Survival Analysis of Interventional Cardiac Catheterizations

Jessie Bierschenk, Holly Cui, Leona Lu, Shari Tian

Introduction

Background

For diagnosing or treating heart problems, cardiac catheterizations are common procedures used by doctors. In this procedure, a thin, flexible tube (catheter) is guided through a blood vessel to the heart to diagnose (diagnostic catheterization) or treat (interventional catheterization) certain heart conditions, [such as clogged arteries or irregular heartbeats; this method renders important informaiton about the heart muscle, heart valves, and blood vessels in the heart.](https://www.mayoclinic.org/tests-procedures/cardiac-catheterization/about/pac-20384695#:~:text=Cardiac%20catheterization%20(kath%2Duh%2D,clogged%20arteries%20or%20irregular%20heartbeats) This research specifically explores the use of catheterizations as a clinical intervention in the context of heart disease and cardiac events among human patients. Even with interventional cardiology existing as a well-established field of medicine, access to such care and resources is not universal. Today, many heart conditions can be treated during an interventional heart catheterization that would have otherwise needed a separate, more intense surgery. We aim to use data from Duke Hospital, a well-established medical institute, to better inform doctors/hospitals of the factors to consider in the context of (interventional) catheterizations in the hopes of motivating an introduction or improvement of the practice.

To gain a deeper understanding of variables associated with this procedure, we begin by analyzing a dataset that includes adult patients catheterized at Duke Hospital starting in 1985 through 2013, and only includes patients that have evidence of significant extant coronary artery disease. The data contain medical and demographic records for 39,098 unique patients, with one record per procedure. In total, there are 83,320 observations, each representing a unique procedure.

The data include records of procedures involving both diagnostic and interventional catheterizations. While almost all procedures in the dataset involve diagnostic catheterizations, only about 33.7% involve interventional catheterizations. A study done in 2010-2011 shows that within approximately 85% of the cardiac catheterization laboratories in the United States, "on-site cardiac surgery was not available in 83% of facilities performing fewer than 200 [interventional catheterizations] annually, with these facilities representing 32.6% of the facilities reporting, but performing only 12.4% of the [interventional catheterizations] in this data sample." This finding highlights a potential lack of interventional catheterization procedures available as well as the limited amount of relevant research on interventional procedures. Motivated by this, our analysis aims to focus on analyzing patient survival rates associated with procedures involving an interventional catheterization vs. patient survival rates associated with procedures that did not involve an interventional catheterization, all the while controlling for and analyzing other factors associated with patient survival. By doing so, we hope to provide information to be used by doctors and hospitals that they can use to improve or introduce the practice of interventional catheterizations within their health services. This would allow clinicians and medical institutions to better isolate the association between interventional catheterization procedures and patient survival rates, thus motivating future efforts to reduce wasteful and inefficient care variations and implement effective, continuous quality improvement processes.

Our primary (1) and secondary (2) research questions are as follows:

(1) When controlling for other patient/procedure characteristics, what is the association between patient survival and the use of interventional coronary catheterizations?

(2) In regards to patient data that a physician can access on the day of a given procedure, what factors show evidence for an association with patient survival?

In addition to informing the use of interventional catheterizations across healthcare systems, our research aims to answer these questions in order to give physicians a more comprehensive understanding of appropriate or optimal procedures for a given patient. This means that we want to synthesize information to inform physicians about past research on survival rates of patients receiving an interventional catheterization while controlling for other variables. With this, we will be focusing on variables associated with information that a physician can access about a patient in real-time and not on information that can only be gathered retrospectively; these variables will be our predictor variables.

Exploratory Data Analysis

Response Variable Our event of interest is survival time or days to last known alive. There are 180 patients who died at time zero and 23 patients with no record of having died by the end of follow-up, but who have no follow-ups. We believe these 23 patients may have moved away or "disappeared" after the first procedure. Because our research question is interested in patients who survive the procedure and do not have other critical conditions that would lead to an immediate death, we will drop all 203 patients because they provide no further information in our study. We plot

Table 1: Interventional vs. Diagnostic Catheterization Frequencies

	Had Inverventional Cath.		Did Not Have Inverventional Cath.	
Had Diagnostic Cath.	26770	(32.1%)	52873	(63.5%)
Did Not Have Diagnostic Cath.	1321	(1.6%)	2356	(2.8%)

the survival probability with patients who had interventional coronary catherizations and those who didn't. Figure 1 shows an association between patients who had interventional catheterizations and higher survival probability.

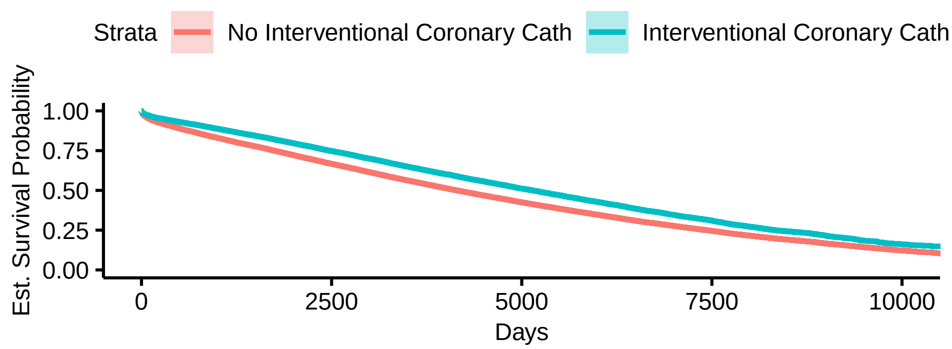


Figure 1: Estimated Survival Probability Between Patients With and Without Interventional Catheterizations

Table 1 shows the breakdown of catheterization procedure types within our dataset. 95.6% of procedures in the this data set involved a diagnostic catheterization, yet only 33.7% involve an interventional catheterization; such discrepency motivates further investigation into interventional catheterizations. Most procedures involved a diagnostic and not an interventional catheterization. The next most frequent procedure involved both an interventional and diagnostic catheterization. Less than 5% of the data did not include a diagnostic catheterization. This distribution allows us to explore the differences in survival among patients who did vs. did not receive an interventional procedure.

Figure 2 takes the catheterization details of each procedure (observation) in the dataset as well as the respective patient's survival status by the end of follow-up. Across all procedure types, there were more procedures associated with patients that died than those that survived by the end of the follow-up. The procedure groups that involved a diagnostic catheterization had lower or about the same proportion of deceased (of respective patient by end of follow-up) than procedure groups that did not involve one. Such pattern is not as obvious with procedures including interventional procedures. Similarly, the cardinal number of procedures related to a patient that died versus didn't was much closer for procedures with a diagnostic and interventional catheterization compared to other procedures. With this lack of clear associations between use of interventional catheterizations in procedures and vital status of respective patient, our research will continue to explore other factors associated with patient survival after interventional catheterization.

Methodology

Selected (and Manipulated) Variables

Catheterization Identification and Linking We will include the sequential catheterization number variable in our analysis because the number of previous catheterizations a patient has had is one indicator of their clinical presentation. Higher sequential catheterization number signifies more previous catheterizations, which would indicate greater severity of a patient's cardiac condition at the time of the procedure. Since current research on cardiac catheterizations risks and complications suggests that a patient's clinical presentation influences complication rate following cardiac catheterizations, we will account for sequential catheterization number in our model. There are no missing data for this variable. For the purposes of interpreting our model output, we will subtract 1 from every observation of sequential catheterization number so that the first visit equates to a vlue of 0.

Demographic Characteristics According to current research on cardiac catheterizations risks and complications, the complication rate of catheterizations is also dependent on the demographics of the patient. Thus, for our analysis, we will try to include all demographic variables in our model. These include age, gender, and race of the patient.

The patient age variable does not have any missing data. From Figure 1A, we can see that there is a pattern between the age of the patient and whether or not they died at the end of follow-up. Thus, we will include age in our model.

