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 19, 2019
To: "Floriane JOCHUM" [redacted]
From: "The Green Journal" em@greenjournal.org
Subject: Your Submission ONG-19-333

RE: Manuscript Number ONG-19-333

Externally validated score of labor induction with cervical ripening from a prospective cohort study

Dear Dr. JOCHUM:

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 May 10, 2019, we will assume you wish to withdraw the manuscript from further consideration.

REVIEWER COMMENTS:

Reviewer #1: This is a secondary analysis of a labor induction data base from 94 French maternity The objective of the study was to build a scoring system for determining the C/S risk after IOL with cervical ripening agent. Several factors were determined to have a strong impact on the predictive nature of the scoring system to include; Height, BMI, EGA, parity, dilation, effacement, station, macrosomia, PROM and FHR status. The model was validated using part of their database and an external database as well. The authors conclude that they have developed an easy-to-use scoring system to determine the risk of C/S with labor induction with cervical ripening. Ways in which this manuscript could be improved include:

1. Line 58: Is grammatically awkward, I would rewrite.

2. Lines 57-69: The methods section of the abstract needs to be condensed, it is far too long.

3. Lines 159-160: Why is this? I would elaborate or list reference.

4. Lines 172-173: Why did you stratify this way? Why not using previous SVD or not? Or use parity as continuous variable?

5. Lines 184-186: Why did you make this split in your database? What was the method for determining this split?

6. Lines 208-210: Has this method been described before? Or was there a method to determine that 6 was the best multiplier?

7. Lines 346-347: Have other studies split parity the way your study chose to? Or is there variation?

8. Line 353-354: I would expand this to include other etiology for higher cesarean rate? Perhaps postdates is predictive of uterine dysfunction?

9. Line 363-365: Have you, or do you plan to, make this available via the internet or app?

Reviewer #2: Precis: Appropriate in length, succinct yet thorough

Abstract: Abstract is concise and easy to read

Introduction

1* The introduction appropriately frames the context of the research

Material and Methods

2* appropriate, thorough

Results

3* no comment

Discussion

4* The discussion highlights the strengths and weaknesses of the study well.

5* I think it is an excellent study that corroborates and improves upon current literature. Research/publication on this area of predicting cesareans rates during induction is relatively new and not common so this is an excellent contribution to the field

Tables and Figures

6* Figure 1 is a nice summary of the study design

References

7* References appear contemporary and appropriate for study

Reviewer #3: This manuscript is a secondary analysis of a prospective multicenter, observational, population-based cohort study of labor induction practices in France, known as the MEDIP study. The authors use this large induction data set to first derive, and then internally validate a model for prediction of cesarean delivery regardless of indication, using Bayesian methods to apply assumptions from prior literature to the model. They apply this model to a publicly available external data set, and compared the discriminative power of the model when applied to derivative, internal, and external validation cohorts with and without the use of priors (sensitivity analysis). By comparing relative AUCs in the current study with those from prior publishes works (Levine, et al and the traditional Bishop score), the investigators conclude that their externally-validated predictive model using a 50-point score is better at predicting cesarean delivery compared to the modified Bishop or Levine scores.

There are several questions about the study methods and analysis plan that need to be addressed prior to publication. Please refer to and use the methods and checklist described in the TRIPOD statement for reporting a prediction model, and please reference this in the methods of the study.

Ref: Collins GS, Reitsma JB, Altman DG, Moons KG, TRIPOD Group. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. The TRIPOD Group. Circulation. 2015;131(2):211-219. doi:10.1161/CIRCULATIONAHA.114.014508

Additionally:
Title: The current title says nothing about predicting cesarean delivery, and I would suggest rewording to convey this.

Abstract: This is overall a faithful summary of the manuscript. Several comments below regarding manuscript text, study design and analysis could be incorporated into the abstract. Additionally:

1. Objective: Both primary and secondary objectives could be addressed succinctly here.

2. Methods: Consider including a short explanation of the rationale for Bayesian approach in the abstract. Also consider a sentence that describes comparing the discriminative power of the model applied to internal and external validation cohorts by comparing area under the receiver operating characteristic curve. Subsequently include the use of priors in the Bayesian approach for the sensitivity analysis.

3. Results: consider moving the sentence about the 50-point scoring system to the methods section

Introduction:

1. Lines 118-119. Why do the authors propose that no score has emerged as standard of care for prediction of cesarean delivery following induction with cervical ripening?

Methods

1. Lines 143-144. Induction without cervical ripening is an intervention, but Bishop score is a characteristic, so they are not mutually exclusive. Were there any patients with Bishop <= 6 who were induced without cervical ripening? If so, are these patients included in the currently study?

2. Line 147-148. Was there an established practice for cervical ripening in all patients with Bishop Score <=6 in the original study? If so, please describe the guidelines or induction protocol followed.

3. Line 154-155. The authors refer to a prospective, multicenter, observational cohort MEDIP study and describe the eligible participants used to include in the current study, but the study referenced "for details of the protocols" is a survey of administrative or hospital leaders on delivery room practices. Does a published protocol exist for how the prospective data were collected, and could the authors reference this or provide more details in the current manuscript?

