

# Predicting the need for tracheostomy in infants with severe bronchopulmonary dysplasia

Hannah Eglinton

November 2023

## Abstract

**Background:** Bronchopulmonary dysplasia (BPD) poses a significant challenge in the care of preterm infants, with severe cases often needing tracheostomy. This study aims to develop models for predicting tracheostomy in neonates with severe BPD using respiratory parameters at 36 and 44 weeks post-menstrual age (PMA).

**Methods:** A multicenter, retrospective case-control study was conducted, involving 996 infants born at less than 32 weeks PMA. Clinical data were collected at birth, 36 weeks PMA, 44 weeks PMA, and at discharge. Multiple imputation addressed missing data. Four logistic mixed-effects models were developed, with 30% of observations reserved for validation. Models were evaluated using area under the receiver operating characteristic curve (AUC), Brier scores, precision, recall, and F1 score.

**Results:** The prevalence of tracheostomy was 15%. Respiratory support level, prenatal corticosteroids, and their interactions emerged as significant predictors. The 44-week model outperformed the 36-week model, showcasing the importance of respiratory parameters closer to discharge. Prenatal corticosteroids were associated with an increased tracheostomy risk.

**Conclusion:** Prenatal corticosteroids, respiratory support level, fraction of inspired oxygen, and medication for pulmonary hypertension were selected as predictors for tracheostomy. However, class imbalance and small sample size necessitate cautious interpretation. Future research should validate these models in larger, diverse cohorts, addressing these limitations for enhanced generalizability.

## Introduction

Prematurity remains a leading cause of neonatal morbidity and mortality, with bronchopulmonary dysplasia (BPD) as the most prevalent complication in this population. Defined by the chronic impairment of pulmonary development, BPD predominantly affects infants born prematurely, particularly those requiring prolonged mechanical ventilation and oxygen therapy. Severe BPD (sBPD), characterized by the need for invasive respiratory support at 36 weeks post menstrual age (PMA) affects 10,000 to 15,000 infants per year in the United States [1].

Severe BPD can have a lasting effect on infants, as 75% of neonates with sBPD will require ventilatory support when they are discharged from the hospital. In these cases, the patient needs a tracheostomy, a surgical hole in the neck that allows them to be hooked up to a ventilator. The decision to place a tracheostomy in patients is complex. While it is associated with improved outcomes in the patients who need it, tracheostomy also introduces additional risks, including heightened susceptibility to respiratory infections and an increased risk of death [2].

Clinicians grapple with the challenge of determining the optimal timing for tracheostomy placement in neonates with sBPD. While early intervention helps the patient and parents adjust to tracheostomy well before discharge, leading to better outcomes and improved growth, clinicians do not want to introduce the risks associated with tracheostomy to patients who will not need it [2]. Existing attempts to predict the need

for tracheostomy have predominately relied on baseline demographics and clinical diagnosis and have not provided prediction at different postmenstrual ages or used detailed respiratory parameters.

This research seeks to address this gap by leveraging clinical data collected at two timepoints – 36 and 44 weeks PMA – to develop predictive models for tracheostomy outcomes in neonates with BPD. 36 weeks marks a crucial junction in the neonatal period, as it is the point at which preterm born infants are considered to have reached their term-equivalent age [3]. The 44-week timepoint provides insights closer to anticipated discharge. The implications of such a model are substantial, not only for guiding clinicians through informed decision-making but also for refining the ongoing debate surrounding tracheostomy placement in neonates with severe bronchopulmonary dysplasia.

## Methods

### Study population

The study population for this research was derived from the BPD Collaborative Registry, a multi-center consortium of interdisciplinary BPD programs located in the United States and Sweden. The registry includes infants whose gestational age at birth was less than 32 weeks and who had sBPD as defined by 2001 NHLBI criteria (requiring  $>30\%$  inspired oxygen or positive pressure ventilation at 36 weeks). In the registry, standard demographic and clinical data were collected at four time points; birth, 36 weeks PMA, 44 weeks PMA, and at discharge.

In this study, we included data across ten participating centers affiliated with the BPD Collaborative. The study population consisted of 996 infants born at less than 32 weeks PMA with sBPD at 36 weeks. Data collected at birth included maternal ethnicity, weight, gestational age, length, head circumference, delivery method, whether they received prenatal corticosteroids, whether they received a complete course of prenatal corticosteroids, maternal chorioamnionitis, gender, whether they were small for gestational age (SGA), and whether the infant received surfactant in the first 72 hours after birth.

Data collected at the 36 week and 44 week PMA timepoints included weight, ventilation support level, fraction of inspired oxygen ( $\text{FiO}_2$ ), peak inspiratory pressure (PIP), positive and expiratory pressure (PEEP), and medication for pulmonary hypertension. The three ventilation support levels were defined as (1) no respiratory support or supplemental oxygen, (2) non-invasive positive pressure, or (3) invasive positive pressure. Data collected at discharge included the gestational age at discharge and whether a tracheostomy was placed. Gestational age at discharge was not considered because the purpose of this model is to allow clinicians to predict tracheostomy at 36 weeks or 44 weeks PMA, before the patients has been discharged. Data were also recorded on whether the infant had died.

### Candidate Variable Selection

#### Birth Covariates

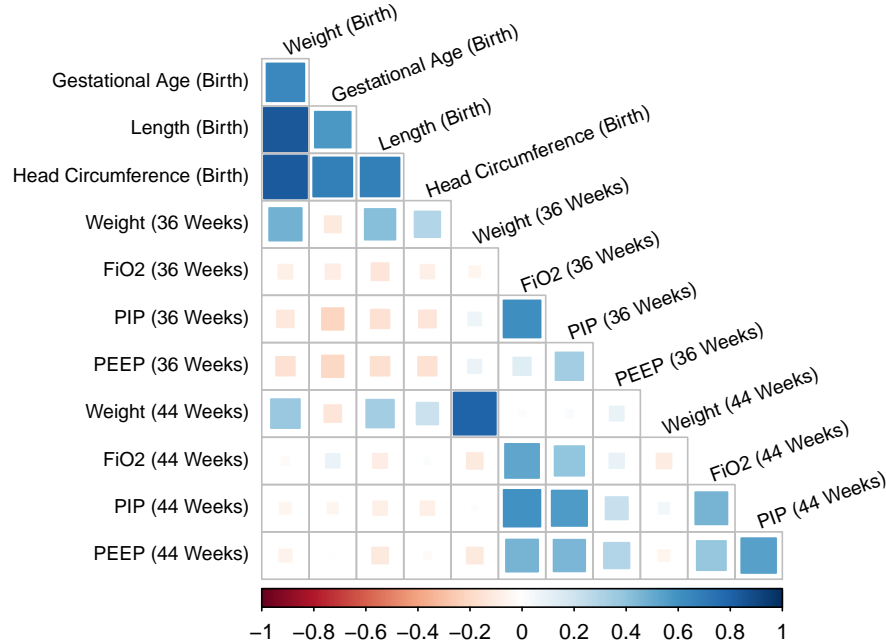
The birth weight, gestational age, length, and head circumference were highly correlated, with pairwise correlations between 0.68 and 0.84 (Figure 1). To avoid multicollinearity in our model, we chose to include only birth weight since it was the most highly related to tracheostomy ( $p = 0.004$ ) and had no missing data. The SGA indicator was also included to capture the relationship between size and gestational age at birth.