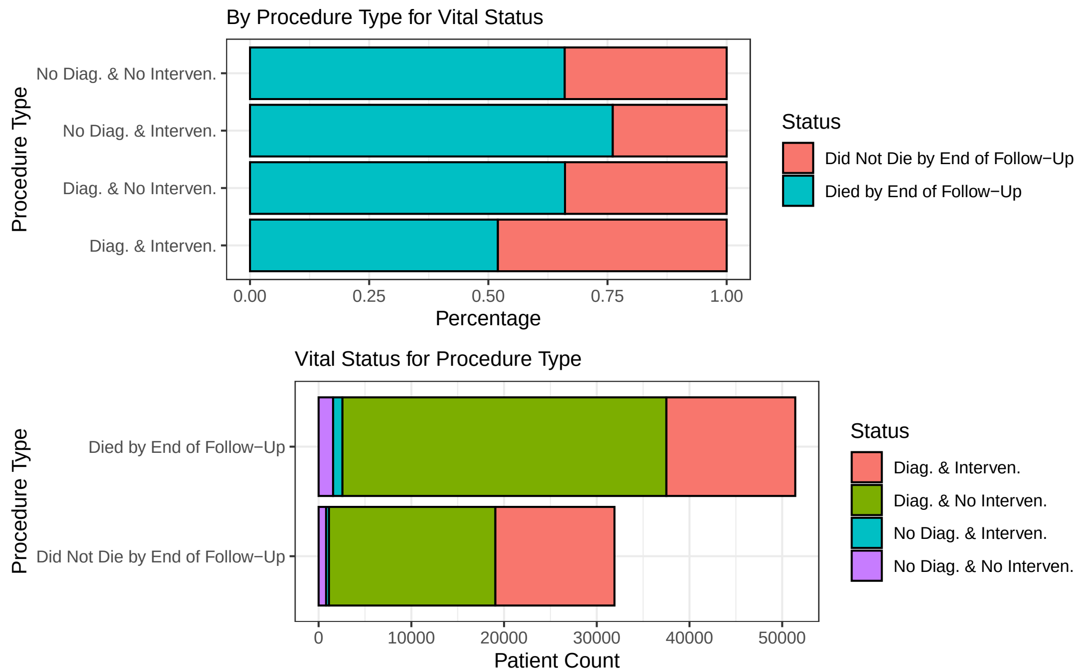


Figure 2: Breakdown of Vital Status and Procedure Type

The patient gender variable also does not have missing data. From Figure 1A, it appears that a larger percentage of women in the data set died at the end of follow-up compared to men. Given this difference, we will include gender in our model.

There is some missingness for the patient race variable. Due to the desire to keep as many observations as possible, and the impracticality imputing race through prediction, we will categorize all missing values for race into an "unknown" category. From Figure 1A, there doesn't appear to be much difference in the vital status outcomes among White, African American, and other-race identifying patients. However, the patients of unknown race seem to have a lower proportion of deaths than the other three categories. Given this observation, we will still include race in our model.

Patient History In order to achieve our main objective of giving physicians a more comprehensive understanding of survival outcomes following interventional catheterization procedures, patient history variables are of great importance. This section includes justifications for the inclusion of variables encompassing patients' history of selected diseases, days to closest previous visits for one or several related diseases, and syndrome severity levels.

Acute Coronary Syndrome(ACS) Status Upon Presentation: According to current documentation, ACS is a term used to describe a range of conditions associated with sudden, reduced blood flow to the heart. This means that it is a collection of relevant syndromes rather than only one specific problem. The variable value is number-coded, each of which indicates the patient's syndrome status, including No ACS, STEMI, Non-STEMI, MI Unspecified, and Unstable Angina (STEMI is short for ST-Elevation Myocardial Infarction, where MI is Myocardial Infarction.) However, we have two other individual variables that also indicate history of MI and Angina. Through comparisons, we can see that ACS is a standard status indicator upon presentation in each visit, which is unrelated to patient history, and independent of MI. More specifically, even if the patient was checked STEMI upon a visit, the doctor may not record it as MI in patient's history. Similarly, even if there is Unstable Angina presented during visits, it is not the factor that decides the history of Angina. Thus, to simplify the variable levels, we decide to split ACS into two separate dummy variables, STEMI and Unstable Angina upon presentation. The scheme is stated as whenever the observation has recorded STEMI in ACS during the visit, STEMI variable is set to 1, otherwise 0. The same logic applies to Unstable Angina in ACS.

Number of previous Myocardial Infarction(MI): There are three inter-connected patient history myocardial infarction (MI) variables: history of MI, days to closest previous MI, and number of previous MIs. History of MI is a binary variable indicating if the patient has had an MI before. Days to previous MI numerically shows the time lapse from the closest MI occurrence, where 0 indicates it happened on the same day of visit, and NA means no previous MI is identified. Number of previous MIs shows the total number of MIs before the current visit ranging from 0 (no MI occurs) to any positive integer (at least one MI occurs). Since all three variables have overlapping information coded, we ran correlation analysis and discovered that each of the two pairs are perfectly correlated. Thus, to grasp maximum information while keeping the least amount of predictors in our model, we will choose number of previous MIs as the best option among the three, due to the fact that it can both indicate whether a patient has a history of MI (==0) and capture the frequency of the patient experiencing MIs (>=1).

Congestive Heart Failure (CHF) Severity: CHF severity is defined as the worst dyspnea class in the previous 2 weeks, similar to ACS. We will choose to include CHF severity in our analysis over the history of Congestive Heart Failure (CHF). The history variable is a binary indicator of whether the patient had CHF previously, while the severity variable consists of values ranked based on no CHF to severe CHF syndromes. From Figure 2A, we can see that an observation having a history of CHF may also have no severity; however, there were no observations in which a patient presented with symptoms but did not have a history of CHF. Thus, we will keep CHF Severity and not CHF History in our analysis.

Other disease history: To maximize patient history information of relevant diseases, we will also include other binary variables that would be accessible to a doctor upon a patient visit: history of Angina, Cerebrovascular Disease, Chronic Obstructive Pulmonary Disease, Diabetes, Hypertension, Hyperlipidemia, and Smoking. We want to explore how these variables may be associated with patient survival.

Vital Signs Because our analysis focuses on patient information that a doctor can analyze on the day of a procedure/check-up, we will include vital sign variables in our model analysis. These variables include systolic blood pressure (mmHg), diastolic blood pressure (mmHg), and heart rate (bpm).

Because systolic and diastolic blood pressure appear to have a strong linear relation, and in a medical context are looked at together, we will use information from an article describing normal and abnormal blood pressure based on systolic and diastolic blood pressure readings and categorize the readings into levels: "Normal", "Elevated", "Stage 1 Hypertension", "Stage 2 Hypertension", "Hypertensive Crisis", and "Missing" for readings where there were missing values for both systolic and diastolic blood pressure. As seen in Figure 8A, most of the missingness are from records of procedures in the earliest year group (1985-1990), and because of this, we will make a "Missing" category rather than try to impute it. There were 131 observations with values for one blood pressure type but not the other. For these we looked at the record we did have to sort into the appropriate category.

Heart rate: As seen in Figure 8A, most of the missing observations for heart rate are from the earliest year grouping in the study (1985-1990). Because there does not seem to be any other consistent trend among this missing data, we assume that this missingness may be attributed to a lack of consistent recording during those early, less technologically-advanced years, or it may be harder to find data from longer ago. Because of this, we will make a "Missing" category rather than try to impute it. We categorize the heart rate into groups based on scientific discussion of heart rates. According to the British Heart Foundation a normal resting heart rate for an adult is between 60-100 bpm. We will hold this as a category and make appropriate ranges for lower than normal and higher than normal heart rate. Any missing data will go into a "missing" category.

Physical Examination Like vital signs, physical examination results are also patient information that a physician can analyze on the day of a visit. Thus, we will include physical examination variables in our model analysis.

Third heart sound will be included as it is a patient data point that a doctor can gather from a patient on the day of a procedure. As seen in Figure 8A, over 95% of the missing data is from the first year group or records (1985-1990). Because of this, we will categorize the 2.9% of missing data as "missing."

We want to take into account the BMI of the patient, serving as a gauge of the patient's overall health. There isn't a BMI variable in the data set, so we will create one from the height and weight variables. However, there is some missingness for the patient height and weight in the data set. We will assume these data are MAR because they are likely missing due to the clinician failing to record these physical exam characteristics for certain patient visits. Given this assumption, we will impute these data using the Multiple Imputation via Chained Equations (MICE) algorithm. Specifically, we will create a temporary subset of the data with just patient age, gender, race, height, and weight because we believe that these demographic characteristics are likely to be the most correlated with and thus most informative for predicting height and weight. We will then use predictive mean matching to impute both patient height and weight in this subset before storing them back in the original data frame. While BMI calculations are not adjusted for gender, patient sex can affect BMI. For instance, males may be generally taller than females, and females may have more body fat compared to males. Figure 3A visualizes the relationship between BMI and vital status at the end of follow-up for males and females separately. For better visualization, the figure only includes those observations with BMI between 10 and 50. The figure shows a wider interquartile range of BMI values for female patients compared to male patients for both vital status outcomes. Additionally, in both vital status outcome categories, the median BMI for females is slightly larger than the median BMI for males. This justifies adding an interaction between BMI and gender in our model. BMI will be mean centered in our model to be appropriately interpreted.

Catheterization Results We want to include as much information as we can that would be accessible to a doctor on a patient's visit. Thus, we will try to include all catheterization results and handle missingness mostly by categorization methods. Our selection process is described below.

Coronary dominance may be an important assessment in determining appropriate procedures for patients. There are only 50 observations with missing data for coronary dominance. More than half of these results were procedures that did not include a diagnostic or interventional catheterization. As seen in Figure 8A, most of this missingess includes procedures from 1985-1990, and missingess may be attributed to higher difficulty finding records from longer ago. None of the missing values were associated with procedures that used both interventional and diagnostic catheterizations. However, because our research aims to explore interventional catheterizations and look at them in contect with diagnostic interventions, we will remove these 50 missing observations.

