

# Evaluation and Validation of a Prediction Model for Extubation Success in Very Preterm Infants

**Rebecca Dryer**

Emory University School of Medicine

**Anand Salem**

Emory University School of Medicine and Children's Healthcare of Atlanta

**Vivek Saroha**

Emory University School of Medicine and Children's Healthcare of Atlanta

**Rachel Greenberg**

Duke University <https://orcid.org/0000-0003-4156-8543>

**Matthew Rysavy**

University of Texas Health Science Center at Houston

**Sanjay Chawla**

Emory University School of Medicine and Children's Healthcare of Atlanta <https://orcid.org/0000-0002-7299-2737>

**Ravi Patel** (✉ [rmpatel@emory.edu](mailto:rmpatel@emory.edu))

Emory University School of Medicine and Children's Healthcare of Atlanta <https://orcid.org/0000-0002-7444-5550>

---

## Article

**Keywords:** IQR, AUROC, bronchopulmonary dysplasia (BPD), TRIPOD

**Posted Date:** September 15th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-894089/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at Journal of Perinatology on September 24th, 2022. See the published version at <https://doi.org/10.1038/s41372-022-01517-z>.

# Abstract

## Objective

To evaluate the performance of a publicly available model predicting extubation success in very preterm infants.

## Study Design

Retrospective study of infants born < 1250 g at a single center. Model performance evaluated using the area under the receiver operating curve (AUROC) and comparing observed and expected probabilities of extubation success, defined as survival  $\geq$  5 d without an endotracheal tube.

## Results

Of 177 infants, 120 (68%) were extubated successfully. The median (IQR) gestational age was 27 weeks (25–28) and weight at extubation was 915 g (755–1050). The model had acceptable discrimination (AUROC 0.72 [95% CI 0.65–0.80]) and adequate calibration (calibration slope 0.96, intercept – 0.06, mean observed-to-expected difference in probability of extubation success – 0.08 [95% CI -0.01, -0.15]).

## Conclusions

The extubation success prediction model has acceptable performance in an external cohort, supporting its potential utility in clinical decision-making. Additional studies are needed to determine if its use can improve outcomes.

# Introduction

Invasive mechanical ventilation in very preterm infants is associated with adverse outcomes such as bronchopulmonary dysplasia (BPD) and neurodevelopmental impairment(1). Reducing the duration of invasive mechanical ventilation by early extubation may reduce these risks, but extubation failure is associated with an increased risk of BPD, death and intracranial hemorrhage(2, 3).

Recently, a model to estimate the probability of successful extubation in very preterm infants was developed(3) and made available online at [www.extubation.net](http://www.extubation.net). The model includes six variables associated with extubation success: gestational age at birth, chronological age at extubation, pre-extubation blood gas pH, pre-extubation FiO<sub>2</sub>, and highest respiratory severity score (RSS) in the first 6 hours of age. The area under the receiver operating characteristic curve (AUROC) for the model was 0.77 in the development cohort. However, the performance of the model in external settings is unknown.

We aimed to assess the performance of the extubation success prediction model in a cohort of infants cared for at an academic U.S. neonatal intensive care unit (NICU). Additionally, we aimed to evaluate the

performance of alternative prediction models and examine differences in characteristics and outcomes between infants with extubation success, compared to failure.

## Methods

### Study design

We conducted a retrospective observational study that included intubated infants with a birth weight < 1250 g admitted from 8/1/08 – 7/31/17 at Emory University Hospital Midtown in Atlanta, Georgia, with initial extubation occurring within 60 days of birth. These criteria were used to match the selection criteria of the cohort in which the model was derived(3). Infants with first extubation after 60 postnatal days, prior extubation, a first extubation that was unplanned, or with data missing for the extubation model or outcome were excluded. The study was approved by the Emory University Institutional Review Board with a waiver of informed consent and was performed in accordance with the Declaration of Helsinki. The TRIPOD Checklist for Prediction Model Validation was used to ensure transparent reporting of findings(4) (Supplementary Table 1).

