

# 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)\*
- Email correspondence between the editorial office and the authors\*

*\*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](mailto:obgyn@greenjournal.org).

**Date:** Dec 20, 2018  
**To:** "Lee Reicher" [REDACTED]  
**From:** "The Green Journal" em@greenjournal.org  
**Subject:** Your Submission ONG-18-2108

RE: Manuscript Number ONG-18-2108

Prediction of success in external cephalic version for breech presentation at term

Dear Dr. Reicher:

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

#### REVIEWER COMMENTS:

Reviewer #1: This is a retrospective study using clinical information to predict success rate after ECV in 250 women at 36-41 week' gestation. All procedures were performed by the same experienced physician. After using creative Condition inference classification tree analysis , they found that parity , BMI , and fore bag size in cm are the best predictors for success. They found that fore bag size, BMI, and parity are the best predictors for ECV success. I have few comments

1. Despite the impressive results , the findings are not generalizable to clinical practice since all procedures were performed by one experienced provider.
2. What are the years of the study? the authors should include the time period.
3. The authors provide data on mean BMI. Do they have data on number of those with AFI < 5 or deepest vertical pocket < 2 cm.? how do they compare to the fore bag data.?
4. Data relating to parity and BMI are not surprising , and data on fore bag size is a proxy for lack of engagement.
5. The average BMI for the patient population is way different than that in the US. I am interested to see how the Fore bag size will predict ECV in those with BMI > 40.
6. The authors should provide more data on obstetric outcomes other than admission to NICU. It is surprising that no complications are included.

Reviewer #2: This is a retrospective cohort study designed to develop a prediction model to predict success after external cephalic version for breech presentation at term. The study identified 3 independent predictors including: fore-bag size, BMI and prior vaginal birth.

#### Main issues:

- 1- There was already 2 other studies (Kok, Am J Perinatol, 2011;28:103-110; Burgos. Aust N Z J Obstet Gynecol. 2012;52:59-61) that addressed the same question, why do the authors believe their study had better accuracy? Please provide the C-index of the final proposed model including the 3 selected variables ( BMI, vaginal deliveries, and size of the

fore-bag) so the readers can compare the accuracy of this study model to the previously published studies!

2- Please review and utilize the STROBE statement to improve the reporting of the study!

3- The presence of a fore-bag is not well standardized objective measure, was that measured a priori and retrospectively reviewed or was that re-measured by the study investigators. Was that measure used in any prior studies? Please provide a reference for how to calculate this variable and how accurate and reliable is that measure?

Specific Issues:

1- Introduction: What was the prior factors published in the literature and included in prior models?

2- Methods:

a. Please provide a reference for the use of nifedipine before ECV!

b. All patients included in the study was done by a single obstetrician who is very experienced in ECV (> 500); how this apply to the ECV practice in the US and in other places, how does this affect the external validity of the findings of the study?

c. The C-tree method to develop the final model is not a common statistical method, please explain why this was used instead of the regular multivariate regression?

d. Please provide a reference for the stat done to develop the model and remove specific formula included! How did the author select the included 3 variables for the final model? What was the criteria "statistically and epidemiologically utilized" ; did the author evaluated prior published risk factors for success/failure and included in their analysis?

e. Why did the author exclude women who had AFI < 8 , what is the reference cited for this cut off?

f. Please provide a power analysis based on the sample size included in the study (250 patients) to allow the reader to understand if the study was powered to study the research question in hand.

3- Results, tables and figures:

a. Please include a flow chart starting from all the patients who attempted ECV at the study institution and clarify any exclusions! It is not clear to me if there are other physician who do ECV or it is only one provider in the whole institution?

b. The final model presented in table 3 included 11 variables, please include a final model with only the 3 "selected" factors ( prior vaginal delivery, BMI <29 and Forebag >1 cm) and please provide the C-statistic for that model so it can be compared to prior published models (Kok, Am J Perinatol, 2011;28:103-110; Burgos. Aust N Z J Obstet Gynecol. 2012;52:59-61)

c. Table 1: Why did the authors include "Neonatal intensive care unit" in the clinical and demographic, this is an outcome?

d. Table 1: please include the median parity and the range and do not use parity as a continuous variable because it is usually skewed! Please consider a cut off.

e. Table 3 please present the univariate regression analysis for all the factors and have a separate table with only the final variables included in the model (Prior vaginal delivery, Fore-bag >1 cm and BMI <29 kg/m<sup>2</sup>)

f. Please add a figure starting from all the ECV done at your institution during the study duration and identify the reason for any exclusion?

g. Figure 1 can be removed

h. Figure 2 is not clear to me why the author mentioned in the abstract BMI <29 and they used BMI <25 for the figure, please be consistent and please explain what is the importance of this figure?

i. Figure 3 please replace this decision tree with Nomogram that can used to estimate the probability of success/failure of ECV based on the final model!

4- Discussion:

a. Please start with the main finding of the study

b. To be able to compare this study to the prior studies, please add a multivariate regression analysis and provide the C-statistic of the final model that only include ( BMI, parity and size of Fore-bag!

Reviewer #3:

1. Summary:

This is a retrospective cohort study for women at 36-41 weeks gestation with malpresentation who opted for an attempt at external cephalic version at a single institution between February 2016 through July 2018. Different variables were analyzed based on ECV success and a predictive model was created based on those variables. Presence of a forebag was identified as the most important variable for successful ECV in this predictive model.

2. Novelty:

Though there have been other predictive models presented in the literature, this study presents novel variables to consider clinically.

### 3. Methodology:

- a. Was the forebag measurement obtained using transabdominal or transvaginal ultrasound? I would specify this in the article.
- b. Excluding nuchal cords (which are very common and not a true contraindication for ECV) limits the applicability of this predictive model.
- c. This clinical predictive model needs to be externally validated.
- d. Were other forms of malpresentation (i.e. transverse lie) included in this study, or strictly breech presentation? Please specify in article.
- e. For the model, the formula for accuracy includes TP, TN, FP, and FN which are defined in the article. However, it is not specified what "negative class" and "positive class" are defined as.
- f. You performed in table 3 a logistic regression; did you also consider using a logistic regression classifier instead of a decision tree in order to create your model?
- g. What were the hyperparameters that you used in training your model?

### 4. Significance:

- a. This is overall a nicely designed study that seeks to address an important clinical issue.

### 5. Presentation:

- a. Wording is a bit confusing at times; this is a retrospective study, so patients were identified, not "eligible" (see line 33).

### STATISTICAL EDITOR COMMENTS:

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

lines 103-106: Need to provide data showing what were the frequency of each of the 4 conditions re: terminating the session. Were reasons iii or iv associated with any of the variables in the final model?

lines 175-182: Need to clarify whether the  $\pm$  intervals after the proportions represented SD or 95% CIs. Should include a table of sensitivity, specificity and accuracy, along with CIs, for the test version and validation set version.

Were the parity, forebag dimension and BMI tested for collinearity?

Table 1: This cites the data for all 250 women, both the training and the model validation sets. Need to provide information re: each separately to ensure that they were representative of the same population. Parity can only have integer values, so should cite as median(IQR or range) or as categories and test non-parametrically.

Table 2: Estimated fetal weight should be rounded to nearest gram, not .01 gram. For amniotic fluid volume index, estimated fetal wgt and fore-bag, should format as mean(SD) and cite that as footnote to Table.

Table 3: Need to include crude, then multivariate ORs and give justification for the number of variables in the final model, to ensure that the model is not over fitted. Need to cite referent, for example, is age per year, or is estimated fetal wgt per gram or what are time units for time in breech presentation, or is fore-bag per 1 cm etc for all variables.

For model justification including 3 variables, need to show the utility of a model including only parity or BMI or fore bag, then the additional precision gained by including first parity, then BMI then fore-bag in step-wise fashion.

Figures: Need to clarify whether these are from all or just from training set of data. If just from training set, then need to separately show the results of the validation set.

### EDITOR COMMENTS:

1. Thank you for your submission to Obstetrics & Gynecology. In addition to the comments from the reviewers above, you are being sent a notated PDF that contains the Editor's specific comments. Please review and consider the comments in this file prior to submitting your revised manuscript. These comments should be included in your point-by-point response cover letter.

