PHP 2550: Project 2

Due: November 12th by 11:59pm

Abstract

This project, in collaboration with Dr. Chris Schmid from the Biostatistics Department at Brown, targets a vital area in neonatal healthcare: determining the optimal timing for tracheostomy in neonates with severe bronchopulmonary dysplasia (sBPD). Tracheostomy, essential for aiding breathing in these infants, has lacked clear, data-driven guidelines for its timing. Utilizing a comprehensive national dataset, we analyzed demographic, diagnostic, and respiratory parameters of infants with sBPD at two critical post-menstrual ages – 36 and 44 weeks.

Our analysis revealed that models incorporating 44-week data were superior in predicting the need for tracheostomy or the risk of death compared to those using only up to 36-week data. The best-performing models, using Lasso and Forward Selection at 44 weeks, achieved an accuracy of 86%. These model's success is attributed to key predictors such as birth length, delivery method, prenatal steroid use, ventilation support level, and inspired oxygen fraction, underscoring the intricate relationship between prenatal and postnatal health in sBPD outcomes.

This work lays a critical data-driven foundation for future model development, aiming to enhance clinical decision-making and improve outcomes for vulnerable neonates. It marks a first step in addressing the challenge of tracheostomy timing in sBPD, providing a springboard for further investigation and model refinement in neonatal care.

1. Introduction

1.1 Background on Tracheostomy in Severe BPD

Severe bronchopulmonary dysplasia (sBPD) in neonates presents a complex medical challenge, often requiring advanced respiratory support. In certain cases, this support includes tracheostomy, a surgical intervention to establish a direct airway through an incision in the windpipe. The decision to perform a tracheostomy in these fragile patients is complex, balancing immediate respiratory needs with long-term developmental considerations. The timing of this intervention is crucial; too early, and it may expose the infant to unnecessary surgical risks; too late, and it might delay essential respiratory support.

1.2 Dataset Overview

The dataset for this project is a dense compilation of data from neonatal intensive care units (NICUs) across the US, encompassing a wide range of variables. These include demographic information such as maternal race (mat_race) and ethnicity (`mat_ethn`), medical details like birth weight (`bw`), gestational age (`ga`), and various measures of neonatal health and treatment at 36 and 44 weeks post-menstrual age (PMA). These measures include weight at specific weeks (`weight_today.36`, `weight_today.44`), levels of ventilation support (`ventilation_support_level.36`, `ventilation_support_level_modified.44`), and other critical respiratory parameters (e.g., `inspired_oxygen.36`, `p_delta.36`).

1.3 Project Overview

The project began with thorough data cleaning, establishing a solid foundation for subsequent analysis. This was followed by an exploratory data analysis (EDA) phase, where we examined patterns and correlations between various predictors and the key outcome—tracheostomy or death in neonates with severe bronchopulmonary dysplasia (sBPD).

To tackle the issue of missing data, we implemented multiple imputation and removed cases where large amounts of 44-week variables were missing. Our analytical approach was divided into settings, focusing on the data until the post-menstrual ages of 36 and 44 weeks, to account for the progression of neonatal health. For each setting, we developed models using Ridge, Lasso, and Forward Stepwise Logistic Regression, fitting and evaluating these methods across each of the five imputed datasets. The models were then averaged to ensure generalizability and robustness. Finally, we assessed the model evaluation metrics we used such as accuracy, F1-score, specificity, sensitivity, AUC, and Brier score. This varied assessment provided crucial insights for clinical decision-making and underscored the significance of data-driven analysis in neonatal healthcare.

2. Methods

2.1 Data Preprocessing

For data preprocessing, data was meticulously cleaned, structured, and prepared to ensure its suitability for regression analysis. This involved transforming various variables into binary or multilevel factor formats for analytical consistency. For instance, the `Death` variable was recoded to '1' for 'Yes' and '0' for 'No'. Other binary predictors, such as prenatal steroid use (`prenat_ster`), complete prenatal steroids (`com_prenat_ster`), maternal chorioamnionitis (`mat_chorio`), small for gestational age status (`sga`), surfactant usage (any_surf), delivery method (del_method), and gender, were similarly encoded. Ventilation support levels at 36 and 44 weeks post-menstrual age (ventilation_support_level.36 and ventilation_support_level_modified.44) were categorized as factors.