### Predictors and outcome

Data were collected on the six predictors in the extubation success calculator: gestational age at birth, chronological age at extubation, weight at extubation, pre-extubation FiO<sub>2</sub>, pre-extubation pH (closest to extubation, regardless of timing), and the highest respiratory severity score (RSS) in the first six hours of age. RSS was calculated as the product of the mean airway pressure and FiO<sub>2</sub>. Additional data available on infant and maternal demographics, surfactant and caffeine exposure, and several aspects of pre- and post- extubation ventilatory support were collected. Extubation success was defined as survival for ≥ 5 days without an endotracheal tube, consistent with the derivation cohort(3). BPD was graded according to severity using published criteria(5). Retinopathy of prematurity receiving treatment was defining as receipt of laser or intravitreal anti-VEGF therapy. Severe intraventricular hemorrhage, defined as grade 3 or 4, and periventricular leukomalacia were ascertained based on ultrasound reports by a pediatric neuroradiologist. Necrotizing enterocolitis was defined as modified Bell Stage IIA or greater. No blinding of predictor or outcome assessment was performed.

### Statistical methods

The sample size was pragmatic with an intent to have at least 10 outcome events per parameter in the extubation model. No imputation for missing data was performed. To assess model performance, we generated probabilities of extubation success using the parameters from the published extubation success calculator, available at [www.extubation.net](http://www.extubation.net). Next, we generated receiver operating characteristic curves using predicted probabilities and calculated the AUROC as a measure of model discrimination for extubation success. We also created calibration plots to compare predicted vs. observed probabilities of extubation success by deciles of predicted probability(6, 7). Within each decile, the observed proportion of infants with extubation success and corresponding 95% binomial confidence intervals were reported. The

calibration plot included a fitted linear model slope and intercept. In addition, we estimated the observed-to-expected mean difference in extubation success.

To determine how model parameters compared from the original published extubation success calculator to our cohort, we recalibrated the Gupta et al. logistic regression model for our cohort and compared odds ratios of model parameters from the initial derivation cohort(3) to those generated from this validation cohort. Subsequently, we evaluated how a more parsimonious model or expanded model changed discrimination, by comparing the AUROC among different models that removed variables with odds ratio estimates that were different from the derivation cohort (Reduced model) or included variables that were significantly different between extubation success groups that were not in the original model (Expanded model). We also included the evaluation of a recently published two-variable model that included gestational age at birth and pre-extubation MAP(8), as well as gestational age only, given the strong prognostic influence of gestational age on neonatal outcomes and the goal to determine the incremental prognostic value of the other variables in the calculator.

To explore optimal cut-offs of predicted probability of extubation success considering both sensitivity and specificity, Youden's J(9) and Euclidean distances were plotted by predicted probability of extubation success and a cut-point was selected based on the highest Youden's J and lowest Euclidean distance. Characteristics between infants with extubation success and failure were compared using chi-square or Fisher's exact tests for categorical variables, independent sample t-tests for means and Wilcoxon rank-sum tests for medians. The mean probability of extubation success at time of extubation was compared by severity of BPD using linear regression, with and without adjustment for gestational age, to determine the independent association of extubation success probability after accounting for gestational age. All other outcome comparisons were unadjusted, to only account for potential differences after considering random variation, and were not intended for causal inference. Analyses were performed using SPSS version 27 (IBM) and GraphPad Prism 8. A two-sided  $P < 0.05$  was considered statistically significant.

## Results

Among a total of 182 infants meeting inclusion criteria, we had complete data for the extubation prediction model in 177 infants, of which 120 (68%) had extubation success (**Supplementary Fig. 1**). Infants who failed extubation had a lower birth weight and gestational age ( $P < 0.001$ , Table 1). There were no significant differences in maternal race, infant sex, multiple gestation, mode of delivery, Apgar score at 1 and 5 minutes, highest MAP and  $\text{FiO}_2$  in the first 12 hours of age, receipt of surfactant within 2 hours of age, or RSS in the first 6 hours of age by extubation success groups (Table 1).

