Author’s response to reviews

Title: Progesterone in women with arrested premature labor. A report of a randomised clinical trial and updated meta-analysis.

Authors:
Stephen Wood (stephen.wood@albertahealthservices.ca)
Yakov Rabi (Jack.Rabi@ahs.ca)
Selphie Tang (selphie.tang@albertahealthservices.ca)
Rollin Brant (rollin@stat.ubc.ca)
Susan Ross (sjross@ualberta.ca)

Version: 2 Date: 30 Apr 2017

Author’s response to reviews:

Response to 2nd reviews:

Ewoud Schuit (Reviewer 2): I appreciate that many of my concerns have been addressed.

I do have some minor comments regarding the responses of the authors.

#7. my comment: According to Kernan et al. JCE 1999 it is advised to adjust for covariates that is stratified for in the randomization. This means regression analyses should be performed instead of the Mann-Whitney U-test, student's t-test, and Fisher's exact test.

Authors' response: To our knowledge this is not universally applied in clinical trials. Additionally, we feel that more complex analysis makes it far more difficult for the average clinician to understand and therefore trust the results. Ultimately, we performed the analysis that we decided on apriori and to change this after the fact would not be appropriate. We appreciate the advice for future trials.

My response: although it may be true that clinician may prefer simple t-tests over regression analysis, I do not think this should be an argument in refraining from regression analysis. I'm fine with leaving the t-tests in there, but it would be reassuring to know that such regression analyses would results in the same treatment effect estimates. Can the authors perform these analyses and then mention in the results section that results were similar when assessing treatment effects using regression analysis with the treatment allocation and stratification variables as covariates in the regression model?
As requested we have added this at the end of the results section.

#10. My comments: Risk of bias was assessed "using standard criteria", but items, e.g. reporting of outcomes, seem to be missing from the criteria. A better risk of bias tool may be the one developed by the Cochrane Collaboration.

Authors' response: We prefer a simplified system which has been described in the literature and we have used in a previous meta-analysis2. It is not manifestly different from the Cochrane "risk of bias" tool. It is worth noting that the risk of bias assessment is largely ignored in Cochrane reviews as to our knowledge among the obstetric reviews there are none with subgroup analysis by study quality.

My response: I'd really prefer the authors to use Cochrane's risk of bias tool as this is the current standard for assessing risk of bias of RCTs for meta-analyses. I'm not sure what the authors mean by their statement that "the risk of bias assessment is largely ignored in Cochrane reviews as to our knowledge among the obstetric reviews there are none with subgroup analysis by study quality". Using Cochrane's risk of bias tool is very common in systematic reviews and meta-analyses, including those published by the Cochrane Collaboration, regardless of specialty.

Response: Again, although I appreciate the suggestion we do not feel we should change the methodology, posteriori, we used to assess the trials. First, we decided apriori to use these criteria. Secondly, we have referenced a manuscript from the BMJ published by a prominent researcher in the field of critical appraisal Dr D Altman which recommends and outlines the criteria we used1,2. We agree that there are many indicators of trial quality but this in itself leads to difficulty in “scoring” quality. In any event, Dr Altman’s argument is that there are key indicators that have been shown in empirical research, to have an impact on the results of trials. These are: allocation concealment and randomization, blinding, and handling of patient attrition in the analysis (loss to follow up and Intention to treat analysis). Based on this work and our own experience in systematic reviews we decided, as we have in the past3, to grade trials as high or low quality based on these criteria. These criteria allowed us to classify trials as high or low quality and do a separate analysis by quality. I agree that the Cochrane Risk of Bias tool is commonly used but Cochrane does not have a monopoly on quality assessment. My issue with the “Cochrane Risk of Bias” is that it does not have clear criteria for what constitutes a low or high quality trial. I note that Dr Schuit has published meta-analyses. However, even in his own work I cannot find clear indicators of how he used his risk of bias assessments beyond commenting on them. I have provided with this re-re-submission the article of Altman and Juni. In return perhaps Dr Schuit could find me a Obstetrics review in the Cochrane collection that explicitly made use of their risk of bias assessment to perform a sensitivity analysis or subgroup analysis by quality. I would like to have one for teaching purposes. To date what I uniformly encounter are reviews that reliably perform risk of bias assessments but then they simply ignore them and proceed to pooled analysis including all the trials good and bad. To me, this renders the “risk of bias” assessment largely pointless. Finally, I am sure that if Dr Schuit reviewed the criteria we used he would find that they essentially capture the main elements of the cocharane risk of bias assessment which he “prefers” and therefore we disagree that some of them seem to be missing from our criteria. Comparing risk of bias (Cochrane Handbook version 5) to our criteria- Selection bias: adequate allocation concealment (note the Cochrane handbook confuses
selection bias with confounding when they state that selection bias is controlled by sequence generation or randomization, selection bias is controlled by allocation concealment, randomization controls for confounding, again I only mention this as Cochrane reviews have developed an aura comparable to papal encyclicals). Performance bias: blinding, Detection bias: blinding, Attrition bias: >20% losses to follow up. Reporting bias: Intention to treat analysis (ITT). In fact Cochrane does not mention ITT specifically a vital aspect of trial design and interpretation!!!! Therefore, is the one “missing” important elements. (I think the only explanation is that it may get in the way of the highly productive sausage machine that they have developed and encouraged).

