OBSTETRICS & GYNECOLOGY

NOTICE: This document contains correspondence generated during peer review and subsequent revisions but before transmittal to production for composition and copyediting:

- Comments from the reviewers and editors (email to author requesting revisions)
- Response from the author (cover letter submitted with revised manuscript)*

**The corresponding author has opted to make this information publicly available.*

Personal or nonessential information may be redacted at the editor's discretion.

Questions about these materials may be directed to the *Obstetrics & Gynecology* editorial office: obgyn@greenjournal.org.

Date: Jan 16, 2020
To: "Tomomi Kotani" itoto@med.nagoya-u.ac.jp
From: "The Green Journal" em@greenjournal.org
Subject: Your Submission ONG-19-2052

RE: Manuscript Number ONG-19-2052

Antenatal Corticosteroids and Outcomes in Preterm Twins: A Propensity Score-Matched Cohort Study

Dear Dr. Kotani:

Your manuscript has been reviewed by the Editorial Board and by special expert referees. Although it is judged not acceptable for publication in Obstetrics & Gynecology in its present form, we would be willing to give further consideration to a revised version.

If you wish to consider revising your manuscript, you will first need to study carefully the enclosed reports submitted by the referees and editors. Each point raised requires a response, by either revising your manuscript or making a clear and convincing argument as to why no revision is needed. To facilitate our review, we prefer that the cover letter include the comments made by the reviewers and the editor followed by your response. The revised manuscript should indicate the position of all changes made. We suggest that you use the "track changes" feature in your word processing software to do so (rather than strikethrough or underline formatting).

Your paper will be maintained in active status for 21 days from the date of this letter. If we have not heard from you by Feb 06, 2020, we will assume you wish to withdraw the manuscript from further consideration.

REVIEWER COMMENTS:

Reviewer #1: This reasonably well-written manuscript uses propensity score matching to compare the benefits of antenatal corticosteroids (ACS) in singleton versus twin gestations. In general the benefits were similar in singletons and twins, with the notable exception of RDS, where ACS did not appear to benefit twins. In general, however, the results support current ACOG recommendations regarding use of ACS in both singleton and twin pregnancies. Specific comments:

Propensity score modeling: The propensity score seems to be reasonably done, but important details are either missing or hard to find in the paper.

1) What matching algorithm was used? Greedy matching? Nearest neighbor matching? Weighting? With or without replacement? There are several other matching algorithms as well.

2) It's hard to figure out how many ACS exposed could not be matched to an unexposed pregnancy because nobody had a score close enough. This is important, because the "causal question" applies only to an exposed pregnancy where there was a comparable unexposed one. This has an equivalent in clinical trials: If a treatment is absolutely indicated or absolutely contraindicated in someone (even if it's for reasons unrelated to the study outcome), then we can't say whether the treatment improves outcome in those excluded people. This does not diminish the validity of the comparison of those who did enroll in the trial (or comparably, who could be matched), but if a lot of women were excluded because a comparable woman could not be found (or because in a trial the treatment was absolutely indicated or contraindicated), then the generalizability of the results is limited.

3) In the last paragraph of the intro, I'm not a fan of the concept that propensity score analysis (matching, weighting or other) truly mimics a randomized trial. As a biostatistician I know always says "If you did not measure it, then you cannot control for it, no matter how fancy your methods are". In fact, there was a paper, I think in JAMA, by an author names Stukel (or Stuekel or something like that) about 20 years ago that compared PS matching with conventional regression and found that if the database is large enough, the methods are equivalent. PS does have some possible advantages (it is particularly helpful where there are lots of exposures but few outcomes, and it's helpful as I describe in (2) to identify treated people for whom a comparable untreated person cannot be found), but on balance in many settings it's over-rated.

4) It's not clear to me exactly how the interaction was done. Did they pool everyone into a conventional model, or did they maintain the matching and included the interaction term of treatmet*multiplicity (or treatment*chorionicity) in a conditional logistic model, or did they use an unconditional model? If the latter, did they include only the matched pairs, or did they include everyone-- including the pregnancies that could not be matched? The latter seems to destroy the advantage of PS matching.

5) similarly, regarding figure 2, did the authors test for effect modification by gestational age with interaction terms? That's the preferred way, but maybe I missed what was done when I read the paper.

Data quality
1) was there any quality control in the data? That is, did someone go to the original medical records to verify a sample for accuracy of coding of, say, hypertensive disorders? diabetes? RDS? nonreassuring fetal status? etc?

2) there were a lot of incomplete records (missing data). How did those pregnancies differ from the others? Could multiple imputation have been used?

3) what fraction of kids were lost to follow-up and so had no intermediate term outcomes? How did they differ from those with data present?

4) I see a major problem with how kids got into the database itself. The inclusion criteria were gestation <32 weeks AND birthweight <1500 grams. In the early 1990s, Cody Arnold and Michael Kramer published a paper in the Am J Epidemiol pointing out that the median weight at 32 weeks is already >1500 grams. Basically, that means that the cutoff will allow in probably all babies <28 weeks, but as gestational age advances, the 1500 gram cutoff means that an increasing number of the included babies are already IUGR. And since twins tend to be smaller than singletons, particularly after around 26 weeks, more twins than singletons will be allowed into the cohort. The only way to solve this problem is to base inclusion on gestational age, regardless of birthweight. The authors absolutely need to read the Arnold paper and decide whether this invalidates their entire study, and if not, what its impact on the results might be. I admit I'm concerned, but am not sure what the effect on the results might be. Regardless, it needs to be noted as a problem.

5) Another problem is that this dataset, by the very way it was collected, cannot mimic a randomized trial. That's because in a trial, one would take women at high risk of preterm birth (at <32 weeks) in the next week and would randomize them to ACS or placebo. However, at least some of those randomized women would not deliver in the next week, and in fact might go past 32 weeks or even make it to term. Yet under intent to treat, you would have to include those pregnancies in your assessment of study outcomes! However, because this study was based on preterm babies, those women who managed to stay pregnant past 32 weeks are excluded by definition, which is a gross violation of intent-to-treat. If, in that hypothetical trial, steroids did not impact the time to delivery, than this failure would lead to random misclassifications, or a loss of generalizability. But if treatment impacted the time to delivery, then severe bias might be introduced. Given the way the dataset was assembled, I don't think this is fixable.

6) how often was chorionicity missing?

Other
The enrollment ran from 2003-2015. That's a long time, and there might be time trends in survival and other outcomes. The PS matching should probably also include matching for years, or at least the authors need to document that the birth dates of the exposed and unexposed kids were comparable.

Between them, tables 2 and 4 make 32 comparisons, and 1 was found to be "significant" (RDS between singletons in twins). Unless there's a strong (ideally a priori) rationale for why the benefits of ACS on RDS, and no other outcome, should differ between twins and singletons, the authors need to acknowledge multiple comparisons and de-emphasize this finding.

Minor points
1) Line 73-- has it really become "standard of care" to use ACS in imminent preterm birth up to 37 weeks? ALPS study results notwithstanding, not everyone is ready to jump on the "steroids at 34-36 week bandwagon", and my reading of the ACOG recommendation is that it allows for a bit of 'wiggle room' at that time.

2) Lines 215-247. The paragraphs stress the differences and then go on to say that the interactions were not significant. The emphasis needs to be reversed. The differences should be de-emphasized, and the lack of significant interaction needs more emphasis.

Reviewer #2: The authors conducted a large, population-based study of VLBW infants in which they evaluated neonatal morbidity and mortality according to corticosteroid administration during pregnancy. Developmental testing was further performed in a subset of the children at age 3. The manuscript is very well written. Comments and questions follow.

1. Title. Would consider including something about the comparison between singletons and twins (or even singletons and monochorionic and dichorionic twins), as this is a strength of the study.

2. Precis. As there is no established definition medium-term outcomes, might phrase this differently, e.g. major morbidities prior to hospital discharge, as well as cerebral palsy and developmental impairment.

3. Abstract. The abstract summarizes the manuscript well.
a. As some may read only the abstract, would try to incorporate some key numerical data (%, odds ratios) into the abstract.
b. In lines 50-51, might mention something about developmental testing of the children (as otherwise readers might not otherwise appreciate that you performed this).
c. In line 54, are you referring to twin pregnancies or twin infants?

4. Introduction. This section is well written and appropriate in length and content.
Minor. In lines 77-82, the authors report that while a Cochrane systematic review demonstrated reductions in adverse outcomes in twins who received antenatal corticosteroids, the WHO questions its effectiveness. The Cochrane review was published 2 years after the WHO recommendations (WHO did not have the benefit of the Cochrane review when making their recommendations).

