

## PEER REVIEW HISTORY

BMJ Paediatrics Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

### ARTICLE DETAILS

<b>TITLE (PROVISIONAL)</b>	Predictors of extubation success in infants considered extubation ready: A population-based study of infants below a gestational age of 26 weeks
<b>AUTHORS</b>	Ohnstad, Mari Oma Stensvold, Hans Jørgen Pripp, Are Hugo Tvedt, Christine Raaen Jelsness-Jørgensen, Lars-Petter Astrup, Henriette Eriksen, Beate Horsberg Klingenberg, Claus Mreihil, Khalaf Pedersen, Tanja Rettedal, Siren Selberg, Terje Reidar Solberg, Rønnaug Støen, Ragnhild Rønnestad, Arild E

### VERSION 1 – REVIEW

<b>REVIEWER</b>	Reviewer name: Dr. Pracha Nuntnarumit Institution and Country: Mahidol Univ, Ireland Competing interests: None
<b>REVIEW RETURNED</b>	06-Jun-2022

<b>GENERAL COMMENTS</b>	<p>The authors report the results of a national retrospective cohort study to determine the characteristics that predict successful first-time extubation in extremely preterm infants born before 26 weeks of gestation. The major findings are that a pre-extubation FiO<sub>2</sub> of less than 0.35, higher 5-minute Apgar scores, greater GA, female gender, and higher postnatal age at extubation are associated with successful extubation.</p> <p>Overall this is a well written and well organized manuscript. There are few questions as follows:</p> <p>Methods:</p> <p>Did the participating 10 NICUs have their own extubation criteria? If each participating NICU had its own criteria/strategy for extubation during the study, is it possible that it would have an impact on the outcomes measured.</p> <p>Results:</p> <p>The authors mentioned that success rate using the 7-day definition decreased from 55% to 44%.</p> <p>Can the authors provide further information on the predictive factors for successful extubation at day 7 after extubation. Did the characteristics that predicted successful extubation on day 3 hold true on day 7?</p> <p>Table 2. showed that postnatal age (PNA) at extubation of infants with successful extubation were lower than those failed extubation (median (IQR) 5 (2-19) vs. 7 (3-16) in CV group, p-value 0.14). I'm</p>
-------------------------	--

	curious as to why a higher PNA at extubation was found to be a predictor of successful extubation in a multivariable analysis.
--	--

<b>REVIEWER</b>	Reviewer name: Dr. Sanjay Chawla Institution and Country: Wayne State Univ, Ireland Competing interests: None
<b>REVIEW RETURNED</b>	02-Jun-2022

<b>GENERAL COMMENTS</b>	<p>Mari and colleagues report data on extubation readiness for extreme preterm infants. The aim of the study was to investigate first extubation attempts among extremely premature (EP) infants and to explore factors that may increase the quality of clinical judgement of extubation readiness.</p> <p>I have following comments:</p> <ol style="list-style-type: none"> <li>1. Overall the research question is interesting. Even though few studies have been done in this area and authors acknowledge and present this in the background and provide rationale to do the study in current population.</li> <li>2. Successful extubation was defined as no re-intubation within 72 hours after extubation. Most experts in this area advocate to use 5 to 7 days for this definition.</li> <li>3. The background and flowchart of patients has been well presented in the figure.</li> <li>4. Why did you consider FiO2 and Apgar score as categorical variables instead of continuous variables?</li> <li>5. Please provide details on how internal validation was done</li> <li>6. Please give details on what factors were considered in the model to begin with? How were these variables chosen? Were these decided a priori based on clinical significance?</li> <li>7. Discussion should include limitations of observational design of the study which can affect the results as only patients considered ready for extubation by clinicians were included.</li> </ol>
-------------------------	--

<b>REVIEWER</b>	Reviewer name: Dr. Peter Flom Institution and Country: Peter Flom Consulting, United States Competing interests: None
<b>REVIEW RETURNED</b>	16-Jun-2022