#12. My comments: The authors indicate that they will use both fixed and random effects models, but from the methods it is unclear what criteria are used for the choice on the type of model, e.g. based on the level of statistical heterogeneity above a certain threshold.

Authors' response: Using fixed effects if no heterogeneity is standard and we do not think requires explanation.

My response: "no heterogeneity" is still rather vague. Although the authors state that the choice of fixed vs. random does not require explanation the authors present a fixed effects M-H analysis in figure 2, while the i2 = 83%. Based on this heterogeneity, the analysis should have been performed using a random effects model. Please re-check all analyses on whether the choice of a fixed effects model was appropriate. I still think the authors should specifically state when a fixed (e.g. i2 < 50%) and when a random (e.g. i2 => 50%) effects model was used.

We have changed figures 2, 3 and 4 as suggested to random effects models. The results do not change fundamentally and as one might expect lead to the smaller studies being more influential. The odds ratios are more favorable but the conclusions are unchanged. We have left the analyses by trial quality a fixed effect model as the detected heterogeneity suggests that the unpoled estimates would be the most reliable in any event. See explanation in text 116-118. Overall our approach to meta-analysis in general articulates what we think is caution in performing and interpreting pooled analysis. We are generally most comfortable with Fixed Effects models as we cannot really know the true distribution of trial results and dislike the increased influence of small studies. However, Dr Schuit is correct that we overlooked that it is the accepted standard to use a random effects model when heterogeneity is significant and we have made the changes suggested.

#20. My comment: Line 140-141. Was there statistical heterogeneity between the strata of the subgroups? This information should be added to the text. Authors' response: Information is provided in the figures 2-7. We feel this would be very redundant to add to text.

My response: I disagree. Such information would be very informative as it indicates whether the treatment effect differs significantly across subgroups. Presenting a (significant) treatment effect within a subgroup without presenting the heterogeneity of the effect across subgroups may lead readers to believe that this significant effect is different from the effect in other subgroup, while it in fact may not be different. Therefore I think this p-value of heterogeneity among subgroups should be added to the text.
We have added them to the text as requested and leave the interpretation to the reader. Line 155,159,161.

#23. My comment: Line 164. Remove "adequate".

Authors' response: Word is not found in line 164.

My response: please see sentence: "Several groups have found that poor adequate allocation concealment and blinding influence results, usually, by increasing positive treatment effects25-27." In this sentence, adequate is redundant.

Thank you line changes occurred with revision. Have corrected.

#24. My comment: Line 173. Low power is obviously an issue of the RCT, but indeed not of the meta-analysis.

Authors' response: This is a matter of opinion. Not sure what change the reviewer is suggesting.

My response: My apologies for not being more clear. What I hinted to was that the authors could distinguish between the RCT and meta-analysis in their assessment of power. Considerin that the RCT was stopped early, power of the RCT is obviously lower than planned. THis, however is compensated in the meta-analysis by combining data from multiple studies. The authors could explicitly mention this in their discussion.

We don’t agree that low power is not an issue with the meta-analysis. In total there are only 1400 subjects. The meta-analysis does not entirely compensate for this.

