Survival Analysis of Post-Myocardial Infarction Patients

Research: Alvein, Parametric: Orr, Non-Parametric: Pham

5/22/2020

Abstract

Background: The rates of myocardial infarction is becoming an increasing common occurrence in the United States. Rapid development of medical technology and knowledge have led to an decline in myocardial infarction fatalities (Gu, et al 1999). However, there is much to be learned regarding the survival probabilities of patients following an infarction episode.Some studies have already examined the effects of externalities on the survival rates of these patients (Rimm, et al. 1995).

Objective: Our goal is to provide detailed survival statistics of patients during a post-myocardial infarction time period with specific concern addressed to age, ventricular activity, and physiological cardiac state. Using these variables, we aim to provide succinct information on the current state of the dataset as well as provide robust predictors for the future estimates of survival for future patients.

We aim to fit non-parametric (Kaplan-Meier) and parameters curves to describe the data as well as choose a regression model to be used for predictive survivability.

Methods: Data from 133 post-myocardial infarction patients measure the time in months until death in a one year monitoring period of follow-up. We use a combination of non-parametric (Kaplan-Meier) and parametric methods (Weibull, Log-Normal, Log-Logistic, Cox PH) to determine estimates of survival among gender and physiological cardiac state (contraction depth, muscular activity, anatomical status). We fit multiple distributions over the dataset to provide current-state information of the patient dataset. Then, we regress multiple models and use combination of Akaike Information (AIC) statistics, logistic ratio tests, and residual analysis to determine model adequacy.

Results: Initial non-parametric Kaplan-Meier curve shows a median survival time of ~30 months for all age groups. We choose a Weibull regression fit (tentative) for predictive model as we have favorable AIC, ratio, and residual indicators out of all of our model.

Conclusion: Thus, for predictive model we found the Weibull regression fit to be the most ideal candidate for modeling survivability for patient groups. Additionally, when examining the survival times for the Kaplan-Meier step curve, we see that the younger age groups do survive as well as their older counterparts. Given our limited sample size for that population, we recommend continued studies into external effects of the post-myocardial episode survival.

Introduction

Heart disease has become the leading cause of US deaths among all racial and ethnic groups (Heron 2019, Fryar 2012). In 2009 cardiovascular disease represented nearly 64% of all cardiac related deaths (Dalen, et. Al. 2014). These myocardial infarction – commonly known as heart attacks - are becoming largely common among all U.S. demographic populations. As such, researchers are looking to understand the underlying causes of these episodes. Specifically, increases in cardiovascular disease (CVD) cases have been largely attributed to many risk factors such as high levels of low-density lipoproteins (LPL), high blood pressure, and smoking (Fryar 2012).

These variables are often the results of lifestyle choices and effects of poverty. The prevalence of the disease has closely been followed a large body of conducted researchers aiming to reduce either the number of these cases or reduce the mortality of the specific myocardial infarction rates. Between 1980 and 2002, mortality rates saw a decrease of approximately 49% (Wilmot 2015). Decreases in mortality was common through the world better medical intervention techniques and increase awareness of healthier lifestyle choices became more prevalent (Goldman, et al. 1984).

Unsurprisingly, as more patients survive CVD related infarction episodes, more detail has been paid to understand the survivability the time period following an episode. Wall motion score (a measure of heart contractility during cycling) was significantly higher in those that survived versus those that died (Kan, et. Al. 1986). We hope to examine several factors that determine survivability among these patients. In addition to wall motion score, we hope to stratify and understand the relationships between time to event (death) measurements compared to general heart health and age. Our goal is to describe the survivability of our dataset and provide a model to predict the factors that determine survivability in the one-year period following a myocardial infarction episode.

Dataset

Our data was obtained our data set from Kaggle via the Reed Institute. The data set contains 133 total patient observations across 8 variables: status at the end of the survival period, age, presence of pericardial effusion, fractal shortening, EPSS, wall motion score, wall motion index, and alive at the end of one year. Three patient survival times were not given; thus, we elected to remove those values to develop the most accurate portrayal of survival times.

Since the time of myocardial infarction varies (depending if a patient joined the study prior to the start), some patients were followed for less than a year. This provides a clear censoring and truncation. We discuss the nature of censoring in the following section.

At this point, 40 points of data were missing from the total dataset. A random forest algorithm (see: missForest package) was employed to iteratively impute values. With this in mind, our predictive and summary models will have less than ideal accuracy.

We then classify continuous variables into groups for stratification.

Age is divided into three groups with 0 denoting younger than 55 years, 1 denoting 55 to 70 years, and 2 denoting older than 70 years. Pericardial effusion is already grouped into binary values with 0 denoting the absence of fluid while 1 denotes the presence of the effusion. Finally, wall motion score is divided into three groups: 0 denoting scores less than 12, 1 denoting scores of 12 - 17, and 2 denoting scores greater than 14. These groups are summarized in the table below:

Indicator Age Effusion WMS 1 0 < 55 Years Fluid is absent < 12 2 1 55-70 Years Fluid is present 12-14 3 2 > 70 Years > 14

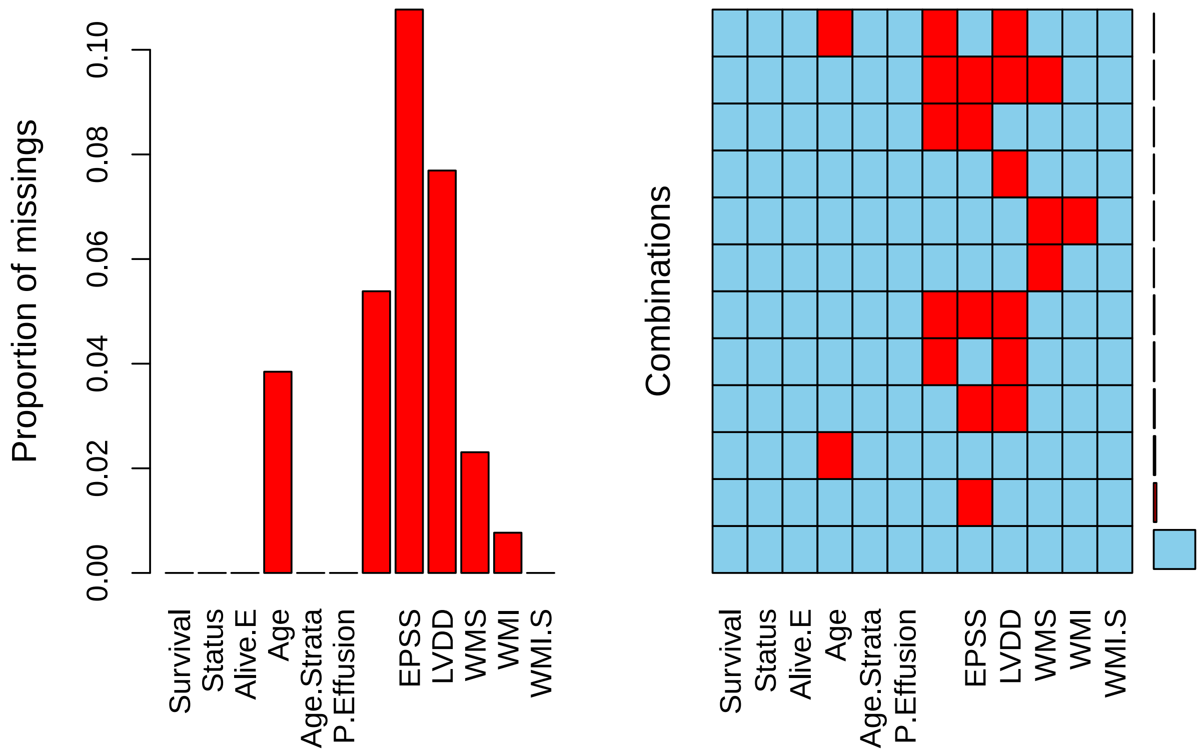
Table 1: Stratification Groupings

Indicator	Age	Effusion	WMS
0	< 55 Years	Fluid is absent	< 12
1	55-70 Years	Fluid is present	12-14
2	> 70 Years		> 14

The reader may find a summary of tables and original dataset in the appendix of this paper.