The dataset included both a binary indicator of whether prenatal corticosteroids were taken as well as an indicator of whether a complete course of corticosteroids was taken. We have chosen to include only the former, as the latter had a high prevalence of missing data (19.4%, Table 1) and was not associated with tracheostomy ( $p > 0.9$ ).

Information on the mother’s ethnicity was also collected. We have chosen to exclude this from the model derivation, as it does not have a biological basis and its inclusion could potentially lead to discrimination if the model returns different results depending on ethnicity. Since tracheostomy status was not significantly associated with ethnicity ( $p = 0.3$ ), excluding this variable is unlikely to introduce bias.

There was a high prevalence of missing data for whether the infants received surfactant within the first 72 hours (43.5%, Table 1). Since this variable was not highly associated with the outcome ( $p = 0.14$ ), it was excluded from the model. Thus, the birth variables included as candidates in model selection were birth weight, SGA, prenatal corticosteroids, delivery method, maternal chorioamnionitis, and sex assigned at birth.

**Figure 1. Correlation Between Continuous Covariates**



### 36 and 44 Week Covariates

Variables collected at 36 and 44 weeks PMA included weight, ventilation support level, fraction of inspired oxygen (FiO2), peak inspiratory pressure (PIP), positive and inspiratory pressure (PEEP), and whether the infant was medicated for pulmonary hypertension. Since all of these variables were highly associated with tracheostomy ( $p < 0.001$ ) and had only moderate pairwise correlations (between  $r = -0.09$  and  $r = 0.57$ ), we decided to include all as candidates in the model.

There were large proportions of missing data for each variable collected at the 44 week timepoint (Table 1). Although these had high proportions of missing data, the fact that each variable was also collected at 36 weeks indicates that multiple imputation may accurately predict values at 44 weeks. We thus used multiple imputation to fill in these values – however, due to this close relationship between 36 and 44 week variables, we did not fit both timepoints in the same models. Instead, we fit separate models for the 36 week data and the 44 week data. Clinicians can then choose which model to use based on the data that they have available.

### Center

The data were collected from ten centers, some of which were referral hospitals with a higher proportion of sick infants. To account for the difference in center characteristics while retaining the transportability of the model, we included the centers as a random intercept in mixed-effects models rather than as a fixed effect.

Table 1: Prevalence of Missing Data

Variable	Percent Missing
Maternal Ethnicity	5.7%
Weight (Birth)	0%
Gestational Age (Birth)	0%
Length (Birth)	7.8%
Head Circumference (Birth)	7.7%
Delivery Method	0.3%
Prenatal Corticosteroids	3.5%
Complete Prenatal Steroids	19.4%
Maternal Chorioamnionitis	6.2%
Sex Assigned at Birth	0.4%
SGA	1.5%
Surfactant in First 72 Hours	43.5%
Weight (36 Weeks)	9.2%
Ventilation Support Level (36 Weeks)	3%
FiO2 (36 Weeks)	9.2%
PIP (36 Weeks)	12.9%
PEEP (36 Weeks)	11.7%
Medication for Pulmonary Hypertension (36 Weeks)	3%
Weight (44 Weeks)	44.8%
Ventilation Support Level (44 Weeks)	42.6%
FiO2 (44 Weeks)	45%
PIP (44 Weeks)	45%
PEEP (44 Weeks)	44.8%
Medication for Pulmonary Hypertension (44 Weeks)	42.6%
Tracheostomy	0%

## Outcome

The outcome of interest was whether a tracheostomy was eventually needed. Death was not included as an outcome of interest because we cannot assume that death was due to BPD, tracheostomy, or an unrelated complication. Infants born prematurely often experience a range of comorbidities that increase the risk of death, none of which are reported in this data set. Further, death was relatively rarely observed ( $N = 54$ ) making it a highly imbalanced outcome with a low class size.

## Data Preprocessing

Multiple imputation was performed using the `mice` package in R. The MICE algorithm iteratively imputes missing values for each variable based on the observed values of other variables in the dataset. We created five imputed datasets to ensure robustness in our imputation process. The imputation model included all covariates from the original dataset, including covariates that were excluded as candidate variables for model selection. Continuous variables were imputed using predictive mean matching and categorical variables were imputed using logistic regression.

Multiple imputation requires an assumption that missingness is missing at random (MAR). In other words, missingness is related to other measured variables in the analysis model, but not to the underlying values of the incomplete variable. We imputed data for delivery method, prenatal corticosteroids, maternal chorioamnionitis, sex, SGA, weight at 36 and 44 weeks, ventilation support level, FiO2, PIP, PEEP, and medication for pulmonary hypertension. Much of the missingness in these variables appeared to be associated with center where the data were collected, rather than their underlying values. Further, there is no inherent reason to believe that the missing values were systematically higher or lower than observed values for any of the these variables, supporting the assumption that missingness is independent of the variable values. This aligns with the MAR assumption, indicating that the probability of missing data is related to other measured variables in the data.

The dataset was evaluated for outliers, though no infeasible values were detected. Continuous candidate variables (weight and respiratory measures) were normalized to maintain consistency in scale and allow for meaningful comparisons during analysis.

## Regression Models

After removing highly correlated variables and variables with a high prevalence of missing data, both the 36 and 44 week models had 14 candidate variables. This is still too high-dimensional for logistic regression, so we used further variable selection procedures. Two common options for optimized variable selection are best subset selection and lasso regression.

Lasso regression penalizes zero coefficients and shrinks the values of nonzero coefficients. Thus, lasso regression is a method of both variable selection and regularization. Best subsets, on the other hand, penalizes nonzero coefficients but does not shrink down the values of the coefficients. Both lasso and best subsets are optimized and find the best overall solution. However, best subsets can be computationally expensive as it is solving a non-convex problem. An alternative is forward stepwise selection, which starts with an empty model and iteratively adds one new variable that best improves the fit. This results in  $p$  models, each with a different number of variables included in the model, the best of which can be chosen by cross-validated predicted error.

Since we planned to account for the various centers using a mixed-effects regression model, we chose a variable selection technique that did not shrink the coefficient values. This way, we could use variable selection to identify our subset of covariates, which could then be entered into a mixed-effects model (which assumes no regularization). Although forward stepwise selection is often used as a quicker alternative to best subset selection, recent packages in R have improved the computational feasibility of running best subsets. We used the package `L0Learn` to fit best subset selection on our data. The best subsets optimization problem can be expressed as so:

$$\min \frac{1}{n} \sum_{i=1}^n (y_i - X_i \beta)^2 + \lambda \sum_{j=1}^p 1\{\beta_j \neq 0\}$$

## Model Development

Thirty percent of the observations ( $N = 304$ ) were set aside as a validation data set. The remaining observations ( $N = 692$ ) were used to train the models. The training and validation sets each contained equal proportions of observations with a tracheostomy. We developed four types of models: (1) birth and 36-week data, including no interactions; (2) birth and 36-week data, including all interactions with ventilation support level; (3) birth and 44-week data, including no interactions; (4) birth and 44-week data, including all interactions with ventilation support level.

A best subset regression algorithm was used to select the variables included in each of the four types of models. First, fivefold cross-validation was used to choose the optimal  $\lambda$  value that minimized logistic loss. Best subset models using the optimal  $\lambda$  values were then fit on each of the five imputed datasets. Any variables that were selected when using at four of the five imputed datasets were included in subsequent mixed-effects models.

