BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (http://bmjopen.bmj.com/site/about/resources/checklist.pdf) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

**ARTICLE DETAILS**

| TITLE (PROVISIONAL) | Association of antenatal corticosteroids with morbidity and mortality among preterm multiple gestations: meta-analysis of observational studies |
|---------------------|-------------------------------------------------------------------------------------------------------------------------------------|
| AUTHORS             | Dongxin, Lin; Dazhi, Fan; Chen, Gengdong; Luo, Caihong; Guo, Xiaoling; Liu, Zhengping                                                                 |

**GENERAL COMMENTS**

The authors have performed a comprehensive meta-analysis of neonatal outcomes in multiple gestations following ACS administration. ACS administration and its impact on neonatal outcomes in multiples has previously been limited and the strength of impact in multiples (vs singletons) has not been thoroughly investigated, thus a meta-analysis of prior studies is valuable.

Specific recommendations and comments below:

**Abstract**
Line 19 (and lines 98)---I would consider listing mortality also a primary outcome (it's listed in the manuscript title) as well as RDS. It seems approximately ½ of the included studies had either mortality or RDS as a primary outcome and this would add consistency.

**Results**
Line 22—delete “regarding” before “by”
Results, line 25—include the number of subjects collectively included in the 16 studies
Line 39---Point 4: consider rephrasing. The heterogeneity did not materially alter what in some subgroups? Unclear what this is referring to, did not alter the outcomes/endpoints?

**Introduction**
Line 43-44---It is curious you have quoted a reference for PTB specifically in the Netherlands. I would consider using a more global estimate that refers to increasing rates of PTB.

**Methods**
Lines 92-93--Choice of subgroup analyses are important.

**Results**
Line 123---Please report the number of subjects collectively included in the 16 studies
Line 149-150—the OR for multicenter studies crosses 1 (0.67-1.0), meaning nonsignificant, so this is not significantly decreased. Please correct this.

• It would be helpful to include weighted-averages describing the study population (mean or median gestational age, mean or median gestational latency/admin-to-delivery time), if available from the studies.

• I would consider removing the OR reported the studies that had no information on ACS completion (for RDS, mortality, etc). It is hard to interpret what this means, as maybe those studies did have ACS completeness vs maybe they didn’t but, inferring that there is some sort of combination of these two possibilities, doesn’t really tell us anything additional of clinical value. Reporting the significant OR for ACS completion is important and valuable to report, but I would remove the converse/unknown of unsure ACS completion. I have similar thoughts to excluding reference to the OR for studies with unknown/unreported administration-to-delivery time interval. However, it is still important to include in your discussion that for the overall analysis (i.e. non subgroup analysis), ACS completion likely attributed to heterogeneity in outcomes.

• Supplementary figures—could include the publication bias/Egger’s test graph for the primary outcome(s) to visually demonstrate no publication bias

• Table 3—nice comprehensive summary of all the OR for outcomes, subgroup analyses. Well-presented.

• Sensitivity analyses is well-done and appreciated.

Discussion

Lines 197—please rephrase/clarify this statement. I think you are referring to prior evidence of the efficacy of ACS among singletons has been abundant, but evidence among multiple births has been sparse and limited to small observational studies. Adding these clarifiers will strengthen the impact of this statement.

Lines 202-204—it is important that the authors have offered a biologic basis for why ACS administration and effect on outcomes would differ in singletons vs multiples. I would recommend expanding on this biologic proposition as it strengthens the importance of why this manuscript’s findings are significant and important for publication.

Lines 221—222—Please combine these two sentences “First, the majority…” and “And the overall quality of the evidence…” I would change the period after “avoided” to a comma and de-capitalize the “And”

Line 223—heterogeneity did not materially alter what? The outcomes/study endpoints?

• In the limitations, the authors should emphasize large heterogeneity weakens the validity of the study conclusions, not just that heterogeneity is present due to retrospective nature, meta-analysis of heterogeneous studies, etc. Almost all I^2 values are >50% and often >80-90% ---this is very high, and certainly weakens the study findings. While an RCT would be useful, it would be unlikely to be feasible due to ethical reasons.

