Tracheostomy/Death Prediction Model in Neonates with Severe Bronchopulmonary Dysplasia at 36-week Gestational Age

Han Ji

2023-11-13

Abstract

Background: Bronchopulmonary dysplasia (BPD) is a chronic lung condition affecting 10,000 to 15,000 premature infants annually in the United States. Despite advances in neonatal care, the number of severe BPD cases remain steady, particularly among extremely low birth weight (ELBW) infants.

Methods: This study addresses the critical need for effective prediction models for composite BPD outcomes - the potential need for tracheostomy and death. Based on the records of infants with severe BPD across the US and Sweden from BPD Collaborative Registry (n = 985), we developed three predictor models: Lasso regression with and without center and multilevel lasso regression with center as a random effect. Multiple imputation was performed to address missing data, and 5-fold cross-validation was performed to optimize the hyperparameter.

Results: The fitted models are evaluated on the validation set split from the original data, and the multilevel lasso model has the best performance (AUC = 0.910). Our findings reveal significant variability in outcomes across different centers, underscoring the importance of considering clinical setting heterogeneity. Meanwhile, the estimates agree that ventilation support, inspired oxygen, prenatal corticosteriod, and hospital discharge gestational age are the most significant predictors, consistent with the existing literature. However, limitations such as potential bias due to variable missing proportions across centers and the challenge of predicting outcomes for centers not represented in the training data were acknowledged. Also, one of the predictors hospital discharge gestational age may need an additional model to simulate.

Conclusion: This study provides a prediction model for BPD outcomes with easily accessible clinical measurements, and it accommodates the complexity of clinical settings. These results potentially lead to individualized care for premature infants with severe BPD at early stages.

Introduction

Bronchopulmonary dysplasia (BPD) is a chronic lung condition primarily affecting premature infants, marked by an imbalance between lung injury and repair during lung development [1, 2]. As the most common respiratory morbidity in preterm infants, BPD impacts nearly 10,000 – 15000 neonates annually in the United States. Despite significant advancements in managing extremely low birth weight (ELBW) infants, the incidence of BPD over the past two decades hasn't decreased but remained relatively. The pathogenesis of BPD is complex and varied by genetic and epigenetic factors for different individuals. Gestational age and birth weight are two of the strongest predictors: the risk of BPD is directly proportional to the degree of prematurity and low birth weight. For instance, infants born at 23 weeks have a significantly higher incidence and severity of BPD compared to those born at 28 weeks. Chorioamnionitis is also a frequent cause of preterm births, which likely results in serious complications in mother and babies but also BPD [3]. On the other hand, there are life-saving interventions including oxygen supplementation and mechanical ventilation, but they can potentially disrupt lung development and cause damage for neonates, and thus mechanical ventilation should be avoid when possible [1]. To prevent this damage but also help infants acquire necessary oxygen, tracheostomy can be planted in patients with severe BPD when discharged from hospital [2]. Although tracheostomy is still hooked with a ventilator, it has benefits in infant lung and neural development including improving ventilator synchrony, weaning sedation requirements, etc. However, tracheostomy is associated with an increased risk of death, meaning whether using tracheostomy should be carefully considered. Unal et al. conducted case studies with 9 neonates admitted to neonatal intensive care units (NICU), and they showed tracheotomy, if carried out correctly, makes infant nursing easier with low complication [4]. Adaikalam et al. have developed a tracheostomy prediction model for neonatal BPD via lung and airway MRI, while MRI might not be available for neonates at many hospitals. At present, no prediction models for tracheostomy based on easily accessible clinical measurements and diagnosis. This study aims to develop a tracheostomy prediction model in neonatal BPD with easily accessible measurements.

Methods

Study setting and population

Study participants were drawn 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. The registry includes infants whose gestational age is less than 32 weeks and who have severe bronchopulmonary dysplasia (sBPD) (defined by 2001 NHLBI criteria; specifically, FiO2 3 0.3 or positive pressure ventilation (invasive or non-invasive) at 36-weeks PMA). In the registry, standard demographic and clinical data are collected at four time points: birth, 36 weeks PMA, 44 weeks PMA and discharge. For this study, we queried the registry for patients with BPD and complete growth data between January 1 and July 19, 2021. At the time of analysis, 10 BPD Collaborative centers had contributed data meeting study inclusion criteria. [2]

Model derivation

Lasso regression models was chosen to perform variable selection, and two separate models were developed with center as a factor and without center. This modeling part was performed using `glmnet` package in R [5]. A multilevel lasso regression model was developed by treating the data clustered by center, and a random effect was fit for center while fixed effects for the rest. This modeling part was performed using `glmmLasso` package in R [6]. One-hot encoding was applied to the train data, while center indicator was excluded from normalization for the multilevel lasso.

Imputation and cross-validation

Multiple imputation was applied following the guidance by He [7]. Specifically, we replaced the missing values for surfactant usage indicator because its missing proportion is too high. Also, we used the original outcome variables tracheostomy and death in imputation formula to retain the association between predictors and the outcome rather than the composite outcome. The data set was split into train set (n = 660) and validation set (n = 325) by 2:1 ratio sampled by center. Multiple imputation with 5 imputed sets was done first in the train set, and the same imputation model was fit on the test set separately. This prevents the imputed values in the train set borrow information from test set and vice versa. All imputation was done by `mice` package in R [8].

5-fold cross-validation was performed to estimate the optimal hyperparameter λ for all three models. The fold index was generated by center and remained the same. The cross-validation for two lasso regression models was done by function `cv.glmnet`. For the multilevel lasso regression, a grid search method was applied: for each validation set, AUC was calculated for 20 λ ranging evenly from 0 to the max possible value where all coefficients shrink to 0. The optimal λ was obtained by the highest average AUC. We then repeat this grid search within a smaller range bound the neighboring values of the optimal λ.

Model performance

Models were examined on the validation set based on their sensitivity, specificity, AUC, and F-score. The optimal threshold for prediction was determined by Youden Index, where the maximum of sensitivity + specificity was achieved [9]. The final validation set was composed of 5 imputed sets together, which is equivalent to evaluating the models on each imputed set and taking average.

Results

Study population characteristics

Between January 1 and July 19, 2021, a total of 999 records of neonates with BPD were in the registry. Of these, 3 were excluded due to duplicate records, 5 were excluded because the their center size (n = 4; n = 1) were too small in this study, 2 were excluded due to missing death indicators, and 4 were excluded due to outliers in gestational age at discharge. Because of the large missing proportions, we removed all variables related to 44-week measurements. Maternal race was excluded because its values in the data didn't match the codebook. The remaining variables have demographic information, measurements of infants at birth and at 36 weeks, supporting factors like steroids, tracheostomy usage, and death. We combined tracheostomy and death into a composite outcome - whether infants had tracheostomy or died - because the number of cases for each is too low.

Figure 1 and 2 summarize the above characteristics stratified by center, and centers differ from each other significantly regarding the baseline and the outcome. Center 2 has the highest number of infants admitted (n = 630), while center like 16 only has 38. Infants in center 1 have a low median gestational age, short birth length, and birth weight, which is associated with the high proportion in composite outcome; however, infants in center 5 have worse baseline characteristics but has a much lower proportion in composite outcome. There is no obvious trend agreed across centers regarding the association between characteristics and the outcome.

