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: Apr 17, 2020
To: "Stephanie A. Leonard" [redacted]
From: "The Green Journal" em@greenjournal.org
Subject: Your Submission ONG-20-592

RE: Manuscript Number ONG-20-592

An Expanded Obstetric Comorbidity Scoring System for Predicting Severe Maternal Morbidity

Dear Dr. Leonard:

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).

Due to the COVID-19 pandemic, your paper will be maintained in active status for 30 days from the date of this letter. If we have not heard from you by May 17, 2020, we will assume you wish to withdraw the manuscript from further consideration.

REVIEWER COMMENTS:

Reviewer #1: This study developed and internally and externally validated an obstetrics comorbidity index against CDC-defined severe maternal morbidity (including and excluding transfusion-only cases). The targeted causal inference approach is interesting and novel to this area. Presenting both internally and externally validated data was appropriate and rigorous.

1. While the statistical approach (targeted causal inference approach, precision-recall calibration) is exciting, the description in methods is quite technical and can be improved to help a largely clinical audience understand the method.
2. Related, the Discussion should include limitations of the targeted causal inference approach
3. Missing supplemental Table 3
4. Why do you think that the calibration curve levels off (more prominent in the non-SMM outcome group) >50%? Figure 3 - fit
5. Why is the precision recall AUC from the morbidity index from the Easter paper so much lower than that of the morbidity index proposed in this paper? Table 3
6. Minor: Methods - Spell out TRIPOD and add related check list if possible.

Reviewer #2: Objective is well stated and consistent:To develop and validate an expanded obstetric comorbidity score for predicting severe maternal morbidity (SMM) that can be applied consistently across contemporary U.S. patient discharge datasets.
Introduction is well written
Materials and Methods well defined
Results: well defined

Discussion
Within in any comorbidity is a range of severity, example an percreta that involves the bladder is much more severe than an accreta, maternal age of 45-50 is clinically more worrisome than 35
A 3 medication hypertension vs a small amount of Labetalol used for c hypertenson.
Is there any way to control for severity within a comorbidity group?
Is thee room for a modifier for severity? Please discuss

Reviewer #3: The authors developed and validated obstetric comorbidity scoring system to predict the outcomes of severe

maternal morbidity during hospitalization for childbirth. Overall, the authors conducted a well-designed study (given the limitations of the dataset used). The manuscript is written well and easy to read. However, the study suffers from one major limitation with reduced my enthusiasm.
A major limitation is the lack of a subset in whom the predictive value of the scoring system was validated by chart review. As described by the authors in the limitation section, ICD code based databases do have limitations and projects like this should consider validating predictive capacity of the scale in clinical data (using chart review).

Minor concerns
1. The authors stated that anemia complicating childbirth and cardiomyopathy may be ambiguous in identifying preexisting conditions versus complications arising during childbirth. They found 92% of anemia complicating childbirth and 0.3% of cardiomyopathy codes were reported using present-on-admission codes. Cardiomyopathy is a broad category. The authors may consider assessing the percent of each categories of cardiomyopathy (e.g. peripartum cardiomyopathy and alcoholic cardiomyopathy) reported as present on admission.
2. Grouping comorbidities based on rates? For example HELPP syndrome and preeclampsia with severe features, congenital and acquired cardiac diseases, all anemias, and all substance use disorders. I am not sure whether similar rate is sufficient reason to combine these conditions together. If there is statistical support for this action, the authors may consider explaining that.
3. Adjusted for educational attainment and expected method of payment for delivery. How about race? Race is a major factor. For example, sickle cell disease is a problem among African Americans.

STATISTICAL EDITOR COMMENTS:

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

General: Need to provide (could be on-line supplemental) a more complete description of how the variables were weighted to yield the scores. That is, what were the coefficients and constant values for the logistic equations derived from the training set of data?

lines 242-244: These analyses provided true external validation (the previous validation was a random sample from the same data set that was used for model construction). So, these should be made available to the reader, either in primary text or on-line supplemental. SHould also include the calibration curve corresponding to the Optum data.

