PHP2550_Project2_Backup

Jialin Liu

2023-11-13

Abstract

Background- Existing prediction models based on large databases can estimate the likelihood of tracheostomy placement or death given baseline demographics and clinical diagnoses. However, these analyses have not used detailed respiratory parameters and have not provided prediction at different post-menstrual age (PMA). The predictive model is proposed for this purpose in this study.

Methods- We developed and internally validated logistic regression model in predicting the need for tracheostomy in infants with severe bronchopulmonary dysplasia. Variable selection methods were used to find the best subset of variables included in the logistic model, which include the lasso and forward stepwise selection methods. Both models were internally validated, and the discrimination and calibration were estimated.

Results- Notably, the forward stepwise regression model stands out as the top performer, with the highest AUC (91.6%) and the highest sensitivity (61.3%) among the three variable selection methods. While lasso models also demonstrate good performance, the forward stepwise model exhibits a slight edge in predictive accuracy.

Conclusion- The forward stepwise selection had better predictive accuracy than the lasso model in assessment of potential adverse outcome (tracheostomy placement or death) in infants who are born prematurely with chronic lung disease.

Introduction

Bronchopulmonary dysplasia (BPD), also known as chronic lung disease, causes long-term breathing problems in newborn babies especially for those who are born prematurely. As the most common complication of prematurity, this disease affects estimated 10,000-15,000 infants each year in the United States, which is caused many multifactorial individual characteristics both from genetic and epigenetic aspects and substantial impact infant's susceptibility. Compared to the healthy lung tissue which can support normal breathing, the lungs with BPD have fewer and larger the tiny air sacs of the lung (alveoli), causing tissue destruction (fibrosis and metaplasia) within the lungs and usually showing signs of respiratory distress, such as breathing quickly and grunting. This deficit in pulmonary vascular development has no cure, but it can be treated and most babies go on to live a long and healthy life. There are four levels of severity of BPD, in particular, 75% of babies with grade 3 BPD are always dependent on a ventilator at 36 weeks gestational age when they are discharged from the hospital. To allow babies to be hooked up to a ventilator for a long time, they needs a tracheostomy that is a surgical hole in the neck and tube inserted in the trachea to allow them in breath and out breath to lungs. Since some studies show that tracheostomy associated with improved outcomes within 4 months of age and a list of benefits to performing a tracheostomy, up to 12% babies with severe grade 3 BPD are required to have a tracheostomy. However, risks associated with a tracheostomy are also existing, which include increased risk of death compared to no tracheostomy, accidentally cannula obstruction or abscission, and increased rates of infection on skin, trachea and lungs.

Existing prediction models based on large databases can accurately estimate the likelihood of tracheostomy placement or death given baseline demographics and clinical diagnoses. However, these analyses have not used detailed respiratory parameters and have not provided prediction at different post-menstrual age (PMA). Accurate prediction the need for tracheostomy at early PMA would have implications for counseling of families and appropriate timing og tracheostomy placement, which is an active area of debate in severe BPD (sBPD). Motivated by deficiency in the previous work, models are designed to determine who really needs a tracheostomy, and when is an ideal time frame to refer a patient for tracheostomy. We will be using clinical data collected from multicenter, retrospective case-control study and recorded infants who are born at ≤ 32 weeks and their respiratory support at 36 and 44 weeks PMA. Outcomes of interest (tracheostomy or death) are recorded at the time point when they were discharged from hospitals. We developed and validated two prediction models and compared the performance of their predictability with respect to eventual needs for tracheostomy or death prior to discharge.

Methods

Study Population

This study analyzed data from a national data set of demographic, diagnostic, and respiratory parameters of infants with sBPD admitted to collaborative Neonatal intensive care units (NICUs) across multiple centers. The data consists of 999 participants who are born at ≤ 32 weeks and their corresponding 30 factors and outcomes of eventually healthy status at(or before) discharge. The rest of this section is devoted to describing summary statistics, procedures for preprocessing data, exploring any potential relationships between variables as well as missing values before building models with variable selection. We will conduct brief exploratory data analysis in terms of three aspects: birth and demographic variables, respiratory support variables, weight and tracheostomy placement at 36 and 44 weeks.

