Myocardial Infarction and Mitral Regurgitation: Predicting Severity and Survivability

Nick Peckover, RDCS

Problem Description and Executive Summary

Following an acute myocardial infarction (AMI), the most common and relied upon surgical treatment is a percutaneous coronary intervention (PCI). Based on the location and severity of the myocardial infarction (MI), mitral valve regurgitation (MR) may develop or worsen. An echocardiogram is used to diagnose the severity of the MR, and is graded on a scale of mild, moderate, or severe. Mild MR is usually treated as inconsequential, but moderate and severe MR require careful attention and management. Moderate MR can progress to severe MR without careful management. Severe MR can lead to critical heart failure and death in a matter of days.

In resource stretched hospital systems, prediction of hemodynamically significant MR would allow the highest risk patients to receive valuable and hard-to-find resources quickly. In this report, I use this motivation and information gathered from coronary angiograms and echocardiograms to answer two key questions:

1. What are the differences in survival outcomes and survival probabilities for each grade of MR?
2. Can a black box model using non-echocardiogram data predict which patients have hemodynamically significant (moderate to severe) MR?

The results for these questions are summarized below:

Table 1: Results for Question 1: Survival Analysis

MR Severity	Risk relative to No MR	Model coefficient	p-value
Mild	1.3277	0.2835	0.26646
Moderate	1.1103	0.1046	0.82339
Severe	5.4984	1.7045	0.00413

Table 2: Results for Question 2: Classification

Accuracy	0.8755
Balanced Accuracy	0.7151
Sensitivity	0.5294
Specificity	0.9009
F1	0.3674
ROC AUC	0.7572

		Reality	
		None-Mild	Moderate-Severe
Predicted	None-Mild	***209***	8
	Moderate-Severe	23	***9***

Project Background and Exploratory Data Analysis

I use the 2021 Sharma data (available in the appendix) for this project. The data are observational and collected from 1000 patients suffering from AMI, treated with PCI, on whom an echocardiogram was performed prior to hospital discharge. The data were collected from December 2015 to December 2019 and include 122 potential variables of interest, 41 of which are used in the final analysis.

For both research questions, the distribution of the severity of MR has great bearing on the analysis. "No MR" is the overwhelmingly likely outcome in this dataset, and Severe MR is the least. There are a mere 9 cases of Severe MR, while No MR comprises 703 cases.

For the purposes of the classification task, the classes are consolidated into "None-Mild" and "Moderate-Severe". This further exacerbates the class imbalance but is in line with our goals: to use tools other than echocardiograms (which determine the MR grade) for risk stratification.

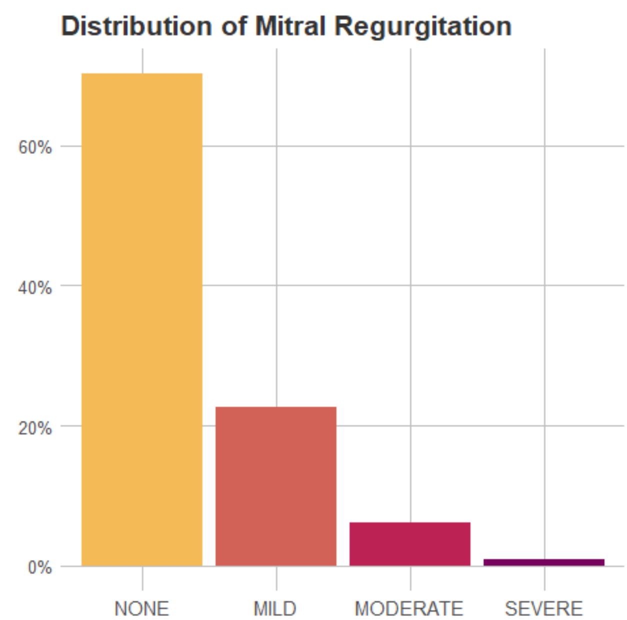


Figure 1: Mitral Regurgitation Proportions

There are 122 variables in this data set, and 70 of them include information regarding patient history and disease details. Relevant examples include gender, age, history of smoking, stroke, myocardial infarction, hypertension and family history of coronary artery disease. The rest contain post discharge and readmission information. The data set is comprised of mostly categorical variables, where factors are largely imbalanced and missingness is high. Of the 70 possibly usable variables, 12 have missing information so great that they cannot be used for rigorous analysis, including several variables with more than 50% of the observations missing. 41 variables will be used for the classification and prediction task. Of these, 15 variables are missing more than one observation. The average number of missing observations (given that a variable is useable and at least one observation from said variable is missing) is 26.

