Association between Surgery Time and Outcome

Xiaohan Yang

11/22/2020

Summary

This project is primarily interested in exploring if surgery time factors (ie. surgery hour, day of the week, month, etc.) have an association with mortality within 30 days after surgery. A subordinate interest of this project is to explore if there are other factors associated with 30-day mortality. The original dataset contains 32,001 observations and 25 rows. The methodology is logistic regression. The conclusion is that the surgery scheduled month and hour of a day have an association with surgery outcome.

Introduction

It has been proven by many studies that working performance is largely affected by human factors, such as fatigue, scheduling, or the inadequacy of sleep. People tend to be exhausted by shift work and prolonged duty and hence become more error-prone. The scheduled worktime plays an important role in working effectiveness. There is no reason to assume that hospital personnel is immune to such performance degradation. This project investigates if the time factors of general surgery (increasing hour of a day, increasing day of a week, middle of the year versus other months, and moon phase) are independently related to 30-day mortality using the logistic regression model. Meanwhile, this project also explores if other factors (patient's demographics, baseline diseases, surgical risk indices) influence 30-day mortality.

Data

This dataset comes from a 2011 study and contains 32,001 observations and 25 columns. Each observation represents a general surgery patient at the Cleveland Clinic between January 2005 and September 2010. This dataset consists only of general surgeries scheduled at routine workdays and workweeks and does not include emergency surgeries.

The response variable is the binary indicator of 30-day mortality. The other 24 predictors can be grouped into 4 categories: 1) surgical time predictors (hour, day of week, month, and moon phase). 2) patients demographic predictors (age, gender, race, BMI). 3) binary indicators of 8 baseline diseases. 4) surgical risk indices (asa_status, baseline_charlson, complication_rsi, etc.). Amongst all predictors, age, BMI, hour, mortality_rsi, and complication_rsi are continuous numerical variables. All the rest of the variables are categorical.

It is worth noticing that this dataset has missing values in 4 columns. Column gender has 3 missing cells; race has 480; asa_status has 8; BMI has 3290. There are about 10% of the observations have missing values, so the issue is not too serious. I applied the multiple imputation strategy (mice, set m=10) on the entire dataset, and randomly selected one of the completed datasets, and ran the diagnostics (Appendix A). The imputation is reasonable and is of high quality. I will use the imputed dataset for the rest of the analysis.

EDA

To begin with, I checked the distribution for all numerical variables. Complication_rsi, mortality_rsi, and hour are not normally distributed (Appendix B). Complication_rsi and mortality_rsi are risk indices for in-hospital complication and death and have values less than 0. It is difficult to do log or square transformation on such variables, so I categorized them using 0 as a threshold. The other variable hour represents the hour of a day, so it does not make much sense to make it numerical. I also categorized it into 2 levels: “AM” and “PM”, using 12 pm as a threshold.

After binning the three variables just mentioned, the remaining numerical variables do not show a correlation with 30-day mortality in boxplots. Moving on to EDA for categorical predictors, I first explore the distributions of surgical time predictors because they are the key interest of this project. The conditional probability for 30-day death is slightly different across months (Appendix D). I suspect that there is a relationship between the two. Yet the p-value in chi-squared test for month is not significant, which says these two variables are independent. Despite the result of the chi-squared test, we should put the variable month in our model later, because it is one of the factors we are most interested in.

As for other categorical predictors, gender, asa_status (American Society of Anesthesiologist Physical Status), baseline_charlson (Charlson Comorbidity Index) also demonstrated strong correlations with 30-day mortality, and they all have significant p-values in chi-squared test. Additionally, whether a patient has baseline cancer, cardiovascular or pulmonary diseases also seem to be correlated to 30-day death, which makes sense because these diseases are relatively more fatal.

Multicolinearity

Then I explored variable multicollinearity. Inferring from domain knowledge, I think there might be a multicollinearity issue between age and baseline diseases, since some diseases are more likely to occur among senior people. I investigated the multicollinearity between age and most other predictors and found the distribution of age vs baseline_cvd are different between the dead group and survived group (Appendix E). We may want to explore this interaction more, later in the model.

To sum up, the EDA suggests that gender, mortality_rsi, complication_rsi, month, asa_status, baseline_charlson, baseline_cancer, baseline_cvd, baseline_pulmonary should have relationships with 30-day mortality. The interaction between age and baseline_cvd should also be worth exploring. We should consider include these variables in the model later.

