ6 the relevant studies are described in greater detail than for the KQs for which there are a greater number of relevant studies.	We have tried to balance the level of detail.
TEP Member 1	Results	Results section meets all criteria.	Thank you.
TEP Member 2	Results	Results sections had sufficient data to understand the studies.	Thank you.
TEP Member 2	Results	The studies were described well for the intevention---impact of screening but less well defined for the studies of senstivity adn specifiicty. They were well described for the "setting" quesiton.	We have revised our description of the test characteristics studies.
TEP Member 2	Results	The tables are a little difficult to understand quickly. The SOE were very clear in some but less clear in others. The Table for outcomes of interventions---how you determine "imprecise" is not really very clear.	We have revised the tables for clarity.

Commentator & Affiliation	Section	Comment	Response
TEP Member 2	Results	The figures for senstivity and specificty are very clear. The fact that BECK did all of the BECK studies is a little worrisome. Also no comment about the cost of using these tools in clinical practice was made. Beck charges.	Explicit consideration of costs was outside the scope of the report. Although financial (or intellectual) conflict of interest resulting in biased results is possible, there were insufficient studies to directly compare whether results differed in studies where the designer of the instrument (with or without financial incentives) was a co-author compared to those where they had no involvement.
TEP Member 2	Results	Again the Yawn 2012 study should be reviewd and included if possible. Does have 12 month outcomes.	The Yawn 2012 article was identified during the search update and has been included in the final report.
TEP Member 3	Results	It would be reasonable to tone down negative comments about the new haven study since this was not an RCT	The quality criteria are not meant to be pejorative, but to simply convey the relative lack of certainty about the conclusions resulting from aspects of study design or reporting that increase the risk of a biased result.
TEP Member 4	Results	yes. My only gripe about the tables is that the term "SOE" is used in the tables without ever being defined. Took me a few minutes to realize this was "Strength of Evidence".	We have provided a footnote in each of the Discussion section SOE tables to define the acronym.
TEP Member 5	Results	The authors used ClinicalTrials.gov as a method for identifying unpublished studies. I would regard that method as unsatisfactory. The authors should have contacted investigators directly.	As much as possible, EPCs are asked to maintain anonymity during the review process to avoid the potential for undue external influence. In lieu of contacting a large number of individual investigators, the EPC convened groups of stakeholders (Key Informants and Technical Expert Panel members) representing a wide range of interests in the topic area to provide input on the project. These groups included researchers and other experts familiar with current work in the field who were able to alert the EPC team to ongoing research of relevance to the report. Requests made to industry representatives seeking information on interventions and potential unpublished data (as described in the ES and main report Methods) are handled directly through AHRQ's Scientific Resource Center.

Commentator & Affiliation	Section	Comment	Response
TEP Member 5	Results	The authors are probably aware of the trial by Barbara Yawn that was recently published. Again, I remain concerned that the older literature was not included. This particularly affects conclusions for KQ-2	The Yawn 2012 article was identified during the search update and has been included in the final report. The search dates, which were approved by the Key Informants and TEP, were chosen primarily to increase efficiency. We have clarified that our primary goal for KQ 2 was to identify any evidence that using specific risk factors, either alone or as part of a predictive model, improved the performance of screening instruments, and that our rating of the evidence is primarily based on this consideration, rather than epidemiological criteria for establishing causality. We have noted that the findings are consistent with other recent reviews, including the IOM report.
TEP Member 5	Results	[pages 50 and 59] The amount of detail presented about the studies seemed sufficient and their characteristics were described at an appropriate level of detail. The Key Points generally were very clear and the Tables were clear. The Figures were more variable. Figure 9, which is actually in the Discussion, was problematic because of the overlap of many of the symbols. A table might be a much more informative way of presenting the information. In Table 14, the abbreviation SOE was not included in the Table note.	Thank you; we agree with the reviewer's suggestion regarding Figure 9. We have eliminated the graph that was Figure 9 in the draft report and replaced it with data in a table format (Table 20). We have clarified the SOE abbreviation in Table 14.
TEP Member 5	Results	[page 32]On page 32 and later the authors discuss the use of clinical predictive models such as the Gail model for breast cancer. To many readers this will be unfamiliar territory and so the whole idea needs more explication. Moreover, in reviewing the Gail model, it stands on its own; it does not seem to be designed to be used in conjunction with a separate screening tool to improve the performance of the screening tool. As a consequence, this discussion is a bit misleading and not real helpful.	The use of predictive models such as the Gail model is included in guidelines for deciding on whether to use more sensitive imaging techniques, such as MRI, for breast cancer screening—increased risk in the model is associated with an improved positive predictive value with the more sensitive but less specific imaging. We have added appropriate references and clarified the discussion.

Commentator & Affiliation	Section	Comment	Response
TEP Member 5	Results	The authors could have done a better job of clarifying the use of the HRSD. It is typically used as an interview with a mental health clinician. It is inconceivable that such an approach would ever be used for screening.	We present the results of the published study, which directly compared the HRSD to the self-administered screening instruments, as well as clarification of how the HRSD is intended to be used. The choice of any screening test, for any condition, is ultimately based on considerations of both test performance and the resources required to perform the test. Conceptually, a more resource-intensive test might be preferable to a self-administered test if the improvement in overall test performance was markedly superior.
TEP Member 6	Results	The report is appropriately detailed. Figures and tables are easy to understand	Thank you.
TEP Member 7	Results	The summary section will probably be sufficent for clinicians, with the detailed review geared more for statisticians .	Acknowledged.
TEP Member 8	Results	Yes, the results level of detail, study characteristics, and key messages are adequate. Tables are informative, but Figure 9 is difficult to see and understand with graph markers so similar and overlapping. In Figure 10, is the red triangle line for 'screen once epds' missing or just completely covered by the gold line? needs improvement or clarification.	Figure 9 has been replaced with a table, and Figure 10 (now Figure 9) has been revised for clarity, with a footnote to explain that the two “Screen Once” strategies overlap almost exactly.

Commentator & Affiliation	Section	Comment	Response
Coyne, James	Results	[pg. 23]K Q1. Performance Characteristics of Screening Instruments The following statement is startling and suggests some real problems in this literature, particularly given the concerns about cutpoints allowed to freely vary so that results capitalize on chance: "because no more than two studies provide results for the same test of the same threshold, we did not perform analyses." The issue is not whether a metaanalysis appropriate, but whether this literature is suitable for evening narrative interpretation without commenting on this problem of studies commonly not specifying cutpoints a priori, but rather post hoc picking the best for a particular sample.	We agree with the reviewer that the choice of threshold is important. Methodologically, the performance of a test at a given threshold should be independent of whether those thresholds were specified a priori or post hoc (i.e., the sensitivity and specificity are independent characteristics). The key issue, which we believe is the reviewer's point, is that the choice of threshold for use in *a specific setting* should be determined based on considerations of positive and negative predictive value appropriate for that setting, given the tradeoffs between false negatives and false positives at both the individual and system level. Our understanding from both the literature and the comments of other reviewers is that there has been little or no formal, explicit consideration of these tradeoffs in the context of postpartum depression. This point was made in the previous AHRQ review, and we have found no indication that it has been resolved.
Coyne, James	Results	[kq1 pg. 23]The inclusion of the Hamilton Rating Scale for Depression is bizarre. It was never intended as a screening instrument, but as a measure of severity for patients already having a diagnosis of depression. It is lengthy and requires a trained interviewer and is certainly not to be administered by self-report. Why would anyone bother to use the Hamilton Rating Scale rather than simply perform a diagnostic interview if this amount of time is to be invested? The common understanding of screening for depression is that it involves self-report measures to identify patients needing further assessment.	We present the results of the published study, which directly compared the HRSD to the self-administered screening instruments. The choice of any screening test, for any condition, is ultimately based on considerations of both test performance and the resources required to perform the test. Conceptually, a more resource-intensive test might be preferable to a self-administered test if the improvement in overall test performance was markedly superior.