Figure 3 shows the positive correlation between birth weight and gestational age, and it seems to have a regression line can determine whether the infant is small for gestational age (SGA) or not. This could imply multi-collinearity in the data since many clinical decisions and measurements are associated. In addition, most infants with composite outcome are not SGA, meaning they should have a better baseline health level but the results are counterintuitive.

Characteristic	1, N = 65[1]	2, N = 629[1]	3, N = 54[1]	4, N = 59[1]	5, N = 40[1]	7, N = 32[1]	12, N = 68[1]	16, N = 38[1]	p-value[2]
Maternal Ethnicity									<0.001
Hispanic or Latino	6 (9.2%)	24 (3.8%)	13 (24%)	5 (8.5%)	8 (20%)	1 (3.1%)	7 (10%)	6 (16%)	
Not Hispanic or Latino	34 (52%)	605 (96%)	39 (72%)	52 (88%)	32 (80%)	4 (13%)	61 (90%)	32 (84%)	
Unknown	25 (38%)	0 (0%)	2 (3.7%)	2 (3.4%)	0 (0%)	27 (84%)	0 (0%)	0 (0%)	
Birth Weight (g)									<0.001
Median (IQR)	652 (549, 785)	770 (610, 968)	720 (588, 859)	790 (645, 975)	593 (515, 666)	695 (540, 863)	720 (589, 923)	788 (650, 1,076)	
Unknown	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	
Gestational Age (wk)									<0.001
Median (IQR)	25 (24, 27)	26 (24, 27)	26 (24, 27)	25 (25, 27)	24 (23, 25)	25 (23, 27)	26 (25, 27)	26 (24, 28)	
Unknown	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	
Birth Length (cm)									<0.001
Median (IQR)	31 (29, 33)	32 (30, 35)	32 (31, 34)	33 (31, 35)	29 (28, 31)	32 (29, 36)	33 (31, 34)	33 (31, 37)	
Unknown	9 (14%)	24 (3.8%)	0 (0%)	1 (1.7%)	1 (2.5%)	6 (19%)	36 (53%)	0 (0%)	
Birth Head Circumference (cm)									<0.001
Median (IQR)	22.25 (21.00, 23.50)	23.00 (21.50, 25.00)	23.50 (21.75, 25.00)	23.50 (22.00, 25.00)	21.00 (20.00, 22.00)	22.00 (20.53, 24.00)	23.00 (21.25, 24.38)	23.50 (21.81, 25.50)	
Unknown	9 (14%)	29 (4.6%)	0 (0%)	2 (3.4%)	0 (0%)	6 (19%)	30 (44%)	0 (0%)	
Delivery Method									0.040
Viginal delivery	16 (25%)	176 (28%)	15 (28%)	18 (31%)	14 (35%)	10 (31%)	17 (25%)	14 (37%)	
Cesarean section	48 (74%)	453 (72%)	39 (72%)	41 (69%)	26 (65%)	22 (69%)	49 (72%)	24 (63%)	
Unknown	1 (1.5%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	2 (2.9%)	0 (0%)	
Prenatal Corticosteriods									<0.001
No	5 (7.7%)	85 (14%)	6 (11%)	12 (20%)	3 (7.5%)	4 (13%)	5 (7.4%)	4 (11%)	
Yes	56 (86%)	543 (86%)	46 (85%)	46 (78%)	37 (93%)	26 (81%)	41 (60%)	33 (87%)	
Unknown	4 (6.2%)	1 (0.2%)	2 (3.7%)	1 (1.7%)	0 (0%)	2 (6.3%)	22 (32%)	1 (2.6%)	
Complete Prenatal Steriods									<0.001
No	15 (23%)	110 (17%)	6 (11%)	18 (31%)	10 (25%)	8 (25%)	17 (25%)	7 (18%)	
Yes	34 (52%)	414 (66%)	40 (74%)	24 (41%)	27 (68%)	15 (47%)	26 (38%)	26 (68%)	
Unknown	16 (25%)	105 (17%)	8 (15%)	17 (29%)	3 (7.5%)	9 (28%)	25 (37%)	5 (13%)	
Maternal Chorioamniontis									<0.001
No	18 (28%)	524 (83%)	28 (52%)	50 (85%)	25 (63%)	28 (88%)	63 (93%)	31 (82%)	
Yes	17 (26%)	105 (17%)	4 (7.4%)	8 (14%)	14 (35%)	3 (9.4%)	5 (7.4%)	2 (5.3%)	
Unknown	30 (46%)	0 (0%)	22 (41%)	1 (1.7%)	1 (2.5%)	1 (3.1%)	0 (0%)	5 (13%)	
Gender									0.7
Female	26 (40%)	248 (39%)	21 (39%)	28 (47%)	17 (43%)	16 (50%)	28 (41%)	20 (53%)	
Male	38 (58%)	379 (60%)	32 (59%)	31 (53%)	23 (58%)	16 (50%)	40 (59%)	18 (47%)	
Unknown	1 (1.5%)	2 (0.3%)	1 (1.9%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	

[1] n (%)

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

Figure 1: Summary table of the characteristics stratified by center.

