Investigating the Association between Duration of Chest Pain Symptoms and the Presence and Severity of Coronary Artery Disease based on a Cardiac Catheterization Procedure

Thu Vu

BIOST 579

Abstract

The primary objective is to investigate the associations between the duration of chest pain symptoms and the presence and severity of coronary artery disease which are defined as Significant and Severe Coronary Disease while adjusting for cholesterol level, sex, and age. Based on the primary objective, the secondary objective is to determine if cholesterol level is an effect modifier which cholesterol level will be dichotomized into two levels, high and otherwise. The target population for both of these objectives are adults aged 20 years and older who are experiencing chest-related pain symptoms. The main statistical methods used are descriptive statistics and multivariate logistic regression. Relevant point estimates and confidence intervals using robust standard errors were examined to answer the above objectives. Statistical significance was determined based on p-values derived from likelihood ratio tests with a significance level of 0.05. The results for the primary analysis found that there is not a significant association between having Significant Coronary Disease and the duration of chest pain symptoms while there is a significant association between having Severe Coronary Disease and the duration of chest pain symptoms adjusting for the other variables. Results for the secondary analysis show that high cholesterol is likely not an effect modifier and is likely a confounder of coronary artery disease instead. The results of the primary analysis might be plausible as patients with Severe Coronary Disease experience more extreme and chronic symptoms compared to patients with Significant Coronary Disease who may be experiencing varying intensity levels of symptoms in which duration fluctuates. The dataset used for analysis comes from an observational study, therefore, no causal relationships are implied. However, based on these results, further investigation with specific recordings of duration, severity and types of coronary disease symptoms and other possible confounders might be worth conducting.

Introduction

In the United States, heart disease is the leading cause of death across most groups and coronary artery disease is the most common type of heart disease. Coronary Artery Disease (CAD) is a condition where plaque builds up within the coronary arteries possibly leading to chest pain and/or a heart attack. Risk factors include age, sex, and high blood cholesterol levels. One common procedure to diagnose or treat CAD is a cardiac catheterization (or cardiac cath) in which a catheter is guided through a blood vessel to the heart to examine the condition of the coronary arteries. There are varying severity levels of CAD. Three Vessel Disease and Left Main Disease are examples of severe forms of CAD and both are associated with high risk of mortality.

Based on the severity level of CAD an individual has, they can experience differing types, magnitude, and duration of symptoms. These symptoms include chest-related pain, fatigue and heart attack. Therefore, one of the main goals that will be addressed is to investigate if there is an association between the severity of CAD and specifically, the duration of chest-related pain symptoms.

Dataset Background

Patients were referred to Duke University Medical Center for chest pain symptoms who were then evaluated with cardiac catheterization. Chest pain symptoms include heart-related symptoms such as shortness of breath, cold sweats, dizziness, pressure/tightness in the chest, searing pain that spreads to arms, etc. The findings of the procedure were summarized as having Significant Coronary Disease or not then determining if it's Three Vessel Disease or Left Main Disease or neither. Significant Coronary Disease means there's a blockage of at least 75% in at least one of the major coronary arteries. Three Vessel Disease means having blockages of all three of the left anterior descending (LAD) coronary artery, the left circumflex coronary artery (LCA), and the right coronary artery (RCA). Left Main Disease means there's a blockage of the left main coronary artery. This is an observational cross-sectional study with 3504 participants. It contains 6 variables with 1249 missing observations:

- sex (0 = Male, 1 = Female)
- age (years)
- cad_dur (days) — duration of chest pain symptoms
- choleste (milligrams per deciliter) — serum cholesterol level
- sigdz (0 = No, 1 = Yes) — Significant Coronary Disease
- tvdlm (0 = No, 1 = Yes) — Severe Coronary Disease (Either Three Vessel Disease or Left Main Disease)

Data obtained from http://hbiostat.org/data courtesy of the Vanderbilt University Department of Biostatistics (Harrell Jr (2002)). Dataset is from the Duke University Cardiovascular Disease Databank.

Scientific Questions

Primary Objective

Among adults aged 20 years and older who are experiencing chest pain symptoms, the main questions that will be addressed are the following:

1. Is having Significant Coronary Disease associated with the duration of chest pain symptoms when controlling for sex, age, and cholesterol levels?

- How does the odds of having Significant Coronary Disease change for every additional day an individual experiences chest pain symptoms?

2. Given that an individual has Significant Coronary Disease, is having severe forms of CAD (Three Vessel Disease or Left Main Disease) associated with the duration of chest pain symptoms when controlling for sex, age, and cholesterol levels?

- Given that an individual has Significant Coronary Disease, how does the odds of having severe forms of CAD (Three Vessel Disease or Left Main Disease) change for every additional day an individual experiences chest pain symptoms?