4. Lines 170, 171, etc. Use of the term "discretized" is somewhat unfamiliar to a general audience.

5. Statistical methods:
a. The authors use a Bayesian approach for analysis, which may be the primary difference when compared to past predictive models published (Levine, et al). Though I'm not an expert in Bayesian statistics, my understanding of the approach is that it allows for integration of prior information with newly obtained data for a "final quantitative summary" of the relationship between a variable and outcome. The approach may be favored when outcomes are rare, and when prior information facilitates a more efficient method to reach conclusions without having to enroll more patients. However, in the current study, outcomes (cesarean delivery) are not rare, and the statistical approach does not appear to contribute meaningful insight to the research objectives. Using the priors does not affect the performance of the model. Could the authors provide justification for using the Bayesian rather than traditional, frequentist approach (with regular old p values), given that most general readers will not take the time to learn or understand its use and limitations?
b. As stated above, please reference the TRIPOD statement/checklist and adhere to the proposed guidelines
c. Lines 208-209. Why was the regression coefficient multiplied by 6?

6. Lines 217-225. Is the information about how the external validation cohort was selected necessary?

Discussion:

1. I would suggest the authors address the use of the current model compared to recent, similar published scoring systems for prediction of cesarean section (Levine, et al). What is the difference in these scoring systems other than use of Bayesian statistics?

2. Line 313. Please address the strengths and limitations of using of Bayesian statistics in your approach.

3. Line 334. Isn't there always a risk of cesarean section when embarking on labor induction, or even during spontaneous labor?

4. Lines 339-360. This section seems to fit better with the section entitled "Clinical Implication" which provides context for the study findings. Rather than summarizing the study findings again and then acknowledging individual parameters that have already been published, could the authors provide a more in-depth assessment of how the current study fills in gaps in the existing literature?

5. How will prospective use of this model not become a self-fulfilling prophecy, whereby a high score subsequently affects the decision to proceed with cesarean delivery, which is the outcome predicted by the score?

Reviewer #4: The article analyzes two large dataset and tries to validate a new score for the prediction of the CS after induction of labor. The population includes only women with need of cervical ripening according to a previous application of the Bishop score.

From the methodological point of view I have some doubt regarding the use of a score after the use of another score, in this case of the Bishop. Of interest the new score includes cervical dilation, cervical consistency, cervical effacement, cervical position and fetal head station, that are all the items of the Bishop score. Three of these variables were included in the final model.

The second point regards the clinical utility of the score. Like for VBAC and admission to vaginal delivery, which is the cut off that we should use to renounce to induction? If the risk is 35-50% an elective CS could substitute the induction? Moreover only a very few number of patients could have a score > 40. I.e. a nullipara, > 41 weeks, with PROM, non reassuring FHR (anomaly of the fetal vitality is confusing), BMI>35, height < 160, cervix closed, no effacement at all, head mobile has a score of 40 and has anyway a probability of vaginal delivery of 50% after induction. These numbers should/can reassure or confuse the obstetrician?

We should consider the effect of the practical application of this score as more relevant that the performance of the score itself, also because the only modifiable variable is the condition of the cervix that is related to advancing GA, that is on the contrary pejorative. Differently from the VBAC the risks of the admission at induction ending in CS are less explored, also

in the prospective of future pregnancies and delivery if the present delivery will be a CS, and from the point of view of the neonate. If we renounce to an indicated induction at 37 weeks for example we should administrate corticosteroids and completely change the condition of the neonate? These points should be discussed in the paper.

STATISTICAL EDITOR'S COMMENTS:

1. lines 189-190: Should enumerate (could be on-line material) the n(%) of missing data for each of the variables. Should similarly cite the missing data for each variable in the internal validation set and the n(%) of complete data for that portion of the data. Also, within both the derivation and the validation data sets, were proportion of patients with complete data statistically equivalent for both those with cesarean and vaginal deliveries?

2. Were the samples n = 1024 and n = 668 the total samples or the samples with complete data? Need to clarify.

3. Fig 3 and Appendix 4 give an incomplete summary of the utility of the model. Should emulate the format in TRIPOD, with an accounting of what proportion of the derivation data set (presumably those with only complete data) comprised each of the 8 subsets represented by Fig 3 (or some other suitable stratification), each with the observed vs the expected probability for that stratum and with the CI for that stratum.

4. Ref:
"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.

5. This reference includes a calibration curve (fig 8), which allows the reader to see the relationship of observed vs predicted probabilities along the spectrum of probabilities from the data, along with confidence intervals for those prediction estimates. An added feature is the display below the x-axis of the relative counts of adverse vs non-adverse outcomes. Alternatively, those could be displayed (similar to survival analysis graphs), with numerical counts of adverse vs non-adverse outcomes at the intervals referred by the graph. The advantage to this level of detail is that it would convey to the reader the strength of association at various model scores, along with their relative uncertainty, reflected by the amount of data available at various cut-points.

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, as well as subsequent author queries. 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:
1. OPT-IN: Yes, please publish my response letter and subsequent email correspondence related to author queries.
2. OPT-OUT: No, please do not publish my response letter and subsequent email correspondence related to author queries.

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.

Any author agreement forms previously submitted will be superseded by the eCTA. During the resubmission process, you are welcome to remove these PDFs from EM. However, if you prefer, we can remove them for you after submission.

