Predicting Tracheostomy or Death in Infants with Severe Bronchopulmonary Dysplasia (BPD)

Project 2: Regression Analysis (Github Link)

Zihan Zhou

2023-12-15

Abstract

Aim: This study aims to develop a model predicting the necessity and timing of tracheostomy in infants with severe Bronchopulmonary Dysplasia (BPD), using a rich dataset from the BPD Collaborative Registry. It examines demographic and clinical variables to identify significant predictors of tracheostomy or death.

Method: Best Subset and Lasso regression were used to select variables for a logistic mixed-effects model, treating center as a random effect. Models were developed for two time points: 36 weeks and 44 weeks. The Brier score and AUC were employed for model assessment.

Result: Both best subset and lasso regression models, using a mixed-effects approach, effectively identified factors influencing tracheotomy/death in infants with severe BPD. The Best Subset model at 36 weeks had a Brier Score of 0.0835 and an AUC of 0.8920, while the Lasso model at 44 weeks showed a Brier Score of 0.0923 and an AUC of 0.8853. However, calibration plots indicated some deviation in the 44-week models.

Conclusions: The project reveals key factors impacting tracheotomy timing in severe BPD infants, highlighting birth metrics, respiratory support, and pulmonary hypertension medication. However, hese models need to be improved considering logistic regression assumptions such as linearity, outlier influence, and multicollinearity.

1. Introduction

Bronchopulmonary dysplasia (BPD) is the most common and one of the prognostically significant consequence of premature birth, affecting between 10,000 and 15,000 infants in the United States each year [1]. Despite advances in understanding its pathophysiology and the development of management strategies aimed at reducing its occurrence, the incidence of BPD has remained unchanged over the years [2]. In cases of severe BPD, particularly Grade 3 (defined by 2001 NHLBI criteria), which requires ventilator dependence at 36 weeks of corrected gestational age, about 75% of these infants still need ventilator support upon hospital discharge, although not permanently. Tracheostomy, which is necessary for 2-4% of infants with BPD and up to 12% of those with severe or Grade 3 BPD, has several benefits such as stable airway maintenance, improved growth, and enhanced age-appropriate interactions. Early tracheostomy within four months, is associated with better outcomes, with infants receiving this procedure earlier (before 120 days) showing lower risks of death or neurodevelopmental impairment at 18-22 months [3]. However, there is still no consensus on the optimal timing for tracheostomy. Previous studies utilizing large databases have successfully predicted the likelihood of tracheostomy or death using basic demographics and clinical diagnosis. However, these studies lacked detailed respiratory data and did not offer predictions at varying postmenstrual ages (PMA).

This project aims to create a regression model to predict the composite outcome of tracheostomy or death, thus guiding to determine the criteria and optimal timing for tracheostomy. Several models will be discussed and compared, including mixed-effect model, best subset selection, and lasso regression and ridge regression.

2. Data

2.1. Data sources

The data comes from the BPD Collaborative Registry, a multi-center consortium of interdisciplinary BPD programs located in the United States and Sweden formed to address gaps in evidence and promote research to enhance the care of children with severe forms of BPD. It specifically records data on infants born before 32 weeks of gestation and diagnosed with severe bronchopulmonary dysplasia (sBPD), as defined by the 2001 NHLBI criteria. This includes infants requiring a fractional inspired oxygen (FiO2) of more than 0.3 or any form of positive pressure ventilation at 36 weeks postmenstrual age (PMA). The registry gathers standard demographic and clinical information at key intervals: at birth, and at 36, 44 weeks PMA, and at discharge.

The data is pre-processed by converting the variables into the appropriate types as specified in the codebook. Out of 28 variables, 24 have missing values. Missing values in the variable of `center` are imputed based on the patients' record IDs. Center 21 only has one observation, so I impute it as center 1. Three duplicate records have also been removed, so the total 999 observations become 996 observations. Since `mat_race` is recorded differently between the data and the codebook, this variable will be excluded from future analyses.

Table 1 presents the challenge of missing data across various variables, with the extent of missingness showing significant variation. Notably, data collected at the 44-week corrected gestational age exhibit the most substantial missingness, which is most likely due to discharge. Additionally, the variable `any_surf`, indicating surfactant administration within the first 72 hours, displays a similar pattern of missingness to the 44-week variables. A significant number of missing entries are also found in the 36-week data. To address these concerns, multiple imputation will be implemented to enable a more robust and complete analysis.