There is a high frequency of missingness in records of maximum stenosis in any graft. A bypass graft is a surgical procedure used to treat coronary heart disease that diverts blood around narrowed or clogged parts of the major artieries to improvles blood flow and oxygen supply to the heart. When comparing the subset of data with missing vs. non missing values for maximum stenosis in any graft, there does not appear to be any major discrepancy in death or symptom severity of associated patient. This missingness may be attributed to the fact that the patient did not receive a graft. Because we will look at stenosis in the other artertie s, we can remove this variable from our analysis and instead, create a binary variable indicating if the patient received a graft or not, assuming that missingness for maximum stenosis in any graft indicates no graft was done.

According to these distributions in Figure 4A, the regions of the heart show similar distributions of maximum stenosis in the catheterization results. However, when plotted against one another (Figure 5A), there is no clear relationship between the maximum stenosis variables except between the left anterior descending artery region and the proximal left anterior descending artery region. With this, we will impute some missing data from the the left anterior descending artery region with data from the proximal left anterior descending artery region and create a new variable containing this information that has stenosis data for the general region of the left anterior descending artery. We will then categorize the numbers by percentage groups (groups of 20 percentiles) and add a category "missing" to assign to the remaining missing data points. We repeat this categorization process for the maximum stenosis values of the other regions and include a "missing" category for missing values. We do not want impute these missing values across the maximum stenosis variables directly because it may be that the test to assess the maximum stenosis for a particular area of the heart was not performed and/or was not necessary for the procedure, potentially introducing a confounding variable that we will try to control for with the inclusion of the "missing" category.

The left ventrical ejection fraction is a measurement of the percentage of blood leaving the heart each time it contracts. According to the Mayo Clinic and the American Heart Association, a normal ejection fraction is about 50% to 75% and a borderline ejection fraction can range between 41% and 50%. As seen in Figure 6A, patients that did not die by the end of follow-up had a more normal distribution of ejection fraction than patients that did die by the end of follow-up. Because we want to control for other variables that may be associated with patient survival, we will include this result in our analysis. We will categorize this variable according to the "normalness" of the result and note if there was a missing value for this (likely there was no test performed to assess the ejection fraction).

We will include the number of significantly diseased vessels in our analysis. This information is something a doctor can analyze from a patient on a given day. Knowing such information may help a physician in deciding which catheterization procedures are appropriate during that time or in the future, and thus, we will include this variable in our model analysis.

Removed Variables - See Appendix For Details

Our research aims to give physicians a more comprehensive understanding of appropriate or optimal procedures for a given patient, and we want to synthesize information to inform physicians about past research on survival rates of patients receiving an interventional catheterization while controlling for other variables. With this, we will be focusing on variables associated with information that a physician can access about a patient in real-time and not on information that can only be gathered retrospectively. This motivates our removal of follow-up variables (information only obtained after our procedure). Additionally, we dropped variables we felt were redundant, not relavant to the ofucs of our analysis, or whose information could be adequately accounted for in other variables.

Model

From EDA, we can't identify a distribution for Days to Last Known Alive. As a result, we want to use distribution-free model who will not make distribution assumptions. We will explore two types of model: Buckley-James censored regression estimator and Cox Proportional hazards models:

$$\text{Cox PH Model: } \lambda_i(t) = \lambda_0(t)\exp(x_i^T\beta)$$
$$\text{Buckley James Model: } Y_i = \alpha + x_i\beta$$

Where $i = 1, ..., n$ represents n censored patients in the data set, β is a $p \times 1$ vector of regression coefficients corresponding to the covariate vector x_i for patient i, and p is the number of covariates used. The coefficents correspond to:

- **Demographic**: Age, Race, Gender
- **Patient History**: Patient's Sequential Cath Number, STEMI, Unstable Angina, History of Angina, CHF Severity, Number of Previous MIs, History of Smoking, History of Cerebrovascular Disease, History of COPD, History of Diabetes, History of Hypertension, History of Hyperlipidemia
- **Catheterization Result**: Maximum Stenosis of the Left Main Artery; Left Circumflex Artery; any GRAFT; anterior descending arterial system; Right Coronary Artery, Left Ventricular Ejection Fraction, Coronary Dominance
- **Vital Sign**: Diastolic Blood Pressure (mmHg), Heart Rate, Systolic Blood Pressure
- **Physical Examination**: BMI, Third Heart Sound

We found strong evidence that adjusted for demographic and patient history, intervention catherization procedures were associated with 2 times longer survival time compare to patients who did not receive the treatment.

Model Assumptions Cox Proportional hazards models have proportional hazards assumptions and linearity. To check for proportional hazard assumptions, we applied a hypothesis test where the null hypothesis states the data hold for proportional hazard rate. Only history of chronic obstructive pulmonary disease and history of smoking fail to reject the hypothesis. Other variables violate the proportional hazard assumption. We use deviance residual to assess outliers and linearity assumptions. From the deviance residual plot (Figure 7A), we can see that the residuals are evenly distributed between 4 and -4, thus satisfying linearity assumptions. While proportional hazard assumption is violated, we can interpret the coefficient as the "average effect" over time points that are observed in a data set. (Allison, 1995).

For Buckley-James, there are assumptions of linearity and homoscedasticity where we assume equal or similar variances in different groups being compared. From the residual plot (Figure 7B), we see the model fails the linearity assumption, with more residuals located below 0. In addition, the plot does not display an even distribution, violating the assumption of homoscedastic. Because Buckley-James model extend least square regression model to right-censored data. Violation of linearity will impact our ability to interpret coefficient.

Because our research goal is to explore association between survival with features in the model, the ability to interpret coefficient is crucial to our research question. As a result, we will apply Cox proportional hazard models instead of Buckley-James Model.

Results

Our full model output can be found in the Appendix (Table 4). A partial output of our cox model is shown below (Table 2):

Table 2: Cox Model Output (Partial)

Variable	Hazard Ratio	95% CI Lower Bound	95% CI Upper Bound	P-Value
Age 80+ (vs. Age 65-69)	2.528	2.430	2.630	<0.001
African American (vs. Caucasian)	1.194	1.163	1.227	<0.001
BMI Points Away from Mean	0.996	0.994	0.999	0.001
Female (vs. Male)	1.049	1.028	1.071	<0.001
Patient's Sequential Cath Number	0.987	0.983	0.991	<0.001
ST-Elevation Myocardial Infarction Present (vs. not present)	0.962	0.937	0.988	0.004
Unstable Angina Present (vs. not present)	0.973	0.953	0.994	0.013
Class I Congestive Heart Failure (vs. none)	1.404	1.344	1.468	<0.001
Class IV Congestive Heart Failure (vs. none)	2.087	2.002	2.176	<0.001
Number of Previous MIs	1.108	1.096	1.120	<0.001
41-60% Maximum Stenosis of the Left Main Artery (vs. 0-21%)	1.165	1.123	1.209	<0.001

Variable	Hazard Ratio	95% CI Lower Bound	95% CI Upper Bound	P-Value
61-80% Maximum Stenosis of the Left Main Artery (vs. 0-21%	1.192	1.139	1.249	<0.001
41-60% Maximum Stenosis of Left Circumflex Artery (vs. 81-100%)	0.907	0.876	0.938	<0.001
61-80% Maximum Stenosis of Left Circumflex Artery (vs. 81-100%)	0.951	0.923	0.980	<0.001
Interventional Catheterization Not Used in Procedure (vs. Used)	1.320	1.284	1.358	<0.001
Third Heart Sound (vs. none)	1.241	1.187	1.297	<0.001
Elevated Blood Pressure (vs. Normal)	0.956	0.924	0.990	0.011
Hypertensive Crisis Blood Pressure (vs. Normal)	1.102	1.017	1.195	0.017
Missing Record Blood Pressure (vs. Normal)	0.952	0.891	1.019	0.155
Stage 1 Hypertension Blood Pressure (vs. Normal)	0.960	0.933	0.987	0.004
Stage 2 Hypertension Blood Pressure (vs. Normal)	0.972	0.948	0.997	0.029
NA	1.616	1.136	2.300	0.008
NA	0.932	0.550	1.579	0.793
NA	NA	NA	NA	NA

At the $\alpha = 0.05$ significance level, here are interesting findings from the results:

For the in-depth description of "holding all other predictors constant, see Appendix (Baseline Metrics)

Compared to patients who received interventional catheterizations, the hazard of death is expected to be multiplied by 1.310 (95% CI: 1.273, 1.347) for patients who didn't receive interventional catheterizations, holding all other predictors constant.

The coefficients for many demographic and physical exam variables are also significant at this level. For instance, the hazard of death for patients over 80 years old is expected to be multiplied by 2.530 (95% CI: 2.432, 2.632) compared to patients between 65 and 69 years old, holding all other predictors constant. Compared to Caucasian patients, the hazard of death is expected to be multiplied by 1.198 (95% CI: 1.166, 1.231) for African American patients, holding all other factors constant. For male patients, for every increase in one point from the average BMI, the hazard of death is expected to be multiplied by 0.997 (95% CI: 0.994, 0.999), adjusting for all other factors. For female patients, for every increase in one point from the average BMI, the hazard of death is expected to be multiplied by 1.040 (95% CI: 1.016, 1.064), adjusting for all other factors. However, this result is not significant at the 0.05 level. Compared to patients with no third heart sound, the hazard of death is expected to be multiplied by 1.239 (95% CI: 1.185, 1.295) for patients with a third heart sound, holding all other factors constant.

