Temporal Patterns in Surgical Patient Outcomes

Mitchelle Mojekwu

2023-12-14

Introduction

Context and Background

Since the rise of the coronavirus disease (COVID-19), there has been a surge in the concern of a complication in healthcare industries worldwide. The shortage of healthcare professionals has been an increasingly discussed topic among citizens and healthcare providers alike as millions of people experienced the threat of a global pandemic. In general, the scarcity of medical industry professionals has led to a decline in the quality and availability of care in hospitals. (Džakula and Danko, 2022) It is widely known and studied that insufficient amounts of sleep and extended work shifts risk negative impacts on one's cognitive and motor performance. (Alhola and Päivi, 2007)Due to the risk of underperformance of healthcare workers, particularly surgeons, there has been a push to limit resident work hours to mitigate potential errors that may impact patients.

Given that the healthcare sector is one of high stakes that often demands medical professionals to work extensive shifts due to short staff and increased patient attendance, the risk of long work hours cannot be ignored. Furthermore, just as extended hours of work may lead to fatigue and a compromise in cognitive and physical tasks, typical workweek progression may lead to similar results. (Sessler et al., 2011) This study aims to address the association between multiple time-related factors and the resulting outcome of patients' operation cases.

These are the main objectives of this case study (Sessler et al., 2011):

1) Is there any association between whether or not a patient's operation results in 30-day mortality and the hour in which the operation occurs? How about the day of the week?

2) Is there any association between whether or not a patient experiences in-hospital complications and the hour in which the operation occurs? How about the day of the day of the week?

Answering these objectives will help healthcare industry leaders better understand the possible risks of further progression in time from morning to evening and Monday to Friday of surgical labor on patient operation outcomes.

The Data

The data for this study was collected and provided by the Cleveland Clinic Lerner Research Institute. The data is formatted as patient-level observations that comprise of 32,001 surgical patients from the Cleveland Clinic between January 2005 and September 2010. The data collection window consisted of the hours of 6 am to 7 pm during the work weekdays. The surgical procedures were limited to elective operations since urgent or semi-urgent procedures are often performed later in the workday even without being officially labeled as emergencies. The variables in the dataset are the patient's age, gender, race, ASA physical status, body mass index, United States Agency for Healthcare Research and Quality's Clinical Classifications Software (AHRQ-CCS) Procedure Category, Charlson Comorbidity Index, 30-Day Mortality Risk Stratification Index, In-Hospital Complications Risk Stratification Index, overall incidence of 30-day Mortality for each AHRQ-CCS procedure category, overall Incidence of In-hospital Complications for each AHRQ-CCS procedure category, operation hour, day of the week, month of the year, phase of the moon, whether or not the patient resulted in having 30-day mortality, whether or not the patient resulted in having in-hospital complications, and

whether or not the patient has the following diseases/disorders: cancer, cardiovascular/cerebrovascular disease, dementia, diabetes, digestive disease, osteoarthritis, psychiatric disorder, pulmonary disease. (some variable definitions in Appendix)

Exploratory Data Analysis

The initial stages of exploratory data analysis consisted of viewing the distribution of each of the patient descriptor potential predictors. The distribution of age, gender, race, ASA Physical status, and body mass index (Appendix: Images 1-5) assisted in identifying the best measure of center and spread for summary statistics.

Summary Statistics of Baseline Risk Factors for Patients

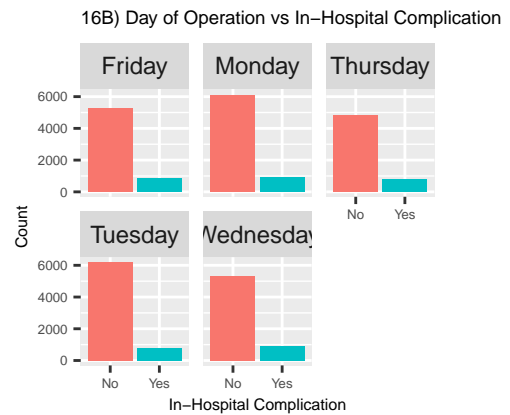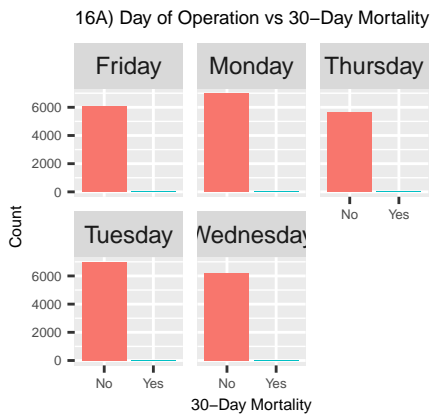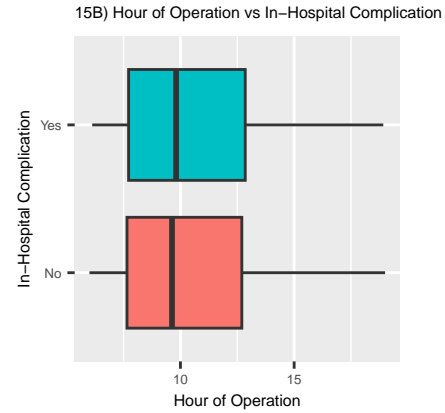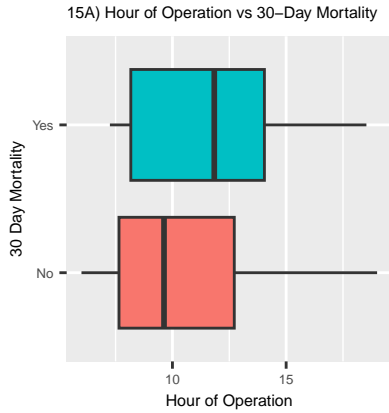| Predictors	Summary.Statistics
Age(mean/SD)	
Age	57.66 ± 15.04
Gender (count/%)	
Male	14768 (46.15 %)
Female	17230 (53.84 %)
Race (count/%)	
Caucasian	26488 (82.77 %)
African American	3790 (11.84 %)
Other	1243 (3.88 %)
ASA Physical Status (count/%)	
I-II	17261 (53.94 %)
III	13677 (42.74 %)
IV-VI	1055 (3.3 %)
BMI (median/IQR)	
Body Mass Index	28.19 [24.6 , 32.81]
Diseases/Disorders (count/%)	
Cancer	10958 (34.24 %)
Cardiovascular/cerebrovascular disease	16176 (50.55 %)
Dementia	242 (0.76 %)
Diabetes	4166 (13.02 %)
Digestive disease	7037 (21.99 %)
Osteoarthritis	5719 (17.87 %)
Psychiatric disorder	2910 (9.09 %)
Pulmonary disease	3493 (10.92 %)
Indexes (median/IQR)	
Charlson Comorbidity Index	0 [0 , 2]
Risk Stratification Index (30-day mortality)	-0.3 [-1.24 , 0]
Risk Stratification Index (in-hospital complications)	-0.27 [-0.84 , 0]

Based on the visualizations and the summary statistics above it is noteworthy that the age of the patients is normally distributed centered around 57.66 years old, there are more female patients, and the racial demographic of the patients is largely Caucasian. The patients tend to be centered more around the lower to middle class of ASA physical status meaning they are mostly in relatively good health. The BMI of the patients is normally distributed but slightly skewed towards higher BMI with a center at 28.19 which is relatively overweight. It is also worth noting that roughly half the patient population has a cardiovascular or cerebrovascular disease, about a third of the patients have cancer, and about a fifth of the patients have a digestive disease. The median Charlson Comorbidity Index of the patient population is roughly 0. This means no comorbidities were found and therefore generally there wasn't more than one disease simultaneously found in one patient. Overall, certain demographics of the patients are quite atypical in comparison to the general population.