Table 1: Summary of Missing Values

Variable	Number	Pct	Variable	Number	Pct
inspired_oxygen.44	448	44.98%	blength	78	7.83%
p_delta.44	448	44.98%	birth_hc	77	7.73%
weight_today.44	446	44.78%	com_prenat_ster	71	7.13%
peep_cm_h2o_modified.44	446	44.78%	mat_chorio	62	6.22%
any_surf	433	43.47%	mat_ethn	57	5.72%
ventilation_support_level_modified.44	424	42.57%	prenat_ster	35	3.51%
med_ph.44	424	42.57%	ventilation_support_level.36	30	3.01%
p_delta.36	128	12.85%	med_ph.36	30	3.01%
hosp_dc_ga	124	12.45%	sga	15	1.51%
peep_cm_h2o_modified.36	117	11.75%	gender	4	0.4%
weight_today.36	92	9.24%	del_method	3	0.3%
inspired_oxygen.36	92	9.24%	Death	2	0.2%

2.2. Demographics and Clinical Diagnosis

Table 2 highlights the variability in tracheostomy rates and mortality rates among the centers. The tracheostomy rates has a range from 0% at Center 20 to 50.72% at Center 12, underscoring the significant disparities in practice. The higher rates of 50.72% and 41.54% in center 12 and 1 may suggest more aggressive treatment approaches or patient profiles requiring such interventions. In contrast, Centers 3, 7, and 16 have rates below 4%, which could reflect variations in demographics, protocols, or resources. The mortality rates ranges from 0% at Center 7, 16 and 20 to 20.29% at Center 12. The incidence of death is more rare compared to tracheostomy.

The pronounced differences across centers are statistically confirmed by a chi-squared test ($p < 0.001$). Importantly, Center 2 accounts for over 60% of the patient cohort, highlighting a multilevel data structure

and underscoring the importance of appropriately accounting for the `center` variable in analysis.

Table 2: Tracheostomy and death Proportions across the Centers

Center	Total	Tracheostomy Proportion	Death Proportion
1	66	42.42%	10.61%
2	630	10.16%	4.60%
3	57	1.75%	1.75%
4	60	18.33%	1.67%
5	40	12.50%	5.00%
7	32	3.12%	0.00%
12	69	50.72%	20.29%
16	38	2.63%	0.00%
20	4	0.00%	0.00%

The project's aim is to construct a regression model that predicts the combined outcome of tracheostomy or death. Consequently, a new composite variable, Y is introduced. In this binary outcome, patients are labeled '1' if they received a tracheostomy or succumbed to their condition, and '0' if they survived without undergoing a tracheostomy. This variable allows for a comprehensive analysis of both tracheostomy and patient survival.

The demographic and clinical diagnosis data presented in Table 4 compare infants across this new composite outcome, revealing significant differences in several metrics in both birth metrics and follow-up assessments.

There was a notable reduction in birth weight compared to those without tracheostomy or death, with a median weight of 670 grams versus 760 grams, respectively, and this difference was statistically significant with a p-value smaller than 0.001. Additionally, 34% of the infants with tracheostomy or death were small for gestational age (SGA), a higher percentage than the 18% for infants with neither trac nor death ($p<0.001$). Delivery method also differed significantly, with more infants in the trach/death group born via cesarean section (79% vs. 70%, $p=0.017$). No significant differences were observed in maternal ethnicity, chorioamnionitis, or gender.

In terms of the follow-up assessments, all of the metrics are significantly different accorss the two groups. A significant contrast was evident at 36 weeks of corrected gestational age, with 73% of the trach/death group requiring invasive positive pressure ventilation, compared to 17% not needing such support ($p<0.001$) in the no trach/death group. By 44 weeks, the need for invasive positive pressure increased to 77% in the tracheostomy group versus just 15% ($p<0.001$). The no trach/death group's non-invasive positive pressure decrease from 69% at 36 weeks to 28% at 44 weeks, while the trach/death group only dropped 6% from 22% to 16%. Additionally, infants with trach/death have higher median peak inspiratory pressures at both 36 and 44 weeks, and a greater proportion required medication for pulmonary hypertension at 44 weeks. The trach/death group also had significantly higher discharge ages.