Table 1  
Baseline Characteristics

Baseline characteristics	Extubation failure (n = 57)	Extubation success (n = 120)	P
Gestational age, mean weeks $\pm$ SD	25.8 $\pm$ 1.6	27.2 $\pm$ 1.9	< 0.001
Gestational age group			0.001
22 to < 25 weeks	19 (33.3%)	18 (15.0%)	
25 to < 28 weeks	30 (52.6%)	55 (45.8%)	
28 to < 31 weeks	8 (14.0%)	47 (39.2%)	
Birth weight, mean $\pm$ SD	773 $\pm$ 172	917 $\pm$ 205	< 0.001
Multiple gestation	18 (31.6%)	31 (25.8%)	0.43
Mode of delivery			0.16
Vaginal	15 (26.3%)	31 (25.8%)	
Routine cesarean	20 (35.1%)	58 (48.3%)	
Emergent cesarean	22 (38.6%)	31 (25.8%)	
Sex			0.56
Male	34 (59.6%)	66 (55.0%)	
Female	23 (40.4%)	54 (45.0%)	
Race / ethnicity <sup>a</sup>			0.18
Black	45 (78.9%)	94 (79.7%)	
White	8 (14.0%)	16 (13.6%)	
Hispanic	2 (3.5%)	0	
Other	2 (3.5%)	8 (6.8%)	
1 min Apgar score, median (IQR)	3 (2–5)	3 (2–5)	0.88
5 min Apgar score, median (IQR)	7 (5–8)	7 (5–8)	0.67

Abbreviations: SD, standard deviation; IQR, interquartile range; RSS, respiratory severity score; FiO<sub>2</sub>, fraction of inspired oxygen; MAP, mean airway pressure.

<sup>a</sup> Missing data for race on 2 infants (both in extubation success group)

Baseline characteristics	Extubation failure (n = 57)	Extubation success (n = 120)	P
Age at caffeine initiation, median (IQR) in hours since birth	13 (5–24)	14 (5–24)	0.86
Early surfactant within 2 hours of birth	18 (31.6%)	35 (29.2%)	0.74
Highest RSS in first 6 hours of birth, mean ± SD	2.5 ± 1.6	2.7 ± 2.0	0.56
Highest FiO <sub>2</sub> in first 12 hours of birth, median (IQR)	0.35 (0.27–0.50)	0.40 (0.28–0.50)	0.51
Highest MAP in first 12 hours of birth, median cm H <sub>2</sub> O (IQR)	8.5 (7.5–9.0)	8.1 (7.0–9.4)	0.54
Abbreviations: SD, standard deviation; IQR, interquartile range; RSS, respiratory severity score; FiO <sub>2</sub> , fraction of inspired oxygen; MAP, mean airway pressure.			
<sup>a</sup> Missing data for race on 2 infants (both in extubation success group)			

### Model validation

The AUROC of the primary 6 variable model was 0.72 (95% CI 0.65–0.80, Fig. 1). A linear fit line of the calibration plot had a slope of 0.96, y-intercept of -0.06 and R<sup>2</sup> of 0.81 (Fig. 1). The mean observed-to-expected difference in probability of extubation success was -0.08 (95% CI -0.01, -0.15). Exploring cut-points of extubation success considering both specificity and sensitivity, the highest Youden's J and lowest Euclidean distance were at a predicted probability of extubation success of 0.80 (sensitivity of 0.63, specificity of 0.74) (**Supplementary Table 2 and Supplementary Fig. 2**). Infants with a predicted probability of extubation success of >0.80 had an 83% incidence of first-attempt extubation success, compared to 52% for infants with a predicted probability of extubation success ≤ 0.80.

### Model modification

Using a model derived in the validation cohort with the 6 variables in the original extubation success calculator, the AUROC was 0.759 (Fig. 2). After removal of RSS as a predictor, the AUROC was 0.744. After expanding the model by adding birth weight, pre-extubation MAP, receipt of peri-extubation steroids and caffeine, the AUROC was 0.787. The AUROC of a gestational-age only model was 0.712 and of a recently published two-variable model based only on gestational age and MAP was 0.721.