Model

I first decided to use a stepwise selection process to find a preliminary model. The null model only includes the null predictor 1, and the full model includes all predictors plus interaction terms age:baseline_cvd. Since there are a lot of variables in the original dataset, I choose AIC instead of BIC because it tends to keep more variables.

The AIC model selects 6 predictors - gender, hour, baseline_osteoart, and 3 surgical risk indices, yet excludes the interaction terms (See Appendix F). The AIC model is mostly consistent with the findings in EDA, so I decided to construct a final model based on the AIC model.

The AIC model does not include month, baseline_cancer, baseline_cvd, baseline_pulmonary, and age:baseline_cvd, which we found worth exploring in the EDA. I added these predictors into the model one at a time and applied F-test to test the significance. It turns out that only adding month improves the AIC model, so I decided to include month in the model.

For the surgical time predictors, the AIC model excludes day of week and moon phase. I tested if adding these two predictors would improve the model. However, the F-test shows adding day of week or moon phase would not make much difference to the model.

Likewise, the interaction term we found interesting in EDA does not improve the model. I also tried 6 other potential interaction terms, but none of them significantly improves the model, which means that my final model would not include any interaction terms.

As the result, the final model has 7 predictors, and the model summary is as follow:

$$logit(\pi_i) = \beta_0 + \beta_1 x_{gender} + \beta_2 x_{asa\_status} + \beta_3 x_{baseline\_osteoart} + \beta_4 x_{baseline\_charlson} + \beta_5 x_{comprsi\_group} + \beta_6 x_{month} + \beta_7 x_{hour\_group}$$

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-8.392	0.4534	-18.51	1.806e-76
genderF	-0.401	0.1789	-2.241	0.025
asa_status2	2.065	0.323	6.392	1.64e-10
asa_status3	3.616	0.3515	10.29	8.1e-25
baseline_osteoartYes	0.1995	0.283	0.7049	0.4809
baseline_charlson1	0.7137	0.3007	2.373	0.01763
baseline_charlson2	0.5569	0.2767	2.013	0.04413
baseline_charlson3	0.2642	0.3899	0.6776	0.498
baseline_charlson4	0.4003	0.4487	0.8921	0.3723
baseline_charlson5	0.2859	0.6393	0.4472	0.6548
baseline_charlson6	1.339	0.439	3.049	0.002295
baseline_charlson7	1.029	0.7773	1.323	0.1857
baseline_charlson8	1.207	0.3346	3.607	0.0003099
baseline_charlson9	1.113	0.4563	2.438	0.01475
baseline_charlson10	1.667	0.5784	2.882	0.003949
baseline_charlson11	2.019	0.8341	2.42	0.01552
baseline_charlson12	-11.65	574.4	-0.02029	0.9838
baseline_charlson13	-11.67	996.1	-0.01171	0.9907
comprsi_group1	2.16	0.2585	8.355	6.552e-17
month2	-0.1816	0.3477	-0.5223	0.6015
month3	-0.5622	0.3791	-1.483	0.1382
month4	-0.4729	0.3781	-1.251	0.211
month5	-0.7167	0.4011	-1.787	0.07397
month6	-0.4562	0.3624	-1.259	0.2081
month7	-0.33	0.3792	-0.8705	0.3841
month8	-1.086	0.4142	-2.622	0.008743
month9	-0.3929	0.3486	-1.127	0.2597
month10	-0.9503	0.4319	-2.2	0.0278
month11	-1.25	0.5106	-2.448	0.01437
month12	-1.231	0.5603	-2.197	0.02802
hour_groupPM	0.3622	0.177	2.046	0.04075

(Dispersion parameter for binomial family taken to be 1)

Null deviance:	1779 on 32000 degrees of freedom
Residual deviance:	1369 on 31970 degrees of freedom

Model Interpretation

Month: Holding everything else constant, comparing to surgical patients in January, a patient who is operated in August is 66% less likely to die in 30 days; a patient operated in October is 61% less likely to die; a patient in November is 71% less likely to die; patient in December is also 71% less likely to die.

Hour of day (binned): Holding everything else constant, comparing to surgical patients operated in the morning, patients operated in the afternoon are 43% more likely to die.