\*\*\*The notated PDF is uploaded to this submission's record in Editorial Manager. If you cannot locate the file, contact Randi Zung and she will send it by email - rzung@greenjournal.org.\*\*\*

- When you write that a study occurred between date 1 and date 2, it literally excludes those boundary dates. For instance, "This study was performed between Feb 2018 and Jan 2019" would mean it was performed from March 2018 to

Dec 2018. Do you instead mean that the study was performed from date 1 to date 2? If so, please edit.

- since this is a single operator (which needs to be included in the abstract) by saying "all candidates" do you mean all candidates to that center or to that Ob GYN and there may have been others at the medical center who went to different MD's for ECV?

- were these variables you selected from a literature review? How were these variables selected?

- how was this known?

- how measured and when relative to the procedure?

- did you include parity in the model? From the results section, it seems that it was not included. Please clarify.

- The reported odds....

- have been

- Please define primary and secondary outcomes.

- Did you consider (and can you add) the secondary outcome of route of delivery?

- note abstract comment about "between". In next sentence, please name the IRB

- none of the patients had any medical or other obstetrical complications? No GDM, hypertension, thyroid problems, etc etc?

- As one reviewer notes, excluding women with AFI < 8 cm already excludes a group of women with relatively low fluid from your model. This is fine but needs to be emphasized in your discussion that these women are excluded off the top.

- Transabdominal or transvaginal scan?

- This limits the generalizability of the model of course--needs to be emphasized in the discussion

- what do you a variable is chosen to be divided? Do you mean it is chosen to be included?

- please clarify "training group" and "test group". By training group, is this the group on which the model was developed and test group is the group you internally validated it on? In any case, you need to explain what the implications of training and test groups are.

- were these the first 75% in the series? if not, how were the groups defined?

- Section of what? ctree is not something we've published about in the past and this needs to be more fully explained. You may want to put some of this explanation in the supplemental digital content. what is the purpose of putting aside 10%? were they ever included in the analysis? Why didn't they just go into the test group? As you can see, I don't understand the analysis and I'm sure others won't as well so further explanation is needed.

- please check instructions for authors about use of abbreviations. CS isn't an acceptable one and will need to be spelled out as cesarean births or cesarean deliveries

- please include here what you mean. Did you use some cut point (Forebag size > 1 cm vs < 1cm and then evaluate that? As presented in the table 2 it looks like a continuous variable. True for all of these. If this is going to be useful, important to indicate how to use this information.

- to be clear, you are referring to "training" here as a description of the way you divided your data set. To avoid confusion that you are describing training to be an Ob Gyn or perform this procedure, could you say "the accuracy in the training group"?

- what is a significant split?

- lower than what?

- you describe on line 221 a poor performance with AUC of 0.67 and a fair performance at 0.70. Are those really that different?

- This is called a primacy claim (your paper is the first or biggest) and must either be deleted or supported by providing

the search terms used, dates, and data bases searched (Medline, Ovid, Pubmed, Google Scholar, etc) in order to substantiate your claim.

- or more lax abdominal walls?

- we

- please see comments above in methods about some points that need to be made in your discussion re: generalizability and need for external validation

General: In addition to the above noted things that need to be added to the discussion of limitations, please note that this is a relatively small validation set and that the low BMI limits generalizability. Did the operator know the results of the information like forebag measurement, EFW, and BMI measurement before attempting the version? If so, this could have made her or him less likely to proceed with what is turning out to be a difficult version attempt. Please also clarify the description of the criteria used to call a "failed version"=again, two of these are very subjective and, particularly if the operator was aware of the pre-version assessments, could have influenced a decision to call a failed version earlier than otherwise. (Unconsciously).

2. 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.

3. 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.

4. 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.

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

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

7. Standard obstetric and gynecology data definitions have been developed through the reVITALize initiative, which was convened by the American College of Obstetricians and Gynecologists and the members of the Women's Health Registry Alliance. Obstetrics & Gynecology has adopted the use of the reVITALize definitions. Please access the obstetric and gynecology data definitions at <https://www.acog.org/About-ACOG/ACOG-Departments/Patient-Safety-and-Quality->

Improvement/reVITALize. If use of the reVITALize definitions is problematic, please discuss this in your point-by-point response to this letter.

8. Because of space limitations, it is important that your revised manuscript adhere to the following length restrictions by manuscript type: Original Research reports should not exceed 26 typed, double-spaced pages (6,500 words). Stated page limits include all numbered pages in a manuscript (i.e., title page, précis, abstract, text, references, tables, boxes, figure legends, and print appendixes) but exclude references.

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

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

10. Provide a précis on the second page, for use in the Table of Contents. The précis is a single sentence of no more than 25 words that states the conclusion(s) of the report (ie, the bottom line). The précis should be similar to the abstract's conclusion. Do not use commercial names, abbreviations, or acronyms in the précis. Please avoid phrases like "This paper presents" or "This case presents."

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

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

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

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

14. Line 235: 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.

15. 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](http://edmgr.ovid.com/ong/accounts/table_checklist.pdf).

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

17. Figures 1-3 may be resubmitted as-is.

18. 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/ifaauth.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.

19. 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, that each author has given approval to the final form of the revision, and that the agreement form signed by each author and submitted with the initial version remains valid.

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

Sincerely,

Nancy C. Chescheir, MD  
Editor-in-Chief

2017 IMPACT FACTOR: 4.982

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

---

In compliance with data protection regulations, please contact the publication office if you would like to have your personal information removed from the database.



January 9, 2019

To

Nancy C. Chescheir, Editor-in-Chief

Obstetrics & Gynecology journal

Thank you for your enlightening remarks and favorable consideration for publication of our manuscript entitled: "Prediction of success in external cephalic version for breech presentation at term" (ONG-18-2108). Attached hereby is our response to the reviewers' and editor's remarks. We addressed each and every comment.

**Reviewer #1**

*1. Despite the impressive results, the findings are not generalizable to clinical practice since all procedures were performed by one experienced provider.*

Thank you for this comment. The study was conducted in a single tertiary center, by a single experienced physician who was totally unaware of parameters later determined to be predictive of successful version. In addition, by focusing on ECV performed by a single provider, we were able to offset any bias or impact that might stem from differences in technique or experience by a range of providers. The entire procedure, from the pre-monitoring phase, the approach to therapy, and the manual rotation maneuver itself was the same for all patients. The generalization issue was discussed by the authors prior to publication and we agreed that the advantages of a single performer outweigh the lack of generalization. Nevertheless we added this important issue to the limitation section (Discussion) of our study.

*2. What are the years of the study? The authors should include the time period.*

"We conducted a retrospective cohort study of pregnant women with singleton pregnancies and breech presentations who opted for an ECV attempt at a single,

tertiary, university-affiliated medical center between February 2016 and July 2018." We have added this information to the abstract and methods.

*3. The authors provide data on mean BMI. Do they have data on number of those with AFI < 5 or deepest vertical pocket < 2 cm.? how do they compare to the fore bag data.?*

Patients with an amniotic fluid index (AFI) <8 cm were excluded from the study, thus we do not have data about those patients.

*4. Data relating to parity and BMI are not surprising, and data on fore bag size is a proxy for lack of engagement.*

Previous scoring systems have addressed fetal, uterine and maternal parameters as predictors for success and identified maternal parameters such as BMI and uterine parameters such as parity as strong predictors. Previously identified fetal parameters such as lack of engagement and palpable fetal head are usually subjective and therefore, may introduce bias into the predictive model. Here we introduce the fore bag size as a novel, easily obtainable, objective fetal parameter which can be easily measured by an abdominal ultrasound and may be used to predict version attempt success.

*5. The average BMI for the patient population is way different than that in the US. I am interested to see how the Fore bag size will predict ECV in those with BMI > 40.*

The mean BMI among pregnant Israeli women is 26.1 with a 95% range of 22-29. The average BMI in our study was similar. We do not have sufficient data to predict the success of ECV in those with a BMI > 40. It would certainly be worthwhile and interesting to apply this model in further studies in additional populations.