Of those 999 patients' records, some duplicated patient ID with information have been detected and removed for further analysis. To make sure completeness of outcome variables, we found two places of missingness in `Death` outcome variable. One of them should be corrected as "No" death since this patient was discharged from hospital at 43 weeks without tracheostomy placement, thus it is reasonable to replace this missing value with known outcome based on our basic speculation. The another missing value in `Death` cannot be deduced as missing value appeared in hospital discharge gestational age `hosp_dc_ga`, without this supporting information, we couldn't make assumption upon these bunch lack of data so that we removed this particular patient from our observational data. For conciseness and fewer number of models needed to construct, we combined two outcomes of interest, `Trach` and `Death`, into one final outcome about healthy status that refers to "Yes(1)" when babies neither had tracheostomy placement nor died at(or before) discharge, and conversely, "No(0)" represents adverse outcome of health, equally saying, babies either had tracheostomy or died.

For hospital discharge `hosp_dc_ga` variable, there exists some values that lie far much away from the main body of observations and may distort summaries of the distribution. For example, some cases showed hospital discharge gestational ages are greater than 300 weeks, which seems to be irrational in this study, since, based on boxplot and interquartile range, most often patients were discharged around 40-50 weeks. Therefore, we planned to simply remove those few cases. Additionally, the data has been collected from multiple centers, we wanted to check if number of observations are evenly distributed and balanced. We found that center 20 and 21 only consisted with a total of 5 patients, which sample size are too small to conduct further analysis on these two centers. So we decided to remove those 5 participants in center 21 and 20 from the whole dataset. In addition, levels in maternal race variable did not align with specified categories shown in the code book. Without explanation for this error, we removed this variable. For model simplicity, we considered respiratory and diagnostic related variables measured at 36 weeks only, and further analyses will be considered to add later time points and give a more comprehensive insight into an ideal time point to refer patients for tracheostomy.

Table 1 describes summary statistics for a part of participant birth and demographic variables, stratified by

outcome of interest `adverse_outcome`, with the sample size for healthy group being 806 (noted as N = 806) and for group who either had tracheostomy placement or died prior to discharge (N = 182). Birth variables, which include birth weight (in g) `bw`, birth length (in centimeters) `blength`, and head circumference at birth `birth_hc`, show statistically significant differences between healthy and non-healthy groups since all p-values are greatly less than significance level ($\alpha = 0.05$). In particular, we could observe that those babies who had adverse outcome are high likely to have lower birth weights (mean of 757g) and smaller hear circumference (mean of 22.89cm), probably due to prematurity, than those who hadn't tracheostomy placement or died at discharge birth weights (mean of 816g) and larger head circumference (mean of 23.25cm). Delivery Method `del_method` was reported as categorical data with two methods: 1 represents for vaginal delivery and 2 represents for cesarean section. Higher percentage of babies who were delivered by cesarean section experienced adverse outcome than those who were delivered by vaginal method, with a significance difference between two groups based on a small p-value from Chi-squared test. One notable thing is that `any_surf` to record if the infant receive surfactant in the first 72 hours consisted with 44% of missingness in the original data set, thus it is crucial to carefully criticize whether the assumptions of multiple imputation are likely to hold and this variable cannot be reasonably imputed from the other available data by multiple imputation method.

