Comparative Effectiveness Research Review Disposition of Comments Report

Research Review Title: *Progestogens for Prevention of Preterm Birth*

Draft review available for public comment from July, 2010 to August, 2010.

Research Review Citation: Likis FE, Andrews JC, Woodworth AL, Velez Edwards DR, Jerome RN, Fonnesbeck CJ, McKoy JN, Hartmann KE. Progestogens for Prevention of Preterm Birth. Comparative Effectiveness Review No. 74 (Prepared by the Vanderbilt Evidence-based Practice Center under Contract No. 290-2007-10065-I). AHRQ Publication No. 12-EHC105-EF. Rockville, MD: Agency for Healthcare Research and Quality. September 2012. Available at: 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=1239
Published Online: September 7, 2012

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #6	Abstract	The structured abstract concludes that "progestogen treatment reduces the risk of preterm birth in singleton pregnancies in women with prior preterm birth". The second sentence acknowledges that the strength of evidence in support of this is low. However, the first sentence can be taken out of context and interpreted as an endorsement of 17-alpha hydroxyprogesterone caproate when a sister agency of AHRQ, the FDA, has not approved this drug for reasons of safety and efficacy. The information pertinent to this is on the website of the FDA and is in the public domain. I believe it is important that there be consistency among agencies if possible.	We have updated the report to include new information concerning the safety of 17-OHP. As above, we find the aggregate estimates of benefit among those with prior preterm birth to be squarely on the side of benefit: effect size in meta-estimate: OR = 0.66; 95% BCI: 0.53, 0.82 and have changed the strength of evidence rating to moderate. (See Table 21). In the interim the FDA has approved Makena a 17OHP injection for prevention of preterm birth as anticipated. We share concerns about lack of safety data and have amplified that in the final version of the report adding discussion of the Combs paper and the general need for larger datasets to assess harms such as IUFD and neonatal death risks. None of the synthesis of data is intended as an endorsement.
Peer Reviewer #6	Abstract	The structured abstract is silent about the safety risks of 17-alpha hydroxyprogesterone caproate uncovered by the FDA. I think it would be responsible to include such information in the PubMed abstract and describe what has transpired with the RCT described in the previous point.	We will update the report to include new information concerning the safety of 17-OHP. The new study by Combs added 81 triplet pregnancies and reported a statistically significant higher rate of intrauterine fetal death with progestogen treatment (7.7 percent versus zero percent; p = 0.01). In this study, the intrauterine fetal death rate was only reported in combination with the miscarriage rate. Thus the 7.7 percent result includes some miscarriages after 16 weeks. When taken in total with other trials of IM 17OHP (Meis, 2003; Fuchs, 1960; Caritis, 2009; Rouse 2007, Hartikainen-Sorri 1980) there is not a net excess of fetal deaths in the IM compared to placebo groups. And this extends to other formulations. The exact numbers are included in Tables 12, 14, and 15. (As an aside this is a group for which we indicate moderate evidence that there is no benefit for use.)

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #6	Abstract	The most important issue is the language in the conclusion of the structured abstract. It does not reflect the narrative in the summary. Moreover, it can be misinterpreted by readers. In addition, the abstract does not contain any information about the safety signal of 17-alpha hydroxyprogesterone caproate. A balanced, fair and objective review must contain this information.	We hope the conclusion now maps more easily to the tone and content of the text and the results of the meta-analysis. With regard to safety we have updated the report as discussed above. The abstract now includes this overview acknowledging limitations in size of study populations to date in understanding risk: “No data were available from large registries often developed for surveillance of rare outcomes like fetal death.”
Peer Reviewer #7	Abstract	Found the structured abstract very helpful: it "cut to the chase" and was clear about what could and could not be substantiated--made it easier to read the rest of the document.	Thank you
Peer Reviewer #2	Executive Summary	Page 15: Regarding the executive summary, the authors attempt to address the number of injections in question KQ4, highlighting a database study in which 5 or more injections were associated with prolonged gestation, but fewer did not. This outcome is confounded by preterm birth; those delivering preterm will have fewer injections. Please consider deleting this section.	Excellent point. Added the following consideration: "However, this analysis is not adjusted for confounding by gestational age at birth, leaving interpretation inconclusive."
Peer Reviewer #2	Executive Summary	Page 16: Question KQ5 could but does not include the co-intervention of omega 3 fatty acid supplement. In reference 75, all patients were treated with progestogen and were randomized to omega 3 supplementation or placebo.	Thank you. Harper 2010 has been added to this section.
Peer Reviewer #2	Executive Summary	Page 17: In the Summary Strength of Evidence and Findings, the authors allude to the fact that the largest trial found no evidence of effectiveness. This is true, but treatment was with vaginal gel rather than intramuscular or suppository therapy. A similar number of subjects were exposed to intramuscular progestogen in the Meis trial which did show benefit of treatment for women with a prior preterm birth.	We will add descriptors about formulation when we note the O'Brien study.
Peer Reviewer #3	Executive Summary	Page 13, Line 51This sentence was confusing. The KQ next to the number made it look like a sub question. This can be fixed by saying 54 articles to KQ2, 21 articles to KQ3 etc.	Thank you. We have edited this for clarity.
Peer Reviewer #3	Executive Summary	Page 14: you mention that quality assessment was done on the articles. Was this a standard quality tool or did you make one up? If a standard tool reference, if not, reference the appendix where it can be found. It is possible it is missing her by design because this is the executive summary, if so ignore this comment.	Thank you. A reference to the appendix has been added.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #3	Executive Summary	Page 14, Line 21: This is unclear. How can latency be “conflicting results” and PTB be consistently supported. If the treatment works shouldn’t both be congruent. This may be a wording problem vs. needing a better explanation of how that can be true.	Changed for clarity.
Peer Reviewer #3	Executive Summary	Page 14, Line 35: would start a new paragraph before “Evidence for all uses”	Changed for clarity.
Peer Reviewer 3	Executive Summary	Page 15: I am concerned about the inclusion of studies that used DES. This is a drug that is known to cause harm. To include it makes it seem like an accepted treatment.	Thank you. These studies met our inclusion criteria and must be reported on. However, we have added the following for clarification: "These studies are noted for completeness, but are not included in the meta-analysis or the SOE assessment."
Peer Reviewer #4	Executive Summary	Page 7, Line 41: I am concerned with the “with the largest trial finding no evidence of effectiveness”. This trial used a type of progesterone that has not to date been found to be effective in any condition – it is likely the formulation not that the treatment is ineffective. I think including this end to the sentence in line 41 is unnecessary and a disservice to the other trials.	We will add descriptors about formulation when we note the O'Brien study.
Peer Reviewer #4	Executive Summary	Page 7, Line 53: These do not list the conclusions – I suggest using some of the sentences from the structured abstract conclusion on page vii.	Thank you. We have updated this section to include concepts from the abstract.
Peer Reviewer #4	Executive Summary	Comment should 17ohp alone be assessed? I know that there have been meta analysis before but to lump all progestins together seems to mix apples and oranges. (addendum it looks like some of this is done on page 52)	Key Questions 1 & 4 directly address differences by indication and formulation. The review now includes a meta-analysis by indication.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #6	Executive Summary	Page 17: The Executive Summary (under Summary – Strengths of Evidence and Findings) indicates that "use of progestogens in singleton pregnancy with prior spontaneous preterm birth is informed by evidence of low strength, based on numbers of small trials using progestogens, with the largest trial finding no evidence of effectiveness". It would seem that this summary of evidence is not supportive of the conclusion of the structured abstract. This is an important issue because readers will probably rely on the abstract in PubMed (rather than in the article) to form an opinion about this important issue. Obstetricians are busy clinicians, and therefore, often rely on the conclusions of peer-reviewed literature (in particular, that of a peer-reviewed abstract). I would ask the authors that the conclusions of the abstract be made consistent with the Summary – Strengths of Evidence and Findings (page 7, or page 17/518 – there are two methods of page numbering).	Changed: Summary Strength of Evidence and Findings Progestogen treatment reduces risk of preterm birth in singleton pregnancies in women with prior preterm birth. Use of progestogens for this indication is informed by evidence of moderate strength, based on small numbers of trials, using different progestogens. Moderate strength of evidence indicates lack of effectiveness for multiple gestations. Evidence is insufficient for all other uses and for understanding factors associated with patient preference and adherence to different routes of birth of progestogens. Across indications, data are sparse to evaluate influence on near term and long-term maternal and infant health outcomes. Overall evidence is insufficient for understanding whether intervention has the ultimately desired outcome of preventing morbidity and promoting normal childhood development.