*6. The authors should provide more data on obstetric outcomes other than admission to NICU. It is surprising that no complications are included.*

No procedure related complications were observed among our patients from the completion of the procedure to their discharge following the procedure. In addition,

this study aimed at identifying possible determinants of success of an ECV using a model based on clinical parameters for successful ECV near term. While our focus was the objective success of ECV based on the predictive value of objective clinical parameters, we found lower rates of NICU admission in the successful group (9.3% vs 20.5%,  $p < 0.05$ ). Further studies including a larger study population might provide information regarding the rate of complications from ECV, which is typically around one percent.

## **Reviewer #2**

### **Main issues**

*1. There was already 2 other studies (Kok, Am J Perinatol, 2011;28:103-110; Burgos. Aust N Z J Obstet Gynecol. 2012;52:59-61) that addressed the same question, why do the authors believe their study had better accuracy? Please provide the C-index of the final proposed model including the 3 selected variables ( BMI, vaginal deliveries, and size of the fore-bag) so the readers can compare the accuracy of this study model to the previously published studies!*

We are grateful to the reviewer for the valuable in-depth analysis of our paper and for the important comments and suggestions for enhancing it. The C-index was calculated for the c-tree model for both the training set [0.928 (0.924-0.931)] and the internal validation set [0.933 (0.863-1)]. We agree that this allows a better comparison measure between our model and the ones previously published. This information was added to the results section in the text

*2. Please review and utilize the STROBE statement to improve the reporting of the study!*

We read the checklist thoroughly and ensured that all of the parameters were present and added any elements that were not present.

3. *The presence of a fore-bag is not well standardized objective measure, was that measured a priori and retrospectively reviewed or was that re-measured by the study investigators. Was that measure used in any prior studies? Please provide a reference for how to calculate this variable and how accurate and reliable is that measure?*

The fore bag measurement is a proxy of lack of engagement which is easily obtainable and reproducible. The measurement is done by the sonographer prior to the ECV by an abdominal scan. The distance between the internal os and the lowest part of the presenting fetal breech is measured. The measurement is very accurate and reliable. To the best of our knowledge, this measurement had never been used as a predictor of ECV success.

## **Specific Issues**

**Introduction:** *What was the prior factors published in the literature and included in prior models?*

Prior models were based on a combination of fetal, uterine and maternal parameters. Fetal parameters included: estimated fetal weight, estimation of fetal engagement, fetal spine direction and palpable fetal head. Uterine parameters included uterine tone, use of uterine relaxants, the amount of amniotic fluid and placental location. Maternal parameters included parity, BMI, and maternal analgesia. We focused on easily obtainable and reproducible parameters representing each of the categories mentioned above

This information was added to the manuscript

## **2. Methods**

a. *Please provide a reference for the use of nifedipine before ECV!*

The use of tocolytic agents to enhance the success rates of ECV is debatable due to lack of incisive evidence. In a Cochrane database review from 2004 it is suggested

that tocolytic agents may reduce the failure rates of ECV (**Hofmeyr GJ**, Interventions to help external cephalic version for breech presentation at term, [Cochrane Database Syst Rev](#). 2004) but the best tocolytic candidate (if at all) is yet to be determined. We use nifedipine as an internal department protocol. Our clinical impression has been that the use of tocolytics prior to the performance of ECV reduces uterine tone and facilitates the procedure. Similar studies also used tocolytics prior to performing the procedure. (Burgos J, A prospective study of the factors associated with the success rate of external, International journal of obstetrics&gynecology, 2010)

*b. All patients included in the study was done by a single obstetrician who is very experienced in ECV (> 500); how this apply to the ECV practice in the US and in other places, how does this affect the external validity of the findings of the study?*

By focusing on ECV performed by a single provider, we were able to offset any bias or impact that might stem from differences in technique or experience by a range of providers. The entire procedure, from the pre-monitoring phase, the approach to therapy, and the manual rotation maneuver itself was the same for all patients. The generalization issue was discussed by the authors prior to publication and we agreed that the advantages of a single performer outstands the downside of generalization. Nevertheless we added this important issue to the limitation of our study.

It would certainly be worthwhile to apply our model across ECV procedures performed by multiple clinicians in order to validate our findings.

*c. The C-tree method to develop the final model is not a common statistical method, please explain why this was used instead of the regular multivariate regression?*

Conditional inference classification trees (ctree) is a type of decision tree. Decision trees represent a form of discriminative models much like logistic regression. By identifying the most informative variables and their corresponding cutoffs, decision trees provide several advantages over logistic regression including: (i) identification of subgroups that have the same outcome of interest, (ii) identification of non-linear relations to the outcome, (iii) identification of variable interactions that better stratify the population according to the outcome, (iv) easily interpretable models that allow users without statistical background to better understand the variables association to

the outcome and implement the model to predict the outcome of new cases. A major drawback of decision trees is over simplification of variable-outcome relationships by setting a single cutoff when a continuous relationship is present. However, we believed the aforementioned benefits outweigh this issue and therefore chose to implement ctree to generate the final decision model.

We have added this explanation to the text.

*d. Please provide a reference for the stat done to develop the model and remove specific formula included! How did the author select the included 3 variables for the final model? What was the criteria "statistically and epidemiologically utilized" ; did the author evaluated prior published risk factors for success/failure and included in their analysis?*

The variables chosen for the final tree model were automatically selected by the ctree algorithm out of a list of demographic and obstetric parameters collected for each case. The ctree algorithm is completely unbiased, which means that no information regarding known risk association was incorporated into its development process. The underlying algorithm incorporates formal testing of statistical hypotheses in order to choose which variables to include, determine their discriminatory value, and the order in which they occur in the tree. A variable is chosen to be included in the final model only if it is significantly associated with the outcome.

We have deleted the accuracy formula from the text and added C-index to the performance measures.

*e. Why did the author exclude women who had AFI < 8 , what is the reference cited for this cut off?*

Several previous studies addressed the issue of AFI threshold for successful ECV and a few cut-offs had been proposed ranging from <10 ([Boucher M1](#), [Bujold E](#), [Marquette GP](#), [Vezina Y](#). The relationship between amniotic fluid index and successful external cephalic version: a 14-year experience. [Am J Obstet Gynecol](#). 2003 Sep;189(3):751-4.) to <5, 5-8 (([Haas DM](#), [Magann EF](#). External

cephalic version with an amniotic fluid index  $\leq 10$ : a systematic review, [J Matern Fetal Neonatal Med.](#) 2005 Oct;18(4):249-52.). Since there is no agreement we have decided upon a threshold of AFI $\leq 8$  as a reasonable cutoff.

*f. Please provide a power analysis based on the sample size included in the study (250 patients) to allow the reader to understand if the study was powered to study the research question in hand.*

We conducted a one sample proportion power analysis and added the following description to the methods section:

A one sample proportion power analysis determined that given  $n=250$  patients, our study had the power to detect even a small-sized effect (0.185) corresponding, for example, to an increase in accuracy from 80% to ~87% with type I error probability of 0.05 and type II error probability of 0.1

### **3. Results, tables and figures**

*a. Please include a flow chart starting from all the patients who attempted ECV at the study institution and clarify any exclusions! It is not clear to me if there are other physician who do ECV or it is only one provider in the whole institution?*

Thank you for this comment, we added the suggested flow chart.

Due to the nature of ECV as a procedure that requires extensive experience, a single provider performs nearly all of the procedures in our department.

*b. The final model presented in table 3 included 11 variables, please include a final model with only the 3 "selected" factors ( prior vaginal delivery, BMI  $\leq 29$  and Forebag  $>1$  cm) and please provide the C-statistics for that model so it can be compared to prior published models (Kok, Am J Perinatol, 2011;28:103-110; Burgos. Aust N Z J Obstet Gynecol. 2012;52:59-61)*

In order to compare the ctree-based model to a more conventional logistic regression model, we performed two multivariate logistic regression analyses on the training set



(75% of the original dataset): one full model with the entire set of variables and another with a partial model including only BMI, fore-bag size and number of previous deliveries. We added a table that describes the results of the two models. We also describe their performance (accuracy and c-index) when tested on the internal validation set. Overall we show that both accuracy and c-index of the ctree model were non-inferior when compared with the logistic regression models (Accuracy: 91.9% vs 90.3% and 91.9%, C-index 0.93 vs 0.969 and 0.967) .

