Predicting the need for tracheostomy in infants with severe bronchopulmonary dysplasia

Abstract

Severe bronchopulmonary dysplasia (sBPD) in neonates, a chronic lung disease affecting premature infants, poses significant healthcare challenges. A crucial aspect of managing sBPD involves determining the need for tracheostomy, a decision that currently lacks standardized criteria and varies across medical centers. This study develops a predictive regression model to forecast the need for tracheostomy or the occurrence of death in infants with sBPD, using a comprehensive dataset from various Neonatal Intensive Care Units (NICUs).

We utilized a national dataset containing demographic, diagnostic, and respiratory parameters of infants diagnosed with sBPD. The data, including crucial parameters measured at 36 and 44 weeks Post-Menstrual Age (PMA), underwent rigorous preprocessing. Our primary outcome was a composite measure of tracheostomy placement or death. We employed the Least Absolute Shrinkage and Selection Operator (LASSO) regression to develop three distinct models based on different time frames and variables.

Our study found that approximately 14.66% of the infants in our dataset underwent tracheostomy. Key predictors identified across the models included maternal ethnicity, delivery method, prenatal steroid use, and specific respiratory support parameters. The 36-week model specifically highlighted the significance of invasive ventilation support at 36 weeks. The performance of the models was assessed using sensitivity, specificity, F1 score, and Brier score, with the 36-week model demonstrating the most balanced performance.

The study successfully identifies critical predictors for tracheostomy in sBPD, including prenatal steroids, respiratory support level, and maternal factors. However, the findings should be interpreted with caution due to potential overfitting and data variability across centers. Future research should focus on validating these models in larger cohorts, addressing class imbalance and sample size limitations for enhanced reliability and generalizability.

Introduction

Severe bronchopulmonary dysplasia (sBPD) in neonates, a chronic lung disease affecting premature infants, presents significant healthcare challenges. Its management is complex due to the disease's multifaceted nature, impacting both short-term clinical decisions and long-term developmental outcomes. A crucial decision in the management of sBPD involves the timing of tracheostomy placement, a procedure that can significantly influence the infant's respiratory support needs and overall health trajectory.

Despite its critical importance, the decision-making process for tracheostomy in sBPD patients lacks standardized criteria, often resulting in varied practices across different medical centers. This variability stems from the individual differences in disease progression, response to treatment, and the influence of a broad spectrum of demographic, diagnostic, and respiratory factors.

In light of these challenges, this research aims to develop a predictive regression model using a comprehensive national dataset. This dataset encompasses a wide array of demographic, diagnostic, and respiratory parameters of infants with sBPD admitted to various collaborative Neonatal Intensive Care Units (NICUs). The model's objective is to predict the composite outcome of tracheostomy placement or death, providing a quantitative tool to aid in clinical decision-making.

Methods

Data Source and Preprocessing

Our analysis utilized a national dataset including demographic, diagnostic, and respiratory parameters of infants diagnosed with severe bronchopulmonary dysplasia (sBPD) across various collaborative NICUs. The data encompassed crucial parameters measured at 36 and 44 weeks Post-Menstrual Age (PMA).

The initial stage of data preprocessing involved removing duplicate records to ensure data uniqueness. We addressed specific discrepancies and errors, such as an erroneous value in the 'center' variable, where '21' was replaced with '20' based on contextual evidence. Missing values in the 'center' variable were imputed, considering the observed patterns in relation to 'record_id' and 'center'. The 'mat_race' variable was excluded due to inconsistencies in coding and the codebook, ensuring data quality for our analysis.

We encountered instances of data irregularity, such as infants reportedly discharged before reaching 36 weeks PMA yet having recorded data at 36 and even 44 weeks. This highlighted potential errors in recorded Hospital Discharge Gestational Age, necessitating careful scrutiny and handling of such data points.