Variable	Smoker	FHxOfCAD	Hyperchol.	HTN	CVA	PreviousMI	IndexTroponin
Missing	36	34	83	7	1	4	87
Variable	PeakTroponin	RCA	Clopidogrel	Ticagrelor	Prasugrel	Eptifibatide	ClinicalHF
Missing	10	1	40	40	40	3	1

Table 3: Missing data by selected classification variables

Using a coronary angiogram, it is possible to grade the severity of a vessel blockage and determine which vessel requires PCI. A common numeric grading scale where 0 is ‘no blockage’ and 100 is ‘completely blocked’ is used to describe the severity of a blockage. Furthermore, it is theorized that the location of the blockage can be a predictor of the severity of MR. The mitral valve is attached to the left ventricular muscle wall through a complicated subvalvular apparatus including papillary muscles. When these papillary muscles are starved of oxygen (such as during AMI) their function may be impaired, leading to mitral valve disfunction and MR. (Dal-Bianco, Beaudoin, Handschumacher, & Levine, 2014) Blockages in the left circumflex artery (LCx) and right coronary artery (RCA) are hypothesized to be of particular importance since these often feed papillary muscles. The Figure 2 shows the severity of the vessel blockage (by vessel location) and the corresponding MR grade (note that due to low counts in Moderate and Severe the categories are combined):

Figure 2: MR Severity by Vessel Location and Blockage Severity. The Moderate and Severe groups are consolidated for readability.

LMS = Left Main Stem Artery; LAD = Left Anterior Descending Artery; RCA = Right Coronary Artery; LCx = Left Circumflex Artery

A subtle pattern that supports the theory of blockage location impact on MR severity is seen. While the absolute counts are small, moderate to severe MR appears more likely to occur with certain vessel blockages than others. The distribution of MR in all 1000 patients regardless of vessel location or blockage severity is 70.3% 'None', 22.7% 'Mild', and 7% 'Moderate-Severe'. Looking at the intervenable blockages (75-94, 95-99, and 100) in any one vessel can give a hint to the effect of blockage location on MR. The left circumflex artery, for example, shows a distribution of 60.9% 'None', 27.2% 'Mild', and 11.8% 'Moderate-Severe' MR among intervenable blockages. Non-intervenable blockages for the same vessel correspond to a split of 76.1% 'None', 19.9% 'Mild', and 4% 'Moderate-Severe' MR.

The few continuous variables we have (7 of the 41 predictors used for the classification task are continuous) are weakly correlated with each other and show very poor class separation. Interested parties may consult the appendix for scatterplot matrices, density plots, and correlations of select variables for the two consolidated groups ("None-Mild" and "Moderate-Severe").

	Age	Creatine Clearance	Index Troponin	Peak Troponin	Stents Used	Time To PCI	Days Post Symp. Onset	Door To Balloon
Min	25	0	0	0	1	65	0	0
1st Qu.	55	57	31	66.25	2	283.1	0	0
Median	66	78	95	521.5	2	1454.7	1	0
Mean	65.2	83.7	513.1	1826.9	2.6	4147.6	2.8	40.6
3rd Qu.	76	106	371	2473.25	3	4992.9	3	36
Max	98	287	9425	11990	9	94559.8	65	1337
Variance	169.5	1475	1251414	7357079	0.955	62963040	30.196	1684.2

Table 4: Summary statistics of available continuous variables

For the continuous variables it is clear that (like the factor variables) the data are often highly skewed, usually containing many low values. This is evidenced quickly in the above table by the presence of a 75th percentile value lower than the mean for a few variables.

Methodology and Analysis

Question 1: What are the differences in survival outcomes and survival probabilities for each grade of MR?