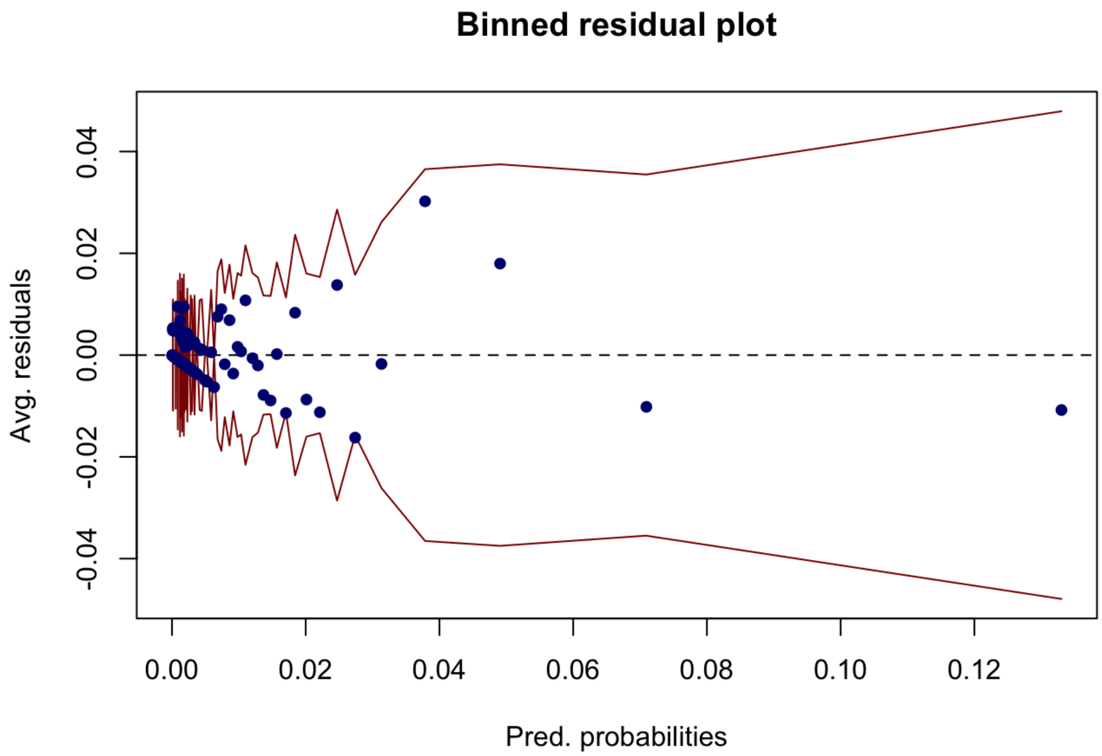
Gender: Holding everything else constant, comparing to male patients, female patients are 33% less likely to die within 30 days of the surgery.

Baseline_osteoart: Holding everything else constant, comparing to patients without osteoarthritis as their baseline disease, patients with osteoarthritis are 22% more likely to die within 30 days of the surgery.

Besides these predictors, three surgical risk indices are also significant, including asa_status (American Society of Anesthesiologist Physical Status), baseline_charlson (Charlson Comorbidity Index), comprsi_group1 (binned in-hospital complication risk). Generally speaking, the higher level of risk a patient has, the more likely he/she dies within 30 days of the surgery.

Model Diagnostics

For model diagnostics, I first drew a binned residual plot. It looks mostly good. Most of the points fall within the 95% bin, and are roughly random, which says that the observations are independent.



VIF scores were also checked for the final model (Appendix G). All variables have VIF scores around 1, so there is no multicollinearity issue.

Moving on to the confusion matrix, I used the mean of 30-day mortality as the threshold. The sensitivity and specificity scores are both high. The model also has a decent accuracy score. The model does well on both fitting the data and predicting.

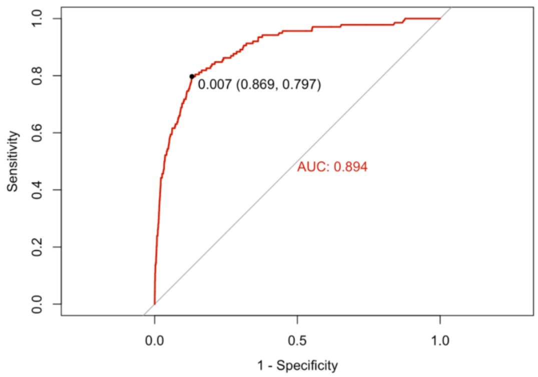
Confusion Matrix

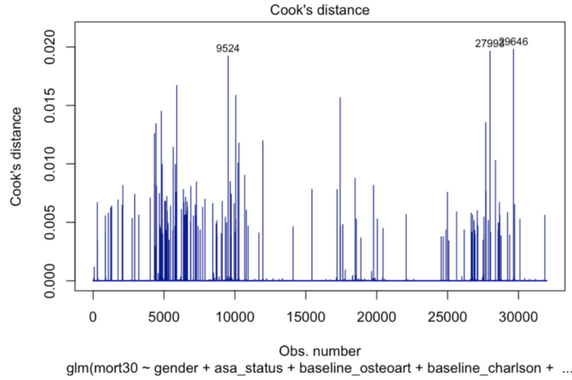
Predicted	Observed	
	No	Yes
No	26757	26
Yes	5106	112

Confusion Matrix Statistics

Accuracy	Sensitivity	Specificity
0.84	0.81	0.84

Moreover, the ROC plot is shown below. The AUC score is 0.89, which is pretty close to 1.