### Comparison of factors between derivation and validation cohort

Odds ratio estimates were similar among the validation cohort and the derivation cohort, with the exception of RSS in the first 6 hours of age, which showed differing directions in risk estimates between the two cohorts (**Supplementary Fig. 3**).

### *Peri-extubation characteristics*

Infants that failed extubation weighed less at time of extubation ( $P = 0.002$ , Table 2), and received a higher  $FiO_2$  ( $P < 0.001$ ), had a lower blood gas pH ( $P = 0.003$ ), higher MAP ( $P = 0.02$ ) and an earlier postnatal age at extubation ( $P = 0.02$ ). There was no significant difference in the interval in hours between the pre-extubation blood gas and time of extubation between groups ( $P = 0.55$ ). Infants who failed extubation were more likely to receive peri-extubation steroids ( $P = 0.009$ ). There was no significant difference between the groups in receipt of caffeine prior to extubation ( $P = 0.06$ ).

Table 2  
Peri-extubation Characteristics

Peri-extubation characteristics	Extubation failure (n = 57)	Extubation success (n = 120)	P
Estimated extubation success probability, mean ± SD	67% ± 19%	81% ± 16%	< 0.001
Postnatal age at extubation, days, median (IQR)	5.0 (2.5–25.5)	3.0 (2.0–6.8)	0.02
Weight at extubation, grams, mean ± SD	846 ± 176	942 ± 191	0.002
FiO <sub>2</sub> before extubation, median (IQR)	0.25 (0.21–0.32)	0.21 (0.21–0.26)	< 0.001
Pre-extubation pH, mean ± SD	7.30 ± 0.07	7.33 ± 0.06	0.003
Interval between pre-extubation blood gas and extubation, hours, median (IQR)	5.6 (1.9–7.0)	3.9 (1.9–7.5)	0.55
Any peri-extubation systemic corticosteroids <sup>a</sup>	7 (12.3%)	2 (1.7%)	0.009
Caffeine before extubation	56 (98.2%)	108 (90.0%)	0.06
Caffeine dose before extubation, mg/kg/day, median (IQR)	5.5 (5.0–8.8)	5.7 (5.0–20.0)	0.32
Mean airway pressure immediately prior to extubation, mean ± SD	7.0 ± 1.5	6.5 ± 1.0	0.02
Type of respiratory support immediately post extubation			0.001
Unsynchronized NIPPV <sup>b</sup>	28 (49.1%)	39 (32.5%)	
SIPAP <sup>c</sup>	18 (31.6%)	17 (14.2%)	
nCPAP	11 (19.3)	52 (43.3%)	

Abbreviations: SD, standard deviation; IQR, interquartile range; FiO<sub>2</sub>, fraction of inspired oxygen; MAP, mean airway pressure; PEEP, positive end expiratory pressure; NIPPV, non-invasive positive pressure ventilation; nCPAP, nasal continuous positive airway pressure; HFNC, high flow nasal cannula; LFNC, low flow nasal cannula; RA, room air.

Values are n (%) unless otherwise indicated.

<sup>a</sup> Receipt on the day of extubation.

<sup>b</sup> Administered through a RAM or Optiflow nasal prong interface.

<sup>c</sup> Bi-level positive airway pressure.

Peri-extubation characteristics	Extubation failure (n = 57)	Extubation success (n = 120)	P
HFNC ( $\geq 2$ L/min)	0	3 (2.5%)	
LFNC ( $< 2$ L/min)	0	8 (6.7%)	
RA	0	1 (0.8%)	
Abbreviations: SD, standard deviation; IQR, interquartile range; FiO <sub>2</sub> , fraction of inspired oxygen; MAP, mean airway pressure; PEEP, positive end expiratory pressure; NIPPV, non-invasive positive pressure ventilation; nCPAP, nasal continuous positive airway pressure; HFNC, high flow nasal cannula; LFNC, low flow nasal cannula; RA, room air.			
Values are n (%) unless otherwise indicated.			
<sup>a</sup> Receipt on the day of extubation.			
<sup>b</sup> Administered through a RAM or Optiflow nasal prong interface.			
<sup>c</sup> Bi-level positive airway pressure.			