To answer this question, survival analysis will be performed. In this data set, there are 919 surviving patients and 81 for which a death was observed during the data collection period. If a death was not recorded during the data collection period, the patient was assumed to be alive. It is also assumed that the censoring time is independent of event time and that each death occurs at a distinct time. The data is collected over a period of four years, and it is assumed that treatment quality does change over this data collection period.

The goal of this analysis is to examine potential differences between survival probabilities between the four grades of MR (None, Mild, Moderate, Severe). A Log-Rank test is performed with a null hypothesis of no difference in survival probabilities between MR types. A Cox Proportional Hazards model is fit. The alpha level for all tests is 5%. R version 4.2.1 is used for all analysis.

Question 2: Can a black box model using non-echocardiogram data predict which patients have hemodynamically significant (moderate to severe) MR?

The goal of this classification task is to better predict the presence of Moderate-to-Severe MR using non-echocardiogram data. This limits the useable variables to those such as lab values, medications, angiogram findings, and demographic information. Potentially strong predictors such as left ventricular ejection fraction (LVEF) were omitted as it is most likely that these data were collected from echocardiograms. Overall, the available predictors are normally very weak predictors of the response, explaining the need for echocardiograms in the first place. Information such as time of death, time to discharge, and multiple predictors where missing data were greater than 10% were also omitted. Since more than one artery can be blocked at a time, two new indicator variables were constructed to signify multiple blockages and multiple interventions.

Missing data were imputed accordingly: continuous variables received predictive mean matching, factor variables with only 2 factors received logistic regression, and factor variables with greater than 2 factors received multinomial logistic regression imputation. The missing data for each variable was imputed 20 times, updating the predicted value at each step according to the appropriate predictive mechanism. The data were split into a testing and a training set of 249 and 751 observations respectively, taking care to retain balance in the response variable (Severity) within both sets. To prevent data leakage, the final imputation model was trained only on the training portion of the data, and the test set missing data were imputed using this training set imputation model.

Because of the large class imbalance (930 observations in “None-Mild” against 70 observations in “Moderate-Severe”), *Synthetic Minority Over-Sampling Technique* (SMOTE) was used. Briefly, this method allows for simultaneous oversampling of the “Moderate-Severe” class (by creating artificial data points in the feature space between two real data points) as well as random undersampling of the “None-Mild” class. The oversampling rate for this task is 400% (to create 212 synthetic date points in the training set) and the undersampling rate is set so as to randomly remove 274 observations in the majority class. This brings the total class counts to 424 None-Mild and 265 Moderate-Severe, from 698 and 53 respectively. SMOTE is performed within cross validation as doing otherwise can potentially optimistically bias the model (Kuhn, 2019) through data leakage. 10-fold cross validation is used to tune the classifier parameters.

Traditionally, predictive accuracy (defined as $\frac{True\ Predictions}{All\ Predictions}$) is used for measuring the performance of a classification model. For this task, the test set is comprised of a 93% majority "None-Mild" class. Any model that prioritizes accuracy can simply predict "None-Mild" for every observation and be correct 93% of the time. However, the real penalty for a False Negative (predicting "None-Mild" when the reality is "Moderate-Severe") is potentially deadly. Therefore, it should be the goal of the classifier to increase the number of "Moderate-Severe" predictions, even at the cost of reduced accuracy stemming from False Positives. A balanced approach is still necessary, however, for large amounts of false positives do not necessarily aid with resource allocation.

The Receiver Operating Characteristic (ROC) Area Under the Curve (AUC) is preferred by some authors for evaluating the performance of these models on an imbalanced dataset with imbalanced error costs (N. V. Chawla, 2002), while some authors recommend the use of Precision-Recall (PR) curves (Davis & Goadrich, 2006). Both metrics describe a balance between the positive and negative error rates that simpler performance metrics cannot. The ROC and PR curves for the final model are shown below. The ROC AUC measures approximately 0.76. The PR AUC measures approximately 0.25. For both metrics a perfect classifier would measure 1.0, and the worst possible classifier would score 0.0. The curves give an idea of where one may be able to set a subjective classification threshold. 'Recall' = 'Sensitivity' and is the proportion of correct positive (i.e. having the disease) classifications to the total number of positive cases: $\frac{True\ Positives}{True\ Positives+False\ Negatives}$. It is possible to make this value equal to 1 by ensuring that there are no false negatives by simply predicting 'Positive' (i.e. Moderate-Severe MR) for every case, however, the tradeoff is that strategy reduces the positive predictive value (the precision), measured $\frac{True\ Positives}{True\ Positives+False\ Positives}$.