Summary Statistics of Case Frequencies (Count/%)

Predictors	Summary.Statistics
Hour of Operation	
6:00	562 (1.76 %)
7:00	10631 (33.22 %)
8:00	3807 (11.9 %)
9:00	1664 (5.2 %)
10:00	2501 (7.82 %)
11:00	2855 (8.92 %)
12:00	2763 (8.63 %)
13:00	2623 (8.2 %)
14:00	2063 (6.45 %)
15:00	1267 (3.96 %)
16:00	745 (2.33 %)
17:00	356 (1.11 %)
18:00	163 (0.51 %)
19:00	1 (0 %)
Day of the Week	
Monday	7005 (21.89 %)
Tuesday	7008 (21.9 %)
Wednesday	6266 (19.58 %)
Thursday	5635 (17.61 %)
Friday	6087 (19.02 %)
Month	
January	2670 (8.34 %)
February	2506 (7.83 %)
March	2697 (8.43 %)
April	2698 (8.43 %)
May	2654 (8.29 %)
June	2994 (9.36 %)
July	2325 (7.27 %)
August	3177 (9.93 %)
September	3208 (10.02 %)
October	2689 (8.4 %)
November	2544 (7.95 %)
December	1839 (5.75 %)
Phase of Moon	
New Moon	7708 (24.09 %)
First Quarter	8100 (25.31 %)
Full Moon	8051 (25.16 %)
Last Quarter	8142 (25.44 %)

The table above as well as the Images 9-12 (Appendix) display the distribution of the patient's operations by measures of time. The distribution of operation cases by hour shows that a third of operation cases took place from 7-8 am and a steady flow of operation cases took place from 10 am to 2 pm. The distribution of operation cases over the days of the week, shows there are only slightly more cases at the beginning of the week. The distribution of cases over months has some differences with the highest concentrations being in August and September. Lastly, the distribution of operations over moon phases is quite uniform.

The last significant portion of EDA was visualizing the possible relationship between the time-related predictors and the response variables of whether or not a patient had in-hospital complications and 30-day mortality. Image 15A) shows that operations of patients that have 30-day mortality tend to happen slightly later in the day than operations of patients that do not have 30-day mortality. Image 15B) shows that operations of patients who have in-hospital complications tend to be relatively in the same range of time as operations of patients who do not have in-hospital complications. Image 16A) shows there is a noticeably decreasing number of patients without 30-day mortality as the week goes on. Image 16B) shows there is also a decreasing number of patients without in-hospital complications as the week goes on.

15A) Hour of Operation vs 30–Day Mortality

15B) Hour of Operation vs In–Hospital Complication

16A) Day of Operation vs 30–Day Mortality

16B) Day of Operation vs In–Hospital Complication

Methodology

Model Choice: Logistic Regression Model

To model the relationship between the measures of time (hour of the operation, day of the week) in which the surgical operations took place and the response variables of 30-day mortality and in-hospital complication, I decided to create two logistic regression models. The reason behind this choice is due to the response variables being two separate binary outcomes. I decided to utilize the variables that the research objective highlighted such as hour of operation and day of the week, but I also accounted for other variables. In both models, I added the patients' age, Body Mass Index, ASA Physical Status, and Charlson Comorbidity Index because these are predictors known to be associated with the general wellness and health complications of a person which would likely explain some variability in the response variables of 30-day mortality and in-hospital complication. In the logistic regression model with 30-day mortality as the response variable I accounted for the Risk Stratification Index (30-day mortality) to account for procedures and diagnoses the patient undergoes. This RSI allows the 30-day mortality outcome to be compared equally across institutions. This was also done in the logistic regression model with in-hospital complication as the response variable but instead, I accounted for the Risk Stratification Index (in-hospital complications). Additionally, in both models, I accounted for the overall incidence of the response variables for each AHRQ-CCS Procedure Category. This adjustment was implemented to guarantee thorough control for potential confounding factors linked to variations in the exposures of interest concerning the different procedure types of each patient. (Sessler, Daniel I., et al, 2011)

Model 1: 30-Day Mortality

$$log(\frac{\pi_i}{1-\pi_i}) = \beta_0 + \beta_1 \times RSIMortality_i + \beta_2 \times OverallIncidenceMortality_i + \beta_3 \times hour_i$$

$$+ \beta_4 \times (day_i = Monday) + \beta_5 \times (day_i = Thursday) + \beta_6 \times (day_i = Tuesday) + \beta_7 \times (day_i = Wednesday)$$

$$+ \beta_8 \times Age_i + \beta_9 \times BMI_i + \beta_{10} \times (ASAstatus_i = I - II) + \beta_{11} \times (ASAstatus_i = III)$$

$$+ \beta_{12} \times (ASAstatus_i = IV - VI) + \beta_{13} \times CCI_i + \epsilon_i$$

π = probability of a patient having 30-day mortality

where $Y_i|X_i\beta \sim Bernoulli(\pi_i)$ and $\epsilon_i \sim N(0,\sigma)$

Model 2: In-Hospital Complication

$$log(\frac{\pi_i}{1-\pi_i}) = \beta_0 + \beta_1 \times RSIComplication_i + \beta_2 \times OverallIncidenceComplication_i + \beta_3 \times hour_i$$

$$+ \beta_4 \times (day_i = Monday) + \beta_5 \times (day_i = Thursday) + \beta_6 \times (day_i = Tuesday) + \beta_7 \times (day_i = Wednesday)$$

$$+ \beta_8 \times Age_i + \beta_9 \times BMI_i + \beta_{10} \times (ASAstatus_i = I - II) + \beta_{11} \times (ASAstatus_i = III)$$

$$+ \beta_{12} \times (ASAstatus_i = IV - VI) + \beta_{13} \times CCI_i + \epsilon_i$$

π = probability of a patient having In-Hospital Complication

where $Y_i|X_i\beta \sim Bernoulli(\pi_i)$ and $\epsilon_i \sim N(0,\sigma)$

The coefficients are β_0 for the intercept, β_1 for the Risk Stratification Index (30-Day Mortality)/(In-Hospital Complication), β_2 for the Overall Incidence of 30-Day Mortality/In-Hospital Complication for Each AHRQ-CCS Procedure Category, β_3 for operation hour, β_4-β_7 for the days of the week with the baseline being Friday, β_8 for patients' age, β_9 for patients' body mass index, β_{10}-β_{12} for ASA Physical Status,β_{13} for Charlson Comorbidity Index, and ϵ_i for the residual error for the ith patient.

Model Assumptions and Diagnostics

The model assumptions needed to fit these models were linearity of the log odds, independence of the observations, absence of perfect multicollinearity, and the data being obtained from a random process. Image 17) shows there is linearity between the predictor hour of operation and the log odds of 30-day mortality. Image 18) shows there is linearity between the hour of operation and the log odds of in-hospital complications. Image 19) shows there is no -1/+1 correlation between the numerical predictors therefore no perfect multicollinearity is present.