*c. Table 1: Why did the authors include "Neonatal intensive care unit" in the clinical and demographic, this is an outcome?*

Changes were made for table 1

*d. Table 1: please include the median parity and the range and do not use parity as a continuous variable because it is usually skewed! Please consider a cut off.*

Corrected

*e. Table 3 please present the univariate regression analysis for all the factors and have a separate table with only the final variables included in the model (Prior vaginal delivery, Fore-bag >1 cm and BMI <29 kg/m<sup>2</sup>)*

We have added a table (table 2) describing both the univariate analysis and the multivariate regression analysis results for both a model with all the factors and a model with only the factors identified by the ctree. In both models, fore-bag size, BMI and parity are considered as continuous variables.

*f. Please add a figure starting from all the ECV done at your institution during the study duration and identify the reason for any exclusion?*

We added the suggested flow chart

*g. Figure 1 can be removed*

Figure 1 was added to the text in order to visualize the demographic, clinical and sonographic characteristics differences between the study groups. The figure can be removed from the text if you find it redundant

*h .Figure 2 is not clear to me why the author mentioned in the abstract BMI <29 and they used BMI<25 for the figure, please be consistent and please explain what is the importance of this figure?*

Figure 2 was removed.

*i. Figure 3 please replace this decision tree with Nomogram that can used to estimate the probability of success/failure of ECV based on the final model!*

We thank the reviewer for this comment as it encouraged us to revisit the utility and benefit of the final selected decision tree model. Nomograms are graphical representations of regression analysis. Due to previous issues raised in this review, we have added a performance comparison between the ctree-based decision algorithm and two multivariate logistic regression models (one with all the factors and another with only BMI, fore bag size and number of previous deliveries). We show that the ctree accuracy is non-inferior to both models and that there is only a mild difference in overall discriminatory performance. We believe that the intuitive interpretability of the ctree model justifies its use over a nomogram in this context and will represent an easier more accessible decision aid for clinicians.

#### **4 Discussion**

*a. Please start with the main finding of the study*

A paragraph describing the main findings was added to the beginning of the discussion section.

*b. To be able to compare this study to the prior studies, please add a multivariate regression analysis and provide the C-statistic of the final model that only include ( BMI, parity and size of Fore-bag!*

As we describe in a previous comment, we performed two multivariate logistic regression analysis on the training set (75% of the original dataset): one full model with the entire set of variables and another with a partial model including only BMI, fore-bag size and number of previous deliveries. We added a table that describes the results of the two models. We also describe their performance (accuracy and c-index) when tested on the internal validation set. Overall we show that that both accuracy and c-index of the ctree model were non-inferior when compared with the logistic regression models (Accuracy: 91.9% vs 90.3% and 91.9%, C-index 0.93 vs 0.969 and 0.967) . A ROC curve figure was added to facilitate visual comparison as well.

### **Reviewer #3**

#### **1. Summary**

*This is a retrospective cohort study for women at 36-41 weeks gestation with malpresentation who opted for an attempt at external cephalic version at a single institution between February 2016 through July 2018. Different variables were analyzed based on ECV success and a predictive model was created based on those variables. Presence of a forebag was identified as the most important variable for successful ECV in this predictive model.*

#### **2. Novelty**

*Though there have been other predictive models presented in the literature, this study presents novel variables to consider clinically.*

#### **3. Methodology**

*a. Was the forebag measurement obtained using transabdominal or transvaginal ultrasound? I would specify this in the article.*

The forebag measurement obtained using transabdominal ultrasound. We specified this in the text.

*b. Excluding nuchal cords (which are very common and not a true contraindication for ECV) limits the applicability of this predictive model.*

Candidates for ECV went through a pre-procedure evaluation that included the presence of nuchal cord. Although this finding is not a contraindication for an attempt we conveyed the information to the patients and decided to abort the attempt. Altogether we had only three candidates that were excluded due to nuchal cord hence the impact was negligible.

*c. This clinical predictive model needs to be externally validated.*

Thank you for this comment. We are aware of this limitation of our study. Our model holds in internal validation. We encourage further studies to evaluate the performance of this prognostic model in other populations. We added this limitation to the discussion section of the manuscript.

*d. Were other forms of malpresentation (i.e. transverse lie) included in this study, or strictly breech presentation? Please specify in article.*

Patients with transverse lie have excellent success rates of ECV. We focused on parameters to predict the success rates of strictly breech presentation in order to refrain from performing unpleasant procedures with a very low success rates, hence, all other forms of malpresentations were excluded

*e. For the model, the formula for accuracy includes TP, TN, FP, and FN which are defined in the article. However, it is not specified what "negative class" and "positive class" are defined as.*

As per a request from a different reviewer, we have removed the accuracy calculation formula. We have added this clarification to the text: The tree-based model was trained to predict the outcome of an ECV attempt given maternal and

ultrasonographic features collected prior to the attempt. A positive result was defined as a post version attempt cephalic presentation, and a negative result was defined as breech presentation after version attempt

*f. You performed in table 3 a logistic regression; did you also consider using a logistic regression classifier instead of a decision tree in order to create your model?*

We thank the reviewer for this suggestion, which was also raised by an additional reviewer. We agree that the tree-based model should be compared against more conventional methods such as multivariate logistic regression. We therefore performed two multivariate logistic regression analyses on the training set (75% of the original dataset): one full model with the entire set of variables and another with a partial model including only BMI, fore-bag size and number of previous deliveries. We added a table that describes the results of the two models. We also describe their performance (accuracy and c-index) when tested on the internal validation set. Overall we show that both accuracy and c-index of the ctree model were non-inferior when compared with the logistic regression models (Accuracy: 91.9% vs 90.3% and 91.9%, C-index 0.93 vs 0.969 and 0.967) . A ROC curve figure was added to facilitate visual comparison as well.

*g. What were the hyperparameters that you used in training your model?*

Training was done using the caret R package. The package uses the default ctree parameter setting and the only parameter tuning was performed over the ‘mincriterion’ parameter. As mentioned in the text, a split is implemented when the criterion exceeds the value given by mincriterion. The final selected value was 0.938 which means that the p-value must be smaller than 0.062 in order to split a node.

#### **4. Significance**

*a. This is overall a nicely designed study that seeks to address an important clinical issue.*

## 5. Presentation:

a. *Wording is a bit confusing at times; this is a retrospective study, so patients were identified, not "eligible" (see line 33).*

Thank you for your comment. Changes were made.

## **STATISTICAL EDITOR COMMENTS:**

*lines 103-106: Need to provide data showing what were the frequency of each of the 4 conditions re: terminating the session. Were reasons iii or iv associated with any of the variables in the final model?*

The association between the termination reasons and the collected variables was evaluated using a multivariate logistic analysis and no variable was found to be significantly associated with the termination reason.

*lines 175-182: Need to clarify whether the  $\pm$  intervals after the proportions represented SD or 95% CIs. Should include a table of sensitivity, specificity and accuracy, along with CIs, for the test version and validation set version.*

"All accuracy and C-index values are reported together with a 95% confidence interval" – added to the text . This table describes these metrics for each model on the validation set and may be added to the main text if requested.

		Full logistic regression	Partial logistic regression
	Ctree	model	model
Sensitivity	0.975	0.925	0.925
Specificity	0.818	0.909	0.864
PPV	0.907	0.949	0.925
NPV	0.947	0.870	0.864
Accuracy (95%	0.919 (0.822-0.973)	0.919 (0.822-0.973)	0.903 (0.801-0.964)

CI)

AUC (95% CI) 0.933 (0.864-1) 0.969 (0.923-1) 0.967 (0.927-1)

*Were the parity, forebag dimension and BMI tested for collinearity?*

In order to assess collinearity we tested pairwise correlation between the aforementioned variables and found that BMI is negatively associated with forebag and parity. Although significant, both correlations are below 0.5 and the variables have a calculated variance inflation factor (VIF) <3. We describe these correlations in the text and stress that decision trees are not as susceptible to collinearity as logistic regression.