After best subset regression was run for each of the model types, mixed-effects models were fit on the full data using selected covariates for each model type. Each of these models included a random intercept term for the hospital center ID.

## Evaluation Metrics

Models were evaluated using the withheld validation set. Model discrimination was measured using AUC and model calibration was measured using Brier scores. A higher AUC value indicates better discrimination performance, capturing the model’s ability to distinguish between positive and negative cases. The Brier score measures the mean squared difference between predicted probabilities and actual outcomes. Lower Brier scores indicate better calibration, reflecting the model’s accuracy in predicting class probabilities.

Given the observed high class imbalance in the dataset, we recognized the need for metrics that specifically account for imbalanced classification scenarios. Therefore, in addition to AUC and Brier scores, we reported precision, recall, and F1 score. Precision measures the accuracy of positive predictions, recall evaluates the model’s ability to capture all positive instances, and F1 score provides a balance between precision and recall. We used the F1 scores to determine the best 36-week and 44-week models.

To account for potential variability introduced by imputed data sets, all metrics were calculated separately for each of the five imputed data sets. Subsequently, the results were averaged to provide an estimation of model performance across multiple imputations.

## Results

### Study Population Characteristics

Prevalence of tracheostomy in the BPD Collaborative cohort was 15%. A summary of the candidate variables by tracheostomy status is reported in Table 2. Based on Wilcoxon rank sum tests and Pearson’s Chi-squared tests, the weight at birth, delivery type, prenatal corticosteroids, SGA, weight at 36 weeks, all respiratory support variables, and medication for pulmonary hypertension were significantly different between those who did have a tracheostomy and those who did not. Maternal chorioamnionitis, sex assigned at birth, and weight at 44 weeks did not differ significantly between those that did and did not have a tracheostomy in this dataset.

Table 2: Candidate Variable Summary by Tracheostomy Status

Characteristic	No Tracheostomy (N = 850)	Tracheostomy (N = 146)	p-value
Weight at Birth	814 (295)	761 (303)	0.004
Cesarean Section Delivery	594 (70%)	114 (79%)	0.035
Prenatal Corticosteroids	712 (86%)	123 (94%)	0.011
Maternal Chorioamnionitis	138 (17%)	22 (16%)	0.8
Sex Assigned at Birth	498 (59%)	86 (59%)	>0.9
Small for Gestational Age	161 (19%)	42 (30%)	0.005
Weight at 36 Weeks	2,132 (410)	2,024 (432)	0.025
Ventilation Support Level at 36 Weeks			<0.001
None	111 (13%)	6 (4.7%)	
NIPPV	560 (67%)	29 (23%)	
IPPV	168 (20%)	92 (72%)	
Fraction of Inspired Oxygen at 36 Weeks	0.32 (0.13)	0.49 (0.20)	<0.001
Peak Inspiratory Pressure at 36 Weeks	4 (9)	15 (12)	<0.001
Positive and Exploratory Pressure at 36 Weeks	6 (3)	8 (3)	<0.001
Medication for Pulmonary Hypertension at 36 Weeks	41 (4.9%)	25 (20%)	<0.001
Weight at 44 Weeks	3,667 (665)	3,550 (750)	0.2
Ventilation Support Level at 44 Weeks			<0.001
None	262 (57%)	7 (6.3%)	
NIPPV	128 (28%)	18 (16%)	
IPPV	71 (15%)	86 (77%)	
Fraction of Inspired Oxygen at 44 Weeks	0.32 (0.13)	0.44 (0.19)	<0.001
Peak Inspiratory Pressure at 44 Weeks	5 (12)	21 (16)	<0.001
Positive and Exploratory Pressure at 44 Weeks	3 (4)	9 (4)	<0.001
Medication for Pulmonary Hypertension at 44 Weeks	48 (10%)	51 (46%)	<0.001

<sup>1</sup> Mean (SD); n (%)

<sup>2</sup> Wilcoxon rank sum test; Pearson’s Chi-squared test

Data were collected from ten hospital centers, with the number of observations per center ranging from just 1 patient to 630. Some centers were referral centers, which see a higher proportion of sicker patients. The proportions of each respiratory support level per center are reported in Table 3. Centers 1 and 12 had

higher prevalences of subjects requiring invasive positive pressure and more subjects receiving a tracheostomy, suggesting that these may be referral centers.

This dataset has a hierarchical structure, as it was collected from separate centers with unique characteristics. Therefore, a mixed-effects model is appropriate as it allows for the incorporation of both fixed effects, capturing the trends in respiratory and other clinical parameters, and random effects, accommodating the variability and correlations within hospital centers. By incorporating the hospital centers as a random effect in the mixed-effects model, the analysis becomes more generalizable, enabling the model to be applied to centers outside our data.

Table 3: Hospital Center Characteristics

Characteristic	1, N = 65	2, N = 630	3, N = 57	4, N = 60	5, N = 40	7, N = 32	12, N = 69	16, N = 38	20, N = 4	21, N = 1
Support Level (36 Wks)										
None	8 (12%)	50 (7.9%)	5 (8.8%)	8 (13%)	0 (0%)	22 (69%)	1 (1.4%)	22 (58%)	1 (25%)	0 (0%)
NIPPV	22 (34%)	425 (67%)	35 (61%)	34 (57%)	31 (78%)	8 (25%)	17 (25%)	14 (37%)	2 (50%)	1 (100%)
IPPV	34 (52%)	146 (23%)	16 (28%)	18 (30%)	9 (23%)	2 (6.3%)	32 (46%)	2 (5.3%)	1 (25%)	0 (0%)
Missing	1 (1.5%)	9 (1.4%)	1 (1.8%)	0 (0%)	0 (0%)	0 (0%)	19 (28%)	0 (0%)	0 (0%)	0 (0%)
Support Level (44 Wks)										
None	11 (17%)	198 (31%)	12 (21%)	0 (0%)	19 (48%)	10 (31%)	12 (17%)	5 (13%)	2 (50%)	0 (0%)
NIPPV	18 (28%)	97 (15%)	7 (12%)	0 (0%)	9 (23%)	0 (0%)	13 (19%)	0 (0%)	1 (25%)	1 (100%)
IPPV	32 (49%)	96 (15%)	1 (1.8%)	0 (0%)	3 (7.5%)	2 (6.3%)	22 (32%)	0 (0%)	1 (25%)	0 (0%)
Missing	4 (6.2%)	239 (38%)	37 (65%)	60 (100%)	9 (23%)	20 (63%)	22 (32%)	33 (87%)	0 (0%)	0 (0%)
Tracheostomy Needed	27 (42%)	64 (10%)	1 (1.8%)	11 (18%)	5 (13%)	1 (3.1%)	35 (51%)	1 (2.6%)	0 (0%)	1 (100%)

<sup>1</sup> n (%)

## Model Development

Using variables selected from best subset selection, four logistic mixed-effects models were fit on the training data using a random intercept for hospital center ID.

Model 1 (36 week, no interactions) included fixed effects for respiratory support level and normalized FiO<sub>2</sub>. Model 2 (36 week, interactions) included fixed effects for the prenatal corticosteroids indicator, respiratory support level at 36 weeks, normalized FiO<sub>2</sub> at 36 weeks, and the interaction between respiratory support level and normalized FiO<sub>2</sub>. Model 3 (44 week, no interactions) included fixed effects for the prenatal corticosteroids indicator and respiratory support level at 44 weeks. Model 4 (44 week, interactions) included fixed effects for prenatal corticosteroids, respiratory support level at 44 weeks, an indicator for medication for pulmonary hypertension at 44 weeks, and the interaction between respiratory support level and medication for pulmonary hypertension. A summary of the model coefficients are reported in Table 4.