Secondary Objective

The additional questions that will be addressed are the following:

3. Does the association between having Significant Coronary Disease and the duration of chest pain symptoms when adjusting for sex and age differ for differing values of cholesterol levels, high levels versus low levels?
4. Given that an individual has Significant Coronary Disease, does the association between having severe forms of CAD (Three Vessel Disease or Left Main Disease) and the duration of chest pain symptoms when adjusting for sex and age differ for differing values of cholesterol levels, high levels versus low levels?

Statistical Methods

The main methods we plan to use to address the scientific questions are: multivariate logistic regression for analysis and hypothesis testing for inference. For questions 1 and 3, the response variable will be the indicator variable for Significant Coronary Disease, *sigdz*. For questions 2 and 4, the response variable will be the indicator variable for Severe Coronary Disease, *tvdlm*. For all questions, the covariates will be: *cad_dur*, *sex*, *age*, and *choleste*. Addition to these covariates, for questions 3 and 4, there will be an interaction term between *cad_dur* and *choleste* as a dichotomized variable, high and low levels of serum cholesterol.

The software for data analysis that will be used is R Version 4.1.2.

The target population for the scientific questions will be adults aged 20 years and older who are experiencing chest pain symptoms. Hence, observations for participants younger than 20 years old will be omitted. For both objectives, descriptive statistics will be computed for each variable with the observed data. Mean, median, standard deviation, maximum and minimum will be reported for the continuous variables while proportions/percentages will be reported for the categorical variables. Additionally, a histogram will be generated to visualize the covariate of interest, duration of chest pain symptoms. Based on this plot, the covariate will be data-transformed as necessary. For the primary objective, box-plots will be presented to visualize the comparison of having coronary disease and the duration of chest pain. For the secondary objective, these plots will be stratified by cholesterol levels. Based on the descriptive statistics, independent variables with a large amount of missing data will be plotted against each quantitative covariate to identify patterns while for categorical covariates, proportions/percentages will be reported. If identified as missing at random, missing data will be handled by regression imputation. If identified as missing not at random, depending on the specific patterns, the appropriate imputation methods will be implemented.

The primary objective is to investigate the association between the duration of chest pain symptoms and the presence and severity of CAD. To accomplish this, multivariate logistic regression will be used to fit each model. The response variable will indicate the presence and severity of CAD. The potential confounders that will be adjusted for in each model are: sex, age, and cholesterol level. The baselines used are: 0 days for duration of chest pain symptoms, males are the reference group for sex, 20 years old for age, and 0 mg/dl for cholesterol level. Relevant point estimates and confidence intervals using robust standard errors will be reported. Statistical significance will be determined based on p-values derived from likelihood ratio tests with a significance level of 0.05.

The secondary objective is to determine if cholesterol level is an effect modifier. Similar to above, multivariate logistic regression will be used with the same response variable and baselines. Relevant point estimates, confidence intervals, and p-values will also be reported. Patients with cholesterol greater than or equal to 239 mg/dl are considered high levels. Therefore, cholesterol will be dichotomized at 239 mg/dl and will be an indicator variable for high cholesterol levels or not (low cholesterol levels). Thus, addition to the covariates, age and sex, there will be an interaction term between duration of chest pain symptoms and high cholesterol level.

Model 1: To examine the association between Significant Coronary Disease and duration of chest pain symptoms, multivariate logistic regression will be used with a response variable indicating presence of Significant Coronary Disease or not. The covariates that will be adjusted for are: age (age minus 20 as baseline is 20 years old), sex, and cholesterol level. A likelihood ratio test will be used to test for significant association between Significant Coronary Disease and duration of chest pain symptoms.

Model 2: If an individual has Severe Coronary Disease, then they also have Significant Coronary Disease. Therefore, a subset of the data which only include patients with Significant Coronary Disease will be used to examine the association between Severe Coronary Disease and duration of chest symptoms. Multivariate logistic regression will be used with a response variable indicating presence of Severe Coronary Disease or not. Similar to Model 1, the same

covariates will be adjusted for and a likelihood ratio test will be used to test for significance.

Models 3 & 4: To determine if cholesterol level is an effect modifier, we will use Models 1 & 2 and add another covariate which is an interaction term between duration of chest pain symptoms and high cholesterol levels. A likelihood ratio test will then be used to test the significance of this interaction variable.

Table 1 displays visualizations of the planned statistical models for the analyses.