### *Clinical outcomes*

Infants who failed extubation had an increased severity of BPD or death ( $P < 0.001$ , Table 3). In addition, among infants with extubation attempts prior to 36 weeks' postmenstrual age, the probability of extubation success at first extubation attempt was inversely correlated with the severity of BPD (**Supplementary Fig. 4**). Infants who failed extubation had a longer length of stay ( $P = 0.003$ ), longer duration of ventilation ( $P < 0.001$ ), were more likely to receive postnatal steroids ( $P < 0.001$ ). Other outcomes are shown in Table 3.

Table 3  
Outcomes of Infants by Initial Extubation Failure or Success

Outcome	Extubation failure (n = 57)	Extubation success (n = 120)	P
BPD Grade (Jensen criteria) or death <sup>a</sup>			< 0.001
None	9 (17.0%)	49 (42.6%)	
I	21 (39.6%)	52 (45.2%)	
II	18 (34.0%)	11 (9.6%)	
III	1 (1.9%)	3 (2.6%)	
Death	4 (7.5%)	0	
Length of stay, days, median (IQR)	95 (76–117)	79 (63–96)	0.003
Total duration of mechanical ventilation, days, median (IQR)	30 (16–51)	4 (1–11)	< 0.001
Number of endotracheal mechanical ventilation courses			N/A <sup>b</sup>
1	0	85 (70.8%)	
2	39 (69.6%)	26 (21.7%)	
3	15 (26.8%)	9 (7.5%)	
4	2 (3.6%)	0	
Postnatal systemic corticosteroids <sup>c</sup>	17 (30.4%)	9 (7.7%)	< 0.001
Retinopathy of prematurity receiving treatment	2 (4.2%)	1 (0.9%)	0.21
Severe intraventricular hemorrhage (grade 3 or 4)	17/54 (31.5%)	8/118 (6.8%)	< 0.001
Periventricular leukomalacia	5/51 (9.8%)	5/114 (4.4%)	0.29
Necrotizing enterocolitis (Bell stage II or greater)	8 (14.0%)	27 (22.5%)	0.19
<sup>a</sup> Evaluated among 53 infants with extubation failure and 115 with extubation success.			
<sup>b</sup> No statistical comparisons performed because groups were defined by need for reintubation.			
<sup>c</sup> Three infants enrolled in a blinded, randomized trial of hydrocortisone vs. placebo treatment were excluded.			

## Discussion

The extubation prediction model in our validation cohort had an AUROC of 0.72, compared to an AUROC of 0.77 in the previously published derivation cohort(3). By convention, an AUROC of 0.5 is non-informative, 0.7 to 0.8 is considered acceptable, 0.8 to 0.9 is considered excellent and more than 0.9 is considered outstanding, although this depends on the clinical context(10, 11). Based on these categorizations, both the Gupta et al. study(3) and the results from our validation study demonstrate acceptable model performance. Based on the calibration data, the model slightly underestimated extubation success on average in our cohort (8% lower than expected), but the calibration slope and intercept supported acceptable calibration.

In our study, 68% of very preterm infants were successfully extubated on initial attempt and those that failed extubation had a higher severity of BPD. By comparison, extubation success in the Gupta et al.(3) study was 73%. Notably, the association between the highest RSS in the first 6 hours of life and extubation success differed between the two cohorts. This may reflect differences in early respiratory care, including surfactant administration and approaches to mechanical ventilation, which may influence RSS. Because this variable is included in the original model, we assessed the impact of its removal and found only a small reduction in model discrimination.