Commentator & Affiliation	Section	Comment	Response
Coyne, James	Results	[pg. 32] K Q2. Effect of Individual Factors on Screening Performance This seems to be little value to this tabulation of risk factors unless some quantification can be provided as to how they moderate the prevalence of depression in the performance of screening instruments. Differences in absolute risk of preferable to relative risks. Otherwise, too much importance is attached to seemingly important differences in relative risk to do not translate into much differences in absolute risk.	We agree that measures of absolute risk are more important for both clinical and policy purposes than relative risk. However, quantifying these risks requires three specific estimates—the relative risk, the prevalence of the risk factor in the population under consideration, and the overall risk of the condition. In particular, the prevalence of the risk factors will vary widely between specific populations.
Coyne, James	Results	[pg. 42] K Q4. Comparative Benefits of Screening, and K Q5. Comparative Harms of Screening There is too little information provided here to allow the reader to do anything but accept the reports conclusions without an opportunity to critically evaluate the basis for them. The evaluations of screening need to be independent of the evaluation of other enhancements of care, including notably collaborative care models of depression. It is known from the literature concerning screening in general medical settings and primary care, that any effectiveness of screening depends on enhanced care, whereas it is not clear that enhancements of care for depression depends on having screening as an element. Furthermore, evaluations of screening need to distinguish between effects for patients already detected as depressed and effects for patients not otherwise detected.	The discussion of the impact of setting has been expanded, with the inclusion of the recent Yawn 2012 paper, which demonstrates high levels of receipt of services in a coordinated care setting. We agree that the effectiveness of screening is dependent not only on the test characteristics of the particular screening instrument, but on receipt of appropriate followup of patients with positive screening results. This is true of any screening test, for any condition.
Jesse, D. Elizabeth	Results	Specify BDI or BDI-II in your tables. Clarify your statements ie employment and PPD or do you mean PPDS.	We have clarified BDI/BDI-II. The tables indicate that employment status has some association with PPD.
Peer Reviewer 1	Summary/Discussion/Conclusion	Yes, the findings from the review of the studies are clearly stated. There is tremendous confusion among consumers and policymakers regarding screening. This report is important in having intelligent discussion between consumer advocates, healthcare professionals, and researchers. Future research must follow the recommendations from this report if progress is going to made in this field.	Thank you.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 1	Summary/Discussion/Conclusion	Possible typos: Page 57, lines 23/24 "that that the" page 61 line 38 "is" instead of "in" [in the phrase “in the test characteristics”]	Thank you; we have corrected these errors.
Peer Reviewer 2	Summary/Discussion/Conclusion	Implications are clear, and as we often find disappointing in terms of guiding clinical care decisions. Given the number of organizations that promote routine screening paradigms and the challenges of timely referral, diagnosis, and treatment, the literature is disappointingly sparse. However, have to say I am not surprised.	Acknowledged.
Peer Reviewer 2	Summary/Discussion/Conclusion	Discussion, [p.53-54 Table 16] Number of subjects is missing from a few rows on this table.	We have added the subject numbers for the remaining rows.
Peer Reviewer 2	Summary/Discussion/Conclusion	Discussion [p. 60 Figure 10]. There is a symbol for “Screen once EPDS” in the legend but no data in the graph; Also see the footnote for this table ‘[To be added]”	Figure 10 (now Figure 9) has been revised for clarity, with a footnote to explain that the two “Screen Once” strategies overlap almost exactly.
Peer Reviewer 2	Summary/Discussion/Conclusion	Discussion [p. 62 l. 37] Do you mean “weighing of harms vs benefits” not “weighting”?	Yes, we have corrected the text.
Peer Reviewer 3	Summary/Discussion/Conclusion	In general, it is disappointing that the state of the evidence is no better in 2012 than in 2005. I acceptthe rationale on p 56, lines 51-57 that more studies are be in progress, but it also seems that either 1)too many studies are executed without proper foresight and planning or 2) we have honed our methodology for evidence evaluation to such an extent that it goes beyond our capabilities to pay for and execute studies we will rely upon. This report makes a nice case for the research gaps and seems to need a clear call for a consensus conference to address them (p 63, General Gaps)	Thank you.
Peer Reviewer 3	Summary/Discussion/Conclusion	P 59, Figure 9: This seems to lend more evidence for the two-step screen proposed by Gjerdingen et al., and again, makes me wonder why this approach is not recommended in abstract and conclusions.	Although the modeling results are intriguing (and intuitively make sense), we believe that a formal recommendation should be based on more direct evidence, which should be a high research priority.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 4	Summary/Discussion/Conclusion	I believe the authors have "missed the boat" re: making recommendations that are clinically meaningful. The inclusion criteria are too restrictive and hence there are not enough studies included to generate conclusive results.	As noted in previous responses, our task was not to make clinical recommendations, but to review the evidence, within constraints of time and other criteria defined by the key informants and TEP.
Peer Reviewer 5	Summary/Discussion/Conclusion	[page 56] "what is already known" should include dsicussion of what is known about treatment for dpression in women at other stages of life.	We have added discussion, primarily based on the USPSTF review and recommendations, of the evidence for screening and treatment of adults in general, although a detailed description is outside of the scope of this report.
Peer Reviewer 5	Summary/Discussion/Conclusion	With growing adoption of depression screening by state Medicaid programs and otehr organizations, there should be some discussion of how observational data on screening and treatment can be used to help to fill in the evidence gaps. This fits with the discussion of potential uses of simulation modeling to address questions regarding timing of screening. (Figures 9 and 10 were hard to read in black & white print copy).	We have revised our discussion of the policy implications, including revision of the tables and figures regarding the modeling results.
Peer Reviewer 6	Summary/Discussion/Conclusion	The conclusions of this report are clearly stated but if they are indeed that it is unclear whether screening translates into improved mental health for patients , then the next steps in terms of policy and implications for practice need to be even more explicitly stated	We have revised this discussion.
Peer Reviewer 7	Summary/Discussion/Conclusion	Discussion section was clear. Table 15 was a great addition and helped concretize more complex discussions. Description of patient characteristics was well presented and easily translated in new research and practice.	Thank you.
Peer Reviewer 7	Summary/Discussion/Conclusion	[Page 56 lines 47-51 and Page 57 line 10] I would like to see a more concise description of US study populations and how they relate to population demographics. The reports explains that many studies recruited populations that did not reflect community characteristics. I would like to see the report explain how they were difference - by race, income, etc.	We have added a qualitative discussion of how the U.S. studies compare to the other included studies.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 7	Summary/Discussion/Conclusion	In terms of applicability, the report correctly acknowledges that differences in culture, gender roles, and health systems affects applicability. the report's applicability cound be increased by summarizing the subgroup of US studies and how/if their results/contributions differ than studies conducted in other countries.	We have added a qualitative discussion of how the U.S. studies compare to the other included studies.
TEP Member 1	Summary/Discussion/Conclusion	Section KQ 4 (Page 9 line 50 and following) does a good job of discussing benefits of screening in terms of outcomes based on the limited data available.	Thank you.
TEP Member 1	Summary/Discussion/Conclusion	The article makes a good point that before further research is done, there should be standardization of variables to collect along with how to measure and report them (page 91 [top=p. 63 bottom] line 28-29).	Thank you.
TEP Member 2	Summary/Discussion/Conclusion	This is well written and clear.	Thank you.
TEP Member 2	Summary/Discussion/Conclusion	As much as I would like the findings to be different, even with the Yawn study I think they will still be insufficient for recommendations by groups like the USPSTF.	Acknowledged.
TEP Member 2	Summary/Discussion/Conclusion	Some consideration might be added for the issue of screening and care in the sme site versus the need in all peadiatric practices and most OB/GYN pracitces for women to be referred out for evaluation and therapy initiation. This may be a major consideration for future studies. Identify ways to keep women's care within the same site.	The study by Yawn et al. (2012), included subsequent to the draft review, provides additional evidence relevant to this point and the discussion has been expanded.
TEP Member 3	Summary/Discussion/Conclusion	findings are clearly stated	Thank you.
TEP Member 4	Summary/Discussion/Conclusion	clear - conclusions are drawn about the effectiveness of the service, which is different from the usual SER that is done for the USPSTF. I realize this CER is not the same thing; I'm just noting that.	Acknowledged.