Results

Table 1: 30-Day Mortality Model Coefficients/Summary Stats

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-17.708	544.451	-0.033	0.974
Risk Stratification Index (30-Day Mortality)	1.101	0.083	13.293	0.000
Overall Incidence of 30-Day Mortality for Each AHRQ-CCS Procedure Category	72.340	16.802	4.305	0.000
Hour of Operation	0.043	0.032	1.336	0.181
day:Monday	-0.088	0.302	-0.290	0.772
day:Thursday	-0.367	0.319	-1.148	0.251
day:Tuesday	0.123	0.295	0.416	0.677
day:Wednesday	-0.388	0.322	-1.204	0.228
Age	0.028	0.008	3.575	0.000
Body Mass Index	0.003	0.012	0.295	0.768
ASA Physical Status:I-II	8.324	544.450	0.015	0.988
ASA Physical Status:III	9.324	544.450	0.017	0.986
ASA Physical Status:IV-VI	10.301	544.450	0.019	0.985
Charlson Comorbidity Index	0.011	0.035	0.323	0.746

The results of the model in which 30-day mortality is the response variable show that the only statistically significant predictors in the model are the Risk Stratification Index (30-day mortality), the Overall Incidence of 30-Day Mortality for each AHRQ-CCS Procedure Category, and the patient's age. This model shows that after adjusting for critical risk factors, the variability of whether a patient experiences 30-day mortality after a surgical case is not significantly explained by the day of the week or the hour in which the operation took place. In terms of the `hour` of operation coefficient, the model shows that for each additional hour in a workday, the log-odds of a patient experiencing 30-day mortality after an operation are expected to increase by 0.043, holding all other predictors constant. Furthermore, for each additional hour in a workday, the odds of a patient experiencing 30-day mortality after an operation are expected to multiply by a factor of approximately 1.044 (exp(0.043)) holding all other predictors constant. This means that the general odds a patient results in 30-day mortality increase as the workday progresses. Since the baseline for the `day of the week` variable is Friday, it is best to interpret the Monday coefficient to gauge the differences between the beginning of the week and the end of the week. The log-odds of a patient having 30-day mortality after an operation are expected to be 0.088 less on Mondays than on Fridays, holding other predictors constant. Furthermore, the odds of a patient having 30-day mortality on Mondays is expected to be 0.916 (exp(-0.088)) times the odds of that on Fridays, holding all else constant. This means the odds of a patient having 30-day mortality are a bit lower at the beginning of the week than at the end of the week.

Table 2: In-Hospital Complications Model Coefficients/Summary Stats

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-3.420	1.110	-3.081	0.002
Risk Stratification Index (In-Hospital Complications)	0.299	0.016	19.313	0.000
Overall Incidence of In-Hospital Complications for Each AHRQ-CCS Procedure Category	6.250	0.186	33.588	0.000
Hour of Operation	-0.004	0.006	-0.707	0.480
day:Monday	-0.061	0.058	-1.055	0.292
day:Thursday	0.004	0.060	0.065	0.948
day:Tuesday	-0.148	0.059	-2.512	0.012
day:Wednesday	-0.024	0.058	-0.409	0.682
Age	0.011	0.001	8.287	0.000
Body Mass Index	0.009	0.003	3.442	0.001
ASA Physical Status:I-II	-0.338	1.100	-0.307	0.759
ASA Physical Status:III	-0.013	1.100	-0.012	0.991
ASA Physical Status:IV-VI	-0.047	1.103	-0.043	0.966
Charlson Comorbidity Index	-0.004	0.009	-0.391	0.696

The results of the model in which in-hospital complication is the response variable show that the only statistically significant predictors in the model are the Risk Stratification Index (in-hospital complication), the Overall Incidence of In-Hospital Complication for each AHRQ-CCS Procedure Category, the days of the week Tuesday and Friday, the patient's age, and the patient's BMI. This model shows that after adjusting for critical risk factors, the variability of whether a patient experiences in-hospital complication is somewhat significantly explained by the day of the week but not the hour in which the operation took place. In terms of the `hour` of operation coefficient, the model shows that for each additional hour in a workday, the log odds of a patient experiencing in-hospital complication are expected to decrease by 0.004, holding all other predictors constant. Furthermore, for each additional hour in a workday, the odds of a patient experiencing in-hospital complications are expected to multiply by a factor of approximately 0.996 (exp(-0.004)) holding all other predictors constant. This means that the general odds a patient results in in-hospital complications roughly stay the same as the workday progresses. In terms of the Monday `day of the week` variable, the log odds of a patient having in-hospital complications are expected to be 0.061 less on Mondays than on Fridays, holding other predictors constant. Furthermore, the odds of a patient having in-hospital complications on Mondays are expected to be about 0.941 (exp(-0.061)) times the odds of on Fridays, holding all else constant. This means the odds of a patient having in-hospital complications are slightly lower at the beginning of the week than at the end of the week.

Discussion

1) Hour of Operation & 30-Day Mortality

In terms of EDA, the box plots displaying the relationship between the hour of operation and the status of 30-day mortality in Image 15A) show that the patients whose operations result in a 30-day mortality status had operations that took place in hours of a slightly later range than patients whose operations did not result in a 30-day mortality status. Also recalling the *Summary Statistics of Case Frequencies* table above, the distribution of most operations happened in the morning, so it is noteworthy that even though a small proportion of operation cases happened later in the day, they still largely comprised of patients resulting in a 30-day mortality cases in comparison to patients who did not result in 30-day mortality. (Appendix: Image 9) Though the logistic regression model deemed the hour of operation not a statistically significant predictor in the variability of patients' 30-day mortality status, the model still showed that the general odds a patient results in 30-day mortality do increase as the workday day progresses. Overall, this shows that there is an association between the hour of an operation and whether or not a patient result in having 30-mortality. Though this association is not statistically significant, given the context of operations being high stakes and the response variable being a critical measure of one's life, it is imperative to highlight the relationship found between hour of operation and 30-day mortality.

1) Day of Week & 30-Day Mortality

In the EDA, the bar plots displaying the relationship between the day of the week and the status of 30-day mortality in Image 16A) show slight decreases in the number of patients who do not have 30-day mortality as the week progresses up until Friday. But it is also noteworthy that this pattern matches the decrease in the total patient population (Appendix: Image 10) as well as the decrease in percentages of the total patient population as seen in the *Summary Statistics of Case Frequencies* table above. This means the EDA shows there is not a visible relationship between the 30-day mortality status and the day of the week, only the number of patients and the day of the week. The logistic model deemed the day of the week not a statistically significant predictor in the variability of patients' 30-day mortality status but the model still showed there was a slight decrease in the odds of a patient having 30-day mortality as the week progresses. This may be worth noting given the context of this study being a high-risk environment.

2) Hour of Operation & In-Hospital Complications

The EDA of Image 15B) shows box plots of the relationship between the hour of operation and the status of patients' in-hospital complications. This image shows that the hours of operations for both statuses take place at relatively the same range of time. This means that there is not a very visible association between the time of operation and whether or not a patient experiences in-hospital complications. The logistic regression model did not express the hour of operation to be a statistically significant predictor in the variability of patients' in-hospital complication status. The model also expressed that the general odds a patient experiences in-hospital complication roughly stays the same as the workday progresses which is reflected in the EDA. Overall, this shows that there is not a noteworthy association between the hour of an operation and whether or not a patient experiences in-hospital complications.

2) Day of Week & In-Hospital Complications

In terms of EDA, Image 16B) shows bar plots of the relationship between the day of the week and the status of in-hospital complications. There looks to be a decreasing number of patients without in-hospital complications as the week goes on up until Friday. The visual also shows slight decreases in the number of patients with in-hospital complications as the week progresses except on Wednesday. The pattern in the number of patients without in-hospital complications aligns with the pattern in the total patient population by day of the week (Appendix: Image 10) as well as the pattern in percentages of the total patient population as seen in the *Summary Statistics of Case Frequencies* table above. However, the pattern in the number of patients with in-hospital complications on the different days of the week displays some differences. This means the EDA shows there is somewhat of a visible relationship between the in-hospital complications status and the day of the week. This aligns with the logistic regression model that shows Tuesday and Friday being statistically significant predictors of the variability of in-hospital complication where the odds of a patient experiencing in-hospital complication is lower on Tuesdays than on Fridays. Overall, this shows that there is a significant association between the day of the week of the operation and whether or not a patient experiences in-hospital complications.