Imputation

In addition to the two rows that we removed, we further modified the dataset. The provided data contains 40 missing values that we chose to impute using the random forest algorithm methods in the missForest R package. Below is a summary of the missing data:



missForest iteration 1 in progress...done! missForest iteration 2 in progress...done! missForest iteration 3 in progress...done!

We leverage the missForest package that uses algorithmic process used here uses a modified k-nearest neighbor (KNN) approach. Using a training data set, the routines of the algorithm predicts the missing values trained on the observed parts of the dataset (Stekhoven 2012). The process checks each iteration for an acceptable amount of error. If an iteration produces an error that is smallest than that last iteration, then the algorithm continues to function. This progress stops when an error is larger than the previous iteration. Refer to Stekhoven, et. al 2012 for more detail.

We used the missFortune package to run up to 500 iterations. Each iteration was allotted 1000 trees for the random forest algorithmic approach.

Following imputation, we verify the imputation accuracy using the normalized root mean squared error as an indicator of accuracy (NRMSE, Oba et al. (2003)). The general performance of our imputed dataset can be expressed by:

$$NRMSE = \sqrt{\frac{mean\left(\left(X^{true} - X^{imp}\right)^2\right)}{var\left(X^{true}\right)}}$$

Where X is a matrix of our dataset. Being a random forest iterative process, each imputed dataset will be different from each other. For our particular seed and iterations, we obtained a NRMSE value of 0.1442 - that is our inputted values have an estimate 14.42% deviation from estimated true accuracy.

The full imputed dataset may be found in the appendix of this paper. As well as references to the authors who created the algorithm.

Censoring

Our dataset has numerous censored valued - that is, valued that cannot be recorded due the constraint of the study design. In our data set, we are examining the survival after a heart attack, that is, the event of interest is death given that a patient has had already survived a heart attack (left truncation).

We have fixed start and end dates for when the data was collection. Some patients joined when the study began. Others joined later after the start date. Because of this, we cannot accurately determine how long a patient survived after our observation period is over. In addition, there are some patients that have been lost to follow up or may have died due to the onset of other unrelated factors. These data present themselves as being randomly right censored.

Methodology

Here, we briefly review the methodology and theory behind our analysis techniques for context.

Non-Parametric: Kaplan Meier

We use Kaplan-Meier (KM) survival estimators to model a step curve for the survival of our censored dataset. The KM estimator is an adjustment of an empirical survival function to reflect the presence of right-censored observations (Tableman & Kim, 2004). The estimator can be described in the following equation:

$$\hat{S}(t) = \prod_{y_{(i)} \leq t}^{k} p_i = \prod_{i=1}^{k} (\frac{n_i - d_i}{n_i})$$

Where n_i is the number alive before time y_i and d_i is the number of events during during that interval. In our case, y_i is the specific patient being observed, n_i is the number of patients alive at time y_i. With $k = 131$, our KM equation is:

$$\hat{S}(t) = \prod_{i=1}^{131} (\frac{n_i - d_i}{n_i})$$

We use this equation to estimate the survival at each time interval. We conduct this analysis for the whole data set and then choose to stratify on age, pericardial effusion presence, and wall motion score. We also include cumulative hazard estimators based on the KM fit.

Cumulative Hazard Estimator

We calculate the hazard of our Kaplan-Survivor function by observing standard cumulative hazard estimate (shown below):

$$\hat{H}(t) = -logS(t) = -log \prod_{y_{(i)} \leq t} \frac{d_i - n_i}{n_i}$$

Intuitively, the relationship of the observed hazard is the negative log of the survival function at each interval. We can clearly see a graphical relationship between our survival by examining our hazard plots in the results section. There was the possibility of using Nelson-Aalen's approximation for hazard, but we find that the computation is trivial.

Parametric Estimates: Log-Normal Distribution

Parametric Estimates: Log-Logistic Distribution

Parametric Estimates: Weibull Distribution

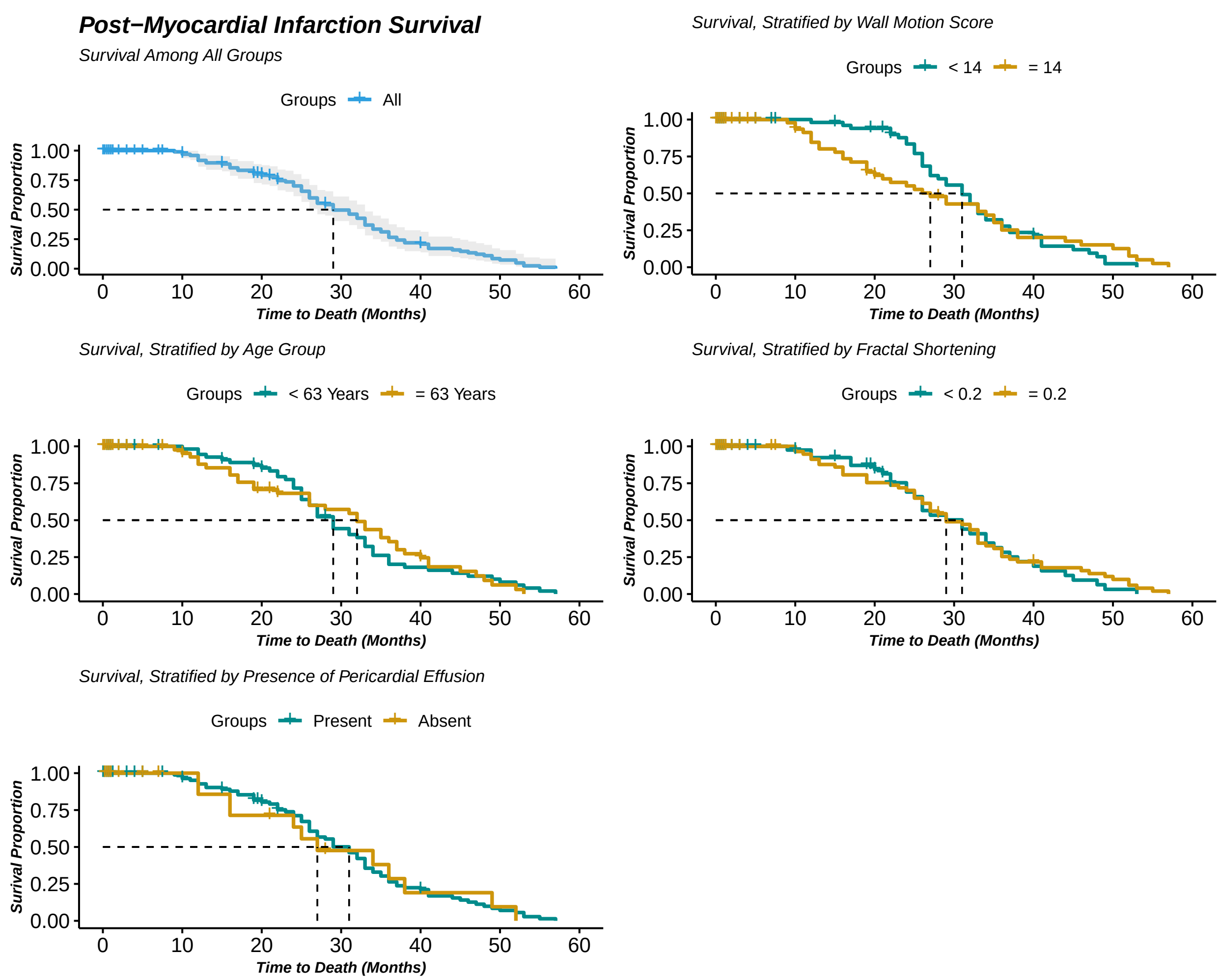
Hazard Estimates

Regression & Model Selection

Results

Non-Parametric: Kaplan-Meier Survival Estimates

Output from our Kaplan-Meier estimators give us the following curve (full KM estimator table can be found in the appendix).