To address our research questions effectively, we created a composite outcome variable to represent the combined incidence of tracheostomy or death. This approach was necessitated by the limited cases exclusively involving death ('Death Only') or both tracheostomy and death ('Trach and Death'). The outcome was defined as '0' for scenarios involving neither death nor tracheostomy and '1' for either outcome.

Data cleaning also involved rectifying anomalies and removing non-essential information. For example, an implausibly low recorded weight of 3 grams at 44 weeks was marked as missing. We discovered and removed duplicate records. Missing center values were filled based on the record_id, but due to limited data from some centers and to enhance model generalizability, the center variable was ultimately removed. Hospital Discharge Gestational Age (hosp_dc_ga) was also excluded as it would not be known in the practical use of these models. Lastly, irrelevant variables such as record_id, Death, Trach, and mat_race (due to data collection issues) were also omitted from the dataset for analysis.

2.2 Exploratory Data Analysis

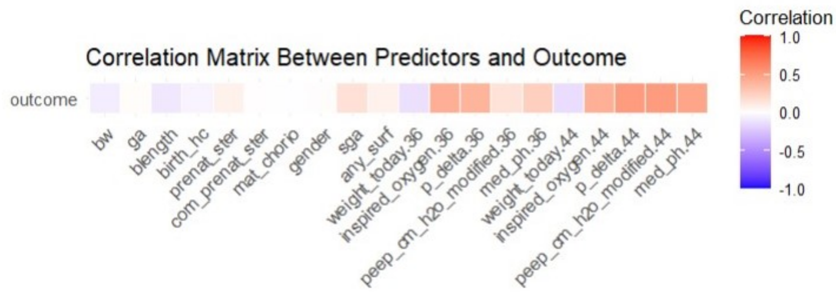
Characteristic	**0**, N = 811[1]	**1**, N = 183[1]
Maternal Ethnicity		
Hispanic	703 (92%)	160 (94%)
Not Hispanic	64 (8.3%)	10 (5.9%)
Birth weight (g)	817 (285)	756 (340)
Obstetrical gestational age	26 (2)	26 (2)
Birth length (cm)	33 (4)	32 (4)
Birth head circumference (cm)	23.25 (2.65)	22.88 (3.29)
Delivery Method		
C-section	564 (70%)	143 (79%)
Vaginal	245 (30%)	39 (21%)
Prenatal Corticosteroids	679 (86%)	154 (92%)
Complete Prenatal Steroids	499 (76%)	109 (76%)
Maternal Chorioamnionitis	132 (17%)	28 (17%)
Male	473 (59%)	110 (60%)
SGA	142 (18%)	61 (34%)
Surfactant	374 (81%)	87 (88%)
Weight at 36 wks	2,142 (393)	1,981 (507)

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

Characteristic	**0**, N = 811[1]	**1**, N = 183[1]
Ventilation at 36 wks		
Invasive	140 (17%)	119 (73%)
Non-Invasive	553 (69%)	36 (22%)
None	109 (14%)	7 (4.3%)
Frac of Inspired Oxygen at 36 wks	0.31 (0.12)	0.49 (0.21)
Peak Inspiratory Pressure at 36 wks	4 (8)	16 (12)
PEEP at 36 wks	6 (3)	7 (3)
Medication for PH at 36 weeks	32 (4.0%)	33 (20%)
Weight at 44 wks	3,704 (619)	3,473 (782)
Ventilation at 44 wks		
Invasive	53 (12%)	104 (78%)
Non-Invasive	124 (28%)	22 (16%)
None	261 (60%)	8 (6.0%)
Frac of Inspired Oxygen at 44 wks	0.31 (0.11)	0.45 (0.20)
Peak Inspiratory Pressure at 44 wks	4 (11)	22 (16)
PEEP at 44 wks	3 (4)	9 (3)
Medication for PH at 44 wks	33 (7.5%)	66 (49%)