<b>GENERAL COMMENTS</b>	<p>I confine my remarks to statistical aspects of this paper. The basic approach is fine, but I have some issues to resolve before I can recommend publication.</p> <p>Abstract: Do not rely so much on stat. sig. Look at effect size. Stat sig is conflated with sample size, and, HFOV had a much smaller sample.</p> <p>p. 6 line 23 ff: Good. A priori reasoning is the best way to build a model.</p> <p>line 54 Is APGAR score "subjective"? I am not a neonatologist, but most of the components seem pretty objective to me.</p> <p>p 7 line 58 Why were these done separately? Why not include this as a variable?</p> <p>p. 8 line 15 Backward selection (also forward and stepwise) cannot be recommended. The standard errors will be too small, the p values too low, and the parameter estimates biased away from 0. For details, examples, and proofs, see Harrell *Regression</p>
-------------------------	--

	<p>Modeling Strategies*. In addition, nonsignificant variables may be good mediators (control variables). Since you chose variables for good reasons, you should look at them all.</p> <p>p, 10 line 54 What is "optimism corrected R2" ? I've never heard of it. Some explanation would be good. But R2 measures do not really exist for logistic regression. There are various pseudo R2, but none are perfect.</p> <p>Peter Flom</p>
--	---

## VERSION 1 – AUTHOR RESPONSE

Response to the Editors and reviewers of the Manuscript: bmjpo-2022-001542 "Predictors of extubation success in infants considered extubation ready:

A population-based study of infants below a gestational age of 26 weeks"

Date submitted: 11/05/2022

Dear Editors and reviewers,

Thank you for valuable contributions to the article, and the effort and time spent during the review. We sincerely believe your comments have helped us in improving our manuscript. Our responses to the comments are listed point by point below, and the revisions are highlighted in the article using the track and change function (document: "Main Document - marked copy").

Response to comments to author from the Editor in Chief

Comment number 1

"Title delete "in infants considered extubation ready""

"Title replace "infants" with "neonates""

Reply: We have made the change. We changed the title to: "Predictors of extubation success:

A population-based study of neonates below a gestational age of 26 weeks".

Please see page 1.

## Response to comments to author from the Associate Editor

### Comment number 1

“The term ‘gender’ is used in the abstract and manuscript, I assume the authors mean sex instead of gender?”

Reply: The assumption is correct, and we have replaced “gender” with “sex” throughout the manuscript.

### Comment number 2

“The CRIBII score was included as a variable scribing illness severity score. However, was infection as variable also assessed separately?”

Reply: We thank the Associate Editor for pointing this out. Maternal factors such as preeclampsia, HELLP-syndrome, chorioamnionitis and prolonged rupture of membranes may certainly be of importance for the infant’s respiratory trajectory after birth. The Norwegian Neonatal Network (NNN) consist of clinical data for infants admitted to a Norwegian neonatal unit and does not contain complete data related to maternal health. Practically all extremely premature infants born below 26 weeks GA are treated with antibiotics the first 1-3 days. There is uncertainty related to the validity of the diagnosis of the perinatal sepsis-variable. We have included this important aspect to the study limitation in the manuscript. Please see page 13.

### Comment number 3

“The ventilation index and respirator severity score were calculated. The authors added references for this, but it might be relevant for the reader to shortly explain these parameter more clearly. Can the authors expand on this?”

Reply: We agree with the Associate Editor and have included a short explanation of these two parameters to clarify. Please see page 7.

### Comment number 4

“The use of respiratory stimulating drugs in the unit, and how this was in the study cohort is lacking. Was caffeine used (if yes, which doses received the different groups), was theophylline, doxapram or other compounds used?”

Reply: We agree that respiratory stimulating drugs are important when considering extubation outcome in this population. In table 2 we have provided information about the administration of caffeine at the day of extubation and doses by extubation outcome. In this study cohort practically, all infants received caffeine at the day of extubation. Theophylline, doxapram or other compounds are rarely used in the Norwegian neonatal units. Doxapram was infrequently used in a few of the participating units, but this variable was not obtained as part of the data collection in this study. Please see table 2 for information about administration of caffeine in the study cohort.

#### Comment number 5

“In the discussion, the authors mention the use of corticosteroids on the day of extubation. Can they provide more details on this topic? What is their general protocol in the units? Did practices differ? How was the use in the study population (frequency, dosing, etc). It might be useful for the readers to provide more details on this topic.”

Reply: We agree and have provided more details on the common use of postnatal corticosteroid therapy in the Norwegian NICUs. Please see page 13, at the end of first section.

Response to reviewer: 1 Dr. Sanjay Chawla, Wayne State Univ

#### Comment number 1

“Overall the research question is interesting. Even though few studies have been done in this area and authors acknowledge and present this in the background and provide rationale to do the study in current population.”