Peer Reviewer #8	Executive Summary	Page 17 Line 42: Page 7, 3rd sentence in the section "Summary Strength of Evidence and Findings:" there must be a typo where the text states "different routs of 'birth' of progestogens."	Corrected. Deleted "of birth".
Peer Reviewer #1	Introduction	presents the dilemna clinical and policy makers face re this therapy	Thank you.
Peer Reviewer #2	Introduction	The structured abstract is concise and clear, delineating the methods and results accurately.	Thank you.
Peer Reviewer #2	Introduction	The authors suggest that there are no maternal factors that have been shown to modify the effects of progestogen treatment. This is true with respect to preterm labor, twin and triplet pregnancies, and this should be stated for these specifically. There are also data regarding race, number of prior preterm births, and gestational age of prior preterm births. This could also be stated specifically. There are some data that suggest women with a short cervical length, with or without a prior preterm birth or other high risk factors, may benefit from progestogen therapy. The authors may wish to modify the initial comment regarding "no maternal factors".	A number of maternal factors are known to be predictors of preterm birth risk. Key question 3 specifically dealt with potential modification of the effect of treatment by individual characteristics like, severity of prior PTB and cervical length. These factors have not been appropriately modeled in clinical trials data to understand if they behave as effect modifiers.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #2	Introduction	An important aspect is that not all progestogen treatment is the same. The authors lump intramuscular, intravaginal, and oral therapy, suggesting that there is no consistency among studies. Importantly, there is relative consistency among studies of intramuscular progestogens. Similarly, five studies of vaginal suppositories demonstrated pregnancy prolongation but none of gel therapy did. There is also consistency among the trials of vaginal gel (progesterone vaginal gel did not work in either study). There appear to be consistent patterns according to type of progestogen and route of administration.	Key Questions 1 & 4 directly address differences by indication and formulation. The review now includes a meta-analysis by indication. The findings of that analysis in the discussion of "KQ3, KQ4, and KQ5: Modifiers of Outcomes" find benefit from all formulations assessed – oral, IM, vaginal. This may well obscure some further distinctions between gel vs cream or other nuances, but suggest that formulation is unlikely to be an explanation for lack of benefit is some trials. We also not that this is not conclusive since there are no head-to-head comparisons to determine if there are differences.
Peer Reviewer #3	Introduction	Page 22, Line 54: I would take out the part about how many PTB may be prevented with progestin's. That number is based on the effect size of 17P, which is under debate in this document. I think it is fine to say how many with prior preterm birth might be eligible, but again, this reports is supposed to be determining what populations have evidence for using progestins. I think these sentences may suggest a bias on the part of the report.	Changed for clarity. "In the United States, approximately 133,000 expectant mothers annually have a history of preterm birth and are potential candidates for progestogens. If the results of the largest US trial for that indication are used, an estimated 10,000 preterm births might be prevented annually by use of progestogens."
Peer Reviewer #4	Introduction	As above, very dense report	Abstract, executive summary, and subsequent articles will be of more use. We agree the full formal reports serve better as extensive documentation than reading material.
Peer Reviewer #5	Introduction	This section provides concise background on the scope of the clinical problem and variety of approaches utilized clinically that are seen in the various studies.	Thank you.
Peer Reviewer #6	Introduction	The introduction is an adequate justification for the report. However, the appropriateness of using 37 weeks as a primary endpoint was not considered an adequate surrogate for infant and neonatal outcome by an FDA Panel. This matter needs to be considered and discussed.	See tables 4 through 16 for examples. We abstracted both maternal and infant/fetal outcomes from an a priori list of outcomes developed with the report's Technical Expert Panel. We share the same concern as the reviewer that when it comes time to aggregate results <37 weeks becomes the default.
Peer Reviewer #7	Introduction	Appreciate the cost ($) data as well as clinical--key information for policy development. Cost data well done and credible. Cost calculations avoided the pitfall of making questionable associations and cause and effect relationships--well done	Thank you.
Peer Reviewer #8	Introduction	Nice introduction sets the stage for what will be addressed within the study, as well as the main thrust of the issues.	Thank you.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #8	Introduction	Page 22, Line 15: On page 12, under the section “Use of Progestogens,” the text states “The exact mechanism for pharmaceutical effects is not well understood.” While this is broadly true, there are certainly some theoretical underpinnings for mechanisms, including the idea that progesterone functions as an anti-inflammatory. The inflammatory pathway, as one of the pathways to preterm birth, is well-recognized. Perhaps there should be at least some mention of some of the current thoughts about the mechanisms of the pharmaceutical effects of progestogens.	Thanks – the text includes that “Some evidence suggests that progestogens act by inhibiting an inflammatory pathway that may be the impetus for preterm birth.”
Peer Reviewer #8	Introduction	Page 23. Line 55: Page 13 under “treatment options,” next to last sentence: I don’t believe it is correct to state that “only medroxyprogesterone acetate is currently available in the United States.” Prometrium (progesterone, USP) is also widely available in the U.S., correct?	Sentence corrected. "of these three,…"
Peer Reviewer #9	Introduction	Appropriate and includes the relevant information.	Thank you.
Peer Reviewer #10	Introduction	Two Introductions are included in the manuscript—one in the executive summary and one in the overall presentation. Both do a reasonably good job of outlining the problem, describing the use of progestogens and introducing the key questions.	Thank you.
Peer Reviewer #10	Introduction	Page 22, Line 47: the authors state, “to endorse use of progestogens for women with prior spontaneous preterm births... Other progestogens may also be effective”. They need to provide clarification.	Sentence corrected. "…to endorse the use of 17 OHP"
Peer Reviewer #10	Introduction	Page 24, Line 52the authors state, “crystalline progesterone and natural progesterone . . .” I think crystalline would be natural progesterone.	Language reflects the terminology used in the literature.
Peer Reviewer #1	Methods	have used standard criteria and statistical techniques	Thank you.
Peer Reviewer #2	Methods	Methods for this document are clearly specified, allowing clear understanding of the criteria for inclusion and exclusion of relevant studies. The statistical methods applied appear to be appropriate.	Thank you.
Peer Reviewer #2	Methods	Of note, the grading system includes the presence or absence of a flow diagram as a criterion for quality. Review of the grading for the two largest studies (Meis and O’Brien) reveals the Meis study was declared to have "fair" quality solely because it lacked a flow diagram. However, the flow diagram in the O’Brien study provides no additional information to that provided in the Meis paper's text. This criterion seems somewhat arbitrary and not quality driven.	The Meis study has been changed to good. The flow diagram is a required Consort diagram (and we find sometimes reveals differences from text and tables); we wish editors would uniformly allow this figure to be included. The grading is on quality of the data as presented not the study as performed.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #4	Methods	Uncertain why Northen followup study not included in report	It is included as part of a family with the Meis trial meaning that analyses that build from the same population are noted, counted and discussed if they add new information to the particular area of discussion. Otherwise they are cited as part of the family or inter-related papers with the "parent" paper given primary position.
Peer Reviewer #4	Methods	Major concern that the analyses were not performed based on type of progesterone used. Given that the metaanalyses done in the 1960s were similar until evaluating based on specific progesterone type, it is clear that the type of progesterone used is critical. Simply to say that the quality is low and imprecise when you are mixing all kinds of trials and types of progesterones together is meaningless.	Key Question 4 directly addresses differences by formulation and now includes a meta-analysis by indication.
Peer Reviewer #5	Methods	The criteria for study inclusion and exclusion are specifically described and are appropriate. The search strategy is fairly standard. The methods for data synthesis are outlined appropriately, as are the specific criteria for quality assessment. The statistical methods for meta-analysis is described sufficiently.	Thank you.