Methodology Evaluation

To assess the validity of my methodology process, I assessed the accuracy of my logistic regression models. The accuracy score for the model in which 30-day mortality was the response variable is 0.9962, meaning the model did a great job of classifying patients' 30-day mortality status based on the model choice and predictors used. The accuracy score for the in-hospital complications model was lower at 0.8629 meaning the model did relatively well in classifying patients' in-hospital complication status with the model. Another good metric of accuracy is the model sensitivity since it measures how many classifications were correct out of all the patients that actually had 30-day mortality or in-hospital complications. This is crucial because, in the context of possible death or critical complications, it is imperative to find out all patients who are likely at risk. The sensitivity score for both models is approximately 0.99 which is very helpful in terms of accurately classifying those who do have a positive mortality and complication status. (Appendix: Image 20A and 20B) I also produced PR curves to display the relationship between the sensitivity and specificity which displays the models being more in favor of the recall (sensitivity) than the precision (specificity) which is best given the fact that the context strives more for a higher sensitivity. (Appendix: Image 21A and 21B)

An alternative method that could have been useful would have been utilizing a logistic mixed model to account for the clustering effect that may be produced by patients being treated by the same surgeon but the surgeon data was not provided in this data set.

**Limitation, Implications, and Conclusions**

One limitation that this data provides is that it is not representative of the typical population. Though the data may reflect a population of patients who need surgeries in Cleveland, demographic factors and health descriptors such as race, age, BMI, and disease/disorder statuses were not equally stratified nor close to the general average. This makes the models less generalizable to other patient populations. Another limitation is that the data is quite outdated. There are possible changes in medical demand, surgical practices, diagnostic criteria, as well as treatment options in the past 13 years that make this study less applicable right now. Lastly, the data does not provide a metric that measures the start time of the surgeon's shift to accurately account for previous labor in the day.

This study has provided evidence that there is a slight association between whether a patient experiences a 30-day mortality status of a patient and the hour in which they receive an operation as well as a statistically significant association between whether a patient experiences in-hospital complications and the day on which they receive the operation. More specifically, the increase in the odds of having a positive mortality status increase slightly as the workday progresses and there is an increase in the odds of a patient's complication status in some of the later weekdays. Therefore this study indicates the possible negative effects of a surgeon's extended workshift and workweek. If given the chance to approach this study differently, I'd consider incorporating data from other clinics with more diverse patient profiles to make a model that reflects broader cases. I'd also incorporate profile data on the surgeon running each operation to account for variability in the response relating to surgical errors, experience level, and previous hours worked before the current patient. Altogether, this study provides insight into the relationship between surgical cases over metrics of time that would assist researchers and medical leaders in further understanding the detrimental outcomes of continuous extensive medical labor.

Appendix

Extra EDA Visuals
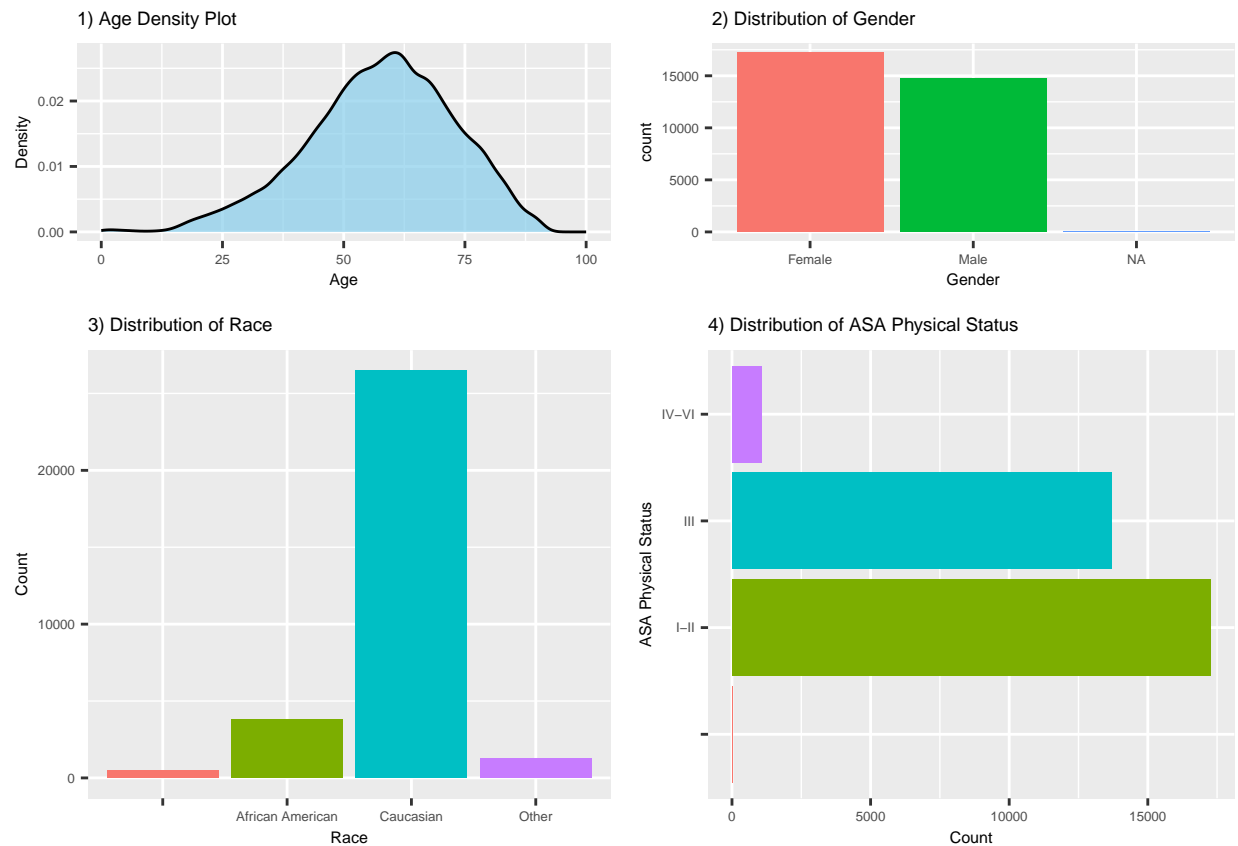
Image 1) Age

Since the distribution of the patients' age is relatively normal and unimodal and not really that skewed to the left the best summary statistic is the mean and the best metric of spread is the standard deviation.

Image 2) Gender

There appears to be a bit more operation cases done on women than men.

Image 3) Race

There are multidude more operation cases done on Caucasian patients than any other racial demographics.

Image 4) ASA Physical Status

*Note: ASA classification uses a grading system of I (one) through V (five) I identifying a person in good health and V as a person with a severe, life-threatening condition. The sixth (VI) status identifies deceased organ donors.