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. Clinical trials submitted to the journal as of July 1, 2018, must include a data sharing statement. The statement should indicate 1) whether individual deidentified participant data (including data dictionaries) will be shared; 2) what data in particular will be shared; 3) whether additional, related documents will be available (eg, study protocol, statistical analysis plan, etc.); 4) when the data will become available and for how long; and 5) by what access criteria data will be shared

(including with whom, for what types of analyses, and by what mechanism). Responses to the five bullet points should be provided in a box at the end of the Methods section.

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.

5. 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.

6. 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.

7. 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. Please provide a word count.

8. 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.

9. 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.

10. 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.

11. 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 via Editorial Manager for Obstetrics & Gynecology at http://ong.editorialmanager.com. It is essential that your cover letter list point-by-point the changes made in response to each criticism. Also, please save and submit your manuscript in a word processing format such as Microsoft Word.

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 May 10, 2019, we will assume you wish to withdraw the manuscript from further consideration.

Sincerely,

The Editors of Obstetrics & Gynecology

2017 IMPACT FACTOR: 4.982

2017 IMPACT FACTOR RANKING: 5th out of 82 ob/gyn journals

_____

May 10, 2019
Editorial Office – Obstetrics and Gynecology

Dear Editor, Dear reviewers,

We would like to sincerely thank you for your precious comment.

The manuscript was extensively revised along with reviewers's comment. Please find below an answer item by item. We highlighted the changes in our document.

I, Floriane Jochum, affirm 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 registered have been explained.

We hope this version will be considered of scientific interest to be published in Obstetrics and Gynecology.

Sincerely yours

Floriane Jochum, on behalf of all the authors.

REVIEWER COMMENTS:

Reviewer #1: This is a secondary analysis of a labor induction data base from 94 French maternity The objective of the study was to build a scoring system for determining the C/S risk after IOL with cervical ripening agent. Several factors were determined to have a strong impact on the predictive nature of the scoring system to include; Height, BMI, EGA, parity, dilation, effacement, station, macrosomia, PROM and FHR status. The model was validated using part of their database and an external database as well. The authors conclude that they have developed an easy-to-use scoring system to determine the risk of C/S with labor induction with cervical ripening. Ways in which this manuscript could be improved include:

1. Line 58: Is grammatically awkward, I would rewrite.

We modified this sentence:

« This study is a secondary analysis of data collected in the prospective multicenter observational French population-based cohort study MEDIP, the primary objective of which was to obtain national data regarding labor induction practices »

2. Lines 57-69: The methods section of the abstract needs to be condensed, it is far too long.

We shortened the abstract, even if we provided some more details regarding the Bayesian approach as requested by the reviewer #3.

3. Lines 159-160: Why is this? I would elaborate or list reference.

We understand that law is different in other countries than France… The French law of 1978 "Informatique et libertés" provides that "it is prohibited to collect or process personal data which reveal, directly or indirectly, racial or ethnic origins, political, philosophical or religious opinions or trade union membership of persons". We have added a reference of this law in the manuscript.

4. Lines 172-173: Why did you stratify this way? Why not using previous SVD or not? Or use parity as continuous variable?

Thank you for your comment. First, just to be noted that we excluded patients with previous c-section, so the parity only relates to vaginal deliveries. As described in the TRIPOD statement, a linear functional relationship is the most popular approach for keeping the continuous nature of a predictor, but the log linerarity must be checked or it will conduct to a missspecified model. Keeping the parity as a continuous variable was not possible in our study, because the log linearity was not verified.

Categorization allows to avoid strong assumptions about the relationship between the predictor and outcome. However, this comes at the expense of throwing away information. The information loss is obviously greatest when the predictor is dichotomized (2 categories).

Categorizing a continuous variable into 3 or more groups reduces the loss of information. After checked the number of patients in each group, we decided to divide parity into 3 groups.

	Numbers	Proportions
0	647	63.18 [60.15 ; 66.14]
1	238	23.24 [20.69 ; 25.95]
2	90	8.79 [7.13 ; 10.69]
3	33	3.22 [2.23 ; 4.5]
4	12	1.17 [0.61 ; 2.04]
5	3	0.29 [0.06 ; 0.85]
6	1	0.1 [0 ; 0.54]
Total	1024	100

	Effectifs	Proportions
0	647	63.18 [60.15 ; 66.14]
1	238	23.24 [20.69 ; 25.95]
2	139	13.57 [11.53 ; 15.83]
Total	1024	100

5. Lines 184-186: Why did you make this split in your database? What was the method for determining this split?

To our knowledge, the best method for splitting a database is not very well codified. Most of the time, a 70/30% division is performed, but as we had a large database, we decided to perform a 60/40% division. This has allowed us to have a training set of over 1000 patients with still a large internal validation set.

6. Lines 208-210: Has this method been described before? Or was there a method to determine that 6 was the best multiplier?

Thank you for this question. We created our score by following the rules for creating a weighted score using the beta coefficients (Hemalkumar et al. 2016). One of the usual ways is to divide every beta coefficients by the smallest, but this would have created a score that would have been too difficult to use. As it is described, we multiplied the coefficients by the same number in order

to have a score that is easy to use. Multiplying by 6 allowed us to have a score out of 50, and therefore an easy to use score.