*Table 1: This cites the data for all 250 women, both the training and the model validation sets. Need to provide information re: each separately to ensure that they were representative of the same population. Parity can only have integer values, so should cite as median(IQR or range) or as categories and test non-parametrically.*

The following table describes the training and validation set and may be added to the main text if requested. Parity is now described using median and IQR and tested using the Kruskal–Wallis test.

	level	Training set	Validation set	p
n		188	62	
Age (mean (sd))		(5.39) 30.60	(4.86) 31.81	0.118
Parity (median		,0.00] 1.00	,0.00] 1.00	
[IQR])		[2.00	[2.00	0.777
AFI (mean (sd))		(2.44) 13.68	(2.21) 13.65	0.949
(%) Back Direction	front	(21.8) 41	(14.5) 9	0.239
	left	(41.0) 77	(37.1) 23	
	right	(37.2) 70	(48.4) 30	



EFW (mean (sd))		(200) 2799	(187) 2780	0.517
BMI (mean (sd))		(3.34) 25.63	(3.28) 25.57	0.908
(%) Previous CS	0	(91.0) 171	(90.3) 56	1
	1	(9.0 ) 17	(9.7 ) 6	
Placenta direction				
(%)	back	(54.8) 103	(56.5) 35	0.935
	front	(45.2) 85	(43.5) 27	
Time in breech presentation (mean (sd))		(1.44) 3.49	(1.60) 3.56	0.729
Forebag (mean (sd))		(0.98) 1.42	(0.91) 1.40	0.904
End presentation				
(%)	breech	(35.1) 66	(35.5) 22	1
	cephalic	(64.9) 122	(64.5) 40	
(%) Birth type	CS	(33.0) 62	(33.9) 21	0.828
	ND	(59.6) 112	(56.5) 35	
	VE	(7.4 ) 14	(9.7 ) 6	
Apgar5 (mean (sd))		(0.76) 9.40	(0.66) 9.60	0.067
(%) NICU	0	(87.8) 165	(83.9) 52	0.569
	1	(12.2) 23	(16.1) 10	
Birth week (mean (sd))		(0.97) 39.27	(1.04) 39.11	0.274

*Table 2: Estimated fetal weight should be rounded to nearest gram, not .01 gram. For amniotic fluid volume index, estimated fetal wgt and fore-bag, should format as mean(SD) and cite that as footnote to Table.*

Thank you for this comment, we have updated the tables accordingly.

*Table 3: Need to include crude, then multivariate ORs and give justification for the number of variables in the final model, to ensure that the model is not over fitted. Need to cite referent, for example, is age per year, or is estimated fetal wgt per*

*gram or what are time units for time in breech presentation, or is fore-bag per 1 cm etc for all variables.*

Multivariate analysis was ran on the training data (75% of the original data) for both the complete set of variables and the three variables identified by the ctree model (fore-bag, BMI and parity). The results of both models are described in table 2. We have added the univariate OR for each variable to the table as well. A refernt is added for each variable in the table.

*For model justification including 3 variables, need to show the utility of a model including only parity or BMI or fore bag, then the additional precision gained by including first parity, then BMI then fore-bag in step-wise fashion.*

In this study, we describe the ctree model in which tree expansion occurs in a sequential manner - a given node is split only if it results in significant sub-grouping of the outcome. The multivariate regression described in this study was performed mainly in order to validate the performance of the tree model. For this purpose, two models were generated, one with the complete set of variables and the other with only the three variables used in the tree model.

In order to further demonstrate the association of the three selected variables with the outcome, we have performed the following analysis (may be added to the main text if requested):

a four stage hierarchical multiple regression was conducted with fore-bag entered at stage one of the regression, BMI entered at stage two, parity at stage three and all remaining variables at the last stage. The hierarchical multiple regression revealed that adding fore-bag, BMI and parity contributed significantly to the regression ( $\chi^2(1, N=186)=126.2$   $p<0.005$ ,  $\chi^2(1, N=185)=27.8$   $p<0.005$ ,  $\chi^2(1, N=184)=26.9$   $p<0.005$ , respectively). Finally, the addition of the remaining variables to the regression model did not result in a significant improvement of the model ( $\chi^2(1, N=176)=3.5$   $p=0.9$ ).

*Figures: Need to clarify whether these are from all or just from training set of*

*data. If just from training set, then need to separately show the results of the validation set.*

We have added annotation describing what dataset was used to generate each figure

### **EDITOR COMMENTS**

*- When you write that a study occurred between date 1 and date 2, it literally excludes those boundary dates.*

We mean that the study was performed between Feb 2018 and Jan 2019, inclusive

*- Since this is a single operator (which needs to be included in the abstract) by saying "all candidates" do you mean all candidates to that center or to that Ob GYN and there may have been others at the medical center who went to different MD's for ECV?*

Thank you for this comment. We included this in the abstract. Since there is only one operator in that center we mean "all candidates" to the center.

The study was conducted in a single tertiary center, by a single experienced physician who was totally unaware of parameters later determined to be predictive of successful version. we added this important issue to the limitation of our study. All candidates to our center were seen by a single operator. No other MD's attempted ECV's on candidates that were referred to our center with a breech presentation.

*- were these variables you selected from a literature review? How were these variables selected?*

This issue was also addressed by one of the reviewers.

Prior models were based on a combination of fetal, uterine and maternal parameters.

Fetal parameters included: estimated fetal weight, estimation of fetal engagement, fetal spine direction and palpable fetal head. Uterine parameters included uterine tone,

use of uterine relaxants, the amount of amniotic fluid and placental location. Maternal parameters included parity, BMI, and maternal analgesia. We focused on easily obtainable and reproducible parameters representing each of the categories mentioned above

*- how was this known?*

In Israel, routine fetal growth estimation by ultrasound is performed between 30 and 33 weeks' gestation. We therefore used these scan as our reference point regarding the length of time the fetus was in breech presentation. In many cases the 30-33 weeks ultrasound was followed by a later scan at 34-36 weeks in order to confirm malpresentation prior to ECV.

*- how measured and when relative to the procedure?*

The fore-bag is the size of the pocket of amniotic fluid preceding the fetal presenting part measured by an abdominal ultrasound in centimeters from the internal os to the low-most fetal breech (in the midline).

*- did you include parity in the model? From the results section, it seems that it was not included. Please clarify.*

Parity was included in the model and we have rephrased the paragraph describing the decision tree predictions adding the predicted effect of parity:

Reviewing the decision tree, patients with no fore-bag had a predicted probability of 2.6% and 10% for successful version in the training and internal validation cohorts respectively. Patients with a BMI larger than 29 kg/m<sup>2</sup>, had a low probability for version success, regardless of fore-bag size. The version outcome in patients with a BMI  $\leq 29$  was significantly associated with the fore-bag size. When fore-bag size was higher than 1cm, versions were likely to succeed (97% and 96.3% in the training and internal validation cohorts, respectively). When fore-bag size was 1cm, outcome differed between nulliparous and multiparous patients, with nulliparous patients having a much lower probability of success (24% vs 91% and 0% vs 81% in the training and validation cohorts, respectively).

- *The reported odds...*
- *have been*

Thank you for this comments. Changes were made.

- *Please define primary and secondary outcomes.*

Thank you for this valuable comment. We defined primary and secondary outcome in the method section.

- *Did you consider (and can you add) the secondary outcome of route of delivery?*

Thank you for this valuable comment. we have added this paragraph to the text:  
We continue to check which of the collected variables are significantly associated with the secondary outcome of final route of delivery. A multivariate logistic regression analysis including all the collected variables identified post ECV attempt cephalic presentation to be the only variable significantly associated with a vaginal route of delivery (OR 864 [95-14848];  $P < 0.005$ ).

- *Note abstract comment about 'between". In next sentence, please name the IRB*

The study was approved by the local institutional review board (TLV-0749-18), and all the participating patients provided written informed consent. Changes were made in the abstract.

- *None of the patients had any medical or other obstetrical complications? No GDM, hypertension, thyroid problems, etc etc?*

We did not evaluate in this study the influence of obstetric complications on the success of the ECV. This data can be provided if necessary.