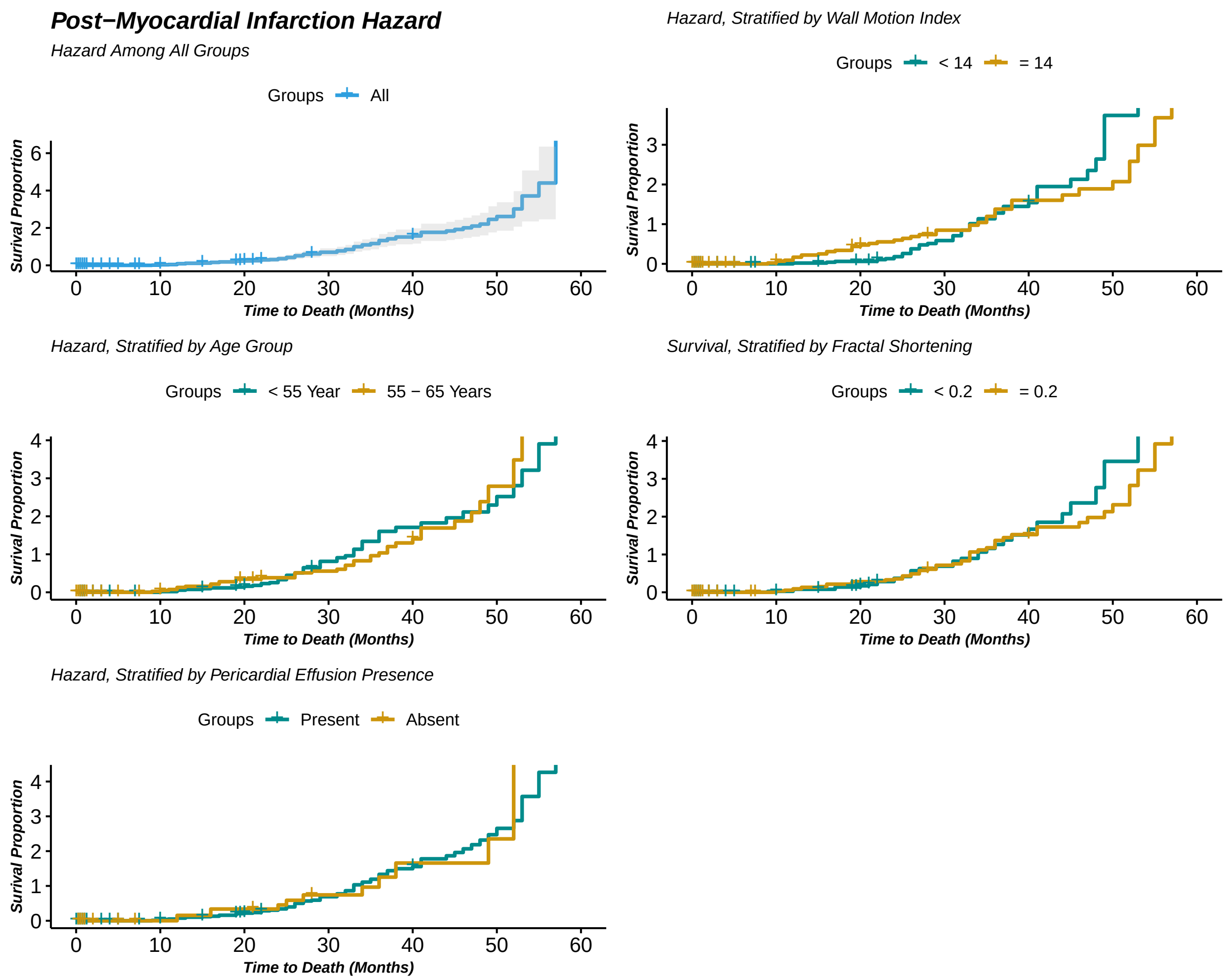


Here is the survival curve for all groups within our dataset. With time spanning to a maximum of 57 months, we have a mean survival time of approximately 30.5 months. The median survival time is approximately 29 months a 95% lower and upper confidence limits of 27 and 33 months, respectively. We see that a majority of our censored values have very short survival times. This is an intuitive finding as patients who joined the study late have relatively unknown survival (hence, their right censored nature. A summary of the results from our KM fit is below:

Table 2: Kaplan-Meier Results

	Records	Events	Mean	Median	Median 0.95 LCL	Median 0.95 UCL
All	130	88	30.53	29	27	33
Age:<55	66	51	30.47	29	26	33
Age: 55-65	64	37	30.60	32	26	37
P.Eff: Absent	106	76	30.63	31	27	33
P.Eff: Present	24	12	29.94	27	24	NA
WMS: < 14	62	46	32.17	31	27	34
WMS: $= 14$	68	42	28.61	27	20	35
Fractal Shortening: < 0.2	61	33	30.45	31	26	36
Fractal Shortening: $= 0.2$	69	55	30.44	29	26	33

Cumulative Hazard Estimates



Hazard stratified by pericardial effusion shows relatively similar results between both groups. Age stratified risk shows more risk for the younger groups. Wall motion scores stratification shows a slight increase in risk past the median for higher scores.

Parametric: Log-Normal Model

Parametric: Log-Logistic Model

Parametric: Weibull Model

Cox Proportional Hazard

Model Diagnostics

AIC, BIC, and Confidence Intervals

Residual Analysis/QQ Plot

Discussion

[Add F-shortening discussion]

Explanation of results

In our non-parametric analysis, median survival times for nearly all of the stratification elements show similar results. The exception of these results are patients with pericardial effusion present and high wall motion scores. Well functioning hearts generally do not have this effusion present as [insert additional explanation of the pathology]. Wall motion scores reflecting heart muscle activity outside of the 12-14 range indicate abnormal heart function. This too, may be hinting at the survivability of post-infarction patients.

An interesting observation of non-parametric analysis when stratified on age shows a lower survivability in the younger age groups. Intuitively, we would expect higher survivability given that younger patients have more robust bodies. An explanation could be the externalities attached to having a heart attack at such young ages. For example, if a patient is young and at risk for heart attacks, then it's likely that that patient is already at risk for other factors, increasing the relative risk for that group.

Summary of Limitations

Very clearly, our data is smaller than we hoped for. To project a large regression fit to reliably predict variables, we would require a much larger dataset that also includes the aforementioned sex discrepancies. Additionally, none of our variables have a significant effect on our model. This suggests that either (1) our data set is too small to develop a concrete prediction model or (2) our chosen recorded variables actually do not effect the survivability of our patients. In examination of stratification shows relative unequal distribution among groups. For example, when examining pericardial effusion, we compare 106 records of with effusion absent to 24 records of effusion present. Such comparisons are unequal comparisons.

Conclusion

We can conclude that future studies should definitely expand on collection of variables that could potentially influence survivability. Interesting effects may include poverty, diet, ethnicity, race, sex, and even work place stress/effects.

Appendix

References

Andrikopoulos, G. K., Tzeis, S. E., Pipilis, A. G., Richter, D. J., Kappos, K. G., Stefanadis, C. I., . . . Chimonas, E. T. (2006). Younger age potentiates post myocardial infarction survival disadvantage of women. International Journal of Cardiology, 108(3), 320–325. doi: 10.1016/j.ijcard.2005.05.016

Fryar CD, Chen T-C, Li X. Prevalence of uncontrolled risk factors for cardiovascular disease: United States, 1999–2010 pdf icon[PDF-494K]. NCHS data brief, no. 103. Hyattsville, MD: National Center for Health Statistics; 2012. Accessed May 9, 2019.

Ford, E. S., & Capewell, S. (2007). Coronary Heart Disease Mortality Among Young Adults in the U.S. From 1980 Through 2002. Journal of the American College of Cardiology, 50(22), 2128–2132. doi: 10.1016/j.jacc.2007.05.056

Goldman, L. (1984). The Decline in Ischemic Heart Disease Mortality Rates. Annals of Internal Medicine, 101(6), 825. doi: 10.7326/0003-4819-101-6-825 Gu K, Cowie CC, Harris MI. Diabetes and Decline in Heart Disease Mortality in US Adults. JAMA. 1999;281(14):1291–1297. doi:10.1001/jama.281.14.1291 Heron, M. Deaths: Leading causes for 2017 pdf icon[PDF – 3 M]. National Vital Statistics Reports;68(6). Accessed November 19, 2019.