5. Methods. Clearly presented.
a. Monochorionic twins are at increased risk for specific complications (TTTS, TAPS) that confer significantly increased risks for neonatal morbidity. Was there a mechanism to address this in the propensity analysis?
b. Dichorionic twins are more prevalent than monochorionic twins (approximately 4 times more prevalent), but the authors present comparable numbers of each type of twin. Might explain why.
c. Minor. Short-term morbidities are defined in lines 134-142 and then listed again in lines 148-151. Would list these in just 1 place in this section -- can include brief definitions in parentheses as needed.

6. Results.
In figure 1, fewer than half of the pregnancies evaluated for short-term outcomes had assessment of medium-term outcomes. Was this because of the propensity score matching or loss to follow-up? If the former, please provide all available medium-term outcome data (for those who had short-term data). If the latter, please address more fully in the discussion.

7. Discussion.
a. General comment. Might try to emphasize what the study adds (a little more). Suggest emphasizing the CP and developmental quotient <70.
b. Despite the propensity risk matching, pregnant women who were able to receive ACS may differ (in risk) from those who did not have the opportunity to receive ACS. Might discuss reasons why women did not receive ACS.
c. Minor. Lines 270-272, 278-280 repeat content already presented and might be deleted.

8. Figures and tables. Complete, straightforward to follow.

Reviewer #3: This is an important retrospective study using a homogeneous Japanese population from the years 2003-2015 reporting on the impact of steroid therapy on singleton and twin gestations between 24 weeks and 32 weeks of gestation on short and medium term adverse outcomes. This is a large population based study that had significantly to our knowledge of the effects of steroid therapy on premature twin gestations. Although it is not a randomized control trial propensity score matching was performed to minimize bias associated with retrospective non-randomized studies and the results show evidence of a favorable impact on the use of antenatal corticosteroid therapy and twins in both major short and medium term adverse complications, regardless of chorionicity.

The following are my comments:

1. This is a rather homogeneous Japanese population and therefore it would be difficult to use any study conclusions to other populations. Evidence for this is that there is no information on the ethnicity/Race of the women in the study cohort. However, there really is no reason to think that the study results would not apply to other populations since they are consistent with other reports, and plausible.

2. The study time period is a bit broad but I do not believe that there was any thing of significance that occurred in NICU care neonatal care that could have significantly altered the results throughout the study time period. I would like the authors to make some comments whether the neonatal critical care was standardized across all of the neonatal intensive care units submitting information to the database. I am quite certain it is but I believe it should be stated or addressed in the methods section. In addition the care was the same for both singleton and twin gestations and therefore this adds to the validity of the results.

3. The paper is rather well written, concise and easy to follow. It is well referenced.

4. Because of the complexity of their statistical approach I recommend consultation with a statistician to evaluate the correctness of their analysis.

5. Several minor spelling or in order. And lines to 225 and 287 the word incubation was used when I believe it should be

intubation.

6. The paper entitled "Antenatal Betamethasone for Women at Risk for Late Preterm Delivery", Cynthia Gyamfi-Bannerman, M.D., Elizabeth A. Thom, Ph.D., Sean C. Blackwell, M.D., Alan T.N. Tita, M.D., Ph.D., Uma M. Reddy, M.D., M.P.H., George R. Saade, M.D., Dwight J. Rouse, M.D., David S. McKenna, M.D., Erin A.S. Clark, M.D., John M. Thorp, Jr., M.D., Edward K. Chien, M.D., Alan M. Peaceman, M.D., et al., for the NICHD Maternal-Fetal Medicine Units Network*.April 7, 2016, N Engl J Med 2016; 374:1311-1320, should be referenced in the second paragraph of the introduction section.

7. In reviewing the exclusion criteria for that isoimmunization was not a stated exclusion. I asked this since we know that isoimmunization can affect lung maturity. Was isoimmunization also excluded? Were cases of complications specific to monochorionicity excluded?

STATISTICAL EDITOR COMMENTS:

The Statistical Editor makes the following points that need to be addressed:

General: A significant proportion of both singletons and twins were omitted from the propensity matching process. The Authors should address this by (1) corroborating the analysis using logistic regression on the complete data set, while adjusting for all known maternal factors that could have affected the outcomes (2) modifying the matching algorithm so as to allow more entries into the match, while preserving minimal differences at baseline (3) comparing those included in the analyzed cohorts vs those excluded.

Another potential problem is the decrease in participants from the short to the medium term outcomes (~ 50% loss of respondents in each subset). The onus is on the Authors to demonstrate that the medium term outcome groups are representative of the original cohorts. The matching at the time of short-term analysis does not guarantee that the medium-term cohorts are also appropriately matched.

The analysis of twin outcomes is always more complicated than that of singletons. First, each set of twins shares many risk characteristics, so one cannot assume independence and ignore any correlation of outcomes within a twin set. If analyzed as independent events, the sample size is inflated and inference testing is erroneous. The Authors must adjust for intra twin correlation of events. Also, the 1st vs the 2nd twin are known to have different risk profiles for their outcomes, so pooling the outcomes of all twins may have obscured any intra twin difference. Should also analyze the outcomes for 1st and 2nd twins separately.

Fig 1: Should elaborate on the incomplete medical record details. Which variables were missing and does that mean that all variables were complete among those analyzed? If not, should enumerate all missing data.

EDITOR'S COMMENTS:

Thank you for this very clearly written manuscript. I am very appreciative of your use of an English language service to allow you to submit such a well-written paper.

We no longer require that authors adhere to the Green Journal format with the first submission of their papers. However, any revisions must do so. I strongly encourage you to read the instructions for authors (the general bits as well as those specific to the feature-type you are submitting). The instructions provide guidance regarding formatting, word and reference limits, authorship issues, and other things. Adherence to these requirements with your revision will avoid delays during the revision process, as well as avoid re-revisions on your part in order to comply with the formatting. As an example, we do not use subheadings in the discussion section such as “main findings”.

Line 50-51: Can you define what you mean by short- and medium-term outcomes? In the results section, please provide data to support your declarative statements. This will of course require editing of your abstract to remain within the 300 word limit.

P Values vs Effect Size and Confidence Intervals
While P values are a central part of inference testing in statistics, when cited alone, often the strength of the conclusion can be misunderstood. Whenever possible, the preferred citation should be in terms of an effect size, such as odds ratio or relative risk or the mean difference of a variable between two groups, expressed with appropriate confidence intervals. When such syntax is used, the P value has only secondary importance and often can be omitted or noted as footnotes in a Table format. Putting the results in the form of an effect size makes the result of the statistical test more clinically relevant

and gives better context than citing P values alone.
This is true for the abstract as well as the manuscript, tables and figures.

Please provide absolute values for variables, in addition to assessment of statistical significance.

We ask that you provide crude OR's followed by adjusted OR's for all relevant variables.

Line 73: ACS is standard of care in US for expected delivery within 7 days at 34 weeks or less but not up to 37 weeks. Not everyone has embraced the ALPS trial use of late preterm birth steroids from 34-37. Please edit this sentence.

Line 87: Please edit out the "to our knowledge" or similar wording. As the readers cannot gauge the depth and breadth of your knowledge, this phrase does not add significant meaning. You can either reference your literature search details (database searched and search terms used) that informed your knowledge, or you could say something noting that your cited references provide limited information about this point.

Line 89: As with the abstract notation, please define short- and medium-term time frames you used.

Line 100: Please note that your study was conducted from date 1 to date 2, not between those dates. As written, it would exclude the dates given .

Line 101: Are the level designation for NICU's in Japan the same as that in the US? Please so state or describe differences.
Line 104: Is higher order triplets and more? Please clarify.

Who enters data into this database? How is it validated?

Please explain how the data for the 3 year child hood status was obtained.

Line 119: Defined or define? Have these guidelines changed?

Line 128: The journal style does not support the use of the virgule (/) except in mathematical expressions. Please remove here and elsewhere.

Line 135 :Have you spelled out RDS previously? If not, please do so here.

139: What do you mean "clinical syndromes"?

Line 142: Were the psychological tests done by the child's local pediatrician and reported to the data base or were the children brought to a centralized location for testing?

Statistics: The statistical editor and one of your reviewers provide a great deal of feedback about necessary changes and additions needed in your description of your analysis and some of the analytic approach. Please address all of these.

Line 176: How was chorionicity assigned?

Line 218: I hope you discuss this in the discussion-interesting that there is a higher rate of chronic lung disease in these children. Not intuitive. Same for line 219—that rate of RDS didn't change. All will require discussion.

Line 225:What do you mean by "incubation at birth". Is this an "auto correct" and should be "Intubation"? Please address as well on line 287 and elsewhere this may show up (tables, etc).

Line 261-262. Although it is well known that ACS treatment offers a broad range of benefits to exposed singleton fetuses, the main reason it is given is to promote pulmonary maturity beyond expected for gestational age. Before you discuss the neuroprotective and other benefits, please add some discussion of the lack of observed benefit for RDS and chronic lung disease. This is a very important clinical distinction and should be presented early in your discussion.