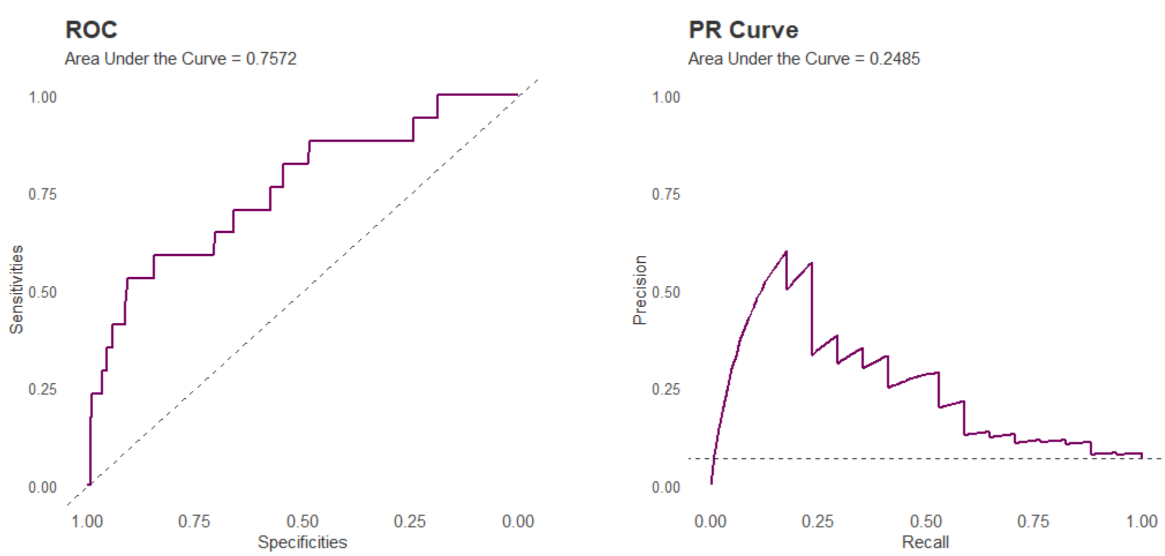


Figure 3: ROC and PR Curves for the final classification model. The dashed lines show the theoretical results of the "no information" model.

Multiple algorithms were tested to find the best performer, including Random Forest, logistic regression, support vector machines, and a traditional gradient boosting machine. The final model uses the popular *eXtreme Gradient Boosting* (XGBoost) algorithm, an ensemble tree-based method that learns slowly by creating sequential decision trees, using the errors from the previous tree to improve the fit. (Chen & Guestrin, 2016) This is a flexible method that performs well with the large amount of categorical variables in this data set. While XGBoost can normally handle missing information well, the imputed data was retained to train the model. R version 4.2.1 is used for all analysis.

Results

Question 1: What are the differences in survival outcomes and survival probabilities for each grade of MR?

Severity	n	Observed	Expected	$\frac{(O-E)^2}{E}$	$\frac{(O-E)^2}{V}$
None	703	51	56.89	0.610	2.053
Mild	227	22	18.48	0.673	0.872
Moderate	61	5	5.02	0.0001	0.0001
Severe	9	3	0.61	9.36	9.443

Table 5: Log-Rank Test Results

The Chi-squared test statistic for the full Log-Rank test is 10.7 with an associated p-value of 0.01. We can reject the null hypothesis of no difference at the 5% level and conclude that at least one of the survival curves is different.