Table 1: Participants Baseline Demographics Variables

Characteristic	**Missing**	**0**, N = 806[1]	**1**, N = 182[1]	**p-value**[2]
bw	0 (0%)			<0.001
Mean (Maximum, SD)		816 (2,725, 284)	757 (2,615, 341)	
blength	77 (7.8%)			0.002
Mean (Maximum, SD)		33 (48, 4)	32 (45, 4)	
birth_hc	76 (7.7%)			0.013
Mean (Maximum, SD)		23.25 (36.00, 2.65)	22.89 (38.30, 3.29)	
del_method	3 (0.3%)			0.020
1		243 (30%)	39 (22%)	
2		561 (70%)	142 (78%)	
prenat_ster	33 (3.3%)			0.027
No		112 (14%)	13 (7.8%)	
Yes		677 (86%)	153 (92%)	
sga	15 (1.5%)			<0.001
Not SGA		656 (83%)	117 (66%)	
SGA		139 (17%)	61 (34%)	
any_surf	430 (44%)			0.088
No		89 (19%)	12 (12%)	
Yes		370 (81%)	87 (88%)	

[1]n (%)

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

Table 2: Summary Statistics of Respiratory Variables at 36 Weeks by Center

Characteristic	**1**, N = 55	**2**, N = 630	**3**, N = 56	**4**, N = 59	**5**, N = 40	**7**, N = 32	**12**, N = 68	**16**, N = 38	**(Missing)**, N = 10	p-value
weight_today.36										0.021
Mean (SD)	2,073 (440)	2,135 (408)	2,113 (426)	2,120 (339)	1,922 (401)	2,169 (409)	2,044 (484)	2,220 (409)	1,992 (618)	
N missing (% missing)	12 (22%)	36 (5.7%)	3 (5.4%)	6 (10%)	0 (0%)	1 (3.1%)	29 (43%)	0 (0%)	5 (50%)	
ventilation_support_level.36										<0.001
0	7 (13%)	50 (8.1%)	5 (9.1%)	8 (14%)	0 (0%)	22 (69%)	1 (2.0%)	22 (58%)	1 (11%)	
1	19 (35%)	425 (68%)	34 (62%)	34 (58%)	31 (78%)	8 (25%)	16 (33%)	14 (37%)	3 (33%)	
2	29 (53%)	146 (24%)	16 (29%)	17 (29%)	9 (22%)	2 (6.2%)	32 (65%)	2 (5.3%)	5 (56%)	
inspired_oxygen.36										<0.001
Mean (SD)	0.43 (0.21)	0.32 (0.14)	0.31 (0.09)	0.40 (0.12)	0.36 (0.13)	0.36 (0.10)	0.40 (0.19)	0.35 (0.11)	0.40 (0.06)	
N missing (% missing)	14 (25%)	36 (5.7%)	2 (3.6%)	3 (5.1%)	0 (0%)	1 (3.1%)	29 (43%)	0 (0%)	4 (40%)	
p_delta.36										<0.001
Mean (SD)	7 (8)	5 (11)	7 (8)	5 (6)	4 (7)	0 (1)	9 (7)	1 (5)	6 (8)	
N missing (% missing)	17 (31%)	39 (6.2%)	7 (12%)	15 (25%)	13 (32%)	1 (3.1%)	32 (47%)	0 (0%)	3 (30%)	
peep_cm_h2o_modified.36										<0.001
Mean (SD)	7 (4)	6 (2)	8 (3)	6 (3)	9 (2)	2 (3)	7 (2)	3 (4)	7 (5)	
N missing (% missing)	18 (33%)	41 (6.5%)	11 (20%)	6 (10%)	0 (0%)	1 (3.1%)	34 (50%)	0 (0%)	6 (60%)	
med_ph.36										<0.001
0	42 (76%)	596 (96%)	52 (95%)	49 (83%)	37 (92%)	30 (94%)	45 (92%)	34 (89%)	9 (100%)	
1	13 (24%)	25 (4.0%)	3 (5.5%)	10 (17%)	3 (7.5%)	2 (6.2%)	4 (8.2%)	4 (11%)	0 (0%)	
hosp_dc_ga										<0.001
Mean (SD)	60 (NA)	53 (18)	46 (21)	NA (NA)	54 (18)	45 (7)	54 (14)	41 (3)	NA (NA)	
N missing (% missing)	54 (98%)	0 (0%)	0 (0%)	59 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	10 (100%)	