• Do the authors propose an explanation for why a multicenter vs single center finding may find more or less significant impact of ACS on the outcomes, besides heterogeneity in multicenter studies? It is unclear clinically why multicenter vs single center actually has an impact on ACS administration and outcomes, and thus unclear to me why this subgroup analysis is of value other
than to point out more significant heterogeneity in a multicenter study.

This article by Lin et al. shows the effects of antenatal steroid administration on the outcomes of preterm multiples by meta-analysis and may provide clinically useful information. However, some corrections are needed.

p.6, Eligible criteria
Since the outcomes of late preterm infants is clearly better than that of infants born at 34 weeks’ gestation or less or very preterm infants, it is better not to include the article by Ben-David et al. in this analysis.

p.10, Characteristics of included studies
1) The reference numbers shown in Table 1 do not match the reference numbers shown at the end of the text.
2) Gestational age is a major contributor to mortality and complications, whether single or multiple, and with or without prenatal steroids. It seems inappropriate that the gestational ages shown in Table 1 are shown by range alone. By showing the mean and standard deviation of gestational age for each cited article, we can see the distribution of gestational age in each study.

P.17-21, Table 2 and 3.
1) It should be clarified whether the grading of IVH, NEC, and ROP, which are indicators of morbidities, are the same in each cited article.
2) Similarly, it should be clarified whether the definition of BPD is the same in each cited article.
3) Neonatologists are also interested in how prenatal steroid administration affects the development of hemodynamically significant PDAs in preterm multiples. There is any consensus in past papers, so please add it to your analysis.

Discussion
If possible, please comment a little on whether the effects of antenatal steroid administration are clearly different between single and multiple births.

The statistical methods are OK, although to use the random-effect model to solve a problem of heterogeneity is not according to a proper knowledge of summarizing methods in meta-analysis. The fact that many do that does not mean it is correct. The assumptions of the fixed and the randoms effects model are different.

Regarding publication bias methods, the Beggs does not add anything to the Egger’s procedure (its statistical power is much lower than that of the Egger’s one). The p-values to consider a
 VERSION 1 – AUTHOR RESPONSE

Reviewer: 1  
Dr. Stephanie Blankenship, Northwestern University Feinberg School of Medicine  
Comments to the Author:  
The authors have performed a comprehensive meta-analysis of neonatal outcomes in multiple gestations following ACS administration. ACS administration and its impact on neonatal outcomes in multiples has previously been limited and the strength of impact in multiples (vs singletons) has not been thoroughly investigated, thus a meta-analysis of prior studies is valuable.

Specific recommendations and comments below:

Abstract

1. Line 19 (and lines 98)—I would consider listing mortality also a primary outcome (it’s listed in the manuscript title) as well as RDS. It seems approximately ½ of the included studies had either mortality or RDS as a primary outcome and this would add consistency.

Response: Thank you for suggestion. We have corrected the description in the Abstract and the Methods. (Line 20 and 99)

2. Line 22—delete "regarding" before "by"

Response: Thank you for reminder. We have corrected this. (Line 22)

3. Results, line 25—include the number of subjects collectively included in the 16 studies

Response: The number of subjects was added in the Abstract. (Line 25)

4. Line 39—Point 4: consider rephrasing. The heterogeneity did not materially alter what in some subgroups? Unclear what this is referring to, did not alter the outcomes/endpoints?

Response: Thank you for suggestion. This sentence has been rephrased (Line 40).

“The high heterogeneity remained in some subgroups, which suggested the presence of unexplained heterogeneity.”

Introduction

5. Line 43-44—It is curious you have quoted a reference for PTB specifically in the Netherlands. I
would consider using a more global estimate that refers to increasing rates of PTB.

**Response:** We had added references from other countries and modified the sentence (Line 44-45).

“The rate of PTB among multiple gestations remains at a high level in many countries [5-8].”

Methods

6. Lines 92-93--Choice of subgroup analyses are important.

**Response:** Thank you for comment. We believed the ACS completeness and administration-to-delivery are important factors of the effect of ACS administration among multiple gestations. Moreover, regarding the study design, we thought multi- or single center may play a role in the bias of results. Therefore, the above-mentioned variables were considered in the subgroup analyses.