Characteristic	**1**, N = 65[1]	**2**, N = 629[1]	**3**, N = 54[1]	**4**, N = 59[1]	**5**, N = 40[1]	**7**, N = 32[1]	**12**, N = 68[1]	**16**, N = 38[1]	p-value[2]
Small for Gestational Age									0.004
Not SGA	38 (58%)	502 (80%)	40 (74%)	53 (90%)	32 (80%)	24 (75%)	51 (75%)	31 (82%)	
SGA	26 (40%)	117 (19%)	11 (20%)	5 (8.5%)	8 (20%)	8 (25%)	17 (25%)	7 (18%)	
Unknown	1 (1.5%)	10 (1.6%)	3 (5.6%)	1 (1.7%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	
Surfactant used in the first 72 hrs									<0.001
No	4 (6.2%)	70 (11%)	2 (3.7%)	4 (6.8%)	2 (5.0%)	3 (9.4%)	9 (13%)	7 (18%)	
Yes	36 (55%)	265 (42%)	51 (94%)	11 (19%)	33 (83%)	3 (9.4%)	47 (69%)	9 (24%)	
Unknown	25 (38%)	294 (47%)	1 (1.9%)	44 (75%)	5 (13%)	26 (81%)	12 (18%)	22 (58%)	
Weight at 36 weeks (g)									0.011
Median (IQR)	2,100 (1,843, 2,354)	2,150 (1,865, 2,405)	2,130 (1,933, 2,445)	2,100 (1,860, 2,400)	1,943 (1,723, 2,138)	2,200 (1,925, 2,420)	2,020 (1,785, 2,180)	2,273 (2,069, 2,472)	
Unknown	17 (26%)	36 (5.7%)	3 (5.6%)	6 (10%)	0 (0%)	1 (3.1%)	29 (43%)	0 (0%)	
Ventilation Support at 36 weeks									<0.001
No respiratory support or suppl. oxygen	8 (12%)	49 (7.8%)	5 (9.3%)	8 (14%)	0 (0%)	22 (69%)	1 (1.5%)	22 (58%)	
Non-invasive positive pressue	22 (34%)	425 (68%)	34 (63%)	34 (58%)	31 (78%)	8 (25%)	16 (24%)	14 (37%)	
Invasive positive pressure	34 (52%)	146 (23%)	14 (26%)	17 (29%)	9 (23%)	2 (6.3%)	32 (47%)	2 (5.3%)	
Unknown	1 (1.5%)	9 (1.4%)	1 (1.9%)	0 (0%)	0 (0%)	0 (0%)	19 (28%)	0 (0%)	
Fraction of Inspired Oxygen at 36 weeks									<0.001
Median (IQR)	0.35 (0.30, 0.49)	0.27 (0.23, 0.35)	0.30 (0.25, 0.35)	0.40 (0.30, 0.50)	0.33 (0.26, 0.43)	0.35 (0.32, 0.38)	0.35 (0.28, 0.45)	0.35 (0.27, 0.39)	
Unknown	18 (28%)	36 (5.7%)	2 (3.7%)	3 (5.1%)	0 (0%)	1 (3.1%)	29 (43%)	0 (0%)	
Peak Inspiratory Pressure at 36 weeks (cm H2O)									<0.001
Median (IQR)	2 (0, 14)	0 (0, 0)	0 (0, 15)	4 (0, 9)	0 (0, 9)	0 (0, 0)	12 (0, 15)	0 (0, 0)	
Unknown	20 (31%)	39 (6.2%)	5 (9.3%)	15 (25%)	13 (33%)	1 (3.1%)	32 (47%)	0 (0%)	
Positive and exploratory pressure at 36 weeks (cm H2O)									<0.001
Median (IQR)	8 (6, 9)	7 (6, 8)	8 (7, 10)	6 (6, 7)	9 (8, 10)	0 (0, 5)	6 (6, 7)	0 (0, 8)	
Unknown	24 (37%)	41 (6.5%)	9 (17%)	6 (10%)	0 (0%)	1 (3.1%)	34 (50%)	0 (0%)	
Medication for Pulmonary Hypertension at 36 weeks	13 (20%)	25 (4.0%)	3 (5.7%)	10 (17%)	3 (7.5%)	2 (6.3%)	4 (8.2%)	4 (11%)	<0.001
Hospital Discharge Gestational Age									<0.001
Median (IQR)	60 (60, 60)	47 (42, 55)	44 (41, 45)	NA (NA, NA)	52 (48, 54)	43 (39, 47)	51 (47, 59)	40 (39, 43)	
Unknown	64 (98%)	0 (0%)	0 (0%)	59 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	
Trachoetomy	27 (42%)	64 (10%)	1 (1.9%)	11 (19%)	5 (13%)	1 (3.1%)	35 (51%)	1 (2.6%)	<0.001
Death	7 (11%)	29 (4.6%)	0 (0%)	1 (1.7%)	2 (5.0%)	0 (0%)	14 (21%)	0 (0%)	<0.001
Composite Outcome	34 (52%)	84 (13%)	1 (1.9%)	12 (20%)	7 (18%)	1 (3.1%)	41 (60%)	1 (2.6%)	<0.001

[1] n (%)

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

Figure 2: (Continued) Summary table of the characteristics stratified by center.

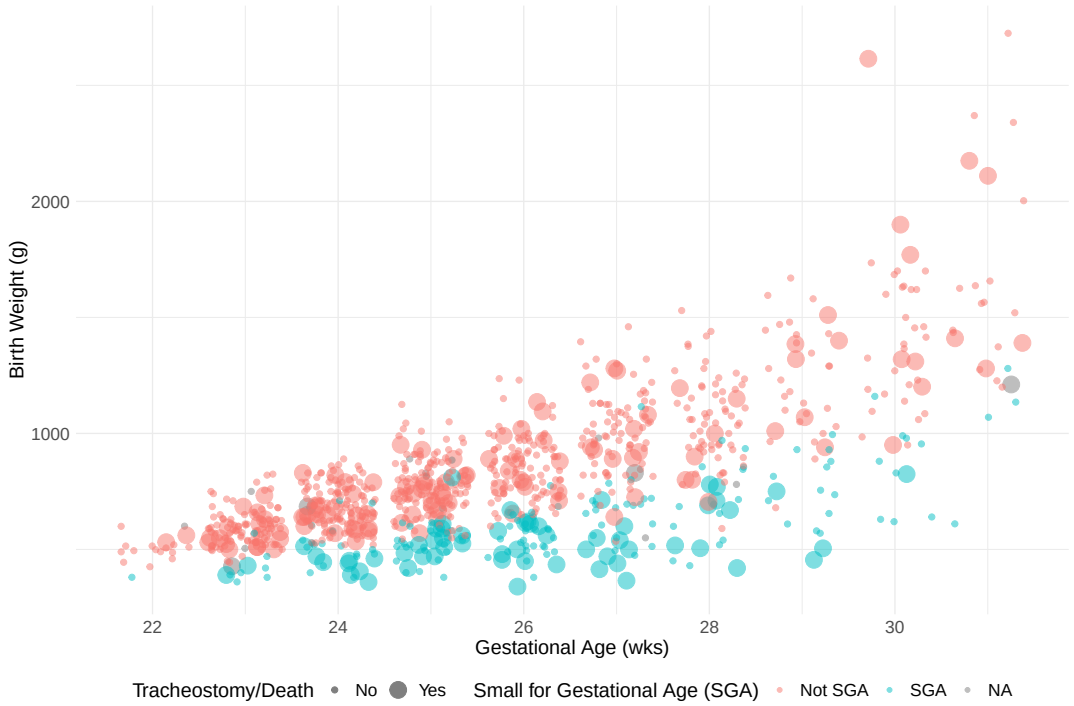

Figure 3: Association between birth weight and gestational age at birth. Points are colored by SGA and scaled by the composite outcome.

Model Performance

The model coefficients are obtained by averaging the coefficients from each imputed sets (Table 1). For consistency, we calculate the intercept for each center by adding the intercept and the individual center intercept. We see the overall trend among coefficients is consistent across three models as the sign and the relative magnitude are comparable. No variables are selected out by multilevel lasso regression, but they are close to 0 if these variables are selected out by the lasso regression with or without center. We see the predictors prenatal corticosteriods, invasive ventilation support at 36 weeks, inspired oxygen at 36 weeks, and gestational age have a high positive estimate across three models. The predictor non-invasive ventilation support at 36 weeks have a high negative estimate.

The most significant difference among these three models comes from the center intercepts due to the model derivation. The lasso regression without using center as a predictor has the same intercept. The one uses center and the multilevel lasso model have varying intercepts, but we see the difference tends to be larger in the multilevel one. For example, the maximum difference in intercepts is about 0.7 in the first lasso model, but the difference between center 2 and center 12 is more than 1 in multilevel lasso.

Table 1: Coefficient estimates from 3 models. The intercepts for Lasso models are calculated by combining the overall intercept and the individual center intercept, while the model without center doesn't have individual intercepts. Coefficients are rounded to three digits for display.