We defined extubation success as survival without an endotracheal tube for 5 days to be consistent with the definition used in the original study by Gupta et al(1, 3). Other definitions of extubation success exist, with a 3 to 7 days duration of observation being most common and rates of extubation success varying based on definition(12). All of the infants in this study received caffeine, although some received caffeine after extubation. In addition, some received peri-extubation steroids. Inclusion of these factors into an expanded model suggested they might slightly improve model performance, although the association between these factors and extubation success may not be generalizable to other centers. A recent study reported that a two-factor model that included gestational age and mean airway pressure predicted extubation success with an AUROC of 0.77(8). This two-variable model in our cohort had an AUROC of 0.72. In both this study and the study by Kidman et al.(8), pre-extubation MAP was associated with extubation success. By contrast, the study by Gupta et al.(3) reported no association with pre-extubation MAP and extubation success. In addition, we found an association between peri-extubation corticosteroid use and extubation success, which was not observed in the study by Kidman et al.(8) These differences may be due to differences in practices among centers, and highlight the importance of external validation of predictive models used in clinical practice. Additional studies of model performance of this and other extubation readiness models are needed in other cohorts of similar infants in order to determine which factors might be universally associated with extubation readiness and which factors are impacted by local center practices. It would be useful to assess the association with RSS within 6 hours of birth and extubation success in additional studies, given the differences observed between the original Gupta et al. (3) study and this study. In order to make the calculator easier to use, a reduced model without RSS may be more practical and not sacrifice model performance. The severity of BPD or death was associated with a lower predicted probability of extubation success in our cohort, even after adjustment for gestational

age. This suggests that the other factors in the model, aside from gestational age, may contribute to the severity of BPD. These findings are consistent with the association between extubation failure and BPD reported in prior studies(3, 8). However, other factors associated with both BPD and extubation success may either confound or explain the observed association.

Additional studies are needed to determine how use of the extubation success calculator may improve outcomes. One potential approach would be to wait until an infant reaches a sufficiently high predicted probability of extubation success before attempting extubation. Our study suggests a threshold of > 80% could be considered, although the sensitivity and specificity at this threshold may not be sufficient for clinical decision making. Additionally, the extubation prediction model may be useful to evaluate heterogeneity in treatment effect(13) of interventions to improve extubation success(14).

There are some limitations to our study. Our findings may not be generalizable to infants that weigh > 1250 grams at birth, have previously been extubated, have an unplanned extubation, or are extubated after 60 postnatal days. Additionally, over 90% of infants in our study received any dose of caffeine within 24 hours before extubation, and the association between pre-extubation caffeine use and extubation success may differ for centers with more variable use. Finally, differences in outcomes among infants who had extubation success, compared to extubation failure, may be explained by many factors, including confounders that were not accounted for in our study. Importantly, our goals were not causal inference on the association between extubation success and clinical outcomes, but a focus on the external validation of the prediction model, and differences reported in outcomes should be viewed accordingly.

In conclusion, a publicly available prediction model for first-attempt extubation success in very preterm infants demonstrated acceptable performance in this external cohort of infants, supporting its potential utility in clinical practice. Additional studies are needed to determine if its use can improve outcomes.

## Declarations

**Availability of Data and Materials:** The corresponding author can be contacted for source data.

**Funding:** None

**Author Contributions:** BD conceived of and designed the study, acquired data and drafted the manuscript, AS and VS acquired data and critically revised the manuscript for important intellectual content, RG, MR, and SC guided analyses and interpretation of data and critically revised the manuscript for important intellectual content and RP conceived of and designed the study, analyzed the data, drafted the manuscript and provided study supervision.