- *As one reviewer notes, excluding women with AFI < 8 cm already excludes a group of women with relatively low fluid from your model. This is fine but needs to be emphasized in your discussion that these women are excluded off the top.*

Several previous studies addressed the issue of AFI threshold for successful ECV and a few cut-offs had been proposed ranging from  $<10$  ([Boucher M1](#), [Bujold E](#), [Marquette GP](#), [Vezina Y](#). The relationship between amniotic fluid index and successful external cephalic version: a 14-year experience. [Am J Obstet Gynecol](#). 2003 Sep;189(3):751-4.) to  $<5$ , 5-8 (([Haas DM](#), [Magann EF](#). External cephalic version with an amniotic fluid index  $< \text{or} = 10$ : a systematic review, [J Matern Fetal Neonatal Med](#). 2005 Oct;18(4):249-52.)). Since there is no agreement we have decided upon a threshold of  $\text{AFI} < 8$  as a reasonable cutoff.

*- Transabdominal or transvaginal scan?*

All scans performed for the evaluation of fore bag size were transabdominal ultrasounds.

*- This limits the generalizability of the model of course--needs to be emphasized in the discussion*

We have added another paragraph about limitations of this study.

*- what do you a variable is chosen to be divided? Do you mean it is chosen to be included?*

We have added this explanation to the text:

A variable is chosen to be incorporated into the decision tree only if it is significantly associated with the outcome. Association is tested by univariate analysis according to the variable tested (ANOVA and Chi Square for numeric and categorical variables, respectively). For additional information please see Torsten Hothorn, Kurt Hornik and Achim Zeileis (2006). Unbiased Recursive Partitioning: A Conditional Inference Framework. *Journal of Computational and Graphical Statistics*, 15(3), 651--674]

*- please clarify "training group" and "test group". By training group, is this the group on which the model was developed and test group is the group you internally validated it on? In any case, you need to explain what the implications of training and test groups are.*

*- were these the first 75% in the series? if not, how were the groups defined?*

*- Section of what? ctree is not something we've published about in the past and this needs to be more fully explained. You may want to put some of this explanation in the supplemental digital content. what is the purpose of putting aside 10%? were they ever included in the analysis? Why didn't they just go into the test group? As you can see, I don't understand the analysis and I'm sure others won't as well so further explanation is needed.*

This paragraph has been rephrased -

The tree-based model was trained to predict the outcome of an ECV attempt given maternal and ultrasonographic features collected prior to the attempt. A positive result was defined as a post version attempt cephalic presentation, and a negative result was defined as breech presentation after version attempt. The entire cohort was initially randomly divided into two groups. The first group, termed “training set”, comprised of 75% of the patients and used for model development. The second group, termed “internal validation set”, comprised of the remaining 25% and was used for performance evaluation. Cross validation was used during model development in order to estimate model performance under different parameter settings. Briefly, for each tested parameter, the training set was randomly divided into 10 sections. One section was put aside, and a classification tree was built upon the combined remaining nine sections. The performance of the constructed tree was estimated on the isolated section which was not used during model development. This form of evaluation tests how well a developed model performs on new datasets. This process was repeated five times, and the average of the combined estimates was recorded. The parameter setting with the best performance estimated by cross validation was used in the final decision tree model. Performance was compared between different values of the minimal 1-p-value used to decide whether a variable will be added to the tree (e.g mincriterion). The final model was developed using a mincriterion of 0.938. Since the outcome groups were comparable in size, the overall accuracy (i.e., the sum of correct ECV outcome predictions divided by the total number of predictions) was used for performance estimation. The C-index was calculated as an additional measure of



performance on both the training and the validation set. All accuracy and C-index values are reported together with a 95% confidence interval.

Finally, model performance was validated on the test set (the remaining 25%) that was not used during model development.

We have also added a reference in the text for the article describing the algorithm in greater detail.

*- please check instructions for authors about use of abbreviations. CS isn't an acceptable one and will need to be spelled out as cesarean births or cesarean deliveries*

These issues were corrected in the text.

*- please include here what you mean. Did you use some cut point (Forebag size > 1 cm vs < 1cm and then evaluate that? As presented in the table 2 it looks like a continuous variable. True for all of these. If this is going to be useful, important to indicate how to use this information.*

We re-ran the multivariate logistic regression analysis using only the training set and rephrased this paragraph:

As decision tree models may oversimplify variable-outcome relations, a multivariate logistic regression-based model was generated using the training set as well (Table 2). Three variables were identified to be significantly associated with successful ECV: Fore-bag size (cm) (OR 153 [31.5-1439.4]), BMI (kg/m<sup>2</sup>) (OR 0.6 [0.4-0.8]) and number of previous deliveries (OR 6 [2.6-17.3]). An additional logistic regression model was developed using only these three variables. When tested on the internal validation set, the accuracies of the complete and partial models were 91.9% and 90.3%, respectively and the C-indices were 0.969 and 0.967, respectively.

This sentence was added to the methods section: Forebag size, BMI, age, parity, estimated fetal weight, and time in breech presentation were all included as continuous variables and the remaining variables as categorical.

*- to be clear, you are referring to "training" here as a description of the way you*

*divided your data set. To avoid confusion that you are describing training to be an Ob Gyn or perform this procedure, could you say "the accuracy in the training group"?*

This section was rephrased in the text:

The tree was developed by analyzing 75% of the patients (the training set) and evaluated on the remaining 25% (the internal validation set) (Table XX). The final, best-performing tree included decision tree based on fore-bag size, BMI, and number of deliveries. The C-Index for the final tree during model training was 0.928 (0.924-0.931) and the accuracy for correctly predicting version outcome was 92.2% (91%-93.4%). Evaluating the performance of the developed tree on the internal validation subgroup yielded a C-Index of 0.933 (0.863-1) and the prediction accuracy was 91.9% (86.5%-97.3%)

*-what is a significant split?*

This paragraph has been rephrased :

Figure 3 summarizes the decision tree for estimating the probability of a successful ECV attempt (cephalic presentation after version). Model training and development was performed using 75% of the original cohort (training set). The final tree divides the cohort into five subgroups. Each subgroup is represented as a terminal node. Node size is indicated above each terminal node. A multiple-testing-adjusted P value was given for each split, specifying the strength of association between the predictor and the outcome variable conditioned on the previous splits. The fore-bag was identified as the best discriminator among all the tested variables ( $P < 0.001$ ).

*- lower than what?*

The average BMI in our population was lower than the average BMI in previous studies conducted and therefore our ECV success rate was higher.

*- you describe on line 221 a poor performance with AUC of 0.67 and a fair performance at 0.70. Are those really that different?*

We have rephrased this section in the text

*- This is called a primacy claim (your paper is the first or biggest) and must either be deleted or supported by providing the search terms used, dates, and data bases searched (Medline, Ovid, Pubmed, Google Scholar, etc) in order to substantiate your claim.*

A Pubmed and Google scholar search for the following terms between 1960 until today was performed: fore-bag, forebag, fore bag, amniotic sac, amniotic fluid preceding breech, amniotic fluid ecv. We didn't find any paper about the relationship between fore-bag and ECV.

*- or more lax abdominal walls?*

*- we*

Thank you. Changes were made.

*- please see comments above in methods about some points that need to be made in your discussion re:  
generalizability and need for external validation*

*General: In addition to the above noted things that need to be added to the discussion of limitations, please note that this is a relatively small validation set and that the low BMI limits generalizability. Did the operator know the results of the information like forebag measurement, EFW, and BMI measurement before attempting the version? If so, this could have made her or him less likely to proceed with what is turning out to be a difficult version attempt. Please also clarify the description of the criteria used to call a "failed version"=again, two of these are very subjective and, particularly if the operator was aware of the pre-version assessments, could have influenced a decision to call a failed version earlier than otherwise. (Unconsciously).*

We chose to collect data regarding eight objective parameters, the influence of which on ECV success was unknown to the examiner prior to the subsequent retrospective statistical analysis performed. While any experienced clinician is certainly influenced by their own impression of the potential success of the procedure prior to the attempt,

the provider attempted ECV in all patients who met the inclusion criteria and a small fore-bag measurement, EFW and BMI values were not among the exclusion criteria and therefore did not influence the decision to undertake ECV in the first place.

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

4. 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 have been explained

*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.*

Lee Reicher and Sharon Maslovitz entered all the data, we checked several times that the data is accurate by validation in 2 separate data bases.