Most patients of operation cases reside in the I-II ASA status and then the III status comes in a close second. There are not many patients that reside in the IV-VI status.
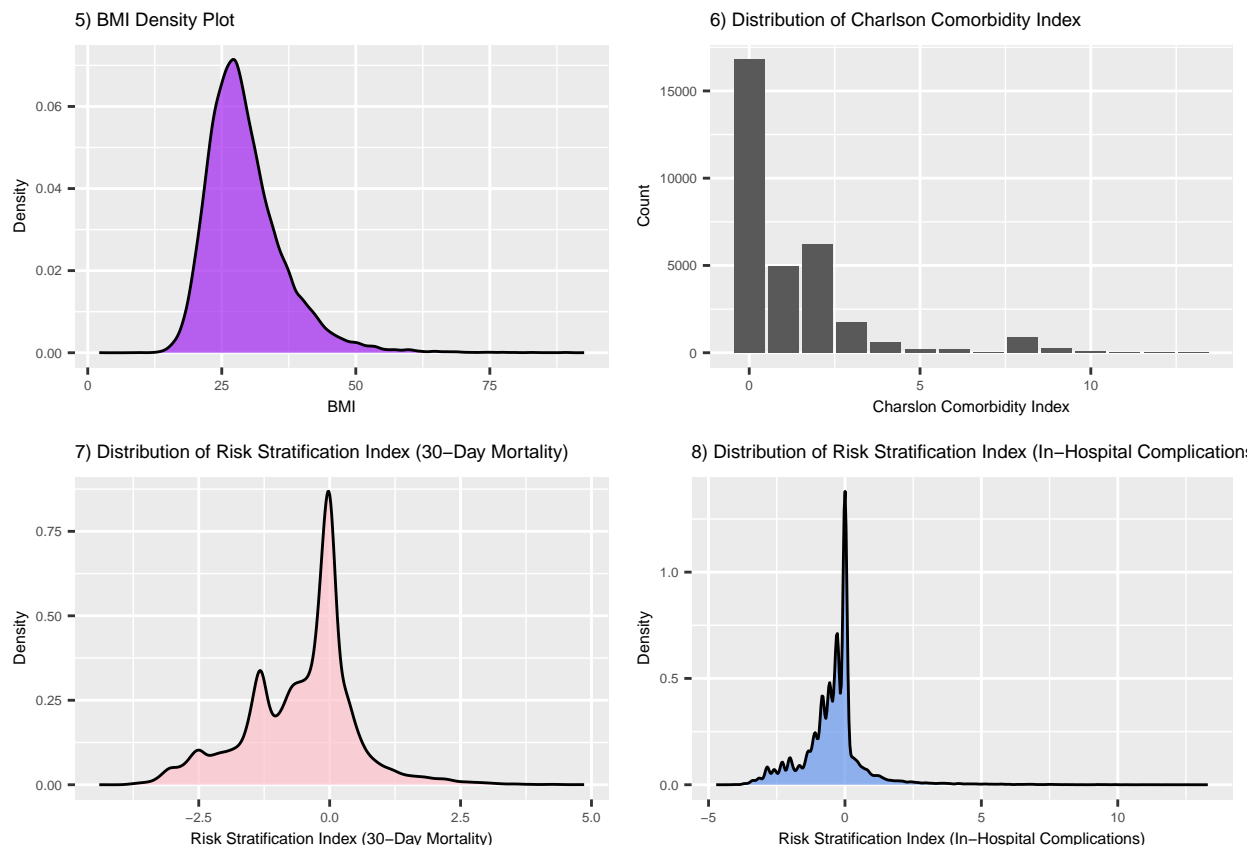
Image 5) BMI

Since the distribution of the patients' BMI is slightly skewed to the right the best summary statistic is the median and the best metric of spread is the IQR.

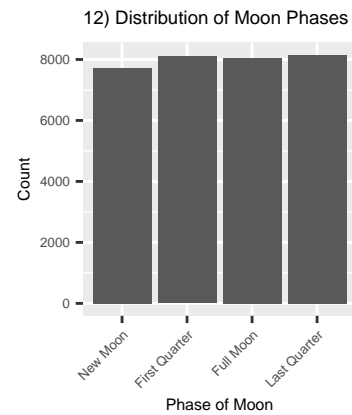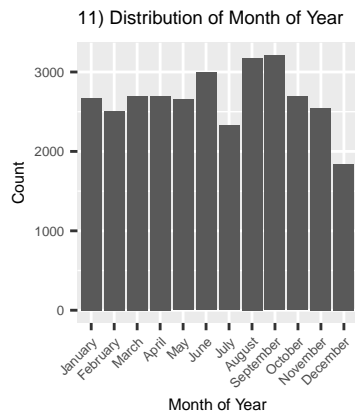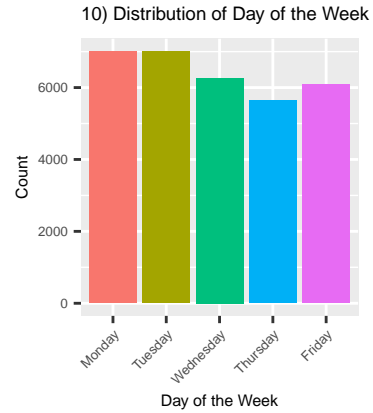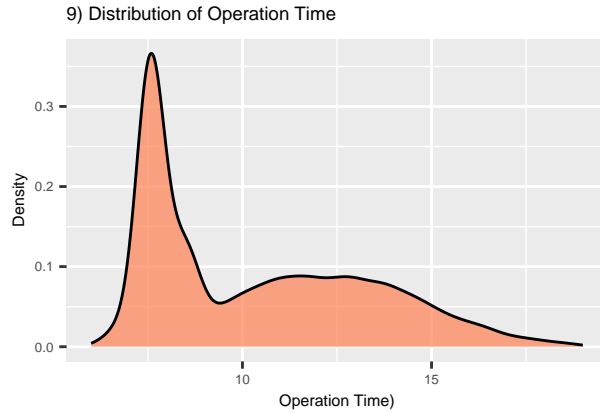*Image 6) Charlson Comorbidity Index*

*Note: The Charlson comorbidity index (CCI) is the most widely used comorbidity index used to determine survival rate (1yr and 10yr) in patient with multiple comorbidities. As the CCI level increases the probability of one year of survival decreases.

Since the distribution of the patients' Charlson Comorbidity Index is skewed to the right the best summary statistic is the median and the best metric of spread is the IQR.

Image 7) Risk Stratification Index (30-Day Mortality)

*Note: The Risk Stratification Index (RSI) is an open source, nationally validated, risk stratification methodology that permits outcomes such as duration of hospitalization and mortality to be compared equally across institutions.

The density chart of the distribution of RSI (30-day mortality) is multimodal and skewed to the left so the median and IQR would be more accurate measures of center and spread.

Image 8) Risk Stratification Index (In-Hospital Complications)

The density chart of the distribution of RSI (in-hospital complications) is multimodal and skewed to the left so the median and IQR would be more accurate measures of center and spread.

9) Distribution of Operation Time

10) Distribution of Day of the Week

11) Distribution of Month of Year

12) Distribution of Moon Phases

Image 9) Operation Hour

The distribution is bi-modal with a huge concentration of operation hours at around 8:00 am and a much smaller concentration at 12:30 pm.

Image 10) Day of the Week

The distribution of the days the of operation cases tend have a higher concentration in the beginning of the week.