Commentator & Affiliation	Section	Comment	Response
TEP Member 5	Summary/Discussion/Conclusion	I was struck by the low number of studies in Table 16 that addressed risk factors. This point has already been made and the problem could have been largely or partially remedied by including older literature.	We have included some discussion of the consistency of these findings with older literature as summarized in other reviews, such as the IOM report.

Commentator & Affiliation	Section	Comment	Response
TEP Member 5	Summary/Discussion/Conclusion	Overall, the authors did a good job with the Discussion. Table 15 was very useful in illustrating the importance of population prevalence in determining the rates of true and false positive and false negative given set levels of sensitivity and specificity. The section on Applicability was particularly useful and sobering. The authors rightly point to the differences in the US and European health care systems as a factor in limiting the generalizability of findings of European studies. The section on Implications for clinical and policy decision making still suffers from a lack of serious consideration of the psychological treatment literature for subthreshold depression both in the general population and postpartum women. This is a very important issue and has been glossed over by the authors. The section on simulation modeling is very useful but it could be improved by making it a bit more concrete for the lay reader. Also, it was not clear to me what the authors had in mind for the second more specific test. As noted earlier, the Figure is less helpful than it could be. A Table could be added with the actual numbers would be less cluttered and more informative. Similarly, the paragraph describing the simulation represented in Figure 10 could be improved by making it a bit more concrete. Nevertheless, the discussion of the importance of false positives and false negatives was useful and important. It might be worth elaborating a bit on the issue of why a woman would not pursue help in some way if she screened positive. Of course, there are many possible explanations. One important explanation is that many or most women understand the context that surrounds their responses to a screening tool and are able to weigh the potential benefits of pursuing versus not pursuing further help.	Thank you. 1) We have revised Table 15 to further illustrate the population impact of the interaction of sensitivity, specificity, and prevalence. 2) We have expanded the discussion of applicability to point out broader differences in other social systems that may interact with maternal depression and child outcomes. 3) We have included a broader discussion of this issue, although our focus, confirmed by the TEP, was always on major depression. 4) We have revised the section on modeling, and replaced Figure 9 with a table, as suggested by the reviewer. 5) We have expanded the discussion of the implications of false positives and false negatives to include both longer term child outcomes, and the potential resource issues created by screening in a setting where false-positive rates are so high.
TEP Member 5	Summary/Discussion/Conclusion	The authors did a commendable job of identifying gaps in research that bear on the various research questions.	Thank you.

Commentator & Affiliation	Section	Comment	Response
TEP Member 5	Summary/Discussion/Conclusion	[page 65] The last paragraph – Conclusions – lands with a great big thud. It would seem that the authors are conceding that nothing has been learned from their work and that ultimately if asked by a clinician whether it would be useful for he/she to start perinatal depression screening, the authors would simply shrug their shoulders and say "I don't know." The USPSTF has a recommendation for depression screening in primary care, a recommendation which includes postpartum women. The work that underlies that recommendation and the larger societal context has to be recognized in the discussion. This review will have research implications but it will also have policy, clinical and social implications and these cannot be ignored in the narrative discussion.	We have revised the discussion to address these points.
TEP Member 6	Summary/Discussion/Conclusion	The implications are clear. It really has not changed too much from the report in 2005. More questions arise, which is typical. The report clearly reflects the areas that need more study or why there is little change.	Thank you.
TEP Member 7	Summary/Discussion/Conclusion	I am not an expert in this area but I beleive the autors discussed in suffcient detail that problems in reviewing this area of science, and they make the case for their assessment of insuffcient evidence.	Thank you.
TEP Member 8	Summary/Discussion/Conclusion	The discussion/conclusion is clear and thorough, with useful suggestions for future research, especially the need for long-term follow-up.	Thank you.
TEP Member 9	Summary/Discussion/Conclusion	Yes. The future research section is clear. The section beginning on page 91 is very perfect.	Thank you.
TEP Member 10	Summary/Discussion/Conclusion	General comment: One recommendation should be that the KQs should be integrated into NIH's National Child Health Study and/or HRSA's National Survey of Chilren's Health.	We have added references to including more data in these surveys.

Commentator & Affiliation	Section	Comment	Response
Coyne, James	Summary/Discussion/Conclusion	[pg. 57]Applicability Repeating the usual claims about the EPS as being considered "standard of care," is rather misleading when it is followed by statements of the range of cutpoints and lack of description of testing protocols. Overall, this report can be seen as lending support to the continued use of the EPDS when the literature so poor. How can one recommend the instrument without specifying a cutpoint? Arguably, instruments on a validated, cutpoints on particular instruments are.	We have emphasized the importance of choosing a screening test based on specific considerations regarding positive and negative predictive value.
Coyne, James	Summary/Discussion/Conclusion	[pg. 63] Research Gaps I am disappointed that this report does not clarify that screening is best evaluated in a context where patients' access to care is determined by their screening score and the comparison is patients having access to the same care without screening provided. It appears that some of the studies made treatment available to patients assigned to screening that was not otherwise available to patients in the control group so that one cannot distinguish between screening and availability of treatment.	The ideal evaluation for any screening strategy, regardless of condition, involves randomizing subjects to screening or no screening, with treatment in screened subjects contingent on test results, and treatment in unscreened subjects contingent on clinical signs or symptoms detected by the patient or the provider in the context of usual care. Such a study design incorporates all the important aspects of screening, from test performance to appropriate followup of test results to effective treatment. To the best of our ability to determine from the published papers, none of the included studies denied care to unscreened women who either sought care because of symptoms or were referred by providers. We agree with the reviewer that such a study design would be uninformative, as well as unethical.