Reply: We thank the reviewer for the positive feedback and is pleased that the rationale for conducting the study is clear.

#### Comment number 2

“Successful extubation was defined as no re-intubation within 72 hours after extubation. Most experts in this area advocate to use 5 to 7 days for this definition.”

Reply: We acknowledge the comment by the reviewer and are aware of the heterogeneity regarding the definition of extubation success/failure ranging from 2-7 days after extubation. We chose a definition of 72 hours without the need of re-intubation to prevent inclusion of re-intubations caused by other factors than respiratory distress syndrome (RDS). In addition, previous results from this study-population comparing the 72-hour and 7-day time frames, postnatal age at the first successful extubation was equal in 88, 80, and 82% of the infants born at GA 22–23, 24, and 25 weeks, respectively (Ohnstad, 2020) . Furthermore, another study showed that among neonates who experienced extubation failure within 5 days, most (75%) were re-intubated less than 2 days after extubation (Chawla, 2017). A recent review from Shalish et al (2022) recommend reporting extubation failure rates at 48-72 hours and at 7 days postextubation. Based on this recommendation, we included success rates for 7 days without the need of re-intubation.

Comment number 3

“The background and flowchart of patients has been well presented in the figure.”

Reply: We thank the reviewer for the positive feedback.

Comment number 4

“Why did you consider FiO2 and Apgar score as categorical variables instead of continuous variables?”

Reply: We thank the reviewer for pointing this out. For the FiO2 variable we considered that dichotomous variables were more clinically relevant than a continuous variable. We considered the Apgar score as a nominal categorical variable. The cut off set at Apgar 5 at 5 minutes of age was also considered based on clinical relevance.

Comment number 5

“Please provide details on how internal validation was done”

Reply:

We thank the reviewer for the comment. All multivariable linear regression models were internally validated by bootstrapping using 1000 bootstrap samples to assess overfitting. Overfitting is a statistical term to describe a better model performance in the development sample than in new

samples with other subjects. The calibration slope from the internal validation with bootstrapping (where 1.00 is ideal and estimates below 1.00 indicate overfitting) assesses the degree of overfitting. It was used to provide shrinkage factors for adjusting regression coefficients and adjusted model intercepts for use in prediction formulas. A prediction model with adjusted regression coefficients (shrinkage) has according to statistical theory a better performance in a new sample than the regression model with the original coefficients. Further, from the results after internal validation we estimate the so-called optimism-corrected model performance measures. The statistical literature use several terms to describe model performance assessed by internal validation:

- Apparent model performance is estimated directly from dataset that was used to develop prediction model. This is model characteristics (for example area under the curve /c-statistics or pseudo-R<sup>2</sup>) from the original regression analysis in the study sample.
- Training model performance is the average performance in each of the bootstrap samples with replacement (for example the average area under the curve from the 1000 bootstrapped samples)
- Test model performance is the average performance of the models from the bootstrap samples when applied to the original study sample.
- Optimism is the average difference between model performance in bootstrap data and test performance in original dataset.
- Optimism-corrected model performance is obtained by subtracting the average optimism after internal validation from the apparent performance.

Internal validation was conducted using the R package rms. We also repeated the internal validation calculation with the user-developed stata command bsvalidation (Fernandez-Felix BM, García-Esquinas E, Muriel A, Royuela A, Zamora J. Bootstrap internal validation command for predictive logistic regression models. The Stata Journal. 2021;21(2):498-509), that gave very similar results.

Two general references on internal validation of prediction models:

Steyerberg EW. Clinical prediction models: a practical approach to development, validation, and updating. 2nd ed: Springer; 2019.

Steyerberg EW, Bleeker SE, Moll HA, Grobbee DE, Moons KG. Internal and external validation of predictive models: a simulation study of bias and precision in small samples. J Clin Epidemiol. 2003;56(5):441-447.

Based on the reviewer's comment we have revised the section describing internal validation. Please see page 8.

Comment number 6

“Please give details on what factors were considered in the model to begin with? How were these variables chosen? Were these decided a priori based on clinical significance?”

Reply: We agree with the reviewer that how the candidates were chosen for the model needs to be clarified in the manuscript. We have revised the section describing the logistic regression modeling. Please see page 8.