Peer Reviewer #**6**	Methods	The search is explicitly stated, but clearly missed important information which is in the hands of the Federal Governnment (FDA). I do not think it is appropriate to ignore this information, which is also in the public domain. The composition of the Technical Expert Panel is not disclosed - similarly, there is no information about whether the TEP had conflicts of interest or whether they were knowledgeable about the safety signal and did not bring it to the attention of the authors. The safety signal uncovered by the FDA was an excess of fetal death in patients exposed to 17P. There is now an RCT that has reported a significant increase in fetal death, strengthening the concern first identified by the FDA. This information was available in December 2009, presented in February 2010, and may have been known by members of the TEP. However, the authors may not have had a way to know this. This is particularly concerning because the identity of the TEP is not disclosed	Technical Expert Panel members will be listed in the final report. Their potential conflicts of interest were reviewed before they served in this capacity. We will review this study and determine its inclusion in the review and/or add it to the discussion of the review. Earlier in this document (and in greater detail in the final report) we address the Combs paper. In the interim, the FDA has weighed the content of the literature presumably concurring with the data we describe above showing no excess of fetal deaths across all the related trials. In early February 2011, the FDA approved an IM formulation of 17OHP for prevention of preterm birth.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #6	Methods	In the Methods section, the authors state that they had a "Technical Expert Panel" and called attention to Appendix E; yet, Appendix E does not have the composition of the "Technical Expert Panel", and also, Appendix E refers to the reference list of excluded studies (page 506/518). I believe that the composition of this Technical Expert Panel and whether they could have conflicts of interest would enhance the transparency and confidence that the public would have from a Government agency such as AHRQ. This is particularly important because the document states that "this report may be used in whole or in part, as the basis for development of clinical practice guidelines and other quality enhancement tools, or as a basis for reimbursement and coverage policies." Therefore, this document could be invoked by professional organizations and other members of the community to support an intervention which has, at best, weak evidence and potential risk identified by the FDA.	Technical Expert Panel members will be listed in the final report. Their potential conflicts of interest were reviewed before they served in this capacity.
Peer Reviewer #7	Methods	Researchers hit all the critical criteria for inclusion and exclusion. Appreciate clear rationale stated for inclusion and exclusion of each study. Search strategies are complete and eliminate the potential for bias--well done. Definitions and diagnostic criteria are appropriate, complete and reflect the probable consensus of clinical providers (if they were polled)on a large scale. Cannot comment in detail on statistical methods--not my area of expertise but to my limited knowledge they look appropriate and proper without overstating the results.	Thank you.
Peer Reviewer #8	Methods	The authors are methodical, thorough, and explicit in their methods. The rationale is appropriate and clear. Nice job!	Thank you.
Peer Reviewer #9	Methods	Methods appear correct and have resulted in appropriate articles being obtained. Statistical methods appear correct to me	Thank you.
Peer Reviewer #10	Methods	In the Methods sections the authors present inclusion and exclusion criteria, all of which seem logical. It was initially somewhat disappointing that English was required as an inclusion for the study. However, the authors later explain this exclusion quite appropriately and satisfactorily.	Thank you.
Peer Reviewer #10	Methods	the scores are very nicely presented. However, it is very difficult to find where these scores are actually listed in relationship to the specific studies. This should be addressed in this section.	Corrected. Inserted a note to see Appendix D for the quality score table.
Peer Reviewer #1	Results	While this report is lengthy and the reader burdened by the basic ignorance we have about the epidemiology and mechanisms of preterm birth,the results are clear and compartmentalized according to conventional wisdom.	Thank you.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #2	Results	Page 46: Throughout the Results (particularly on page 36 and 38) the authors describe differences between progestogen treatment and controls as either "higher" or "lower" when statistical significance is not reached, or even "nearly" reached. The authors are strongly discouraged from taking this approach in an objective evaluation. The findings are either statistically different or significance was not reached. If the authors wish to discuss more subjective “near significance,” this should be restricted to the discussion and the power of the findings delineated. Near significance should not be asserted if statistical significance is not reached and there is adequate power to evaluate the question of a clinically meaningful difference in outcomes.	Two facets are of separate importance in the summary of findings of individual studies: the point estimate of the association and the precision of that estimate. As many in the clinical epidemiology and statistical community, we treat those as distinct characteristics of the results. So statements like "The rate of tocolysis was higher with progestogen treatment in a trial comparing intramuscular 17OHP with placebo (17.3 percent versus 15.9 percent; RR = 1.09; 95 percent CI: 0.70, 1.69) and a trial comparing vaginal progesterone with placebo (11.3 percent versus 10.3 percent; OR = 1.12; 95 percent CI: 0.67, 1.86)" are fully accurate. These sections describe the point estimates - both in the direction of 17OHP participants being more likely to receive tocolysis - while also noting lack of significance. Significance levels chosen are arbitrary and vary by field. Reliance of meeting that level in order to note findings short-circuits the full description of the literature by relegating rare outcomes and underpowered studies to the "not-significant" scrap heap while the full picture may be information across the full body of literature. If we do not discuss the direction and magnitude of non-significant findings we are leaving part of the work of synthesis unattended to. We concur with the assessment that "near" and "approaching" are not helpful - as if the data were striving to become more - and have reworded related sentences.
Peer Reviewer #2	Results	Page 54, Table 18: the authors present the range of outcomes with the number of studies reporting in parentheses, the authors also report the N for placebo injection, vaginal, and oral therapy in the first header row. This makes the N unclear but suggests for example that there are 23 placebo controlled studies of which only three evaluated reaction or discomfort to the suppository etc. Please clarify.	The data in the table are presented in a way to describe the N in the header to represent the number of studies of a particular type. The numbers in parenthesis in each cell represents the number of studies that reported the outcome.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #2	Results	Page 58: Regarding newborn outcomes, the authors have elected to provide meta-estimates for neonatal deaths and birth weight, but not for newborn morbidities such as respiratory distress syndrome, necrotizing enterocolitis, intraventricular hemorrhage, sepsis, the need for ventilation and retinopathy of prematurity (each presented in tables 13-14.) The authors are encouraged to provide meta-estimates for these outcomes as reduction of infant morbidity and mortality is the major reason for attempted pregnancy prolongation and prematurity prevention. The criteria for performance or non performance of meta-estimates (formal metanalysis) should be described in the Methods section.	The criteria are listed in the Principal Findings and Considerations section. "All other maternal, fetal, or neonatal outcomes were reported by fewer studies or had incompatible definitions not appropriate for aggregate estimates."
Peer Reviewer #2	Results	Page 64: regarding injections of 17 hydroxyprogesterone the authors combine studies of singleton therapy with those for multiple gestations. Given that no studies of multiple gestations has demonstrated benefit and a that large fraction of singleton studies did find benefit, and also given the likely different mechanisms of preterm birth in these cohorts, the authors are encouraged to compare the evidence for singleton and multifetal gestations separately.	This section of the report presents the results of KQ4, in which we discuss the effectiveness of the progestogen use by route and formulation, regardless of indication. In the results for KQ1 the literature is summarized by indication (prior preterm, multiples, current preterm labor, etc).
Peer Reviewer #2	Results	Page 65: regarding vaginally administered progestogens, the authors combine vaginal gel and suppositories but comment in the last sentence that neither trial of gel found benefit. Again, it may be appropriate to present data separately for gel and for suppositories unless the authors can make the case that these are equivalent treatments.	Table 19 includes the precise form (gel, capsule, or suppository) and related odds ratios. We have attempted to discuss the types clearly in the text. The specific sentence about “neither trial of gel” does refer to two trials of that form. Across other forms, the studies lack similarity in indication and use. We take some reassurance that providing summary evidence is sufficient since the meta-estimates for prevention of preterm birth show benefit for the class of vaginal agents as a group.
Peer Reviewer #2	Results	The characteristics and progestogen dosing regimens within the included studies are described in detail, as are their geographic origination and funding.	Thank you.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #2	Results	There is some inconsistency in descriptions. For example, for the meis trial, the placebo is described repeatedly as castor oil, but the placebos for other studies are not described. The O'Brien study is considered to have a low dropout rate. However, only those who received their first dose were included. The actual dropout rate is unknown	We note castor oil as a placebo in the Meis trial due to its use as a method to induce labor. Descriptions of other placebo agents are included in the evidence tables in Appendix C. We are unable to report past the information provided in the publication; however the O'Brien (2007) paper does include information in Figure 1 that reports pre-randomization loss and post-randomization loss by arm.