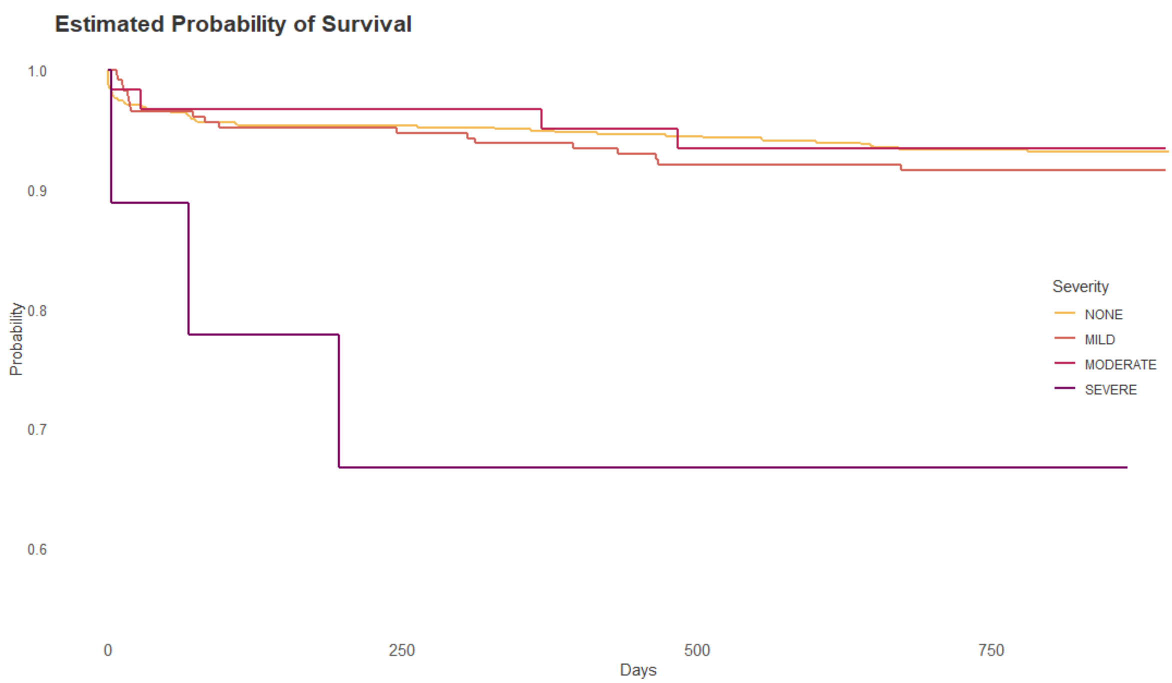


Figure 4: Kaplain-Meier Survival Curves for each MR grade

A Cox Proportional Hazards model is also fit.

MR Severity	Model coefficient	Risk relative to No MR	Relative Risk 95% CI Lower Bound	Relative Risk 95% CI Upper Bound	p-value
Mild	0.2835	1.3277	0.8053	2.189	0.26646
Moderate	0.1046	1.1103	0.4431	2.782	0.82339
Severe	1.7045	5.4984	1.7153	17.626	0.00413

Table 6: Cox Proportional Hazards Model Results

There is not a statistically significant difference between the survival rates of Mild or Moderate MR relative to the baseline of No MR. The risk of death for those with Severe MR is nearly 5.5 times that of those with no MR at all. In this sample, 67% of patients with Severe MR survived 1000 days, as opposed to 93% of those with no MR. The 95% Confidence Interval (CI) for relative risk for Severe MR is (1.7153, 17.626).

Question 2: Can a black box model using non-echocardiogram data predict which patients have hemodynamically significant (moderate to severe) MR?

The performance of the final method on the final test set is shown below. Correct predictions are highlighted.

		Reality	
		None-Mild	Moderate-Severe
Predicted	None-Mild	***209***	8
	Moderate-Severe	23	***9***

Table 7: Confusion Matrix of classifier results on testing set

The model correctly predicts 53% of the “Moderate-Severe” cases while limiting false positives to 9% of the total predictions. The model’s probabilities for each case of “Moderate-Severe” are listed in Table 8.

Case	11	34	36	43	98	123	127	144	166
P(MODs)	0.136	0.520	0.216	0.160	0.166	0.642	0.050	0.693	0.246
Case	173	182	189	203	206	238	244	249	
P(MODs)	0.411	0.806	0.551	0.722	0.855	0.816	0.821	0.060	

Table 8: Classifier assigned probabilities for each "Moderate-Severe" case in the test set.