For patient history, we found similar patterns under $\alpha = 0.05$ significance level. Patients with mild Congestive Heart Failure (CHF) is expected to have a higher estimated hazard of death compared to non-CHF patient, as the hazard ratio is roughly multiplied by 1.403 (95% CI: 1.343, 1.467), adjusting for all other factors. As the severity of CHF increases, for example, to the highest classified level (Class IV), increased hazard of death is estimated with baseline multiplied by 2.081 (95% CI: 1.996, 2.169), holding all other factors constant. Moreover, for every increase in one previous MI the patient had experienced, the risk of death is expected to be multiplied by 1.107 (95% CI: 1.095, 1.119), controlling for all other factors. Similar increased hazard of death compared to baseline has also been observed in patients with a history of Smoking, Cerebrovascular Disease, Chronic Obstructive Pulmonary Disease, Diabetes, and Hypertension, with all other factors controlled. For patients presenting ST-Elevation Myocardial Infarction (STEMI) and Unstable Angina upon visits, the hazard ratio is not being greatly impacted. Both factor estimates reside around 1, holding all other factors constant. This suggests that the syndrome check upon presentation is making a smaller contribution to the difference in the hazard of death after adjusting for other factors. Another interesting output of the model is that with each additional sequential catheterization procedure, the hazard of death is expected to be multiplied by 0.988 (less than 1) after controlling for all other factors. However, it is important to note that an increased number of sequential catheterizations may also indicate increased access to health care–although there is no evidence for this it is an interesting aspect to consider.

For vital signs variables, increasing levels of heart rate, in general, signifies higher hazard of death than baseline heart rate of 60-100 bpm (consider a "normal" resting heart rate). For example, heart rate of 101-120 bpm has an estimated hazard ratio of baseline multiplied by 1.389 (95% CI: 1.325, 1.457), holding all other factors constant. And heart rate of 121-140 bpm is expected to be the baseline hazard multiplied by 1.537 (95% CI: 1.380, 1.712), controlling for all other factors. Surprisingly, blood pressure variables have a hazard risk of death slightly less than 1 compared to baseline, holding all other factors constant.

For catheterization results, we also find the pattern in regards to the hazard of death. Take maximum stenosis of left main artery as an example: compared to the baseline of 0-21% maximum stenosis and adjusting for all other factors, patients with 41-60% and 61-80% have an increased estimated hazard ratio by a factor of 1.166 and 1.194, respectively. If we set the baseline as 81-100% maximum stenosis of left circumflex artery (range with highest number of observations in the dataset) and holding all other factors constant, patients with 41-60% and 61-80% have a lower estimated hazard ratio by a factor of 0.907 and 0.951, respectively.

In general, most factors show promising results as either confirming the effect of interventional catheterization or informing the significance of the variables. However, there are a small portion of factors that generates counter-intuitive results. For example, for both men and women, the coefficient for BMI suggests that with each additional point BMI, the risk of death is expected to be multiplied by a factor less than 1 after controlling for all other factors. As higher than average BMIs often indicate poor health and espically poor heart health according to a John Hopkins study, this finding is surprising.

Limitations and Next Step

A significant limitation to this analysis lies in the missingness of the original data set. Missingness in certain catheterization results such as the maximum stenosis percents in various regions of the heart or found in a bypass graft were unexplained by the original data description. Without more information, it is impossible to know if this missingess was due to the fact that these regions were not explored by the catheterization procedure or if the doctor decided that the respective area was not important to analyze given the status of a patient's disease or symptoms.For future studies, it will be important to add a variable recording if a certain region of the heart was not documented/recorded intentionally or not. This will allow for a better understanding of a patient's condition, and in turn, their medical needs. Furthermore, as addressed throughout the paper and illustrated in Figure 8A, many missing observations are associated with records from 1985-1990. Such missingess may suggest the potential of a confounding factor in the analysis related the reliability and/or accuracy of results from this period. Further research may want to verify information from this time period or gather more recent data instead.

In addition, data set only include patients at Duke Hospital, which limits the study result to Durham or places with similar demographic made up. For example, the patient race variable only has "Caucasian," "African American," "Other," and "Unknown" levels. Furthermore, over 80% of the observations are Caucasian patients. The lack of observations and specification for other races in the data set limits the generalizability of our results. In order to create the BMI variable, we imputed around 800 observations for both patient height and weight using the MICE algorithm on a subset of the data that included patient height, weight, age, gender, and race. This method introduced bias into the analysis, as we assumed that the data for patient height and weight was missing at random and created the subset based on attributes we predicted to be associated with height and weight.

Furthermore, because data set violate the proportional hazard assumption for Cox model, the model coefficient can only interpret as the "average effect" over time points. To address such violation, we can use two methods: (i) introducing interactions of selected covariates with function of time, (ii) stratification model. For further analysis on the study, we are looking into increase the scope of the study to increase generalizability, modified Cox model to address non-proportional hazard, and verify missing information from the data set.

Discussion

When controlling for other patient/procedure characteristics, what is the association between patient survival and the use of interventional coronary catheterizations?

We see an association between interventional coronary catheterization procedure with higher survival probability when controlling for other factors (see appendix for baseline metrics), the hazard rate is 1.32 times higher compared to patients who don't have interventional coronary catherization procedure. In the face of insufficient availabiity of interventional catheterizations (as mentioned in the introduction), such finding provides evidence for doctors and/or health are providers to seek out means to deliver more effective and reliable interventional catheterization procedures.

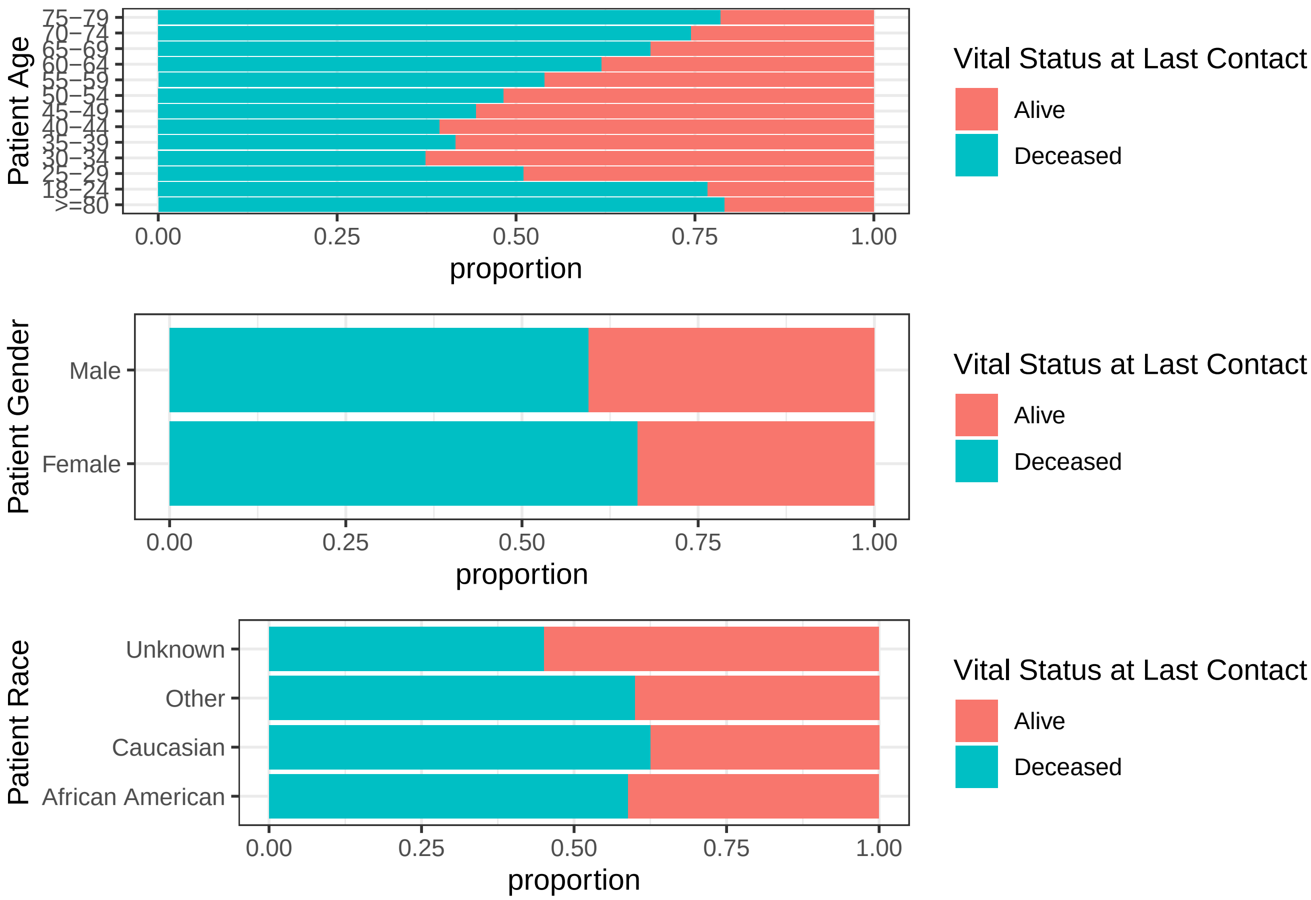
In regards to patient data that a physician can access on the day of a given procedure, what factors show evidence for an association with patient survival?