Kan, G., Visser, C., Kooler, J., & Dunning, A. (1986). Short and long term predictive value of wall motion score in acute myocardial infarction. British Heart Journal, 56, 422-427. Oba S, Sato MA, Takemasa I, Monden M, Matsubara K, Ishii S. A Bayesian missing value estimation method for gene expression profile data. Bioinformatics. 2003;19(16):2088-2096. doi:10.1093/bioinformatics/btg287

Rimm, E. B., Stampfer, M. J., Giovannucci, E., Ascherio, A., Spiegelman, D., Colditz, G. A., & Willett, W. C. (1995). Body Size and Fat Distribution as Predictors of Coronary Heart Disease among Middle-aged and Older US Men. American Journal of Epidemiology, 141(12), 1117–1127. doi: 10.1093/oxfordjournals.aje.a117385

Salzberg, S. (1988). Exemplar-based learning: Theory and implementation (Technical Report TR-10-88). Harvard University, Center for Research in Computing Technology, Aiken Computation Laboratory (33 Oxford Street; Cambridge, MA 02138).

Sia, Y. T., Parker, T. G., Liu, P., Tsoporis, J. N., Adam, A., & Rouleau, J. L. (2002). Improved post-myocardial infarction survival with probucol in rats: Effects on left ventricular function, morphology, cardiac oxidative stress and cytokine expression. Journal of the American College of Cardiology, 39(1), 148–156. doi: 10.1016/s0735-1097(01)01709-0

Stekhoven, D. J., & Buhlmann, P. (2011). MissForest–non-parametric missing value imputation for mixed-type data. Bioinformatics, 28(1), 112–118. doi: 10.1093/bioinformatics/btr597

Tableman, M., & Kim, J. S. (2004). Survival analysis using S: analysis of time-to-event data. Boca Raton, Florida: Chapman & Hall.

Wilmot, K. A., O'Flaherty, M., Capewell, S., Ford, E. S., & Vaccarino, V. (2015). Coronary Heart Disease Mortality Declines in the United States From 1979 Through 2011CLINICAL PERSPECTIVE. Circulation, 132(11), 997–1002. doi: 10.1161/circulationaha.115.015293

Dataset Variable Summary

Table 3: Summary of Dataset Covariates

Variable	Label	Definition
Survival	Survival	The number of months the patints survived, post-myocardial infarction.
Status	Status	Censorship status. 0 denotes that a patient is a censored while 1 denotes that a patient is uncensored.
Alive at the end of Survival Period	Alive.E	Binary variable. 0 denotes that patient is alive at the end of the survival period while 1 indicates that a patient is still alive.
Patient Age	Age	The age in years when a myocardial infarction occurs.
Age Group	Age.S	0 denotes younger than 55 years . 1 denotes 55 - 70 years. 2 denotes older than 70
Pericardial Effusion	P.Effusion	Binary variable. Pericardial effusion is excess fluid surrounding the heart. Though excess is not harmful, it is sometimes indicates a porly functioning heart. 0 denotes that pericardial effusion is absent while 1 denotes that fluid is present.
Fractional Shortening	F.Shortening	Fractional shortening is a measure of contractility around the heart. Generally, lower numbers are considered to be abnormal.
E-Point Septal Separation	EPSS	E-point septal separation is an addition measure of heart contractivity. Larger numbers are considered to be abnormal.
Left Ventricular End-Diastolic Dimension	LVDD	Left ventricular end-diastolic dimension is the measure of the heart at the end of disatole. The larger this value is indicates a larger heart. Larger hearts are generally in poor health.
Wall Motion Score	WMS	Wall motion score is a measure of how the segments of the left ventricle are moving during systol.

Table 3: Summary of Dataset Covariates *(continued)*

Variable	Label	Definition
Wall Motion Index	WMI	Wall motion index is the wall motion score divided by the number of segments that are moving. Normally, 12-13 segments can be seen in an echocardiogram.
Wall Motion Strata	WMS.S	0 denotes score less than 11, 1 denotes score 12-14, 2 denotes score greater than 14

Original Dataset

Table 4: Original Dataset

Survival	Status	Alive.E	Age	Age.Strata	P.Effusion	F.Shortening	EPSS	LVDD	WMS	WMI	WMI.S
11.00	1	0	71.00	2	0	0.260	9.000	4.600	14.00	1.000	0
19.00	1	0	72.00	2	0	0.380	6.000	4.100	14.00	1.700	1
16.00	1	0	55.00	1	0	0.260	4.000	3.420	14.00	1.000	0
57.00	1	0	60.00	1	0	0.253	12.062	4.603	16.00	1.450	1
19.00	0	1	57.00	1	0	0.160	22.000	5.750	18.00	2.250	1
26.00	1	0	68.00	2	0	0.260	5.000	4.310	12.00	1.000	0
13.00	1	0	62.00	1	0	0.230	31.000	5.430	22.50	1.875	1
50.00	1	0	60.00	1	0	0.330	8.000	5.250	14.00	1.000	0
19.00	1	0	46.00	0	0	0.340	0.000	5.090	16.00	1.140	0
25.00	1	0	54.00	1	0	0.140	13.000	4.490	15.50	1.190	0
10.00	0	1	77.00	2	0	0.130	16.000	4.230	18.00	1.800	1
52.00	1	0	62.00	1	1	0.450	9.000	3.600	16.00	1.140	0
52.00	1	0	73.00	2	0	0.330	6.000	4.000	14.00	1.000	0
44.00	1	0	60.00	1	0	0.150	10.000	3.730	14.00	1.000	0
0.50	0	1	62.00	1	0	0.120	23.000	5.800	11.67	2.330	1
24.00	1	0	55.00	1	1	0.250	12.063	4.290	14.00	1.000	0
0.50	0	1	69.00	2	1	0.260	11.000	4.650	18.00	1.640	1
0.50	0	1	62.53	1	1	0.070	20.000	5.200	24.00	2.000	1
22.00	0	1	66.00	2	0	0.090	17.000	5.819	8.00	1.333	1
1.00	0	1	66.00	2	1	0.220	15.000	5.400	27.00	2.250	1
0.75	0	1	69.00	2	0	0.150	12.000	5.390	19.50	1.625	1
0.75	0	1	85.00	2	1	0.180	19.000	5.460	13.83	1.380	1
0.50	0	1	73.00	2	0	0.230	12.733	6.060	7.50	1.500	1
5.00	0	1	71.00	2	0	0.170	0.000	4.650	8.00	1.000	0
48.00	1	0	64.00	1	0	0.190	5.900	3.480	10.00	1.110	0
29.00	1	0	54.00	1	0	0.300	7.000	3.850	10.00	1.667	1
29.00	1	0	35.00	0	0	0.300	5.000	4.170	14.00	1.000	0
29.00	1	0	55.00	1	0	NA	7.000	NA	2.00	1.000	0
0.25	0	1	75.00	2	0	NA	NA	NA	NA	1.000	0
36.00	1	0	55.00	1	1	0.210	4.200	4.160	14.00	1.560	1
1.00	0	1	65.00	2	0	0.150	NA	5.050	10.00	1.000	0