Peer Reviewer #2	Results	While the results of studies are described as being "conflicting" the formulations and mode of administration are different between these. The authors are asked to reconsider whether studies of different drug formulations and rout of administration should be considered as equivalent and combined for comparison purposes.	The report includes a full and varied analysis of the use of progestogen, including separate analysis by route, formulation, and indication. Meta-analysis results also suggest that all major formulation types are effective lessening our worry that formulation conceals or creates appearance of benefit.
Peer Reviewer #3	Results	Page 45: many of these outcomes are not hard health outcomes but physician behavior mediated events. That might be worth commenting on (tocolysis, antenatal hospitalization, determination of PTL).	Good point, we have now noted the role of providers in "creating" these outcomes: "The most clinically significant and frequently reported outcomes for complications during pregnancy and mode of birth are presented in Tables 4-7, of note each of these is mediated by the care provider as part of the process of care; none are patient reported or longer term."
Peer Reviewer #3	Results	Page 49, Line 46: most of the other paragraphs start with the progestin group compared to placebo. This paragraph compares placebo to progestin thus reversing the language. Try to keep comparisons consistent through the paragraphs so people can skim for "high" or "lower" and understand the relationship.	Corrected.
Peer Reviewer #4	Results	Page 43: I would argue that although the DaFonseca 2003 trial was mixed, over 90% of the patients had a prior PTB and they should be analyzed with those trials. Why isn't daFonseca included in lines 40-48 of page 33	The description of the studies noted here includes only RCTs with greater than 200 participants. Da Fonseca only had 142 participants. It does not meaningfully change estimates and after meta-analysis we have changed the strength of evidence to moderate for use in prior preterm birth while maintaining mutually exclusive categorization of the studies.
Peer Reviewer #4	Results	Page 67, Line 17: why isn't the Northen study cited (the followup of the Meis trial) published in Obstetrics & Gynecology?	It is included as part of a family with the Meis study. (See longer related note above.)

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #4	Results	Page 76, Line 34: “Two studies with interventions initiated before 20 weeks….have conflicting findings” – again I am concerned with this statement in that different formulations were used – consistently, the gel has not shown benefit	The report includes a full and varied analysis of the use of progestogen, including separate analysis by route, formulation, and indication. We provide the view through each of these “lenses” separately.
Peer Reviewer #4	Results	Page 82: when combining all types of progesterone the evidence is inconsistent and imprecise. Strongly urge evaluation by type of progesterone.	We benefitted from reviewers strong interest in greater consideration of type of progesterone in this and several prior comments. We conducted a meta-analysis by formulation that shows within the limitations of the literature that the type of progestogens is not likely to be driving the results of the studies – all are effective in aggregate; indication seems a stronger driver of outcomes.
Peer Reviewer #4	Results	Page 85: text says 3 RCTs in twins – meta estimate graph only displays two	Thank you, this has been updated.
Peer Reviewer #4	Results	How were the studies rated? Page 25 describes the rating system, but I would like to see the data of why the Meis trial is rated “fair” (p33 line 14) whereas the O’Brien is rated “Good” (p33) and Rouse is “good” by the same group as the Meis trial. What puts the Meis trial in a lower tier?	Ratings on individual criteria are located in Appendix D. As noted above Meis has been changed to good on the basis of consideration that Figure 1 data is in text and was excluded by editorial constraints.
Peer Reviewer #4	Results	By not analyzing by type of progesterone the results are not useful.	See notes above related to new analyses by formulation. Indication appears to have more weight than formulation as all formulation. We had key questions organized by indication and formulation and we now provided the related meta-analysis results.
Peer Reviewer #5	Results	The amount of detail in this section is extensive and appropriate, with the characteristics of the studies clearly defined and described. The interpretations are clear and are appropriate for the data. Figures, tables, and appendices are extensive and adequate, but easy to read and navigate to address the reader's questions. No studies that are relevant appear to be overlooked, and included studies meet the entry criteria.	Thank you

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #6	Results	Page 58: 1. The data regarding spontaneous abortion for the Meis et al. trial (which is the basis for the conclusion in the abstract) is missing. Specifically, Table 13 reports fetal and neonatal outcomes for women with a history of preterm birth. This table does not include the data for the 5 miscarriages that occurred in the trial. I believe that the source of this is an incorrect reading of Tables 2 and 3 on the paper published in the *New England Journal of Medicine*. However, this was the basis for the safety signal discovered by the FDA.	We only report IUFDs and not miscarriages. Miscarriage was very uncommonly reported in this literature (likely related to gestational age at initiation) and was not sufficiently commonly reported to be presented in summary tables.
Peer Reviewer #6	Results	Page 66: There is some evidence that 17-alpha hydroxyprogesterone caproate may increase the risk of gestational diabetes. However, this is not included in the section of Harms of Progestogen Treatment – yet, this information was reviewed on page 56 in narrative form, and in Table 18.	Table 17 is in the harms section KQ2 (which was Table 18 at the time of review draft). It includes the data the reviewer refers to about the incidence of gestational diabetes by study type (RCT, other observational). The literature on the topic is sparse with wide ranges that overlap the rates in placebo groups. Text the reviewer notes is also within the harms results.
Peer Reviewer #6	Results	The section about potential for harm and risk does not adequately cover the risk of fetal death or gestational diabetes.	We agree these are important outcomes; however the literature is limited. Table 17 reflects data reported and each table by indication in KQ1 for fetal/neonatal outcomes also includes this information when it was reported by the authors. We have noted in multiple locations that lack of sufficiently large studies to understand this risk is concerning and problematic for clinical care.
Peer Reviewer #7	Results	Opening section of KQ1 has the best list of clinical concerns that I have seen. I hope future research takes these clinical questions as their measurement points. Answering these questions in future research will move the care of women at risk for preterm labor giant steps forward. Tables are extremely helpful. Tables include critical information without unnecessary information. List of included and excluded studies seems complete. Appreciate the clear statement of reasons for inclusion or exclusion of each study.	Thank you.
Peer Reviewer #8	Results	Throughout the results section, when discussing studies in the text that are NOT placed within any of the tables, it would be helpful to provide the sample size for those studies	This has been corrected.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #8	Results	• It is not made clear if the included studies address spontaneous preterm birth ONLY or have included medically induced preterm births. None of the tables make this indication either. It would be important to clarify this, perhaps several times within the report.	Some studies differentiate these, but many do not. We have reported the spontaneous preterm births when available.
Peer Reviewer #8	Results	When meta-estimates have been conducted, it would be helpful to place these results at the bottom of the corresponding tables for a quick ability to compare these with the individual results.	Corrected. Meta-estimates added to results section. (Check with KH)
Peer Reviewer #8	Results	Table 10 on page 43: To be consistent, the * that indicates statistical significance needs to be located in the same spot for each study. For example, the Facchinette et al RCT places the * on the control group, whereas the others place the * on the intervention group statistics.	This has been corrected.
Peer Reviewer #8	Results	Table 10, page 43: The text indicates that the mean GA difference in the Borna, et al study is statistically significant. However, the table does not place the * on the statistic.	This has been corrected.
Peer Reviewer #8	Results	Page 46, text and table 12: Regarding the Suvonnakote study, the text states “The rate of preterm birth at less than 37 weeks was also significantly lower with progestogen treatment in a RCT comparing intramuscular 17OHP with no treatment (85.71 percent versus 51.28 percent, p = 0.0036). However, the table reports 14.3* and 48.7. These are very different in number, as well as in direction!	This has been corrected.
Peer Reviewer #8	Results	Table 12: Regarding the da Fonseca, et al study, the table reports 28.5% PTB < 37 weeks for the placebo group, but the texts reports 38.5%.	This has been corrected.
Peer Reviewer #8	Results	Table 14: The Borna, et al study results for RDS don’t match what has been reported in the text on page 51, which states “Respiratory distress syndrome was the only condition for which there was a significantly lower rate (p = 0.021) among infants whose mothers received vaginal progesterone suppositories.” The table reports a HIGHER rate of RDS in the progesterone group, not a lower rate.	This has been corrected.