Results

Descriptive Analysis

First, we will evaluate our findings from our descriptive analyses. From **Table 2**, about 67% of the patients in the data set has Significant Coronary Disease while about 33% does not. Also, the mean days of chest pain is higher for those with Significant Coronary Disease than those who don't. Similarly, the proportion of those who have Significant Coronary Disease who also has high cholesterol is higher than those who don't have Significant Coronary Disease, but also have high cholesterol. In **Table 3**, of the 67% who has Significant Coronary Disease, about 48% of those cases has Severe Coronary Disease and about 52% does not. Additionally, the mean difference for duration of chest pain between the two groups, those with Severe Coronary Disease and without, has a larger, notable mean difference than the two groups in **Table 2**. **Figure 2** helps illustrate this notable mean difference between the two groups.

To achieve our primary objectives, we want to observe our covariate of interest, duration of chest pain. Therefore, in **Figure 1**, the data recorded for duration of chest pain is right-skewed. To normalize the covariate for valid results, we will log transform it and add 1 as there are cases where patients experience 0 days of chest pain. We will now use this transformed variable in future analyses. **Figure 3** is similar to **Figure 2**, but is now stratified by high cholesterol. It appears all groups have a noticeable median difference except for those who have Significant Coronary Disease and high cholesterol (**Figure 3**). This exception might mean high cholesterol is an effect modifier and is worth investigating further.

Missing Data Analysis

Now, we will examine our missing data analyses. In **Figure 4**, it appears 6% of the total observed data are missing. The majority of the missing values are cholesterol levels and a few missing values indicating Severe Coronary Disease. **Figure 5** illustrates the possible missing patterns for cholesterol levels. However, in this figure, it appears the patterns for missing versus not missing values are similar. Also, for a subset that include only the missing cholesterol levels, 33% of those patients with missing values are females. This is similar to the proportion in **Table 1**, 31% of the patients in the entire observed data are females. Therefore, the missingness of cholesterol levels appears to not have a distinct pattern based

on the other covariates. Therefore, it's reasonable to assume that this data is missing at random due to no prior knowledge behind the missingness of the data. We will implement multiple stochastic regression imputation for the missing values of cholesterol levels, and logistic regression imputation for the missing values indicating Severe Coronary Disease. Thus, we will now implement our planned statistical methods to answer our primary and secondary scientific questions with these imputed values.

Primary Analysis

Based on the results of the descriptive and missing data analyses, we fitted multiple logistic regression to models 1 and 2 in **Table 1** to answer our primary objectives. From **Table 4**, when comparing two populations with the same cholesterol level, sex, and age but who differ in duration of chest pain symptoms by 1 log-day, we estimate the group who experiences more log-days of chest pain symptoms has odds of having Significant Coronary Disease 1.018 times greater than those who experiences less log-days (95% CI: 0.961 - 1.079). We did not find a statistically significant association between having Significant Coronary Disease and the duration of chest pain symptoms after adjusting for other variables (LRT p-value = 0.536).

From **Table 5**, when comparing two populations with the same cholesterol level, sex, and age but who differ in duration of chest pain symptoms by 1 log-day, we estimate the group who experiences more log-days of chest pain symptoms has odds of having Severe Coronary Disease 1.304 times greater than those who experiences less log-days (95% CI: 1.23 - 1.382). We did find a statistically significant association between having Severe Coronary Disease and the duration of chest pain symptoms after adjusting for other variables (LRT p-value $<$ 0.001).

Secondary Analysis

Similarly, we fitted multiple logistic regression to models 3 and 4 in **Table 1** to answer our secondary objectives. From **Table 6**, when comparing two populations with the same cholesterol level, sex, and age but who differ in duration of chest pain symptoms by 1 log-day, the estimated difference in odds of having Significant Coronary Disease between two groups that have high cholesterol is 0.96 times lower than the difference between two groups that don't have high cholesterol (95% CI: 0.855 - 1.079). We did not find this interaction between chest pain symptoms and high cholesterol to be statistically significant after adjusting for other variables (LRT p-value = 0.495).

From **Table 7**, when comparing two populations with the same cholesterol level, sex, and age but who differ in duration of chest pain symptoms by 1 log-day, the estimated difference in odds of having Severe Coronary Disease between two groups that have high cholesterol is 1.058 times greater than the difference between two groups that don't have high cholesterol (95% CI: 0.943 - 1.187). Likewise, we did not find this interaction between chest pain symptoms and high cholesterol to be statistically significant after adjusting for other variables (LRT p-value = 0.341).

Discussion

Based on the results of the primary analysis, we found that there is a significant association between having Severe Coronary Disease and the duration of chest pain symptoms, but there is not a significant association between having Significant Coronary Disease and the duration of chest pain symptoms. We previously mentioned that if an individual has Severe Coronary Disease, they also have Significant Coronary Disease, so this is an interesting result as we might expect to find a significant association for both groups. There may be many reasons for these results. One reason is that individuals with Severe Coronary Disease are experiencing a more extreme form of CAD as compared to those with only Significant Coronary Disease. These individuals are likely experiencing more acute symptoms that last longer compared to mild symptoms where duration might fluctuate. Additionally, within this dataset the definition of Significant Coronary Disease is that there's a blockage of at least 75% in at least one of the major coronary arteries. Therefore, two individuals with Significant Coronary Disease might be experiencing very different types and severity of symptoms while two individuals with Severe Coronary Disease are likely to experience similar types and severity of symptoms.