Table 4: Original Dataset *(continued)*

Survival	Status	Alive.E	Age	Age.Strata	P.Effusion	F.Shortening	EPSS	LVDD	WMS	WMI	WMI.S
1.00	0	1	52.00	1	1	0.170	17.200	5.320	14.00	1.170	0
3.00	0	1	NA	2	0	NA	12.000	NA	6.00	3.000	1
27.00	1	0	47.00	0	0	0.400	5.120	3.100	12.00	1.000	0
35.00	1	0	63.00	1	0	NA	10.000	NA	14.00	1.170	0
26.00	1	0	61.00	1	0	0.610	13.100	4.070	13.00	1.625	1
16.00	1	0	63.00	1	1	NA	NA	5.310	5.00	1.000	0
1.00	0	1	65.00	2	0	0.060	23.600	NA	21.50	2.150	1
19.00	1	0	68.00	2	0	0.510	NA	3.880	15.00	1.670	1
31.00	1	0	80.00	2	0	0.410	5.400	4.360	NA	1.000	0
32.00	1	0	54.00	1	0	0.350	9.300	3.630	11.00	1.222	0
16.00	1	0	70.00	2	1	0.270	4.700	4.490	22.00	2.000	1
40.00	1	0	79.00	2	0	0.150	17.500	4.270	13.00	1.300	1
46.00	1	0	56.00	1	0	0.330	NA	3.590	14.00	1.000	0
2.00	0	1	67.00	2	1	0.440	9.000	3.960	17.50	1.450	1
37.00	1	0	64.00	1	0	0.090	NA	NA	12.00	2.000	1
19.50	0	1	81.00	2	0	0.120	NA	NA	9.00	1.250	0
20.00	0	1	59.00	1	0	0.030	21.300	6.290	17.00	1.310	1
0.25	0	1	63.00	1	1	NA	NA	NA	23.00	2.300	1
2.00	0	1	56.00	1	1	0.040	14.000	5.000	NA	NA	1
7.00	0	1	61.00	1	1	0.270	NA	NA	9.00	1.500	1
10.00	1	0	57.00	1	0	0.240	14.800	5.260	18.00	1.380	1
12.00	1	0	58.00	1	0	0.300	9.400	3.490	14.00	1.000	0
1.00	0	1	60.00	1	0	0.010	24.600	5.650	39.00	3.000	1
10.00	1	0	66.00	2	0	0.290	15.600	6.150	14.00	1.000	0
45.00	1	0	63.00	1	0	0.150	13.000	4.570	13.00	1.080	0
22.00	1	0	57.00	1	0	0.130	18.600	4.370	12.33	1.370	1
53.00	1	0	70.00	2	0	0.100	9.800	5.300	23.00	2.300	1
38.00	1	0	68.00	2	0	0.290	NA	4.410	14.00	1.167	0
26.00	1	0	79.00	2	0	0.170	11.900	5.150	10.50	1.050	0
9.00	1	0	73.00	2	0	0.120	NA	6.780	16.67	1.390	1
26.00	1	0	72.00	2	0	0.187	12.000	5.020	13.00	1.180	0
0.50	0	1	59.00	1	0	0.130	16.400	4.960	17.83	1.370	1
12.00	1	0	67.00	2	1	0.110	10.300	4.680	11.00	1.000	0
49.00	1	0	51.00	1	0	0.160	13.200	5.260	11.00	1.000	0
0.75	0	1	50.00	1	0	0.140	11.400	4.750	10.00	2.500	1
49.00	1	0	70.00	2	1	0.250	9.700	5.570	5.50	1.100	0
47.00	1	0	65.00	2	0	0.360	8.800	5.780	12.00	1.000	0
41.00	1	0	78.00	2	0	0.060	16.100	5.620	13.67	1.367	1
0.25	0	1	86.00	2	0	0.225	12.200	5.200	24.00	2.180	1
33.00	1	0	56.00	1	0	0.250	11.000	4.720	11.00	1.000	0
29.00	1	0	60.00	1	0	0.120	10.200	4.310	15.00	1.670	1
41.00	1	0	59.00	1	0	0.290	7.500	4.750	13.00	1.080	0
26.00	1	0	50.00	1	0	0.060	30.100	5.950	21.50	2.390	1
15.00	1	0	54.00	1	0	0.217	17.900	4.540	16.50	1.180	0
0.25	0	1	68.00	2	0	0.220	21.700	4.850	15.00	1.150	0
0.03	0	1	NA	2	0	0.260	19.400	4.770	21.00	2.100	1

Table 4: Original Dataset *(continued)*

Survival	Status	Alive.E	Age	Age.Strata	P.Effusion	F.Shortening	EPSS	LVDD	WMS	WMI	WMI.S
12.00	1	0	64.00	1	0	0.200	7.100	4.580	14.00	1.000	0
32.00	1	0	63.00	1	0	0.200	5.000	5.200	8.00	1.000	0
32.00	1	0	65.00	2	0	0.060	23.600	6.740	12.00	1.090	0
27.00	1	0	54.00	1	1	0.070	16.800	4.160	18.00	1.500	1
23.00	1	0	62.00	1	0	0.250	6.000	4.480	11.00	1.000	0
0.75	0	1	78.00	2	0	0.050	10.000	4.440	15.00	1.360	1
0.75	0	1	61.00	1	0	NA	NA	NA	28.00	2.330	1
34.00	1	0	52.00	1	0	0.140	25.000	6.210	11.50	1.150	0
1.00	0	1	73.00	2	0	0.050	14.800	4.140	15.50	1.410	1
21.00	0	1	70.00	2	1	0.160	19.200	5.250	11.00	1.000	0
55.00	1	0	55.00	1	0	0.280	5.500	4.480	22.00	1.830	1
15.00	0	1	60.00	1	0	0.180	8.700	4.560	13.50	1.040	0
0.50	0	1	67.00	2	0	0.155	11.300	5.160	13.00	1.000	0
35.00	1	0	64.00	1	0	0.300	6.600	4.360	14.00	1.270	0
53.00	1	0	59.00	1	0	0.344	9.100	4.040	9.00	1.000	0
33.00	1	0	46.00	0	0	0.272	16.500	5.360	12.67	1.060	0
33.00	1	0	63.00	1	0	0.250	5.600	3.870	18.00	1.500	1
40.00	0	1	74.00	2	0	0.200	4.800	4.560	12.50	1.040	0
33.00	1	0	59.00	1	0	0.500	9.100	3.420	18.00	1.500	1
5.00	0	1	65.00	2	1	0.160	8.500	5.470	16.00	1.450	1
4.00	0	1	58.00	1	0	0.170	28.900	6.730	26.08	2.010	1
31.00	1	0	53.00	1	0	0.170	NA	4.690	10.00	1.000	0
33.00	1	0	66.00	2	0	0.200	NA	4.230	12.00	1.000	0
22.00	1	0	70.00	2	0	0.380	0.000	4.550	10.00	1.000	0
25.00	1	0	62.00	1	0	0.258	11.800	4.870	11.00	1.000	0
1.25	0	1	63.00	1	0	0.300	6.900	3.520	18.16	1.510	1
24.00	1	0	59.00	1	0	0.170	14.300	5.490	13.50	1.500	1
25.00	1	0	57.00	1	0	0.228	9.700	4.290	11.00	1.000	0
24.00	1	0	57.00	1	0	0.036	7.000	4.120	13.50	1.230	0
0.75	0	1	78.00	2	0	0.230	40.000	6.230	14.00	1.400	1
3.00	0	1	62.00	1	0	0.260	7.600	4.420	14.00	1.000	0
27.00	1	0	62.00	1	0	0.220	12.100	3.920	11.00	1.000	0
13.00	1	0	66.00	2	0	0.240	13.600	4.380	22.00	2.200	1
36.00	1	0	61.00	1	0	0.270	9.000	4.060	12.00	1.000	0
25.00	1	0	59.00	1	1	0.400	9.200	5.360	12.00	1.000	0
27.00	1	0	57.00	1	0	0.290	9.400	4.770	9.00	1.000	0
34.00	1	0	62.00	1	1	0.190	28.900	6.630	19.50	1.950	1
37.00	1	0	NA	2	0	0.260	0.000	4.380	9.00	1.000	0
34.00	1	0	54.00	1	0	0.430	9.300	4.790	10.00	1.000	0
28.00	0	1	62.00	1	1	0.240	28.600	5.860	21.50	1.950	1
28.00	1	0	NA	2	0	0.230	19.100	5.490	12.00	1.200	0
17.00	1	0	64.00	1	0	0.150	6.600	4.170	14.00	1.270	0
38.00	1	0	57.00	1	1	0.120	0.000	2.320	16.50	1.375	1
31.00	1	0	61.00	1	0	0.180	0.000	4.480	11.00	1.375	1
12.00	1	0	61.00	1	1	0.190	13.200	5.040	19.00	1.730	1
36.00	1	0	48.00	0	0	0.150	12.000	3.660	10.00	1.000	0