Table 2: Although the aRR and aRD are useful metrics for relative risk, should also include the proportion of women with the various factors listed, so that the absolute risk differences can be put in context.

Fig 2: Could include this figure in supplemental material.

Fig 3: Need to change these calibration curves to conform to the TRIPOD guidelines. See "Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): Explanation and Elaboration" by K.G.M. Moons, D.G. Altman, J.B. Reitsma, J.P.A. Ionnidis, P. Macaskill, E.W. Steyerberg, A.J. Vickers, D. F. Ransohoff and G. S. Collins, Annals of Internal Medicine 2015:162:W1-W73. The observed scores should be divided into equal counts of 10 or 20 groups, which are plotted as their predicted vs observed means, with CIs for the observed proportion with SMM or SMM without transfusions. This format will convey to the reader the strength of association at various model scores, along with their relative uncertainty. The format used now in Fig 3 does not convey enough the information implied by the density plot shown along the present x-axis. That is, the least certainty is available at higher scores.

MANUSCRIPT EDITOR COMMENTS:

1. Label each supplemental item an "Appendix," number them consecutively (no matter if the item is text, figure, or a table), and cite them in order at first mention in the manuscript.

EDITOR 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. 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.

4. 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.

5. 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).

6. 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; Case Reports, 125 words; Current Commentary articles, 250 words; Executive Summaries, Consensus Statements, and Guidelines, 250 words; Clinical Practice and Quality, 300 words; Procedures and Instruments, 200 words. Please provide a word count.

7. 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.

8. 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.

9. 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%").

10. A sentence in your Discussion reads, "Our study is also the first, to our knowledge, to use ICD-10-CM diagnosis and procedure codes to study comorbidities and SMM in the U.S." 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.

11. 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.

12. Figures should be saved as high-resolution TIFF files. The minimum requirements for resolution are 300 dpi for color or black and white photographs, and 600 dpi for images containing a photograph with text labeling or thin lines.

Art that is low resolution, digitized, adapted from slides, or downloaded from the Internet may not reproduce.

13. To ensure a quality experience for those viewing supplemental digital content, the journal's publisher suggests that authors submit supplemental digital files no larger than 10 MB each. The exceptions to this rule are audio or video files, which are acceptable up to 100 MB. When submitting text files or tables as supplemental digital content with your revisions, please do not submit PDFs.

14. 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.

***

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 30 days from the date of this letter. If we have not heard from you by May 17, 2020, we will assume you wish to withdraw the manuscript from further consideration..

Sincerely,

The Editors of Obstetrics & Gynecology

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

_____

Stanford MEDICINE

June 1, 2020

Editorial Board
Obstetrics & Gynecology

Dear Editorial Board:

We appreciate the feedback provided by reviewers and editors on our manuscript, "An expanded obstetric comorbidity scoring system for predicting severe maternal morbidity." We have addressed all of the comments, as detailed in the attached pages. All changes to the submitted manuscript are shown with tracked changes. We believe these revisions have led to a stronger manuscript, which we hope is acceptable for publication in *Obstetrics & Gynecology*.

Sincerely,

Stephanie A. Leonard

RESPONSES TO REVIEWER AND EDITOR COMMENTS

Reviewer #1: This study developed and internally and externally validated an obstetrics comorbidity index against CDC-defined severe maternal morbidity (including and excluding transfusion-only cases). The targeted causal inference approach is interesting and novel to this area. Presenting both internally and externally validated data was appropriate and rigorous.

1. While the statistical approach (targeted causal inference approach, precision-recall calibration) is exciting, the description in methods is quite technical and can be improved to help a largely clinical audience understand the method.

Response: We appreciate this input and have revised the Methods section to be less technical. We have added an Appendix (Appendix 3) to the Online Supplemental Material with the technical details. The appendix is referenced in the Methods at the end of the section "Development of Comorbidity Index": "More details on the development of the comorbidity index, including sensitivity analyses, are available in Appendices 3-4."