Based on the results of the secondary analysis, we found that the interaction between duration of chest pain symptoms and high cholesterol for both Significant and Severe Coronary Disease is not statistically significant. Addition to this result, high cholesterol is a risk factor of CAD and can be the cause of many CAD symptoms. Therefore, it's likely high cholesterol is not an effect modifier and likely a possible confounder of coronary disease.

Furthermore, there are many limitations in this dataset. The duration of chest pain symptoms is vaguely defined in days. The dataset does not specify the severity and types of symptoms and whether or not the duration of symptoms were recorded based on consecutive days or days in a certain timeframe. Also, it's possible that an individual experiencing 5 minutes of searing chest pain and an individual experiencing shortness of breath are recorded as one day. This is not a similar comparison and can impact the interpretation of the results. Moreover, this is an observational study, so results are exploratory in nature and does imply any causal relationships. It's also likely there are other possible confounders such as smoking status, diet, exercise and more that were not recorded and adjusted for in the statistical analysis.

Tables & Figures

Table 1: Planned Multivariate Logistic Regression Models

	Model 1	**Model 2**	**Model 3**	**Model 4**
Response	$sigdz$	$tvdlm \mid sigdz = 1$	$sigdz$	$tvdlm \mid sigdz = 1$
Covariates	*cad.dur* *sex=Female* *age-20* *choleste*	*cad.dur* *sex=Female* *age-20* *choleste*	*cad.dur* *sex=Female* *age-20* *choleste=high* *cad.dur* × *choleste = high*	*cad.dur* *sex=Female* *age-20* *choleste=high* *cad.dur* × *choleste = high*

Table 2: Summary Statistics for observed data grouped by indicator of Significant Coronary Disease

	Doesn't Have Significant Coronary Disease	**Has Significant Coronary Disease**	**Overall**
Proportion (N)	0.333 (1167)	0.667 (2334)	1 (3501)
Sex (Female)			
Proportion (N)	0.546 (637)	0.198 (462)	0.314 (1099)
Age (years)			
Mean (SD)	48.9 (9.72)	54 (9.51)	52.3 (9.89)
Median [Min, Max]	49 [20.0, 80.0]	54 [25.0, 82.0]	52 [20.0, 82.0]
Duration of Chest Pain (days)			
Mean (SD)	36.5 (53.9)	46.3 (59.9)	43.0 (58.2)
Median [Min, Max]	13 [0, 372]	22 [0, 416]	18 [0, 416]
Cholesterol Level (mg/dl)			
Mean (SD)	220 (44.3)	235 (52.8)	230 (50.6)
Median [Min, Max]	216 [89, 500]	230 [29, 576]	225 [29, 576]
Missing (%)	401 (34.4%)	844 (36.2%)	1245 (35.6%)
Has High Cholesterol (>238 mg/dl)			
Proportion (N)	0.292 (224)	0.432 (643)	0.384 (867)
Missing (%)	401 (34.4%)	844 (36.2%)	1245 (35.6%)

Table 3: Summary Statistics for observed data grouped by indicator of Severe Coronary Disease given individual has Significant Coronary Disease

	Doesn't Have Severe Coronary Disease	**Has Severe Coronary Disease**	**Overall**
Proportion (N)	0.515 (1203)	0.484 (1129)	1 (2334)
Sex (Female)			
Proportion (N)	0.227 (273)	0.167 (188)	0.198 (462)
Age (years)			
Mean (SD)	52.4 (9.45)	55.8 (9.25)	54.0 (9.51)
Median [Min, Max]	52 [25, 82]	56 [25, 81]	54 [25, 82]
Duration of Chest Pain (days)			
Mean (SD)	34.6 (48.8)	58.7 (67.7)	46.3 (59.9)
Median [Min, Max]	12 [0, 364]	35 [0, 416]	22 [0, 416]
Cholesterol Level (mg/dl)			
Mean (SD)	233 (51)	237 (54.6)	235 (52.8)
Median [Min, Max]	230 [105, 576]	230 [29, 500]	230 [29, 576]
Missing (%)	436 (36.2%)	406 (36.0%)	844 (36.2%)
Has High Cholesterol (>238 mg/dl)			
Proportion (N)	0.425 (326)	0.438 (317)	0.432 (643)
Missing (%)	436 (36.2%)	406 (36.0%)	844 (36.2%)