Variables in the dataset were relabeled for clarity and interpretability. This included assigning descriptive labels to key variables such as patient ID, medical center, maternal ethnicity, birth weight, gestational age, and various respiratory support parameters.

Outcome Measurements

In this study, we defined the primary outcome as a composite measure of tracheostomy placement or death. This decision was informed by clinical considerations that both outcomes are pivotal and often interrelated in the management of sBPD. The composite nature of this outcome accounts for the severity spectrum of sBPD, recognizing that tracheostomy is a significant intervention often associated with more severe cases and may also be an indicator of the risk of mortality.

Upon examination of the dataset, we found that approximately 14.66% of the infants underwent tracheostomy. Our investigation into the relationship between tracheostomy and death revealed a notable distribution of death occurrences, signifying the gravity of these outcomes within this patient cohort. By combining these outcomes into a single composite measure, we increased the statistical power of the study, allowing for more robust analysis in a dataset with a relatively low event rate.

To operationalize this composite outcome, we isolated instances where 'Death' was missing and excluded these cases from the analysis to maintain data integrity. We then created a binary variable, 'trach_death', coding for the presence of either tracheostomy or death.

Figure 1 depicts the percentage distribution of the composite outcome within the study population. The bar chart clearly delineates the relative frequency of the two groups within the composite outcome, with one group experiencing neither tracheostomy nor death ('0') and the other group having at least one of these events ('1').

Exploratory Data Analysis

The EDA revealed a diverse range of birth weights and gestational ages, reflecting the broad spectrum of infant conditions within the dataset. The birth weight of the infants ranged from a minimum of 340 grams to a maximum of 2725 grams, with a median of 745 grams. The gestational age spanned from 22 to 31 weeks.

The respiratory support parameters at 36 and 44 weeks PMA showed significant variability. At 36 weeks, the median value for inspired oxygen (FiO2) was 0.30, with a subset of infants requiring high levels of oxygen support. The distribution of ventilation support levels at 36 weeks PMA indicated that a substantial proportion of infants required some form of respiratory support, whether non-invasive or invasive.

The variable 'center' represents the medical center where each infant in the dataset received care. This categorical variable is significant because it may encapsulate a range of unmeasured factors related to the quality of care, resources available, or specific protocols that can affect patient outcomes. As such, data coming predominantly from center 2, with centers 12 and 4 also contributing significantly, could indicate a potential

center effect where these facilities' specific practices might disproportionately influence the model's results. This suggests that center-specific variations need to be accounted for during the analysis to ensure that the findings are not biased by these institutional differences. The summary of the variables by outcome can be seen in the Table 1.