Regression coefficient based scoring system should be used to assign weights to the risk index
Hemalkumar B. Mehtaa, Vinay Mehtab,*, Cynthia J. Girmanc, Deepak Adhikaria, Michael L. Johnsond
J Clin Epidemiol. 2016 Nov;79:22-28. doi: 10.1016/j.jclinepi.2016.03.031. Epub 2016 May 13

We rephrased the explanation section and added the reference for the method:
"To create a regression coefficient-based scoring system, the beta coefficient were multiplied by a unique number (6) and then rounded off in order to derive weights, as previously described [27]. Therefore we obtained a simplified score, easy to calculate, out of 50 points."

7. Lines 346-347: Have other studies split parity the way your study chose to? Or is there variation?

The parity variable has been used in very different ways in previous published studies. Several studies have used it as a continuous variable, but as explained above, log linearity was not verified in our study. Still, to avoid losing too much data, we decided to stratify it into 3 groups.

8. Line 353-354: I would expand this to include other etiology for higher cesarean rate? Perhaps postdates is predictive of uterine dysfunction?

Yes, we agree and we added what you suggested:
"Another explanation would be that the postdate might be predictive of uterine dysfunction".

9. Line 363-365: Have you, or do you plan to, make this available via the internet or app?

That is a great idea, we will definitely consider to build a dedicated app!
In the meantime, we believe that the simplified score filled in a sheet of paper is already pretty convenient.

Reviewer #2: Precis: Appropriate in length, succinct yet thorough

Abstract: Abstract is concise and easy to read

Introduction

1* The introduction appropriately frames the context of the research

Material and Methods

2* appropriate, thorough

Results

3* no comment

Discussion

4* The discussion highlights the strengths and weaknesses of the study well.

5* I think it is an excellent study that corroborates and improves upon current literature. Research/publication on this area of predicting cesareans rates during induction is relatively new and not common so this is an excellent contribution to the field

Tables and Figures

6* Figure 1 is a nice summary of the study design

References

7* References appear contemporary and appropriate for study

Thank you very much for your positive comment.

Reviewer #3: This manuscript is a secondary analysis of a prospective multicenter, observational, population-based cohort study of labor induction practices in France, known as the MEDIP study. The authors use this large induction data set to first derive, and then internally validate a model for prediction of cesarean delivery regardless of indication, using Bayesian methods to apply assumptions from prior literature to the model. They apply this model to a publicly available external data set, and compared the discriminative power of the model when applied to derivative, internal, and external validation cohorts with and without the use of priors (sensitivity analysis). By comparing relative AUCs in the current study with those from prior publishes works (Levine, et al and the traditional Bishop score), the investigators conclude that their externally-validated predictive model using a 50-point score is better at predicting cesarean delivery compared to the modified Bishop or Levine scores.

There are several questions about the study methods and analysis plan that need to be addressed prior to publication. Please refer to and use the methods and checklist described in the TRIPOD statement for reporting a prediction model, and please reference this in the methods of the study.

Ref: Collins GS, Reitsma JB, Altman DG, Moons KG, TRIPOD Group. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. The TRIPOD Group. Circulation. 2015;131(2):211-219. doi:10.1161/CIRCULATIONAHA.114.014508

Thank you, we followed every single point of the TRIPOD checklist and we added an attached document to our submission.

We also added the reference for the TRIPOD checklist as suggested:

“We followed the TRIPOD checklist for this work[16].”

Additionally:

Title: The current title says nothing about predicting cesarean delivery, and I would suggest rewording to convey this.

Yes, we agree and we changed the title accordingly:

“Externally validated score to predict cesarean delivery after labor induction with cervical ripening”

Abstract: This is overall a faithful summary of the manuscript. Several comments below regarding manuscript text, study design and analysis could be incorporated into the abstract. Additionally:

1. **Objective: Both primary and secondary objectives could be addressed succinctly here.**

We made the appropriate change: “To build a score to predict the risk of cesarean section after labor induction with cervical ripening, from a population-based dataset especially designed for induction analysis purpose, and to compare its predictive capacities with other already existing scores.”

2. **Methods: Consider including a short explanation of the rationale for Bayesian approach in the abstract. Also consider a sentence that describes comparing the discriminative power of the model applied to internal and external validation cohorts by comparing area under the receiver operating characteristic curve. Subsequently include the use of priors in the Bayesian approach for the sensitivity analysis.**

Thank you for your comment. We added a short explanation of the rationale of Bayesian approach and the use of priors: “Statistical analyses were performed using a Bayesian approach, allowing the use of priors i.e. previous results published in the literature.”

However, we did not add details on the comparison of ROC curves since the reviewer #1 suggested to shorten this section.

3. Results: consider moving the sentence about the 50-point scoring system to the methods section

We made the change.

Introduction:

1. Lines 118-119. Why do the authors propose that no score has emerged as standard of care for prediction of cesarean delivery following induction with cervical ripening?

Even if the previous published scores are of value, to our knowledge, none is considered as a consensual standard. Moreover, with the exception of the score published by Levine et al., none of these scores were especially dedicated to induction of labor with cervical ripening. We rephrased the sentence and added details on the specificity of cervical ripening:

"Several induction scores have been proposed to replace the Bishop score[9–15] but none has emerged as a consensual standard. Moreover, with the exception of the score published by Levine et al., none of these scores were especially dedicated to induction of labor with cervical ripening."