Table 2 provides data on respiratory-related variables measured at 36 weeks across different centers with an additional column for missing center name. For weight at 36 weeks `weight_today.36` variable, statistically significant p-value (0.021) is measured to assess a tendency of weights in at least one of the groups to be different than weights in at least one of the other group. The mean weight ranges from 2,073 grams in center 1 to 1,922 grams in center 5, with varying levels of missing data, potentially indicating variability in the mean weights at 36 weeks across multiple centers. For categorical ventilation support level at 36 weeks `ventilation_support_level.36` with three levels (0 means no respiratory support; 1 means non-invasive positive pressure; and 2 means invasive positive pressure) shows a significant difference across different centers ($p < 0.001$) by the Pearson's Chi-squared test. Even though Chi-squared test is the most commonly used test for assessing difference in distribution of a categorical variable between two or more independent groups (`center` here), the Chi-squared approximation to the distribution of the test statistic relies on the counts being roughly normally distributed. Because many of cell sizes are very small being less than 5 observations, the approximation may be poor. Instead, we run the Chi-squared test with computation of the simulated p-values, and resulted in a strong association between `center` and `ventilation_support_level.36` variables. The mean fraction of inspired oxygen at 36 weeks `inspired_oxygen.36` also varies significantly ($p < 0.001$) across centers. Some centers reported higher proportion of missingness, such as 43% in center 12. The peak inspiratory pressure (cmH_2O) at 36 weeks shows significant difference in at least one center ($p < 0.001$) compared to others, and a still high percentage of missingness observed in center 12, as well as in positive and exploratory pressure (cmH_2O) at 36 weeks `peep_cm_h2o.36` variable. For medication for pulmonary hypertension at 36 weeks `med_ph.36`, we could reject the null hypothesis and conclude that there is a strong association between center and medication for pulmonary hypertension given a significantly small p-value. Lastly, the mean hospital discharge gestational age `hosp_dc_ga` appears to have significant differences across centers, with a 100% missing rate in center 4 and 98% of missingness in center 1. There is notable variability of missingness in these respiratory variables at 36 weeks across centers, for example, center 12 consisted with almost half missing values in variables related to inspiratory and exploratory pressure and inspired oxygen, and some centers did not record hospital discharge gestational age at all. These missing data rates are concerning, which could impact the reliability of the statistical analyses.

When considering interaction terms for a logistic regression model, we consider interaction between center and respiratory support variables since there might be center-specific variation or patient population differences that associates with the level of respiratory support provided.

Multiple Imputation

Multiple imputation is applied to address missing values by creating 5 complete datasets. This method is trying to handle with each missing entry by estimating multiple reliable values such as regression models, running analysis across those completed dataset, such as mean values over each imputed datasets, and finally aggregating all previous analyses results, such as taking average over those mean values, and analyzing how far they spread out in terms of standard deviations and confidence intervals to identify deviance between imputed missing values. We constructed train-test splitted data and fitted the imputation model on the train data and apply the model to impute the test data. Given 5 complete train datasets, we will run lasso and forward stepwise regression for variable selection and use combined 5 test data as a validation dataset to assess performance of each model.

Selection of Variables for the Model

Table 3: Coefficients of Lasso Regression Models

	Lasso1	Lasso2
(Intercept)	-6.067	-6.067
center2	-0.651	-0.651
center3	-1.353	-1.353
center12	1.006	1.006
center16	-0.074	0.000
mat_ethn2	0.260	0.260
birth_hc	0.042	0.042
prenat_sterYes	0.448	0.448
com_prenat_sterYes	0.006	0.000
mat_chorioYes	0.000	0.000
sgaSGA	0.143	0.143
any_surfYes	0.187	0.000
weight_today.36	0.000	0.000
ventilation_support_level.362	1.232	1.232
inspired_oxygen.36	1.765	1.765
p_delta.36	0.000	0.000
peep_cm_h2o_modified.36	0.006	0.000
med_ph.361	0.039	0.000
hosp_dc_ga	0.053	0.053
center2:ventilation_support_level.361	-0.010	0.000
center3:ventilation_support_level.361	-0.246	-0.246
center5:ventilation_support_level.361	-0.415	-0.415
center12:ventilation_support_level.361	-0.003	0.000
center7:inspired_oxygen.36	-0.090	0.000
center2:p_delta.36	-0.001	0.000
center4:p_delta.36	-0.007	0.000
center12:p_delta.36	0.012	0.000
center2:peep_cm_h2o_modified.36	-0.031	-0.031
center4:med_ph.361	0.954	0.954
center12:med_ph.361	-0.336	-0.336