I then checked outliers. The largest cook's distance value is around 0.02 – still inside the 95% bin. The binned residual plot also confirms that there is no serious outlier issue. Hence, I decide not to remove any points.

In a word, all the model diagnostics show that the final model fit the data well and does a decent job in prediction. There is also no serious outlier issue.

Conclusion

There is enough evidence that 30-day mortality is associated with the month a surgery takes place, and with whether the surgery is scheduled in AM or PM. General surgical patients have a significantly less likelihood of 30-day death in August, October, November, and December. Patients have more likelihood of 30-day death if they are operated in the afternoon. From this result, we can infer that hospital personnel might be less effective in the afternoon.

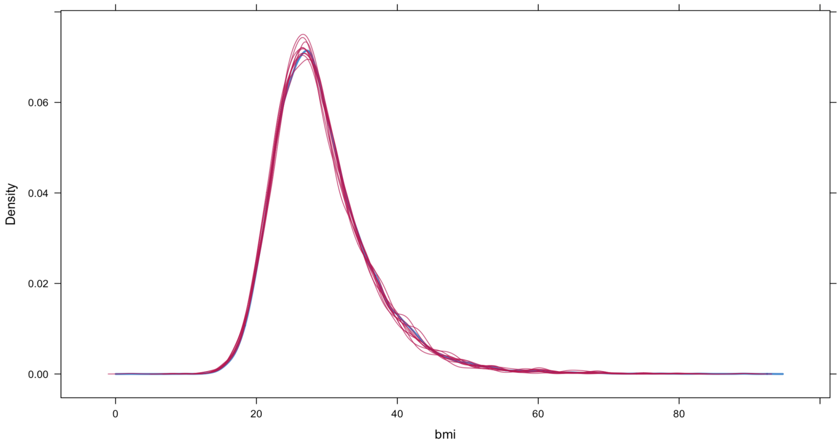
On the other hand, no association was found for day of week or moon phase and 30-day mortality. These two factors are not statistically significant at a 0.1 level.

Other factors that associate with mortality in 30 days include gender, baseline osteoarthritis, and surgical risks.

Lastly, there is a potential limitation of this analysis. The observations in this dataset are selected only from scheduled general surgeries that happen on routine workday and workweek. The dataset does not include emergency surgeries, which are more likely to happen late at night or on weekends. The latest operation hour in this dataset is only 7 pm. During late nights and weekends, it is likely the hospital staff are more easily affected by fatigue. Since the dataset does not include all types of surgeries in the hospital, our conclusion might be very biased.