Methods

1. Lines 143-144. Induction without cervical ripening is an intervention, but Bishop score is a characteristic, so they are not mutually exclusive. Were there any patients with Bishop <= 6 who were induced without cervical ripening? If so, are these patients included in the currently study?

We excluded the cases of induction without cervical ripening and we believe it's one of the strength of our study: "Our study excluded all patients with twin pregnancies, non-cephalic presentations, previous cesarean section, premature deliveries, and induction without prior cervical ripening."

2. Line 147-148. Was there an established practice for cervical ripening in all patients

with Bishop Score <=6 in the original study? If so, please describe the guidelines or induction protocol followed.

Thank you for your comment, it's really noteworthy to provide details on this point. We added a sentence in the methods section: "To be noted that protocols for cervical ripening varied between the included maternities." Please, see also a comment in the discussion section: "Induction practices varied between maternity units but we did not find any significant difference between practices and the risk of cesarean section."

3. Line 154-155. The authors refer to a prospective, multicenter, observational cohort MEDIP study and describe the eligible participants used to include in the current study, but the study referenced "for details of the protocols" is a survey of administrative or hospital leaders on delivery room practices. Does a published protocol exist for how the prospective data were collected, and could the authors reference this or provide more details in the current manuscript?

Thank you for your comment. The original article including details on the protocol is currently in press (Blanc-Petitjean P et al., Gynecol Obstet Fertil Senol May 2019). We added this reference and we also added details on the protocol since it is not published yet. Clearly, it's a population based-cohort study (referenced on clinicaltrials.gov: NCT 02477085), not a survey!

"Briefly, MEDIP is a prospective multicenter observational French population-based cohort study. The data collected come from several sources, including data specific to the induction of labor completed prospectively by the physician or midwife in charge of the patient at the time of the procedure. To be noted that protocols for cervical ripening varied between the included maternities."

4. Lines 170, 171, etc. Use of the term "discretized" is somewhat unfamiliar to a general audience.

We modified this.

5. Statistical methods:

a. The authors use a Bayesian approach for analysis, which may be the primary difference when compared to past predictive models published (Levine, et al). Though I'm not an expert in Bayesian statistics, my understanding of the approach is that it allows for integration of prior information with newly obtained data for a "final quantitative summary" of the relationship between a variable and outcome. The approach may be favored when outcomes are rare, and when prior information facilitates a more efficient method to reach conclusions without having to enroll more patients. However, in the current study, outcomes (cesarean delivery) are not rare, and the statistical approach does not appear to contribute meaningful insight to the research objectives. Using the priors does not affect the performance of the model. Could the authors provide justification for using the Bayesian rather than traditional, frequentist approach (with regular old p values), given that most general readers will not take the time to learn or understand its use and limitations?

It’s true that that Bayesian statistics are especially relevant in case of small effectives but not only. Bayesian analyses are an appropriate alternative to the frequentist methods. Several papers tackled the comparison between Bayesian and frequentist methods in health data (for example, Dunson D. Commentary: practical advantages of Bayesian analysis of epidemiologic data. American Journal of Epidemiology 2001, 153(12): 1222-6.). Finally an increased number of published papers reports Bayesian results.

Hence, we choose to rely on Bayesian methods by using Markov chains with Monte Carlo integrations (McMC) to estimate posterior distribution for each coefficient. With those distributions, the probability for a coefficient to be strictly positive or strictly negative can be evaluated, contrary to point estimates that are not informative on the complete distribution. Furthermore, no assumption is required on the probability for a statistic to be greater than a theoretical quantile, i.e. no p-value is retrieved. Results are presented by describing posterior distributions, and conclusions are formulated in terms of probability for the coefficient to be strictly positive. If this

probability is close to 1 (resp. close to 0), we'll conclude that the probability of the tested hypothesis is very high (resp. very small). If this probability is close to 50%, the distribution is half negative, half positive, and we'll conclude that the coefficient is not different from 0. Conclusions are then more informative than simple p-values.

For seek of clarity, we add a paragraph in the methods section:

"Bayesian statistics give probabilistic statements on the clinical question of interest, i.e. the probability that an effect is present, given the data. The results are expressed as odds ratios (OR), with their respective 95% credibility interval (CI), which is the range of values in which the OR lies with 95% probability. In this study, the Bayesian results provide the probability that the OR of cesarean section is higher than 1. Very high value of this probability (larger than 0.9 i.e. 90%) can be considered as statistically significant. If this probability is close to 1 i.e. 100%, we'll conclude that the probability of the tested hypothesis is very high. If this probability is close to 0.5 i.e. 50%, the distribution is half negative, half positive, and we'll conclude that the OR is not different from 1."

We hope that this paragraph makes understanding of Bayesian statistics better.

b. As stated above, please reference the TRIPOD statement/checklist and adhere to the proposed guidelines

We added this reference.

c. Lines 208-209. Why was the regression coefficient multiplied by 6?

Thank you for this question. We created our score by following the rules for creating a weighted score using the beta coefficients (Hemalkumar et al. 2016). One of the usual ways is to divide every beta coefficients by the smallest, but this would have created a score that would have been too difficult to use. As it is described, we multiplied the coefficients by the same number in order to have a score that is easy to use. Multiplying by 6 allowed us to have a score out of 50, and therefore an easy to use score.