Prenatal corticosteroids, normalized FiO<sub>2</sub>, invasive positive pressure (IPPV) at 36 and 44 weeks, and medication for pulmonary hypertension (MedPH) at 44 weeks were associated with an increased risk of tracheostomy across all models. The interaction between IPPV at 44 weeks and MedPH at 44 weeks were associated with a decreased risk of tracheostomy, modulating each covariate's individual positive effect. Likewise, the individual effect of FiO<sub>2</sub> was only negative when the interaction between respiratory support and FiO<sub>2</sub> was included (Model 2).

Table 4: Model Coefficients

Covariate	Model 1	Model 2	Model 3	Model 4
Intercept	-2.34 (SD 2.33)	-3.05 (SD 2.34)	-4.34 (SD 2.46)	-4.56 (SD 2.19)
Prenatal Corticosteroids	-	1.03	1.31	1.24
NIPPV (36 Wks)	-0.28	-0.56	-	-
IPPV (36 Wks)	1.07	0.81	-	-
FiO2 (36 Wks)	0.46	-0.27	-	-
NIPPV*FiO2 (36 Wks)	-	1.58	-	-
IPPV*FiO2 (36 Wks)	-	0.67	-	-
NIPPV (44 Wks)	-	-	0.58	0.39
IPPV (44 Wks)	-	-	2.60	2.91
MedPH (44 Wks)	-	-	-	1.13
NIPPV*MedPH (44 Wks)	-	-	-	0.35
IPPV*MedPH (44 Wks)	-	-	-	-1.47

NIPPV: Requires non-invasive positive pressure respiratory support.

IPPV: Requires invasive positive pressure respiratory support.

FiO2: Fraction of inspired oxygen needed.

MedPH: Medication for pulmonary hypertension.

Table 5: Model Intercept by Center

Center ID	Model 1	Model 2	Model 3	Model 4
1	-0.94	-1.54	-2.89	-3
2	-2.67	-3.2	-4.67	-4.76
3	-6.01	-6.7	-8.19	-8.2
4	-1.84	-2.68	-3.6	-3.81
5	-2.22	-3.19	-4.01	-4.15
7	-3.19	-3.94	-5.26	-5.53
12	-0.79	-1.39	-2.44	-2.64
16	-3.29	-4.01	-5.54	-5.69
20	-3.59	-4.37	-5.94	-6.03
21	2.1	1.4	0.08	-1.11

Table 4 reports each model’s intercept as its mean and standard deviation, whereas Table 5 reports the individual random intercepts for each hospital center ID. The intercepts for each center in the mixed-effects model represent the baseline level for the probability of tracheostomy (in the logit scale). Center 3 tended to have the lowest random intercept across all models, suggesting that, on average, this center had lower levels of the response variable compared to the overall average. This aligns with the patterns observed in Table 3, where Center 3 exhibited a low incidence of tracheostomy (1.8%).

Conversely, Center 21 had the highest random intercept across models, implying that this center tended to have higher levels of tracheostomy placement compared to the overall average. This finding is consistent with the data in Table 3 as well, where Center 21 had a 100% incidence of tracheostomy, albeit with a sample size of only one patient.

## Model Evaluation

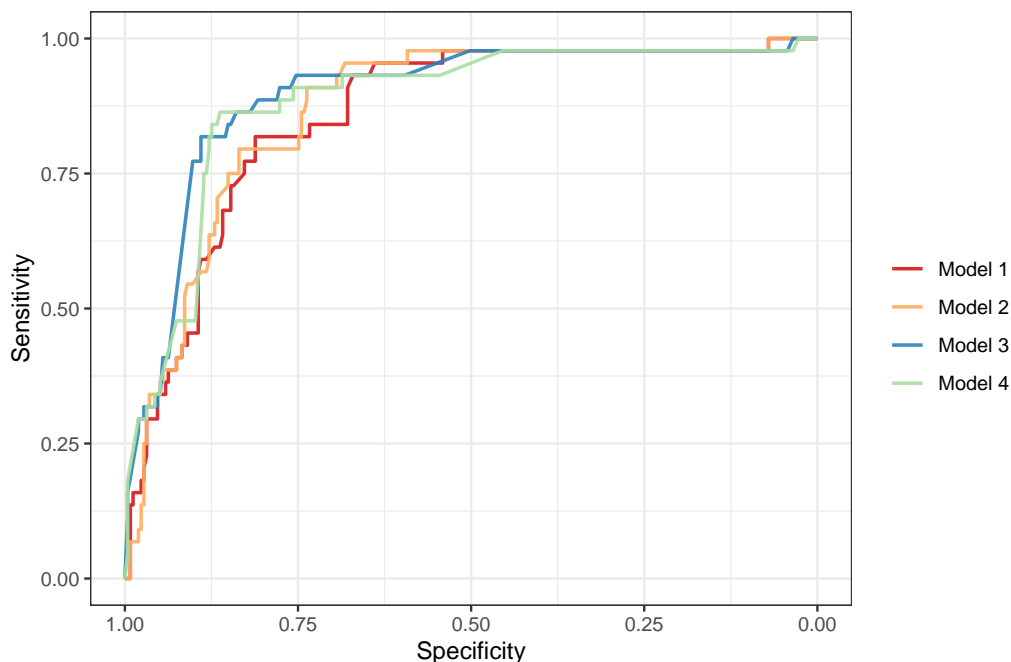
The optimal threshold, AUC, Brier score, precision, recall, and F1 score for each model is reported in Table 6. The optimal threshold represents the cut-off point that maximizes the distance to the identity line in the Receiver Operating Characteristic (ROC) curve. In other words, it is the threshold value that best balances sensitivity and specificity, optimizing the trade-off between correctly identifying positive cases and correctly identifying negative cases. The optimal ROC curves are visualized in Figure 2.



Table 6: Model Evaluation

	Threshold	AUC	Brier Score	Precision	Recall	F1 Score
Model 1	0.186	0.854	0.099	0.417	0.818	0.552
Model 2	0.125	0.863	0.098	0.385	0.877	0.533
Model 3	0.294	0.875	0.089	0.510	0.827	0.631
Model 4	0.243	0.858	0.093	0.492	0.827	0.617

Figure 2. Model ROC Curves



In our models, the optimal thresholds ranged from 0.125 to 0.294. These values are lower than the conventional 0.50 threshold that may be intuitively expected for binary classifications. This may be due to class imbalance in the dataset, where only 15% of observations needed a tracheostomy. The prevalence of observations without tracheostomy could bias the model towards predicting the majority class, prompting the need for lower thresholds to effectively capture the minority class.

Models 1 and 2 used 36-week data. Although Model 2 had a higher AUC and lower Brier score, it had a lower F1 score because it struggled with precision. Due to the class imbalance in the data, we used F1 score as the deciding metric to choose between these models. Thus, we chose Model 1 as the best model for 36-week data. Models 3 and 4 used 44-week data. Model 3 outperformed Model 4 using all metrics, with higher AUC and F1 scores and a lower Brier score. We chose Model 3 as the best model for 44-week data.