Appendix

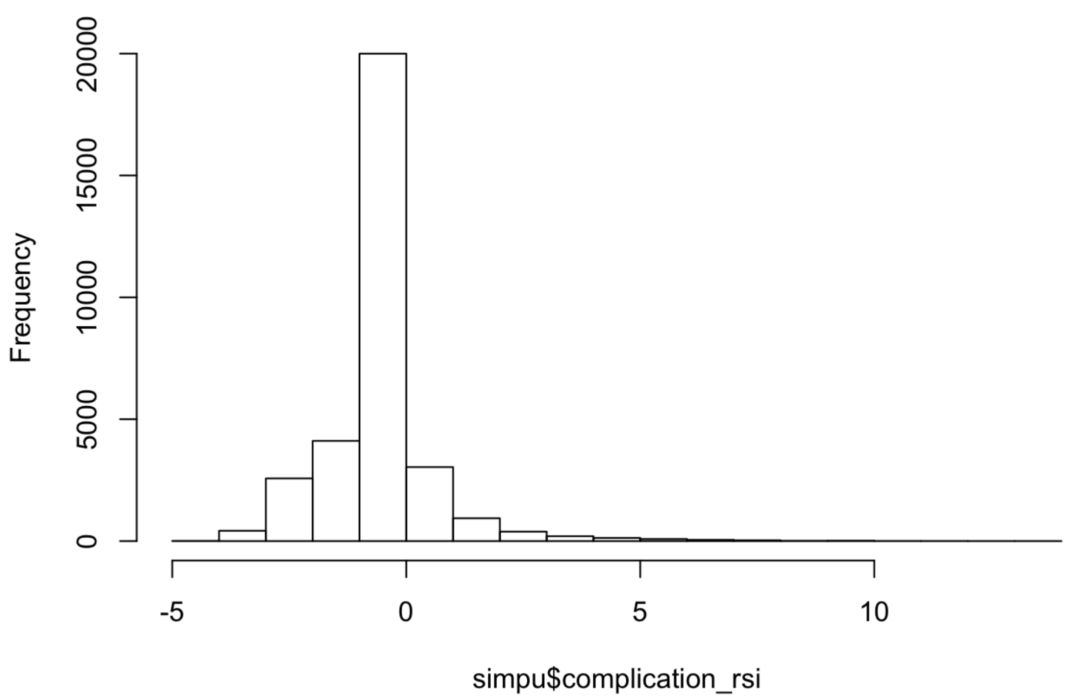
Appendix A: Imputation results



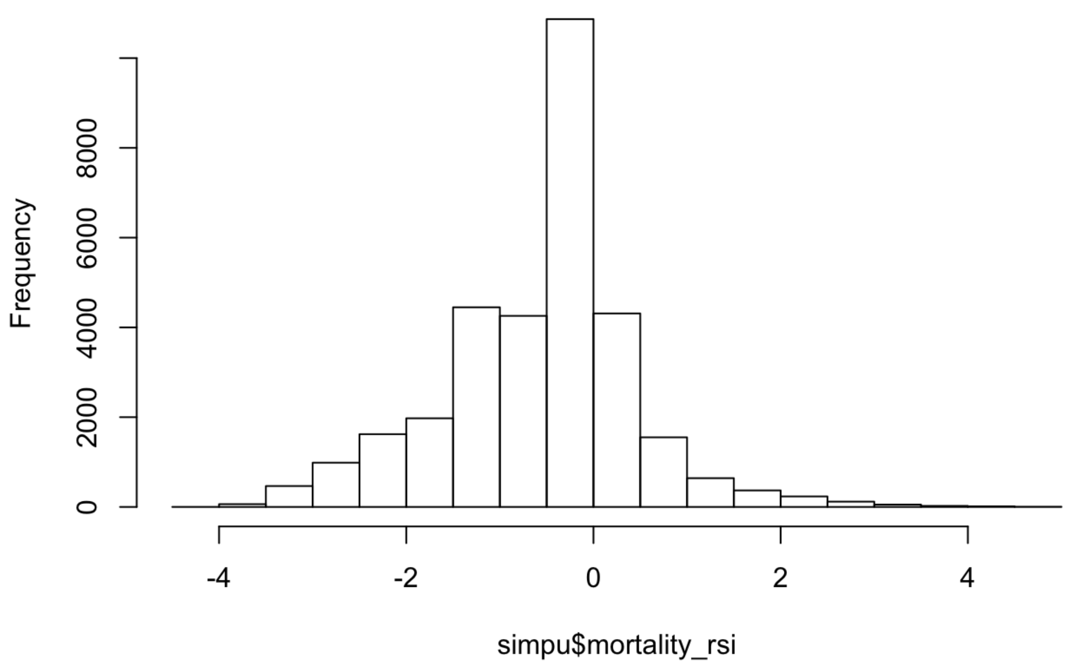
appenA

Appendix B: Distribution of complication_rsi and hour

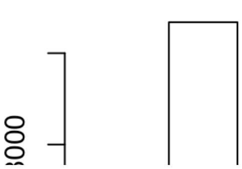