Line 273: Delete "So far".

Lines 290-292: are you making these as possible explanations in general or is it true in your data set that these differences between groups existed? Does your data base have this information in it? Please make this clear.

Line 300: You describe a "Lower odds ratio" but you give an OR of 1.9 which is a higher odds ratio. It's not clear what this sentence is comparing: it reads like you are providing information about patients delivered by cesarean only. You should have only 1 OR for whether RDS differed between those with vs without labor, but you are providing 2 different OR's. Please edit.

Line 311: What ethical concerns if it's unclear if ACS is beneficial to twins?

Line 311: This is known as a primacy claim: yours is the first, biggest, best study of its kind. In order to make such a

claim, please provide the data bases you have searched (PubMED, Google Scholar, EMBASE for example) and the search terms used. IF not done, please edit it out of the paper.

Please edit your Figure 2 to use colors rather than shading differences for the different gestational ages. It's easier to interpret and does not cost you anything.

EDITORIAL OFFICE COMMENTS:

1. The Editors of Obstetrics & Gynecology are seeking to increase transparency around its peer-review process, in line with efforts to do so in international biomedical peer review publishing. If your article is accepted, we will be posting this revision letter as supplemental digital content to the published article online. Additionally, unless you choose to opt out, we will also be including your point-by-point response to the revision letter. If you opt out of including your response, only the revision letter will be posted. Please reply to this letter with one of two responses:
A. OPT-IN: Yes, please publish my point-by-point response letter.
B. OPT-OUT: No, please do not publish my point-by-point response letter.

2. As of December 17, 2018, Obstetrics & Gynecology has implemented an "electronic Copyright Transfer Agreement" (eCTA) and will no longer be collecting author agreement forms. When you are ready to revise your manuscript, you will be prompted in Editorial Manager (EM) to click on "Revise Submission." Doing so will launch the resubmission process, and you will be walked through the various questions that comprise the eCTA. Each of your coauthors will receive an email from the system requesting that they review and electronically sign the eCTA.

Please check with your coauthors to confirm that the disclosures listed in their eCTA forms are correctly disclosed on the manuscript's title page.

3. Our journal requires that all evidence-based research submissions be accompanied by a transparency declaration statement from the manuscript's lead author. The statement is as follows: "The lead author* affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained." *The manuscript's guarantor.

If you are the lead author, please include this statement in your cover letter. If the lead author is a different person, please ask him/her to submit the signed transparency declaration to you. This document may be uploaded with your submission in Editorial Manager.

4. Title: Please delete "Propensity Score–Matched Cohort Study" from your title.

5. In order for an administrative database study to be considered for publication in Obstetrics & Gynecology, the database used must be shown to be reliable and validated. In your response, please tell us who entered the data and how the accuracy of the database was validated. This same information should be included in the Materials and Methods section of the manuscript.

6. Please submit a completed STROBE checklist with your revision.

Responsible reporting of research studies, which includes a complete, transparent, accurate and timely account of what was done and what was found during a research study, is an integral part of good research and publication practice and not an optional extra. Obstetrics & Gynecology supports initiatives aimed at improving the reporting of health research, and we ask authors to follow specific guidelines for reporting randomized controlled trials (ie, CONSORT), observational studies (ie, STROBE), meta-analyses and systematic reviews of randomized controlled trials (ie, PRISMA), harms in systematic reviews (ie, PRISMA for harms), studies of diagnostic accuracy (ie, STARD), meta-analyses and systematic reviews of observational studies (ie, MOOSE), economic evaluations of health interventions (ie, CHEERS), quality improvement in health care studies (ie, SQUIRE 2.0), and studies reporting results of Internet e-surveys (CHERRIES). Include the appropriate checklist for your manuscript type upon submission. Please write or insert the page numbers where each item appears in the margin of the checklist. Further information and links to the checklists are available at http://ong.editorialmanager.com. In your cover letter, be sure to indicate that you have followed the CONSORT, MOOSE, PRISMA, PRISMA for harms, STARD, STROBE, CHEERS, SQUIRE 2.0, or CHERRIES guidelines, as appropriate.

7. Standard obstetric and gynecology data definitions have been developed through the reVITALize initiative, which was convened by the American College of Obstetricians and Gynecologists and the members of the Women's Health Registry Alliance. Obstetrics & Gynecology has adopted the use of the reVITALize definitions. Please access the obstetric and gynecology data definitions at https://www.acog.org/About-ACOG/ACOG-Departments/Patient-Safety-and-Quality-Improvement/reVITALize. If use of the reVITALize definitions is problematic, please discuss this in your point-by-point response to this letter.

8. Because of space limitations, it is important that your revised manuscript adhere to the following length restrictions by

manuscript type: Original Research reports should not exceed 22 typed, double-spaced pages (5,500 words). Stated page limits include all numbered pages in a manuscript (i.e., title page, précis, abstract, text, references, tables, boxes, figure legends, and print appendixes) but exclude references.

9. Specific rules govern the use of acknowledgments in the journal. Please note the following guidelines:

* All financial support of the study must be acknowledged.
* Any and all manuscript preparation assistance, including but not limited to topic development, data collection, analysis, writing, or editorial assistance, must be disclosed in the acknowledgments. Such acknowledgments must identify the entities that provided and paid for this assistance, whether directly or indirectly.
* All persons who contributed to the work reported in the manuscript, but not sufficiently to be authors, must be acknowledged. Written permission must be obtained from all individuals named in the acknowledgments, as readers may infer their endorsement of the data and conclusions. Please note that your response in the journal's electronic author form verifies that permission has been obtained from all named persons.
* If all or part of the paper was presented at the Annual Clinical and Scientific Meeting of the American College of Obstetricians and Gynecologists or at any other organizational meeting, that presentation should be noted (include the exact dates and location of the meeting).

10. Please remove the causal language from your Precis and other sections. Outcomes should be framed as "associations," not "effects." This should be edited through your submission.

11. The most common deficiency in revised manuscripts involves the abstract. Be sure there are no inconsistencies between the Abstract and the manuscript, and that the Abstract has a clear conclusion statement based on the results found in the paper. Make sure that the abstract does not contain information that does not appear in the body text. If you submit a revision, please check the abstract carefully.

In addition, the abstract length should follow journal guidelines. The word limits for different article types are as follows: Original Research articles, 300 words; Reviews, 300 words. Please provide a word count.

12. Only standard abbreviations and acronyms are allowed. A selected list is available online at http://edmgr.ovid.com/ong/accounts/abbreviations.pdf. Abbreviations and acronyms cannot be used in the title or précis. Abbreviations and acronyms must be spelled out the first time they are used in the abstract and again in the body of the manuscript.

13. The journal does not use the virgule symbol (/) in sentences with words. Please rephrase your text to avoid using "and/or," or similar constructions throughout the text. You may retain this symbol if you are using it to express data or a measurement.

14. In your Abstract, manuscript Results sections, and tables, the preferred citation should be in terms of an effect size, such as odds ratio or relative risk or the mean difference of a variable between two groups, expressed with appropriate confidence intervals. When such syntax is used, the P value has only secondary importance and often can be omitted or noted as footnotes in a Table format. Putting the results in the form of an effect size makes the result of the statistical test more clinically relevant and gives better context than citing P values alone.

If appropriate, please include number needed to treat for benefits (NNTb) or harm (NNTh). When comparing two procedures, please express the outcome of the comparison in U.S. dollar amounts.

Please standardize the presentation of your data throughout the manuscript submission. For P values, do not exceed three decimal places (for example, "P = .001"). For percentages, do not exceed one decimal place (for example, 11.1%").

15. Line 311: We discourage claims of first reports since they are often difficult to prove. How do you know this is the first report? If this is based on a systematic search of the literature, that search should be described in the text (search engine, search terms, date range of search, and languages encompassed by the search). If on the other hand, it is not based on a systematic search but only on your level of awareness, it is not a claim we permit.

16. Please review the journal's Table Checklist to make sure that your tables conform to journal style. The Table Checklist is available online here: http://edmgr.ovid.com/ong/accounts/table_checklist.pdf.

17. The American College of Obstetricians and Gynecologists' (ACOG) documents are frequently updated. These documents may be withdrawn and replaced with newer, revised versions. If you cite ACOG documents in your manuscript, be sure the reference you are citing is still current and available. If the reference you are citing has been updated (ie, replaced by a newer version), please ensure that the new version supports whatever statement you are making in your manuscript and then update your reference list accordingly (exceptions could include manuscripts that address items of historical interest). If the reference you are citing has been withdrawn with no clear replacement, please contact the editorial office for assistance (obgyn@greenjournal.org). In most cases, if an ACOG document has been withdrawn, it should not be referenced in your manuscript (exceptions could include manuscripts that address items of historical interest). All ACOG documents (eg, Committee Opinions and Practice Bulletins) may be found via the Clinical Guidance & Publications page at https://www.acog.org/Clinical-Guidance-and-Publications/Search-Clinical-Guidance.