Table 4: Original Dataset *(continued)*

Survival	Status	Alive.E	Age	Age.Strata	P.Effusion	F.Shortening	EPSS	LVDD	WMS	WMI	WMI.S
17.00	1	0	NA	2	0	0.090	6.800	4.960	13.00	1.080	0
21.00	1	0	61.00	1	0	0.140	25.500	5.160	14.00	1.270	0
7.50	0	1	64.00	1	0	0.240	12.900	4.720	12.00	1.000	0
41.00	1	0	64.00	1	0	0.280	5.400	5.470	11.00	1.100	0
36.00	1	0	69.00	2	0	0.200	7.000	5.050	14.50	1.210	0
22.00	1	0	57.00	1	0	0.140	16.100	4.360	15.00	1.360	1
20.00	1	0	62.00	1	0	0.150	0.000	4.510	15.50	1.409	1

Imputed Dataset

Table 5: Imputed Dataset

Survival	Status	Alive.E	Age	P.Effusion	F.Shortening	EPSS	LVDD	WMS	WMI	Age.s	WMS.s	F.Short.s	LVDD.s	EPSS.s
11.00	1	0	71.00	0	0.26	9.00	4.60	14.00	1.00	1	1	1	0	0
19.00	1	0	72.00	0	0.38	6.00	4.10	14.00	1.70	1	1	1	0	0
16.00	1	0	55.00	0	0.26	4.00	3.42	14.00	1.00	0	1	1	0	0
57.00	1	0	60.00	0	0.25	12.06	4.60	16.00	1.45	0	1	1	0	1
19.00	0	1	57.00	0	0.16	22.00	5.75	18.00	2.25	0	1	0	1	1
26.00	1	0	68.00	0	0.26	5.00	4.31	12.00	1.00	1	0	1	0	0
13.00	1	0	62.00	0	0.23	31.00	5.43	22.50	1.88	0	1	1	1	1
50.00	1	0	60.00	0	0.33	8.00	5.25	14.00	1.00	0	1	1	1	0
19.00	1	0	46.00	0	0.34	0.00	5.09	16.00	1.14	0	1	1	1	0
25.00	1	0	54.00	0	0.14	13.00	4.49	15.50	1.19	0	1	0	0	1
10.00	0	1	77.00	0	0.13	16.00	4.23	18.00	1.80	1	1	0	0	1
52.00	1	0	62.00	1	0.45	9.00	3.60	16.00	1.14	0	1	1	0	0
52.00	1	0	73.00	0	0.33	6.00	4.00	14.00	1.00	1	1	1	0	0
44.00	1	0	60.00	0	0.15	10.00	3.73	14.00	1.00	0	1	0	0	0
0.50	0	1	62.00	0	0.12	23.00	5.80	11.67	2.33	0	0	0	1	1
24.00	1	0	55.00	1	0.25	12.06	4.29	14.00	1.00	0	1	1	0	1
0.50	0	1	69.00	1	0.26	11.00	4.65	18.00	1.64	1	1	1	0	0
0.50	0	1	62.53	1	0.07	20.00	5.20	24.00	2.00	0	1	0	1	1
22.00	0	1	66.00	0	0.09	17.00	5.82	8.00	1.33	1	0	0	1	1
1.00	0	1	66.00	1	0.22	15.00	5.40	27.00	2.25	1	1	1	1	1
0.75	0	1	69.00	0	0.15	12.00	5.39	19.50	1.62	1	1	0	1	1
0.75	0	1	85.00	1	0.18	19.00	5.46	13.83	1.38	1	0	0	1	1
0.50	0	1	73.00	0	0.23	12.73	6.06	7.50	1.50	1	0	1	1	1
5.00	0	1	71.00	0	0.17	0.00	4.65	8.00	1.00	1	0	0	0	0
48.00	1	0	64.00	0	0.19	5.90	3.48	10.00	1.11	1	0	0	0	0
29.00	1	0	54.00	0	0.30	7.00	3.85	10.00	1.67	0	0	1	0	0
29.00	1	0	35.00	0	0.30	5.00	4.17	14.00	1.00	0	1	1	0	0
29.00	1	0	55.00	0	0.27	7.00	4.57	2.00	1.00	0	0	1	0	0
0.25	0	1	75.00	0	0.18	12.09	5.04	12.27	1.00	1	0	0	1	1
36.00	1	0	55.00	1	0.21	4.20	4.16	14.00	1.56	0	1	1	0	0
1.00	0	1	65.00	0	0.15	11.61	5.05	10.00	1.00	1	0	0	1	1
1.00	0	1	52.00	1	0.17	17.20	5.32	14.00	1.17	0	1	0	1	1
3.00	0	1	68.34	0	0.15	12.00	5.27	6.00	3.00	1	0	0	1	1
27.00	1	0	47.00	0	0.40	5.12	3.10	12.00	1.00	0	0	1	0	0
35.00	1	0	63.00	0	0.19	10.00	4.43	14.00	1.17	1	1	0	0	0
26.00	1	0	61.00	0	0.61	13.10	4.07	13.00	1.62	0	0	1	0	1
16.00	1	0	63.00	1	0.20	9.83	5.31	5.00	1.00	1	0	1	1	0
1.00	0	1	65.00	0	0.06	23.60	5.66	21.50	2.15	1	1	0	1	1
19.00	1	0	68.00	0	0.51	7.44	3.88	15.00	1.67	1	1	1	0	0
31.00	1	0	80.00	0	0.41	5.40	4.36	11.68	1.00	1	0	1	0	0
32.00	1	0	54.00	0	0.35	9.30	3.63	11.00	1.22	0	0	1	0	0
16.00	1	0	70.00	1	0.27	4.70	4.49	22.00	2.00	1	1	1	0	0
40.00	1	0	79.00	0	0.15	17.50	4.27	13.00	1.30	1	0	0	0	1
46.00	1	0	56.00	0	0.33	8.05	3.59	14.00	1.00	0	1	1	0	0
2.00	0	1	67.00	1	0.44	9.00	3.96	17.50	1.45	1	1	1	0	0