Histogram of simpu$complication_rsi

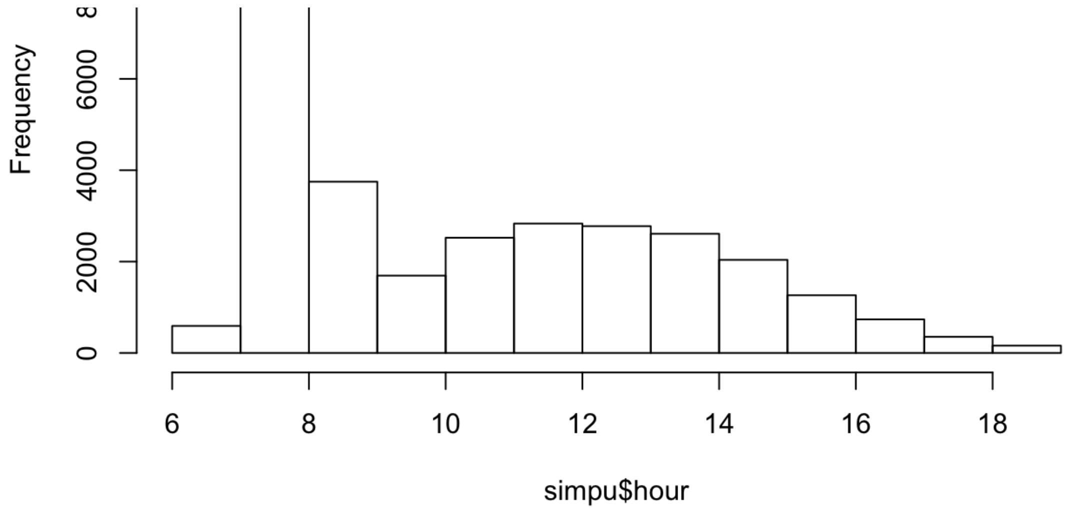


Histogram of simpu$mortality_rsi

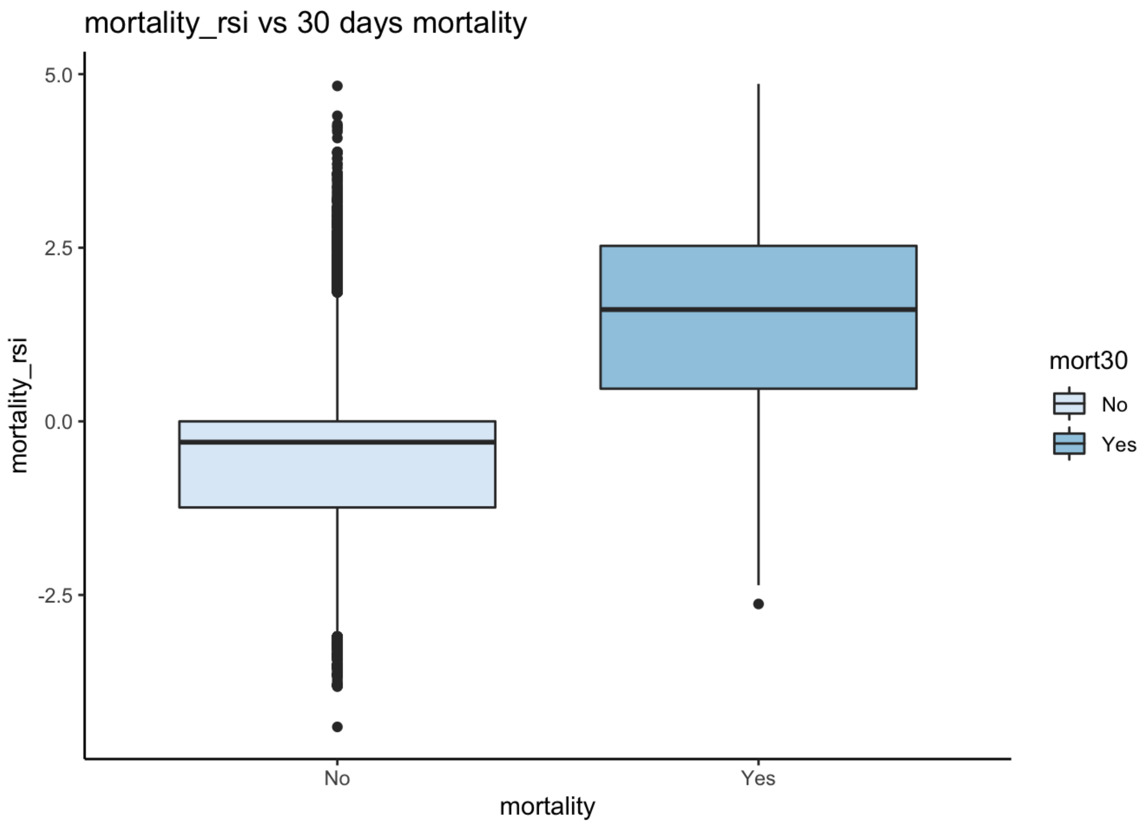


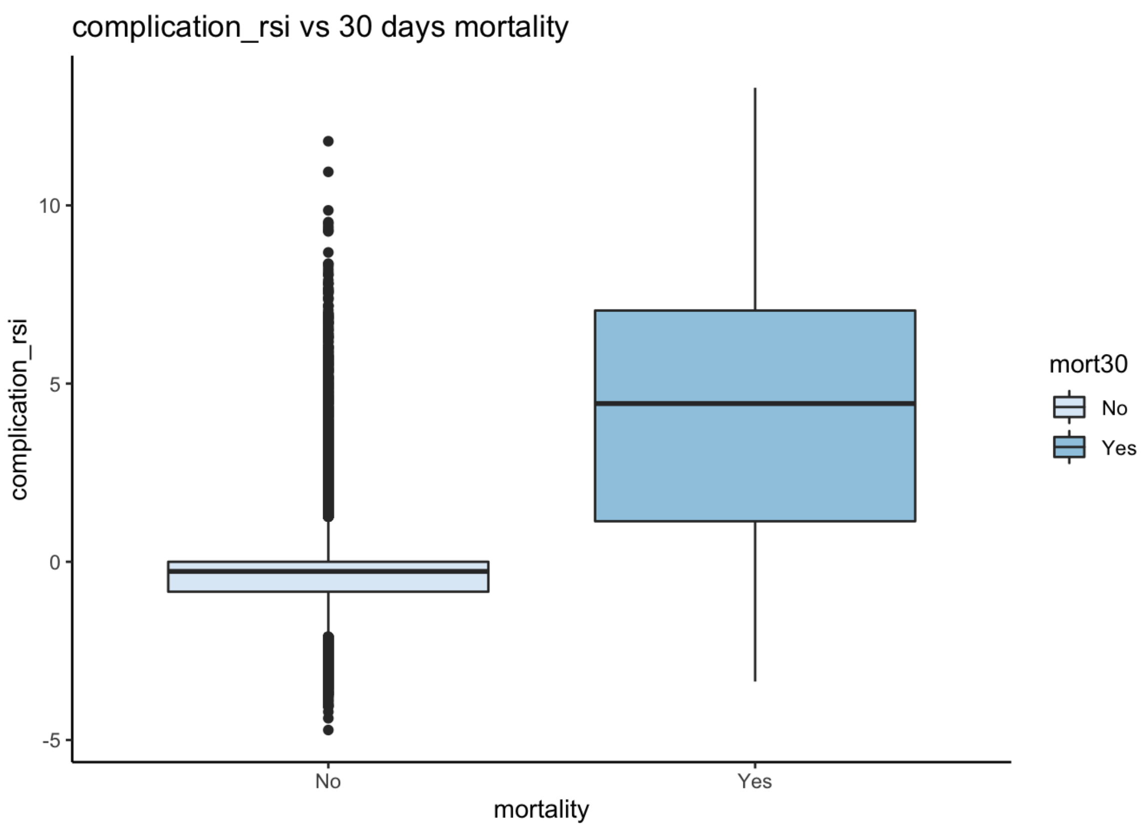