	Lasso (with center)	Lasso (without center)	Multilevel Lasso
Center 1	-2.010	-1.863	-1.107
Center 2	-2.415	-1.863	-2.047
Center 3	-2.281	-1.863	-1.853
Center 4	-2.010	-1.863	-1.530
Center 5	-2.011	-1.863	-1.587
Center 7	-2.025	-1.863	-1.673
Center 12	-1.751	-1.863	-1.039
Center 16	-2.013	-1.863	-1.654
Not Hispanic or Latino	0.000	0.000	0.033
Birth Weight	0.000	0.000	0.042
Gestational Age	0.000	0.000	0.001
Birth length	0.000	0.000	-0.042
Birth Head Circumference	0.004	0.000	0.013
Delivery Method: Cesarean section	0.000	0.000	0.013
Prenatal Corticosteriods: Yes	0.210	0.187	0.176
Complete Prenatal Steriods: Yes	0.000	0.000	-0.003
Maternal Chorioamniontis: Yes	0.000	0.000	0.007
Gender: Male	0.000	0.000	0.004
SGA	0.000	0.002	0.021
Surfactant: Yes	0.000	0.000	-0.027
Surfactant: No	0.000	0.000	0.001
Weight at 36 weeks	0.000	-0.002	-0.013
Ventilation Support at 36 weeks: Non-invasive	-0.012	-0.047	-0.131
Ventilation Support at 36 weeks: Invasive	0.475	0.539	0.356
Inspired Oxygen at 36 weeks	0.300	0.378	0.304
Peak Inspiratory Pressure at 36 weeks	0.082	0.000	0.189
Positive and exploratory pressure at 36 weeks	0.000	0.000	0.024
Medication for Pulmonary Hypertension at 36 weeks	0.000	0.002	0.025
Hospital Discharge Gestational Age	0.945	0.825	0.713

On the validation data, we see the multilevel lasso model achives the highest AUC (0.910), which is slightly higher than the lasso model with center (0.904) and much higher than the one without center (0.884) (Table 2). It also has a highest F-score among three models. At thier optimal thresholds, we see multilevel lasso has the highest specificity, and the lasso without center has the highest sensitivity, but this performance difference can be due to the difference in threshold.

In ROC curves, we see multilevel Lasso has the best performance at high specificity compared to the other two, and it becomes comparable with the lasso model with center at high sensitivity. The lasso model without center tends to be the worst (Figure 3).

Table 2: Model performance on the evaluation data set. The values are obtained at the optimal threshold of each model determined by Youden's J statistics.

	Lasso (with center)	Lasso (without center)	Multilevel Lasso
Threshold	0.144	0.127	0.294
Sensitivity	0.839	0.882	0.800
Specificity	0.828	0.761	0.877
AUC	0.904	0.884	0.910
F-score	0.630	0.583	0.670

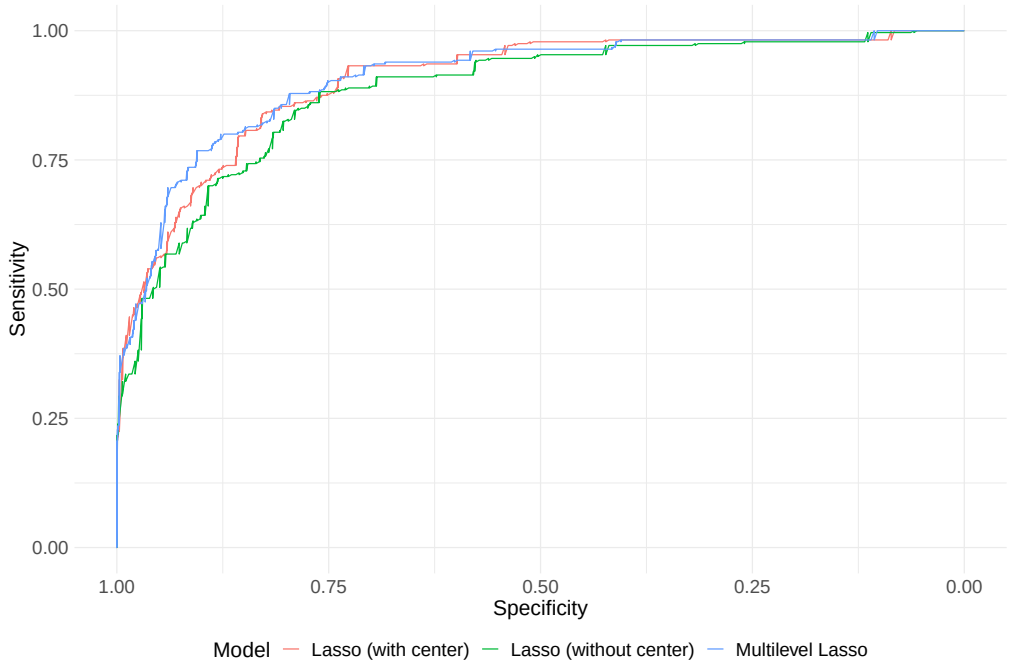

Figure 4: ROC curves for 3 models on the evaluation set.

Discussion

This analysis of premature infants with BPD illustrates the successful development of prediction model on tracheotomy/death composite outcome. Our multilevel lasso model result in a high prediction accuracy, and it not only obtains different baseline estimates for different centers but also account for the random errors across centers. Our model fitting setting illustrates three different scenarios: pooling all observations by not incorporating centers, partial pooling by using centers as predictors, and partial pooling by using centers as random effects. To obtain a higher accuracy, it is necessary to incorporate center information. In the exploratory analysis, center with similar baseline covariates can have significantly different outcomes, like center 1 and center 5. This baseline difference and difference in sample sizes could be related to the hierarchy of hospitals. For example, a tertiary referral hospital may have more advanced instruments and experiences for neonatal care while it may receive more premature infants with serious complications. Multilevel modeling can potentially provide a more precise accurate for baseline level for each center by estimating between-center and within-center variability. Although no variables are selected out by the multilevel modeling, we can set low coefficients to be 0 to get a simpler model but with comparable performance.

Also, our estimates for the 36-weeks measurements are consistent with the previous literature. We see the coefficients for invasive ventilation support and inspired oxygen are positive and large, and these two varaiables are associated with the lung undevelopment. The predictor for prenatal corticosteriods also has a

high positive coefficient, which is consistent with the beneficial effects of steriods in infant development [10]. The coefficient of non-invasive ventilation support is negative for all three models, implying that infants with non-invasive ventilation unlikely need tracheotomy at discharge. The predictor hospital discharge gestational age has the highest coefficient value, meaning the longer the stay is, the worse the health outcome the infatns have. Surprisingly, gestational age, birth length, and birth weight are not included in the two lasso models, and their coffecients are close to 0 in multilevel lasso model. This can be due to that other varaibles of clinical decisions are associated these measurements at birth and thus absorb their effects, which is similar to the collinearity we observed in exploratory analysis.