Based on these metrics using the validation data, both models that included only main effects with no interactions outperformed the more complex models that included interaction terms. This emphasizes the importance of model simplicity, as the inclusion of interaction terms in these cases did not contribute to model performance when fit on validation data.

Both 44-week models outperformed the 36-week models. We found that the 44-week variables were more highly associated with tracheostomy placement, so it makes sense that these are better predictors in the model. However, the 44-week model should still be interpreted with caution because it used a large proportion of imputed data which may not be as accurate as the 36-week data.

## Model Coefficients and Interpretation

The final 36-week model can be expressed as so:

Table 7: Coefficients and Odds Ratios for Selected 36-Week Model

Covariate	Coefficient	Odds Ratio (95% CI)	p-value
Intercept	-2.34 (SD 2.33)		
NIPPV (36 Wks)	-0.28	0.76 (0.52, 1.10)	0.145
IPPV (36 Wks)	1.07	2.92 (2.05, 4.16)	<0.001
FiO2 (36 Wks)	0.46	1.58 (1.44, 1.73)	<0.001

Having a respiratory support level of NIPPV at 36 weeks did not have a significant effect on tracheostomy probability in comparison to needing no respiratory support. The odds ratio for IPPV at 36 weeks versus no respiratory support was 2.92, indicating that those needed IPPV support had 2.92 times higher odds of tracheostomy than those that didn't need respiratory support at 36 weeks. Additionally, a one-standard deviation increase in FiO2 was associated with 1.58 times higher odds of tracheostomy occurring. In the observed data, FiO2 had a mean of 0.34 and a standard deviation of 0.15. .

As an example, a patient from the average center with no respiratory support and inspired oxygen fraction of 0.34 (average) would have a risk of tracheostomy of  $e^{-2.34}/(1 + e^{-2.34}) = 0.088$ . A patient from the average center requiring IPPV support and an inspired oxygen fraction of 0.49 (average + 1SD) would have a tracheostomy risk of  $e^{-2.34+1.07+0.46}/(1 + e^{-2.34+1.07+0.46}) = 0.308$ , which using the optimal threshold of 0.186 would classify them as needing a tracheostomy.

The final 44-week model can be expressed as so:

Table 8: Coefficients and Odds Ratios for Selected 44-Week Model

Covariate	Coefficient	Odds Ratio (95% CI)	p-value
Intercept	-4.34 (SD 2.46)		
Prenatal Corticosteroids	1.31	3.72 (3.03, 4.57)	<0.001
NIPPV (36 Wks)	0.58	1.78 (1.53, 2.08)	<0.001
IPPV (36 Wks)	2.60	13.46 (11.74, 15.43)	<0.001

All effects in this model were statistically significant. Under this model, the odds of tracheostomy for those who received prenatal corticosteroids were 3.72 times higher than those who didn't receive prenatal corticosteroids. The odds of tracheostomy for patients needing NIPPV support was 1.78 times higher than those needing no respiratory support. The odds of tracheostomy for patients needing IPPV support was 13.46 times higher than those needing no respiratory support.

Both the 36 and 44 week models selected respiratory support level as an important predictor. While the 36-week model used fraction of inspired oxygen as an additional predictor, the 44-week model used prenatal corticosteroids as a predictor. The highest odds ratio was observed for IPPV vs. no respiratory support at 46 weeks, suggesting that this is the most influential variable in whether the patient will need a tracheostomy.

## Discussion

This study aimed to develop predictive models for the eventual need for tracheostomy in neonates with BPD using clinical data at 36 or 44 weeks PMA. The findings suggest that the inclusion of respiratory support parameters, especially at 44 weeks, enhances the predictive accuracy of the models. This aligns with the clinical understanding that the severity of BPD is partially defined by the level of respiratory support required, with more severe cases requiring a tracheostomy. Specifically, neonates relying on invasive positive pressure

(IPPV) at 36 weeks would be considered as “severe” BPD according to the Jenson (2019) definition [4] and had a significantly higher odds of needing a tracheostomy according to our model.

The findings also emphasize the importance of timing in assessing respiratory parameters. The 44-week model, incorporating data closer to the anticipated discharge, outperformed the 36-week model. This suggests that the respiratory status of neonates in the weeks leading up to discharge is a more informative predictor than earlier measures.

The inclusion of prenatal corticosteroids as a significant predictor in the 44-week model is noteworthy, with an odds ratio of 3.72. While corticosteroids are a standard intervention to enhance lung maturity in preterm infants, their association with an increased risk of tracheostomy warrants further investigation. This finding could be attributed to the complex interplay between lung development, corticosteroid effects, and the underlying pathology of sBPD.

While this study contributes insights into which variables predict the need for tracheostomy, several limitations should be acknowledged. First, the dataset exhibits class imbalance, with only 15% of observations requiring tracheostomy. This imbalance may introduce bias, potentially influencing the model toward predicting the majority class. The need for lower thresholds in the optimal threshold values, ranging from 0.125 to 0.294, reflects this challenge.

Second, the relatively small sample size, especially considering class imbalance and the hierarchical structure of the data collected from different hospital centers, could limit the generalizability of the findings. While mixed-effects models account for this variability, larger and more diverse datasets would strengthen the external validity of the predictive models.

Third, the retrospective nature of the dataset restricts the establishment of causal relationships. The absence of detailed comorbidity data introduces potential confounding factors that could influence the outcomes. Prospective studies with comprehensive data collection, including comorbidities and long-term follow-up, would provide a more nuanced understanding of the factors influencing tracheostomy outcomes in neonates with sBPD.

Fourth, the reliance on imputed data, especially for the 44-week variables, introduces another layer of uncertainty. While multiple imputation was employed to mitigate missing data, the accuracy of predictions based on imputed values may vary. Future research should aim to collect complete and accurate data at both timepoints to enhance the reliability of the models.

Despite these limitations, the current study lays the groundwork for further investigations into the predictive factors of tracheostomy in neonates with BPD. Future research should focus on overcoming these limitations, integrate the 36-week and 44-week variables into a single model, and validate the models in larger and more diverse cohorts to ensure their transportability across varied settings.

## Conclusions

This study provides an analysis of predictive factors for tracheostomy in neonates with bronchopulmonary dysplasia. The incorporation of respiratory parameters, especially at 44 weeks, enhances the accuracy of predictive models. Respiratory support level remains a key predictor, and the role of prenatal corticosteroids in influencing tracheostomy risk warrants further exploration. The findings have implications for refining clinical decision-making and fostering a nuanced understanding of the factors shaping the respiratory trajectory of infants with severe bronchopulmonary dysplasia.

## References

- [1] Jensen, E. A., & Schmidt, B. (2014). Epidemiology of bronchopulmonary dysplasia. *Birth defects research. Part A, Clinical and molecular teratology*, 100(3), 145–157. <https://doi.org/10.1002/bdra.23235>
- [2] McKinney, R., & Levin, J. (October 16, 2023). Predicting the need for tracheostomy in infants with severe bronchopulmonary dysplasia. Presentation.