Regression coefficient based scoring system should be used to assign weights to the risk index

Hemalkumar B. Mehtaa, Vinay Mehtab,*, Cynthia J. Girmanc, Deepak Adhikaria, Michael L. Johnsond

J Clin Epidemiol. 2016 Nov;79:22-28. doi: 10.1016/j.jclinepi.2016.03.031. Epub 2016 May 13

We rephrased the explanation section and added the reference for the method:

“To create a regression coefficient-based scoring system, the beta coefficient were multiplied by a unique number (6) and then rounded off in order to derive weights, as previously described [27]. Therefore we obtained a simplified score, easy to calculate, out of 50 points.”

6. Lines 217-225. Is the information about how the external validation cohort was selected necessary?

We believe this information is necessary indeed. We finally came out with a single cohort for external validation but we performed a systematic research of all obstetrical databases publicly available. In this way, we avoid the bias of selecting only the cohort(s) in which our score shows better performance.

Discussion:

1. I would suggest the authors address the use of the current model compared to recent, similar published scoring systems for prediction of cesarean section (Levine, et al). What is the difference in these scoring systems other than use of Bayesian statistics?

We believe that the strenghts of our study include also that we used a prospective multicenter national-based cohort especially designed for the purposes of induction analysis. Comparing to Levine, our scoring system is easy to use and does not require any app. Above all, prediction of cesarean section is slightly better.

Please, see the following paragraph in the discussion section: “We developed an easy-to-use, externally validated and efficient score to predict cesarean section after labor induction with cervical ripening. The calculation of the score is very easy and allows to display the probability of

cesarean section. The score was devised using a population-based multicenter cohort especially designed for the purposes of induction analysis. This 50-point score shows good discrimination and calibration in both internal and external validation. AUC for prediction of cesarean section was better than both the modified Bishop score and Levine score."

2. Line 313. Please address the strengths and limitations of using of Bayesian statistics in your approach.

Thank you for this comment. We believe Bayesians statistics are an appropriate alternative to the frequentist methods. The only limitation is that, unfortunately, the bayesian statistics are not well known by many readers. As you suggested, we added a whole paragraph in the methods section to make the bayesian statistics understandable.

3. Line 334. Isn't there always a risk of cesarean section when embarking on labor induction, or even during spontaneous labor?

We modified this sentence: "We opted to make the occurrence of cesarean section our primary outcome because, in the end, the primary interest of caregivers when embarking on induction is the risk of caesarean delivery."

4. Lines 339-360. This section seems to fit better with the section entitled "Clinical Implication" which provides context for the study findings. Rather than summarizing the study findings again and then acknowledging individual parameters that have already been published, could the authors provide a more in-depth assessment of how the current study fills in gaps in the existing literature?

In the discussion section, the interpretation section relates to the comparison between the variables included in our score and the published literature regarding those same parameters. As you suggested, we added a comment on how our study fills in gaps in the existing litterature: "To our knowledge, with the exception of the score published by Levine *et al*. [16], there is no existing score especially dedicated to induction of labor with cervical ripening. In comparison to

the score published by Levine *et al*.[16], our scoring system is easy to use and does not require any informatized calculator or app. Above all, prediction of cesarean section is slightly better."

5. How will prospective use of this model not become a self-fulfilling prophecy, whereby a high score subsequently affects the decision to proceed with cesarean delivery, which is the outcome predicted by the score?

This is a very important point, and we believe that this is somehow the main pitfall of any prediction score. Still, the purpose here is to provide the more accurate information to allow a better patient counselling. We believe that to ignore the chances of success of an induction of labor would be an even more dangerous pitfall.

Reviewer #4: The article analyzes two large dataset and tries to validate a new score for the prediction of the CS after induction of labor. The population includes only women with need of cervical ripening according to a previous application of the Bishop score.

From the methodological point of view I have some doubt regarding the use of a score after the use of another score, in this case of the Bishop. Of interest the new score includes cervical dilation, cervical consistency, cervical effacement, cervical position and fetal head station, that are all the items of the Bishop score. Three of these variables were included in the final model.

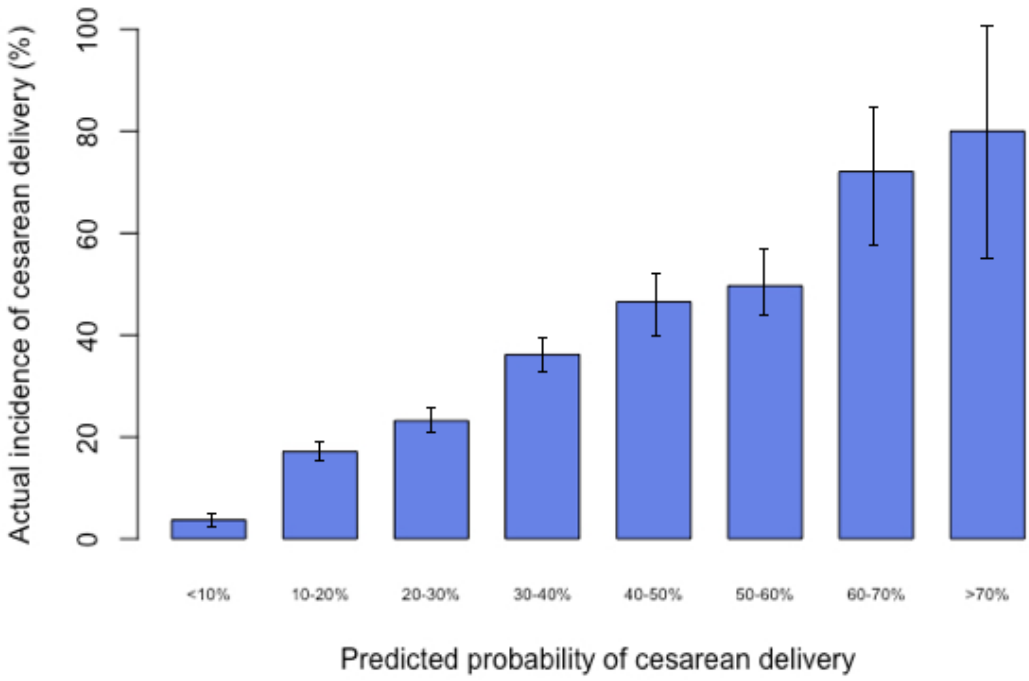
We understand your point and indeed, for our model building, we did not used the Bishop score as a score. The Bishop score was only used to characterize the cervix. The different parameters of the Bishop score were analyzed independently. Finally, only 3 turned out to be significantly associated with the occurrence of cesarean section.