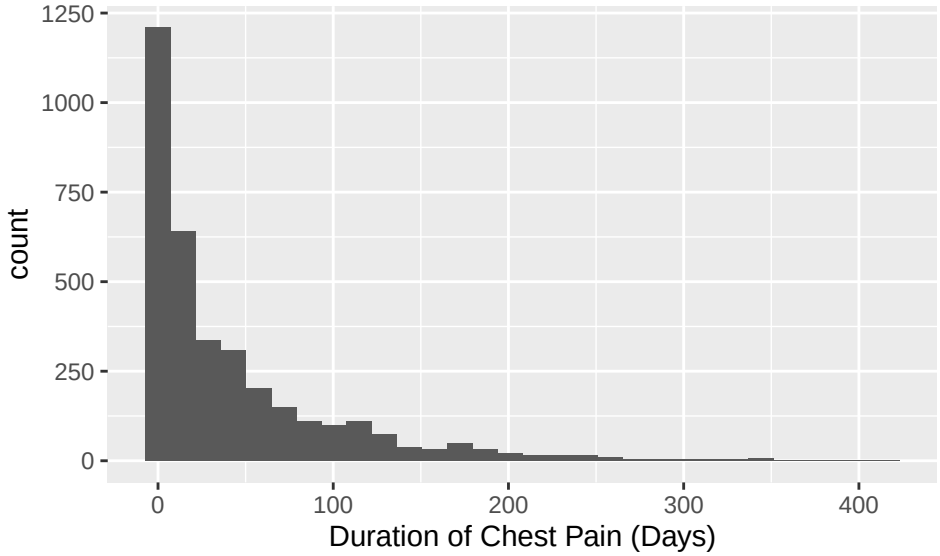

Figure 1: Histogram of Duration of Chest Pain

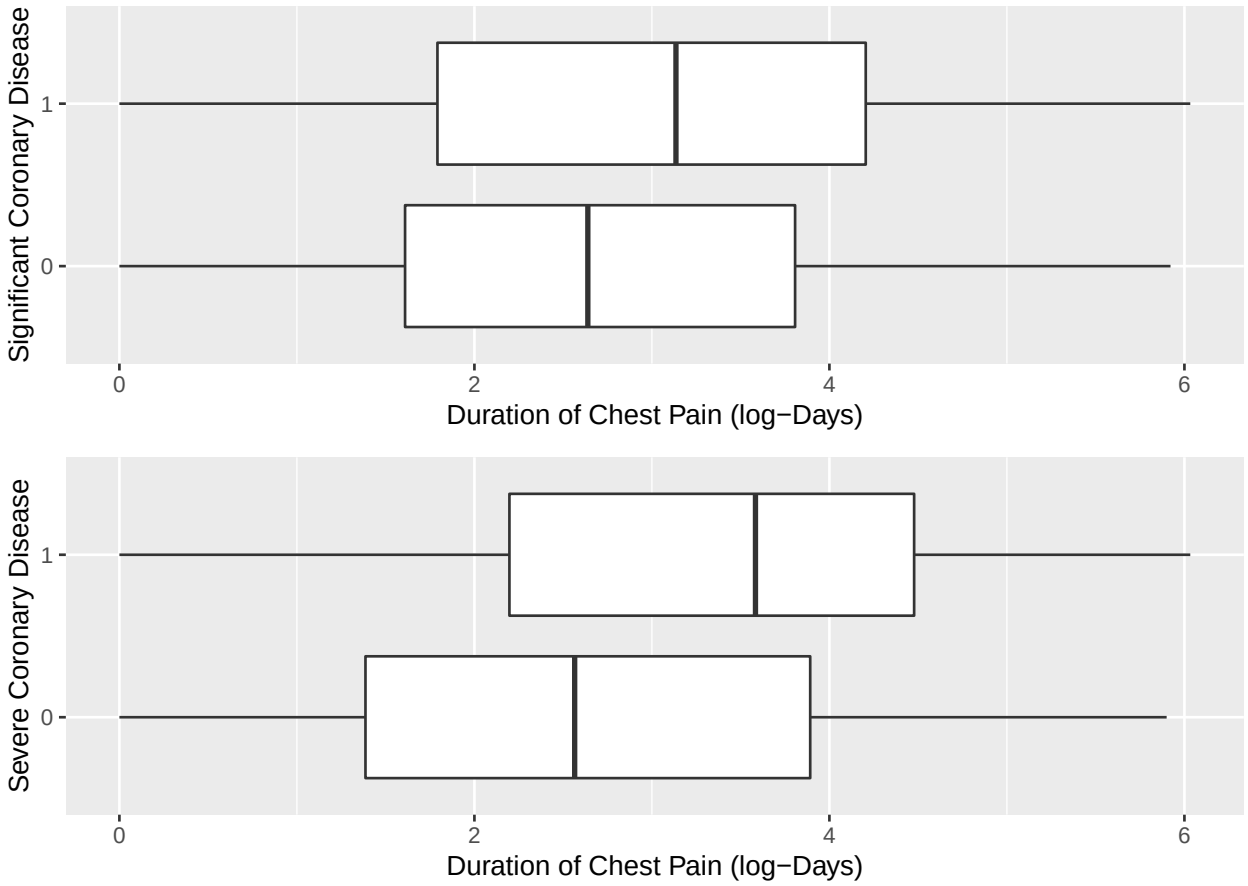

Figure 2: Boxplots of Significant/Severe Coronary Disease vs log-Duration of Chest Pain