3. Methods

3.1. Multiple Imputation

Multiple imputation will be used to address missing data. This method creates several imputed datasets, introducing variability in the imputed values to reflect the uncertainty around the true values. The imputation phase involves applying algorithms repetitively to generate values for the missing data. Subsequently, in the analysis phase, each dataset undergoes standard statistical analysis as though it were complete. The final phase, pooling, aggregates the results from all datasets, yielding parameter estimates that incorporate

Table 3: Demographics and Clinical Diagnosis in Infants

Characteristic	**No**, N = 811	**Yes**, N = 183	**p-value**
Maternal Ethnicity			0.3
Hispanic or Latino	64 (8.3%)	10 (5.9%)	
Not Hispanic or Latino	703 (92%)	160 (94%)	
Birth Weight (g)	760 (610, 950)	670 (540, 835)	<0.001
Obstetrical Gestational Age	25 (24, 27)	25 (24, 27)	0.6
Birth Length (cm)	32 (30, 35)	31 (29, 34)	0.002
Birth Head Hircumference (cm)	23.00 (21.50, 25.00)	22.00 (21.00, 24.00)	0.010
Delivery Method			0.017
Vaginal delivery	245 (30%)	39 (21%)	
Cesarean section	564 (70%)	143 (79%)	
Prenatal Corticosteroids			0.024
0	113 (14%)	13 (7.8%)	
1	679 (86%)	154 (92%)	
Complete Prenatal Steroids			0.2
0	269 (35%)	46 (30%)	
1	499 (65%)	109 (70%)	
Maternal Chorioamnionitis			0.9
0	633 (83%)	139 (83%)	
1	132 (17%)	28 (17%)	
Gender			0.7
Female	334 (41%)	73 (40%)	
Male	473 (59%)	110 (60%)	
Small for Gestational Age			<0.001
Not SGA	658 (82%)	118 (66%)	
SGA	142 (18%)	61 (34%)	
Received Surfactant			0.095
0	89 (19%)	12 (12%)	
1	374 (81%)	87 (88%)	
Weight at 36 Weeks	2,150 (1,880, 2,408)	1,997 (1,694, 2,260)	<0.001
Ventilation Support at 36 Weeks			<0.001
No respiratory support or supplemental oxygen	109 (14%)	7 (4.3%)	
Non-invasive positive pressure	553 (69%)	36 (22%)	
Invasive positive pressure	140 (17%)	119 (73%)	
Fraction of Inspired Oxygen at 36 Weeks	0.29 (0.23, 0.35)	0.45 (0.34, 0.60)	<0.001
Peak Inspiratory Pressure (cm H2O) at 36 Weeks	0 (0, 0)	14 (2, 24)	<0.001
Positive and Exploratory Pressure (cm H2O) at 36 Weeks	7 (6, 8)	8 (6, 9)	<0.001
Medication for Pulmonary Hypertension at 36 Weeks			<0.001
0	770 (96%)	129 (80%)	
1	32 (4.0%)	33 (20%)	
Weight at 44 Weeks	3,765 (3,299, 4,143)	3,555 (3,070, 3,995)	0.009
Ventilation Support at 44 Weeks			<0.001
No respiratory support or supplemental oxygen	261 (60%)	8 (6.0%)	
Non-invasive positive pressure	124 (28%)	22 (16%)	
Invasive positive pressure	53 (12%)	104 (78%)	
Fraction of Inspired Oxygen at 44 Weeks	0.28 (0.25, 0.32)	0.40 (0.30, 0.60)	<0.001
Peak Inspiratory Pressure (cm H2O) at 44 Weeks	0 (0, 0)	19 (10, 37)	<0.001
Positive and Exploratory Pressure (cm H2O) at 44 Weeks	0 (0, 8)	8 (8, 10)	<0.001
Medication for Pulmonary Hypertension at 44 Weeks			<0.001
0	405 (92%)	68 (51%)	
1	33 (7.5%)	66 (49%)	
Hospital Discharge Gestational Age	45 (42, 52)	64 (50, 91)	<0.001

[1] n (%); Median (IQR)

[2] Pearson's Chi-squared test; Wilcoxon rank sum test

the variability across the imputations. In this project, models are fitted to 5 complete datasets, and the resulting estimates are then pooled, thus derivating final models.