Variable importance measures for the 15 most important variables are displayed below. Note that these results are scaled to a maximum of 100 and a minimum of 0.

Variable	Smoker: NON	MultiBlock: YES	Clopidogrel: YES	LCx: 100	AGE	Gender: M
Importance	100	66	62	48	29	25
Variable	LCxIntervention: YES	LADOther: 75-94	ECG: 2	LADProx: 75-94	DaysPostSympt-Procedure	Creatinine-Clearance
Importance	20	12	9	8	7	6

Table 9: Variable Importance Measures

Discussion of Results; Limitations and Recommendations

Question 1: What are the differences in survival outcomes and survival probabilities for each grade of MR?

Survival analysis makes an important assumption about patient survival: namely that deaths were recorded accurately and appropriately. The data collection details for this study are not known, and this assumption may turn out to be inaccurate. The largest limitations of this analysis are related to the sample size. Since Severe MR is rare (only 9 cases in this sample of 1000 patients), care should be taken when interpreting the results, especially the risk relative to baseline. While the results are statistically significant, the 95% confidence interval is very wide due to the small sample. Compare the size of the Severe 95% CI (1.7153, 17.626) to the Mild 95% CI with a sample size of 227 (0.8053, 2.189). To develop more precise estimates, more cases of Severe MR need to be collected and analyzed.

Question 2: Can a black box model using non-echocardiogram data predict which patients have hemodynamically significant (moderate to severe) MR?

When interpreting the results from this classification model, it may be tempting to blindly follow the model recommendation for any given patient. A better method may be to determine a subjective threshold for risk tolerance, then look closely at the probabilities output by the model. Currently, the model classifies any observation with probability greater than 50% into the corresponding class. Changing this threshold to a lower boundary would allow more true positives, at the cost of also creating significantly more false positives. Changing the threshold to, 15%, for instance, would have caught 15 of the 17 positive cases, at the cost of 109 false positives. It is worth remembering that the cost of a false positive is usually limited to an extra diagnostic test such as an echocardiogram that has a high sensitivity to detect MR. This may be more than worthwhile in order to detect a high percentage of positive cases.

It is important when interpreting these model results to examine the variable importance measures. Unsurprisingly, the presence of multiple blockages and a completed occluded left circumflex artery (LCx) are very important to the model. This is consistent with established theory (Dal-Bianco JP, et al, 2014). One potential strategy to reduce risk is closely examine any patient with left anterior descending artery (LAD) and/or LCx disease. An operator of this model could closely examine any important variable (found at the end of Results) to further aid in risk stratification.

The classification task undertaken proved to be extremely difficult. One major limitation of this data set is the high amount of missing data. Many possible predictors were forced to be discarded due to their missingness. This included potentially strong predictors such as the need for circulatory support devices, complication and failure rates, and B-type natriuretic peptide. For variables that did not need to be discarded, there were still high amounts of missing data that needed to be imputed before the statistical models could be created. While advanced methods were utilized, there is no substitute to having the actual information. As is true of most research, future analyses would benefit from higher data collection standards. The classification task in question may also be possible with different types of data not available in this data set. Real-time blood pressure monitoring, fluid intake and output, and various lab values and ventilator settings would all be theoretically useful to the task. Future work remains to be done to establish whether such a classification tool is more than theoretically possible.

Appendix

R version 4.2.1 and R Studio 2022.07.1 is used for all analysis.

The original data set is available below:

Sharma, Harish (2021), "Mitral Regurgitation Following Acute Myocardial Infarction Treated by Percutaneous Coronary Intervention – Prevalence, Risk factors and Predictors of Outcome", Mendeley Data, V1, doi: 10.17632/3yb74rkp4d.1

Full R Code is available on github

References

Azur, M., Stuart, E. A., Frangakis, C., & Leaf, P. J. (2011). Multiple imputation by chained equations: what is it and how does it work? *International journal of methods in psychiatric research*, 40-49.