18. Figure

Figure 1: This file may be resubmitted as-is with the revision.

Figure 2: Please upload a new version using solid colors as bars. Per journal style, we avoid using patterns within graph bars.

19. Authors whose manuscripts have been accepted for publication have the option to pay an article processing charge and publish open access. With this choice, articles are made freely available online immediately upon publication. An information sheet is available at http://links.lww.com/LWW-ES/A48. The cost for publishing an article as open access can be found at http://edmgr.ovid.com/acd/accounts/ifauth.htm.

Please note that if your article is accepted, you will receive an email from the editorial office asking you to choose a publication route (traditional or open access). Please keep an eye out for that future email and be sure to respond to it promptly.

20. If you choose to revise your manuscript, please submit your revision through Editorial Manager at http://ong.editorialmanager.com. Your manuscript should be uploaded in a word processing format such as Microsoft Word. Your revision's cover letter should include the following:

* A confirmation that you have read the Instructions for Authors (http://edmgr.ovid.com/ong/accounts/authors.pdf), and
* A point-by-point response to each of the received comments in this letter.

If you submit a revision, we will assume that it has been developed in consultation with your co-authors and that each author has given approval to the final form of the revision.

Again, your paper will be maintained in active status for 21 days from the date of this letter. If we have not heard from you by Feb 06, 2020, we will assume you wish to withdraw the manuscript from further consideration.

Sincerely,

Nancy C. Chescheir, MD
Editor-in-Chief

2018 IMPACT FACTOR: 4.965
2018 IMPACT FACTOR RANKING: 7th out of 83 ob/gyn journals

_____

Editor-in-Chief
OBSTETRICS & GYNECOLOGY
Professor Nancy C. Chescheir

RE: Manuscript Number: ONG-19-2052

Dear Dr. Chescheir
Thank you very much for your e-mail (dated 16 Jan 2020) enclosing the reviewers' comments of the manuscript (ID: ONG-19-2052). We also appreciate the time and effort you and each of the reviewers have dedicated to providing insightful and constructive feedback on ways to strengthen our paper. It is with great pleasure that we resubmit our article for further consideration. We have carefully reviewed the comments and have thoroughly revised the manuscript, figures and tables accordingly. Our responses are given in a point-by-point manner below. Changes to the manuscript are shown in track changes in the text. Our manuscript was followed by the STROBE guideline. We hope you will find our responses satisfactory, and hope that you will find this manuscript acceptable for publication in your journal.

Takafumi Ushida, the lead author of this manuscript, affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained. Each author has given approval to the final form of this revision.

In summary, we revised as follow;
1. Methods: We added detail information on propensity score matching, statistics and NRNJ database. In addition, we re-performed propensity score matching and all of the analyses to reflect the reviewers' comments.
2. Results and Discussion: Our study provided additional evidence that the effect of ACS treatment on offspring outcomes did not differ according to birth order of twin. We have revised the part of results and discussion thoroughly according to the reviewers' comments.
3. Figure 2: We uploaded a new version using solid colors.

Responses to the Editor's comments

1: As an example, we do not use subheadings in the discussion section such as "main findings".

Response 1: We removed subheadings in the discussion section.

2: Line 50-51: Can you define what you mean by short- and medium-term outcomes? In the results section, please provide data to support your declarative statements. This will of course require editing of your abstract to remain within the 300 word limit.

Response 2: Thank you for your comment. We defined the short- and medium-term outcomes in the abstract, and method section. And we provided some key numerical data in the results section of abstract (Line 49 and 55-58).

3: P Values vs Effect Size and Confidence Intervals
While P values are a central part of inference testing in statistics, when cited alone, often the strength of the conclusion can be misunderstood. Whenever possible, the preferred citation should be in terms of an effect size, such as odds ratio or relative risk or the mean difference of a variable between two groups, expressed with appropriate confidence intervals. When such syntax is used, the P value has only secondary importance and often can be omitted or noted as footnotes in a Table format. Putting the results in the form of an effect size makes the result of the statistical test more clinically relevant and gives better context than citing P values alone. This is true for the abstract as well as the manuscript, tables and figures. Please provide absolute values for variables, in addition to assessment of statistical significance. We ask that you provide crude OR's followed by adjusted OR's for all relevant variables.

Response 3: We described crude ORs and 95% CI as appropriate in the revised manuscript. In this study, PS matching was successfully performed, and no significant difference in any variables was observed between the two groups. Thus, we did not provide adjusted ORs.

4: Line 73: ACS is standard of care in US for expected delivery within 7 days at 34 weeks or less but not up to 37 weeks. Not everyone has embraced the ALPS trial use of late preterm birth steroids from 34-37. Please edit this sentence.

Response 4: Thank you for your suggestion. As you and another reviewer suggested, we have amended the sentence (Line 74).

5: Line 87: Please edit out the "to our knowledge" or similar wording. As the readers cannot gauge the depth and breadth of your knowledge, this phrase does not add significant meaning. You can either reference your literature search details (database searched and

search terms used) that informed your knowledge, or you could say something noting that your cited references provide limited information about this point.

Response 5: Thank you for your comment. We deleted these words (Line 88).

6: Line 89: As with the abstract notation, please define short- and medium-term time frames you used.

Response 6: As your comment #2, we defined short- and medium-term in the manuscript (Line 90-91).

7: Line 100: Please note that your study was conducted from date 1 to date 2, not between those dates. As written, it would exclude the dates given.

Response 7: We apologize for this error, and we have corrected the sentence (Line 99-100).

8: Line 101: Are the level designation for NICU's in Japan the same as that in the US? Please so state or describe differences.

Response 8: According to the guidelines for perinatal care 8th edition published 2017, the level of NICU in the USA is very similar to that of Japan. We described this in the method section (Line 100-101).

9: Line 104: Is higher order triplets and more? Please clarify.

Response 9: Thank you for your comment. We fixed Figure 1 and the manuscript (Line 103-104).

10: Who enters data into this database? How is it validated?

Response 10: Thank you for the questions. Doctors or medical staff enter clinical information into this database with use of a standard network database operation manual on a yearly basis. Basically, data abstractors choose tabs (yes, no, unknown) related to maternal and neonatal characteristics. The data were sent to the NRNJ Database Center. Database administrators in the Database Center check the data quality of clinical information, and ask the data abstractors at each facility to verify the correction of these data if necessary. In addition, the NRNJ publishes theses data on

web on a yearly basis (Line 111-113).

11: Please explain how the data for the 3 year child hood status was obtained.

Response 11: Basically, children born in extremely and very preterm are followed up every 3-6 months until 3-6 years of age in Japan. Physiological and neurodevelopmental assessments are performed around 36 months of age at each facility. Developmental Quotient is also evaluated by experienced clinical assessors at each facility, and is adjusted by chronological age at test (Line 139-141).

12: Line 119: Defined or define? Have these guidelines changed?

Response 12: We apologize for misunderstanding. The guideline recommends to use two intramuscular doses of 12 mg of betamethasone given 24 hours apart as ACS treatment (Line 119-120).

13: Line 128: The journal style does not support the use of the virgule (/) except in mathematical expressions. Please remove here and elsewhere.

Response 13: We removed the virgule in the manuscript and tables.

14: Line 135: Have you spelled out RDS previously? If not, please do so here.

Response 14: We have spelled out RDS in Line 79.

15: Line 139: What do you mean “clinical syndromes”?

Response 15: We apologize to use a wrong word. Neonatal sepsis was defined as clinical “symptoms” of bacteremia with the presence of a pathogenic bacterium from a blood culture (Line 136).

16: Line 142: Were the psychological tests done by the child’s local pediatrician and reported to the data base or were the children brought to a centralized location for testing?

Response 16: As response #11, the psychological tests were performed by experienced clinical assessors around 36 months of age at each facility (Line 139-141).

17: Statistics: The statistical editor and one of your reviewers provide a great deal of feedback

about necessary changes and additions needed in your description of your analysis and some of the analytic approach. Please address all of these.

Response 17: We carefully reviewed the statistical editor and reviewers comments.

18: Line 176: How was chorionicity assigned?

Response 18: Chorionicity is distinguished by ultrasound in the first trimester. If the chorionicity is unknown, these twins were excluded in this study (n=368).

19: Line 218: I hope you discuss this in the discussion-interesting that there is a higher rate of chronic lung disease in these children. Not intuitive. Same for line 219—that rate of RDS didn't change. All will require discussion.

Response 19: Thank you for your valuable comment. As you pointed out, our study showed increased incidence of CLD and home oxygen therapy in the ASC group in singletons, but not in twins. This is consistent with a previous study which used the same database within a different period (2003-2007) (Miyazaki K, 2015, Arch Gynecol Obstet). This can be attributed to the fact that improved survival rate by ACS treatment might increase the number of severe/serious neonates requiring prolonged mechanical ventilation, oxygen administration, and artificial nutrition. However, we need further research to investigate the reasons. We described in the discussion section (Line 271-276).