We rephrased a sentence in the methods section to make this point clearer: "Cervical parameters were described as in the Bishop score but they were analyzed independently."

The second point regards the clinical utility of the score. Like for VBAC and admission to

vaginal delivery, which is the cut off that we should use to renounce to induction? If the risk is 35-50% an elective CS could substitute the induction? Moreover only a very few number of patients could have a score > 40. I.e. a nullipara, > 41 weeks, with PROM, non reassuring FHR (anomaly of the fetal vitality is confusing), BMI>35, height < 160, cervix closed, no effacement at all, head mobile has a score of 40 and has anyway a probability of vaginal delivery of 50% after induction. These numbers should/can reassure or confuse the obstetrician?

You're right, it's pretty rare that the risk of cesarean after a labor induction with cervical ripening is higher than 50%. In order to provide data about the numbers of patients concerned by each stratum of risk, we have added this information in the calibration plot (modified Figure 3.) Having a score > 40 points relates to a risk of cesarean section of more than 50%. That concerns only 212 patients i.e. 5% of the patients induced in our cohort.



Cesarean delivery	66	121	153	217	113	81	31	8
Vaginal delivery	1747	587	508	384	130	82	12	2

We believe it's not possible (and maybe not even desirable) to define a cut off for which an induction must not be performed. The purpose here is to provide the more accurate information to

allow a better patient counselling. The choice will depends on the patients characteristics, especially the indication for induction and the gestationnal age, and of course the patient wishes.

We should consider the effect of the practical application of this score as more relevant that the performance of the score itself, also because the only modifiable variable is the condition of the cervix that is related to advancing GA, that is on the contrary pejorative. Differently from the VBAC the risks of the admission at induction ending in CS are less explored, also in the prospective of future pregnancies and delivery if the present delivery will be a CS, and from the point of view of the neonate. If we renounce to an indicated induction at 37 weeks for example we should administrate corticosteroids and completely change the condition of the neonate? These points should be discussed in the paper.
Thank you for this comment. We've added a paragraph discussing these point in the clinical implication section, in the discussion section: "So the purpose of this score is to provide an accurate information about risk of cesarean section, not to define a cut-off above which a cesarean section should substitute the induction. Several clinical parameters should also be taken into consideration, including the indication for induction. To be noted that this score does not assess neonatal well-being, nor the implications for a subsequent pregnancy. Those points should be included in the counselling."

STATISTICAL EDITOR'S COMMENTS:

1. lines 189-190: Should enumerate (could be on-line material) the n(%) of missing data for each of the variables. Should similarly cite the missing data for each variable in the internal validation set and the n(%) of complete data for that portion of the data. Also, within both the derivation and the validation data sets, were proportion of patients with complete data statistically equivalent for both those with cesarean and vaginal deliveries?
Thank for your comment. The n(%) of missing data for each variables in the derivation set and internal validation set was added in the Appendix 3.