Image 11) Month of the Year

The distribution of the months of operation cases show peaks in June and near the end of the summer (August and September).
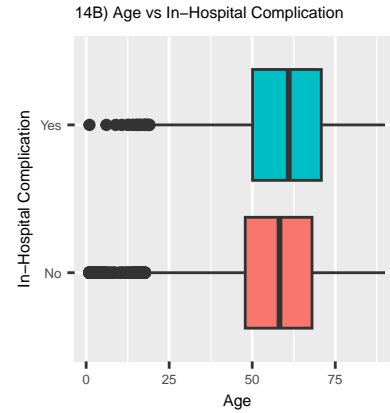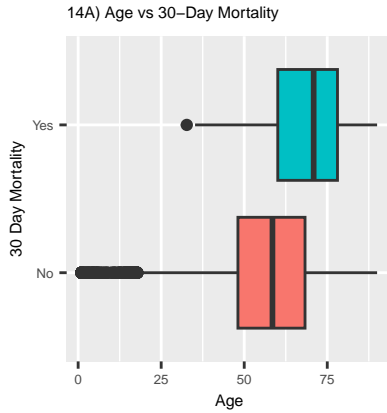
Image 12) Phase of Moon

The distribution of the moon phases of operation cases show a relatively uniform distribution with the least cases occuring during new moon phases.

13) Distribution of AHRQ–CCS Procedure Category
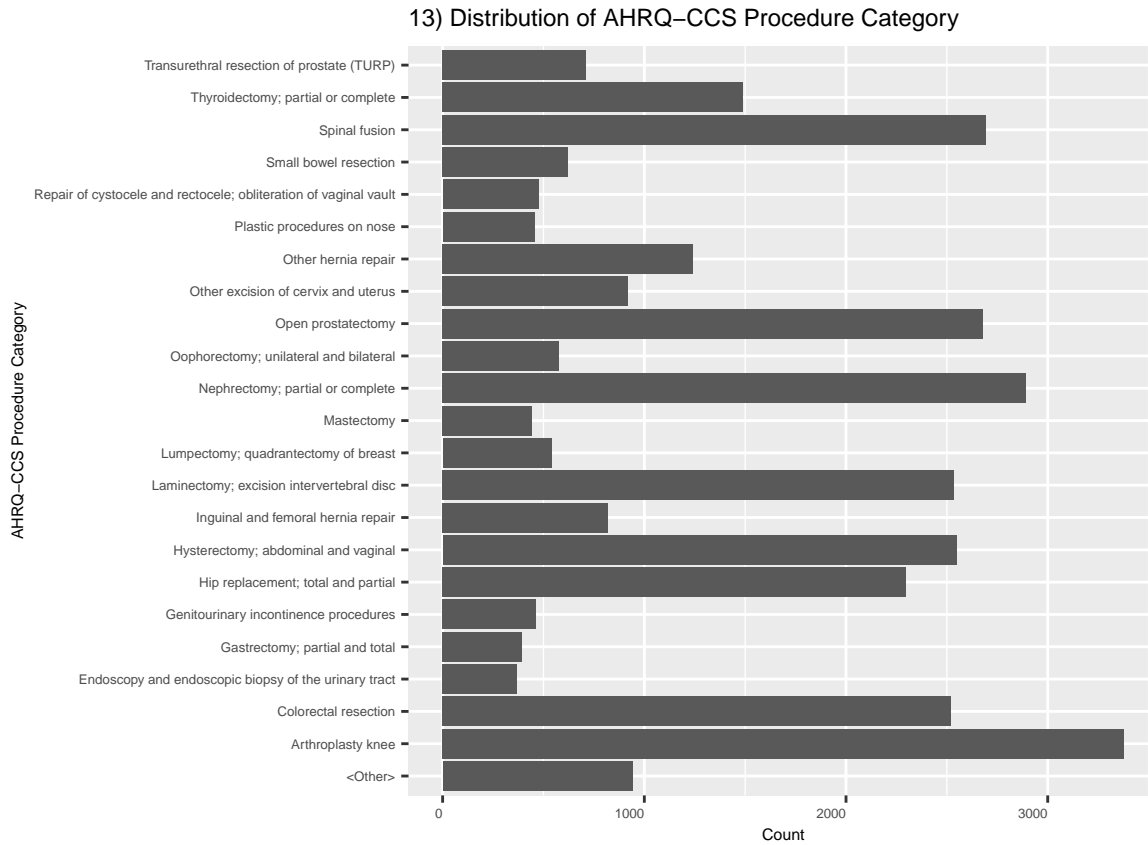
Image 13) AHRQ-CCS Procedure Category

The distribution of procedure type shows there are multiple sporadic procedure types that have very high concentrations of operations and multiple that have low concentrations.
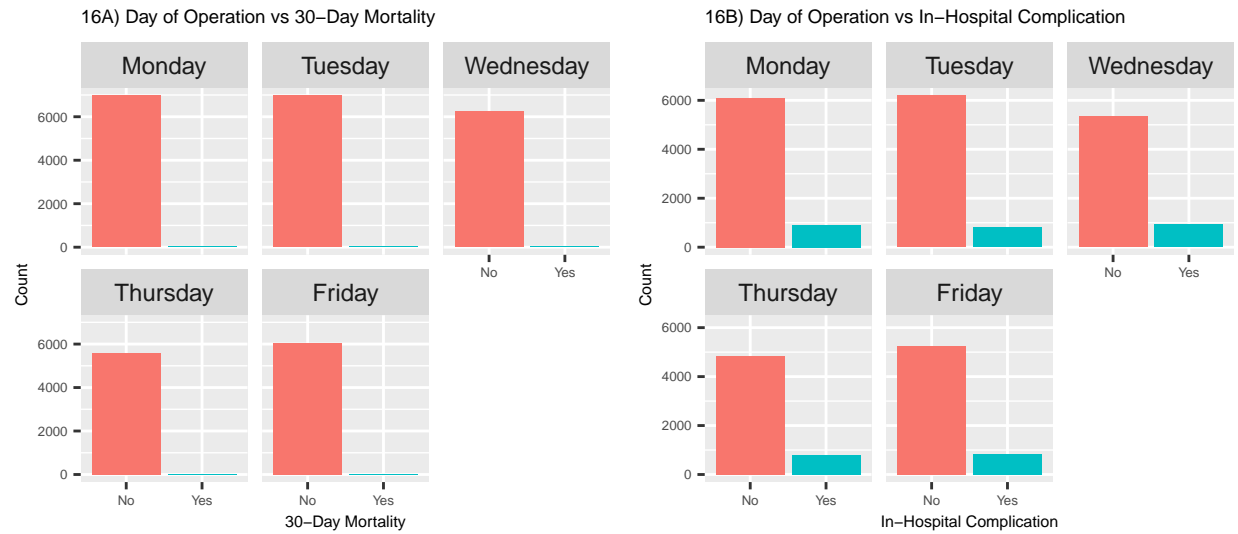
14A) Age vs 30–Day Mortality

14B) Age vs In–Hospital Complication

15A) Hour of Operation vs 30–Day Mortality

15B) Hour of Operation vs In–Hospital Complication
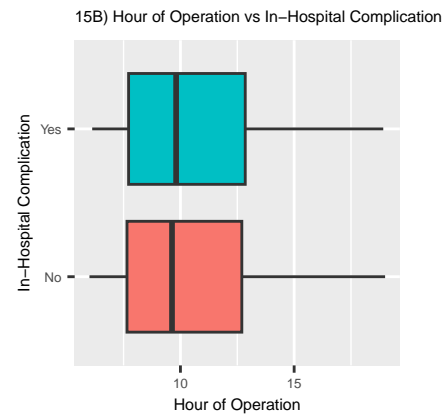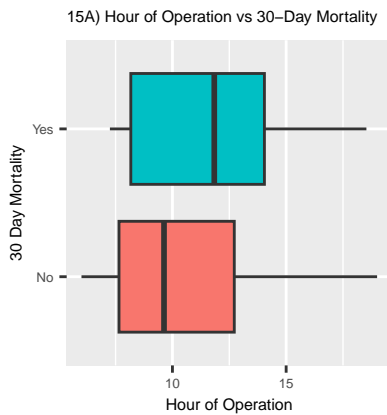
Image 14) Age vs 30 Day Morality and Age vs In-Hospital Complication

The first box plot shows that patients that have 30-day mortality are more generally of an older age than patients who do not have 30-day mortality. The second box plot shows that patients that have in-hospital complications tend to be only slighty older than those who do not.