- [3] Dassios, T., Williams, E., Hickey, A., Bhat, R., Greenough, A. (2022). Mortality after 36 weeks postmenstrual age of extremely preterm infants in neonatal care: The impact of growth impairment and bronchopulmonary dysplasia. *Early Human Development*, 171. <https://doi.org/10.1016/j.earlhumdev.2022.105618>.
- [4] Jensen, E. A., Dysart, K., Gantz, M. G., McDonald, S., Bamat, N. A., Keszler, M., Kirpalani, H., Laughon, M. M., Poindexter, B. B., Duncan, A. F., Yoder, B. A., Eichenwald, E. C., & DeMauro, S. B. (2019). The Diagnosis of Bronchopulmonary Dysplasia in Very Preterm Infants. An Evidence-based Approach. *American journal of respiratory and critical care medicine*, 200(6), 751–759. <https://doi.org/10.1164/rccm.201812-2348OC>

## Code Appendix:

```
knitr::opts_chunk$set(echo = FALSE)
library(tidyverse)
library(gtsummary)
library(corrplot)
library(knitr)
library(mice)
library(L0Learn)
library(lme4)
library(caret)
library(pROC)

# Read in and process data
df <- read.csv("project2.csv") %>%
  distinct() %>%
  select(-c(record_id, hosp_dc_ga, Death, mat_race)) %>%
  mutate_at(c("center", "mat_ethn", "del_method", "prenat_ster",
              "com_prenat_ster", "mat_chorio", "gender", "sga", "any_surf",
              "ventilation_support_level.36", "med_ph.36",
              "ventilation_support_level_modified.44", "med_ph.44", "Trach"),
            as.factor)

# Define ventilation support levels
levels(df$ventilation_support_level.36) <- c("None", "NIPPV", "IPPV")
levels(df$ventilation_support_level_modified.44) <- c("None", "NIPPV", "IPPV")

# Fill in missing 'center' values (based on patient ID)
df$center[is.na(df$center)] <- 1

# Save `center` as a vector
center_vec <- df$center

# Define variable names for correlation plot
df_eda <- df %>%
  rename(`Maternal Ethnicity` = mat_ethn,
         `Weight (Birth)` = bw,
         `Gestational Age (Birth)` = ga,
         `Length (Birth)` = blength,
         `Head Circumference (Birth)` = birth_hc,
         `Delivery Method` = del_method,
         `Prenatal Corticosteroids` = prenat_ster,
         `Complete Prenatal Steroids` = com_prenat_ster,
         `Maternal Chorioamnionitis` = mat_chorio,
         `Sex Assigned at Birth` = gender,
         `SGA` = sga,
         `Surfactant in First 72 Hours` = any_surf,
         `Weight (36 Weeks)` = weight_today.36,
         `Ventilation Support Level (36 Weeks)` = ventilation_support_level.36,
         `FiO2 (36 Weeks)` = inspired_oxygen.36,
         `PIP (36 Weeks)` = p_delta.36,
         `PEEP (36 Weeks)` = peep_cm_h2o_modified.36,
         `Medication for Pulmonary Hypertension (36 Weeks)` = med_ph.36,
```

```

`Weight (44 Weeks)` = weight_today.44,
`Ventilation Support Level (44 Weeks)` = ventilation_support_level_modified.44,
`FiO2 (44 Weeks)` = inspired_oxygen.44,
`PIP (44 Weeks)` = p_delta.44,
`PEEP (44 Weeks)` = peep_cm_h2o_modified.44,
`Medication for Pulmonary Hypertension (44 Weeks)` = med_ph.44,
`Tracheostomy` = Trach) %>%
select(-center)

# Create correlation matrix
cor_mat <- df_eda %>%
  select(`Weight (Birth)`, `Gestational Age (Birth)`, `Length (Birth)`,
    `Head Circumference (Birth)`, `Weight (36 Weeks)`,
    `FiO2 (36 Weeks)`, `PIP (36 Weeks)`, `PEEP (36 Weeks)`,
    `Weight (44 Weeks)`,
    `FiO2 (44 Weeks)`, `PIP (44 Weeks)`, `PEEP (44 Weeks)`) %>%
  cor(use = "complete.obs")

# Plot correlation matrix
corrplot(cor_mat, method = "square", type = "lower", diag = FALSE,
  tl.cex = 0.75, tl.col = "black", tl.srt = 20)

# Create table with prevalence of missing data
data.frame(variable = names(colMeans(is.na(df_eda))),
  perc_missing = paste0(round(100*colMeans(is.na(df_eda)),1),"%"),
  row.names = NULL) %>%
  kableExtra::kbl(format = "latex", caption = "Prevalence of Missing Data",
    col.names = c("Variable", "Percent Missing"),
    booktabs = T) %>%
  kableExtra::kable_styling(latex_options = c("HOLD_position"), font_size = 8)

# Define candidate variables
candidate_vars.36 <- c("center", "bw", "del_method", "prenat_ster", "mat_chorio", "gender",
  "sga", "weight_today.36", "ventilation_support_level.36",
  "inspired_oxygen.36", "p_delta.36", "peep_cm_h2o_modified.36",
  "med_ph.36", "Trach")

candidate_vars.44 <- c("center", "bw", "del_method", "prenat_ster", "mat_chorio", "gender",
  "sga", "weight_today.44", "ventilation_support_level_modified.44",
  "inspired_oxygen.44", "p_delta.44", "peep_cm_h2o_modified.44",
  "med_ph.44", "Trach")

### Multiple Imputation
# Set the number of imputations
m = 5

# Split data into test and validation sets
set.seed(1)
test_index <- createDataPartition(df$Trach, p = 0.3, list = FALSE)
train_index <- seq(1:nrow(df))[-test_index]
traindata <- df[train_index, ]
testdata <- df[test_index, ]

```

```

# Run multiple imputation
imp.train <- mice(select(traindata, -center), m = m, print = FALSE, seed = 123)
imp.test <- mice.mids(imp.train, newdata = select(testdata, -center), print = FALSE)

# Save imputed training data sets
train.36 <- vector("list", m)
train.44 <- vector("list", m)
for (i in 1:m) {
  df_imp <- mice::complete(imp.train, i)
  df_imp$center <- center_vec[train_index]

  train.36[[i]] <- df_imp %>%
    select(all_of(candidate_vars.36)) %>%
    mutate_at(c("bw", "weight_today.36", "inspired_oxygen.36", "p_delta.36",
               "peep_cm_h2o_modified.36"), function(x) (x - mean(x))/sd(x))

  train.44[[i]] <- df_imp %>%
    select(all_of(candidate_vars.44)) %>%
    mutate_at(c("bw", "weight_today.44", "inspired_oxygen.44", "p_delta.44",
               "peep_cm_h2o_modified.44"), function(x) (x - mean(x))/sd(x))
}

# Save imputed validation data sets
val.all <- vector("list", m)
for (i in 1:m) {
  val.all[[i]] <- mice::complete(imp.test, i)
  val.all[[i]]$center <- center_vec[test_index]
}

# Best subset selection
best_subset <- function(df_list, interaction = FALSE, week) {

  coefs_ls <- vector("list", m)
  for (i in 1:m) {

    # Process data
    df_bestfit <- df_list[[i]] %>%
      rename(y = Trach) %>%
      select(-center)

    # Define model matrix
    if (interaction == FALSE) {
      bf_X <- model.matrix(y~., data = df_bestfit)[,-1]
    } else if (week == 36) {
      bf_X <- model.matrix(y~.*ventilation_support_level.36, data = df_bestfit)[,-1]
    } else {
      bf_X <- model.matrix(y~.*ventilation_support_level_modified.44, data = df_bestfit)[,-1]
    }

    # Define outcome
    bf_y <- df_bestfit$y

    # Run 5-fold CV