Peer Reviewer #8	Results	Page 51: Would be helpful to report the rates for LBW in the Borna, et al study along with the p value in the text.	This has been added to the text.
Peer Reviewer #8	Results	Page 53: Citation needed (Dudas, et al study?) for the statement “One case-control study found lower mean birth weight in infants whose mothers were treated with intramuscular 17OHP, which was significant when unadjusted (p=0.002) but lost significance when adjusted (p = 0.09).	This has been corrected.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #8	Results	Page 56: Text states (second paragraph): “none of those receiving oral progesterone or placebo reported development of PIH,” but the table 18 reports 0-29% PIH in placebo.	This has been corrected.
Peer Reviewer #8	Results	Page 56: “postpartum bleeding disorders” and “bleeding problems” need to be defined, i.e. do the authors mean postpartum hemorrhage?	This has been updated.
Peer Reviewer #8	Results	• Page 57: last paragraph states “Others, like cesarean, are entangled with preterm birth and would require additional modeling within study data to evaluate for any independent effect of the drug on risk.” The authors may also want to add “multiple gestation.”	We analyzed singletons and multiple gestations separately and have decided not to add "multiple gestation" to this sentence for clarity.
Peer Reviewer #8	Results	• Page 62, section on BMI, last sentence of first paragraph of that section states “The entanglement of BMI with both risk of preterm birth and potentially with the biological activity of risk of treatment make if (sic) an important target for understanding modification of progestogen treatment outcomes.” It would be appropriate to also mention the risk of medically induced preterm birth secondary to BMI-associated PIH and GDM.	Corrected. "The entanglement of BMI and its related morbidities…"
Peer Reviewer #8	Results	• Page 64: reference made to table 4, but where is table 4? Same for reference to table 4 at bottom of page 65. Tables 19 and 20 are present, but not 4. This is confusing.	Table 4 is referenced multiple times throughout the report to draw attention to the information it contains. Added "…in Results chapter" to text.
Peer Reviewer #8	Results	• Table 21: Is this a correct p value (p = 0.98) for the Hobel, et al study regarding “Favors Progestogens?” If so, it should be in the NS column, right?	This has been corrected.
Peer Reviewer #9	Results	The results are certainly detailed (the 500 pages is a testament to that). Correspondingly, figures, tables, and appendices are appripriate and I can't identify any relevant articles that have been missed.	Thank you.
Peer Reviewer #10	Results	The Results section is reasonably well organized. The authors attempt to logically follow through the key questions. The characteristics of the studies are well described and I believe that the messages come through quite clearly. The figures are useful. I do not believe the authors overlooked studies although there have been studies recently published that identify progesterones for the use of short cervix that the authors mention but do not include. It would be useful to include these in the overall analysis. I also repeat that I believe inclusion of discussions regarding the use of progesterone, as a tocolytic for preterm labor is not appropriate in this presentation.	Thank you. An update to capture recently published literature will be done and studies that meet inclusion criteria will be added. Preterm labor studies met the criteria for the review.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #10	Results	Page 37 of 518, Table 3, 34 of the 58 randomized control trials have used 17-hydroxyprogesterone. In keeping with my prior objection, I suggest that progestogens should not be presented in combination as effective or not effective but rather the specific preparation for the specific indication should be included as efficacious.	The report includes a full and varied analysis of the use of progestogen, including separate analysis by route, formulation, and indication.
Peer Reviewer #10	Results	Page 38 of 518, a second Table 3 is presented. In this table it would be useful to include the number of patients in the study and the quality of the study.	For spacing and organization purposes, we have presented quality in Appendix D. We have added the total N to Table 3.
Peer Reviewer #10	Results	Page 44 of 518, Table 5, it would be important to include the drug that was used in these studies as part of this table.	The table cannot accommodate any more information. All drugs are listed in Table 3.
Peer Reviewer #10	Results	Page 50 of 518, Table 9--what is the meaning of the dagger symbol included for example, in the Rittenberg et al study under PTB ≤ to 32 weeks?	Thank you. This has been corrected.
Peer Reviewer #10	Results	Page 52 of 518, although I don't think the data presented is necessary, the text states that the two studies show significant differences in gestational age with treatment but only one of these is marked as statistically significant in Table 10.	Thank you. This has been corrected.
Peer Reviewer #10	Results	Page 53 of 518, the statement is made that "the meta-estimate combining these three trials is a relative risk of 0.44 (95 percent CI: 0.26, 0.76)". Again as stated above, I don't believe this should be included. However, this rather substantial and impressive observation, if it is included, must address why it shouldn't influence therapy.	We agree. The methodologic problems with these studies are addressed in the Discussion chapter.
Peer Reviewer #10	Results	Page 56 of 518, the authors present a comparison of the results of treatments in groups with mixed risk factors. This data does not appear to be useful. It is clear from the remainder of the presentation that not all forms of preterm birth are prevented by progesterone therapy. Certainly it is not effective in multiple gestations and the only evidence of success is in women with previous preterm birth. The conglomeration of several different indications actually results in meaningless data that might best be excluded from the presentation. There is virtually no way to interpret the meaning of this data as the identity and characteristics of the mixed risk factors are unknown.	We agree that the data of studies with mixed indications are difficult to interpret and note the limitations in the Results and Discussion chapters. However, these studies met our inclusion criteria and therefore their results are included in this report.
Peer Reviewer #10	Results	Page 47 of 518, the authors describe a study of an asymptomatic short cervix determined on ultrasound in which the progesterone treated group seemed to have a clear advantage. Perhaps in combination with the more recent short cervix studies, this could be a useful point to provide to clinicians considering the use of progesterone.	An update to capture recently published literature will be done and studies that meet inclusion criteria will be added.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #10	Results	Page 64 of 518, lines 18-20, the statement is made that the majority of 17 hydroxyprogesterone caproate studies "initiated treatment between 16 and 21 weeks gestation with a range of 15 to 36 weeks". Is it reasonable to include the study that enrolled patients at 36 weeks in this analysis?	This study met our inclusion criteria and therefore its results are included in this report.
Eisenberg Center	Results	Page 35: There is no Table 4.	This has been corrected.
Eisenberg Center	Results	Page 56: The Da Fonseca 2003 article supplies the mean GA data (p 128/308); could it be included here?	This has been added.
Eisenberg Center	Results	Page 82, Table 22: Rating reconsideration, Prior PTB; PTB rate at <37 weeks is scored as Low evidence, but there is a meta-estimate with reasonable confidence intervals and consistency across studies. Of the four studies, 3 are rated "good" quality and one fair. We are curious to know why this is rated low, and are wondering if you are considering changing this rating to moderate. And, if 2 other trials (one fair, one good) are included in the analysis (as discussed on page 42), does the evidence level improve?	Table 21 may provide a useful summary. There are two processes at work: 1) grading the quality of an individual paper; and 2) assessing strength of the evidence for effectiveness of the agent to achieve specific outcomes. The latter rests on assessing risk of bias, consistency of findings, directness (which is "direct" when there are RCTS), and precision of the estimates. The risk of bias is informed by quality of the studies. Table 21 summarizes our considerations by indication for effectiveness in preventing preterm birth, improving birth weight, and preventing fetal/neonatal death. Indeed after conduct of a new meta-analysis added in the final version we changed precision of the information about use of progesterone to "fair" from "imprecise" and the overall rating for use of progestogens to prevent preterm birth to moderate. We are nonetheless distressed that the literature pivots on only 4 RCTS, one with inconsistent findings, and that we have not had the opportunity to answer crucial questions about effectiveness and safety. We sincerely hope the rating of moderate does not imply that the final answer is in.

Commentator & Affiliation	Section	Comment	Response
Eisenberg Center	Results	Page 82, Table 22: Clarification: Prior PTB; fetal /neonatal death is scored as insufficient evidence, lack of precision to estimate, but Figure 4 gives the meta-estimate of 4 trials. We'd like to better understand your rating on this. Also on closer examination, it appears that Mahji 2009 perhaps should not have been included in this meta-analysis because, you report in Table 13 that Mahji and colleagues did not report fetal or neonatal deaths in their study.	Overall answer is as above. We do not provide a meta-analysis for fetal death but do consider the precision across studies. Overall the size of studies has been insufficient to have adequate power to detect differences and no one study is compelling. Thus precision within the studies and across all available studies is lacking or imprecise in the table ratings that contribute to strength of evidence summaries.