2. Related, the Discussion should include limitations of the targeted causal inference approach

Response: We have expanded in the Discussion on the limitations of the targeted causal inference approach. The section now includes the following: "The study analyses were limited in part by the very large sample size and number of comorbidities considered, which prevented incorporation of a wider variety of machine-learning algorithms than those we considered in a sensitivity analysis for these reasons. The analytical approach also required dichotomization of the comorbidities, which prevented assigning different scores for different degrees of severity for a given comorbidity."

3. Missing supplemental Table 3

Response: We greatly appreciate the Reviewer catching this error and have updated the Online Supplemental Material and ensured all tables and figures are included.

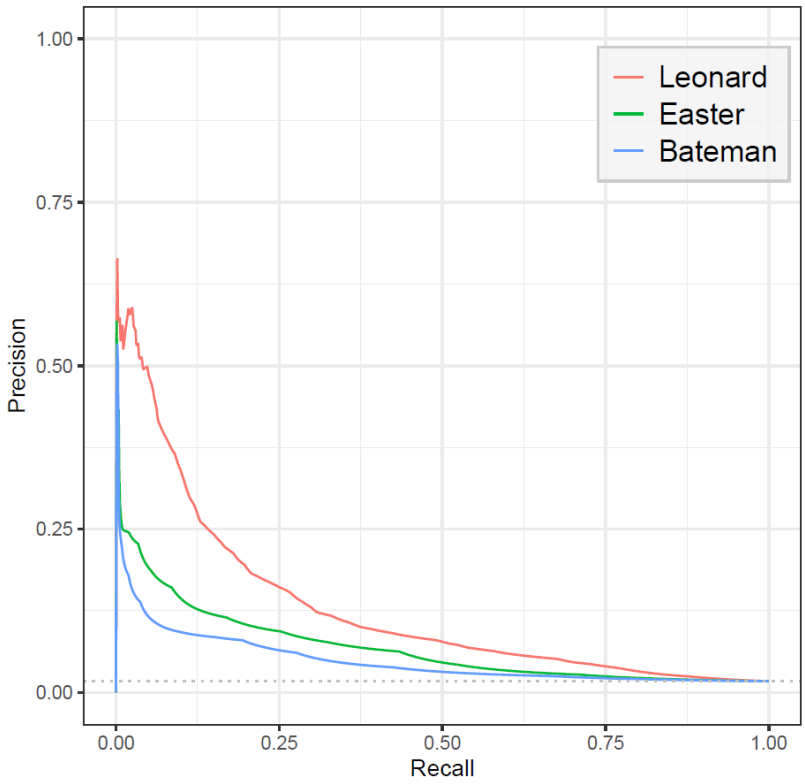
4. Why do you think that the calibration curve levels off (more prominent in the non-SMM outcome group) >50%? Figure 3 – fit

Response: We believe the limited number of subjects with high predicted risk of SMM (given the relative rarity of SMM, and non-transfusion SMM even more so) and the additive nature of the score together cause some information to be lost, causing the calibration curve to level off at high predicted risks. There is likely heterogeneity among these very high-risk patients, such as some having one or two very high-score comorbidities (e.g., accreta, pulmonary hypertension) and others having multiple low-score comorbidities. We had already discussed as a team that this will be one of our next areas of work to better understand comorbidity patterns in these very high-risk patients.

5. Why is the precision recall AUC from the morbidity index from the Easter paper so much lower than that of the morbidity index proposed in this paper? Table 3

Response: Below we show a plot with the precision-recall curves stacked for the 3 comorbidity index versions. (We have added the plot and explanatory text to the manuscript – Figure 3). It can be seen that our index has a higher level of precision (positive predictive value) for any level of recall (true positive rate). Notably, we have substantially higher precision in the 0-20% recall range. This shows that if we look at each score separately and select the highest score-specific threshold that captures 10-20% of

patients with SMM (i.e. recall is 10 – 20%), our score will result in a higher percentage of those selected patients having SMM (i.e. precision). Whereas the other two scores effectively need a bigger net (lower relative threshold) to capture that many positive cases (recall), which results in them less efficiently also flagging a larger number of negative cases (lower precision).