Comment number 7

“Discussion should include limitations of observational design of the study which can affect the results as only patients considered ready for extubation by clinicians were included.”

Reply: We agree that this is an important point regarding the overall study design including extubation attempts when clinicians considered the patient extubation ready. We have further specified this in the discussion. Please see page 13.

Response to reviewer: 2 Dr. Pracha Nuntnarumit, Mahidol Univ

Comment number 1

“Methods:

Did the participating 10 NICUs have their own extubation criteria?

If each participating NICU had its own criteria/strategy for extubation during the study, is it possible that it would have an impact on the outcomes measured.”

Reply: We agree with the reviewer that different extubation criteria among each participating unit would possibly have an impact on the outcomes measured. We have made changes to clarify in the section where we describe the extubation criteria reported from the participating units. Please see page 9.

Comment number 2

“Results:

The authors mentioned that success rate using the 7-day definition decreased from 55% to 44%.



Can the authors provide further information on the predictive factors for successful extubation at day 7 after extubation. Did the characteristics that predicted successful extubation on day 3 hold true on day 7?"

Reply: We appreciate the reviewer's insightful suggestion and agree that it would be useful to demonstrate whether the characteristics that predict successful extubation on day 3 hold true in day 7. However, such an analysis is beyond the scope of our paper, which aimed to investigate predictors of first extubation attempt using the 72-hour definition. We chose a definition of 72 hours without the need of re-intubation to prevent inclusion of re-intubations caused by other factors than respiratory distress syndrome (RDS). However, we decided to report success rates using the 7-day definition based on recommendations from the recent published review from Shalish et al (2022), but we did not further investigate prediction models with the 7-day definition.

#### Comment number 3

"Table 2. showed that postnatal age (PNA) at extubation of infants with successful extubation were lower than those failed extubation (median (IQR) 5 (2-19) vs.7 (3-16) in CV group, p-value 0.14). I'm curious as to why a higher PNA at extubation was found to be a predictor of successful extubation in a multivariable analysis."

Reply: This observation is correct. The unadjusted results of postnatal age (PNA) show a non-significant difference between the infants failed and succeeded the extubation attempt. In the multivariable analysis the PNA is adjusted for multiple factors, and this is probably the reason why a higher PNA at extubation turns out to be a predictive factor in this study-cohort. Estimated coefficients from univariable and multivariable regression models are only identical if the independent (predictor) variables are uncorrelated. Thus, depending on the correlation structure between the variables, variable with significant coefficients in a univariable model could be non-significant in a multivariable model and visa versa.

Response to reviewer: 3 Dr. Peter Flom, Peter Flom Consulting

#### Comment number 1

"Abstract:

Do not rely so much on stat. sig. Look at effect size. Stat sig is conflated with sample size, and, HFOV had a much smaller sample."

Reply:

Yes, we agree with the reviewer and emphasize more the effect size and statistical significance. As stated, the p-value (and statistical significance) is dependent on the sample size. We have made revision in the abstract. Please see page 4.

Comment number 2

“p. 6 line 23 ff: Good. A priori reasoning is the best way to build a model.”

Reply:

Yes, we agree with the comment from the reviewer. We have tried to emphasize a priori reasoning compared to a data-driven approach in model building.

Comment number 3

“line 54 Is APGAR score "subjective"? I am not a neonatologist, but most of the components seem pretty objective to me.”

Reply: We thank the reviewer for the comment. The APGAR score characterizes the infant's physiologic condition immediately after birth, and is scored by either the physician, midwife or other available health personnel with sufficient competence. Elements of the score, such as tone, color and reflex irritability can be subjective and previous studies have identified a substantial interobserver (O'Donnell, 2006) and international (Siddiqui, 2017) variability in assessing the APGAR score.

Comment number 4

“p 7 line 58 Why were these done separately? Why not include this as a variable?”

Reply: We thank the reviewer for the comment with suggestion of including the ventilation mode as a variable. However, indications of ventilatory support and weaning strategies are principally different in CV compared to HFOV. In addition, HFOV is commonly used as a rescue treatment for infants where CV does not provide sufficient respiratory support which means that these two groups of infants differ considerably. We could have excluded infants extubated from HFOV. Nevertheless, we have chosen to present the results based on the relatively high number of infants extubated from HFOV compared to previous studies in this field not being able to describe extubation outcomes for these infants (Kidman, 2021). We recognize that this aspect should be highlighted in

the manuscript, so we clarified this in the discussion. Please see page 13, last section in the discussion.