Commentator & Affiliation	Section	Comment	Response
Coyne, James	Summary/Discussion/Conclusion	[pg. 63] Research Gaps Whether screening increases referrals is a poor proxy for whether it actually improves patient outcomes. In many fragmented and dysfunctional systems -as much of routine care is- referrals are unlikely to be completed and when they are completed, they are unlikely to lead to engagement with evidence-based treatments of sufficient intensity to improve outcomes. One only gets a dim sense of this issue when the low rate of completion of referrals and the only marginal improvement associated with screening is mentioned on page 47.	We agree that increasing referrals is not a particularly useful surrogate. We have included additional evidence published subsequent to the draft report which directly addresses this question.
Coyne, James	Summary/Discussion/Conclusion	[pg. 63] Research Gaps We need to keep in perspective that questions of screening for perinatal depression need to be evaluated in the context of what is typically inadequate routine care, in terms of the absence of minimal efforts to monitor symptoms and side effects and make appropriate adjustments in treatment over time. Why put more women into grossly inadequate care, rather than attempt to improve the quality of that care? Why bother to screen patients if there is such a module effect on receipt of treatment, putting aside issues of the quality of that treatment?	We agree with the reviewer that without adequate diagnosis and treatment, there is no point to screening no matter what the performance characteristics of the test.
Jesse, D. Elizabeth	Summary/Discussion/Conclusion	I am sorry I had to stop here. I cut and pasted my comments from abstract I believe your conclusion is overreaching with to few studies (and even fewer from US). Just as a screening instrument for diabetes is sensitive, the person would still be ill without a plan to address the illness/symptoms. Screening for symptoms of RISK for PPD is a great first step but there needs to be further evaluation and diagnosis of PPD and a plan, ie therapy, support network, follow up. Bests of luck!	We agree that appropriate followup of screening test results is critical for the overall success of screening.
Peer Reviewer 2	Appendix E	Appendix E Might be more useful to have ordered by KQ if possible and by quality within KQ.	We considered a KQ-based ordering approach, but with the significant number of studies that are applicable to multiple KQs, we believe it to be easiest for users to navigate the table if the studies are presented in alphabetical order.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 1	General: Quality of the report	Superior	Thank you.
Peer Reviewer 1	General: Clarity and usability	Based upon how well the report has been organized, the conclusions should lead stakeholders to discuss how they can make informed policy/practice decisions.	Thank you.
Peer Reviewer 1	General	Yes, this report is clinically meaningful, the target population and audience explicitly defined and the key questions appropriate, well defined and stated.	Thank you.
Peer Reviewer 2	General: Quality of the report	Good	Thank you.
Peer Reviewer 2	General: Clarity and usability	Yes, the main point which I would thumbnail much as the conclusion in the abstract is clear and clearly supported without being overblown.	Thank you.
Peer Reviewer 2	General	The key questions tackle the issues that are precisely those which the clinical care provider finds most pressing: How well do screening tools work? Is there reason to prefer one over another? When should we screen? Does it help to screen? Could we be causing harm if we implement a screening program? Ambitious scope. Like the Analytic Framework. Nicely done.	Thank you.
Peer Reviewer 3	General: Quality of the report	Good	Thank you.
Peer Reviewer 3	General: Clarity and usability	This is the second AHRQ Evidence Report I've reviewed. The structure and organization are excellent.	Thank you.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 3	General	I found this report to be frustrating. The methodology is clearly described and of the highest quality. However, it's methodology that only the most accomplished academics will understand or care about. The report is discouragingly short on guidance that will be meaningful to clinicians. The overall conclusion for example: "The available evidence did not allow us to draw any conclusions about the balance of benefits and harms of screening for postpartum depression, or whether specific tools or strategies would result in a more favorable balance."	Our task in conducting the review is to evaluate and synthesize the evidence and provide an assessment of the level of certainty regarding the answers to particular key questions; the EPC does not make recommendations or guidelines. Using the framework for the review agreed upon by the Technical Expert Panel, our judgment was that, for the majority of questions, the evidence does not allow conclusions (put another way, there is still considerable uncertainty about the effectiveness of screening for postpartum depression in improving outcomes for mothers and children). We share the reviewer's frustration that the available evidence does not provide meaningful guidance for clinicians; unfortunately, this is too often the case for many clinical conditions. One of the goals in any evidence review is to identify areas of uncertainty that could be resolved by additional research, which we have attempted to do in this report.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 3	General	Thank you for the privilege of reviewing this report. The key questions are appropriate and the methodology is peerless. In the spirit of constructive peer review, however, I feel the report is not as clinically meaningful as it could be. I am an academic and I understand the limits of evidence. However, I believe there are nuggets of evidence scattered throughout the document that deserve greater emphasis and might be clinically meaningful. Conclusions could be phrased with more conviction, for example: -Screening is associated with fewer depressive symptoms and improvement in overall mental health [low level evidence, p 42] -All screening tests had acceptable sensitivity/specificity characteristics, and when directly compared showed no significant performance differences. [p 29 and elsewhere]	Thank you. We understand the reviewer's desire for the report to be as clinically actionable as possible, but do not wish to overstate any conclusions considering the limited evidence available. We have made the following alterations: The discussion of the evidence of screening effectiveness has been revised, especially in light of the recent RCT published by Yawn et al. (2012). We believe that the determination of "acceptable" sensitivity and specificity should be based on explicit considerations about the clinical and resource implications of the tradeoff. Although the reported test characteristics are similar, there are few direct comparisons, and again, the "significance" of differences depends on the tradeoffs.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 3	General	-A two-question test with follow-up diagnostic testing is a promising strategy [p 23 and consistent with UK guidelines p 3] -In populations where prevalence of depression is low, screening strategies will yield false positive results [p 60] -Women at highest risk are those with a history or family of depression, and those with poor social support. (p) -Overall rates of referral and treatment are low, but better among women screened in the third trimester or at delivery [p 45]	We have reemphasized this point. We have provided a quantitative example of the effects of prevalence. We have clarified our summary of risk factors. Based on additional evidence identified in our updated search, we have revised our discussion of factors affecting referral/treatment rates.
Peer Reviewer 4	General: Quality of the report	Good	Thank you.
Peer Reviewer 4	General: Clarity and usability	The writing of the report is clear and clearly presented. However, the results are not helpful. They essentially say--- we do not know-- and that further research is needed. I believe that if the authors had been less restrictive, there would have been a different approach taken in the inclusion/exclusion criteria and perhaps different conclusions would have been reached.	The scope of the report and key questions arose from the original request for the review, with significant input from the topic nominator, AHRQ, and the Technical Expert Panel. All aspects of the report protocol, including the inclusion/exclusion criteria, were reviewed by the TEP and revised based on their input. Although we agree that different inclusion/exclusion criteria might have led to different conclusions, we cannot speculate what those might have been without more specific suggestions about alternative criteria.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 4	General	I believe that this report is not as clinically meaningful as it should be and that the inclusion criteria were too restrictive for the articles included. It is clear that this report does not have perinatal psychiatry expertise and that the authors do not have a clear understanding of the field and the clinically meaningful aspects of the perinatal depression. It is disappointing that a perinatal psychiatry expert was not part of the team and this is a significant limitation of the report.	One of the members of the review team is a psychiatrist whose practice largely consists of perinatal and postpartum patients; we also received significant input from a psychiatric social worker whose practice is limited to pregnant and postpartum patients. As noted in both the report and other responses, the inclusion/exclusion criteria were developed with significant input from the Key Informant group and the TEP, which included multiple members with experience in perinatal psychiatry. We would also argue that many of the clinical issues involved with evaluating screening strategies, which is the focus of this report, are more relevant to the clinicians who will be performing the screening. We would certainly be willing to discuss any limitations resulting from these criteria, but, without a specific recommendation from the reviewer about alternative inclusion/exclusion criteria or what clinically meaningful aspects of perinatal depression were overlooked on the basis of the inclusion/exclusion criteria, we cannot address the reviewer's general concerns. We have added additional discussion specifically of the potential impact of the study publication date criteria.
Peer Reviewer 5	General: Quality of the report	Good	Thank you.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 5	General: Clarity and usability	Given the sparse amount of reserach meeting the requirements for this evidence review, can the authors be more clear about recommending standard tools and measurement approaches? For example, since the EPDS and the PPDS see m to perform better than other tools -- should there be recommendation to include one of these even if using another tool as well?	The evidence review is primarily meant to provide a synthesis of the evidence for other groups to use in formulating recommendations. Given the relative paucity of direct comparisons of different screening tools, we do not believe the evidence justifies recommendations regarding choice of a specific instrument. Rather than making recommendations based on the test characteristics of a given instrument at a particular threshold, we believe a more useful approach for developing recommendations is to specifically consider the tradeoffs between sensitivity, specificity, screening frequency, and available resources, then identify a particular test and threshold which optimizes these tradeoffs.
Peer Reviewer 5	General	The report is clear and well-written. The approach is well defined. There is limited discussion to the clinical distinction between postpartum depression and major depression in women during other periods of life. What consideration was given to the relevance of evidence about the efficacy and safety of depression screening in the general population for postpartum depression? Given that the study criteria included information on screening during the first 12 months after delivery (which i think is quite reasonable) and that the DSM V is likely to expand the time frame for defining postpartum depression, there should at least be a discussion of whether postpartum women were likely to be included in the general population evidence used to make the existing screening and treatment recommendations.	Thank you. These are very interesting points. Because the EPC approach specifically starts with defining the particular population covered by the review, we did not extensively review the literature on screening in the general population, or directly compare test characteristics across populations. To the extent that the overall effectiveness of screening is dependent on having mechanisms in place to ensure appropriate referral and receipt of services, we believe that the perinatal period and first 12 months postpartum are unique enough to justify focusing on this population. The extent to which postpartum women, particularly late postpartum women, might have been included in the evidence used to make existing general population recommendations is an interesting question, which would require expanding the scope of the review to specifically look at the inclusion/exclusion criteria of the studies cited by, for example, USPSTF.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 6	General: Quality of the report	Good	Thank you.
Peer Reviewer 6	General: Clarity and usability	This was a herculean effort on the part of these investgators . The recommendations (none) and conclusions nonetheless leave the reader who sifts through the voluminous text unsatisfied .	The evidence review is primarily meant to provide a synthesis of the evidence for other groups to use in formulating recommendations. Although the available data were not sufficient to make definitive conclusions regarding optimal screening tools, we believe our conclusions reflect the state of the evidence.
Peer Reviewer 6	General	This is an absolutely exhaustive report which addreses a myriad of issues relevant to safety and effectiveness of screening for postpartum depression . The report reviews numerous studies which relate to the key questions identified a priori as relevant to the matter of screening for postpartum depression . The methodologic issues of concern are exhaustively identified and discussed across the available studies globally since the last AHRQ document on this subject .	Thank you.
Peer Reviewer 6	General	The report while exhaustive , complete and exhausting to read , is of unclear clinical benefit . The conclusions are straightforward and what might be expected : that screening for postpartum depression is not of clear benefit based on available outcome studies in the literature. What is clear is that available screens have reasonable sensitivity and specificity and it is not certain whether improved screens with even greater specificity would be of benefit to patients .	One of the major limitations of this literature is that there does not appear to be a consensus on the definition of “reasonable” sensitivity and specificity. For most screening strategies, the thresholds for sensitivity and specificity (or the choice of a particular test from two or more alternatives with different characteristics) are derived from a consideration of the relative benefits and harms (and often costs) associated with true positive, false positive, true negative, and false negative results, and the likelihood of these different results given a specific pretest probability. We have provided an additional quantitative example of the impact of sensitivity and specificity on the number of true and false positive results (Discussion section, Table 15).