We believe our results may help doctors in informing patients and their families of the risk of procedures (like interventional catheterizations) given characteristics about the patient that are obtainable by a doctor. For example, given the patient is treated with interventional coronary catherization, patients within higher age group, more severe

Congestive Heart Failure, History of smoking, diabetes, Cerebrovascular Disease, and Chronic Obstructive Pulmonary Disease have higher hazard rate. Among patients history, patients with diabetes have the highest hazard rate (1.587 times those without diabetes). During the procedure, patients with higher percentage of maximum stenosis in all artery (left anterior, left circumflex, left main, right coronary) have higher hazard rate. As for patients' vital sign, patients who have third heart sound and high heart rate are associated with higher hazard rate. These are just some of the take-aways from this analysis

Appendix

Fig 1A: Breakdown of Vital Status and Patient Demographics



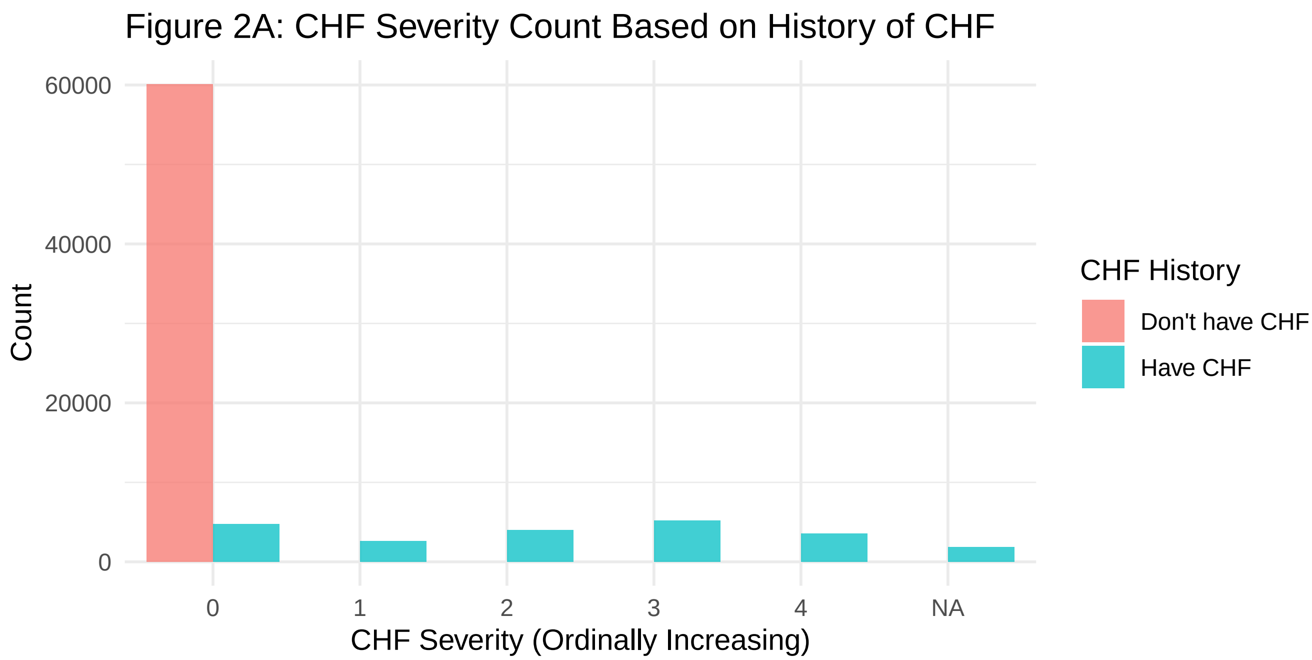


Figure 2A: CHF Severity Count Based on History of CHF
Count
60000
40000
20000
0
0
1
2
3
4
NA
CHF Severity (Ordinally Increasing)
CHF History
Don't have CHF
Have CHF

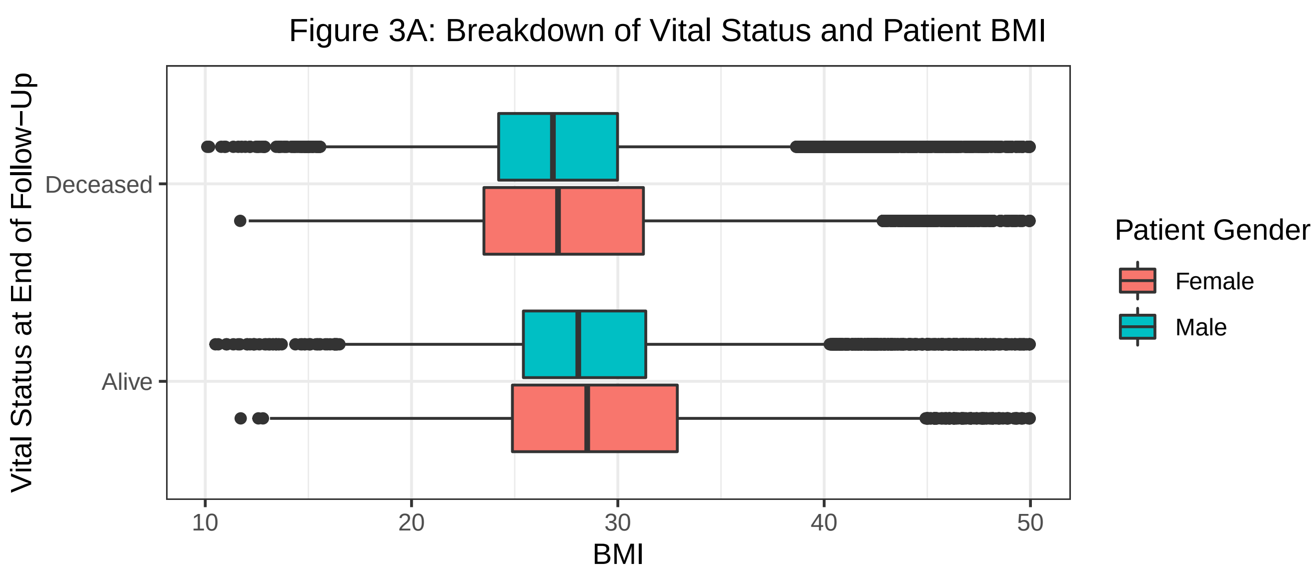


Figure 3A: Breakdown of Vital Status and Patient BMI
Vital Status at End of Follow−Up
Deceased
Alive
10
20
30
40
50
BMI
Patient Gender
Female
Male

Figure 4A: Maximum Percent Stenosis by Heart Region

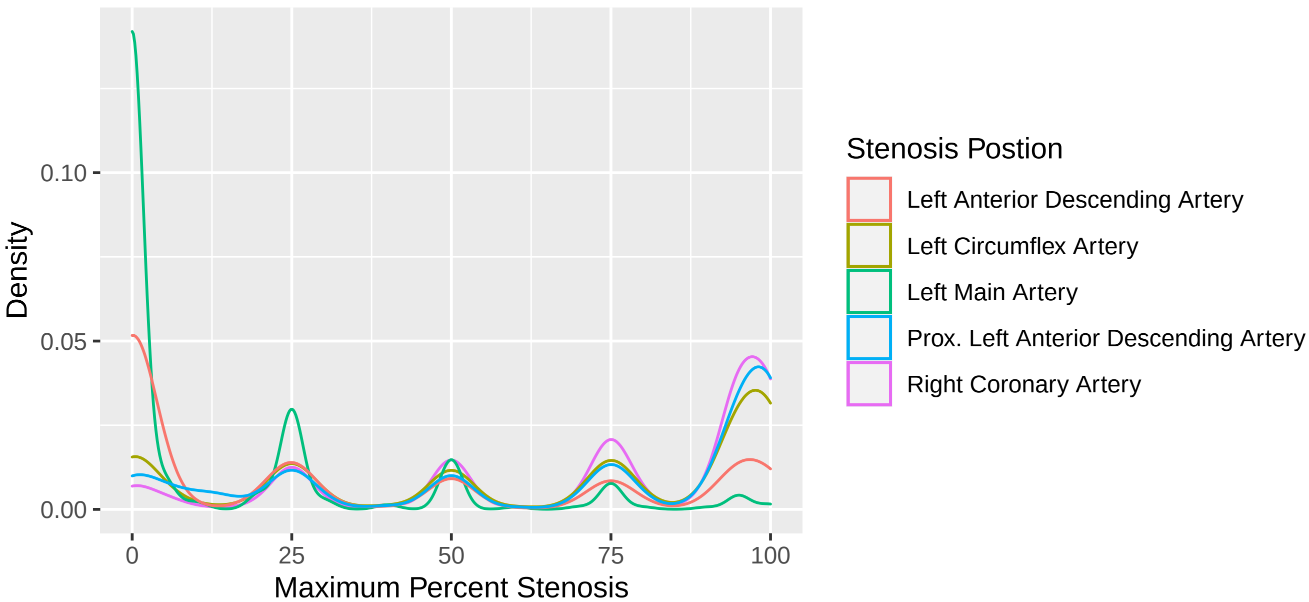
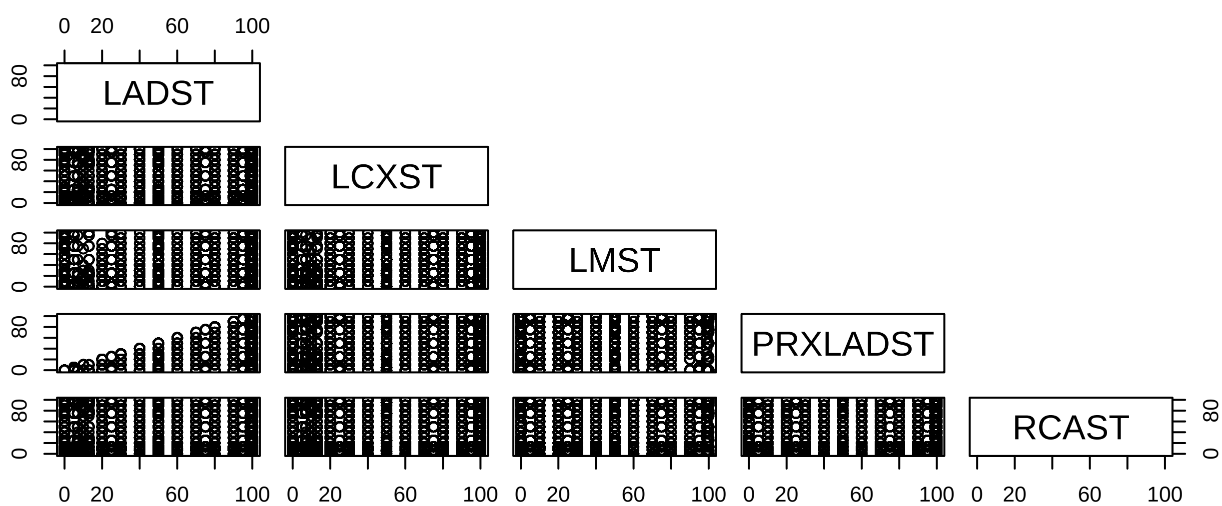