Image 15) Hour vs 30 Day Morality and Hours vs In-Hospital Complication

The first box plot shows that operations of patients that have 30-day mortality tend to happen later in the day than operations of patients that do not have 30-day mortality. The second boxplot shows that operations of patients that have in-hospital complications tend to be relatively around the same hour in the day as operations of patients that do not have in-hospital complications.

Image 16) Day of Week vs 30 Day Mortality and Day of the Week vs In-Hospital Complication

In bar plot 16A) there is a noticeably decreasing number of patients without 30-day mortality as the week goes on. In bar plot 16B) there is also a decreasing number of patients without in-hospital complication as the week goes on.

Model Diagnostics

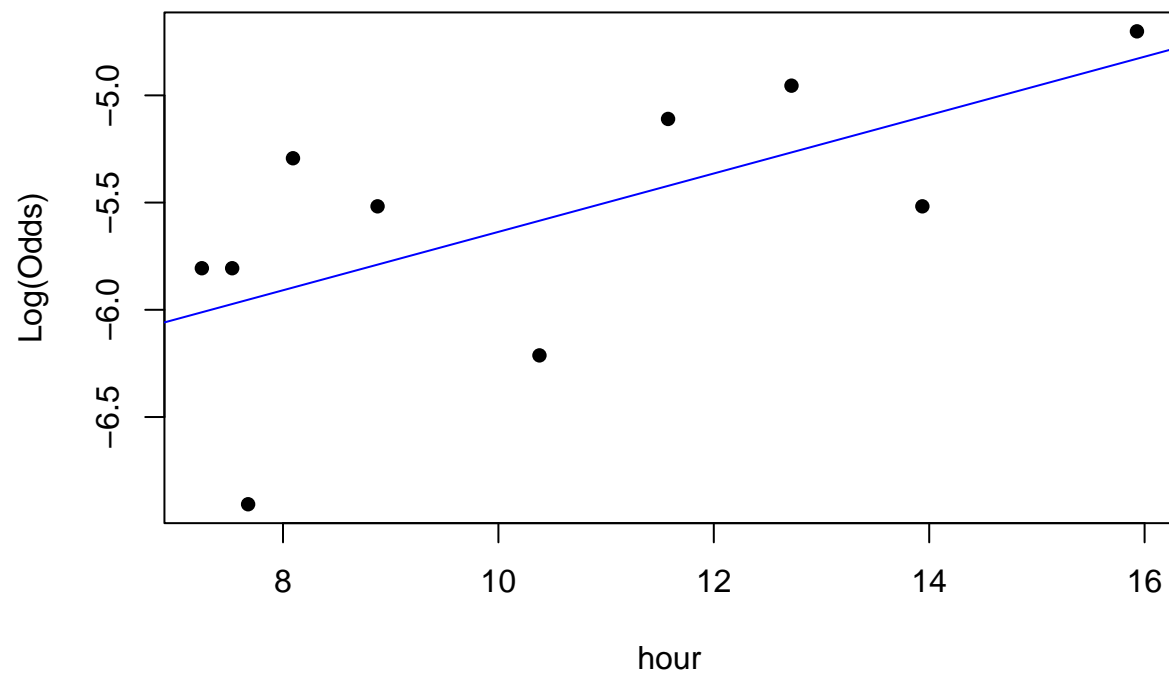*Image 17) Linearity: Logit Plot of Hour of Operation vs 30-Day Mortality*

Image 18) Linearity: Logit Plot of Hour of Operation vs In-Hospital Complication
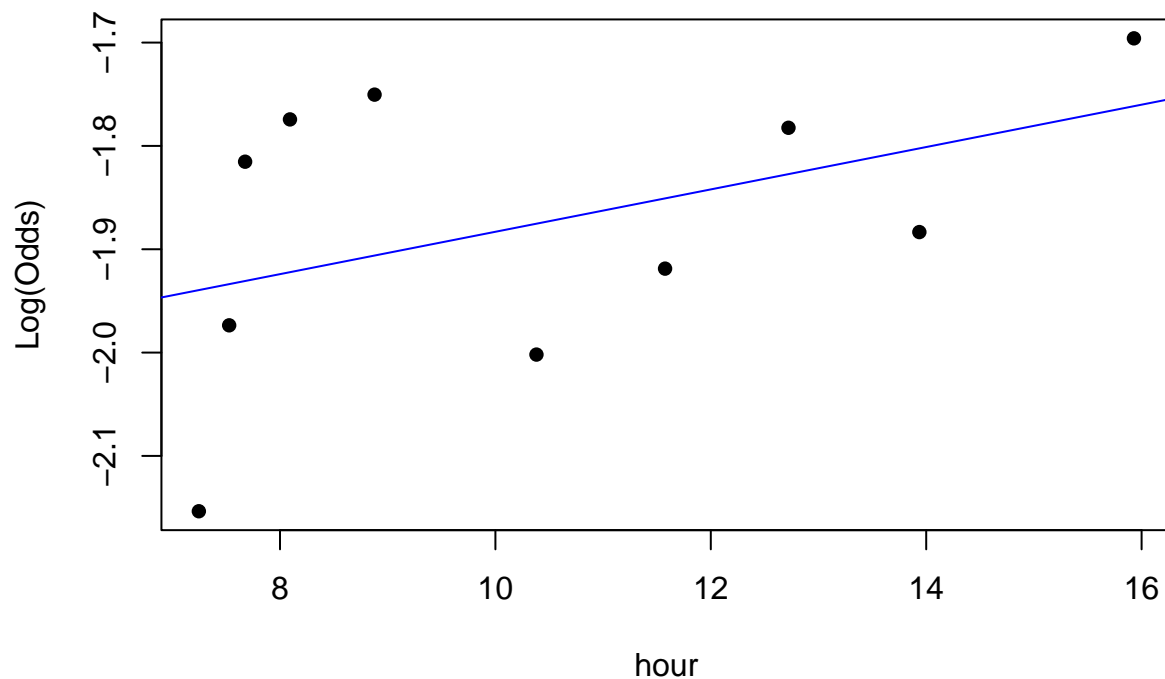
Image 19) Multicollinearity: Correlation Matrix

```
##                 mortality_rsi ccsMort30Rate       hour age bmi
## mortality_rsi       1.0000000    0.14957489 0.12163220  NA  NA
## ccsMort30Rate       0.1495749    1.00000000 0.03523460  NA  NA
## hour                0.1216322    0.03523460 1.00000000  NA  NA
## age                        NA            NA         NA   1  NA
## bmi                        NA            NA         NA  NA   1
## baseline_charlson   0.2535465    0.03504596 0.02680998  NA  NA
##                 baseline_charlson
## mortality_rsi          0.25354649
## ccsMort30Rate          0.03504596
## hour                   0.02680998
## age                            NA
## bmi                            NA
## baseline_charlson      1.00000000
```