Dal-Bianco, J., Beaudoin, J., Handschumacher, M., & Levine, R. (2014). Basic Mechanisms of Mitral Regurgitation. *The Canadian Journal of Cardiology*, 971–981.

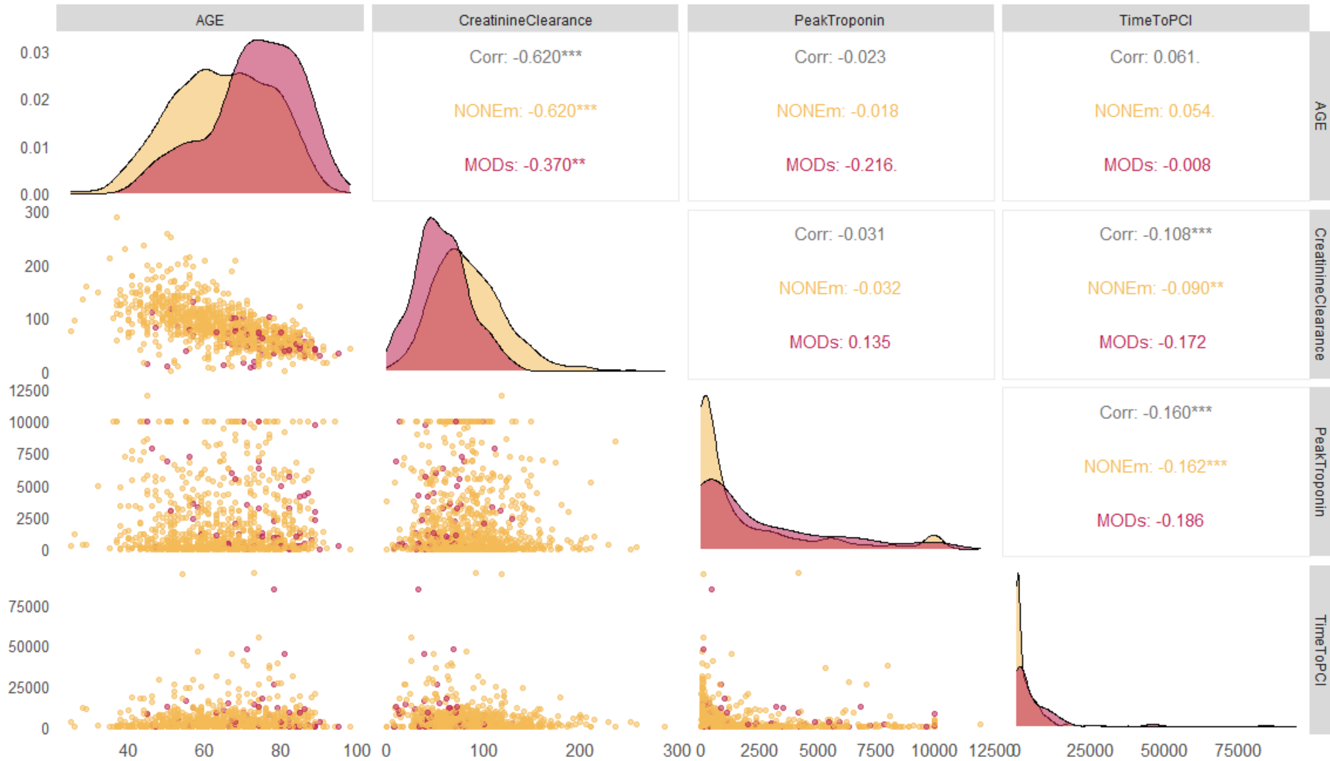
Davis, J., & Goadrich, M. (2006). The relationship between Precision-Recall and ROC curves. *ICML '06: Proceedings of the 23rd international conference on Machine learning*, 233-240.

Gareth James, D. W. (2021). *An Introduction to Statistical Learning.* New York: Springer.