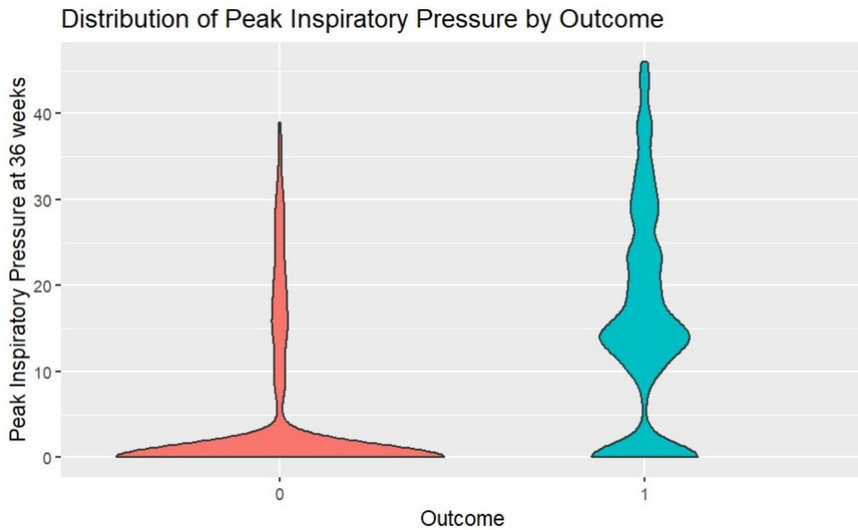
[1] n (%); Mean (SD)

Our exploratory data analysis began with an examination of descriptive statistics of predictors between our different outcomes. As seen in Figure 1, we compared key variables between neonates who did not require tracheostomy or did not experience death (Outcome 0) and those who did (Outcome 1). This comparison encompassed a range of predictors in the dataset. Notable distinctions were observed in infant weight, small for gestational age (SGA) status, and respiratory support metrics such as Peak Inspiratory Pressure and Ventilation at both 36 and 44 weeks. These differences provided preliminary insights into the factors that might influence the likelihood of tracheostomy or death.

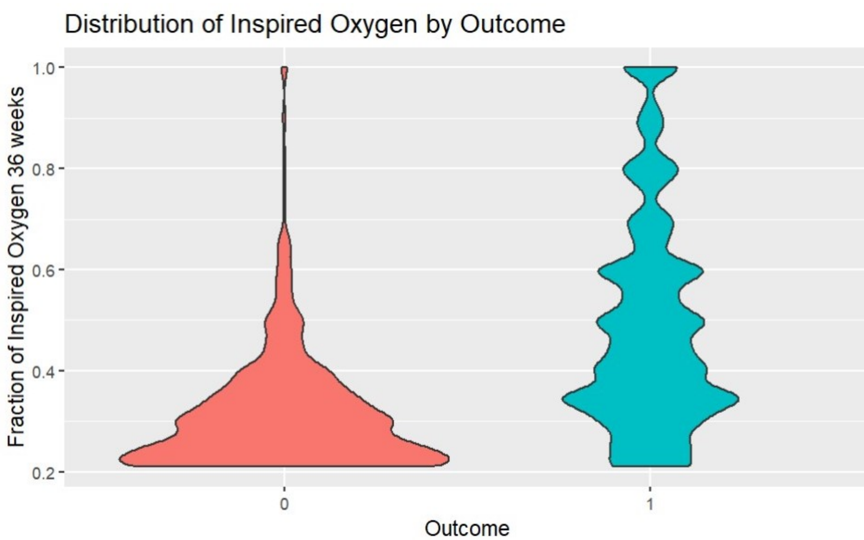


In our correlation analysis, shown in the plot above, we assessed the relationship between these predictors and the composite outcome. The results indicated that variables measured at 36 weeks had a significant correlation with the outcome, and this relationship was even more pronounced with the 44-week variables. This finding suggests the increasing importance of respiratory parameters as neonates progress in age.

To visually illustrate these differences in predictors between the outcomes, we generated two plots focusing on key respiratory measures.



This plot displays the distribution of peak inspiratory pressure at 36 weeks post-menstrual age in relation to the outcome. For infants classified under Outcome 0, the distribution is primarily concentrated at lower pressure values, with a few outliers at higher pressure levels. However, in the Outcome 1 group, the distribution is more evenly spread, with a noticeable presence of medium to high values. This pattern indicates a higher likelihood of adverse outcomes (tracheostomy or death) associated with increased inspiratory pressures at this age.



The second plot examines the fraction of inspired oxygen at 36 weeks. Similar to the findings on inspiratory pressure, infants in the Outcome 0 group predominantly fall into the lower fraction range, as evidenced by a left-skewed distribution. In contrast, the distribution for infants with Outcome 1 is less skewed and encompasses a broader range of higher oxygen fraction values.