Model Accuracy Check

Image 20A) Confusion Matrix and Summary Statistics for 30-Day Mortality Model

```
## Confusion Matrix and Statistics
##
##
## predicted_classes_mort30    0    1
##                        0 5717   19
##                        1    3    2
```

```
##
##                 Accuracy : 0.9962
##                   95% CI : (0.9942, 0.9976)
##      No Information Rate : 0.9963
##      P-Value [Acc > NIR] : 0.640608
##
##                    Kappa : 0.1527
##
##  Mcnemar's Test P-Value : 0.001384
##
##              Sensitivity : 0.99948
##              Specificity : 0.09524
##           Pos Pred Value : 0.99669
##           Neg Pred Value : 0.40000
##               Prevalence : 0.99634
##           Detection Rate : 0.99582
##     Detection Prevalence : 0.99913
##        Balanced Accuracy : 0.54736
##
##         'Positive' Class : 0
##
```

Image 20B) Confusion Matrix and Summary Statistics for In-Hospital Complications

```
## Confusion Matrix and Statistics
##
##
## predicted_classes_complication    0    1
##                              0 4897  738
##                              1   49   57
##
##                 Accuracy : 0.8629
##                   95% CI : (0.8537, 0.8717)
##      No Information Rate : 0.8615
##      P-Value [Acc > NIR] : 0.3888
##
##                    Kappa : 0.0971
##
##  Mcnemar's Test P-Value : <2e-16
##
##              Sensitivity : 0.9901
##              Specificity : 0.0717
##           Pos Pred Value : 0.8690
##           Neg Pred Value : 0.5377
##               Prevalence : 0.8615
##           Detection Rate : 0.8530
##     Detection Prevalence : 0.9815
##        Balanced Accuracy : 0.5309
##
##         'Positive' Class : 0
##
```
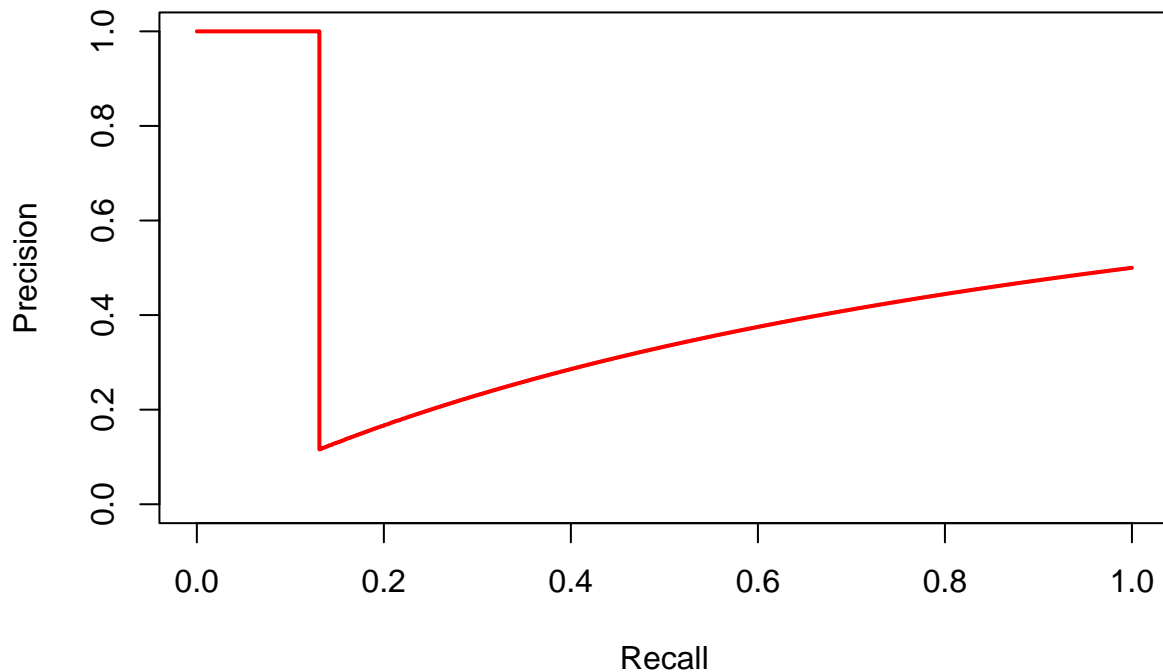
Image 21A) Precision-Recall (PR) curves) for 30-Day Mortality Model

Precision−Recall Curve for 30−Day Mortality
AUC = 0.3110238
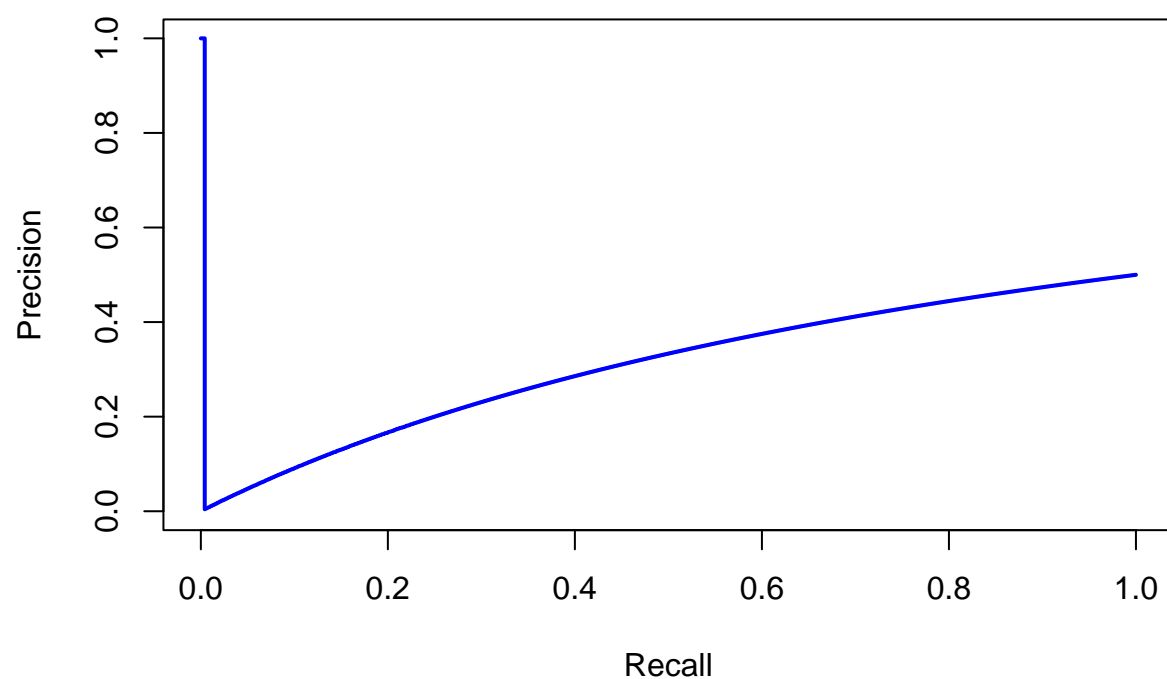
Image 21B) Precision-Recall (PR) curves for In-Hospital Complications

Precision–Recall Curve for In–Hospital Complications
AUC = 0.4300009

Work Cited

- Alhola, Paula, and Päivi Polo-Kantola. "Sleep deprivation: Impact on cognitive performance." Neuropsychiatric disease and treatment vol. 3,5 (2007): 553-67.
- Džakula, Aleksandar, and Danko Relić. "Health workforce shortage - doing the right things or doing things right?." Croatian Medical Journal vol. 63,2 (2022): 107-109. doi:10.3325/cmj.2022.63.107
- Sessler et al."Operation Timing and 30-Day Mortality After Elective GeneralSurgery". Anesth Analg 2011; 113: 1423-8
- Sessler, Daniel I., et al. "Operation timing and 30-day mortality after elective general surgery." Anesthesia & Analgesia, vol. 113, no. 6, 2011, pp. 1423–1428, https://doi.org/10.1213/ane.0b013e3182315a6d.