Our study has several limitations. First, the missing proportion varies for different variables at different centers. Our multiple imputation based on the full data may result in biased estimates for these variables. Second, our performance metrics accout for the imbalance in the outcome but not accout for the imbalance in the center size. Our model may yield good performance overall but incorrect predictions in some small centers. Third, our multilevel lasso and the lasso model with center both require sufficient existing observations for one center. They can't do prediction on a new data belonging to a center that doesn't exist in the train set. Finally, our model used hospital gestational age at discharge as an important predictor, but this value cannot be obtained normally at 36 weeks or even further. It may require another model to predict on the hospital stay, which can affect our model prediction accuracy.

The multilevel predictive models can be extended. For example, we can consider random slopes for predictors related to clinical decisions - different centers may have different general guidline when giving some medication and doing some tests. This is also associated with the missing proportion from the exploratory analysis as it is possible that some measurements won't be conducted in some centers usually.

Conclusion

This paper develops three prediction models on tracheotomy or death for premature babies with different approaches, and multilevel lasso model has the best performance if the center information can be obtained with a relatively large sample size. This approach accounts for the heterogeneity of clinical setting of different hospitals. The clinical application of these models is to estimate whether premature infants require additional care and resouces if they likely result in tracheotomy or death in the future. Our exprctation is that the clinical care can be individualized for children with a lower mortality rate.

References

[1] Kalikkot Thekkeveedu R, Guaman MC, Shivanna B. Bronchopulmonary dysplasia: A review of pathogenesis and pathophysiology. Respir Med. 2017 Nov;132:170-177. doi: 10.1016/j.rmed.2017.10.014. Epub 2017 Oct 24. PMID: 29229093; PMCID: PMC5729938.

[2] Paul, Alice. "PHP 2550: Practical Data Analysis." Class lecture, Brown University, Providence, RI, October 16, 2023.

[3] Jain VG, Willis KA, Jobe A, Ambalavanan N. Chorioamnionitis and neonatal outcomes. Pediatr Res. 2022 Jan;91(2):289-296. doi: 10.1038/s41390-021-01633-0. Epub 2021 Jul 1. PMID: 34211129; PMCID: PMC8720117.

[4] Unal S, Bilgin LK, Gonulal D, Akcan FA. Optimal Time of Tracheotomy in Infants: Still a Dilemma. Glob Pediatr Health. 2015 Jan 30;2:2333794X15569300. doi: 10.1177/2333794X15569300. PMID: 27335940; PMCID: PMC4784622.

[5] Friedman, J. H., Hastie, T., & Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of Statistical Software, 33(1), 1–22. https://doi.org/10.18637/jss.v033.i01

[6] Groll A (2023). *glmmLasso: Variable Selection for Generalized Linear Mixed Models by L1-Penalized Estimation*. R package version 1.6.3, https://CRAN.R-project.org/package=glmmLasso.

[7] He Y. Missing data analysis using multiple imputation: getting to the heart of the matter. Circ Cardiovasc Qual Outcomes. 2010 Jan;3(1):98-105. doi: 10.1161/CIRCOUTCOMES.109.875658. PMID: 20123676; PMCID: PMC2818781.

[8] Stef van Buuren, Karin Groothuis-Oudshoorn (2011). mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software, 45(3), 1-67. DOI 10.18637/jss.v045.i03.

[9] Ruopp MD, Perkins NJ, Whitcomb BW, Schisterman EF. Youden Index and optimal cut-point estimated from observations affected by a lower limit of detection. Biom J. 2008 Jun;50(3):419-30. doi: 10.1002/bimj.200710415. PMID: 18435502; PMCID: PMC2515362.

[10] InformedHealth.org [Internet]. Cologne, Germany: Institute for Quality and Efficiency in Health Care (IQWiG); 2006-. Pregnancy and birth: Before preterm birth: What do steroids do? 2008 Mar 14 [Updated 2018 Mar 22]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK279568/#