Figure 5A: Relation of Maximum Stenosis by Heart Region



6A: Left Ventrical Ejection Fraction by Associated Patient Vital Status

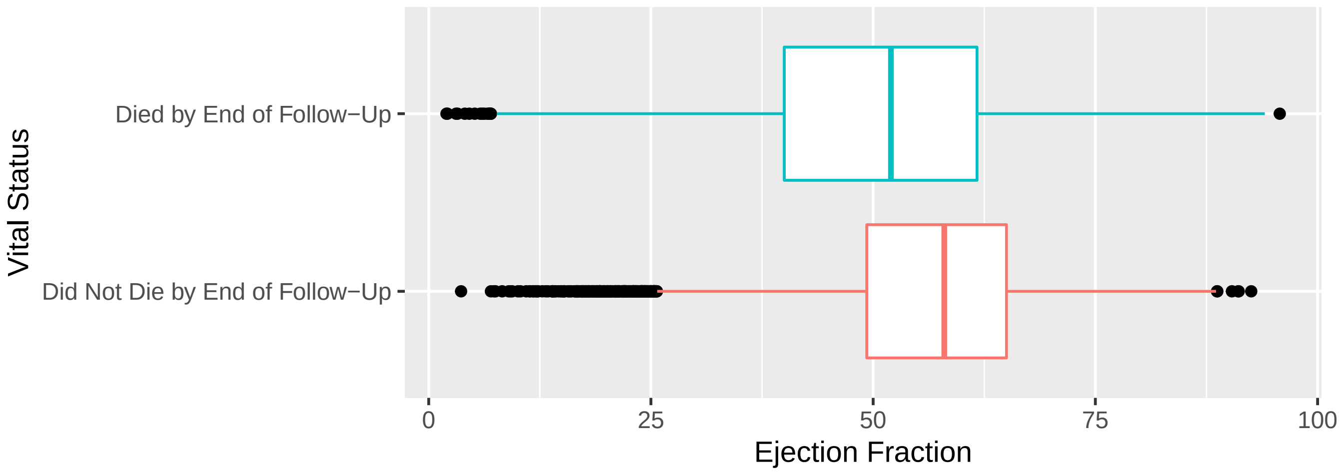


Figure 7A: Residual Plot for Cox Model

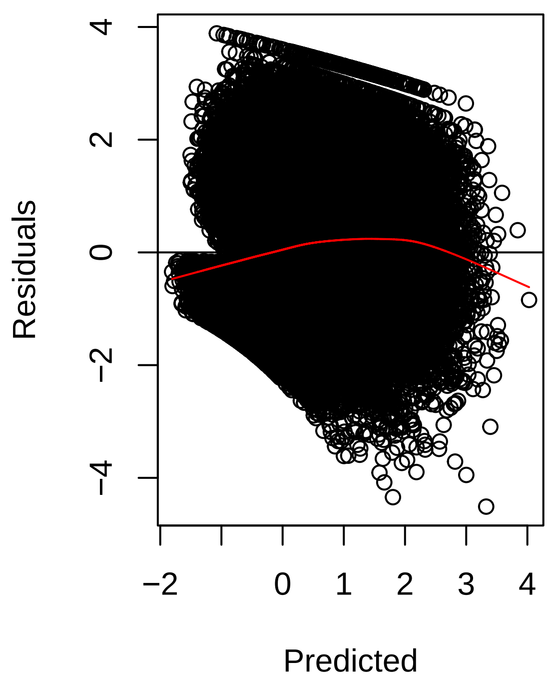
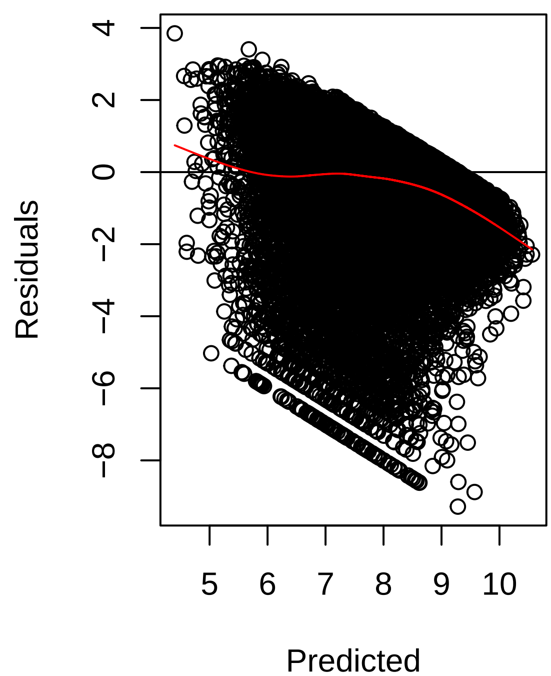
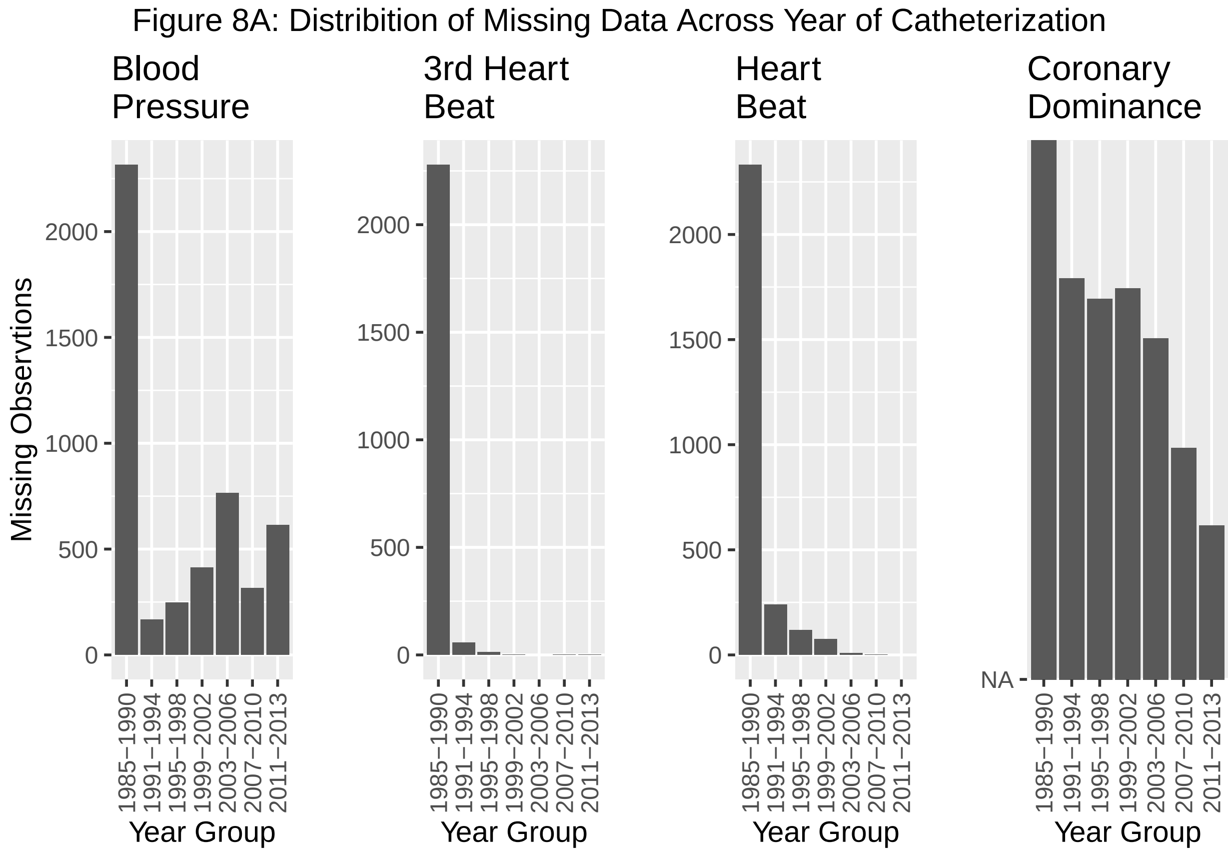