Characteristic	N	No, N = 811[1]	Yes, N = 183[1]
Medical Center	994		
1		31 (3.8%)	34 (19%)
2		545 (67%)	84 (46%)
3		55 (6.8%)	2 (1.1%)
4		47 (5.8%)	12 (6.6%)
5		33 (4.1%)	7 (3.8%)
7		31 (3.8%)	1 (0.5%)
12		28 (3.5%)	41 (22%)
16		37 (4.6%)	1 (0.5%)
20		4 (0.5%)	1 (0.5%)
Maternal Ethnicity	937		
Hispanic or Latino		64 (8.3%)	10 (5.9%)
Not Hispanic or Latino		703 (92%)	160 (94%)
Birth weight	994	817 (285)	756 (340)
Birth Gestational age	994	26 (2)	26 (2)
Birth length	916	33 (4)	32 (4)
Birth head circumference	917	23.25 (2.65)	22.88 (3.29)
Delivery Method	991		
Vaginal delivery		245 (30%)	39 (21%)
Cesarean section		564 (70%)	143 (79%)
Prenatal Corticosteroids	959		
No		113 (14%)	13 (7.8%)
Yes		679 (86%)	154 (92%)
Complete Prenatal Steroids	801		
No		159 (24%)	34 (24%)
Yes		499 (76%)	109 (76%)
Maternal Chorioamnionitis	932		
No		633 (83%)	139 (83%)
Yes		132 (17%)	28 (17%)
Gender	990		
Female		334 (41%)	73 (40%)
Male		473 (59%)	110 (60%)
Small for GA	979		
Not SGA		658 (82%)	118 (66%)
SGA		142 (18%)	61 (34%)
Surfactant in first 72h	562		
No		89 (19%)	12 (12%)
Yes		374 (81%)	87 (88%)
36 week weight	902	2,142 (393)	1,981 (507)
36 weeks ventilation support	964		
No support		109 (14%)	7 (4.3%)
Non-invasive pressure		553 (69%)	36 (22%)
Invasive pressure		140 (17%)	119 (73%)
36 weeks FiO2	902	0.31 (0.12)	0.49 (0.21)
36 weeks PIP	866	4 (8)	16 (12)
36 weeks PEEP	877	6 (3)	7 (3)
36 weeks Meds for PH	964		
No		770 (96%)	129 (80%)
Yes		32 (4.0%)	33 (20%)
44 weeks weight	550	3,695 (643)	3,473 (782)
44 weeks ventilation support	572		
No support		261 (60%)	8 (6.0%)
Non-invasive pressure		124 (28%)	22 (16%)
Invasive pressure		53 (12%)	104 (78%)
44 weeks FiO2	548	0.31 (0.11)	0.45 (0.20)
44 weeks PIP	548	4 (11)	22 (16)
44 weeks PEEP	550	3 (4)	9 (3)
44 weeks Meds for PH	572		
No		405 (92%)	68 (51%)
Yes		33 (7.5%)	66 (49%)
Hospital Discharge GA	871	49 (24)	73 (30)
Tracheostomy	994		
No		811 (100%)	37 (20%)
Yes		0 (0%)	146 (80%)
Death	994		
No		811 (100%)	129 (70%)
Yes		0 (0%)	54 (30%)

[1] n (%); Mean (SD)

Table 1. Summary of Variables by Outcome

In terms of correlations, a strong positive relationship exists between birth-related variables: birth weight (bw), birth length (blength), and head circumference (birth_hc), suggesting that these variables tend to increase together as expected (Figure 2). The decision to retain only birth weight for further analysis is informed by its strong correlation with other size metrics and its clinical relevance as a comprehensive indicator of neonatal health, often used to predict outcomes in infant development and survival. This approach also

mitigates the risk of multicollinearity, which could distort the regression model by overstating the importance of related predictors.

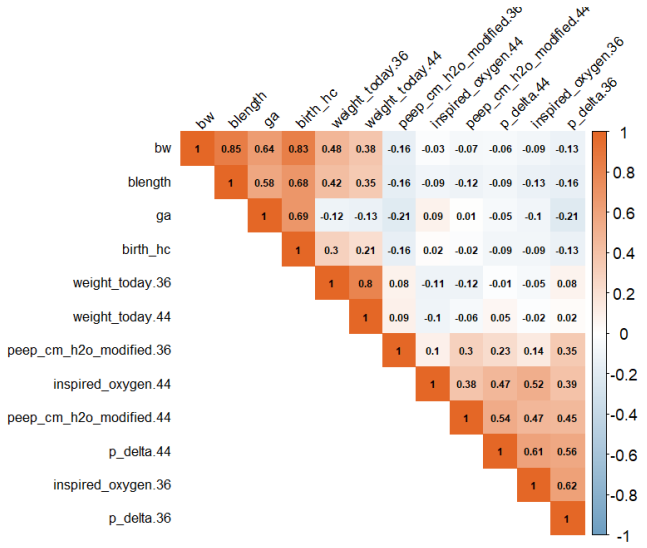

Figure 1. Correlation of continuous variables

Missing data

The presence of missing data in the dataset is an important consideration for the analysis. Notably, a subset of variables, particularly those capturing information at 44 weeks post-menstrual age (PMA), demonstrate a relatively high percentage of missingness, with variables like inspired_oxygen.44 and p_delta.44 each missing in approximately 44.9% of cases. This suggests that data collection at this time point may have been particularly challenging, which could be due to various factors such as the timing of discharge or differential follow-up practices across the NICUs.