Results

7. Line 123---Please report the number of subjects collectively included in the 16 studies

**Response:** Thank you for suggestion. There were 36,973 multiple gestation infants included in this study. (Line 124)

8. Line 149-150—the OR for multicenter studies crosses 1 (0.67-1.0), meaning nonsignificant, so this is not significantly decreased. Please correct this.

**Response:** Thank you for reminder. We have corrected this sentence (Line 156).

“Single center studies (0.47; 95% CI: 0.19-0.90) obtained a significantly decreased pooled OR (0.47; 95% CI: 0.19-0.90), however, multicenter studies did not (0.82; 95% CI: 0.67-1.00)”

9. It would be helpful to include weighted-averages describing the study population (mean or median gestational age, mean or median gestational latency/admin-to-delivery time), if available from the studies.

**Response:** Thank you for suggestion. Since some included studies did not provide information on the mean or median of gestational age at delivery and administration-to-delivery interval, we were only able to supplement the information in Table 1 if the information was available from the studies.

10. I would consider removing the OR reported the studies that had no information on ACS completion (for RDS, mortality, etc). It is hard to interpret what this means, as maybe those studies did have ACS completeness vs maybe they didn’t but, inferring that there is some sort of combination of these two possibilities, doesn’t really tell us anything additional of clinical value. Reporting the significant OR for ACS completion is important and valuable to report, but I would remove the
converse/unknown of unsure ACS completion. I have similar thoughts to excluding reference to the OR for studies with unknown/unreported administration-to-delivery time interval. However, it is still important to include in your discussion that for the overall analysis (i.e. non subgroup analysis), ACS completion likely attributed to heterogeneity in outcomes.

Response: Thank you for comments. Both ACS completeness and administration-to-delivery interval are essential factors of the effect of ACS. The subgroup analyses reported significant pooled ORs for complete ACS or interval ≤7 days but insignificant pooled ORs for unclear ACS completeness or interval in some outcomes (RDS, mortality and PVL). These results were important to show the priority of optimal ACS regarding completeness and administration-to-delivery interval. In real-world clinical practice, the difficulties of predicting PTB among multiple gestations may usually contribute to suboptimal ACS. In this regard, the efficacy of ACS with unclear completeness and administration-to-delivery interval, which was deemed to involve suboptimal ACS, is also informative.

11. Supplementary figures---could include the publication bias/Egger’s test graph for the primary outcome(s) to visually demonstrate no publication bias

Response: Thank you for suggestion. The Egger’s publication bias plots for each outcome were included in the supplementary materials (Figure S15-21).

12. Table 3---nice comprehensive summary of all the OR for outcomes, subgroup analyses. Well-presented.

Response: Thank you for comment.

13. Sensitivity analyses is well-done and appreciated.

Response: Thank you for comment.

Discussion
14. Lines 197—please rephrase/clarify this statement. I think you are referring to prior evidence of the efficacy of ACS among singletons has been abundant, but evidence among multiple births has been sparse and limited to small observational studies. Adding these clarifiers will strengthen the impact of this statement.

Response: Thank you for suggestion. We have rephrased this sentence in the revised manuscript. (Line 218-219)

15. Lines 202-204—it is important that the authors have offered a biologic basis for why ACS administration and effect on outcomes would differ in singletons vs multiples. I would recommend
expanding on this biologic proposition as it strengthens the importance of why this manuscript’s findings are significant and important for publication.

Response: Thank you for suggestion. We have expanded on the biologic plausibility of ACS in the manuscript. (Line 223-226)

“ACS has been confirmed to be effective to prepare fetal lung for air breathing through multiple mechanisms, including the induction of protein and enzymes, the acceleration of antioxidant production and induction of beta-receptor expression in alveolar cells as well as the acceleration of parenchymal change [40]. Though ACS has been internationally recommended in clinical practice, there is no difference in the guidelines for administration of ACS between multiple gestations and singletons [11 41]. From the view of pharmacokinetics, however, shorter half-life and faster clearance of corticosteroids in multiple pregnancies might raise some doubts in the effectiveness of ACS among multiple pregnancies when using the regime as same as for singletons [42].”