2.3 Missing Data

Variables with More Than 10% Missing Data

Variable	Proportion (%)	n
inspired_oxygen.44	44.87	446
p_delta.44	44.87	446
weight_today.44	44.77	445
peep_cm_h2o_modified.44	44.67	444
any_surf	43.46	432
ventilation_support_level_modified.44	42.45	422
med_ph.44	42.45	422
com_prenat_ster	19.42	193
p_delta.36	12.88	128
peep_cm_h2o_modified.36	11.77	117

From the table above we can see that, approximately 40% of the data lacked complete information for the 44-week variables, this pattern of missing data typically occurred together. This can be attributed to various factors such as early hospital discharge, death, or information not being recorded. The variable `any_surf` also exhibited a significantly high rate of missing data but did not follow the same pattern as the other variables and was subsequently excluded from the analyses since imputing this variable would introduce large bias. Beyond these, only three other variables—`comp_prenat_ster`, `p_delta.36`, and `peep_cm_h2o_modified.36`—had missing values exceeding 10%. Notably, our primary outcome variable and all other predictors showed minimal to no missing data.

Given the project's aim to develop and compare models for 36 and 44 weeks, the decision was taken to omit observations where all data for the 44-week variables were missing. This was done to prevent the potential introduction of significant bias that might arise from imputing a large fraction of missing data for these variables. This was also imperative to ensure that the models fit could be properly compared, which might be a concern if the number of complete observations for the 36-week models was much larger than the 44-week models. This led to reduction in the sample size from 994 after data cleaning to a final 572 observations. Additionally, it is important to note that this adjustment altered the balance of outcomes in our dataset. Initially, the distribution was 82% for outcome = 0 and 18% for outcome = 1. After omitting certain observations, the final proportions shifted to 77% for outcome = 0 and 23% for outcome = 1.

To address the remaining missing data, multiple imputation was utilized to create five completed datasets. Following this, models were then applied to each of these datasets, and the results were averaged to ensure a balanced representation in our analysis. This methodology was key in enhancing the robustness of our findings and minimizing potential bias introduced by missing data.

2.4 Model Development

In the model development phase of this project, a total of six models were trained: three using the 36-week data and three using the 44-week data. Each of these models was fit to each of the five imputed datasets. The models selected for this study were Lasso, Ridge, and Forward Stepwise Logistic Regression, each chosen for their unique attributes in handling regression analysis and variable selection.

Lasso Regression and Ridge Regression both promote simplicity in models by shrinking coefficients, however, Lasso also promotes sparsity in models by outright setting certain variables to have coefficients of 0 removing them from the model. For both Lasso and Ridge 10-fold cross-validation was employed to determine the optimal lambda value, the tuning parameter that controls the strength of the penalty applied to each model. Lastly, Forward Stepwise Regression was also used for variable selection when creating our model. This method incrementally added variables based on the AIC, focusing on the variables that would lead to the most significant model improvements.

By utilizing these models, the analysis aimed to identify the most significant predictors for tracheostomy or death in neonates with sBPD, while maintaining model sparsity, key for model interpretation. To perform model selection, each of the five imputed datasets was used to fit the models. After this, the coefficients and evaluation metrics were averaged across these datasets, to further ensure that the results were robust and generalizable across different potential scenarios in the dataset.

3. Results

3.1 Model Interpretation

36-Week Models

Variable	Fwd Step	Ridge	Lasso
(Intercept)	-9.5272	-6.2342	-6.4373
mat_ethn	0.0000	0.0633	0.0000
bw	-0.0003	~0.0000	0.0000
ga	0.2025	0.0959	0.1060
blength	0.0000	-0.0077	0.0000

Variable	Fwd Step	Ridge	Lasso
birth_hc	0.0000	0.0186	0.0000
del_method	0.0000	0.1314	0.0050
prenat_ster	0.3053	0.3078	0.1557
com_prenat_ster	0.0000	0.0849	0.0089
mat_chorio	0.0000	0.0885	0.0084
gender	0.0000	-0.0792	0.0000
sga	0.0000	0.1865	0.0677
weight_today.36	-0.0006	-0.0004	-0.0005
ventilation_support_level.36	1.0011	0.4459	0.6978
inspired_oxygen.36	4.3147	2.5446	3.6347
p_delta.36	0.0000	0.0191	0.0058
peep_cm_h2o_modified.36	0.0209	0.0285	0.0104
med_ph.36	0.1263	0.4849	0.2643

The 36-week model coefficients, as presented, illustrate the distinct methodologies of Forward Stepwise, Ridge, and Lasso Logistic Regression in variable selection and weighting. The Forward Stepwise model is notably sparse, including only variables with the most statistically significant contributions, as evidenced by its inclusion of only 8 non-zero coefficients. Ridge Regression, in contrast, tends to shrink coefficients but includes a broader range of variables, unlike Lasso, which combines coefficient shrinkage with variable selection, resulting in approximately 12 non-zero variables.