Table 4: Coefficients of Forward Stepwise Selection Models

	Forward.1	Forward.2	Forward.3	Forward.4	Forward.5	Avg_coef
(Intercept)	-2.544	1.511	-1.047	-1.283	-1.659	-1.004
center2	-2.360	-2.665	-2.531	-2.203	-2.635	-2.479
center3	-7.702	-6.875	-8.189	-6.186	-6.958	-7.182
center4	-1.672	-0.673	-0.893	-1.171	-0.948	-1.071
center5	-1.839	-2.475	-2.200	-1.577	-2.068	-2.032
center7	0.000	-2.171	-1.906	-1.458	-2.068	-1.521
center12	0.000	0.000	0.000	0.000	0.000	0.000
center16	-1.797	-2.564	-1.864	-1.897	-2.219	-2.068
mat_ethn2	0.979	0.934	0.000	1.070	0.795	0.756
bw	0.002	0.003	0.003	0.003	0.003	0.003
ga	-0.152	-0.203	-0.189	0.000	0.000	-0.109
blength	-0.114	-0.158	-0.123	-0.137	-0.176	-0.142
birth_hc	0.176	0.121	0.173	0.000	0.000	0.094
del_method2	0.000	0.000	0.000	0.000	0.000	0.000
prenat_sterYes	1.151	1.262	1.406	1.071	1.519	1.282
com_prenat_sterYes	0.000	0.000	0.000	0.000	0.000	0.000
mat_chorioYes	0.000	0.000	0.000	0.000	0.000	0.000
genderMale	0.000	0.000	0.000	0.000	0.000	0.000
sgaSGA	0.000	0.578	0.000	0.000	0.000	0.116
any_surfYes	1.189	0.000	1.062	0.000	0.000	0.450
weight_today.36	-0.002	-0.001	-0.002	-0.001	-0.001	-0.001
ventilation_support_level.361	0.000	-1.279	0.000	0.000	0.000	-0.256
ventilation_support_level.362	1.815	0.000	1.593	1.882	1.246	1.307
inspired_oxygen.36	0.000	2.754	1.805	0.000	2.217	1.355
p_delta.36	0.000	0.000	0.000	0.000	0.000	0.000
peep_cm_h2o_modified.36	0.000	0.136	0.000	0.000	0.000	0.027
med_ph.361	0.656	0.000	0.000	0.000	0.000	0.131
hosp_dc_ga	0.071	0.058	0.074	0.064	0.073	0.068

After multiple imputation method, we have 5 imputed datasets for which we do cross-validation in each imputed training dataset. Cross-validation helps with overfitting issues and with generalizability of the lasso model. Then since we will have for each fold in cross-validation results for different lambdas, we choose the lambda that has the lowest model error. This means that we will have five sets of estimated coefficients (one for each imputed data set). We considered two possible solutions to generate final lasso model: The first method **Lasso 1** is simply averaging over all 5 sets of coefficients to obtain the final lasso model. The second method **Lasso 2** is counting the number of times each variable being selected, if more than or equal to 3 times the variable was not selected, we will force to remove the variable from the final lasso model. In other words, if some variables occasionally were not selected, we wanted to neglect occasional situations and averaging over all estimated coefficients from 5 imputation datasets. In order to include transformations or interactions in the model, it is needed to have an idea of which interactions or transformations need to be included based on explanatory analysis or prior professional knowledge before applying the cross-validation to the imputed datasets. So first, after we have the imputed data, we specified and placed interaction terms between center and respiratory parameters for predictors of interest, then, we fit the lasso model with the interactions that are of interest and evaluate their significance by running summary of model after cross-validation. Table 3 provides us with a list of final estimated coefficients with interaction terms under lasso models of 10-fold cross-validation on the train imputation datasets. Interaction terms between center and respiratory variables show strong association which impact the outcome of adverse outcome, such as center