3.2. Models

In this project, we compare two models: best subset selection and lasso regression, using a mixed-effects model. For the best subset model, the fixed effects are initially selected through best subset selection. Lasso regression employs **glmmLasso** [4] in R. These methods are classical approaches to variable selection, contributing to the creation of more efficient prediction models. The mixed-effects model incorporates fixed effects of the observed covariates and random effects associated with the intercepts for each center, thereby capturing the variability between centers. This mixed-effects approach is particularly well-suited for data characterized by random variability across different groups—in this project, medical centers. The birth weight is log-transformed to fit the scale.

Data are collected at four distinct time points: at birth, at 36 and 44 weeks postmenstrual age (PMA), and at discharge. The objective is to predict the composite outcome of tracheostomy or death and to determine the optimal timing for tracheostomy. To achieve this, we will develop models for two critical phases: 36 weeks and 44 weeks PMA. The primary outcome to be modeled is a binary variable — either the occurrence of tracheostomy or death versus neither event, so the logistic regression will be used:

$$log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n.$$

The models will be subjected to a train-test split method, where they will be trained on 70% of the dataset and validated on the remaining 30%.

3.2.1 36 Weeks PMA Model

In constructing the model specific to the 36 weeks PMA time point, I excluded all variables that were assessed at 44 weeks. The remaining variables were then using best subset and lasso regression techniques, incorporating center as a random effect in the analysis. The new dataset has 20 covariates.

3.2.2 44 Weeks PMA Model

For the 44 weeks PMA model, a key consideration was the fact that many infants are typically discharged before reaching 44 weeks. Therefore, this model specifically focused on infants who remained hospitalized at this time point. It included all relevant variables assessed at birth, at 36 weeks, and at 44 weeks. Similar to the 36 weeks PMA model, variable selection was performed using best subset and lasso regression, again considering center as a random effect. The new dataset left with 572 observations and 26 covariates.

4. Results

4.1. Model Coefficients

Table 5 presents the selected variables and their corresponding coefficients, as identified by the best subset and lasso regression methods within mixed-effects models, across five imputed datasets. These models account for center as a random effect. Notably, the best subset method tends to select fewer variables compared to the lasso regression models.

In terms of birth statistics, variables including ethnicity, prenatal steroids, and maternal chorioamnionitis are consistently selected across all models. Ethnicity and Prenatal Steroids are associated with positive

coefficients in these models. However, the influence of Maternal Chorioamnionitis varies: it shows positive coefficients in the 36-week models and negative coefficients in the 44-week models.

For 36-week models, there are several common variables between the best subset and lasso models, including ventilation support level, fraction of inspired oxygen, and peak inspiratory pressure (measured in cmH2O) assessed at 36 weeks.

The 44-week models select a broader range of variables. These include birth statistics like gender, delivery method, and prenatal corticosteroids. Additionally, several parameters measured both at 36 and 44 weeks, such as weight, ventilation support level, fraction of inspired oxygen, and medication for pulmonary hypertension, are commonly included in these models.

Table 4: Summary for coefficients

	Best Subset 36 Week	Best Subset 44 Week	Lasso 36 Week	Lasso 44 Week
(Intercept)	-2.811	13.0624	12.1034	32.7004
mat_ethn2	0.8239	0.3867	1.0547	0.2217
log(bw)	-1.0207		0.255	-3.3080
ga	0.1709		-0.0539	-0.1854
com_prenat_ster1	0.7135	0.6926	0.6145	1.1212
mat_chorio1	-0.4528	0.5013	-0.6411	0.6447
ventilation_support_level.361	-0.161	-0.9097	-0.6368	0.1907
ventilation_support_level.362	0.8976	-0.006	0.7673	1.6840
inspired_oxygen.36	4.725	0.8881	4.1071	0.9484
p_delta.36	0.0146		0.0021	-0.0139
del_method2		0.9302	-0.3204	0.9039
prenat_ster1		0.6302	0.6641	-0.3285
genderMale		0.1805	0.0801	0.1248
log(weight_today.36)		-2.4916	-2.6299	-3.3723
med_ph.361		-0.7453	1.0774	-0.9618
log(weight_today.44)		-0.0953		-0.2642
ventilation_support_level_modified.441		1.4666		-1.6485
ventilation_support_level_modified.442		3.845		1.0106
inspired_oxygen.44		0.5989		0.1702
med_ph.441		0.7804		0.7758
blength			0.0628	0.2120
birth_hc			0	0.3784
sgaSGA			0.1968	-0.1242
any_surf1			0.0545	0.0000
peep_cm_h2o_modified.36			0.0424	-0.1460
p_delta.44				-0.0097
peep_cm_h2o_modified.44				0.4055