Table 5: Imputed Dataset *(continued)*

Survival	Status	Alive.E	Age	P.Effusion	F.Shortening	EPSS	LVDD	WMS	WMI	Age.s	WMS.s	F.Short.s	LVDD.s	EPSS.s
37.00	1	0	64.00	0	0.09	12.48	4.75	12.00	2.00	1	0	0	1	1
19.50	0	1	81.00	0	0.12	12.39	5.03	9.00	1.25	1	0	0	1	1
20.00	0	1	59.00	0	0.03	21.30	6.29	17.00	1.31	0	1	0	1	1
0.25	0	1	63.00	1	0.15	17.62	5.23	23.00	2.30	1	1	0	1	1
2.00	0	1	56.00	1	0.04	14.00	5.00	17.30	1.65	0	1	0	1	1
7.00	0	1	61.00	1	0.27	11.22	4.86	9.00	1.50	0	0	1	1	1
10.00	1	0	57.00	0	0.24	14.80	5.26	18.00	1.38	0	1	1	1	1
12.00	1	0	58.00	0	0.30	9.40	3.49	14.00	1.00	0	1	1	0	0
1.00	0	1	60.00	0	0.01	24.60	5.65	39.00	3.00	0	1	0	1	1
10.00	1	0	66.00	0	0.29	15.60	6.15	14.00	1.00	1	1	1	1	1
45.00	1	0	63.00	0	0.15	13.00	4.57	13.00	1.08	1	0	0	0	1
22.00	1	0	57.00	0	0.13	18.60	4.37	12.33	1.37	0	0	0	0	1
53.00	1	0	70.00	0	0.10	9.80	5.30	23.00	2.30	1	1	0	1	0
38.00	1	0	68.00	0	0.29	6.89	4.41	14.00	1.17	1	1	1	0	0
26.00	1	0	79.00	0	0.17	11.90	5.15	10.50	1.05	1	0	0	1	1
9.00	1	0	73.00	0	0.12	21.14	6.78	16.67	1.39	1	1	0	1	1
26.00	1	0	72.00	0	0.19	12.00	5.02	13.00	1.18	1	0	0	1	1
0.50	0	1	59.00	0	0.13	16.40	4.96	17.83	1.37	0	1	0	1	1
12.00	1	0	67.00	1	0.11	10.30	4.68	11.00	1.00	1	0	0	0	0
49.00	1	0	51.00	0	0.16	13.20	5.26	11.00	1.00	0	0	0	1	1
0.75	0	1	50.00	0	0.14	11.40	4.75	10.00	2.50	0	0	0	1	1
49.00	1	0	70.00	1	0.25	9.70	5.57	5.50	1.10	1	0	1	1	0
47.00	1	0	65.00	0	0.36	8.80	5.78	12.00	1.00	1	0	1	1	0
41.00	1	0	78.00	0	0.06	16.10	5.62	13.67	1.37	1	0	0	1	1
0.25	0	1	86.00	0	0.22	12.20	5.20	24.00	2.18	1	1	1	1	1
33.00	1	0	56.00	0	0.25	11.00	4.72	11.00	1.00	0	0	1	0	0
29.00	1	0	60.00	0	0.12	10.20	4.31	15.00	1.67	0	1	0	0	0
41.00	1	0	59.00	0	0.29	7.50	4.75	13.00	1.08	0	0	1	1	0
26.00	1	0	50.00	0	0.06	30.10	5.95	21.50	2.39	0	1	0	1	1
15.00	1	0	54.00	0	0.22	17.90	4.54	16.50	1.18	0	1	1	0	1
0.25	0	1	68.00	0	0.22	21.70	4.85	15.00	1.15	1	1	1	1	1
0.03	0	1	72.91	0	0.26	19.40	4.77	21.00	2.10	1	1	1	1	1
12.00	1	0	64.00	0	0.20	7.10	4.58	14.00	1.00	1	1	1	0	0
32.00	1	0	63.00	0	0.20	5.00	5.20	8.00	1.00	1	0	1	1	0
32.00	1	0	65.00	0	0.06	23.60	6.74	12.00	1.09	1	0	0	1	1
27.00	1	0	54.00	1	0.07	16.80	4.16	18.00	1.50	0	1	0	0	1
23.00	1	0	62.00	0	0.25	6.00	4.48	11.00	1.00	0	0	1	0	0
0.75	0	1	78.00	0	0.05	10.00	4.44	15.00	1.36	1	1	0	0	0
0.75	0	1	61.00	0	0.13	18.21	5.31	28.00	2.33	0	1	0	1	1
34.00	1	0	52.00	0	0.14	25.00	6.21	11.50	1.15	0	0	0	1	1
1.00	0	1	73.00	0	0.05	14.80	4.14	15.50	1.41	1	1	0	0	1
21.00	0	1	70.00	1	0.16	19.20	5.25	11.00	1.00	1	0	0	1	1
55.00	1	0	55.00	0	0.28	5.50	4.48	22.00	1.83	0	1	1	0	0
15.00	0	1	60.00	0	0.18	8.70	4.56	13.50	1.04	0	0	0	0	0
0.50	0	1	67.00	0	0.16	11.30	5.16	13.00	1.00	1	0	0	1	1
35.00	1	0	64.00	0	0.30	6.60	4.36	14.00	1.27	1	1	1	0	0

Table 5: Imputed Dataset *(continued)*

Survival	Status	Alive.E	Age	P.Effusion	F.Shortening	EPSS	LVDD	WMS	WMI	Age.s	WMS.s	F.Short.s	LVDD.s	EPSS.s
53.00	1	0	59.00	0	0.34	9.10	4.04	9.00	1.00	0	0	1	0	0
33.00	1	0	46.00	0	0.27	16.50	5.36	12.67	1.06	0	0	1	1	1
33.00	1	0	63.00	0	0.25	5.60	3.87	18.00	1.50	1	1	1	0	0
40.00	0	1	74.00	0	0.20	4.80	4.56	12.50	1.04	1	0	1	0	0
33.00	1	0	59.00	0	0.50	9.10	3.42	18.00	1.50	0	1	1	0	0
5.00	0	1	65.00	1	0.16	8.50	5.47	16.00	1.45	1	1	0	1	0
4.00	0	1	58.00	0	0.17	28.90	6.73	26.08	2.01	0	1	0	1	1
31.00	1	0	53.00	0	0.17	10.30	4.69	10.00	1.00	0	0	0	0	0
33.00	1	0	66.00	0	0.20	8.12	4.23	12.00	1.00	1	0	1	0	0
22.00	1	0	70.00	0	0.38	0.00	4.55	10.00	1.00	1	0	1	0	0
25.00	1	0	62.00	0	0.26	11.80	4.87	11.00	1.00	0	0	1	1	1
1.25	0	1	63.00	0	0.30	6.90	3.52	18.16	1.51	1	1	1	0	0
24.00	1	0	59.00	0	0.17	14.30	5.49	13.50	1.50	0	0	0	1	1
25.00	1	0	57.00	0	0.23	9.70	4.29	11.00	1.00	0	0	1	0	0
24.00	1	0	57.00	0	0.04	7.00	4.12	13.50	1.23	0	0	0	0	0
0.75	0	1	78.00	0	0.23	40.00	6.23	14.00	1.40	1	1	1	1	1
3.00	0	1	62.00	0	0.26	7.60	4.42	14.00	1.00	0	1	1	0	0
27.00	1	0	62.00	0	0.22	12.10	3.92	11.00	1.00	0	0	1	0	1
13.00	1	0	66.00	0	0.24	13.60	4.38	22.00	2.20	1	1	1	0	1
36.00	1	0	61.00	0	0.27	9.00	4.06	12.00	1.00	0	0	1	0	0
25.00	1	0	59.00	1	0.40	9.20	5.36	12.00	1.00	0	0	1	1	0
27.00	1	0	57.00	0	0.29	9.40	4.77	9.00	1.00	0	0	1	1	0
34.00	1	0	62.00	1	0.19	28.90	6.63	19.50	1.95	0	1	0	1	1
37.00	1	0	69.34	0	0.26	0.00	4.38	9.00	1.00	1	0	1	0	0
34.00	1	0	54.00	0	0.43	9.30	4.79	10.00	1.00	0	0	1	1	0
28.00	0	1	62.00	1	0.24	28.60	5.86	21.50	1.95	0	1	1	1	1
28.00	1	0	69.19	0	0.23	19.10	5.49	12.00	1.20	1	0	1	1	1
17.00	1	0	64.00	0	0.15	6.60	4.17	14.00	1.27	1	1	0	0	0
38.00	1	0	57.00	1	0.12	0.00	2.32	16.50	1.38	0	1	0	0	0
31.00	1	0	61.00	0	0.18	0.00	4.48	11.00	1.38	0	0	0	0	0
12.00	1	0	61.00	1	0.19	13.20	5.04	19.00	1.73	0	1	0	1	1
36.00	1	0	48.00	0	0.15	12.00	3.66	10.00	1.00	0	0	0	0	1
17.00	1	0	69.67	0	0.09	6.80	4.96	13.00	1.08	1	0	0	1	0
21.00	1	0	61.00	0	0.14	25.50	5.16	14.00	1.27	0	1	0	1	1
7.50	0	1	64.00	0	0.24	12.90	4.72	12.00	1.00	1	0	1	0	1
41.00	1	0	64.00	0	0.28	5.40	5.47	11.00	1.10	1	0	1	1	0
36.00	1	0	69.00	0	0.20	7.00	5.05	14.50	1.21	1	1	1	1	0
22.00	1	0	57.00	0	0.14	16.10	4.36	15.00	1.36	0	1	0	0	1
20.00	1	0	62.00	0	0.15	0.00	4.51	15.50	1.41	0	1	0	0	0