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 6	General	The authors note that major US organizations have recognized the potential benefits of screening for postpartum depression but remain undecided as to the optimum screening instrument . The problem is not with finding an appropriate screening tool but as the authors also note , with the lack of adequate systems for managing positive results . The charge to policy makers and decision makers (p.24) is sparse. Do the authors really believe the issue of clinical meaningfulness /relevance is the sensitivity /specificity of available instruments used to screen for postpartum depression ? Or is the fact that follow-up of the women screened was so abysmally low (5-30%) ? This finding should roar from this report and does not.	Our judgment is that there has not been an explicit consideration of the optimal sensitivity and specificity for a screening tool. We agree that the low rate of followup is a major concern and limits the effectiveness of screening, but providing adequate resources to improve appropriate management of women with positive screening tests explicitly requires an estimate of the number of women who will screen positive, and the proportion among those who do screen positive who will be true positives. These are direct functions of the choice of screening instrument. We have provided an additional quantitative example of the impact of sensitivity and specificity on the number of true and false positive results. Indeed, we would argue that these are not mutually exclusive problems—even at "acceptable" levels of specificity in the 80-90% range, the absolute number of false-positive results is quite high, with 3 to 5 times as many false positives as true positives. The potential impact of a relatively low positive predictive value on the resources needed to improve followup is surely worthy of discussion.
Peer Reviewer 6	General	The discussion of the relative benefits and harms of screening is the weakest part of the report . The discussion of potential harm of screening is particularly poor. There is reference to the problem of overlap of normative symptoms of the puerperal woman and the symptoms endorsed by those with postpartum depression . So this clearly can lead to false positive screens , But the discussion of the " harms" of false positives is lacking despite the enormous volume of the report	The lack of discussion of the potential harms of false positives reflects the lack of evidence in the literature. We explicitly reviewed the literature for evidence regarding harms, and found only one study that addressed a potential adverse consequence of screening. We discuss potential harms such as stigmatization in the report, but if studies do not report outcomes for women with false-positive results, we can do no more than suggest that such outcomes be reported in future research. We note that the broader review of screening for depression in adults conducted for the USPSTF reached a similar conclusion. We have expanded this discussion somewhat, and have added the potential systemic effects of a large number of false positives.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 7	General: Quality of the report	Good	Thank you.
Peer Reviewer 7	General: Clarity and usability	Because policy decisions are not always based on best evidence and there are more states considering mandating/encouraging PPD screening, it would be helpful to strengthen the discussion of screening approaches that, while not necessarily evidence-based, at least use available evidence. The suggestion to conduct 2-stage screening - 1st wtih 2 question screener and then following up with full screen is very helpful. Further discussion of how to target high risk populations or the implications of adding a question about prior history of depression to screening tools would be helpful.	We have provided additional discussion of these issues.
Peer Reviewer 7	General	One other concern that I had and was not sure where to express - the report does not discuss how administration method - e.g. self-administered vs. direct administration - might affect screening results.	This is an excellent point. Only one study provided data on self-administered versus directly administered instruments, although the directly administered instrument used (HRSD) was not designed as a screening test, as has been pointed out by several reviewers. Results were broadly similar, and we have made this point in the revised document.
Peer Reviewer 7	General	The report is well written, target population and audience clearly defined, and key questions appropriate and explicitly stated. It is quite detailed but the reader is able to negotiate the specific sections and find desired information - either summarized or in great detail.	Thank you.
TEP Member 1	General: Quality of the report	Superior	Thank you.
TEP Member 1	General: Clarity and usability	Report is well structured and clear. Main points are easy to delineate, but suggested next steps could be better spelled out.	Thank you; we have extensively revised our discussion section to address these points.
TEP Member 1	General	Report does a great job addressing screening for postpartum depression. Relevant populations and screening tools are addressed.	Thank you.
TEP Member 2	General: Quality of the report	Superior	Thank you.