with medication for pulmonary hypertension at 35 weeks, and center with ventilation support level at 36 weeks. Thus, we could identify center-specific variation that associates with the level of respiratory support provided. In addition, main effects of `center` also perform great differences across centers on the outcome of babies healthy status.

Except for lasso regression models, to preserve generalizability and overfitting, we planned to remove all interaction terms between center and other respiratory variables, instead, we only included main effects of covariates and perform forward stepwise selection with cross-validation for each imputed data set. Starting with the empty model and sequentially adding predictors to the model, one at a time, and choosing the best predictor at each step based on a criterion like AIC and BIC, but it can be seen as a "locally optimal", instead of globally optimal in the sense of the best subset. Table 4 provides coefficients for each imputed data set and averaging over 5 coefficients together to obtain the final forward stepwise model. To compare the estimated coefficients using different variable selection or shrinkage methods we found that intercept values vary a lot across different methods, with lasso method having the most negative value (-6.067), which means the log odds of adverse outcome while all other variables are held at zero. For variables with non-zero coefficients, a positive coefficient indicates a positive association between the variable and the odds of outcome. In forward selection model, `pre_nat_sterYes` has a positive average coefficient (1.282), indicating that having a prenatal corticosteroids are 1.282 times the odds of adverse outcome compared to those who did not have prenatal corticosteroids. Conversely, center 4 (-7.182) is negatively associated with the odds of adverse outcome, in other words, center 4 is least likely to have adverse outcome compared to other centers adjusting for remaining variables. For those variables with zero coefficients all 5 times represents the least importance to be include in predicting adverse outcome of patients.

Internal Validation

We validated three models by assessing their performance on internal validation dataset obtained from multiple imputation method. Then we plot ROC curve and use AUC score to evaluate model fitting. According to scores and graphical means, we can look at how well our model differentiates between the two classes. The plot of ROC curve depicts classification model with a bigger area under the ROC curve explains and predicts outcome of interest better. By comparison, the forward stepwise selection model on the test data (AUC = 91.6%) explains the outcome of adverse outcome is better than two lass methods with consideration of interaction terms (AUC = 91.2% for second lasso model, and AUC = 91.1% for first lasso model). All three models have achieved an AUC greater than 0.8, which is a clear sign of good model performance. The AUC values are in the range of 0.911 to 0.916, suggesting that these models can be considered highly effective in distinguishing between the positive and negative classes.