20: Line 225: What do you mean by "incubation at birth". Is this an "auto correct" and should be "Intubation"? Please address as well on line 287 and elsewhere this may show up (tables, etc).

Response 20: We apologize for these errors. We fixed the manuscript and tables.

21: Line 261-262. Although it is well known that ACS treatment offers a broad range of benefits to exposed singleton fetuses, the main reason it is given is to promote pulmonary maturity beyond expected for gestational age. Before you discuss the neuroprotective and other benefits, please add some discussion of the lack of observed benefit for RDS and chronic lung disease. This is a very important clinical distinction and should be presented early in your discussion.

Response 21: Thank you for your comment. As your suggestion, we first discussed the

lack of efficacy of ACS treatment on respiratory morbidities. In addition, we arranged the discussion section accordingly (Line 261-276).

22: Line 273: Delete “So far”.

Response 22: We deleted it.

23: Lines 290-292: are you making these as possible explanations in general or is it true in your data set that these differences between groups existed? Does your data base have this information in it? Please make this clear.

Response 23: Thank you for your question. These possible explanations are not based on our data. To reflect the reviewer #1 comment #13, we need to less emphasis on the difference of incidence of RDS. Thus, we rearrangement discussion section thoroughly considering your suggestion (comment #19 and 21) (Line 264-270).

24: Line 300: You describe a “Lower odds ratio” but you give an OR of 1.9 which is a higher odds ratio. It’s not clear what this sentence is comparing: it reads like you are providing information about patients delivered by cesarean only. You should have only 1 OR for whether RDS differed between those with vs without labor, but you are providing 2 different OR’s. Please edit.

Response 24: Thank you for your comment. We fixed the sentence (Line 267-268).

25: Line 311: What ethical concerns if it’s unclear if ACS is beneficial to twins?

Response 25: We agree with your opinion. We deleted the sentence.

26: Line 311: This is known as a primacy claim: yours is the first, biggest, best study of its kind. In order to make such a claim, please provide the data bases you have searched (PubMED, Google Scholar, EMBASE for example) and the search terms used. IF not done, please edit it out of the paper.

Response 26: We agree with your comment. We fixed the sentence (Line 302).

27: Please edit your Figure 2 to use colors rather than shading differences for the different gestational ages. It’s easier to interpret and does not cost you anything.

Response 27: Thank you for your valuable suggestion. We uploaded a new version using solid colors.

REVIEWER COMMENTS:

Response to Reviewer #1 comments

Propensity score modeling: The propensity score seems to be reasonably done, but important details are either missing or hard to find in the paper.

1: What matching algorithm was used? Greedy matching? Nearest neighbor matching? Weighting? With or without replacement? There are several other matching algorithms as well.

Response 1: Thank you for your comment. We apologize for insufficient statistical methods. We used Greedy nearest neighbor matching without replacement, and one-to-one pair matching in this study. A caliper width of 0.2 of the standard deviation of the logit of the propensity score was used for the developed propensity score, as previously recommended. (Austin *Pharm stat,* 2011). We described the detail in the method section (Line 162-165).

2: It's hard to figure out how many ACS exposed could not be matched to an unexposed pregnancy because nobody had a score close enough. This is important, because the "causal question" applies only to an exposed pregnancy where there was a comparable unexposed one. This has an equivalent in clinical trials: If a treatment is absolutely indicated or absolutely contraindicated in someone (even if it's for reasons unrelated to the study outcome), then we can't say whether the treatment improves outcome in those excluded people. This does not diminish the validity of the comparison of those who did enroll in the trial (or comparably, who could be matched), but if a lot of women were excluded because a comparable woman could not be found (or because in a trial the treatment was absolutely indicated or contraindicated), then the generalizability of the results is limited.

Response 2: Thank you for your comment. We apologize for insufficient data presentation in Figure 1 and result section. After PS matching, 71.5% (9,350/13,073) of singletons with ACS and 66.1% (2,529/3,824) of twins with ACS were matched in this study. We describe this in the manuscript and Figure 1 (Line 195-197). In addition, we compared the baseline characteristics between PS-matched cohort and PS-un-matched

cohort to reflect statistical editor comment general 1 (Supplementary Table 3).

3: In the last paragraph of the intro, I'm not a fan of the concept that propensity score analysis (matching, weighting or other) truly mimics a randomized trial. As a biostatistician I know always says "If you did not measure it, then you cannot control for it, no matter how fancy your methods are". In fact, there was a paper, I think in JAMA, by an author names Stukel (or Stuekel or something like that) about 20 years ago that compared PS matching with conventional regression and found that if the database is large enough, the methods are equivalent. PS does have some possible advantages (it is particularly helpful where there are lots of exposures but few outcomes, and it's helpful as I describe in (2) to identify treated people for whom a comparable untreated person cannot be found), but on balance in many settings it's over-rated.

Response 3: Thank you for your valuable comment as a biostatistician. We rephrased the sentence (Line 92-93).

4: It's not clear to me exactly how the interaction was done. Did they pool everyone into a conventional model, or did they maintain the matching and included the interaction term of treatment*multiplicity (or treatment*chorionicity) in a conditional logistic model, or did they use an unconditional model? If the latter, did they include only the matched pairs, or did they include everyone-- including the pregnancies that could not be matched? The latter seems to destroy the advantage of PS matching.

Response 4: Thank you for your comment. In the interaction analysis, we maintained the matching and included the interaction term of treatment*multiplicity (or treatment*chorionicity or treatment*birth order) in a conditional logistic model. So, we included only the matched pairs in the interaction analysis. We described in the section of methods (Line 176-184) and Figure 1 (in the bottom).

5: similarly, regarding figure 2, did the authors test for effect modification by gestational age with interaction terms? That's the preferred way, but maybe I missed what was done when I read the paper.

Response 5: Figure 2 showed the prevalence of neonatal complications stratified by gestational weeks without adjustment of covariates. We described in the method section (Line 186-188).

Data quality

6: was there any quality control in the data? That is, did someone go to the original medical records to verify a sample for accuracy of coding of, say, hypertensive disorders?, diabetes?, RDS?, nonreassuring fetal status? etc?

Response 6: Thank you for the questions. As we answered (editor comment #10), doctors or medical staff enter clinical information into this database with use of a standard network database operation manual on a yearly basis. Basically, data abstractors choose tabs (yes, no, unknown) related to maternal and neonatal characteristics. The data were sent to the NRNJ Database Center. Database administrators in the Database Center check the data quality of clinical information, and ask the data abstractors at each facility to verify the correction of these data if necessary. In addition, the NRNJ publishes theses data on web on a yearly basis from 2003 (Line 111-113).

7: there were a lot of incomplete records (missing data). How did those pregnancies differ from the others? Could multiple imputation have been used?

Response 7: Thank you for your question. We uploaded a new version of Figure 1 with details of all incomplete medical records. We needed to use only cases with complete maternal data to perform PS matching and for improvement of data reliability. PS matching were unable to perform in cases of incomplete data. In addition, as many as 13 variables were included to generate propensity score for appropriate matching algorithm. That's why incomplete data were more than 5,000. Although multiple imputation is a useful tool, it needs enormous calculation to impute missing data, it does not always perfectly cover a missing data, and theoretical interpretation of the results after multiple imputation is difficult. So, we did not perform multiple imputation in this study. We analyzed baseline characteristics in singletons and twins with complete data and without complete data (Supplementary Table 1). We described this in the section of method (Line 105-108).

8: what fraction of kids were lost to follow-up and so had no intermediate term outcomes? How did they differ from those with data present?

Response 8: Thank you for your comment. We compared the characteristic and outcomes between children with medium-term outcomes and children without medium-term outcomes after PS matching (Supplementary Table 5 and 6).

9: I see a major problem with how kids got into the database itself. The inclusion criteria

were gestation <32 weeks AND birthweight <1500 grams. In the early 1990s, Cody Arnold and Michael Kramer published a paper in the Am J Epidemiol pointing out that the median weight at 32 weeks is already >1500 grams. Basically, that means that the cutoff will allow in probably all babies <28 weeks, but as gestational age advances, the 1500 gram cutoff means that an increasing number of the included babies are already IUGR. And since twins tend to be smaller than singletons, particularly after around 26 weeks, more twins than singletons will be allowed into the cohort. The only way to solve this problem is to base inclusion on gestational age, regardless of birthweight. The authors absolutely need to read the Arnold paper and decide whether this invalidates their entire study, and if not, what its impact on the results might be. I admit I'm concerned, but am not sure what the effect on the results might be. Regardless, it needs to be noted as a problem.