4.2. Model Comparsion and Evaluation

Since the outcome of the model is binary, Brier score and area under the curve (AUC) are used to evaluate the models on the test datasets, which represent 30% of the data excluding the training set. AUC can be used to compare their discrimination ability, which is the ability of the model to differentiate between people with and without the outcome. Table 4 displays the summary statistics. The Best Subset model at 36 weeks demonstrated a Brier Score of 0.0835, an AUC of 0.8920, a decision threshold at 0.1600, along with a specificity of 0.8215 and a sensitivity of 0.8638. In comparison, the Best Subset model at 44 weeks showed a slightly higher Brier Score of 0.0963 and a higher AUC of 0.8961, with a notably higher threshold of 0.3015, a greater specificity of 0.9209, but a lower sensitivity of 0.7722. The Lasso model at 36 weeks exhibited a Brier Score of 0.0881, an AUC of 0.8736, a threshold of 0.1621, with specificity and sensitivity closely mirroring those of the Best Subset 36 Weeks model. Lastly, the Lasso model for 44 weeks registered a Brier Score of 0.0923, an AUC of 0.8853, the lowest threshold at 0.1281, a specificity identical to the 36 Weeks Lasso model, and the highest sensitivity of 0.8833 among all models.

In order to further compare these models in a intuitive way, the calibration and the receiver operating characteristic (ROC) curve are plotted in Figure 1. Calibration assesses the agreement between observed event frequencies and predicted probabilities. It is visualized through a calibration plot, which contrasts the predicted probabilities against observed event rates, illustrating how well the predictions match actual outcomes. It can be seen that the Lasso model at 36 weeks PMA seems to be closest to the 45-degree line, while both models at 44 weeks PMA exhibit some deviations, suggesting less calibration ability.

Table 5: Metircs for the models

	Best Subset 36 Weeks	Best Subset 44 Weeks	Lasso 36 Weeks	Lasso 44 Weeks
Brier Score	0.0835	0.0963	0.0881	0.0923
AUC	0.8920	0.8961	0.8736	0.8853
Threshold	0.1600	0.3015	0.1621	0.1281
Specificity	0.8215	0.9209	0.8200	0.8216
Sensitivity	0.8638	0.7722	0.8213	0.8833