The patterns of missing data, as visualized in the Figure 2, reveal that missingness is not isolated to single variables but tends to cluster among certain groups, especially those pertaining to the 44-week PMA measurements. This clustering indicates that when one variable is missing, it is common for others within the same subset to also be missing, which can provide insights into the mechanism behind the missing data.

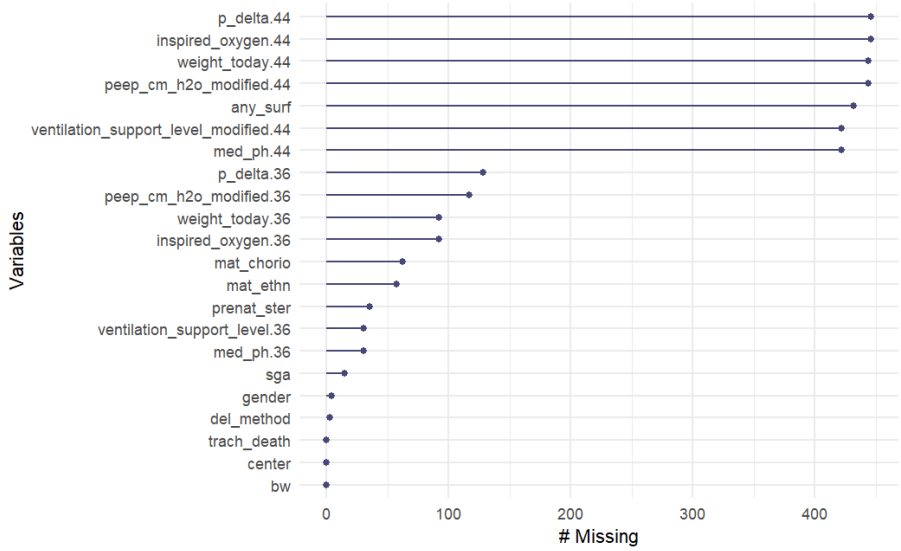

Figure 2. Missing data visualization

Multiple imputation was performed using the mice package in R to address missing data in our analysis. 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 key assumption underlying multiple imputation is that the data are missing at random (MAR), implying that missingness is related to observed data rather than the unobserved data itself. This assumption aligns with the

use of multiple imputation by chained equations (MICE), which is a robust method for handling missing data. It allows us to include all available information in the analysis and reduces potential biases that may be introduced by complete case analysis. MICE imputes missing values based on patterns observed in the rest of the dataset, aiming to provide a more accurate and comprehensive picture of the population under study. In our analysis, missing data seemed connected to the data collection center rather than the variable values, supporting this assumption.

Model Development

The Least Absolute Shrinkage and Selection Operator (LASSO) regression was employed to refine the predictors for the composite outcome of tracheostomy or death among infants with severe bronchopulmonary dysplasia (sBPD). Through this process, certain variables were retained as significant, while others were omitted due to their coefficients being reduced to zero, indicative of their lesser importance or redundancy. Initially, the dataset was divided into training and validation sets. Key predictors identified include maternal ethnicity, delivery method, prenatal steroid use, and specific respiratory support parameters like ventilation support levels and required oxygen levels. Notably, variables such as maternal chorioamnionitis, small for gestational age status, and medication for pulmonary hypertension were significant in some models, indicating their potential impact on outcomes. The variables retained by LASSO align with the significant differences observed in Table 1, thus validating their relevance.

Table 1 presents a bivariate analysis, comparing the characteristics of the infants grouped by the composite outcome of tracheostomy or death (coded as 'No' or 'Yes'). The table is stratified based on this outcome and reports the frequencies and percentages for categorical variables, as well as means and standard deviations for continuous variables. This allows for a clear comparison between the two groups on various demographic and clinical factors.