Eisenberg Center	Results	Page 82, Table 22: Rating reconsideration, Threatened Preterm Labor: The aggregated result of 3 trials indicates a benefit (RR =0.44, CI 0.26, 0.76). We wonder if this should be rated "low" rather than "insufficient." According to our medical content expert, shortened cervix is considered by some to be a symptom of preterm labor, she inquired whether the Fonseca 2007 trial can be discussed as part of the pre-term labor group.	In this domain all outcomes have high risk of bias, inconsistent results, and imprecise (a five-fold width of the confidence bounds is large) estimates despite direct comparisons. That translates in this evaluation of strength to insufficient. Multiple factors not just precision are assessed.
Eisenberg Center	Results	Page 82, Table 22: Clarification: Mixed risk factors: PTB prevention. Evidence quality is rated as “low” in Table 22, but no finding is given. Is there a finding that we can associate with this rating?	We have changed this to insufficient to be consistent with the other ratings.
Eisenberg Center	Results	Page 86, Line 19: Summary of KQ2 results, Harms of Progestogen Treatments. The summary states that “Prospective followup of mothers and children over years has not been reported.” However, it appears that the Northen study of participants from the Mies MFMU network study (reference 20 in the CER) may have been overlooked	We have added discussion of the Northen study to the harms section and clarified that statement in the summary section.
Peer Reviewer #1	Discussion	The future research section is one of the strongest parts. Points out clear paths for researchers to follow.	Thank you
Peer Reviewer #2	Discussion	Page 73: In the last sentence of paragraph 2, page 73 the authors suggest poor precision for understanding of rare outcomes such as IVH and respiratory distress syndrome. However, numerous studies have evaluated these outcomes and they are more frequent than fetal or neonatal death. Please be consistent in presentation. Figure 4 demonstrates the summary estimate for effectiveness of 0.55 (0.299, 1.02). I would suggest that the estimate here is not poor. If those for respiratory distress, IVH, necrotizing enterocolitis and sepsis are significant different, it would be helpful to have this information presented.	Thank you for your comments. The lack of precision noted here relates to the varying definitions of these rare events and their use as intermediate outcomes to determining overall well-being of a child. While RDS was reported in several studies, the authors felt this was an intermediate outcome while fetal/neonatal death was not and a meta-analysis for this outcome was undertaken. The meta-analysis has been redone and new figures are in the report.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #2	Discussion	Page 82, Table 22: In Table 22 the authors suggest the data are sufficient and there is a lack of precision to estimate fetal death, birth weight, fetal neonatal death, etc. Please provide the point estimate and 95% confidence intervals for each. In addition, please provide similar information for the other relevant neonatal morbidities associated with preterm birth as described above (RDS, IVH, necrotizing enterocolitis, and sepsis).	Point estimates and 95% confidence intervals are not consistently reported. Where available, they are in the text.
Peer Reviewer #2	Discussion	Given that the authors have specifically raised the issue of castor oil as the placebo in the intramuscular injection studies, they are encouraged to discuss the relevance of this issue to their interpretation of these studies. Are there data that intramuscular castor oil at this dose has an effect on pregnancy prolongation (or for that matter that larger doses given orally are of concern). This information would help provide context to the apparent efficacy of this intervention given intramuscularly.	Use of castor oil as a placebo is a theoretical concern due to its use as an induction agent. Evidence for the effects of IM castor oil have not been examined in the literature.
Peer Reviewer #4	Discussion	Page 89: The conclusions of the report (p79) do not succinctly describe the findings – they are more a statement of "issues" I suggest using some of the sentences from the structured abstract conclusion on page vii.	We have revised the conclusions section of the report.
Peer Reviewer #4	Discussion	The discussion and conclusion section, especially the latter, do not provide a summary of the findings. I suggest using the end of the structured abstract in the conclusion.	Corrected.
Peer Reviewer #5	Discussion	The summary of the findings is divided into strength of evidence, medication effectiveness, harms, and modifiers of outcomes, and each is presented clearly, with the limitations described in detail. The implications are further summarized in the applicability section, and future research needs listed in terms of both methodologic and content priorities.	Thank you

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #6	Discussion	I do not believe that this implications of the major findings are clearly stated, because there is ambiguity about whether 17P should or should not be used, and that patients need to be informed of the potential risk of fetal death, first identified by the FDA and subsequently found in an RCT. The authors did not count correctly the number of spontaneous abortions (late) in the major trial in which their conclusions are based.	We have updated the report to include new information concerning the safety of 17-OHP. The new study by Combs added 81 triplet pregnancies and reported a statistically significant higher rate of intrauterine fetal death with progestogen treatment (7.7 percent versus zero percent; p = 0.01). In this study, the intrauterine fetal death rate was only reported in combination with the miscarriage rate. Thus the 7.7 percent result includes some miscarriages after 16 weeks. When taken in total with other trials of IM 17OHP (Meis, 2003; Fuchs, 1960; Caritis, 2009; Rouse 2007, Hartikainen-Sorri 1980) there is not a net excess of fetal deaths in the IM compared to placebo groups. And this extends to other formulations. The exact numbers are included in Tables 12, 14, and 15. (As an aside this is a group for which we indicate moderate evidence that there is no benefit for use.) Since this review the FDA has approved Makena a 17OHP injection for prevention of preterm birth as anticipated. We share concerns about lack of large scale safety data and have amplified that in the final version of the report.
Peer Reviewer #7	Discussion	The statement of findings is well done and fairly states the major findings without exaggeration, assumptions, or inference---"just the facts". Very refreshing when reading research summaries. The statement of findings is consistent with the entire paper--I did not find any inconsistency in facts or conclusions. The future research section is easily translated into future research as are each of the Key Question sections. My hope is that researchers will read this report carefully and avoid the problems cited in excluded research and adopt the definitions and clinical questions listed in each section.	Thank you
Peer Reviewer #8	Discussion	The discussion is clear and the conclusions are justified based on the results.	Thank you

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #8	Discussion	The authors adequately address the issue of treatment adherence, as well as participant retention, but there is little discussion about the potential for selection bias related to those women who decline to participate in progestogen studies from the outset. Is there something different about the groups of women who are not included in these studies (i.e. by self selection or by exclusion due to location – urban vs. rural, etc.)?	Thank you for raising this concern. We will add this issue as an area for potential future research.
Peer Reviewer #8	Discussion	Do the authors identify any limitations of their report methods and/or results?	Evidence reviews, this one included, are subject to a range of limitations imposed by the method (can only report on what the search strategy retrieves, what the paper says, etc) and by the innumerable alternatives in approach. We don't have substantive concerns about gaps or weaknesses that limit the utility of the report. No doubt other syntheses and perhaps conclusions are possible, but extensive review and time to reflect have not revealed any crucial flaws.
Peer Reviewer #8	Discussion	Page 71: the table is listed as "table 3." Is that correct? There is also a table 3 on page 29 titled "Summary of progestogen interventions."	Corrected. Table 3 deleted from Discussion, added sentence referring back to Table 3 in the Results chapter.
Peer Reviewer #9	Discussion	Very well done and clear what is missing in data so far obtained.	Thank you
Peer Reviewer #10	Discussion	I think this is one of the better sections of the paper. The authors adequately summarize the data and arrive at clear recommendations. The section on future research is well done and should be useful to guide future research studies.	Thank you.
Peer Reviewer #10	Discussion	Page 85 of 518, the authors state that there is no data to inform whether the effectiveness of progesterone treatment varies among women with prior PPROM cerclage, uterine malformations, or conceptions via assisted reproductive technology, compared to other women. I think the data indicating differential efficacy with different risks for preterm birth is sufficient to allow the authors to make the statement that in future analyses, different risks would need to be considered separately.	We agree. This is addressed in the future research section.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #10	Discussion	Under the State of the Science, under Methodological Priorities the authors list, “Clear specifying of operational definitions for inclusion and exclusion criteria for instance, definition of preterm labor.” I think the data on preterm labor would indicate that this diagnosis is so imprecise that it would not be a good indication for intervention if taken as history in the previous pregnancy. The vast majority of patients with preterm labor go on to deliver at term.	Thank you.