Response 9: Thank you for your valuable comment. We were also aware of the problem of the NRNJ database. According to a sex-specific Japanese neonatal anthropometric chart in 2000, 30 weeks 0 day, 1500g =0.69SD (75.5%tile), 31 weeks 6 day, 1500g =-0.64SD (26.0%tile). Thus, approximately 25% of infants born at 30 gestational weeks, and 50-75% of infants born at 31 gestational weeks were not eligible in this study. As you pointed out, the ideal inclusion criteria of this database is based on the gestational age regardless of birthweight. We could not deny the possibility that this inclusion criteria introduces bias and affects results to some extent. However, the inclusion criteria of the NRNJ database is similar to other neonatal network databases in Italy, Israel, and Spain. And many reports were published using these databases.

10: Another problem is that this dataset, by the very way it was collected, cannot mimic a randomized trial. That's because in a trial, one would take women at high risk of preterm birth (at <32 weeks) in the next week and would randomize them to ACS or placebo. However, at least some of those randomized women would not deliver in the next week, and in fact might go past 32 weeks or even make it to term. Yet under intent to treat, you would have to include those pregnancies in your assessment of study outcomes! However, because this study was based on preterm babies, those women who managed to stay pregnant past 32 weeks are excluded by definition, which is a gross violation of intent-to-treat. If, in that hypothetical trial, steroids did not impact the time to delivery, than this failure would lead to random misclassifications, or a loss of generalizability. But if treatment impacted the time to delivery, then severe bias might be introduced. Given the way the dataset was assembled, I don't think this is fixable.

Response 10: Thank you for your valuable comment. We rephrased the sentence (Line 92-93).

11: how often was chorionicity missing?

Response 11: 368 cases were missing. Approximately 5%.

12: The enrollment ran from 2003-2015. That's a long time, and there might be time trends in survival and other outcomes. The PS matching should probably also include matching for years, or at least the authors need to document that the birth dates of the exposed and unexposed kids were comparable.

Response 12: Thank you for your comment. We did not include calendar year of delivery to generate propensity score in previous version. So in the revised version, we included calendar year of delivery in the PS matching. After PS matching, the birth dates of the exposed and unexposed children were comparable (Table 1). In addition, we also included birth order in the PS matching in the revised version to reflect the statistical editor comments.

13: Between them, tables 2 and 4 make 32 comparisons, and 1 was found to be "significant" (RDS between singletons in twins). Unless there's a strong (ideally a priori) rationale for why the benefits of ACS on RDS, and no other outcome, should differ between twins and singletons, the authors need to acknowledge multiple comparisons and de-emphasize this finding.

Response 13: Thank you for your comment. We discussed with Dr. Akihiro Hirakawa and Dr. Ryo Sadachi (our co-author, specialists for statistics). We have discussed your concern with them. In general, it does not rigorously adjust the multiplicity of testing in an exploratory retrospective study. Considering your comment, we added the policy of multiplicity adjustment in the statistical analysis section of the revised manuscript (Line 185-186).

Minor points

14: Line 73-- has it really become "standard of care" to use ACS in imminent preterm birth up to 37 weeks? ALPS study results notwithstanding, not everyone is ready to jump on the "steroids at 34-36 week bandwagon", and my reading of the ACOG recommendation is that it allows for a bit of 'wiggle room' at that time.

Response 14: Thank you for your comment. We agree with you and other reviewers' comments. We rephrased the sentence (Line 74).

15: Lines 215-247. The paragraphs stress the differences and then go on to say that the interactions were not significant. The emphasis needs to be reversed. The differences should be de-emphasized, and the lack of significant interaction needs more emphasis.

Response 15: Thank you for your comment. As your suggestion, we emphasized on the outcomes with the lack of significant interaction.

Response to reviewer #2

1: Title. Would consider including something about the comparison between singletons and twins (or even singletons and monochorionic and dichorionic twins), as this is a strength of the study.

Response 1: Thank you for your comment. In this study, we performed several kinds of comparison (singleton vs twin, MC vs DC, and 1st and 2nd twins). We feel just "Antenatal Corticosteroids and Outcomes in Preterm Twins" would be better.

2: Precis. As there is no established definition medium-term outcomes, might phrase this differently, e.g. major morbidities prior to hospital discharge, as well as cerebral palsy and developmental impairment.

Response 2: Thank you for your comment. As you and another reviewer comments, we defined the short- and medium-term outcomes in the abstract and manuscript.

3: Abstract. The abstract summarizes the manuscript well.
a. As some may read only the abstract, would try to incorporate some key numerical data (%, odds ratios) into the abstract.
b. In lines 50-51, might mention something about developmental testing of the children (as otherwise readers might not otherwise appreciate that you performed this).
c. In line 54, are you referring to twin pregnancies or twin infants?

Response 3a-c: Thank you for your comments. We added key numerical data in the abstract (Line 55-58). We are referring to twin infants.

4: Introduction. This section is well written and appropriate in length and content.
Minor. In lines 77-82, the authors report that while a Cochrane systematic review demonstrated reductions in adverse outcomes in twins who received antenatal corticosteroids,

the WHO questions its effectiveness. The Cochrane review was published 2 years after the WHO recommendations (WHO did not have the benefit of the Cochrane review when making their recommendations).

Response 4: Thank you for your information. We added information on published year for readers (Line 78 and 82).

5: Methods. Clearly presented.
a. Monochorionic twins are at increased risk for specific complications (TTTS, TAPS) that confer significantly increased risks for neonatal morbidity. Was there a mechanism to address this in the propensity analysis?

Response 5a: Thank you for your comment. Unfortunately, we did not perform PS matching MC twin and DC twin. But, we can compare short- and medium-term outcomes in the MC 1st, MC 2nd, DC 1st, and DC 2nd twins. (Supplementary Table 9 and 10). As you pointed out, we can see that MC twins had increased risks for neonatal morbidity.

b. Dichorionic twins are more prevalent than monochorionic twins (approximately 4 times more prevalent), but the authors present comparable numbers of each type of twin. Might explain why.

Response 5b: Thank you for your comment. According to the previous studies (written in Japanese), the prevalence of DC twins is approximately 2-2.5 times compared with that of MC twins in Japan (probably due to restriction of the number of embryo transfer since 2008). The data included preterm and term twins. However, as you know, MC twins have increased risks for very preterm birth due to discordant twin, TTTS, or TAPS compared with DC twins. That's why the number of MC twins (n=3,115) is similar to DC twins (n=3,431).

c. Minor. Short-term morbidities are defined in lines 134-142 and then listed again in lines 148-151. Would list these in just 1 place in this section -- can include brief definitions in parentheses as needed.

Response 5c: Thank you for your comment. We tried to fix the method section to avoid duplication, but we could not incorporate the definition and outcomes appropriately.

6: Results.

In figure 1, fewer than half of the pregnancies evaluated for short-term outcomes had assessment of medium-term outcomes. Was this because of the propensity score matching or loss to follow-up? If the former, please provide all available medium-term outcome data (for those who had short-term data). If the latter, please address more fully in the discussion.

Response 6: Thank you for your question. Low follow-up rate at 3 years of age was because of loss to follow-up. We described this limitation in the section of discussion (Line 318-320). In addition, we analyzed the baseline characteristics between in singletons and twins with medium-term outcomes and without medium-term outcomes (Supplementary Table 5).

7: Discussion.
a. General comment. Might try to emphasize what the study adds (a little more). Suggest emphasizing the CP and developmental quotient <70.

Response 7a: Thank you for your comment. In revised version, we added an additional evidence that effect of ACS treatment on offspring outcomes did not differ according to birth order of twin (Line 255-260).

b. Despite the propensity risk matching, pregnant women who were able to receive ACS may differ (in risk) from those who did not have the opportunity to receive ACS. Might discuss reasons why women did not receive ACS.

Response 7b: Thank you for your valuable comment. We agree with your opinion. As you pointed out, we need to consider the reason why patients in the non-ACS group did/could not receive ACS treatment (e.g. placental abruption, precipitate delivery, and eclampsia), which might affect offspring outcomes at 3 years of age. We added comment in the part of Limitation (Line 316-318)

c. Minor. Lines 270-272, 278-280 repeat content already presented and might be deleted.

Response 7c: Thank you for your comment. We deleted the sentence to avoid duplication.

8. Figures and tables. Complete, straightforward to follow.

Response: Thank you.

Response to reviewer #3:

1: This is a rather homogeneous Japanese population and therefore it would be difficult to use any study conclusions to other populations. Evidence for this is that there is no information on the ethnicity/Race of the women in the study cohort. However, there really is no reason to think that the study results would not apply to other populations since they are consistent with other reports, and plausible.

Response: We agree with your opinion. As your comment, further research is required to verify our findings in different populations (e.g. different gestational age, race, and region) for improvement of generalizability.