```

```

cvfit <- L0Learn.cvfit(bf_X, bf_y, nFolds = 5, seed = 1, penalty = "L0",
                      nGamma = 1, gammaMin = 0.0001, gammaMax = 0.1,
                      loss = "Logistic")

# Identify optimal lambda
optimalLambdaIndex = which.min(cvfit$cvMeans[[1]])
optimalLambda = cvfit$fit$lambda[[1]][optimalLambdaIndex]

# Fit best subsets using optimal lambda
fit_coef <- coef(cvfit, lambda = optimalLambda,
                 gamma = 1e-7)[,1]

# Save coefficients
coefs_ls[[i]] <- fit_coef
names(coefs_ls[[i]]) <- c("Intercept", colnames(bf_X))
}

# Save coefficients for all imputations in a matrix
coefs <- coefs_ls[[1]]
for (i in 2:m) {
  coefs <- cbind(coefs, coefs_ls[[i]])
}

# Return proportion of imputations that each covariate was selected
return(apply(coefs, 1, function(x) mean(x > 0)))
}

# Run best subset selection for each model type
results.36 <- best_subset(train.36, interaction = FALSE, week = 36)
results_int.36 <- best_subset(train.36, interaction = TRUE, week = 36)
results.44 <- best_subset(train.44, interaction = FALSE, week = 44)
results_int.44 <- best_subset(train.44, interaction = TRUE, week = 44)

# Print covariates that were selected in >80% of imputations
print("36")
results.36[which(results.36 >= 0.8)]
print("36, interaction")
results_int.36[which(results_int.36 >= 0.8)]
print("44")
results.44[which(results.44 >= 0.8)]
print("44, interaction")
results_int.44[which(results_int.44 >= 0.8)]

# Combine training imputations into long dataset
train_long.36 <- train.36[[1]]
for (i in 1:m) {
  train_long.36 <- rbind(train_long.36, train.36[[i]])
}

train_long.44 <- train.44[[1]]
for (i in 1:m) {

```



```

train_long.44 <- rbind(train_long.44, train.44[[i]])
}

# Fit Model 1
mod.36 <- glmer(Trach ~ ventilation_support_level.36 +
  inspired_oxygen.36 +
    (1 | center),
  family = binomial(),
  data = train_long.36)

# Fit Model 2
mod_int.36 <- glmer(Trach ~ prenat_ster + ventilation_support_level.36 *
  inspired_oxygen.36 +
    (1 | center),
  family = binomial(),
  data = train_long.36)

# Fit Model 3
mod.44 <- glmer(Trach ~ prenat_ster +
  ventilation_support_level_modified.44 +
    (1 | center),
  family = binomial(),
  data = train_long.44)

# Fit Model 4
mod_int.44 <- glmer(Trach ~ prenat_ster +
  ventilation_support_level_modified.44 +
  med_ph.44*ventilation_support_level_modified.44 +
    (1 | center),
  family = binomial(),
  data = train_long.44)

# Get evaluation metrics
get_metrics <- function(model) {

  threshold <- rep(NA, m)
  auc <- rep(NA, m)
  brier <- rep(NA, m)
  precision <- rep(NA, m)
  recall <- rep(NA, m)
  f1_score <- rep(NA, m)

  # Iterate through imputed validation sets
  for (i in 1:m) {

    # Get observed and predicted outcomes for validation set
    y_obs <- val.all[[i]]$Trach
    y_pred <- predict(model, newdata = val.all[[i]], type = "response")

    # Save ROC object
    roc_obj <- roc(y_obs, y_pred)

    # Calculate metrics

```

```

threshold[i] <- coords(roc_obj, "best", ret = "threshold")[[1]]
auc[i] <- roc_obj$auc
brier[i] <- mean((y_pred - as.numeric(y_obs) + 1)^2)
precision[i] <- coords(roc_obj, "best", ret = "precision")[[1]]
recall[i] <- coords(roc_obj, "best", ret = "recall")[[1]]
f1_score[i] <- 2 * (precision[i] * recall[i]) / (precision[i] + recall[i])

}

# Return mean metrics across imputed data
return(data.frame(threshold = mean(threshold), auc = mean(auc),
                  brier = mean(brier), precision = mean(precision), recall = mean(recall),
                  f1_score = mean(f1_score)))

}

# Table with patient characteristics
tbl_summary(df, by = Trach,
            include = c(bw, del_method,
                        prenat_ster, mat_chorio, gender, sga, weight_today.36,
                        ventilation_support_level.36,
                        inspired_oxygen.36, p_delta.36, peep_cm_h2o_modified.36,
                        med_ph.36, weight_today.44,
                        ventilation_support_level_modified.44,
                        inspired_oxygen.44, p_delta.44, peep_cm_h2o_modified.44,
                        med_ph.44),
            missing = "no",
            type = list(del_method ~ "dichotomous",
                        prenat_ster ~ "dichotomous",
                        mat_chorio ~ "dichotomous",
                        gender ~ "dichotomous",
                        sga ~ "dichotomous",
                        med_ph.36 ~ "dichotomous",
                        med_ph.44 ~ "dichotomous"),
            value = list(del_method = "2",
                        prenat_ster = "Yes",
                        mat_chorio = "Yes",
                        gender = "Male",
                        sga = "SGA",
                        med_ph.36 = "1",
                        med_ph.44 = "1"),
            label = c(bw ~ "Weight at Birth",
                      del_method ~ "Cesarean Section Delivery",
                      prenat_ster ~ "Prenatal Corticosteroids",
                      mat_chorio ~ "Maternal Chorioamnionitis",
                      gender ~ "Sex Assigned at Birth",
                      sga ~ "Small for Gestational Age",
                      weight_today.36 ~ "Weight at 36 Weeks",
                      ventilation_support_level.36 ~ "Ventilation Support Level at 36 Weeks",
                      inspired_oxygen.36 ~ "Fraction of Inspired Oxygen at 36 Weeks",
                      p_delta.36 ~ "Peak Inspiratory Pressure at 36 Weeks",
                      peep_cm_h2o_modified.36 ~ "Positive and Exploratory Pressure at 36 Weeks",
                      med_ph.36 ~ "Medication for Pulmonary Hypertension at 36 Weeks",