16. Lines 221—222—Please combine these two sentences “First, the majority…” and “And the overall quality of the evidence…” I would change the period after “avoided” to a comma and de-capitalize the “And”

Response: Thank you for suggestion. We have corrected this sentence in the revised manuscript. (Line 246-247)

17. Line 223—heterogeneity did not materially alter what? The outcomes/study endpoints?

Response: Thank you for comment. We have rephrased this sentence in the revised manuscript (Line 249-250). We were referring to that the heterogeneity remained at a high level in some subgroup even when subgroup analyses were performed.

18. In the limitations, the authors should emphasize large heterogeneity weakens the validity of the study conclusions, not just that heterogeneity is present due to retrospective nature, meta-analysis of heterogeneous studies, etc. Almost all I^2 values are >50% and often >80-90% ---this is very high, and certainly weakens the study findings. While an RCT would be useful, it would be unlikely to be feasible due to ethical reasons.

Response: Thank you for suggestion. The stress of high heterogeneity was noted in the limitation. (Line 248)

“Second, the high heterogeneity may weaken the validity of current results. Even though subgroup analyses were performed, the high heterogeneity did not materially alter remained in some subgroups, suggesting the presence of unexplained heterogeneity.”

19. Do the authors propose an explanation for why a multicenter vs single center finding may find
more or less significant impact of ACS on the outcomes, besides heterogeneity in multicenter studies? It is unclear clinically why multicenter vs single center actually has an impact on ACS administration and outcomes, and thus unclear to me why this subgroup analysis is of value other than to point out more significant heterogeneity in a multicenter study.

**Response:** Thank you for comment. From the view of observational study design, multicenter studies would reduce the patient selection bias to a certain degree and partially represented higher level of evidence compared with single center studies. This subgroup analysis may point out the role of patient selection bias in the heterogeneity of the pooled result, which is the reason why we included this variable in subgroup analyses.

Reviewer: 2
Dr. Kazuo Itabashi, Showa Daigaku Daigakuin Igaku Kenkyuka Igakubu

Comments to the Author:
This article by Lin et al. shows the effects of antenatal steroid administration on the outcomes of preterm multiples by meta-analysis and may provide clinically useful information. However, some corrections are needed.

1. p.6, Eligible criteria
Since the outcomes of late preterm infants is clearly better than that of infants born at 34 weeks’ gestation or less or very preterm infants, it is better not to include the article by Ben-David et al. in this analysis.

**Response:** Thank you for suggestion. Based on the inclusion criteria, we included all studies reporting efficacy of ACS among preterm (≤37 weeks) multiple gestations without specifying gestation age range. Although the study of Ben-David et al. included late preterm pregnancies, the pooled results remained stable after excluding this study, which can be seen in the sensitivity analysis (Fig S8).

2. p.10, Characteristics of included studies
1) The reference numbers shown in Table 1 do not match the reference numbers shown at the end of the text.

**Response:** Thank you for reminder. We have corrected this error.

2) Gestational age is a major contributor to mortality and complications, whether single or multiple, and with or without prenatal steroids. It seems inappropriate that the gestational ages shown in Table 1 are shown by range alone. By showing the mean and standard deviation of gestational age for each cited article, we can see the distribution of gestational age in each study.

**Response:** Thank you for suggestion. Since some included studies did not provide information on the
mean or median of gestational age at delivery and administration-to-delivery interval, we were only able to supplement the information in Table 1 if the information was available from the studies.

3. P.17-21, Table 2 and 3.
1) It should be clarified whether the grading of IVH, NEC, and ROP, which are indicators of morbidities, are the same in each cited article.
2) Similarly, it should be clarified whether the definition of BPD is the same in each cited article.

Response: Thank you for suggestion. In the Method, we had stated that the outcomes were defined by the authors of individual studies. As can be seen in Table 1, The grading of IVH, NEC and ROP were not all the same across included studies. Also, the definition of BPD varied in across studies. We added a clarification on this in the manuscript. (Line 131-132)

“The definition or grading of outcomes (IVH, NEC, ROP and BPD) was not all the same across included studies. Most studies included IVH ≥ grade III, NEC ≥ grade II and ROP ≥ grade III.”