Figure 7B: Residual Plot for Buckley−James Model





Removed Variables

We omitted three variables on catheterization identification and linking: days from first catheterization, subject ID, and year of catheterization. We omitted the days from first catheterization because current research on the subject does not suggest any evidence that time passed between catheterizations affects patient survival. We also removed subject ID from the analysis because we don't have any reason to believe that there are prominent differences between each patient that aren't already accounted for by the included demographic, patient history, vital sign, physical exam, and catheterization results variables. Lastly, the year of the catheterization was omitted from analysis because cardiac catheterization methods have not changed greatly between 1985-2013, which is the range included in the data set.

We removed four variables in the patient history section: We substitute the Acute Coronary Syndrome(ACS) Status Upon Presentation variable with two recoded variables STEMI and Unstable Angina based on variable selection discussion. We choose the number of previous Myocardial Infarctions (MIs) over the history of MIs and days to the closest previous MI because it can both indicate whether a patient has a history of MI and capture the frequency of the patient experiencing MIs based on the correlation analysis in previous discussion. For days to the closest previous Coronary Artery Bypass Surgery, only 25.6% of the observations have valid values and only 21.4% of all unique patients have received the surgery. Thus, we decide to remove this variable due to high proportion of missing information and insufficient accountability.

We removed all lab result data from this analysis. Because each lab result can be anywhere from the same day as the catheterization to a year prior to the catheterization, there is no way of differentiating between the relevancy of these results for a given procedure. Additionally, a doctor may not have access to such records/such records may not exist on the day of a given procedure. Therefore, for the purpose of this analysis which focuses on data a doctor may collect the same-day as a procedure, we will not include variables associated with lab results in our analysis.

Since our research question focuses on interventional catheterizations, the other two catheterization procedure variables, type of catheterization (e.g. left or right heart) and diagnostic catheterizations, were omitted from our analysis.

We omitted the follow-up variables in the dataset, which include all variables indicating days to first subsequent medical event or surgery and whether the patient is on follow-up protocol. These variables were omitted because our analysis is aimed at physicians trying to determine whether an interventional catheterization will be worthwhile for their patient. Thus, it is not informative to include any variables that provide information on events occurring after the procedure or visit as this data can only be gathered retroactively.

Baseline Metrics for Model

- When the procedure involved an interventional catheterization.
- When the patient is in the age range of 65-69.
- When the patient is Caucasian.
- When the patient has the average BMI of 28.3865052.
- When the patient is male.
- When it's the patient's first catheterization procedure.
- When ST-Elevation Myocardial Infarction is not present.
- When unstable angina is not present.
- When there are no symptoms of congestive heart failure.
- When there is no records of any pervious MIs.
- When the patient has no history of smoking.
- When the patient has no history of cerebrovascular disease.
- When the patient has no history of chronic obstructive pulmonary disease.
- When the patient has no history of diabetes.
- When the patient has no history of hypertension.
- When the patient has no history of hyperlipidemia.
- When the patient has no history of anginal pain.
- When the procedure did not have a artery bypass graft.
- When the maximum stenosis of the left main artery is 0-21%.
- When the maximum stenosis of the left circumflex artery is 81-100%.
- When the maximum stenosis of the left anterior descending artery is 81-100%.
- When the maximum stenosis of the right coronary artery is 81-100%.
- When there is normal left ventricular ejection.
- When there is right coronary dominance.
- When there is no third heart sound.
- When blood pressure is normal.
- When the heart rate is 60-100 bpm.

Assumptions for Baselines in Model

- Set the Interventional Catheterization baseline to used so that we may look at other variables in this context.
- Set age group 65-69 as baseline because it had the most observations.
- Set caucasion race as baseline since most patients were caucasion.
- Congestive Heart Failure (CHF) Classes (by New York Heart Association (NYHA) classification):
  - Class None: Patient has no CHF.
  - Class I: Patient has cardiac disease but without resulting limitations of ordinary physical activity (e.g., walking one to two level blocks or climbing one flight of stairs). Transient CHF sx (Class V) has been converted to Class I.
  - Class II: Patient has cardiac disease resulting in slight limitation of ordinary physical activity. Patient is comfortable at rest. Ordinary physical activity such as walking more than two blocks or climbing more than one flight of stairs results in limiting symptoms.
  - Class III: Patient has cardiac disease resulting in marked limitation of physical activity. Patient is comfortable at rest. Less than ordinary physical activity causes fatigue, palpitation, dyspnea, or anginal pain.
  - Class IV: Patient has symptoms at rest that increase with any physical activity. Patient has cardiac disease resulting in inability to perform any physical activity without discomfort. Symptoms may be present even at rest. If any physical activity is undertaken, discomfort is increased.
  - Class Missing: No CHF class listed or unable to determine.
- Set baseline of Maximum Stenosis of the Left Main Artery to 21%-40% as that range was the most common within the data set.
- Set baseline of Maximum Stenosis of the Left Anterior Descending Artery Region to 81%-100% as that range was the most common within the data set.
- Set baseline of Maximum Stenosis of the Right Coronary Artery to 81%-100% as that range was the most common within the data set.
- Set baseline of Maximum Stenosis of the Left Circumflex Artery to 81%-100% as that range was the most common within the data set.
- Set baseline of Left Ventrical Ejection Fraction to the Normal range category.
- Set Coronary Dominace baseline to Right because it is observed the most within the dataset.

- Set the Blood Pressure baseline to Normal to compare across categories.
- Set the heart rate (pulse) baseline to 60-100 bpm as that range in considered a “normal” range for adults.

Model Assumptions

Table 3: Proportional Hazard Test

Variable	chisq	df	P-Value
Age Group	811.761	12	<0.001
Race	134.240	3	<0.001
Global	38.674	1	<0.001
Gender	44.221	1	<0.001
Patient’s Sequential Cath Number	18.109	1	<0.001
ST-Elevation Myocardial Infarction Present	164.124	1	<0.001
Unstable Angina Present	131.551	1	<0.001
Class of Congestive Heart Failure	251.019	4	<0.001
Number of Previous MIs	59.317	1	<0.001
History of Smoking	3.169	1	0.075
History of Cerebrovascular Disease	23.371	1	<0.001
History of Chronic Obstructive Pulmonary Disease	0.399	1	0.528
History of Diabetes	30.174	1	<0.001
History of Hypertension	17.525	1	<0.001
History of Hyperlipidemia	98.576	1	<0.001
History of Anginal Pain	137.785	1	<0.001
Had a Coronary Artery Bypass Graft	18.751	1	<0.001
Category of Percent Maximum Stenosis of the Left Main Artery	40.720	5	<0.001
Category of Percent Maximum Stenosis of Left Circumflex Artery	16.639	5	0.005
Category of Maximum Stenosis of Left Anterior Descending Artery	33.907	5	<0.001
Category of Maximum Stenosis of Right Coronary Artery	60.906	5	<0.001
Category of Left Ventricular Ejection Fraction	321.882	3	<0.001
Coronary Dominace	17.226	2	<0.001
Interventional Catheterization Used in Procedure	17.114	1	<0.001
Third Heart Sound	147.933	2	<0.001
Blood Pressure Status	349.455	5	<0.001
Heart Rate bpm Category	242.364	6	<0.001
Global	9.730	1	0.002
Global	2434.013	73	<0.001

Cox Model Output

Table 4: Cox Model Output (Full)