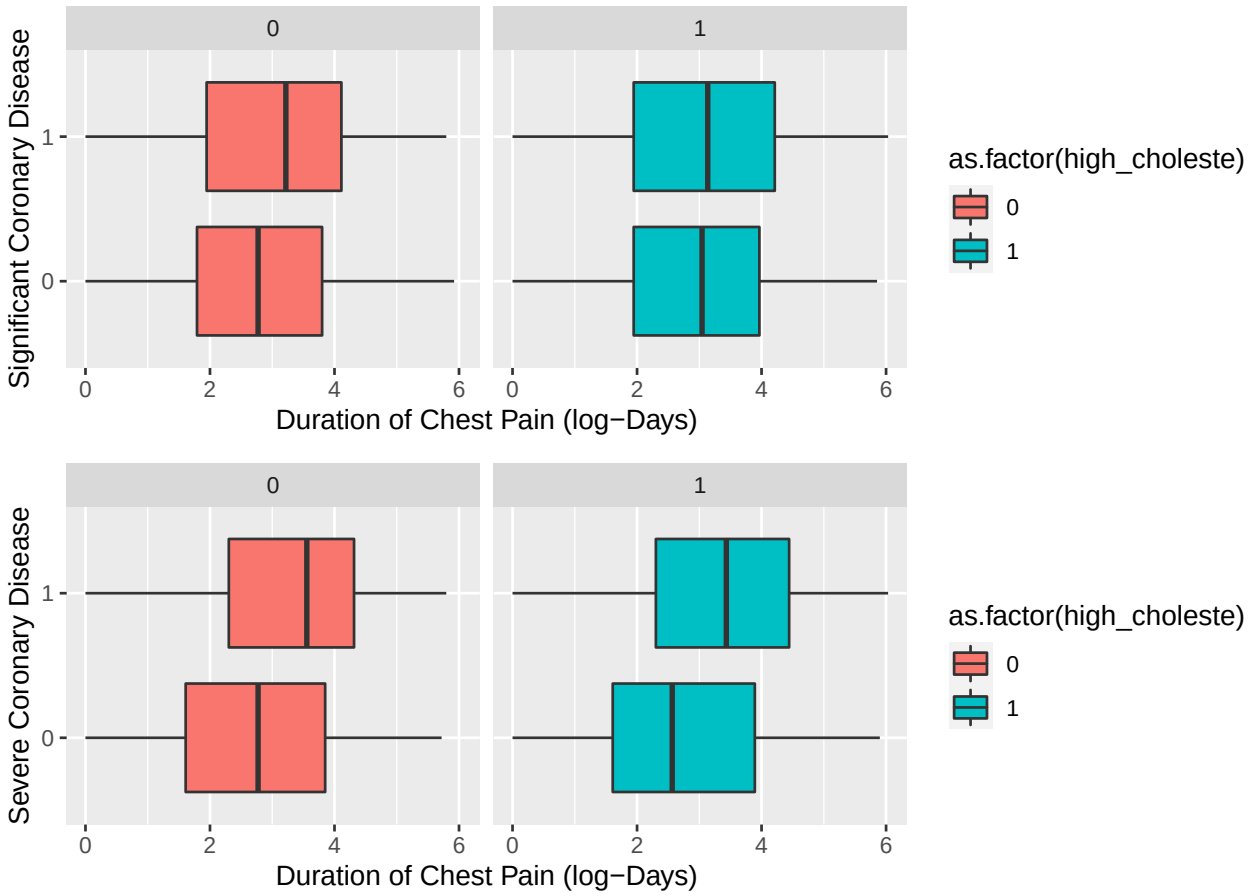

Figure 3: Boxplots of Significant/Severe Coronary Disease vs log-Duration of Chest Pain stratified by High Cholesterol Level