Code Appendix:

```
# remember to set them to FALSE
knitr::opts_chunk$set(echo = F)
knitr::opts_chunk$set(include = F)
knitr::opts_chunk$set(warning = F)
knitr::opts_chunk$set(message = F)
knitr::opts_chunk$set(eval = T)

knitr::opts_chunk$set(fig.height = 8)
knitr::opts_chunk$set(fig.width = 12)
knitr::opts_chunk$set(fig.align="center")

# Packages
library(boot)
library(tidyverse)
library(gtsummary)
library(kableExtra)
library(glmmLasso)
library(mice)
library(pROC)
# read the results
coef.lasso <- readRDS("coef.lasso.RDS")
coef.lasso_wo_center <- readRDS("coef.lasso_wo_center.RDS")
coef.glmmLasso <- readRDS("coef.glmmLasso.RDS")
coef.summary <- readRDS("coef.summary.RDS")
# Read data and basic processing
df_child_original <- read.csv("project2.csv")
df_child <- df_child_original

# delete duplicates
df_child <- df_child %>%
  group_by(record_id) %>%
  slice(1) %>%
  ungroup()

# factorize
df_child$center <- factor(df_child$center)
df_child$mat_ethn <- factor(df_child$mat_ethn,
                            labels = c("Hispanic or Latino",
                                       "Not Hispanic or Latino"))
df_child$del_method <- factor(df_child$del_method,
                              labels = c("Viginal delivery", "Cesarean section"))
df_child$prenat_ster <- factor(df_child$prenat_ster)
df_child$com_prenat_ster <- factor(df_child$com_prenat_ster)
df_child$mat_chorio <- factor(df_child$mat_chorio)
df_child$gender <- factor(df_child$gender)
df_child$sga <- factor(df_child$sga)
df_child$any_surf <- factor(df_child$any_surf)
df_child$ventilation_support_level.36 <-
  factor(df_child$ventilation_support_level.36,
         labels = c("No respiratory support or suppl. oxygen",
                    "Non-invasive positive pressue",
```

```
                    "Invasive positive pressure"))
df_child$Death <- ifelse(df_child$Death == "Yes", 1, 0)

# fill in the ones for centers - all in center 1
# this missing seems no reasons
df_child$center[is.na(df_child$center)] <- 1

# Create the new outcome based on Trach and Death
df_child$outcome <- ifelse(df_child$Trach == 1 | df_child$Death == 1,
                           1, 0)

# the subset we will focus on
df_child_sub <- df_child %>%
  filter(!center %in% c(20, 21)) %>%
  dplyr::select(-c(weight_today.44:med_ph.44)) %>%
  mutate(center = factor(center))

# remove the ones with missing outcomes, two outliers in sga
# 2 more than 300 weeks; 2 discharge wk < 30 but they have measurement at 36 wk
df_child_sub <- df_child_sub %>%
  filter(!is.na(outcome)) %>%
  filter(is.na(hosp_dc_ga) | (hosp_dc_ga < 300 & hosp_dc_ga > 30))

# Drop race - since not sure about its correspondence
df_child_sub <- df_child_sub %>%
  dplyr::select(-mat_race)
# Generate the partial table - save it
df_child_sub %>%
  dplyr::select(-record_id) %>%
  dplyr::mutate_if(is.factor, function(x)
    forcats::fct_explicit_na(x, na_level = "Unknown")) %>%
  dplyr::select(center:gender) %>%
  tbl_summary(by = "center",

              type = all_continuous() ~ "continuous2",
              label = c(mat_ethn ~ "Maternal Ethnicity",
                        bw ~ "Birth Weight (g)",
                        ga ~ "Gestational Age (wk)",
                        blength ~ "Birth Length (cm)",
                        birth_hc ~ "Birth Head Circumference (cm)",
                        del_method ~ "Delivery Method",
                        prenat_ster ~ "Prenatal Corticosteriods",
                        com_prenat_ster ~ "Complete Prenatal Steriods",
                        mat_chorio ~ "Maternal Chorioamniontis",
                        gender ~ "Gender"),
              missing = "no",
              statistic = all_continuous() ~ c("{median} ({p25}, {p75})",
                                               "{N_miss} ({p_miss}%)")) %>%
  add_p(all_categorical() ~ "chisq.test") %>%
    modify_table_body(
    dplyr::mutate,
    label = ifelse(label == "N missing (% missing)",
                   "Unknown",
```

```
                    label)) %>%
    as_gt() %>% gt::gtsave("summary_table_p1.png")
# Generate the partial table 2 - save it
df_child_sub %>%
  dplyr::select(-record_id) %>%
  dplyr::mutate_if(is.factor, function(x)
    forcats::fct_explicit_na(x, na_level = "Unknown")) %>%
  dplyr::select(-c(mat_ethn:gender)) %>%
  tbl_summary(by = "center",
              type = all_continuous() ~ "continuous2",
              label = c(sga ~ "Small for Gestational Age",
                        any_surf ~ "Surfactant used in the first 72 hrs",
                        weight_today.36 ~ "Weight at 36 weeks (g)",
                        ventilation_support_level.36 ~ "Ventilation Support at 36 weeks",
                        inspired_oxygen.36 ~ "Fraction of Inspired Oxygen at 36 weeks",
                        p_delta.36 ~ "Peak Inspiratory Pressure at 36 weeks (cm H2O)",
                        peep_cm_h2o_modified.36 ~ "Positive and exploratory pressure at 36 weeks (cm H2
                        med_ph.36 ~ "Medication for Pulmonary Hypertension at 36 weeks",
                        hosp_dc_ga ~ "Hospital Discharge Gestational Age",
                        Trach ~ "Trachoetomy",
                        outcome ~ "Composite Outcome"),
              missing = "no",
              statistic = all_continuous() ~ c("{median} ({p25}, {p75})",
                                               "{N_miss} ({p_miss}%)")) %>%
  add_p(all_categorical() ~ "chisq.test") %>%
    modify_table_body(
    dplyr::mutate,
    label = ifelse(label == "N missing (% missing)",
                   "Unknown",
                   label)) %>%
    as_gt() %>% gt::gtsave("summary_table_p2.png")
# Generate the full summary table - save it
df_child_sub %>%
  dplyr::select(-record_id) %>%
  dplyr::mutate_if(is.factor, function(x)
    forcats::fct_explicit_na(x, na_level = "Unknown")) %>%
  tbl_summary(by = "center",
              type = all_continuous() ~ "continuous2",
              label = c(mat_ethn ~ "Maternal Ethnicity",
                        bw ~ "Birth Weight (g)",
                        ga ~ "Gestational Age (wk)",
                        blength ~ "Birth Length (cm)",
                        birth_hc ~ "Birth Head Circumference (cm)",
                        del_method ~ "Delivery Method",
                        prenat_ster ~ "Prenatal Corticosteriods",
                        com_prenat_ster ~ "Complete Prenatal Steriods",
                        mat_chorio ~ "Maternal Chorioamniontis",
                        gender ~ "Gender",
                        sga ~ "Small for Gestational Age",
                        any_surf ~ "Surfactant used in the first 72 hrs",
                        weight_today.36 ~ "Weight at 36 weeks (g)",
                        ventilation_support_level.36 ~ "Ventilation Support at 36 weeks",
                        inspired_oxygen.36 ~ "Fraction of Inspired Oxygen at 36 weeks",
```

```
                        p_delta.36 ~ "Peak Inspiratory Pressure at 36 weeks (cm H2O)",
                        peep_cm_h2o_modified.36 ~ "Positive and exploratory pressure at 36 weeks (cm H2
                        med_ph.36 ~ "Medication for Pulmonary Hypertension at 36 weeks",
                        hosp_dc_ga ~ "Hospital Discharge Gestational Age",
                        Trach ~ "Trachoetomy",
                        outcome ~ "Composite Outcome"),
              missing = "no",
              statistic = all_continuous() ~ c("{median} ({p25}, {p75})",
                                               "{N_miss} ({p_miss}%)")) %>%
  add_p(all_categorical() ~ "chisq.test") %>%
    modify_table_body(