#### Comment number 5

“p. 8 line 15 Backward selection (also forward and stepwise) cannot be recommended. The standard errors will be too small, the p values too low, and the parameter estimates biased away from 0. For details, examples, and proofs, see Harrell *\*Regression Modeling Strategies\**. In addition, nonsignificant variables may be good mediators (control variables). Since you chose variables for good reasons, you should look at them all.”

#### Reply:

The rms package developed by Frank Harrell has implemented commands for backward elimination. However, as pointed out by the reviewer, statisticians dispute about the characteristics and usefulness of this procedure and other stepwise regression approaches. We assessed different statistical results during model building, but did not do a formal stepwise regression in the multivariable analysis. Therefore, we have removed the text on “backward selection method” in the revised version since our final models since it is somewhat misleading.

#### Comment number 6

“p, 10 line 54 What is “optimism corrected R<sup>2</sup>”? I've never heard of it. Some explanation would be good. But R<sup>2</sup> measures do not really exist for logistic regression. There are various pseudo R<sup>2</sup>, but none are perfect.”

#### Reply:

Please see above our response at comment number 5 to reviewer number 1 (Dr. Sanjay Chawla, Wayne State Univ) about the different statistical terms and estimates from internal validation of prediction models. The statistical terminology is somewhat technical and we suggest readers in the revised version of the manuscript to consult the references for a more in-depth explanation.

The Nagelkerke R<sup>2</sup> is estimated from logistic regression. This is also specified in the revised version instead of using the term “pseudo R-squared”. The use and interpretation is not equal to R<sup>2</sup> from linear regression, but we think it provides a useful assessment of model performance. It is commonly estimated in many statistical packages.

Finally, we would like to thank the reviewers again for taking time to review our manuscript.

Sincerely,

The authors of manuscript

#### References:

Ohnstad et al. (2020) Duration of Mechanical Ventilation and Extubation Success among Extremely Premature Infants. *Neonatology*. DOI: 10.1159/000513329

Chawla et al. (2017) Markers of Successful Extubation in Extremely Preterm Infants, and Morbidity After Failed Extubation. *Journal of Pediatrics*. DOI: 10.1016/j.jpeds.2017.04.050

Shalish et al (2022). Decision to extubate extremely preterm infants: art, science or gamble? *Archives of Disease in Childhood: Fetal and Neonatal Edition*

Fernandez-Felix BM, García-Esquinas E, Muriel A, Royuela A, Zamora J. Bootstrap internal validation command for predictive logistic regression models. *The Stata Journal*. 2021;21(2):498-509

Steyerberg EW. *Clinical prediction models: a practical approach to development, validation, and updating*. 2nd ed: Springer; 2019.

Steyerberg EW, Bleeker SE, Moll HA, Grobbee DE, Moons KG. Internal and external validation of predictive models: a simulation study of bias and precision in small samples. *J Clin Epidemiol*. 2003;56(5):441-447.

O'Donnell CPF, Kamlin COF, Davis PG, Carlin JB, Morley CJ. Interobserver variability of the 5-minute Apgar score. J Pediatr 2006; 149: 486-9.

Siddiqui A, Cuttini M, Wood R, et al. Can the Apgar score be used for international comparisons of newborn health? Paediatr Perinat Epidemiol 2017; 31: 338-45.

Kidman et al (2021). Predictors and outcomes of extubation failure in extremely preterm infants. Journal of Paediatrics and Child Health. DOI: 10.1111/JPC.15356

#### VERSION 2 – REVIEW

<b>REVIEWER</b>	Reviewer name: Dr. Pracha Nuntnarumit Institution and Country: Mahidol Univ, Ireland Competing interests: None
<b>REVIEW RETURNED</b>	14-Jul-2022
<b>GENERAL COMMENTS</b>	The authors have answered all the questions clearly. I recommend to accept the manuscript for publication

<b>REVIEWER</b>	Reviewer name: Dr. Peter Flom Institution and Country: Peter Flom Consulting, United States Competing interests: None
<b>REVIEW RETURNED</b>	10-Jul-2022
<b>GENERAL COMMENTS</b>	The authors have addressed my concerns and I now recommend publication  Peter Flom