Commentator & Affiliation	Section	Comment	Response
TEP Member 2	General: Clarity and usability	It is usable for both the clinical person, the researcher and the guideline developer. Interesting that only the AAP recommends screening in the US.	Thank you.
TEP Member 2	General	Teh report is clinically meaningful and does consider several clinical issues that should be taken into account such as the issues of screenng in pediatrics.	Thank you.
TEP Member 2	General	However, not much is said about the major problem of screening at a site that requires referral to another site. A recent publication by Yawn et all addressed this--pubished in ?????. Gjeerdigen et all also addressed this and that manuscript is referrenced but that point is not highlighted.	The Yawn 2012 paper referenced by the reviewer was published subsequent to the completion of the draft report. It has been included in the updated literature review, and the implications of its findings are discussed in the revised report.
TEP Member 2	General	The review did also not include a major AHRQ funded study that was not published until July 2012 but should be included if possible since it also provides positive outcomes for depressive symptoms and agrees with another paper that parenting Stress Index is not improved adn is in fact associated with failure for depressive symptoms to improve. Published in Ann of Fam Medicine.	The Yawn 2012 paper referenced by the reviewer was published subsequent to the completion of the draft report. It was included in the updated literature review and the implications of its findings are discussed in the revised report.
TEP Member 2	General	Please consider using term Famaily Physician ===NPs consider themselves practiioners, and you should have comparable status for Family physicians to pediatricians and OB/GYN.	We have revised the reference to family practitioners in the Introduction to the following: “Screening is often focused during pregnancy or the first 3 postpartum months in settings where care is provided to pregnant or postpartum women by providers such as obstetricians, family physicians, or nurse-midwives.” We do include the term “family practitioner” as one of the possible provider examples in KQ 3c. This language was finalized with significant input from the topic nominator, AHRQ, the Key Informants, and the Technical Expert Panel (which included family physicians).
TEP Member 3	General: Quality of the report	Superior	Thank you.

Commentator & Affiliation	Section	Comment	Response
TEP Member 3	General: Clarity and usability	the conclusions are limited by existing data	Acknowledged.
TEP Member 3	General	This is a very thorough review	Thank you.
TEP Member 4	General: Quality of the report	Good	Thank you.
TEP Member 4	General: Clarity and usability	Yes. Very clearly written and a pleasure to read.	Thank you.
TEP Member 4	General	Generally very clear and complete. a few small comments on the document attached for the authors. The KQs are stated well. However - for the purposes of making a policy recommendation - I think that the second half of the AF has been left off. KQ 4-6 address the overall effectiveness of screening, for which the evidence is insufficient. Given that, the AF would include both the test characteristics (KQ 1-3) AND further KQs about the effectiveness and harms of therapy/intervention. So - for hte purposes for which I would want to use this review (the USPSTF) - half of the AF and the KQ's that would come from that half is/are missing.	We agree that questions regarding the effectiveness and harms of downstream therapy/interventions are of significant interest. Such questions are currently being pursued in a second AHRQ comparative effectiveness review on the effectiveness and safety of treatment options. We have added clarification in the report that a second review will address questions regarding treatment.
TEP Member 5	General: Quality of the report	Good	Thank you.
TEP Member 5	General: Clarity and usability	The report is well-structured and organized. The main points are clearly presented. I have noted my reservations about issues that bear on policy and practice decisions already. The implications for research are clear. I fear that the implications for policy and practice are not clear and that the report could be used as a basis for diminishing support for identifying perinatal depressed women and providing help for them and their offspring.	To the extent that policy and practice are (or should be) evidence-based, our conclusion that more specific evidence is needed to inform that policy and practice is quite different than a conclusion that identifying and providing care for perinatally depressed women is not worthwhile (in other words, “absence of evidence is not evidence of absence”). We have emphasized this in the revised discussion section and executive summary.

Commentator & Affiliation	Section	Comment	Response
TEP Member 5	General	Overall the authors of this report did a very good job synthesizing the research that they intended to review. Much of the report is clinically meaningful; however, there were instances of jargon that diminished the effective of the communication.	Thank you; we have made efforts to clarify the language where possible in order to make the findings accessible to a wide range of end users.
TEP Member 5	General	The target population was clear – assuming that it was perinatal women. The audience was not as clear. The report seems to be written for a relatively sophisticated audience. The authors could probably do a better job of writing it in a way that the average clinician and lay person could appreciate. There will be many individuals from advocacy groups and from lay networks who will read this report and try to make some sense of it.	Thank you; we have made efforts to clarify the language where possible. We anticipate the report may be of interest to a wide range of stakeholders and end users, from patients and clinicians to experts in the field. We hope to strike a balance of utility across this spectrum.
TEP Member 5	General	The key questions are appropriate and explicitly. The authors did a good job in this regard.	Thank you.
TEP Member 6	General: Quality of the report	Superior	Thank you.
TEP Member 6	General: Clarity and usability	Very well organized	Thank you.
TEP Member 6	General	The report has again identified there is still much needed information. It is important to identify areas of specific research needed to answer the questions. The key questions are appropriate but as concluded more information is needed.	Thank you; we have noted areas in which further research is needed throughout the report and have attempted to provide additional specific research recommendations in the Research Gaps section of the Discussion.
TEP Member 7	General: Quality of the report	Superior	Thank you.
TEP Member 7	General: Clarity and usability	The report is well structured and organized, with the technical details placed after the summary section.	Thank you.
TEP Member 7	General	This is a well-written report on a complex subject that is importnat to maternal-child health. Yes, the report is clinically meangingful, and the key questions are stated clearly.	Thank you.