Variable	Hazard Ratio	Confidence (2.5%)	Confidence (97.5%)	P-Value	Significant
Age 18-24 (vs. Age 65-69)	1.204	0.846	1.715	0.303	No
Age 25-29 (vs. Age 65-69)	0.544	0.426	0.695	<0.001	Yes
Age 30-34 (vs. Age 65-69)	0.328	0.285	0.377	<0.001	Yes
Age 35-39 (vs. Age 65-69)	0.383	0.355	0.413	<0.001	Yes
Age 40-44 (vs. Age 65-69)	0.352	0.332	0.372	<0.001	Yes
Age 45-49 (vs. Age 65-69)	0.412	0.394	0.430	<0.001	Yes
Age 50-54 (vs. Age 65-69)	0.477	0.460	0.495	<0.001	Yes
Age 55-59 (vs. Age 65-69)	0.584	0.564	0.604	<0.001	Yes
Age 60-64 (vs. Age 65-69)	0.761	0.737	0.785	<0.001	Yes
Age 70-74 (vs. Age 65-69)	1.327	1.288	1.368	<0.001	Yes
Age 75-79 (vs. Age 65-69)	1.764	1.705	1.825	<0.001	Yes
Age 80+ (vs. Age 65-69)	2.528	2.430	2.630	<0.001	Yes

Variable	Hazard Ratio	Confidence (2.5%)	Confidence (97.5%)	P-Value	Significant
Race Unknown (vs. Caucasian)	0.904	0.829	0.986	0.022	Yes
African American (vs. Caucasian)	1.194	1.163	1.227	<0.001	Yes
Other Race (vs. Caucasian)	1.157	1.106	1.211	<0.001	Yes
BMI Points Away from Mean	0.996	0.994	0.999	0.001	Yes
Female (vs. Male)	1.049	1.028	1.071	<0.001	Yes
Patient's Sequential Cath Number	0.987	0.983	0.991	<0.001	Yes
ST-Elevation Myocardial Infarction Present (vs. not present)	0.962	0.937	0.988	0.004	Yes
Unstable Angina Present (vs. not present)	0.973	0.953	0.994	0.013	Yes
Class I Congestive Heart Failure (vs. none)	1.404	1.344	1.468	<0.001	Yes
Class II Congestive Heart Failure (vs. none)	1.410	1.356	1.467	<0.001	Yes
Class III Congestive Heart Failure (vs. none)	1.655	1.597	1.714	<0.001	Yes
Class IV Congestive Heart Failure (vs. none)	2.087	2.002	2.176	<0.001	Yes
Number of Previous MIs	1.108	1.096	1.120	<0.001	Yes
History of Smoking (vs. none)	1.201	1.177	1.226	<0.001	Yes
History of Cerebrovascular Disease (vs. none)	1.377	1.342	1.413	<0.001	Yes
History of Chronic Obstructive Pulmonary Disease (vs. none)	1.526	1.476	1.578	<0.001	Yes
History of Diabetes (vs. none)	1.587	1.555	1.619	<0.001	Yes
History of History of Hypertension (vs. none)	1.159	1.135	1.183	<0.001	Yes
History of Hyperlipidemia (vs. none)	0.825	0.809	0.841	<0.001	Yes
History of Anginal Pain (vs. none)	0.889	0.863	0.916	<0.001	Yes
Had a Coronary Artery Bypass Graft (vs. did not)	0.944	0.919	0.969	<0.001	Yes
Missing Maximum Stenosis of the Left Main Artery (vs. 0-21%)	1.308	1.239	1.381	<0.001	Yes
21-40% Maximum Stenosis of the Left Main Artery (vs. 0-21%)	1.016	0.988	1.044	0.271	No
41-60% Maximum Stenosis of the Left Main Artery (vs. 0-21%)	1.165	1.123	1.209	<0.001	Yes
61-80% Maximum Stenosis of the Left Main Artery (vs. 0-21%	1.192	1.139	1.249	<0.001	Yes
81-100% Maximum Stenosis of the Left Main Artery (vs. 0-21%)	1.371	1.299	1.446	<0.001	Yes
Missing Maximum Stenosis of Left Circumflex Artery (vs. 81-100%)	0.756	0.726	0.788	<0.001	Yes
0-20% Maximum Stenosis of Left Circumflex Artery (vs. 81-100%)	0.836	0.810	0.863	<0.001	Yes
21-40% Maximum Stenosis of Left Circumflex Artery (vs. 81-100%)	0.881	0.852	0.911	<0.001	Yes
41-60% Maximum Stenosis of Left Circumflex Artery (vs. 81-100%)	0.907	0.876	0.938	<0.001	Yes
61-80% Maximum Stenosis of Left Circumflex Artery (vs. 81-100%)	0.951	0.923	0.980	<0.001	Yes
Missing Maximum Stenosis of Left Anterior Descending Artery (vs. 81-100%)	0.862	0.825	0.900	<0.001	Yes
0-20% Maximum Stenosis of Left Anterior Descending Artery (vs. 81-100%)	0.911	0.875	0.948	<0.001	Yes
21-40% Maximum Stenosis of Left Anterior Descending Artery (vs. 81-100%)	0.931	0.899	0.965	<0.001	Yes
41-60% Maximum Stenosis of Left Anterior Descending Artery (vs. 81-100%)	0.961	0.930	0.993	0.018	Yes
61-80% Maximum Stenosis of Left Anterior Descending Artery (vs. 81-100%)	0.986	0.959	1.013	0.304	No
Missing Maximum Stenosis of Right Coronary Artery (vs. 81-100%)	0.947	0.917	0.978	<0.001	Yes

Variable	Hazard Ratio	Confidence (2.5%)	Confidence (97.5%)	P-Value	Significant
0-20% Maximum Stenosis of Right Coronary Artery (vs. 81-100%)	0.855	0.824	0.886	<0.001	Yes
21-40% Maximum Stenosis of Right Coronary Artery (vs. 81-100%)	0.868	0.838	0.899	<0.001	Yes
41-60% Maximum Stenosis of Right Coronary Artery (vs. 81-100%)	0.906	0.873	0.939	<0.001	Yes
61-80% Maximum Stenosis ofRight Coronary Artery (vs. 81-100%)	0.944	0.915	0.973	<0.001	Yes
Abnormal Left Ventricular Ejection Fraction (vs. Normal)	1.555	1.511	1.599	<0.001	Yes
Borderline Abnormal Left Ventricular Ejection Fraction (vs. Normal)	1.204	1.168	1.241	<0.001	Yes
Missing Left Ventricular Ejection Fraction (vs. Normal)	1.577	1.527	1.628	<0.001	Yes
Left Coronary Dominace (vs. Right)	1.169	1.124	1.215	<0.001	Yes
Balanced Coronary Dominace (vs. Right)	1.091	1.027	1.160	0.005	Yes
Interventional Catheterization Not Used in Procedure (vs. Used)	1.320	1.284	1.358	<0.001	Yes
Third Heart Sound (vs. none)	1.241	1.187	1.297	<0.001	Yes
Missing Record Third Heart Sound (vs. none)	0.912	0.812	1.024	0.120	No
Elevated Blood Pressure (vs. Normal)	0.956	0.924	0.990	0.011	Yes
Hypertensive Crisis Blood Pressure (vs. Normal)	1.102	1.017	1.195	0.017	Yes
Missing Record Blood Pressure (vs. Normal)	0.952	0.891	1.019	0.155	No
Stage 1 Hypertension Blood Pressure (vs. Normal)	0.960	0.933	0.987	0.004	Yes
Stage 2 Hypertension Blood Pressure (vs. Normal)	0.972	0.948	0.997	0.029	Yes
Heart Rate 0-49 bpm (vs. 50-99 bpm)	0.836	0.788	0.888	<0.001	Yes
Heart Rate 100-149 bpm (vs. 50-99 bpm)	1.439	1.384	1.497	<0.001	Yes
Heart Rate 150-199 bpm (vs. 50-99 bpm)	1.350	1.114	1.636	0.002	Yes
Heart Rate 200-249 bpm (vs. 50-99 bpm)	1.352	0.955	1.914	0.089	No
Heart Rate 250-299 bpm (vs. 50-99 bpm)	0.779	0.461	1.317	0.351	No
Missing Record Heart Rate (vs. 50-99 bpm)	0.935	0.832	1.052	0.263	No
(BMI Points Away from Mean)*Female	0.998	0.995	1.002	0.338	No

Sources

https://www.jacc.org/doi/10.1016/j.jacc.2012.08.966

https://www.ncbi.nlm.nih.gov/books/NBK531461/

https://www.mayoclinic.org/diseases-conditions/acute-coronary-syndrome/symptoms-causes/syc-20352136

https://www.healthline.com/health/diastole-vs-systole#TOC_TITLE_HDR_1

https://www.nhs.uk/conditions/coronary-artery-bypass-graft-cabg/

https://www.mayoclinic.org/tests-procedures/ekg/expert-answers/ejection-fraction/faq-20058286

https://manual.jointcommission.org/releases/TJC2016A/DataElem0439.html

https://www.mayoclinic.org/healthy-lifestyle/fitness/expert-answers/heart-rate/faq-20057979 #:~:text=A%20normal%20resting%20heart%20rate%20for%20adults%20ranges%20from%2060,to%2040%20beats%20per%20minute

https://www.mayoclinic.org/tests-procedures/cardiac-catheterization/about/pac-20384695#:~:text=Cardiac%20catheterization%20

https://www.ahajournals.org/doi/10.1161/CIRCULATIONAHA.105.592428

https://www.chp.edu/our-services/heart/patient-procedures/interventional-catheterization#:~:text=Some%20catheters%20allow%2