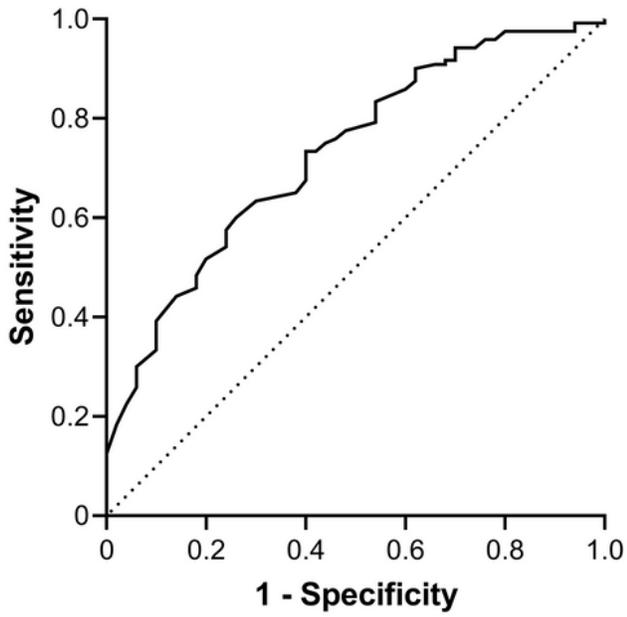
**Acknowledgements:** None

## References

1. Walsh MC, Morris BH, Wrage LA, Vohr BR, Poole WK, Tyson JE, et al. Extremely low birthweight neonates with protracted ventilation: mortality and 18-month neurodevelopmental outcomes. *J Pediatr*. 2005;146(6):798-804.
2. Chawla S, Natarajan G, Shankaran S, Carper B, Brion LP, Keszler M, et al. Markers of Successful Extubation in Extremely Preterm Infants, and Morbidity After Failed Extubation. *J Pediatr*. 2017;189:113-9 e2.
3. Gupta D, Greenberg RG, Sharma A, Natarajan G, Cotten M, Thomas R, et al. A predictive model for extubation readiness in extremely preterm infants. *J Perinatol*. 2019;39(12):1663-9.
4. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ*. 2015;350:g7594.
5. Jensen EA, Dysart K, Gantz MG, McDonald S, Bamat NA, Keszler M, et al. The Diagnosis of Bronchopulmonary Dysplasia in Very Preterm Infants. An Evidence-based Approach. *Am J Respir Crit Care Med*. 2019;200(6):751-9.
6. Alba AC, Agoritsas T, Walsh M, Hanna S, Iorio A, Devereaux PJ, et al. Discrimination and Calibration of Clinical Prediction Models: Users' Guides to the Medical Literature. *JAMA*. 2017;318(14):1377-84.
7. Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. *BMJ*. 2009;338:b605.
8. Kidman AM, Manley BJ, Boland RA, Davis PG, Bhatia R. Predictors and outcomes of extubation failure in extremely preterm infants. *J Paediatr Child Health*. 2021;57(6):913-9.
9. Perkins NJ, Schisterman EF. The Youden Index and the optimal cut-point corrected for measurement error. *Biom J*. 2005;47(4):428-41.
10. Swets JA. Measuring the accuracy of diagnostic systems. *Science*. 1988;240(4857):1285-93.
11. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. *J Thorac Oncol*. 2010;5(9):1315-6.
12. Giaccone A, Jensen E, Davis P, Schmidt B. Definitions of extubation success in very premature infants: a systematic review. *Arch Dis Child Fetal Neonatal Ed*. 2014;99(2):F124-7.
13. Kent DM, van Klaveren D, Paulus JK, D'Agostino R, Goodman S, Hayward R, et al. The Predictive Approaches to Treatment effect Heterogeneity (PATH) Statement: Explanation and Elaboration. *Ann Intern Med*. 2020;172(1):W1-W25.
14. Ferguson KN, Roberts CT, Manley BJ, Davis PG. Interventions to Improve Rates of Successful Extubation in Preterm Infants: A Systematic Review and Meta-analysis. *JAMA Pediatr*. 2017;171(2):165-