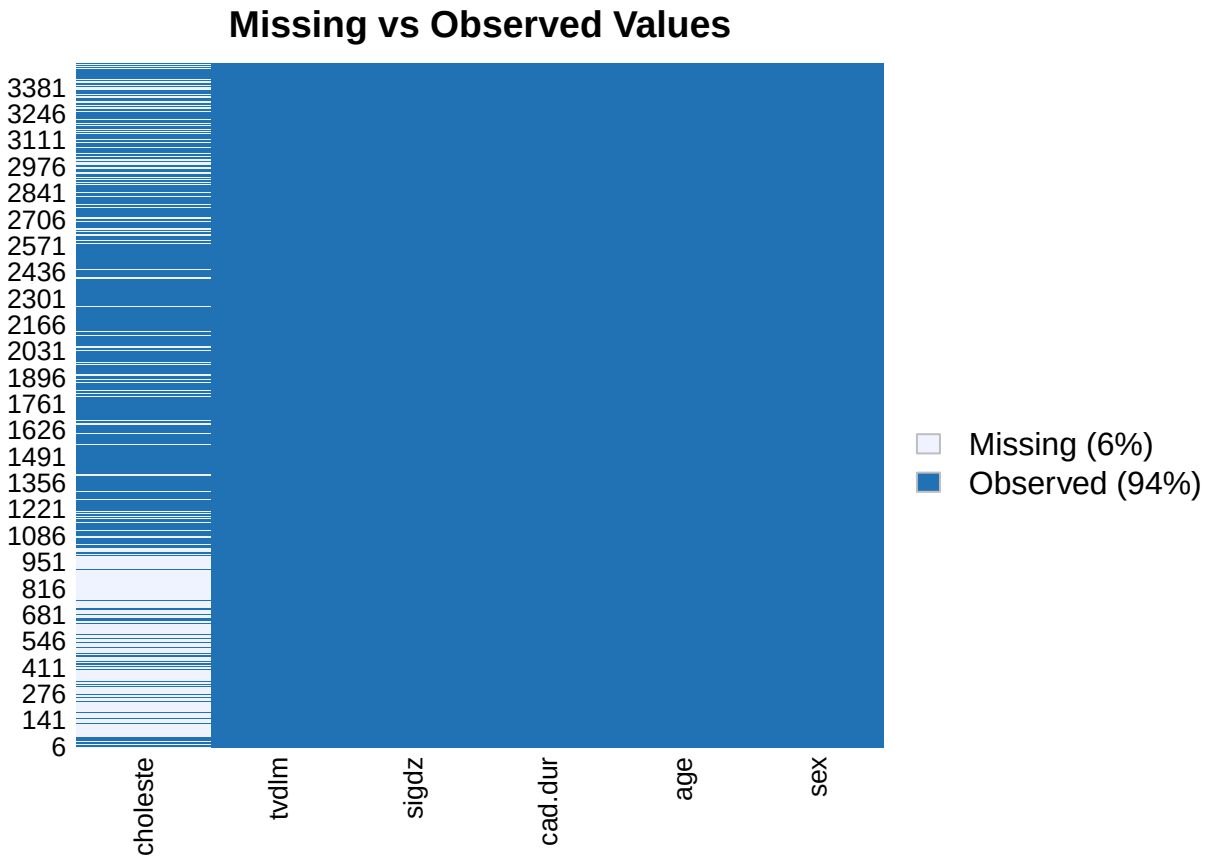

Figure 4: Plot of Missing vs Observed Values

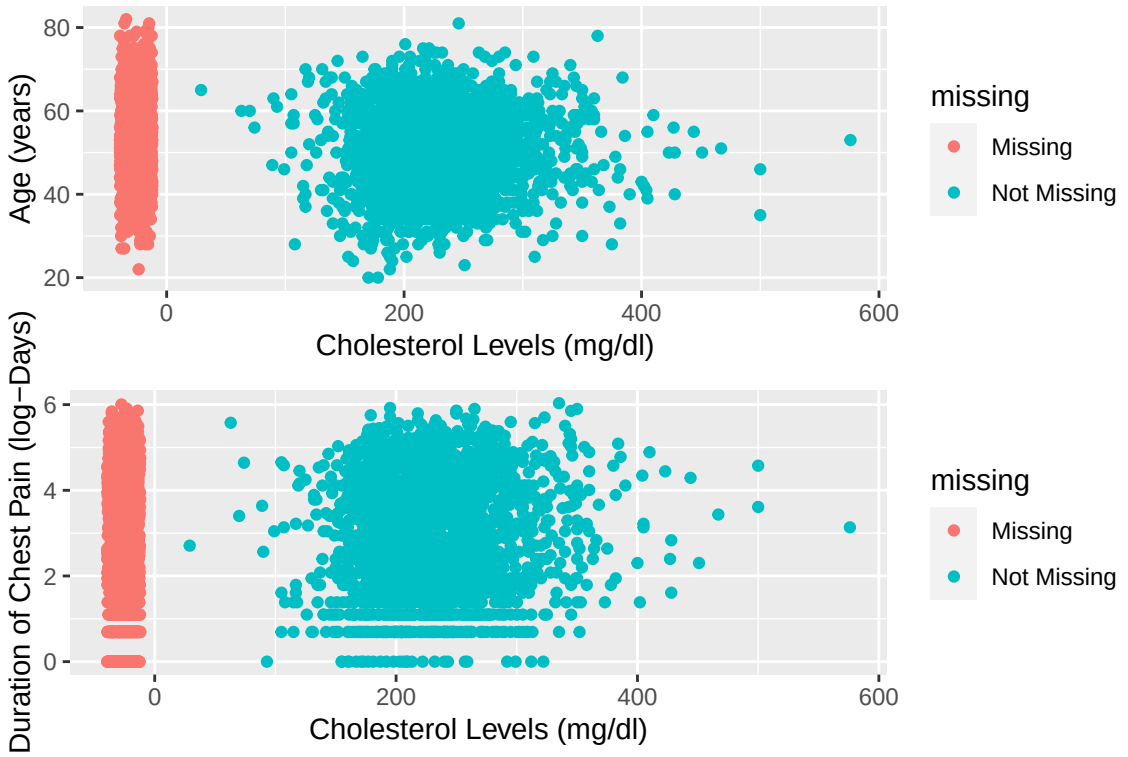

Figure 5: Plots of Cholesterol vs Age & Duration of Chest Pain stratified by missingness

Table 4: Estimated Odds Ratios of Having Significant Coronary Disease

	Estimate	Lower 95% CI	Upper 95% CI	LRT P-Value
Duration of Chest Pain (log-days)	1.018	0.961	1.079	0.536
Cholesterol Level (mg/dl)	1.009	1.007	1.010	–
Sex (Female)	0.123	0.102	0.148	–
Age (years over 20)	1.081	1.071	1.091	–

Table 5: Estimated Odds Ratios of Having Severe Coronary Disease given the individual has Significant Coronary Disease

	Estimate	Lower 95% CI	Upper 95% CI	LRT P-Value
Duration of Chest Pain (log-days)	1.304	1.230	1.382	2.01e-19
Cholesterol Level (mg/dl)	1.002	1.001	1.004	–
Sex (Female)	0.529	0.421	0.665	–
Age (years over 20)	1.038	1.028	1.048	–

Table 6: Estimated Odds Ratios of Having Significant Coronary Disease (with interaction)

	Estimate	Lower 95% CI	Upper 95% CI	LRT P-Value
Duration of Chest Pain (log-days)	1.049	0.978	1.125	–
Cholesterol Level (High)	2.471	1.691	3.611	–
Sex (Female)	0.127	0.106	0.152	–
Age (years over 20)	1.079	1.069	1.089	–
Duration of Chest Pain (log-days):Cholesterol Level (High)	0.960	0.855	1.079	0.495

Table 7: Estimated Odds Ratios of Having Severe Coronary Disease given the individual has Significant Coronary Disease (with interaction)