6. Minor: Methods - Spell out TRIPOD and add related check list if possible.

Response: We have spelled out TRIPOD (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis) and included a reference to the guidelines.

Reviewer #2: Objective is well stated and consistent:To develop and validate an expanded obstetric comorbidity score for predicting severe maternal morbidity (SMM) that can be applied consistently across contemporary U.S. patient discharge datasets.
Introduction is well written
Materials and Methods well defined
Results: well defined

Discussion
Within in any comorbidity is a range of severity, example an percreta that involves the bladder is much more severe than an accreta, maternal age of 45-50 is clinically more worrisome than 35 A 3 medication hypertension vs a small amount of Labetalol used for c hypertenson.
Is there any way to control for severity within a comorbidity group?
Is thee room for a modifier for severity? Please discuss

Response: We very much appreciate and understand the Reviewer's concern about differences in severity of some conditions that were grouped together. We have now added the following sentences to the

limitations in the Discussion section: “Further, vital signs, laboratory values, medications, and severity of a condition are generally not available in patient discharge data but can be informative for predicting obstetric complications… The analytical approach also required dichotomization of the comorbidities, which prevented assigning different scores for different degrees of severity for a given comorbidity.”

Reviewer #3: The authors developed and validated obstetric comorbidity scoring system to predict the outcomes of severe maternal morbidity during hospitalization for childbirth. Overall, the authors conducted a well-designed study (given the limitations of the dataset used). The manuscript is written well and easy to read. However, the study suffers from one major limitation with reduced my enthusiasm. A major limitation is the lack of a subset in whom the predictive value of the scoring system was validated by chart review. As described by the authors in the limitation section, ICD code based databases do have limitations and projects like this should consider validating predictive capacity of the scale in clinical data (using chart review).

Response: We very much appreciate the Reviewer’s comment and suggestion. We are planning a follow-up paper to focus on clinical validation and utility, which we felt was beyond the scope of this initial paper. We have also expanded the limitations section of the Discussion, which describes the lower accuracy and detail of the data compared with chart review.

Minor concerns

1. The authors stated that anemia complicating childbirth and cardiomyopathy may be ambiguous in identifying preexisting conditions versus complications arising during childbirth. They found 92% of anemia complicating childbirth and 0.3% of cardiomyopathy codes were reported using present-on-admission codes. Cardiomyopathy is a broad category. The authors may consider assessing the percent of each categories of cardiomyopathy (e.g. peripartum cardiomyopathy and alcoholic cardiomyopathy) reported as present on admission.

Response: Cardiomyopathy is indeed a broad category and we very much agree that a future area of interest for us would be studying differences in relation to SMM outcomes. A real world limitation is that most types of cardiomyopathy are silent until delivery admission (e.g., alcoholic and drug) or not determined until autopsy. Data limitations are that limited ICD-10-CM diagnosis codes (in particular, there is not a specific code for peripartum cardiomyopathy) and very small numbers due to the rarity of these conditions.

2. Grouping comorbidities based on rates? For example HELPP syndrome and preeclampsia with severe features, congenital and acquired cardiac diseases, all anemias, and all substance use disorders. I am not sure whether similar rate is sufficient reason to combine these conditions together. If there is statistical support for this action, the authors may consider explaining that.

Response: The primary rationale for the groupings was that we had some related comorbidities with similar disease states and low frequencies so combining them when they had similar rates of SMM enabled us to increase sample size, as well as keep the number of total potential predictors reasonable (i.e., under 30). Statistical estimate becomes inconsistent when sample size is too low.

3. Adjusted for educational attainment and expected method of payment for delivery. How about race? Race is a major factor. For example, sickle cell disease is a problem among African Americans.