3) Neonatologists are also interested in how prenatal steroid administration affects the development of hemodynamically significant PDAs in preterm multiples. There is any consensus in past papers, so please add it to your analysis.

Response: Thank you for suggestion. After reviewing the included studies, only three studies evaluated PDAs between no-ACS and ACS groups in the preterm multiples (Kuk et al., Spinillo et al. and Kong et al.).The meta-analysis was waived due to the small number of eligible studies.

Discussion

4. If possible, please comment a little on whether the effects of antenatal steroid administration are clearly different between single and multiple births.

Response: Thank you for suggestion. The pooled results in current study were similar to those obtained from singletons in RDS, IVH, PVL and neonatal death but not in NEC, ROP and BPD. (Line 227)

“These results were similar to previous ones obtained from singletons regarding RDS, IVH, PVL and neonatal death but not regarding NEC, ROP and BPD [10 43-45].”

Reviewer: 3
Dr. Miguel Delgado-Rodríguez, Universidad de Jaén

Comments to the Author:
1. The statistical methods are OK, although to use the random-effect model to solve a problem of heterogeneity is not according to a proper knowledge of summarizing methods in meta-analysis. The fact that many do that does not mean it is correct. The assumptions of the fixed and the random effects model are different.
Response: Thank you for comments. In the previous version of meta-analysis, the choice of a random-effect model or a fixed-effect model was based on the degree of heterogeneity. After we have carefully read the article of Michael Borenstein et al., which explained the key assumptions of random and fixed effect model, we thought that random effect models would be more appropriate in the current meta-analysis because of the large difference in the study size of included studies. We cannot discount a small study by giving it a very small weight (the way in a fixed effect model). This statement was added in the revised manuscript. And also, we modified to perform a random effect model in ROP.

Borenstein M, Hedges LV, Higgins JP, et al. A basic introduction to fixed-effect and random-effects models for meta-analysis. Research synthesis methods 2010;1(2):97-111 doi: 10.1002/jrsm.12[published Online First: Epub Date].

2. Regarding publication bias methods, the Beggs does not add anything to the Egger's procedure (its statistical power is much lower than that of the Egger's one). The p-values to consider a result statistically significant in these procedures should be mentioned.

Response: Thank you for comments. We had represented the P-values for Begg's and Egger's tests in Table 1. Also, the significant P-value in Egger's test was mentioned in the result section. (Line 185).

VERSION 2 – REVIEW

| REVIEWER | Blankenship, Stephanie  
Northwestern University Feinberg School of Medicine, Obstetrics and Gynecology |
|---------------------|---------------------------------------------------------------------|
| REVIEW RETURNED | 02-Jun-2021 |

| GENERAL COMMENTS | The authors have performed a comprehensive meta-analysis of neonatal outcomes in multiple gestations following ACS administration. ACS administration and its impact on neonatal outcomes in multiples has previously been limited and the strength of impact in multiples (vs singletons) has not been thoroughly investigated, thus a meta-analysis of prior studies is valuable. The authors have appropriately incorporated prior revisions. Small grammatical revisions are indicated but content is otherwise acceptable:  
Abstract  
Line 20—Suggest revising to: The primary outcomes were respiratory distress syndrome and mortality…”  
Line 388-point 3. Revise “trails” to “trials”  
Methods  
Line 97—Suggest revising “The primary outcome in the current meta-analysis was respiratory distress syndrome (RDS) and mortality” to “The primary outcomes in the current meta-analysis were respiratory distress syndrome (RDS) and mortality.” |
| Line | Suggestion |
|------|------------|
| 107  | Suggest revising, “Random effects models with the Mantel-Haenszel method was utilized” to “A random effects model with the Mantel-Haenszel method was utilized.” |
| 113  | Add “bias” after “publication” – I believe the authors are referring to publication bias, not just publication. |
| 135  | Delete “as” before “considered good…” |
| 209  | Suggest revising “in despite of the substantial heterogeneity” to “despite the substantial heterogeneity…” |
| 228  | Suggest deleting “an” before “incomplete ACS” OR alternatively, changing it to “an incomplete ACS course” |