    dplyr::mutate,
    label = ifelse(label == "N missing (% missing)",
                   "Unknown",
                   label)) %>%
  as_gt() %>% gt::gtsave("summary_table.png")

# Train Test split by center, ratio 2:1
## any_surf - use "Missing" for NA values since its missing prop. is large
# Also remove record id
df_child_sub <- df_child_sub %>%
  mutate(any_surf = case_when(any_surf == "Yes" ~ "Yes",
                              any_surf == "No" ~ "No",
                              .default = "Missing"))
set.seed(1)
test_index <- (df_child_sub %>%
  group_by(center) %>%
  sample_frac(0.33))$record_id

df_child_test <- df_child_sub %>%
  filter(record_id %in% test_index)

df_child_test_wo_id <- df_child_test %>%
  select(-record_id)

df_child_train <- df_child_sub %>%
  filter(!record_id %in% test_index)

df_child_train_wo_id <- df_child_train %>%
  select(-record_id)

# Do imputation separately
## First impute on train
## Remove composite outcome - since it will cause collinearity with trach/death
m <- 5

imp.train <- mice(df_child_train_wo_id %>% dplyr::select(-outcome),
                  m = m, print = FALSE, seed = 2550)
imp.test <- mice.mids(imp.train, newdata = df_child_test_wo_id  %>% dplyr::select(-outcome))
# Cross validation - needed for lasso
## Define the fold index - keep the same
set.seed(2550)
k <- 5 # 5-fold CV
```

```r
# sample by group
folds <- (df_child_train %>%
  group_by(center) %>%
  mutate(fold_index = sample(1:k, n(), replace = TRUE)))$fold_index
# Loop through each imputed set
coef.lasso <- c()
coef.lasso_wo_center <- c()
coef.glmmLasso <- c()
for (i in 1:m){
  imp.data <- complete(imp.train, i) %>%
    dplyr::select(-c(Trach, Death)) %>%
    mutate(outcome = df_child_train$outcome)
  x.ord <- model.matrix(outcome ~. , data = imp.data)[,-1]
  x.ord <- scale(x.ord,center=TRUE,scale=TRUE)
  x.ord_wo_center <- model.matrix(outcome ~. ,
                                  data = imp.data %>% dplyr::select(-center))[,-1]
  x.ord_wo_center <- scale(x.ord_wo_center, center=TRUE,scale=TRUE)
  y.ord <- imp.data$outcome

  # Lasso
  lasso_mod_cv <- cv.glmnet(x.ord, y.ord, nfolds = k, foldid = folds,
                       alpha = 1, family = "binomial", standardize = FALSE)
  # fit on the full data
  lasso_mod <- glmnet(x.ord, y.ord,
                      alpha=1, family = "binomial",
                      lambda = lasso_mod_cv$lambda.min, standardize = FALSE)

  # Lasso without center
  lasso_mod_cv_wo_center <- cv.glmnet(x.ord_wo_center, y.ord, nfolds = k, foldid = folds,
                       alpha = 1, family = "binomial", standardize = FALSE)
  # fit on the full data
  lasso_mod_wo_center <- glmnet(x.ord_wo_center, y.ord,
                      alpha=1, family = "binomial",
                      lambda = lasso_mod_cv_wo_center$lambda.min, standardize = FALSE)

  # Multilevel Lasso
  ## Set the lambda grid - first calculate the maximum
  ## Make the combined set for glmmLasso function
  xy.ord <- cbind(imp.data$center, x.ord_wo_center)
  xy.ord <- cbind(xy.ord, y.ord)
  colnames(xy.ord)[1] <- "center"
  colnames(xy.ord)[ncol(xy.ord)] <- "outcome"
  xy.ord <- data.frame(xy.ord)
  xy.ord$center <- factor(imp.data$center)

  lambda_max <- max(abs(colSums(x.ord * ifelse(y.ord == 1, mean(y.ord), mean(y.ord)-1))))

  # write out the formula
  glmmLasso_formula <- formula(
    paste0("outcome ~ ", paste0(
      colnames(xy.ord)[!colnames(xy.ord) %in% c("center", "outcome")],
      collapse = " + ")))
```

```r
lambda_set <- seq(0, lambda_max, length.out = 20)
AUC_set <- matrix(nrow = 20, ncol = k)
for (j in 1:20){
  # Go through 5-fold CV
  ## get AUC 5 times and average for each lambda
  for (fold in 1:k){
    # fit on 4 folds
    xy.ord.cv <- xy.ord[folds != fold, ]
    glmmLasso_mod <- glmmLasso(glmmLasso_formula, rnd = list(center=~1),
                lambda=lambda_set[j],
                family = binomial(link = "logit"), data = xy.ord.cv)
    # test on the fold left
    glmmLasso_pred <- predict(glmmLasso_mod,
                              newdata = xy.ord[folds == fold, ],
                              type = "response")
    # AUC
    AUC_set[j, fold] <- as.numeric(auc(xy.ord[folds == fold, "outcome"],
                          glmmLasso_pred))
  }
}
# Average AUC
AUC_avg <- rowMeans(AUC_set)
# then repeat within a smaller range
lambda_smaller_set <- seq(lambda_set[max(1, which.max(AUC_avg) - 1)],
                          lambda_set[min(20, which.max(AUC_avg) + 1)],
                          length.out = 50)
AUC_set_final <- matrix(nrow = 50, ncol = k)
for (j in 1:50){
  # Go through 5-fold CV
  ## get AUC 5 times and average for each lambda
  for (fold in 1:k){
    # fit on 4 folds
    xy.ord.cv <- xy.ord[folds != fold, ]
    glmmLasso_mod <- glmmLasso(glmmLasso_formula, rnd = list(center=~1),
                lambda=lambda_smaller_set[j],
                family = binomial(link = "logit"), data = xy.ord.cv)
    # test on the fold left
    glmmLasso_pred <- predict(glmmLasso_mod,
                              newdata = xy.ord[folds == fold, ],
                              type = "response")
    # AUC
    AUC_set_final[j, fold] <- as.numeric(auc(xy.ord[folds == fold, "outcome"],
                          glmmLasso_pred))
  }
}
# Average AUC
AUC_avg_final <- rowMeans(AUC_set_final)

# run on full model for the optimal lambda estimated
glmmLasso_mod <- glmmLasso(glmmLasso_formula, rnd = list(center=~1),
                lambda=lambda_smaller_set[which.max(AUC_avg_final)],
                family = binomial(link = "logit"), data = xy.ord)
```

```
  # Get coefficients
  coef.lasso <- cbind(coef.lasso, coef(lasso_mod))
  coef.lasso_wo_center <- cbind(coef.lasso_wo_center, coef(lasso_mod_wo_center))
  coef.glmmLasso <- cbind(coef.glmmLasso,
                          c(glmmLasso_mod$coefficients[1], glmmLasso_mod$ranef,
                            glmmLasso_mod$coefficients[-1]))
}

# store the results
saveRDS(coef.lasso, "coef.lasso.RDS")
saveRDS(coef.lasso_wo_center, "coef.lasso_wo_center.RDS")
saveRDS(coef.glmmLasso, "coef.glmmLasso.RDS")
# take average across five imputed sets
coef.lasso.avg <- rowMeans(as.matrix(coef.lasso))
coef.lasso_wo_center.avg <- rowMeans(as.matrix(coef.lasso_wo_center))
coef.glmmLasso.avg <- rowMeans(as.matrix(coef.glmmLasso))

# Adjust for center - for formatting
## Provide intercept for each center
coef.lasso.avg.adjust <- coef.lasso.avg
coef.lasso.avg.adjust[2:8] <- coef.lasso.avg.adjust[2:8] + coef.lasso.avg.adjust[1]

coef.lasso_wo_center.avg.adjust <- c(rep(coef.lasso_wo_center.avg[1], 8),
                                     coef.lasso_wo_center.avg[-1])
coef.glmmLasso.avg.adjust <- c(coef.glmmLasso.avg[2:9] + coef.glmmLasso.avg[1],
                               coef.glmmLasso.avg[10:30])
# make summary dataframe and table
coef.summary <- data.frame(
  matrix(c(coef.lasso.avg.adjust,
           coef.lasso_wo_center.avg.adjust,
           coef.glmmLasso.avg.adjust), ncol = 3)
)
coef.summary <- round(coef.summary, 3)
row.names(coef.summary) <- c("Center 1", "Center 2", "Center 3", "Center 4",
                             "Center 5", "Center 7", "Center 12", "Center 16",
                             "Not Hispanic or Latino", "Birth Weight", "Gestational Age",
                             "Birth length","Birth Head Circumference",
                             "Delivery Method: Cesarean section",
                             "Prenatal Corticosteriods: Yes",
                             "Complete Prenatal Steriods: Yes",
                             "Maternal Chorioamniontis: Yes",