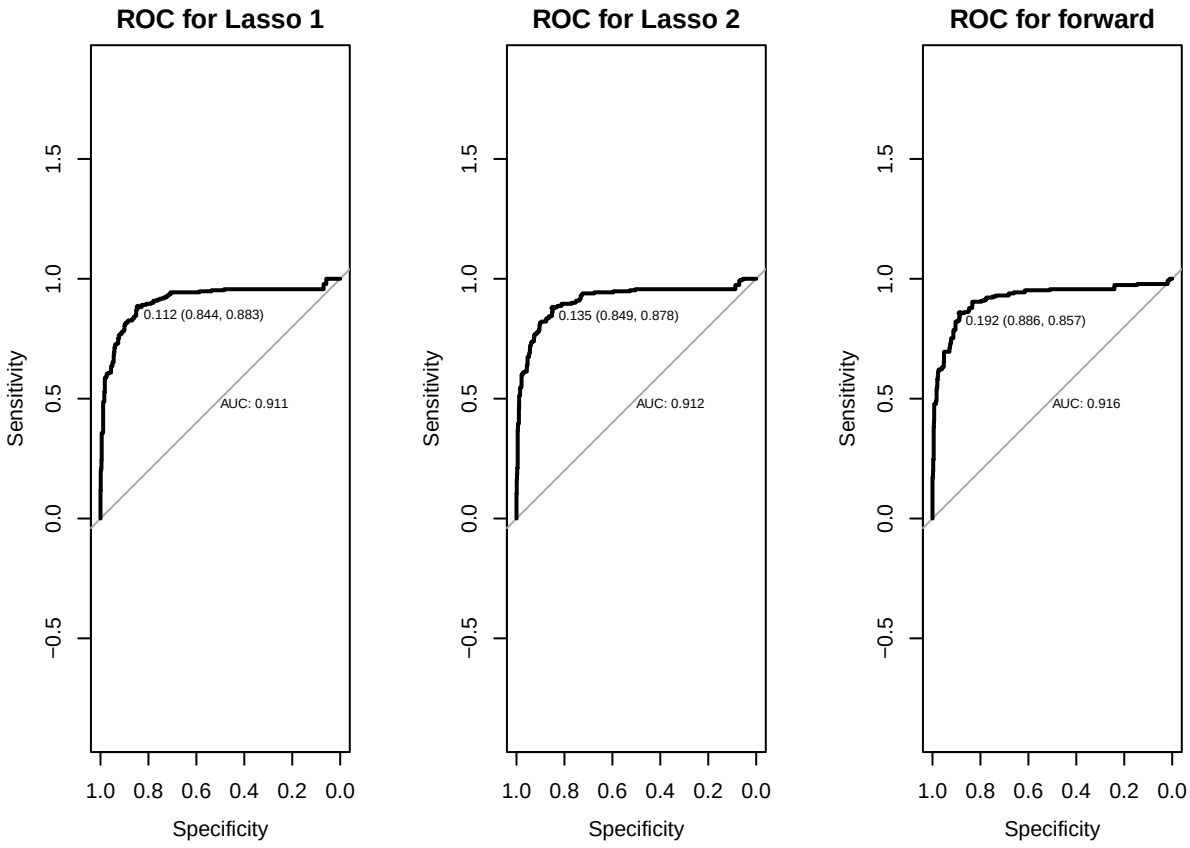


Table 5: Confusion Matix of Test Data under Lasso 1 Estimation

	Pred:0	Pred:1
Actual:0	989	11
Actual:1	119	111

Table 6: Confusion Matix of Test Data under Lasso 2 Estimation

	Pred:0	Pred:1
Actual:0	979	21
Actual:1	104	126

Table 7: Confusion Matix of Test Data under Forward Stepwise Estimation

	Pred:0	Pred:1
Actual:0	975	25
Actual:1	89	141

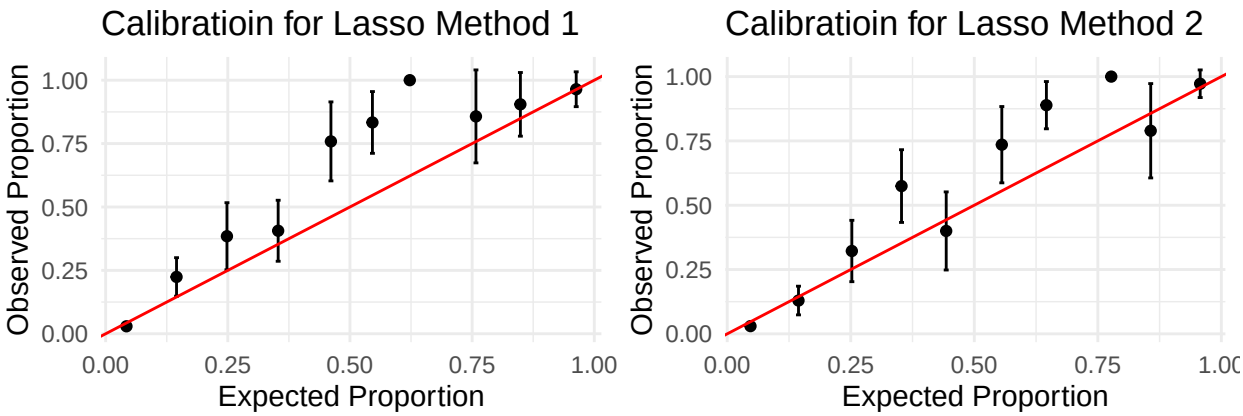
Each row in a confusion matrix represents an actual target, while each column represents a predicted target. Under the first lasso method (averaging over all coefficients) the first row of the matrix evaluated on the test data considers healthy patients (the False class): 989 were correctly classified as no adverse outcome (True