Commentator & Affiliation	Section	Comment	Response
TEP Member 8	General: Quality of the report	Good	Thank you.
TEP Member 8	General: Clarity and usability	--in the executive summary, insert the meaning of the acronym 'SOE' in the title of Table A for clarity --p. ES-16, lines 49-51: the sentence "Many studies had highly selected samples...which limits these findings to broader populations." is missing something, ie that the applicability of the findings to broader populations is limited.	We have inserted strength of evidence (SOE) in all the SOE table captions in the ES and Main. The sentence has been revised to clarify the implications for applicability: “Many studies had highly selected samples due to high rates of nonresponse or attrition during the study period, thus limiting the applicability of the findings to broader populations.”
TEP Member 8	General	The nomination was to evaluate evidence to support the possible adoption of screening for postpartum depression (PPD) as a measure of health services quality, within the scope of CHIPRA, which aims to promote healthy child development. Thus, the intent was to inform policy, but this is also a clinically relevant question and report, as there has been uncertainty about the appropriate role of screening for PPD in peripartum care. The target populations and audience are well-defined. The key questions address the relevant factors and are explicitly stated.	Thank you.
TEP Member 9	General: Quality of the report	Superior	Thank you.
TEP Member 9	General: Clarity and usability	Yes.	Thank you.
TEP Member 9	General	Clear and free of bias. This will be hugely helpful in light of the over reaching recommendations for mandated screening coming from elected officials and their representatives. It's nice to see the state of science presented without political bias.	Thank you.
TEP Member 10	General: Quality of the report	Superior	Thank you.

Commentator & Affiliation	Section	Comment	Response
TEP Member 10	General: Clarity and usability	Yes, I believe the report is well structured and organized. The page numbers need to be reconciled in the Table of Contents but I trust that will happen. Given the findings, the conclusions should inform policy and research planning but I am concerned about its influence on practice. It is vital that the conclusions do not convey that screening, theraputic interventions, and referral for care are not beneficial.	To the extent that policy and practice are (or should be) evidence-based, our conclusion that more specific evidence is needed to inform that policy and practice is quite different than a conclusion that identifying and providing care for perinatally depressed women is not worthwhile (in other words, “absence of evidence is not evidence of absence”). We have emphasized this in the revised discussion section and executive summary.
TEP Member 10	General	[ES-1] This is indeed a very important and exhaustive report. The first sentence on page 8 of 189 is very powerful in its emphasis on the universal need for timely, effective PPD screening. In light of this, it is disappointing that the science does not meet this need.	Acknowledged, thank you.
TEP Member 10	General	The page numbers on the Table of Contents do not correspond to the page of the respective section. For example, the TOC states Abbreviations can be found on page 73. In the PDF version, the 1st page of the Abbreviations section is labeled page 101 of 189 while the Adobe page counter states it is page 102 of 190. This discrepancy made it very difficult to find various reference Tables and Figures.	Acknowledged. The PDF creation process within the draft report submission system leads to duplication of page numbering. The final report will be free of these discrepancies.
TEP Member 10	General	The word "breastfeeding" appears throughout the Exec Summary, report, and tables. It appears without explanation or clarification until (from what I can tell) page 114. Early on in the Exec Summary and the report, define the features of breastfeeding that this review focused on. Initially using terms like "Ever-breastfed during infant's first 12 months" and "Breastfeeding Duration for most recent child (or delivery or postpartum period) will clarify the breastfeeding context for the reader.	We have clarified to provide specific examples—our intent was to include outcomes primarily related to either breastfeeding initiation or duration.
TEP Member 11	General	This report provides a very thorough analysis of these existing studies, and their research gaps that make it difficult to address the Objectives of the report.	Thank you.

Commentator & Affiliation	Section	Comment	Response
TEP Member 11	General	Concerns about this report from the perspective of a clinician: 1) The Conclusions of this report do not expressly state that, based on these studies and their limitations, there is insufficient evidence…There needs to be a statement that studies thus far have not looked at routine screening with systematic follow-up. The report conclusions as stated, have serious implications in that the interpretation may be that postpartum depression screening should not be part of clinical practice. 2) The study uses only major depression as a target, when mild-to moderate depression can also affect maternal and mother-infant dyad function. 3) Rates of referral and follow-up in the studies were low. Therefore conclusions about the efficacy of screening cannot be drawn from this analysis. 4) Referral for psychiatric care is not the only recommended treatment for postpartum depression. 5) The impact of maternal depression on early brain development was not discussed as a possible harm. 6) No outcomes for the social-emotional development of the infant or the motherinfant dyad were included in the studies reviewed. 7) The increased number of infant doctor visits for women who had been screened is seen as a harm, when it may indeed reflect that the dyad sought and received beneficial support.	1) We have included new evidence published subsequent to this draft review that is relevant to this point. 2) We have included a brief discussion of the evidence that depressive symptoms affect maternal function and mother–infant dyad, but this is not equivalent to evidence that treatment specifically improves outcomes, as noted in the IOM report. 3) We would argue that referral and followup are an integral part of “screening”. If adequate rates are not achievable in practice, then screening cannot be effective. 4) We have revised the report to indicate that there are a range of potential treatments, which can occur in a variety of settings. 5) We have included the association of maternal depression with developmental issues, including cognitive development, as a potential harm. 6) We have identified this as a significant limitation of the literature. 7) This is an excellent point, and we have expanded our discussion of the limitations of this easily measurable outcome as a surrogate for either benefit or harm in the absence of data on more clinically relevant outcomes.