Peer Reviewer #1	Usability/Clarity	I think this reprort will be useful to policy makers.	Thank you
Peer Reviewer #2	Usability/Clarity	The presented information is important, generally objective and will be useful to the target audience. The presented tables are clear. A summary table including all meta-estimates might be helpful.	Thank you
Peer Reviewer #5	Usability/Clarity	The report is appropriately structured and organized, with the main findings clearly presented. Unfortunately, the findings are that the strength of evidence is low or insuffient for a majority of the questions addressed, pointing only for the need for further study rather than any clear message for practice change.	Thank you
Peer Reviewer #6	Usability/Clarity	I do not believe that the conclusions can be used to inform policy and/or practice decisions, because they do not take into account the examination of the key trial by the FDA and the concerns of safety and efficacy that have prevented the approval of this drug. Similarly, the document does not state with sufficient clarity that 17P and progesterone are two different compounds with different efficacy and safety profiles.	This publication did not remain in our search because it was an abstract during the search for this review. The update of the treatment trials will look for any new publications and will be compared to our criteria for inclusion in the review.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #7	Usability/Clarity	Reprt very well structured and organized. Main points are repeated when needed. Example: clear statements on what can and cannot be confirmed with evidence. The clear distinction of populations that have evidence of benefit and those that do not (as of this date) inform policy. Financial data on prematurity and cost of treatment with 17OH are excellent guides for where to put public money to get the most benefit for every dollar. Conclusions are clear for clinical practice--as important as the evidence for what works is the lack of evidence or evidence that it does not work in particular populations/risk factors. Hopefully this will deter the medical propensity to apply interventions without solid evidence: "if it works for patient A it must work for patient B" clearly does not apply for 17OH. Information on cointerventions was well written and added clarity that I have not seen in other publications. This is truly a mixed bag and the authors did an excellent job of isolating factors and effects as the data allowed. Very helpful to clinicians and in planning future studies.This information is easily understood, the format walks the "non-researcher" through the research methodology in very readable and understandable steps. This is a great asset to clinicians who have trouble getting "the meat" out of research papers. The organization and the clear statement of analysis makes this usable by all clinicians. Clinicians who "don't let data interfere with how they practice" will have a hard time continuing old practices after reading this--it makes the data understood and the progression to conclusions absolutely clear. The addition of reasons to exclude studies will help clinicians understand why the information in some studies (included) has more relevance and reliability in addressing the questions than the excluded studies. This is a paper I would use to discuss the appropriate use of 17OH in my role on a state Medicaid committee. The policy folks can understand this even though they are not clinicians. The financial sections are of particular value with the policy group.	Thank you
Peer Reviewer #8	Usability/Clarity	A fair amount of text repeats what is found in the tables. Because the text is so dense it may be difficult and time consuming for clinicians to wade through the detail. There is a great deal of information being relayed, so it might be more “user friendly” (and efficient) to allow the readers to use the tables (which we tend to do anyway), and for the authors to avoid repeating the data in the text.	Thank you. There are strict formatting requirements that we must adhere to when constructing full reports.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #8	Usability/Clarity	Summarizing after each section is quite helpful. This might be enhanced by providing a summary subheading at the end of each section (i.e. each KQ report) so these can be easily located within the document.	Thank you. There are strict formatting requirements that we must adhere to when constructing full reports.
Peer Reviewer #9	Usability/Clarity	I think very well structures and organized and conclusions can inform research and policy.	Thank you
Peer Reviewer #10	Usability/Clarity	On the whole, the paper is reasonably easy to follow. It would benefit from the few additions to the tables, which I addressed above and also to refer to tables in some instances where no referrals currently exist. In addition it would be useful to determine the identity of the technical experts. This information is not included in the paper as presented. I certainly think that the conclusions and the recommendations could be used to inform clinical decisions. However, it should be made clear that the data as it stands indicates the success of a particular progestogen for a specific indication and until we are certain that the progestogens are identical, it will be necessary to consider them as independent treatments. Thus, if 17 hydroxyprogesterone is useful for previous preterm birth, it does not automatically follow that vaginal progesterone would be useful since one trial demonstrated that it was not. Similarly, I think it is vital that the authors emphasize that preterm birth that is related to different risks for preterm birth may or may not be prevented by progestin treatment. Some such as previous preterm birth and that related to multiple gestations are obvious but it would be important that future studies address different risks for preterm birth as individual entities rather than lumping them all as “preterm birth”.	Thank you, we agree. The TEP members will be included in the final version.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #3	General	Conclusions in abstract. I think to say that the quality of evidence that 17P reduces PTB is low is overstated.	The strength of evidence grading system is described in detail in the Methods chapter. We assume you mean too stringent by "overstated". At the time of the draft we had not completed the meta-analyses and were faced with four RCTs including only 1,318 women – three found benefit for reducing births <37 weeks and one with 512 participants (largest in the literature did not). That led us to an assessment of inconsistent and imprecise results. However, now that we have completed meta-analyses by indication and dose form, we find the precision of the confidence bounds to be fair and have upgraded the strength of evidence to "moderate" for PTB prevention among women with prior history of PTB.
Peer Reviewer #1	General	The report is on target	Thank you.
Peer Reviewer #2	General	The Evidence Report/Technology Assessment regarding Progestogens for Prevention of Preterm Birth is timely and important in that there are accumulating data regarding the effectiveness of such treatments and clinical practice is rapidly evolving, yet FDA approval for such treatment is lacking and the optimal dosing strategies, target populations, long-term risks and benefits are not well established.	Thank you.
Peer Reviewer #2	General	This document clearly delineates potential target populations for the treatments and also the audiences and constituencies that might be served by this document.	Thank you.
Peer Reviewer #2	General	The key questions are clearly defined and are appropriate.	Thank you.
Peer Reviewer #4	General	This is incredibly dense reading, I cannot put the effort into an exhaustive look at the 500+ pages, sorry. I strongly urge you to come up with a more useful format, if I (someone interested in the topic) cannot bear to read it, few other will.	Abstract, executive summary, and subsequent articles will be of more use. We agree that full formal reports serve better as extensive documentation rather than reading material.
Peer Reviewer #4	General	Concern with combining the different types of progesterone in the metaanalysis – trials in the 60s and earlier when combined showed no benefit as well – it took the Kierse metaanalysis of the high risk women with 17OHP to show benefit. Why not do separate analylsis based on progesterone type? I think it is unfortunate to say the strength of the evidence is low because the estimates were inconsistent, when part of that is due to different formulations. (page vii)	Key Question 4 directly addresses differences by formulation and now includes a meta-analysis by indication.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #5	General	This is a thorough review of a topic which is very important to the field of obstetrics. The manuscript is appropriate for the defined audience of practitioners and policy makers, and will be very helpful in synthesizing the vast amount of published data on the topic. The key questions that are addressed are exactly the issues that needed to be covered, and are stated explicitly.	Thank you.
Peer Reviewer #6	General	The report addresses an important question with clinical implications; however, there is some information which is missing and there is inconsistency between the summary of the evidence and the abstract. Given that this document could be used to provide guidelines for patient care and change reimbursement, this issue need to be addressed. For example, the abstract is citing about the safety signals of 17P, which is fetal death and has been confirmed in the context of an RCT	This publication did not remain in our search because it was an abstract during the search for this review. The update of the treatment trials will look for any new publications and will be compared to our criteria for inclusion in the review.