Response: We gave considerable thought about the inclusion of race/ethnicity and decided not to adjust for it because one of our goals is to create a tool that could be used to evaluate racial/ethnic disparities in SMM independent of differences in comorbidities. We believe if we control for race/ethnicity in the creation of the comorbidity score, it will adjust away such important differences. We thank the Reviewer

for their comment as we then discussed further and believe this rationale applies to socioeconomic disparities as well. We therefore decided to remove any demographic factors from the main analysis. We then replicated the analyses including educational attainment, expected method of payment, and race/ethnicity as confounders and housing or economic instability as a predictor (because it has an ICD-10-CM code). This sensitivity analysis is described in the revised submission (Appendix 3) and results are provided (Appendix 4). We also now note in the Discussion (page 11): "Inclusion of available sociodemographic factors in a sensitivity analysis (Appendix 4) did not meaningfully change results, which warrants further investigation."

STATISTICAL EDITOR COMMENTS:

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

1. General: Need to provide (could be on-line supplemental) a more complete description of how the variables were weighted to yield the scores. That is, what were the coefficients and constant values for the logistic equations derived from the training set of data?

Response: We agree with the Editor's suggestion and have added this description to the supplemental material in Appendix 3. Per Reviewer 1's suggestion, we also moved some of the original Methods section to this appendix as well. The appendix is referenced in the Methods at the end of the section "Development of Comorbidity Index": "More details on the development of the comorbidity index, including sensitivity analyses, are available in Appendices 3-4."

2. lines 242-244: These analyses provided true external validation (the previous validation was a random sample from the same data set that was used for model construction). So, these should be made available to the reader, either in primary text or on-line supplemental. SHould also include the calibration curve corresponding to the Optum data.

Response: We agree and have made the following additions: (1) Index discrimination statistics in Optum data added to Table 3; (2) calibration curves in Optum data added as Appendix 7.

3. Table 2: Although the aRR and aRD are useful metrics for relative risk, should also include the proportion of women with the various factors listed, so that the absolute risk differences can be put in context.

Response: The prevalence of each comorbidity is presented in Table 1. If the Editor feels these numbers would be more suitable in Table 2, we are willing to make that change.

4. Fig 2: Could include this figure in supplemental material.

Response: We have moved Figure 2 (precision-recall ROC curves) to the supplemental material (Appendix 5). We instead created a Figure for the main manuscript (Figure 3) with the PR-ROC curve for SMM for the 3 indices being compared, which we believe is more useful to the reader than the original figure.

5. Fig 3: Need to change these calibration curves to conform to the TRIPOD guidelines. See "Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): Explanation and Elaboration" by K.G.M. Moons, D.G. Altman, J.B. Reitsma, J.P.A. Ionnidis, P. Macaskill, E.W. Steyerberg, A.J. Vickers, D. F. Ransohoff and G. S. Collins, Annals of Internal

Medicine 2015:162:W1-W73. The observed scores should be divided into equal counts of 10 or 20 groups, which are plotted as their predicted vs observed means, with CIs for the observed proportion with SMM or SMM without transfusions. This format will convey to the reader the strength of association at various model scores, along with their relative uncertainty. The format used now in Fig 3 does not convey enough the information implied by the density plot shown along the present x-axis. That is, the least certainty is available at higher scores.

Response: We have updated the manuscript in response to the Editor's comment. The TRIPOD guidelines does suggest the proposed plot as one approach to assess calibration. It additionally states that calibration plots could be "smoothed or by subgroups" (p. 34) and we opted to do a smoothed version in our manuscript. We agree that some readers could appreciate the grouped confidence intervals version so we have created these plots and added them to the supplemental material (Appendix 6). We prefer the smoothed version in the main manuscript because our sample is so concentrated in the very low probability region and it is challenging to interpret the confidence intervals because we used a log transformation. We added to the main manuscript on page 9: "Grouped calibration plots with confidence intervals were also plotted and are provided in Appendix 6."

MANUSCRIPT EDITOR COMMENTS:

1. Label each supplemental item an "Appendix," number them consecutively (no matter if the item is text, figure, or a table), and cite them in order at first mention in the manuscript.

Response: We have made this change.