Key variables like **`ventilation_support_level.36`** and **`inspired_oxygen.36`** are significantly weighted across all models, highlighting their crucial role in predicting outcomes, consistent with their direct relevance to respiratory health in neonates. Other variables such as **`prenat_ster`** (prenatal steroid use), **`ga`** (gestational age), **`peep_cm_h2o_modified.36`** (Positive and exploratory pressure), and **`med_ph.36`** (Medication for Pulmonary Hypertension) also show consistency across models, indicating a greater likelihood of outcome. Interestingly, variables like **`del_method`**, **`sga`**, **`mat_chorio`**, and **`com_prenat_ster`** are selected in Lasso but not in Forward Stepwise, highlighting the different methodologies in handling predictors with varying statistical significance.

44-Week Models

Variable	Fwd Step	Ridge	Lasso
(Intercept)	-7.8588	-6.9090	-5.7616
mat_ethn	0.0000	0.1653	0.0000
bw	0.0000	0.0001	0.0000
ga	0.0000	0.0404	0.0136
blength	0.0172	0.0122	0.0000
birth_hc	0.1034	0.0427	0.0308

Variable	Fwd Step	Ridge	Lasso
del_method	0.1265	0.2185	0.1075
prenat_ster	0.4201	0.3853	0.1953
com_prenat_ster	0.2974	0.2039	0.1515
mat_chorio	0.0000	0.0430	0.0000
gender	0.0000	-0.0751	0.0000
sga	0.0000	0.1730	0.0344
weight_today.36	0.0000	-0.0002	0.0000
ventilation_support_level.36	0.3440	0.3207	0.1877
inspired_oxygen.36	0.4170	0.8507	0.7279
p_delta.36	0.0058	0.0039	0.0003
peep_cm_h2o_modified.36	0.0000	-0.0030	0.0000
med_ph.36	0.0000	0.1355	0.0000
weight_today.44	-0.0008	-0.0003	-0.0004
ventilation_support_level_modified.44	1.9307	0.6656	1.5015
inspired_oxygen.44	0.0000	0.0207	0.0000
p_delta.44	0.0000	0.0221	0.0020
peep_cm_h2o_modified.44	0.0000	0.0915	0.0212
med_ph.44	1.7257	1.2127	1.3782

The 44-week models incorporate additional variables compared to the 36-week models. Forward Stepwise remains the sparsest model, selecting only 11 out of 22 variables as significant. Ridge, maintaining its characteristic of including most variables through coefficient shrinkage, contrasts with Lasso, which again demonstrates a middle ground in terms of variable selection with 14 non-zero variables.

Similar to the 36-week model, key predictors like ventilation support at both 36 and 44 weeks, **`med_ph.36`** (Medication for Pulmonary Hypertension), and infant weight variables at 36 and 44 weeks are integral. Most variables significant in the 36-week model retain their relevance in the 44-week model, but with slightly reduced impacts as indicated by smaller coefficients. This shift in variable importance potentially reflects the developmental changes in neonates health over time.

3.2 Model Performance

36-Week Models

Metric	Fwd Step	Ridge	Lasso
Accuracy	0.8228	0.8263	0.8228
F1 Score	0.4336	0.4198	0.4239
Brier Score	0.1201	0.1237	0.1221

Metric	Fwd Step	Ridge	Lasso
AUC	0.8468	0.8537	0.8512
Sensitivity	0.3250	0.3000	0.3167
Specificity	0.9556	0.9667	0.9578

44-Week Models

Metric	Fwd Step	Ridge	Lasso
Accuracy	0.8596	0.8456	0.8614
F1 Score	0.6643	0.6034	0.6429
Brier Score	0.0931	0.0927	0.0950
AUC	0.9131	0.9213	0.9154
Sensitivity	0.6583	0.5583	0.5917
Specificity	0.9133	0.9222	0.9333

In comparing the Forward Stepwise, Ridge, and Lasso models for both the 36-week and 44-week datasets, each model demonstrates unique strengths and weaknesses. Accuracy, a measure of the overall correctness of the model, is relatively similar across all models, indicating a consistent level of general predictive ability, however is slightly higher among 44-week models. The F1 score, which balances precision and recall, is notably higher in the 44-week models compared to the 36-week models, suggesting improved balance in identifying true positives and negatives at this later time point.