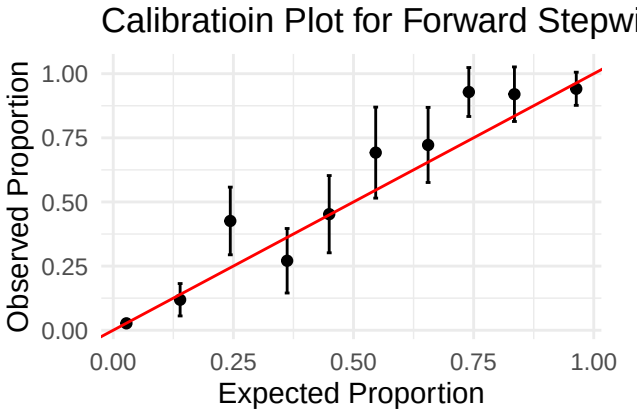
negative), while the remaining 5 was wrongly classified as adverse outcome (False positive). The second row considers unhealthy status, 119 patients are correctly classified in the positive prediction class(True positive), while the False positive was 111. Compared among those confusion matrices under different variable selection methods, we realized that forward stepwise selection made the highest number of true positive prediction (141 versus 126 or 111) on the test data set.

Sensitivity, true positive rate, indicates the percentage of individuals the model correctly predicted patients who had adverse outcome, which is the highest done by forward stepwise selection model. On the contrary, specificity, true negative rate, represents the percentage of individuals the model correctly predicted would not have adverse outcome. And the accuracy-test from the confusion matrix on the test dataset is calculated and is found to be 0.9073 in forward stepwise method, which slighlty outweighs the prediction performance made by lasso method (Accuracy are 0.8983 and 0.8943, respectively).

Table 8: Measurement of Confusion Matix of Test Data

	Sensitivity	Specificity	PositivePredicted	NegativePredicted	Accuracy
Lasso 1	0.4826087	0.989	0.989	0.989	0.8943089
Lasso 2	0.5478261	0.979	0.979	0.979	0.8983740
Forward	0.6130435	0.975	0.975	0.975	0.9073171





Another useful plot is a calibration plot. This type of plot groups the data by the estimated probabilities and compares the mean probability with the observed proportion of observations in class 1. It visualizes how close our estimated distribution and true distribution are to each other. The following three calibration plots all show partially close alignment with a 45-degree line, indicating good calibration. Forward stepwise selection model seems to have the closest alignment with the 45-degree line.

In summary, all three models perform well with respect to accuracy and well-calibration on the internal validation data set. The forward stepwise model shows slightly better performance in both discrimination and calibration compared to the lasso model.

Discussion

At this time our findings have some limitations. First, multiple imputation method can calculate much more unbiased estimates compared to single imputation, however, this technique cannot perform well in case of missing not at random (MNAR) and limitations of sample size. Thus, our multiple imputation may not be highly reliable since possible violation of assumptions. Secondly, to avoid overfitting and generalizability, we should be aware of many potential interaction terms between center and other variables. The reason why lasso methods did not perform as top as forward stepwise models is highly related to overfitting caused by inclusion of many interaction terms in the lasso method. Other than that, forward stepwise method has its own limitation: it can be only seen as a "locally optimal", instead of globally optimal in the sense of the best subset. Thus, further improvement analysis should add best subset method and assess predictive accuracy among these three variable selection methods.