2: The study time period is a bit broad but I do not believe that there was anything of significance that occurred in NICU care neonatal care that could have significantly altered the results throughout the study time period. I would like the authors to make some comments whether the neonatal critical care was standardized across all of the neonatal intensive care units submitting information to the database. I am quite certain it is but I believe it should be stated or addressed in the methods section. In addition the care was the same for both singleton and twin gestations and therefore this adds to the validity of the results.

Response 2: Thank you for your valuable comment. As your suggestion, we described the basic management of neonatal care in the methods section (Line 143-144).

3: The paper is rather well written, concise and easy to follow. It is well referenced.

Response 3: Thank you.

4: Because of the complexity of their statistical approach I recommend consultation with a statistician to evaluate the correctness of their analysis.

Response 4: Our co-authors, Dr Akihiro Hirakawa and Ryo Sadachi are specialists for statistics. They conducted the analyses and confirmed the correctness of our analyses. In addition, we carefully reviewed statistical editor and another statistical reviewer's comments.

5: Several minor spelling or in order. And lines to 225 and 287 the word incubation was used when I believe it should be intubation.

Response 5: We apologize the error. We fixed the word.

6: The paper entitled "Antenatal Betamethasone for Women at Risk for Late Preterm Delivery", Cynthia Gyamfi-Bannerman, M.D., Elizabeth A. Thom, Ph.D., Sean C. Blackwell, M.D., Alan T.N. Tita, M.D., Ph.D., Uma M. Reddy, M.D., M.P.H., George R. Saade, M.D., Dwight J. Rouse, M.D., David S. McKenna, M.D., Erin A.S. Clark, M.D., John M. Thorp, Jr., M.D., Edward K. Chien, M.D., Alan M. Peaceman, M.D., et al., for the NICHD Maternal-Fetal Medicine Units Network*.April 7, 2016, N Engl J Med 2016; 374:1311-1320, should be referenced in the second paragraph of the introduction section.

Response 6: Thank you for your suggestion. However, as editor and other reviewers commented, ACS treatment for women at risk for late preterm delivery has been recognized to be effective recently, but not everyone has embraced the use of late preterm birth steroids from 34-37. To reflect the editor's comment, we rephrased the sentence (Line 74).

7: In reviewing the exclusion criteria for that isoimmunization was not a stated exclusion. I asked this since we know that isoimmunization can affect lung maturity. Was isoimmunization also excluded? Were cases of complications specific to monochorionicity excluded?

R**esponse 7: Thank you for your comment. Unfortunately, the NRNJ database did not cover the information on isoimmunization or TTTS/TAPS. We could not exclude these cases.**

<u>Response to STATISTICAL EDITOR COMMENTS</u>

The Statistical Editor makes the following points that need to be addressed:

General: A significant proportion of both singletons and twins were omitted from the propensity matching process. The Authors should address this by (1) corroborating the analysis using logistic regression on the complete data set, while adjusting for all known maternal factors that could have affected the outcomes (2) modifying the matching algorithm so as to allow more entries into the match, while preserving minimal differences at baseline (3) comparing those included in the analyzed cohorts vs those excluded.

Response: Thank you for your comment. At first, we need to apologize that we could not describe detail method of PS matching in previous version. We used Greedy nearest neighbor matching without replacement, and used one-to-one pair matching in this study. A caliper width of 0.2 of the standard deviation of the logit of the propensity score was used for the developed propensity score, as previously recommended (Austin *Pharm stat,* 2011). Although approximately 29~34% of singletons and twins were excluded by PS matching, we used optimal caliper widths for ideal PS matching. Even if we can change the caliper width >0.2 to include more singletons and twins, there is a risk for increased standardized difference, which may affect several outcomes. (3) As your suggestion, we compared the characteristics between matched and un-matched cohort (Supplementary Table 3).

Another potential problem is the decrease in participants from the short to the medium term outcomes (~ 50% loss of respondents in each subset). The onus is on the Authors to demonstrate that the medium term outcome groups are representative of the original cohorts. The matching at the time of short-term analysis does not guarantee that the medium-term cohorts are also appropriately matched.

Response: Thank you for your comment. We admit that this is one of the limitations in this study. We described this limitation in the section of discussion (Line 318-320). We analyzed the baseline characteristics and outcomes between cases with medium-term outcomes and cases without medium-term outcomes (Supplementary Table 5 and 6). We found that there are several statistically significant differences between the two groups. However, the differences considered to be small in clinical situation.

The analysis of twin outcomes is always more complicated than that of singletons. First, each set of twins shares many risk characteristics, so one cannot assume independence and ignore any correlation of outcomes within a twin set. If analyzed as independent events, the sample size is inflated and inference testing is erroneous. The Authors must adjust for intra twin correlation of events. Also, the 1st vs the 2nd twin are known to have different risk profiles for their outcomes, so pooling the outcomes of all twins may have obscured any intra twin difference. Should also analyze the outcomes for 1st and 2nd twins separately.

Response: Thank you for your valuable comment. We agree with your comment that each set of twins, especially MC twins, shares many risks for subsequent outcomes, and intra twin correlation of events need to be considered when we evaluate the outcomes. Even though we consulted with specialists for statistics of our co-authors, we could not

appropriately adjust the outcomes by suitable methods considering the intra twin correlation in this study because it is quite complicated and may be different depends on the chorionicity, birth order, and neonatal complications. We described this limitation in the section of discussion (Line 320-322). However, as your suggestion, we performed an additional analysis which analyze the offspring outcomes for the 1st and 2nd twins separately. We found that the effect of ACS treatment on offspring outcomes did not differ according to birth order of twin.

Fig 1: Should elaborate on the incomplete medical record details. Which variables were missing and does that mean that all variables were complete among those analyzed? If not, should enumerate all missing data.

Response: We uploaded a new version of Figure 1 with details of all incomplete data. We needed to use only cases with complete maternal data to perform PS matching and for improvement of data reliability. In addition, as many as 13 variables were included to generate propensity score for appropriate matching algorithm. That's why incomplete data were more than 5,000.

Response to EDITORIAL OFFICE COMMENTS:

1. The Editors of Obstetrics & Gynecology are seeking to increase transparency around its peer-review process, in line with efforts to do so in international biomedical peer review publishing. If your article is accepted, we will be posting this revision letter as supplemental digital content to the published article online. Additionally, unless you choose to opt out, we will also be including your point-by-point response to the revision letter. If you opt out of including your response, only the revision letter will be posted. Please reply to this letter with one of two responses:

A. OPT-IN: Yes, please publish my point-by-point response letter.

B. OPT-OUT: No, please do not publish my point-by-point response letter.

Response 1: OPT-IN: Yes, please publish our point-by-point response letter.

2. As of December 17, 2018, Obstetrics & Gynecology has implemented an "electronic Copyright Transfer Agreement" (eCTA) and will no longer be collecting author agreement forms. When you are ready to revise your manuscript, you will be prompted in Editorial Manager (EM) to click on "Revise Submission." Doing so will launch the resubmission process, and you will be walked through the various questions that comprise the eCTA. Each

of your coauthors will receive an email from the system requesting that they review and electronically sign the eCTA.
Please check with your coauthors to confirm that the disclosures listed in their eCTA forms are correctly disclosed on the manuscript's title page.

Response 2: We have confirmed this with co-authors.

3. Our journal requires that all evidence-based research submissions be accompanied by a transparency declaration statement from the manuscript's lead author. The statement is as follows: "The lead author* affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained." *The manuscript's guarantor.

If you are the lead author, please include this statement in your cover letter. If the lead author is a different person, please ask him/her to submit the signed transparency declaration to you. This document may be uploaded with your submission in Editorial Manager.

Response 3: We included this statements in our cover letter. And we uploaded the signed transparency declaration.

4. Title: Please delete "Propensity Score–Matched Cohort Study" from your title.

Response 4: We deleted "Propensity Score–Matched Cohort Study" from the title.

5. In order for an administrative database study to be considered for publication in Obstetrics & Gynecology, the database used must be shown to be reliable and validated. In your response, please tell us who entered the data and how the accuracy of the database was validated. This same information should be included in the Materials and Methods section of the manuscript.

Response 5: Doctors or medical staff enter clinical information into this database with use of a standard network database operation manual on a yearly basis. Basically, data abstractors choose tabs (yes, no, unknown) related to maternal and neonatal characteristics. The data are sent to the NRNJ Database Center. Database administrators in the Database Center check the data quality of clinical information, and ask the data abstractors at each facility to verify the correction of these data if

necessary. In addition, the NRNJ publishes theses data on web on a yearly basis (Line 111-113).