```

```

weight_today.44 ~ "Weight at 44 Weeks",
ventilation_support_level_modified.44 ~ "Ventilation Support Level at 44 Weeks",
inspired_oxygen.44 ~ "Fraction of Inspired Oxygen at 44 Weeks",
p_delta.44 ~ "Peak Inspiratory Pressure at 44 Weeks",
peep_cm_h2o_modified.44 ~ "Positive and Exploratory Pressure at 44 Weeks",
med_ph.44 ~ "Medication for Pulmonary Hypertension at 44 Weeks"),
statistic = list(all_continuous() ~
  c("{mean} ({sd}")))) %>%
modify_header(list(stat_1 ~ "No Tracheostomy (N = 850)",
  stat_2 ~ "Tracheostomy (N = 146)")) %>%
add_p() %>%
as_kable_extra(booktabs = TRUE,
  caption = "Candidate Variable Summary by Tracheostomy Status",
  longtable = TRUE) %>%
kableExtra::kable_styling(font_size = 8,
  latex_options = c("repeat_header", "HOLD_position"))

# Table with characteristics by center
df_center <- df %>%
  select(center, ventilation_support_level.36,
    ventilation_support_level_modified.44, Trach)

df_center$ventilation_support_level.36 <-
  fct_na_value_to_level(df_center$ventilation_support_level.36, "Missing")

df_center$ventilation_support_level_modified.44 <-
  fct_na_value_to_level(df_center$ventilation_support_level_modified.44, "Missing")

tbl_summary(df_center, include = c(ventilation_support_level.36,
  ventilation_support_level_modified.44,
  Trach),
  by = center, missing = "no",
  type = list(Trach ~ "dichotomous"),
  value = list(Trach ~ "1"),
  label = c(ventilation_support_level.36 ~ "Support Level (36 Wks)",
    ventilation_support_level_modified.44 ~ "Support Level (44 Wks)",
    Trach ~ "Tracheostomy Needed")) %>%
as_kable_extra(booktabs = TRUE,
  caption = "Hospital Center Characteristics") %>%
kableExtra::kable_styling(font_size = 8,
  latex_options = c("scale_down", "repeat_header", "HOLD_position"))

# Table with all model coefficients
data.frame(covariates = c("Intercept", "Prenatal Corticosteroids", "NIPPV (36 Wks)",
  "IPPV (36 Wks)", "FiO2 (36 Wks)", "NIPPV*FiO2 (36 Wks)",
  "IPPV*FiO2 (36 Wks)", "NIPPV (44 Wks)", "IPPV (44 Wks)",
  "MedPH (44 Wks)", "NIPPV*MedPH (44 Wks)", "IPPV*MedPH (44 Wks)"),
  model1 = kableExtra::linebreak(c("-2.34\\n(SD 2.33)",
    "-", "-0.28", "1.07",
    "0.46", "-", "-",
    "-", "-", "-", "-",
    "-")),
  model2 = kableExtra::linebreak(c("-3.05\\n(SD 2.34)",

```

```

        "1.03", "-0.56",
        "0.81", "-0.27",
        "1.58", "0.67",
        "-", "-", "-", "-",
        "-")),
    model3 = kableExtra::linebreak(c("-4.34\\n(SD 2.46)",
        "1.31", "-",
        "-", "-",
        "-", "-",
        "0.58", "2.60", "-", "-",
        "-")),
    model4 = kableExtra::linebreak(c("-4.56\\n(SD 2.19)",
        "1.24", "-",
        "-", "-",
        "-", "-",
        "0.39", "2.91", "1.13", "0.35",
        "-1.47")))) %>%
kableExtra::kbl(format = "latex", caption = "Model Coefficients",
    col.names = c("Covariate", "Model 1", "Model 2", "Model 3", "Model 4"),
    booktabs = T, escape = "FALSE") %>%
kableExtra::footnote("MedPH: Medication for pulmonary hypertension.", general_title = "") %>%
kableExtra::footnote("FiO2: Fraction of inspired oxygen needed.", general_title = "") %>%
kableExtra::footnote("IPPV: Requires invasive positive pressure respiratory support.", general_title = "") %>%
kableExtra::footnote("NIPPV: Requires non-invasive positive pressure respiratory support.", general_title = "") %>%
kableExtra::kable_styling(latex_options = c("HOLD_position"), font_size = 8)

# Table with random intercepts
centers <- row.names(coef(mod.36)$center)
cbind(centers,
    round(coef(mod.36)$center[,1],2),
    round(coef(mod_int.36)$center[,1],2),
    round(coef(mod.44)$center[,1],2),
    round(coef(mod_int.44)$center[,1],2)) %>%
data.frame() %>%
kableExtra::kbl(format = "latex",
    col.names = c("Center ID", "Model 1", "Model 2", "Model 3",
        "Model 4"),
    caption = "Model Intercept by Center",
    booktabs = T, digits = 2) %>%
kableExtra::kable_styling(latex_options = c("HOLD_position"),font_size = 8)

# Table with evaluation metrics
rbind(get_metrics(mod.36), get_metrics(mod_int.36),
    get_metrics(mod.44), get_metrics(mod_int.44)) %>%
round(3) %>%
data.frame(row.names = c("Model 1", "Model 2", "Model 3", "Model 4")) %>%
kableExtra::kbl(format = "latex", booktabs = T,
    col.names = c("Threshold", "AUC", "Brier Score",
        "Precision", "Recall", "F1 Score"),
    caption = "Model Evaluation") %>%
kableExtra::kable_styling(latex_options = c("HOLD_position"),font_size = 8)

```

```

# Get ROC object
get_roc <- function(model) {
  y_obs <- val.all[[1]]$Trach
  y_pred_all <- matrix(NA, nrow = length(y_obs), ncol = m)
  for (i in 1:m) {

    # Get observed and predicted outcomes for validation set
    y_pred_all[,i] <- predict(model, newdata = val.all[[i]], type = "response")

  }

  y_pred <- rowSums(y_pred_all)
  roc_obj <- roc(y_obs, y_pred)
}

roc.36 <- get_roc(mod.36)
roc_int.36 <- get_roc(mod_int.36)
roc.44 <- get_roc(mod.44)
roc_int.44 <- get_roc(mod_int.44)

# Plot ROC curves
ggroc(list(Model1 = roc.36, Model2 = roc_int.36, Model3 = roc.44, Model4 = roc_int.44),
      alpha = 0.9, size = 0.75) +
  scale_color_manual(values = c("#D7191C", "#FDAE61", "#2B83BA", "#ABDDA4"),
                    labels = c("Model 1", "Model 2", "Model 3", "Model 4")) +
  theme_bw() +
  labs(x = "Specificity", y = "Sensitivity", title = "Figure 2. Model ROC Curves",
       color = "")

# Table with final 36-week coefficients and ORs
data.frame(covariates = c("Intercept", "NIPPV (36 Wks)", "IPPV (36 Wks)",
                          "FiO2 (36 Wks)"),
           coefficients = c("-2.34 (SD 2.33)", "-0.28", "1.07", "0.46"),
           oddsratio = c("", "0.76 (0.52, 1.10)", "2.92 (2.05, 4.16)",
                         "1.58 (1.44, 1.73)"),
           p_value = c("", "0.145", "<0.001", "<0.001")) %>%
  kableExtra::kbl(format = "latex", caption = "Coefficients and Odds Ratios for Selected 36-Week Model",
                  col.names = c("Covariate", "Coefficient", "Odds Ratio (95% CI)", "p-value"),
                  booktabs = T) %>%
  kableExtra::kable_styling(latex_options = c("HOLD_position"), font_size = 8)

# Table with final 44-week coefficients and ORs
data.frame(covariates = c("Intercept", "Prenatal Corticosteroids", "NIPPV (36 Wks)", "IPPV (36 Wks)"),
           coefficients = c("-4.34 (SD 2.46)", "1.31", "0.58", "2.60"),
           oddsratio = c("", "3.72 (3.03, 4.57)", "1.78 (1.53, 2.08)", "13.46 (11.74, 15.43)",
                         "1.58 (1.44, 1.73)"),
           p_value = c("", "<0.001", "<0.001", "<0.001")) %>%
  kableExtra::kbl(format = "latex", caption = "Coefficients and Odds Ratios for Selected 44-Week Model",
                  col.names = c("Covariate", "Coefficient", "Odds Ratio (95% CI)", "p-value"),
                  booktabs = T) %>%
  kableExtra::kable_styling(latex_options = c("HOLD_position"), font_size = 8)

```