Histogram of simpu$hour





Appendix C: Boxplots of mortality_rsi, complication_rsi vs 30-day mortality





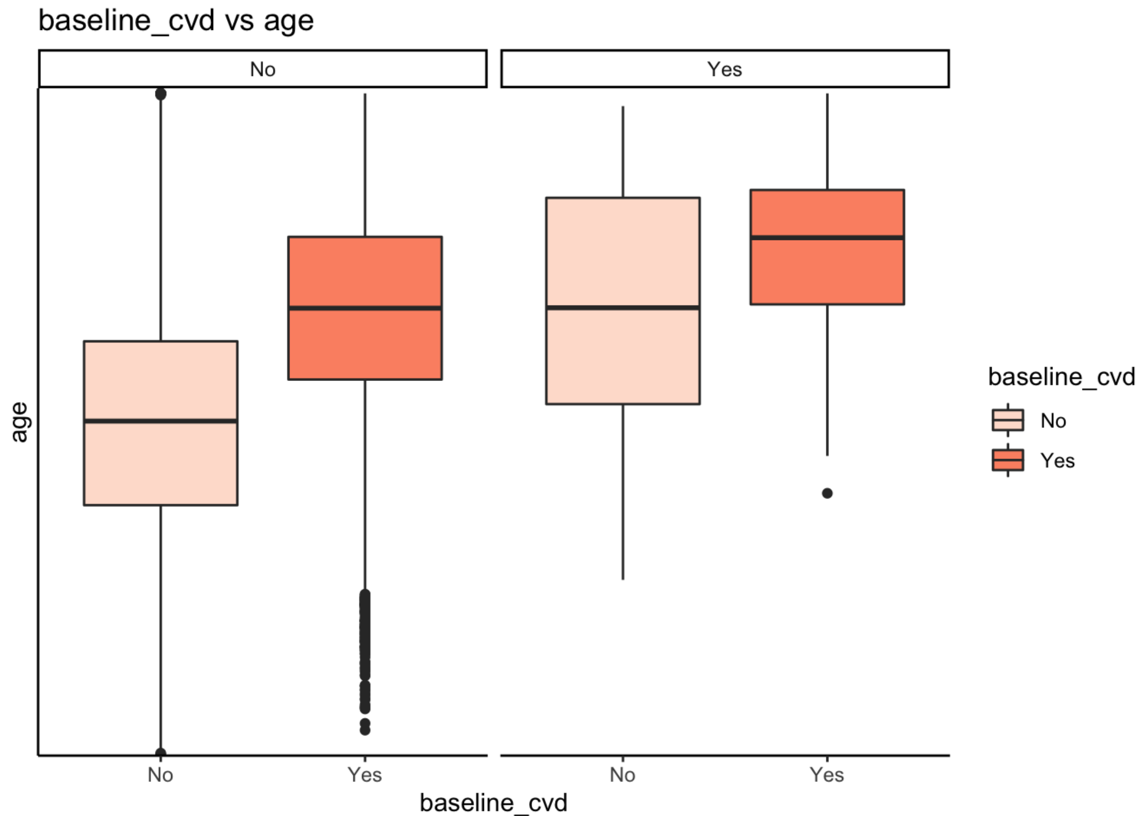
Appendix D: Conditional Probability Across Month