6. We read the checklist thoroughly and ensured that all of the parameters were present and added any elements that were not present.

*10. Provide a précis on the second page, for use in the Table of Contents. The précis is a single sentence of no more than 25 words that states the conclusion(s) of the report (ie, the bottom line). The précis should be similar to the abstract's conclusion. Do not use commercial names, abbreviations, or acronyms in the précis. Please avoid phrases like "This paper presents" or "This case presents."*

A prediction model enables accurate prediction of successful external cephalic version at term and can be used to support patient counseling and decision making before version.

*14. Line 235: 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.*

Thank you, we decided to change this sentence.

**From:** [REDACTED]  
**To:** [Randi Zung](#)  
**Subject:** Re: Your Revised Manuscript 18-2108R1  
**Date:** Friday, January 25, 2019 2:58:55 AM  
**Attachments:** [18-2108R1 ms \(1-25-19v4\).docx](#)

---

Dear Ms Zung,

I attached the requested files

Best regards

Lee

On Wed, Jan 23, 2019 at 9:29 PM Randi Zung <[RZung@greenjournal.org](mailto:RZung@greenjournal.org)> wrote:

Dr. Reicher:

Apologies, there is one additional comment at this time. Please include an in-text citation for where Figure 6 should appear.

Thank you,

Randi

---

**From:** Randi Zung  
**Sent:** Wednesday, January 23, 2019 2:09 PM  
**To:** 'Lee Reicher' [REDACTED]  
**Subject:** RE: Your Revised Manuscript 18-2108R1

Dear Dr. Reicher:

Thank you for your responses. We need your group to complete the attached Authorship Change Form. Each author will need to complete it, and then return it to me. We need this because you added Dr. Yogev during the revision process, and he was not an original author when you first submitted the manuscript for peer review.

We also need Dr. Lavie to complete the electronic Copyright Transfer Agreement still. We can confirm receipt of the form from Dr. Isakov.

On Line 44 of the abstract, Dr. Chescheir added “sonographically measured” to the sentence and you approved it. But in the version you returned to me, it was deleted. If you approve, would you please undelete the addition?

On Line 106, please add the name of the IRB to the sentence. You cannot just say the “local” IRB. We need to know the name of the hospital.

After you have resolved these remaining queries, Dr. Chescheir will review your file.

Please make your edits to the attached manuscript file (v4).

Thank you,  
Randi

**From:** Lee Reicher [REDACTED]  
**Sent:** Wednesday, January 23, 2019 11:58 AM  
**To:** Randi Zung <[RZung@greenjournal.org](mailto:RZung@greenjournal.org)>  
**Subject:** Re: Your Revised Manuscript 18-2108R1

Dear Ms. Zung,

Thank you for your favorable consideration for publication of our manuscript entitled: "Prediction of success in external cephalic version for breech presentation at term" (ONG-18-2108). Attached hereby is our response to your queries:

DS1: Dr Yogev contributed to the study hence his name was added to the byline. We attach a file of author's acknowledgement of the addition.

DS3: STROBE checklist is attached

DS4: Ofer Isakov and Anat Lavi signed the author's agreement form. The signed agreement



was sent to Ms Randi Zung.

DS5: proofs may be sent to Lee Reicher

DS6: Running title: Predictors of successful external cephalic version

DS7: "version" was changes to "procedure" as suggested.

RZ9: 1. overall, 88 procedures were defined as failed attempts. Five due to failure to perform the version after a 30 minutes trial, 42 due to patient's request to abort following major discomfort and 39 versions were ceased by the physician. Thirty seven attempts were aborted due to inability to direct the fetus into vertex after repeated attempts lasting 10-29 minutes and 2 due to repeated decelerations in the fetal heart rate detected by the ultrasound performed by the physician during the procedure. The pre-procedure clinical characteristic of the aborted trials were unrelated to the indication to cease the attempt.

This paragraph may be added to the text if requested.

RZ9: 2. Confidence intervals for the specificity and sensitivity were added to the text.

RZ9: 3. We were not sure whether to include the 4-stage hierarchical model in the text but we will gladly add it to the text if you think it is within the scope of the journal.

NCC10: changes were made as suggested: "from February2016 to July 2018"

NCC12: all abbreviations were spelled out in the first use throughout the manuscript.

NCC13: "sonographically-measured " was added to the text as suggested.

NCC15: "Variables association with ECV success was evaluated using a multivariate logistic regression and a decision tree predicting ECV outcome was developed using 75% of the patients and validated on the remaining 25%." – added to Methods as suggested.

RZ16: the need for external validation is discussed in Discussion: "External validation of a prediction model is essential for its implementation in clinical practice. The only model whose performance has been tested in both internal and external validation thus far was the one developed by Kok et al. [8]. Unfortunately, external validation of their model had a slight underestimation of the probability for a successful ECV. We recognize that our model needs external validation before it can be used in clinical practice."

NCC17: CS was replaced by "cesarean delivery" as suggested.

NCC19: we added a paragraph re the range of success: " The effectiveness of ECV in reducing the rates of cesarean deliveries is supported by a 2015 systematic review of eight randomized trials of ECV at term (n = 1308 women). Compared with women with breech fetuses who had no attempt at ECV, women who attempted ECV reduced the risk of cesarean delivery by approximately 40 percent (RR 0.57, 95% CI 0.40-0.82)."

DS20: IRB name was added as requested.

NCC21: we corrected the terms used in the text according to the revitalize definitions.

NCC23: out intention was to specify that the performer of the versions was not aware of the

data previously collected thus was not influenced by it.

NCC24: Data was collected by the computerized medical files of all patients. Candidates are referred for ECV attempt by care-givers in the community without specifying the institute. They usually have the procedure done at the maternity unit in which they intend to be delivered. All of our patients delivered at our maternity center.

NCC26: redundant parenthesis was removed

NCC28: your correction is correct and added to the text: "The parameter setting with the best mean performance estimated by cross validation was used in the final decision tree model which was then used for internal validation on the remaining 25% of participants"

NCC30: confidence intervals were added as requested.

NCC32: the paragraph relating to the split between the training and the test groups was removed from the Result section and added to the Methods as suggested.

NCC34: we elaborated on using the cutoffs for the fore-bag size and BMI and added this paragraph to the text: "A one centimeter increase in Fore-bag size was associated with better odds of successful ECV attempt (OR 153 [31.5-1439.4]). A one unit increase in BMI ( $\text{kg/m}^2$ ) results in reduced odds of success (OR 0.6 [0.4-0.8]) while the number of previous deliveries increases the odds (OR 6 [2.6-17.3]).

NCC35: A univariate analysis was performed to estimate the association of each variable with the response. The variable with the strongest association is selected and a binary split is implemented.

We were not sure whether to include the next paragraph in the text but we will gladly add it

to the text if you think it is within the scope of the journal:

"A variable is chosen to be incorporated into the decision tree only if it is significantly associated with the outcome. The tree is grown by identifying the variable with the strongest association to the response. If the association is significant, a binary split is implemented. This process is performed recursively for each subpopulation generated by the previous split until none of the variables are significantly associated with the response."

NCC36: an image of the sonographic appearance of a fore-bag was added (Fig 6)

NCC37 : the virgule "/" was removed

NCC38: changed as offered to: "Clinical prognostic factors which have been associated with successful ECVs...."

NCC40: changes as suggested to ".... placental location, cervical dilatation, and station of presentation."

NCC41: changed as offered to: "This may be due to late engagement (often after the onset of labor) or relative laxity of the maternal abdomen in multiparous women"

DS43: the reference was corrected to ACOG publication as suggested.

Best regards,

Lee

On Fri, Jan 18, 2019 at 5:12 PM Randi Zung <[RZung@greenjournal.org](mailto:RZung@greenjournal.org)> wrote:

Dear Dr. Maslovitz and Dr. Reicher:

Your revised manuscript is being reviewed by the Editors. Before a final decision can be made, we need you to address the following queries. Please make the requested changes to the latest version of your manuscript that is attached to this email. **Please track your changes and leave the ones made by the Editorial Office.** Please also note your responses to the author queries in your email message back to me.

1. General: The Editor has made edits to the manuscript using track changes. Please review them to make sure they are correct.