	Estimate	Lower 95% CI	Upper 95% CI	LRT P-Value
Duration of Chest Pain (log-days)	1.278	1.183	1.380	–
Cholesterol Level (High)	0.951	0.648	1.396	–
Sex (Female)	0.548	0.437	0.686	–
Age (years over 20)	1.037	1.027	1.047	–
Duration of Chest Pain (log-days):Cholesterol Level (High)	1.058	0.943	1.187	0.341

References

Clinic, Cleveland. 2018. "Understanding Your Cholesterol Numbers | Cleveland Clinic." Cleveland Clinic. https://my.clevelandclinic.org/health/articles/11920-cholesterol-numbers-what-do-they-mean.

Clinic, Mayo. 2019. "Cardiac Catheterization - Mayo Clinic." Mayoclinic.org. https://www.mayoclinic.org/tests-procedures/cardiac-catheterization/about/pac-20384695.

———. 2020. "Coronary Artery Disease." Mayo Clinic. https://www.mayoclinic.org/diseases-conditions/coronary-artery-disease/symptoms-causes/syc-20350613.

Digirad. 2020. "Triple Vessel Disease: Pathology, Diagnosis and Treatment." Digirad. https://www.digirad.com/triple-vessel-disease-diagnosis-treatment/.

Harrell Jr, Frank E. 2002. "Datasets." hbiostat.org. http://hbiostat.org/data/#vanderbilt-biostatistics-datasets.

Higgins, Peter. 2021. "Cath Dataset Description Document." htmlpreview.github.io. https://htmlpreview.github.io/?https://github.com/higgi13425/medicaldata/blob/master/man/description_docs/cath_desc.html.