                             "Gender: Male", "SGA", "Surfactant: Yes",
                             "Surfactant: No", "Weight at 36 weeks",
                             "Ventilation Support at 36 weeks: Non-invasive",
                             "Ventilation Support at 36 weeks: Invasive",
                             "Inspired Oxygen at 36 weeks",
                             "Peak Inspiratory Pressure at 36 weeks",
                             "Positive and exploratory pressure at 36 weeks",
                             "Medication for Pulmonary Hypertension at 36 weeks",
                             "Hospital Discharge Gestational Age"
                             )
colnames(coef.summary) <- c("Lasso (with center)", "Lasso (without center)",
                            "Multilevel Lasso")
```

```r
knitr::include_graphics("summary_table_p1.png")
knitr::include_graphics("summary_table_p2.png")
# Relationship between gestational age and birthweight
# also whether too small
ggplot(df_child_sub) +
  geom_jitter(aes(x = ga, y = bw, color = sga,
                  size = factor(outcome, labels = c("No", "Yes"))),
              alpha = 0.5) +
  theme_minimal() +
  labs(x = "Gestational Age (wks)", y = "Birth Weight (g)",
       color = "Small for Gestational Age (SGA)",
       size = "Tracheostomy/Death") +
  theme(legend.position = "bottom", legend.text = element_text(size = 14),
        legend.title = element_text(size = 16),
        axis.text = element_text(size = 14),
        axis.title = element_text(size = 16))
coef.summary %>%
  kable(caption = "Coefficient estimates from 3 models. The intercepts for Lasso models are
        calculated by combining the overall intercept and the individual center intercept,
        while the model without center doesn't have individual intercepts. Coefficients are
        rounded to three digits for display.",
  align = "c",
  booktabs = T) %>%
  kable_styling(full_width=T, latex_options = c('HOLD_position'),
                font_size = 8)
# Performance on the test set

# get 5 imputed sets and combine
## equivalent to evaluation on each set and taking average
set.seed(1)
complete.test <- c()
for (i in 1:5){
  complete.test <- rbind(complete.test,
                         complete(imp.test, i))
}
# add outcome back
complete.test <- complete.test %>%
  dplyr::select(-c(Trach, Death))
complete.test$outcome <- rep(df_child_test$outcome, 5)

# add center 1
x.test <- model.matrix(outcome ~. , data = complete.test)[,-1]
x.test <- cbind("center1" = ifelse(complete.test$center == 1, 1, 0),
                x.test)
x.test.scale <- scale(x.test)

# predicted values - and transform into prob
pred.lasso <- inv.logit(x.test.scale[, 9:29] %*% coef.summary[9:29, 1] +
                            x.test[, 1:8] %*% coef.summary[1:8, 1])

pred.lasso_wo_center <- inv.logit(x.test.scale[, 9:29] %*% coef.summary[9:29, 2] +
                            x.test[, 1:8] %*% coef.summary[1:8, 2])
```

```r
pred.glmmLasso <- inv.logit(x.test.scale[, 9:29] %*% coef.summary[9:29, 3] +
                            x.test[, 1:8] %*% coef.summary[1:8, 3])
# matrix(c(pred.lasso, pred.lasso_wo_center, pred.glmmLasso), ncol = 3)
# Calculate performance metrics

## ROC object
roc.lasso <- roc(complete.test$outcome, pred.lasso)
roc.lasso_wo_center <- roc(complete.test$outcome, pred.lasso_wo_center)
roc.glmmLasso <- roc(complete.test$outcome, pred.glmmLasso)

# Find optimal threshold
## Using Youden's J statistics
optimal_index.lasso <- which.max(roc.lasso$sensitivities + roc.lasso$specificities - 1)
optimal_index.lasso_wo_center <- which.max(roc.lasso_wo_center$sensitivities +
                                           roc.lasso_wo_center$specificities - 1)
optimal_index.glmmLasso <- which.max(roc.glmmLasso$sensitivities +
                                     roc.glmmLasso$specificities - 1)

optimal_threshold.lasso <-
  roc.lasso$thresholds[optimal_index.lasso]
optimal_threshold.lasso_wo_center <-
  roc.lasso_wo_center$thresholds[optimal_index.lasso_wo_center]
optimal_threshold.glmmLasso <-
  roc.glmmLasso$thresholds[optimal_index.glmmLasso]
F_score <- function(actual, pred, threshold = 0.5){
  #' Calculate the F-score given the actual outcome and predicted values
  #' @param actual the actual outcome values
  #' @param pred the predicted probabilities from the model
  #' @param threshold the threshold to determine the predicted outcomes based on probabilities
  #' @return the F-score calculated

  pred.outcome <- ifelse(pred > threshold, 1, 0)

  # calculate precision and recall
  TP = sum(actual == 1 & pred.outcome == 1)
  FP = sum(actual == 0 & pred.outcome == 1)
  FN = sum(actual == 1 & pred.outcome == 0)

  precision = TP/(TP+FP)
  recall = TP/(TP+FN)

  # calculate F-score
  return(1/((1/recall + 1/precision)/2))
}
# summary table for optimal metrics
df_summary_metric <- data.frame(
  thresholds = c(optimal_threshold.lasso,
                 optimal_threshold.lasso_wo_center,
                 optimal_threshold.glmmLasso),
  sensitivity = c(roc.lasso$sensitivities[optimal_index.lasso],
                  roc.lasso_wo_center$sensitivities[optimal_index.lasso_wo_center],
                  roc.glmmLasso$sensitivities[optimal_index.glmmLasso]),
  specificity = c(roc.lasso$specificities[optimal_index.lasso],
```

```
                  roc.lasso_wo_center$specificities[optimal_index.lasso_wo_center],
                  roc.glmmLasso$specificities[optimal_index.glmmLasso]),
  AUC = c(roc.lasso$auc, roc.lasso_wo_center$auc, roc.glmmLasso$auc),
  F_score = c(F_score(complete.test$outcome, pred.lasso, optimal_threshold.lasso),
              F_score(complete.test$outcome, pred.lasso_wo_center,
                      optimal_threshold.lasso_wo_center),
              F_score(complete.test$outcome, pred.glmmLasso,
                      optimal_threshold.glmmLasso))
) %>% t()
rownames(df_summary_metric) <- c("Threshold", "Sensitivity", "Specificity",
                                 "AUC", "F-score")
df_summary_metric %>%
  round(3) %>%
  kable(caption = "Model performance on the evaluation data set. The values are obtained at
        the optimal threshold of each model determined by Youden's J statistics.",
        col.names = c("Lasso (with center)", "Lasso (without center)",
                      "Multilevel Lasso"),
  align = "c",
  booktabs = T) %>%
  kable_styling(full_width=T,latex_options = c('HOLD_position'),
                font_size = 8)
# summary df for ROC
df_ROC <- data.frame(sensitivity = c(roc.lasso$sensitivities,
                           roc.lasso_wo_center$sensitivities,
                           roc.glmmLasso$sensitivities),
           specificity = c(roc.lasso$specificities,
                           roc.lasso_wo_center$specificities,
                           roc.glmmLasso$specificities),
           model = c(rep("Lasso (with center)",
                         length(roc.lasso$sensitivities)),
                     rep("Lasso (without center)",
                         length(roc.lasso_wo_center$sensitivities)),
                     rep("Multilevel Lasso", length(roc.glmmLasso$sensitivities))))

# Plot ROC curves
df_ROC %>%
  ggplot() +
  geom_line(aes(x = specificity, y = sensitivity,
                group = model, color = model)) +
  theme_minimal() +
  xlim(1, 0) +
  labs(x = "Specificity", y = "Sensitivity", color = "Model") +
  theme(legend.position = "bottom", legend.text = element_text(size = 16),
        legend.title = element_text(size = 18),
        axis.text = element_text(size = 16),
        axis.title = element_text(size = 18))
```