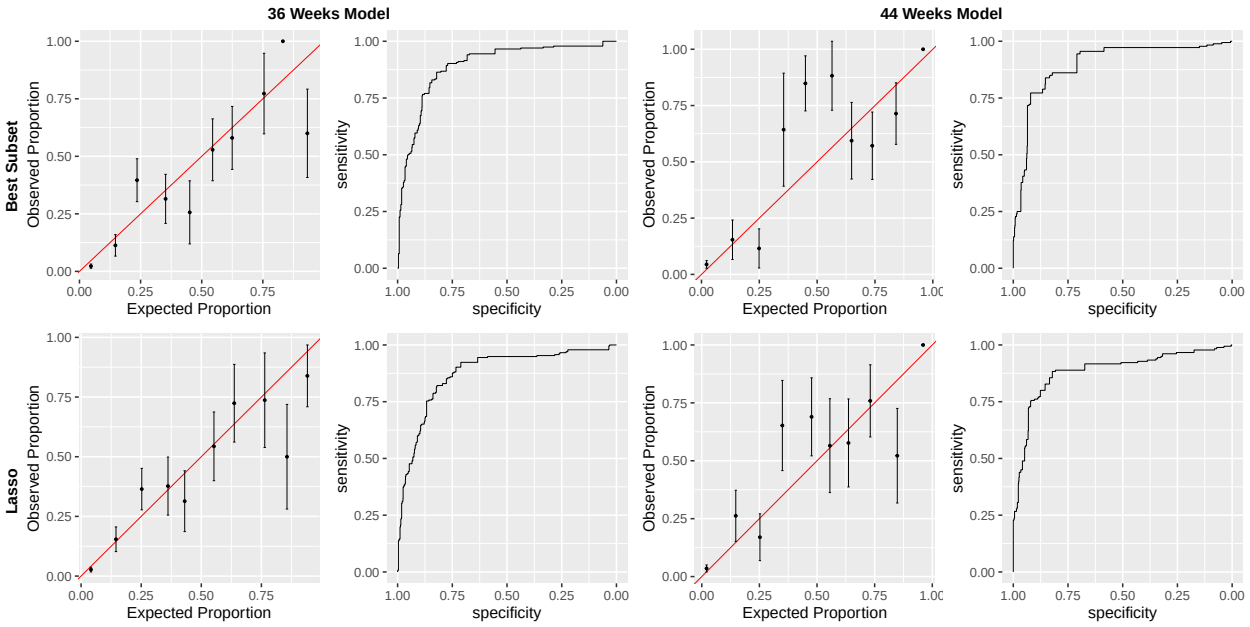

Figure 1: Calibration Plots and ROC Curves for the Models

5. Discussion

The results show that the prediction models using center as a random effect to predict the composite outcome of tracheotomy/death on premature infants with BPD perform well. Both the best subset and lasso regression models exhibit high accuracy at the two critical time points of 36 and 44 weeks. In the exploratory analysis highlights significant variations in demographic statistics and treatment methodologies across different centers, with some adopting more aggressive strategies or catering to patient profiles requiring such interventions.

The commmon variables used across these models, such as prenatal steroids, and maternal chorioamnionitis at birth assessment, ventilation support level, fraction of inspired oxygen, and peak inspiratory pressure (measured in cmH2O) in the follow-up assessment might be vital for doctors to determine the whether to perform tracheotomy.

Moreover, to compare the models across the time point, it can be seen that more variables should be considered at 44 weeks than at the 36 weeks. This variance at different time points suggests that these different variables like medication for pulmonary hypertension are critical for determining the tracheostomy's timing. The variation in coefficients for common variables between the 36-week and 44-week models also indicates changing circumstances of the infants over time.

However, the models in this project are limited by certain assumptions required for logistic regression: the assumption of linearity, the influence of outliers, and the presence of multicollinearity. The four models do not include interaction terms to assess the linearity assumption. The residual plots seem to indicate the presence of some outliers, suggesting potential influential values. Furthermore, the ventilation support variables at 36 and 44 weeks have very high VIF values, indicating potential multicollinearity.

6. Conclusion

The exploratory data analysis and regression models used in this project offer a comprehensive understanding of the factors influencing the necessity and timing of tracheostomy in infants with severe BPD. The models highlight the significance of birth metrics, respiratory support requirements, and medication for pulmonary hypertension as pivotal indicators. However, it is crucial to recognize the limitations presented by logistic regression assumptions, which include the need for linearity, the absence of influential outliers, and multicollinearity, areas where the models require further improvement.

References

[1] E. A. Jensen and B. Schmidt, "Epidemiology of bronchopulmonary dysplasia," *Birth Defects Research Part A: Clinical and Molecular Teratology*, vol. 100, no. 3, pp. 145–157, Mar. 2014, doi: gs4fjw.

[2] B. J. Stoll *et al.*, "Trends in Care Practices, Morbidity, and Mortality of Extremely Preterm Neonates, 1993-2012," *JAMA*, vol. 314, no. 10, p. 1039, Sep. 2015, doi: ghw5jf.

[3] S. B. DeMauro *et al.*, "Developmental Outcomes of Very Preterm Infants with Tracheostomies," *The Journal of Pediatrics*, vol. 164, no. 6, pp. 1303–1310.e2, Jun. 2014, doi: f2rxwj.

[4] A. Groll, "glmmLasso: Variable Selection for Generalized Linear Mixed Models by L1-Penalized Estimation." Aug. 2023.