The results from Table 1 were instrumental in informing the variable selection for the predictive modeling. Variables that demonstrated significant differences between the groups were considered potential predictors. For example, a higher proportion of infants requiring invasive ventilation support at 36 weeks were associated with the 'Yes' outcome, suggesting a critical area for predictive modeling.

Three distinct models were constructed:

1. **All-Inclusive Model:** Incorporated the full range of variables.
2. **36-Week Model:** Focused on data specific to 36-week postnatal age.
3. **44-Week Model:** Concentrated on variables relevant to 44-week postnatal age.

These models were tailored to capture the multifaceted nature of the dataset and the varying temporal aspects of the data.

LASSO models were fitted and validated using training and validation datasets. The analysis included converting categorical variables to numerical using one-hot encoding, fitting the LASSO model on training data, and validating it on a separate dataset. In addition to LASSO, mixed-effects models were fitted for each of the three model types to account for the random effects of different medical centers. These models provided a more nuanced understanding of the fixed and random effects influencing the outcome. In the end, the models' performance was assessed using metrics like sensitivity, specificity, F1 score, and Brier score.

Results

The LASSO regression identified key predictors across three models. In Model 1, predictors like prenatal steroid use (prenat_ster), invasive ventilation support at 36 weeks (ventilation_support_level.36), and inspired oxygen at 36 weeks (inspired_oxygen.36) had significant positive coefficients, indicating a strong association with the outcome. In Model 2, the same predictors were significant alongside a higher weight at 36 weeks being inversely related to the outcome. Model 3 identified similar significant predictors, adjusted for the 44-week timeframe.

The latter analysis focused on fitting Mixed-Effects Models to a dataset to predict the outcome of tracheostomy or death in infants with severe bronchopulmonary dysplasia (sBPD). This methodology acknowledges the complexity of sBPD and the potential influence of different variables at various developmental stages. The function for model developmen integrates both fixed and random effects. The inclusion of the random effect, based on different medical centers (center), is a crucial aspect as it captures the variability in practices or patient populations across these centers.

The summaries of these models in Table 2 provide insights into the significant predictors and the model fit, as indicated by AIC and BIC values. Each model's random effects component underscores the variability due to different medical centers.

Model	AIC/BIC	Random Effects Std.Dev.	Significant Predictors
Model 1 (All Variables)	544.2/607.9	0.9278	Prenatal steroid use, invasive pressure ventilation, inspired oxygen at 36 weeks, medication for PH at 44 weeks
Model 2 (36 Week Variables)	662.8/711.9	1.042	Prenatal steroid use, invasive pressure ventilation, inspired oxygen at 36 weeks, SGA status
Model 3 (44 Week Variables)	575.9/639.6	1.243	Prenatal steroid use, invasive pressure ventilation, inspired oxygen at 44 weeks, SGA status, surfactant requirement

Table 2. Summary of Models

Model Performance Metrics

The performance metrics for the three models in Table 3 highlight different strengths and weaknesses in each model. Model 1 and Model 3 exhibit high specificity (around 95.70%), indicating they are proficient in correctly identifying negative cases but have low sensitivity (30.48%), showing a weakness in detecting positive cases. Their F1 scores, around 45.39%, suggest a moderate balance between precision and recall, but this is not optimal. The Brier scores for these models (0.3287 for Model 1 and 0.3530 for Model 3) indicate moderate accuracy in probabilistic predictions.

On the other hand, Model 2 demonstrates a more balanced and effective performance. It achieves both high sensitivity and specificity (88.89%), indicating it is equally good at identifying both positive and negative cases. Its F1 score of 59.26% is significantly higher than the other two models, reflecting a better balance between precision and recall. Most notably, Model 2 has a much lower Brier score (0.0881), suggesting it has superior accuracy in its probabilistic predictions. This makes Model 2 the most reliable and balanced choice among the three, particularly in scenarios where correctly identifying both positive and negative cases is crucial.