2. Authorship Change: There is an author listed in your record in Editorial Manager who is not listed in the byline: Yariv Yogev. Should Dr. Yogev be listed in the byline? If so, we will require each author to complete the attached form acknowledging this addition. Please also add Dr. Yogev to the byline in the text.

3. Please submit a completed STROBE checklist. A link to the checklist is available at:  
<https://www.editorialmanager.com/ong/default.aspx>

4. Electronic Copyright Agreement: Ofer Isakov and Anat Lavie need to complete our electronic Copyright Transfer Agreement, which was sent to them through Editorial Manager ([EM@greenjournal.org](mailto:EM@greenjournal.org)). Please have them contact me to have the email resent to them if they did not receive it.

5. Corresponding Author Information: Lee Reicher is the corresponding author in Editorial Manager. If your paper is accepted, he will receive the proof. Is this okay, or do you want to change the corresponding author to Sharon Maslovitz?

6. Running title: Please provide a running title of about 45 characters (including spaces).

7. Precis: Please note the edit to this sentence. Do you approve?

8. Line 34 (General comments about the statistics that should be addressed in your abstract, body, and tables as needed): As 2 of your criteria for calling a failed procedure are important but based on human decisions (patient request, provider decision) it is important to enumerate how many failures were based on which reasons. This is particularly important as the provider decision could be based on his knowledge that obese women have a higher failure rate and so he may decide to stop sooner. Please provide the number of failures based on each indication for failure.

Thank you for including the sensitivities, specificities etc, but they all need CIs throughout the submission. Please add these.

The discussion re: the 4-stage hierarchical model is important, thank you for adding it. If you want to include it in the main text, I think that would be appropriate. If not, it should be available as supplemental material for the interested reader.

9. Line 37: Please confirm that this is correct. In your response letter you indicated that these dates were inclusive, so I've changed it to "From" rather than "between".

10. Line 39: Spell out all abbreviations on first use; will need to also be done in the manuscript. Includes BMI, AFI and any others used.

11. Line 41: Please approve or disapprove this additional edit

12. Abstract-Methods: Your abstract methods do not describe the use of a test set and a validation set and needs to do .You mention this in the results but all of that is methods

13. Line 56-57: Please note this comment from the Editor and make the necessary edit to this section and in your manuscript body text.

14. Starting at Line 63: please replace cesarean section with either cesarean birth or cesarean delivery .CS is not an acceptable abbreviation

15. Line 67: Rather than jump into the problems with ECV, perhaps better to explain range of successes first . Also, is there data to suggest how many CS are averted in populations in which ECV is performed

16. Line 92: Please add the name of the IRB to this sentence.

17. Line 96 and elsewhere: Please make sure you are using the reVITALize definitions. This is now known as perlabor rupture of the membranes

They are available at <https://www.acog.org/About-ACOG/ACOG-Departments/Patient-Safety-and-Quality-Improvement/reVITALize?IsMobileSet=false>

18. Line 108: What do you mean by "real time" . "Do you mean analysis of factors measured prior to the attempt

19. Line 116: How was delivery data collected? Since these were referred patients, did some deliver elsewhere?

20. Line 120: There is either missing one parantheses or you don't need one. Please correct this as needed.

21. Line 161-162: Please confirm that the edit made to this sentence is correct.

22. Results: We ask that you provide CI's in lieu of P values. Please make this edit for all of the data you are presenting in this section.

23. Line 187-191 (Sentences highlighted in yellow): Please move these lines to the Methods section.

24. Line 209: I don't understand what you mean be Fore-bag size (cm) .Can you please state something about how these are associated (ie, positively or negatively) .<Were these all analyzed as continuous variables for this or was there a cut off for fore bag size and BMI (ie a binary variable)

25. Line 225: Again, please provide CIs, not just a p-value.

26. Discussion: Would you please provide an image of the forebag measurement? This will need to be included as a figure in the manuscript. Please cite it in the text (renumbering your existing figures if needed) and send your figure file back with your updated manuscript.

27. Line 245: The Journal style doesn't not use the virgule (/) except in numeric expressions. Please edit here and in all instances. Should this be "and" or "or"?

28. Line 246: that" are "associated or which have been associated" ?Are "makes is sound like the results of your analysis

29. Line 255: height of presentation "doesn't make sense. Do you mean degree of non engagement or station

30. Line 304: Would you consider adding, "or relative laxity of the maternal abdomen" to this sentence?

31. Reference 21: Please note the edit to this reference. Did you intend to cite the ACOG publication?

To facilitate the review process, we would appreciate receiving a response by January 22.

Best,

Randi Zung

--

**Randi Zung (Ms.)**

Editorial Administrator | *Obstetrics & Gynecology*

American College of Obstetricians and Gynecologists

409 12th Street, SW

Washington, DC 20024-2188

T: 202-314-2341 | F: 202-479-0830

<http://www.greenjournal.org>



**From:** [REDACTED]  
**To:** [Stephanie Casway](mailto:Stephanie.Casway)  
**Subject:** Re: O&G Figure Revision: 18-2108  
**Date:** Tuesday, January 15, 2019 3:10:26 PM

---

Dear Stephanie,  
We agree to all the changes mention above.

best wishes,  
Lee

On Mon, Jan 14, 2019 at 6:54 PM Stephanie Casway <[SCasway@greenjournal.org](mailto:SCasway@greenjournal.org)> wrote:

Good Morning Dr. Reicher,

Your figures and legend have been edited, and PDFs of the figures and legend are attached for your review. Please review the figures CAREFULLY for any mistakes. In addition, please see our queries below.

AQ1: Note that we did edit terminology in the legend. If these edits are incorrect, please let me know.

AQ2: We have added y-axis labels. If these are incorrect, please let me know.

PLEASE NOTE: Any changes to the figures must be made now. Changes at later stages are expensive and time-consuming and may result in the delay of your article's publication.

To avoid a delay, I would be grateful to receive a reply no later than Wednesday, 1/16.  
Thank you for your help.

Best wishes,

Stephanie Casway, MA  
Senior Production Editor

*Obstetrics & Gynecology*  
American College of Obstetricians and Gynecologists  
409 12th St, SW

Washington, DC 20024

Ph: (202) 314-2339

Fax: (202) 479-0830

[scasway@greenjournal.org](mailto:scasway@greenjournal.org)

**From:** [REDACTED]  
**To:** [Stephanie Casway](mailto:SCasway@greenjournal.org)  
**Subject:** Re: O&G Figure Query: 18-2108  
**Date:** Friday, January 25, 2019 6:14:59 PM

---

Those files are fine.

Thanks,

Lee

ב-25 בינוי 2019, בשעה 21:50, Stephanie Casway <[SCasway@greenjournal.org](mailto:SCasway@greenjournal.org)> כתב/ה:

Hi again Lee,

Attached you will find updated versions of Figure 6 and the legend. Please let me know if these files are okay, or if any edits are needed. Thank you so much for your help!

**From:** Lee Reicher [REDACTED]  
**Sent:** Friday, January 25, 2019 2:50 AM  
**To:** Stephanie Casway <[SCasway@greenjournal.org](mailto:SCasway@greenjournal.org)>  
**Subject:** Re: O&G Figure Query: 18-2108

Dear Ms. Casway,

Sonographic appears of a fore-bag is acceptable

Best regards,

Lee

On Wed, Jan 23, 2019 at 9:27 PM Stephanie Casway <[SCasway@greenjournal.org](mailto:SCasway@greenjournal.org)> wrote:

Good Afternoon Lee,

Thank you for sending Figure 6, we are currently reviewing the figure. Would you be able to send me a version without any text or lines? These items will be added back per journal style. Additionally, did you have a legend you would like us to use, or would "Sonographic appears of a fore-bag" be acceptable?

Thank you so much for your help!

Stephanie Casway, MA  
Senior Production Editor  
*Obstetrics & Gynecology*  
American College of Obstetricians and Gynecologists  
409 12th St, SW  
Washington, DC 20024

Ph: (202) 314-2339

Fax: (202) 479-0830

[scasway@greenjournal.org](mailto:scasway@greenjournal.org)

<18-2108 Fig 6 (1-25-19 v2).pdf>

<18-2108 Legend.pdf>