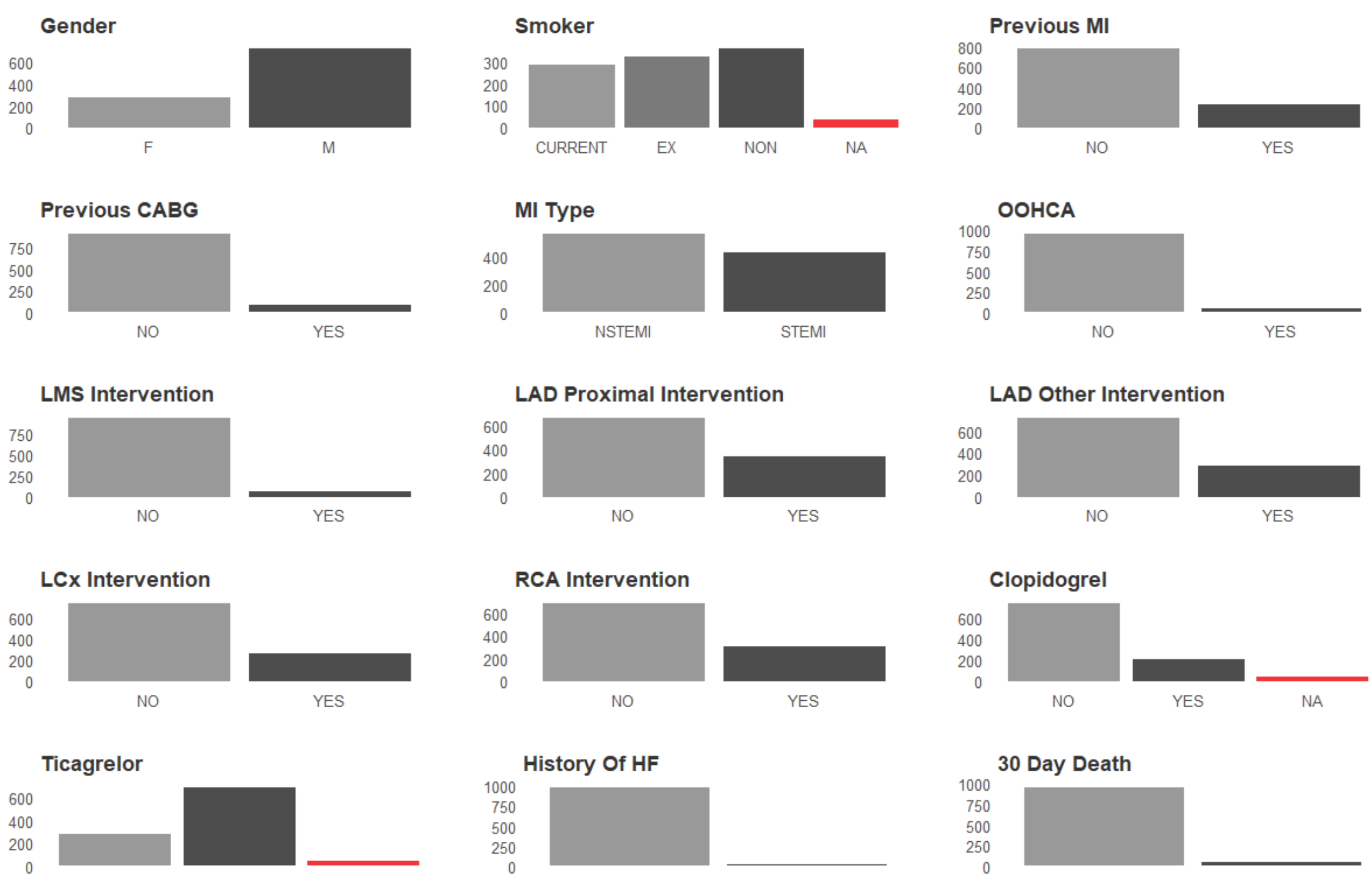
Kuhn, M. (2019). *The caret Package*. Retrieved from caret R Package documentation: https://topepo.github.io/caret/index.html

N. V. Chawla, K. W. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal Of Artificial Intelligence Research*, 321-357.

R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. Retrieved from R Foundation for Statistical Computing: https://www.R-project.org/



Appendix Figure 1: Scatterplot matrix of four important variables. There is poor class separation between the two (consolidated) groups of interest



Appendix Figure 2: Distribution of Factor Variables and Missing Data