## Figures

### A Receiver Operating Characteristic Curve



### B Calibration Plot

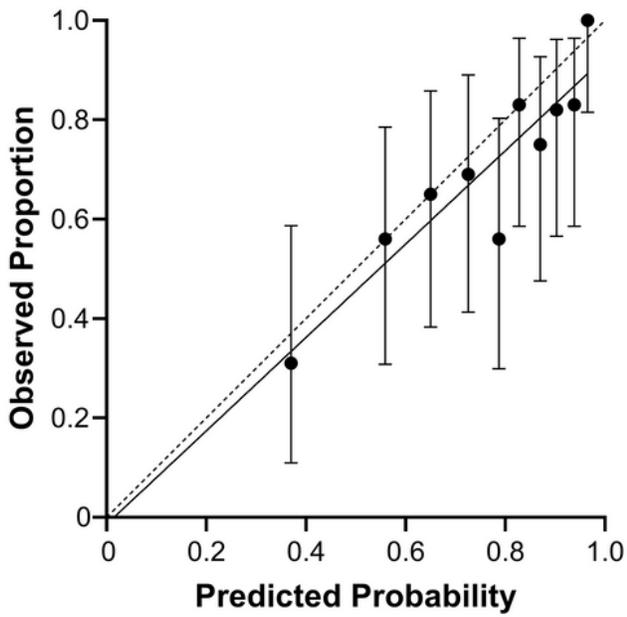
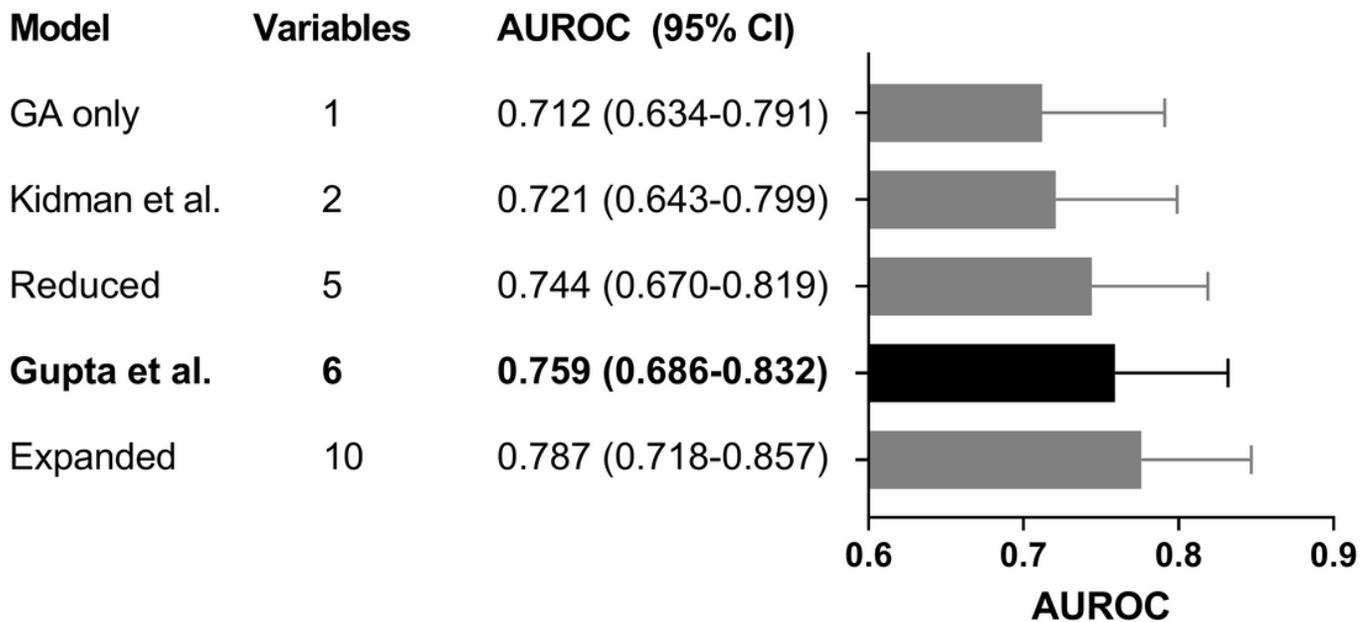


Figure 1

ROC Curve and Calibration Plot Panel A. Receiver operating characteristic (ROC) curve shows sensitivity and 1-specificity for the probability of extubation success. Panel B. The dotted line in the calibration plot represents an exact match between observed and expected probability of extubation success, and the solid black line is fitted to observed data.



**Figure 2**

Classification of Reduced and Expanded Models. The model AUROC values are shown in the figure, with results from variables in the model by Gupta et al.(3) denoted in bold. The GA model only included gestational age, the Kidman et al.(8) model included gestational age and mean airway pressure before extubation, the Reduced model removed the highest RSS within 6 hours of life from variables in the Gupta et al. model, and the Expanded model added mean airway pressure before extubation, birth weight, peri-extubation corticosteroids and caffeine prior to extubation to the variables in the Gupta et al. model.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplement.pdf](#)