Commentator & Affiliation	Section	Comment	Response
TEP Member 11	General	The report does make it clear that studies of routine screening (by the OB and by the infant's PCC) with specific follow-up/referral strategies are needed. As this report infers, serial screening over the period of peaks in incidence of postpartum depression is preferable. Here in North Carolina, primary care clinicians who have been doing such screening for the last 5-10 years report positive responses by mothers for just being asked and enhancement of the clinician-family relationship. Therefore a measure of family satisfaction would also be appropriate. Other outcome measure would include: referral/follow-up rates for the mother and for the dyad; results of screens of the infant's social-emotional development; adherence to well-child visits and preventive care.	We have included discussion of these outcomes. We did not discuss the potential implications of serial screening in the draft report—we have added a discussion of the tradeoffs (more opportunities to detect depression, a recognition of variation in the timing of onset of depression vs a cumulative greater number of false positives).
Block, Robert (American Association of Pediatrics)	General	Although several of our members submitted specific, individual comments online, the Academy is very concerned that the conclusions of this report may be misinterpreted and adversely impact the clinical care provided to maternal-child dyads across the country. While the analysis of the studies considered in the report is thorough, our experts feel strongly that there are major gaps and limitations to the research conducted to date. To the extent that this report draws attention to those gaps and encourages new research to address those limitations in our knowledge, the Academy is supportive.	We have attempted to clarify to the extent possible that the state of the evidence is insufficient to draw a conclusion about overall net benefits, rather than that the evidence points to no benefit.
Block, Robert (American Association of Pediatrics)	General	In particular, the Academy would like to emphasize that the studies published thus far have not looked at routine screening with high, systematic follow-up rates. This is a critical point because the absence of strong empiric support for positive PPD screens leading to improved maternal outcomes is likely due to an inaccessible system of mental health services for postpartum women. Indeed, the sensitivity and specificity of the screens suggest that it is not the screens that are faulty, but an inability of the health system to adequately address the needs identified by the screens.	We agree that this is an important point, and we review evidence published subsequent to the draft report that shows higher followup rates in some settings. As noted in responses to other comments, the resources needed to "adequately address the needs identified by the screens" are a direct consequence of the sensitivity and specificity of the test used for screening.

Commentator & Affiliation	Section	Comment	Response
Block, Robert (American Association of Pediatrics)	General	The Academy would also like to highlight the lack of outcome measures that assess the quality of the mother-infant dyad or the social-emotional development of the infant. Epigenetic, neuroscientific, and epidemiologic data make it clear that any condition that interferes with the serve and return interaction between infant and caregiver (like maternal depression) can have lifelong, even intergenerational effects. The emerging science base for this strong assertion is outlined in the Academy's recent Technical Report "The Lifelong Effects of Early Childhood Adversity and Toxic Stress," and it serves as the foundation for the Academy's Policy Statement on "Early Childhood Adversity, Toxic Stress, and the Role of the Pediatrician: Translating Developmental Science into Lifelong Health." From a life-course perspective, identifying and addressing maternal depression is one of the most powerful interventions a pediatrician can make.	We have emphasized the importance of including this outcome in future research, along with a discussion of the methodological challenges to establishing that screening for postpartum depression with appropriate management of women with positive results itself leads to improved child development.
Block, Robert (American Association of Pediatrics)	General	Furthermore, studies that only target major depression exclude those dyads with mild to moderately depressed mothers. Such dyads may benefit greatly from early intervention, even though the intervention may not constitute formalized psychiatric care (referrals to home visitation programs, support groups, or services that review videotaped dyad interactions should be considered as potentially appropriate interventions).	We agree; although our estimates of test characteristics were based on a diagnostic threshold of major depression, our review of overall effectiveness included a wide range of outcomes. Although we agree that there is certainly potential for improving a range of outcomes, we did not identify any studies that specifically provided evidence that screening using any threshold resulted in improved outcomes for the mother–infant dyad. We have also noted that sensitivity for "major or minor" depression is typically lower, which raises additional issues regarding the tradeoffs between sensitivity and specificity.
Block, Robert (American Association of Pediatrics)	General	The Academy was also concerned that the 2009 Institute of Medicine (IOM) Report Depression in Parents, Parenting, and Children: Opportunities to Improve Identification, Treatment, and Prevention was not cited.	We have added a discussion of the IOM report to both the introduction and discussion sections.

Commentator & Affiliation	Section	Comment	Response
Block, Robert (American Association of Pediatrics)	General	Again, the Academy is supportive of all efforts to encourage new research to address these limitations in our current knowledge. However, the Academy is concerned that the conclusions of this report, as stated, may lead to the misinterpretation that postpartum screening should not be a part of clinical practice. This would be grave miscalculation and could do a great deal of harm. Postpartum depression has serious consequences not only for the mother, but for optimal early brain and child development. In 2010, the AAP published a clinical report on "Incorporating Recognition and Management of Perinatal and Postpartum Depression Into Pediatric Practice," which addresses the impact of maternal depression on the young infant, and emphasizes the role of the primary care clinician in recognizing perinatal depression. On-going questions regarding the standardized utilization of PPD screens (which screen to use, where, and when) and the undeniable need for improved access to evidence-based support services do NOT obviate the very basic need to identify and treat mothers with perinatal depression in order to improve the child's prospects for learning, health, and productivity across the lifespan.	We have attempted to clarify to the extent possible that the state of the evidence is insufficient to draw a conclusion about overall net benefits, rather than that the evidence points to no benefit, and to emphasize the need for additional research to clarify these points.

Commentator & Affiliation	Section	Comment	Response
Block, Robert (American Association of Pediatrics)	General	The Academy encourages AHRQ to emphasize the numerous limitations of the studies under consideration in the current report, to leverage those limitations to encourage new research, and to clarify that the current data do NOT suggest that PPD screening is harmful. Indeed, the increased number of infant doctor visits for women who had been screened may actually reflect dyads who sought and received beneficial support. Most importantly, the Academy believes that the unquestionable damage that PPD can play in disrupting infant development needs to be highlighted to underscore the urgent and dire need for the pediatric, mental health, and research communities to develop and test innovative ways to identify and address PPD. In the meantime, the Academy will continue to encourage our members to identify and assist struggling dyads.	We have attempted to clarify to the extent possible that the state of the evidence is insufficient to draw a conclusion about overall net benefits, rather than that the evidence points to no benefit, and to emphasize the need for additional research to clarify these points.

Commentator & Affiliation	Section	Comment	Response
Coyne, James	General	My overall reaction to this draft report is one of disappointment and a sense of a missed opportunity to raise questions concerning some endemic problems in the literature. Most basically: (1) evaluations of screening instruments should be in terms of a priori cut points and not cut points allowed to freely vary and such evaluations should exclude already identified cases; (2) evaluations of screening need to be distinguished from evaluations of enhancements of care or differential access to treatment based on screening. I would like to have seen a document that explicitly identified problems in the existing literature so that they not perpetuated. We need less research, but more research a better quality. I have taken the liberty of also including a link to a report that my colleagues and I prepared for the Canadian Medical Association Journal about screening in primary care settings that gives these issues more tension and elaboration. https://dl.dropbox.com/u/23608059/Thombs%20-%20CMAJ%20%282011%29.pdf	We appreciate the reviewer sharing the report, and these are excellent points. In regard to (1), we excluded studies that included already identified cases. We have argued more strongly in the report that the evaluation of screening instruments should be based on an explicit consideration of the tradeoffs between false-negative and false-positive results, rather than an arbitrary standard of "acceptable" or "reasonable" sensitivity and specificity. For scaled tests, we believe it is reasonable to vary the threshold for "normal" in order to estimate the test characteristics at each threshold, but we agree with the reviewer that, for evaluation purposes, the choice of threshold should be based on explicit *a priori* considerations of the tradeoff between sensitivity and specificity. With regard to (2), we believe that given appropriate study design, it is possible to evaluate the effectiveness of screening (in terms of identification of true positive cases) and enhancements of care (in terms of receiving appropriate diagnostic and treatment services) simultaneously, since ultimately, the two are integral for evaluating the overall effectiveness of "screening."