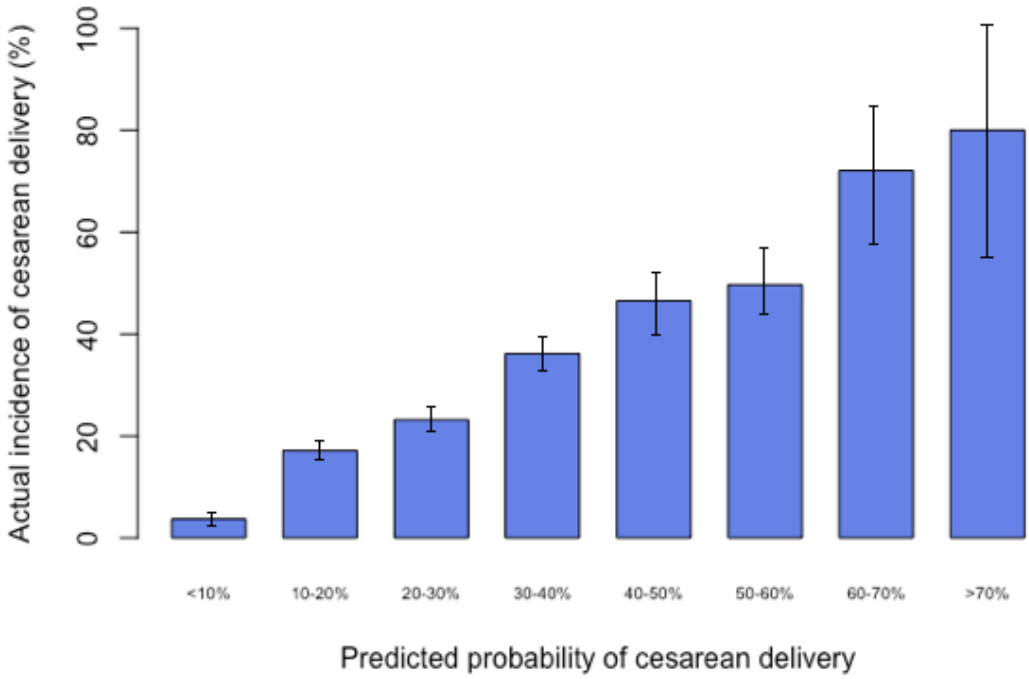
The proportion of patients with complete data is statistically equivalent for both those with cesarean and vaginal deliveries, within both the derivation and the validation data sets.
In the derivation set, the proportion of patients with complete data is 91% with cesarien deliveries, and 91% with vaginal deliveries.
In the internal validation set, the proportion of patients with complete data is 92% with cesarean deliveries, and 92% with vaginal deliveries.

2. Were the samples n = 1024 and n = 668 the total samples or the samples with complete data? Need to clarify.

In the derivation set, n=1024 is the total sample and n=932 (91%) is the sample with complete data.
In the internal validation set, n=668 is the total sample and n=615 (92%) is the sample with complete data.
Clarifications have been added in the article.

3. Fig 3 and Appendix 4 give an incomplete summary of the utility of the model. Should emulate the format in TRIPOD, with an accounting of what proportion of the derivation data set (presumably those with only complete data) comprised each of the 8 subsets represented by Fig 3 (or some other suitable stratification), each with the observed vs the expected probability for that stratum and with the CI for that stratum.

Thank you for your comment. We followed your advice and emulated the format in TRIPOD.



Cesarean delivery	66	121	153	217	113	81	31	8
Vaginal delivery	1747	587	508	384	130	82	12	2

4. Ref:

"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.

This reference was added.

5. This reference includes a calibration curve (fig 8), which allows the reader to see the relationship of observed vs predicted probabilities along the spectrum of probabilities from the data, along with confidence intervals for those prediction estimates. An added feature is the display below the x-axis of the relative counts of adverse vs non-adverse outcomes. Alternatively, those could be displayed (similar to survival analysis graphs), with numerical counts of adverse vs non-adverse outcomes at the intervals referred by the graph. The advantage to this level of detail is that it would convey to the reader the

strength of association at various model scores, along with their relative uncertainty, reflected by the amount of data available at various cut-points.

We modified the figure 3 and added below a table with numerical counts of adverse vs non-adverse outcomes at the intervals referred by the graph.

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, as well as subsequent author queries. 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:

1. OPT-IN: Yes, please publish my response letter and subsequent email correspondence related to author queries.

2. OPT-OUT: No, please do not publish my response letter and subsequent email correspondence related to author queries.

OPT-IN: Yes, please publish my response letter and subsequent email correspondence related to author queries.

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.

Thank you for your explanations.

Any author agreement forms previously submitted will be superseded by the eCTA. During the resubmission process, you are welcome to remove these PDFs from EM. However, if you prefer, we can remove them for you after submission.

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.

I, Floriane Jochum, affirm 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 registered have been explained.

4. Clinical trials submitted to the journal as of July 1, 2018, must include a data sharing statement. The statement should indicate 1) whether individual deidentified participant data (including data dictionaries) will be shared; 2) what data in particular will be shared; 3) whether additional, related documents will be available (eg, study protocol, statistical analysis plan, etc.); 4) when the data will become available and for how long; and 5) by what access criteria data will be shared (including with whom, for what types of analyses, and by what mechanism). Responses to the five bullet points should be provided in a box at the end of the Methods section.

A data sharing statement has been added at the end of the Methods section. MEDIP database belongs to the Assistance Publique-Hôpitaux de Paris (APHP) and individual participant datasharing is not yet available. The protocol is already published on www.clinicaltrials.gov (NCT02477085).

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.

A complete set of information about induction was prospectively collected at the beginning of induction, allowing precise data to be obtained on indications for induction and conditions under which it was performed. Full details on the protocols can be found in the original MEDIP article[18]. Briefly, MEDIP is a prospective multicenter observational French population-based cohort study. The data collected come from several sources, including data specific to the induction of labor completed prospectively by the physician or midwife in charge of the patient at the time of the procedure.

5. 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.

The use of the reVITALize definitions is not problematic to us.

6. 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.

We have less than 5500 words.

7. 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.

The abstract has been checked.

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

It is less than 300 words (273 words exactly).

8. 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.

We explained all the abbreviations used.

9. 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.

The virgule symbol was not used with words in this manuscript.

10. 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.

We reviewed the journal's Table Checklist to make sure that our tables conformed to journal style.

11. 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.

We do not wish to publish our article as open access.