6. Please submit a completed STROBE checklist with your revision.
Responsible reporting of research studies, which includes a complete, transparent, accurate and timely account of what was done and what was found during a research study, is an integral part of good research and publication practice and not an optional extra. Obstetrics & Gynecology supports initiatives aimed at improving the reporting of health research, and we ask authors to follow specific guidelines for reporting randomized controlled trials (ie, CONSORT), observational studies (ie, STROBE), meta-analyses and systematic reviews of randomized controlled trials (ie, PRISMA), harms in systematic reviews (ie, PRISMA for harms), studies of diagnostic accuracy (ie, STARD), meta-analyses and systematic reviews of observational studies (ie, MOOSE), economic evaluations of health interventions (ie, CHEERS), quality improvement in health care studies (ie, SQUIRE 2.0), and studies reporting results of Internet e-surveys (CHERRIES). Include the appropriate checklist for your manuscript type upon submission. Please write or insert the page numbers where each item appears in the margin of the checklist. Further information and links to the checklists are available at http://ong.editorialmanager.com. In your cover letter, be sure to indicate that you have followed the CONSORT, MOOSE, PRISMA, PRISMA for harms, STARD, STROBE, CHEERS, SQUIRE 2.0, or CHERRIES guidelines, as appropriate.

Response 6: We included the STROBE checklist at the end of this letter.

7. Standard obstetric and gynecology data definitions have been developed through the reVITALize initiative, which was convened by the American College of Obstetricians and Gynecologists and the members of the Women's Health Registry Alliance. Obstetrics & Gynecology has adopted the use of the reVITALize definitions. Please access the obstetric and gynecology data definitions at https://www.acog.org/About-ACOG/ACOG-Departments/Patient-Safety-and-Quality-Improvement/reVITALize. If use of the reVITALize definitions is problematic, please discuss this in your point-by-point response to this letter.

Response 7: We followed the reVITALize definitions in this study.

8. Because of space limitations, it is important that your revised manuscript adhere to the following length restrictions by manuscript type: Original Research reports should not exceed 22 typed, double-spaced pages (5,500 words). Stated page limits include all numbered pages in a manuscript (i.e., title page, précis, abstract, text, references, tables, boxes, figure legends,

and print appendixes) but exclude references.

Response 8: Our manuscript is 22 pages long without references (4,970 words without references).

9. Specific rules govern the use of acknowledgments in the journal. Please note the following guidelines:

* All financial support of the study must be acknowledged.
* Any and all manuscript preparation assistance, including but not limited to topic development, data collection, analysis, writing, or editorial assistance, must be disclosed in the acknowledgments. Such acknowledgments must identify the entities that provided and paid for this assistance, whether directly or indirectly.
* All persons who contributed to the work reported in the manuscript, but not sufficiently to be authors, must be acknowledged. Written permission must be obtained from all individuals named in the acknowledgments, as readers may infer their endorsement of the data and conclusions. Please note that your response in the journal's electronic author form verifies that permission has been obtained from all named persons.
* If all or part of the paper was presented at the Annual Clinical and Scientific Meeting of the American College of Obstetricians and Gynecologists or at any other organizational meeting, that presentation should be noted (include the exact dates and location of the meeting).

Response 9: All of the above have been included on the title page of the manuscript.

10. Please remove the causal language from your Precis and other sections. Outcomes should be framed as "associations," not "effects." This should be edited through your submission.

Response 10: We removed the causal language from Precis and other sections.

11. The most common deficiency in revised manuscripts involves the abstract. Be sure there are no inconsistencies between the Abstract and the manuscript, and that the Abstract has a clear conclusion statement based on the results found in the paper. Make sure that the abstract does not contain information that does not appear in the body text. If you submit a revision, please check the abstract carefully.

In addition, the abstract length should follow journal guidelines. The word limits for different article types are as follows: Original Research articles, 300 words; Reviews, 300 words. Please provide a word count.

Response 11: The abstract matches the body of the manuscript and follows the journal guideline.

12. Only standard abbreviations and acronyms are allowed. A selected list is available online at http://edmgr.ovid.com/ong/accounts/abbreviations.pdf. Abbreviations and acronyms cannot be used in the title or précis. Abbreviations and acronyms must be spelled out the first time they are used in the abstract and again in the body of the manuscript.

Response 12: Only standard abbreviations and acronyms were used in the manuscript.

13. The journal does not use the virgule symbol (/) in sentences with words. Please rephrase your text to avoid using "and/or," or similar constructions throughout the text. You may retain this symbol if you are using it to express data or a measurement.

Response 13: We have removed all used of the virgule symbol throughout the text and tables.

14. In your Abstract, manuscript Results sections, and tables, the preferred citation should be in terms of an effect size, such as odds ratio or relative risk or the mean difference of a variable between two groups, expressed with appropriate confidence intervals. When such syntax is used, the P value has only secondary importance and often can be omitted or noted as footnotes in a Table format. Putting the results in the form of an effect size makes the result of the statistical test more clinically relevant and gives better context than citing P values alone.

If appropriate, please include number needed to treat for benefits (NNTb) or harm (NNTh). When comparing two procedures, please express the outcome of the comparison in U.S. dollar amounts.

Please standardize the presentation of your data throughout the manuscript submission. For P values, do not exceed three decimal places (for example, "P = .001"). For percentages, do not exceed one decimal place (for example, 11.1%").

Response 14: We described OR and 95% CI as appropriate in the revised manuscript.

15. Line 311: We discourage claims of first reports since they are often difficult to prove. How do you know this is the first report? If this is based on a systematic search of the

literature, that search should be described in the text (search engine, search terms, date range of search, and languages encompassed by the search). If on the other hand, it is not based on a systematic search but only on your level of awareness, it is not a claim we permit.

Response 15: Thank you for your comment. We deleted the words.

16. Please review the journal's Table Checklist to make sure that your tables conform to journal style. The Table Checklist is available online here: http://edmgr.ovid.com/ong/accounts/table_checklist.pdf.

Response 16: We have reviewed this checklist and updated out tables accordingly.

17. The American College of Obstetricians and Gynecologists' (ACOG) documents are frequently updated. These documents may be withdrawn and replaced with newer, revised versions. If you cite ACOG documents in your manuscript, be sure the reference you are citing is still current and available. If the reference you are citing has been updated (ie, replaced by a newer version), please ensure that the new version supports whatever statement you are making in your manuscript and then update your reference list accordingly (exceptions could include manuscripts that address items of historical interest). If the reference you are citing has been withdrawn with no clear replacement, please contact the editorial office for assistance (obgyn@greenjournal.org). In most cases, if an ACOG document has been withdrawn, it should not be referenced in your manuscript (exceptions could include manuscripts that address items of historical interest). All ACOG documents (eg, Committee Opinions
and Practice Bulletins) may be found via the Clinical Guidance & Publications page at https://www.acog.org/Clinical-Guidance-and-Publications/Search-Clinical-Guidance.

Response 17: We understand.

18. Figure

Figure 1: This file may be resubmitted as-is with the revision.
Figure 2: Please upload a new version using solid colors as bars. Per journal style, we avoid using patterns within graph bars.

Response 18: We uploaded a new version using solid colors as bars.

19. Authors whose manuscripts have been accepted for publication have the option to pay an

article processing charge and publish open access. With this choice, articles are made freely available online immediately upon publication. An information sheet is available at http://links.lww.com/LWW-ES/A48. The cost for publishing an article as open access can be found at http://edmgr.ovid.com/acd/accounts/ifauth.htm.

Please note that if your article is accepted, you will receive an email from the editorial office asking you to choose a publication route (traditional or open access). Please keep an eye out for that future email and be sure to respond to it promptly.

Response 19: We understand.

20. If you choose to revise your manuscript, please submit your revision through Editorial Manager at http://ong.editorialmanager.com. Your manuscript should be uploaded in a word processing format such as Microsoft Word. Your revision's cover letter should include the following:

* A confirmation that you have read the Instructions for Authors (http://edmgr.ovid.com/ong/accounts/authors.pdf), and

* A point-by-point response to each of the received comments in this letter.

If you submit a revision, we will assume that it has been developed in consultation with your co-authors and that each author has given approval to the final form of the revision.

Response 20: Yes, revision of the manuscript has been approved by all co-authors.

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	3
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	3-4
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4-5
Objectives	3	State specific objectives, including any prespecified hypotheses	5
Methods			
Study design	4	Present key elements of study design early in the paper	6

Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	6-7
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	6-7
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case	6-7, 9-11
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	7-9
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	*6-8*
Bias	9	Describe any efforts to address potential sources of bias	6
Study size	10	Explain how the study size was arrived at	6
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	9-10
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	9-10
		(*b*) Describe any methods used to examine subgroups and interactions	9-10
		(*c*) Explain how missing data were addressed	6
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy	11-12
		(*e*) Describe any sensitivity analyses	

We now hope that our paper will be suitable for publication in *OBSTETRICS & GYNECOLOGY* and look forward to hearing from you concerning your editorial decision.

Sincerely,

Tomomi Kotani, M.D., Ph.D. Associate Professor