The Brier score, a measure of the accuracy of probabilistic predictions, shows slight improvements in the 44-week models, indicating more accurate probability estimates for the outcomes. The AUC (Area Under the Curve) reflects the model's ability to discriminate between the two classes (tracheostomy/death vs. none). Higher AUC values in the 44-week models suggest better discriminative power at this stage. Sensitivity, or the true positive rate, is significantly higher in the 44-week models, especially in the Forward Stepwise model, indicating a stronger ability to correctly identify neonates who need tracheostomy or are at risk of death. Specificity, or the true negative rate, is also higher in these models, with Lasso showing the highest specificity.

Comparing the 36-week and 44-week models reveals that the models developed using the 44-week data generally outperform those only using the 36-week data across most metrics. This suggests that as neonates age, the available clinical data provides a clearer and more accurate picture of their health status, enhancing the model's predictive capabilities.

The 44-week Lasso model stands out for its accuracy and specificity, making it particularly adept at minimizing false positives—important in preventing unnecessary surgical interventions. On the other hand, the Forward Stepwise model at 44 weeks shows the best sensitivity, indicating its strength in minimizing false negatives. This is crucial in a clinical context, as a false negative could mean a missed opportunity for timely tracheostomy, potentially leaving a neonate without necessary respiratory support.

In neonatal care, this balance between avoiding unnecessary procedures (false positives) and ensuring necessary interventions are not missed (false negatives) is vital. These results suggest that as neonates approach 44 weeks, healthcare providers using these models could make more informed decisions, potentially improving outcomes by tailoring interventions more precisely. The choice of model in a clinical setting would heavily realy on whether the priority is to minimize risk of unnecessary surgery or to ensure neonates in need of intervention are accurately identified.

4. Limitations

While this report yields important insights, it is important to acknowledge several limitations that may impact the interpretation and applicability of the project's findings. A primary limitation of this study is the significant amount of missing data, leading to a reduced sample size for analysis. From an initial pool of 994 observations, only 572 complete records could be utilized. This reduction in sample size not only raises concerns about the generalizability of the findings but also introduces potential bias. While techniques like multiple imputation were used to mitigate the effects of missing data, these methods cannot fully compensate for the lack of information, especially when the missingness is not random and is potentially related to underlying systemic factors in neonatal care.

Additionally, each regression model used in this study has inherent limitations that must also be considered. The Forward Stepwise model, for instance, may exclude variables that, while not statistically significant, could be clinically important. Ridge and Lasso regressions, known for handling multicollinearity, might oversimplify complex relationships through coefficient reduction. These model-specific constraints highlight the need for careful interpretation of the results and suggest that reliance on any single model may provide an incomplete picture.

Lastly, the study's external validity is another area of concern. The data, derived from specific neonatal intensive care centers, may not fully represent the broader neonatal population or encompass the diversity data collected from outside healthcare environments. This limitation can affect the applicability of the study's conclusions to different settings or populations. Overall, these limitations underscore the need for further research and careful consideration in the clinical application of any findings.

5. Conclusion

This report is the first step in creating a data-driven understanding of tracheostomy timing in neonates with severe BPD. Utilizing a national dataset and three statistical models—Forward Stepwise, Ridge, and Lasso Logistic Regression—we've identified key variables such as ventilation support, inspired oxygen levels, and various other prenatal and postnatal factors that are critical in predicting the need for tracheostomy and associated mortality risks. Significantly, our analysis indicates that data from 44 weeks post-menstrual age yield more accurate predictions than earlier data, emphasizing the dynamic nature of neonatal health.

However, this project should be viewed as a preliminary exploration, highlighting the potential for developing a clinically applicable model in the future. The limitations, including substantial missing data and the specific constraints of each statistical model, weaken the immediate applicability of our findings. Moreover, the report's scope, confined to a particular dataset, may not fully capture the diversity of the neonatal population or different healthcare settings.

Despite these challenges, our work shows a path for future research. It underscores the value of analytical approaches in neonatal care and points to the need for more extensive data collection and nuanced analysis. By continuing this research, the aim should be to refine these predictive models, making them robust enough for clinical application.

Appendix

Github: https://github.com/Kdossal/2550_Project2

Code: https://github.com/Kdossal/2550_Project2/blob/main/Code/Code.qmd