Table continues below

	1	2	3	4	5	6
No	0.9929	0.9932	0.9956	0.9956	0.9962	0.9953
Yes	0.007116	0.006784	0.004449	0.004448	0.003768	0.004676

	7	8	9	10	11	12
No	0.9948	0.9972	0.995	0.997	0.998	0.9978
Yes	0.005161	0.002833	0.004988	0.002975	0.001965	0.002175

Appendix E: Interaction between age and basline_cvd



Appendix F: Summary of AIC Model

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-8.603	0.5037	-17.08	2.078e-65
complication_rsi	0.6888	0.03286	20.96	1.399e-97
asa_status2	1.149	0.3386	3.394	0.0006892
asa_status3	2.06	0.3771	5.464	4.663e-08
baseline_charlson1	0.9057	0.3364	2.692	0.007102
baseline_charlson2	0.8706	0.3165	2.751	0.005946
baseline_charlson3	0.7803	0.4301	1.814	0.06962
baseline_charlson4	0.9661	0.4865	1.986	0.04705
baseline_charlson5	1.32	0.6664	1.981	0.04765
baseline_charlson6	1.39	0.5425	2.562	0.01042
baseline_charlson7	1.435	0.8145	1.762	0.07801
baseline_charlson8	1.695	0.3617	4.686	2.791e-06
baseline_charlson9	1.344	0.5014	2.681	0.007345
baseline_charlson10	2.055	0.6359	3.231	0.001232

	Estimate	Std. Error	z value	Pr(>|z|)
baseline_charlson11	2.709	0.9016	3.005	0.002658
baseline_charlson12	-10.78	568.8	-0.01895	0.9849
baseline_charlson13	-10.81	1007	-0.01073	0.9914
baseline_osteoartYes	0.8119	0.3107	2.613	0.008977
genderM	0.3791	0.2011	1.885	0.05947
hour	0.04529	0.03137	1.444	0.1488

(Dispersion parameter for binomial family taken to be 1)

Null deviance:	1779 on 32000 degrees of freedom
Residual deviance:	1040 on 31981 degrees of freedom

Appendix G: VIF for Final Model

Table continues below

genderM	asa_status2	asa_status3	baseline_osteoartYes
1.026	3.347	3.59	1.097

Table continues below

baseline_charlson1	baseline_charlson2	baseline_charlson3
1.669	1.735	1.358

Table continues below

baseline_charlson4	baseline_charlson5	baseline_charlson6
1.276	1.143	1.302

Table continues below

baseline_charlson7	baseline_charlson8	baseline_charlson9
1.088	1.5	1.266

Table continues below

baseline_charlson10	baseline_charlson11	baseline_charlson12
1.155	1.078	1

Table continues below

baseline_charlson13	comprsi_group1	month2	month3	month4	month5
1	1.13	1.671	1.514	1.508	1.432

month6	month7	month8	month9	month10	month11	month12	hour_groupPM
1.58	1.505	1.398	1.641	1.36	1.23	1.188	1.022