Peer Reviewer #6	General	There is a major issue in that an abstract published in December 2009 in the *American Journal of Obstetrics and Gynecology* reporting the results of a randomized clinical trial of 17-alpha hydroxyprogesterone caproate was not included. This trial was presented at the Annual Meeting of the Society of Maternal-Fetal Medicine in February 2010, and is now published. The results of the trial confirm the existence of a potential safety signal for fetal loss. The FDA detected a safety signal when conducting an independent assessment of a clinical trial of 17-alpha hydroxyprogesterone caproate for the prevention of preterm birth in women with a prior preterm birth. The trial, which has not been included, is important because it found a significant increase in the rate of stillbirth and perinatal death. The authors can find this article in the *American Journal of Obstetrics and Gynecology*, authored by Combs et al. I realize that the full information may not have been available to the authors of this report and technology assessment.	This publication did not remain in our search because it was an abstract during the search for this review. The update of the treatment trials will look for any new publications and will be compared to our criteria for inclusion in the review.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #6	General	I am concerned that the safety signals discovered by the FDA and now strengthened by the finding of a randomized clinical trial are not included because there was an increase in perinatal mortality in triplets ($p<0.01$) and in stillbirths ($p<0.05$). This has implications for practice and the conduction of research. Physicians need to inform patients that there is a potential safety signal with the administration of 17-alpha hydroxyprogesterone caproate. Similarly, there are ongoing randomized clinical trials using 17-alpha hydroxyprogesterone caproate, and the informed consent of these trials must take into account the new evidence of risk confirming the findings of the FDA. I assume that this document was in preparation in December 2009, and therefore, the authors could have access to this information. Indeed, in the Methods section, the authors state that they looked for abstracts of articles published in MedLine and M-Base in English from January 1966-March 2010. The article was published in September 2010 as a full article; nonetheless, this information is of major importance.	This publication did not remain in our search because it was an abstract during the search for this review. The update of the treatment trials will look for any new publications and will be compared to our criteria for inclusion in the review.
Peer Reviewer #6	General	This study has focused on the effects of progestogens – these include natural and synthetic progesterone. The FDA does not consider these compounds equivalent. Indeed, there are reasons to believe that they are not equivalent in clinical medicine. The study of Meis et al. was found to reduce the rate of preterm delivery, defined as <37 weeks using 17-alpha hydroxyprogesterone caproate. In contrast, O'Brien et al. found no evidence of effectiveness of vaginal progesterone in women with a history of previous preterm birth, the same criteria for entry in the Meis et al. trial. Moreover, the safety signal for 17-alpha hydroxyprogesterone caproate has not been found with progesterone. Therefore, there appear to be differences in efficacy and safety.	We do not consider these equivalent, but there is not a basis for choosing among them. It is important to review all progestogens being used for the prevention of preterm birth.
Peer Reviewer #6	General	An issue that requires consideration is that this study focuses on preterm birth, defined as <37 weeks of gestation. However, an FDA advisory panel specifically convened to review the trial that heavily informs this document concluded by overwhelming majority that 37 weeks of gestation was not an adequate surrogate for infant morbidity and mortality. Since delivery at <37 weeks was the primary endpoint of the trial, this trial failed to meet the standard of evidence of the FDA Advisory Board. This is the reason why the efficacy of 17-alpha hydroxyprogesterone caproate is questionable.	See tables 4 through 16 for examples. We abstracted both maternal and infant/fetal outcomes from an a priori list of outcomes. We share the same concern as the reviewer that when it comes time to aggregate results <37 weeks becomes the default.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #6	General	The authors should be aware that there was another study conducted by the Network of Maternal-Fetal Medicine Units using 17-alpha hydroxyprogesterone caproate in which the rate of preterm birth in the control group was substantially lower than that in the Meis et al. trial. Although the first trial was stopped and never published, this information is available on the FDA website and therefore, is in the public domain. I believe that the FDA considered that the rate of preterm birth in the Meis et al. trial was exceedingly high, and raised questions about efficacy. This is the reason why another randomized clinical trial is now in progress to test the efficacy and safety of 17-alpha hydroxyprogesterone caproate. It is understandable that the authors may not have access to that information because it is not in PubMed, but this does not diminish its importance. The transcripts of the meeting are on the FDA website. The name of the Medical Officer that conducted the evaluation is Dr. Barbara Wesley, and she currently works at the FDA.	We will review this study and determine its inclusion in the review and/or add it to the discussion of the review. We will be in contact with Dr. Wesley to get more information.
Peer Reviewer #6	General	I think it is important that the authors are aware that infant outcomes are considered of major importance by the FDA. Moreover, some trials of progestogens are being conducted in Europe, and the primary outcome is not preterm delivery but infant outcome. Similarly, my recollection is that the ongoing trial of 17-alpha hydroxyprogesterone caproate is using infant outcome as the primary outcome rather than preterm delivery. This information is available on clinicaltrials.gov. Its importance is that it reflects a shift in the emphasis from preterm birth as a primary outcome to neonatal and infant outcome. This is also relevant for evidence-based research.	See tables 4 through 16 for examples. We abstracted both maternal and infant/fetal outcomes from an a priori list of outcomes. We share the same concern as the reviewer that when it comes time to aggregate results <37 weeks becomes the default. Infant outcomes are the only reasons to study these interventions. Please see pages 46-53, that show how closely we align with the FDA.
Peer Reviewer #6	General	The argument could be made that the analysis should separate 17-alpha hydroxyprogesterone caproate and progesterone, because they are different compounds	Key Questions 1 &4 directly addresses differences by indication and formulation. The review now includes a meta-analysis by formulation.
Peer Reviewer #6	General	I believe that the considerations of safety and efficacy need to be addressed in a document that has the potential for being used to develop clinical practice guidelines or as a basis for reimbursement and coverage policies. The omission of the safety signal is a major concern. The agencies (DHHS and AHRQ) have called for transparency, and I believe that it is important to know the names of the Technical Expert Panel and whether they had a conflict of interest and informed the authors of this review of information that they knew.	Technical Expert Panel members will be listed in the final report. Their potential conflicts of interest were reviewed before they served in this capacity.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #7	General	Excellent document with clear writing style. This document will be understood by clinicians, researchers and policy makers. The explanation and rationale for every step in the research process supports the credibility of the conclusions. The key questions are appropriate and explicit--it is a pity there is a lack of data to answer them all. But, the list of questions serves as a guide for future research. The paper clearly states the gaps in data and the unanswered questions and the road for future research.	Thank you.
Peer Reviewer #8	General	The issue of progestogen use for preterm birth prevention is a very important one. This report is timely and very much needed given the amount of clinical confusion that exists about type of progestogen, dosage, frequency, and timing of such treatment. This is corroborated by the report's findings of great heterogeneity between research studies, as well as the differences in practice among providers of prenatal care. The review is thorough, accurate, and well-organized. This should provide better direction for clinicians, as well as much needed direction for further research. The report is successful in relaying the fact that the strength of the evidence is low thus far, primarily due to the inconsistency and imprecision of the study estimates. Furthermore, it is made clear that no data exists for long-term outcomes for fetal exposure to progestogens, apparently also a major concern for many providers. The key questions are appropriate and clearly stated.	Thank you.
Peer Reviewer #9	General	I think this is an excellent summary that has good clinical meaning embedded in it. The key questions are appropriate and explicit.	Thank you.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #10	General	The paper is a very important contribution. The authors review the use of several progesterone preparations during pregnancy to reduce preterm birth. I suggest that the review could be simplified if the portions relating to the use of progesterone to stop preterm labor were eliminated. Preterm birth prevention and treating the contractions of preterm labor are clearly two very different topics. Furthermore, the approach of using progesterone or any tocolytic to stop preterm labor has to a large extent been abandoned. In addition, the authors make the point that it is not possible to directly compare the different progestogen preparations because there have been no head-to-head trials. However, the only evidence based indication that they cite is for the use of progesterone to prevent preterm birth in women in whom there is a history of a prior preterm birth. The authors identify two large studies one which used intramuscular 17 hydroxy-progesterone caproate and the other vaginal progesterone. Only the injectable progestogen resulted in prolongation of pregnancy. I found nothing to suggest that the two preparations are equivalent and the lack of success of one and the success of the other would seem to dictate that the recommendation be specifically about the 17-hydroxyprogesterone caproate for this usage. I think the investigators have identified the key questions and these are addressed appropriately.	The point of this review was to examine the evidence of the effectiveness of progestogens to prevent preterm birth. We did not exclude papers based on indication for progestogen treatment and did intend to capture all formulations when authors stated they intended the use of the medication to be to prevent preterm birth.