Model	Sensitivity	Specificity	F1_Score	Brier_Score
Model 1 (All Variables)	0.3047619	0.9569892	0.4539007	0.3286649
Model 2 (36 Week Variables)	0.8888889	0.8888889	0.5925926	0.0880697
Model 3 (44 Week Variables)	0.3047619	0.9569892	0.4539007	0.3530143

Table 3. Performance Metrics for Three Models

Receiver Operating Characteristic (ROC) Curves

The ROC curves depicted in the image represent the trade-off between sensitivity (true positive rate) and 1-specificity (false positive rate) for three different models. A ROC curve plots the full spectrum of trade-offs between the true positive rate and the false positive rate as the classification threshold is varied. Model 2 (Red Curve) exhibits superior performance, as its curve is closest to the top left corner, indicating a higher true positive rate and a lower false positive rate. This is desirable in a good classifier and suggests that Model 2 effectively distinguishes between the classes over a range of thresholds.

Model 1 (Blue Curve) and Model 3 (Green Curve), on the other hand, show similar performance to each other as their curves almost overlap. Both are above the diagonal line of no discrimination, indicating they perform better than random guessing. However, they are outperformed by Model 2, as indicated by their lower positioning which implies lower sensitivity at the same level of specificity. The ROC curves suggest that Model

2 is the most reliable classifier among the three, providing a good balance between correctly predicting positive and negative classes across different thresholds.

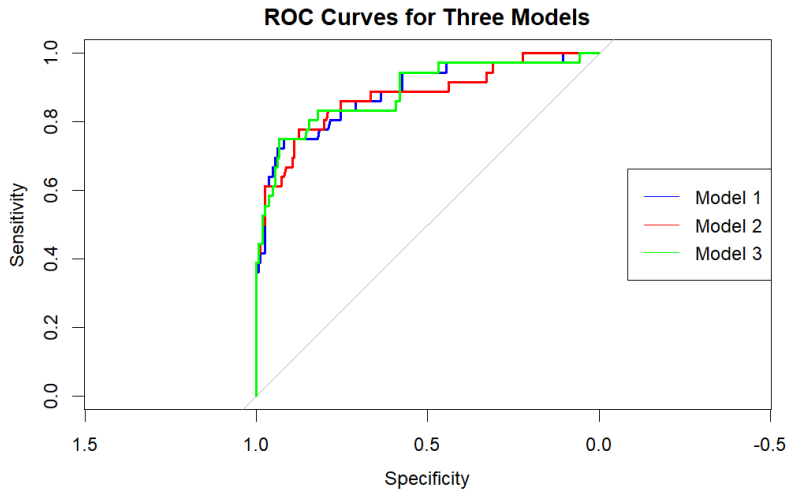

Figure 3. ROC Curves for Three Models

Limitations

While the models demonstrate good predictive performance, there are limitations to consider. First, the data represent a national cohort from multiple centers, which introduces variability that may affect the generalizability of the findings.

Another significant limitation is the risk of overfitting, a common challenge in regression models, especially when dealing with a large number of variables. The study employed LASSO regression for variable selection, which is known for its effectiveness in reducing the risk of overfitting. However, the potential for this issue remains and warrants caution in the interpretation and application of the model.

The study also acknowledges the challenge of missing data, which was addressed through multiple imputation methods. While this approach strengthens the analysis, the nature of imputed data means that the results must be interpreted with caution.

A key concern is the model's convergence issues, as highlighted during the analyses process by warnings indicating failure to converge within the specified tolerance levels. These problems suggest potential issues with the model's complexity, the adequacy of the data, or the choice of algorithm parameters. Such convergence difficulties question the reliability of the model's results, emphasizing the need for further investigation into model structure, data quality, and optimization techniques to enhance the robustness of future predictive models in this area.