Table of Kaplan-Meier Estimators

R Code

```
knitr::opts_chunk$set(echo = TRUE)
knitr::opts_chunk$set(fig.height=4.5, fig.width=7)

library(readxl)
library(knitr)
library(tidyverse)
library(dplyr)
library(kableExtra)
library(survival)
library(survminer)
library(ggplot2)
library(VIM)
library(missForest)
library(ggplot2)
library(ggpubr)

df = data.frame(read_excel("df.xlsx"))
df[df=="?"] = " "

df.new = data.frame(read_excel("df.new.xlsx"))

s.df = Surv(df.new$Survival,df.new$Status)

Indicator = c("0","1","2")
Age = c("< 55 Years", "55-70 Years", "> 70 Years")
Effusion = c("Fluid is absent", "Fluid is present","")
WMS = c("< 12", "12-14", "> 14")

groupings = data.frame(Indicator, Age, Effusion, WMS)
groupings

kable(groupings, caption="Stratification Groupings", align="c") %>%
  kable_styling(position = "center", latex_options="hold_position")


missing.data = aggr(df) #visualize the missing information

set.seed(7522)
df.i = missForest(df, maxiter = 30, ntree = 1000)

round_df <- function(x, digits) {
  # round all numeric variables
  # x: data frame
  # digits: number of digits to round
  numeric_columns <- sapply(x, mode) == 'numeric'
  x[numeric_columns] <-  round(x[numeric_columns], digits)
  x}

df.impute = round_df(df.i$ximp,2) #imputed values table
df.new = df.impute[,c(-5,-12)] #remove incomplete strata from original data
Age.s = ifelse(df.impute$Age < 63,0,1) #new age strata based on imputed data #new age strata based on i
WMS.s = ifelse(df.impute$WMS < 14,0,1) #new WMS strata based on imputed data
```

```r
Fshort.s = ifelse(df.impute$F.Shortening < 0.2,0,1) #new fshort strata based on imputed data
LVDD.s = ifelse(df.impute$LVDD < 4.75,0,1) #new lvdd strata based on imputed data
EPSS.s = ifelse(df.impute$EPSS < 11.1,0,1)#new epss strata based on imputed data


df.new$Age.s = Age.s
df.new$WMS.s = WMS.s
df.new$F.Short.s = Fshort.s
df.new$LVDD.s = LVDD.s
df.new$EPSS.s = EPSS.s

km.plots = list()

km.all = survfit(s.df~1,type="kaplan-meier", data=df.new)
km.plots[[1]] = ggsurvplot(km.all,
                           palette = "#2E9FDF",
                           conf.int = TRUE,
                           title="Post-Myocardial Infarction Survival",
                           subtitle="Survival Among All Groups",
                           font.title=c(14,"bold.italic"),
                           font.subtitle = c(10,"italic"),
                           font.x = c(9, "bold.italic"),
                           font.y = c(9, "bold.italic"),
                           ylab="Surival Proportion",
                           xlab="Time to Death (Months)",
                           surv.median.line = "hv",
                           legend.title = "Groups",
                           legend.labs = "All")


km.age = survfit(s.df~Age.s, type="kaplan-meier", data = df.new)
km.plots[[2]] = ggsurvplot(km.age,
                           palette = c("darkcyan","darkgoldenrod3","darkorange3"),
                           subtitle="Survival, Stratified by Age Group",
                           font.subtitle = c(10,"italic"),
                           font.x = c(9, "bold.italic"),
                           font.y = c(9, "bold.italic"),
                           ylab="Surival Proportion",
                           xlab="Time to Death (Months)",
                           surv.median.line = "hv",
                           legend.title = "Groups",
                           legend.labs = c("< 63 Years","\u2265 63 Years"))

km.effusion = survfit(s.df~P.Effusion, type="kaplan-meier", data = df.new)
km.plots[[3]] = ggsurvplot(km.effusion,
                           palette = c("darkcyan","darkgoldenrod3"),
                           subtitle="Survival, Stratified by Presence of Pericardial Effusion",
                           font.subtitle = c(10,"italic"),
                           font.x = c(9, "bold.italic"),
                           font.y = c(9, "bold.italic"),
                           ylab="Surival Proportion",
                           xlab="Time to Death (Months)",
                           surv.median.line = "hv",
```

```r
                             legend.title = "Groups",
                             legend.labs = c("Present","Absent"))

km.wms = survfit(s.df~WMS.s, type="kaplan-meier", data = df.new)
km.plots[[4]] = ggsurvplot(km.wms,
                             palette = c("darkcyan","darkgoldenrod3","darkorange3"),
                             subtitle="Survival, Stratified by Wall Motion Score",
                             font.subtitle = c(10,"italic"),
                             font.x = c(9, "bold.italic"),
                             font.y = c(9, "bold.italic"),
                             ylab="Surival Proportion",
                             xlab="Time to Death (Months)",
                             surv.median.line = "hv",
                             legend.title = "Groups",
                             legend.labs = c("< 14","\u2265 14"))

km.fshort = survfit(s.df~Fshort.s, type="kaplan-meier", data = df.new)
km.plots[[5]] = ggsurvplot(km.fshort,
                             palette = c("darkcyan","darkgoldenrod3"),
                             subtitle="Survival, Stratified by Fractal Shortening",
                             font.subtitle = c(10,"italic"),
                             font.x = c(9, "bold.italic"),
                             font.y = c(9, "bold.italic"),
                             ylab="Surival Proportion",
                             xlab="Time to Death (Months)",
                             surv.median.line = "hv",
                             legend.title = "Groups",
                             legend.labs = c("< 0.2","\u2265 0.2"))


arrange_ggsurvplots(km.plots, print=TRUE, ncol=2, nrow=3)


ks1 = data.frame(t(summary(km.all)$table))
ks2 = data.frame(summary(km.age)$table)
ks3 = data.frame(summary(km.effusion)$table)
ks4 = data.frame(summary(km.wms)$table)
ks5 = data.frame(summary(km.fshort)$table)

ksall = rbind(ks1, ks2, ks3, ks4,ks5)
km.as = ksall[,c(1,4,5,7,8,9)]
colnames(km.as) = c("Records","Events","Mean","Median","Median 0.95 LCL","Median 0.95 UCL")
rownames(km.as) = c("All","Age:<55", "Age: 55-65", "P.Eff: Absent","P.Eff: Present", "WMS: < 14", "WMS:

kable(km.as, caption="Kaplan-Meier Results",align="c", digits=2) %>%
  kable_styling(position = "center", latex_options="hold_position")
haz.plots = list()

haz.plots[[1]] = ggsurvplot(km.all,
           fun = "cumhaz",
           palette = "#2E9FDF",
           conf.int = TRUE,
           title="Post-Myocardial Infarction Hazard",
```

```r
           subtitle="Hazard Among All Groups",
           font.title=c(14,"bold.italic"),
           font.subtitle = c(10,"italic"),
           font.x = c(9, "bold.italic"),
           font.y = c(9, "bold.italic"),
           ylab="Surival Proportion",
           xlab="Time to Death (Months)",
           legend.title = "Groups",
           legend.labs = "All")

haz.plots[[2]] = ggsurvplot(km.age,
           fun = "cumhaz",
           palette = c("darkcyan","darkgoldenrod3"),
           subtitle="Hazard, Stratified by Age Group",
           font.subtitle = c(10,"italic"),
           font.x = c(9, "bold.italic"),
           font.y = c(9, "bold.italic"),
           ylab="Surival Proportion",
           xlab="Time to Death (Months)",
           legend.title = "Groups",
           legend.labs = c("< 55 Year","55 - 65 Years"))

haz.plots[[3]] = ggsurvplot(km.effusion,
           fun = "cumhaz",
           palette = c("darkcyan","darkgoldenrod3"),
           subtitle="Hazard, Stratified by Pericardial Effusion Presence",
           font.subtitle = c(10,"italic"),
           font.x = c(9, "bold.italic"),
           font.y = c(9, "bold.italic"),
           ylab="Surival Proportion",
           xlab="Time to Death (Months)",
           legend.title = "Groups",
           legend.labs = c("Present","Absent"))

haz.plots[[4]] = ggsurvplot(km.wms,
           fun = "cumhaz",
           palette = c("darkcyan","darkgoldenrod3"),
           subtitle="Hazard, Stratified by Wall Motion Index",
           font.subtitle = c(10,"italic"),
           font.x = c(9, "bold.italic"),
           font.y = c(9, "bold.italic"),
           ylab="Surival Proportion",
           xlab="Time to Death (Months)",
           legend.title = "Groups",
           legend.labs = c("< 14","\u2265 14"))

haz.plots[[5]] = ggsurvplot(km.fshort,
           fun = "cumhaz",
           palette = c("darkcyan","darkgoldenrod3"),
           subtitle="Survival, Stratified by Fractal Shortening",
           font.subtitle = c(10,"italic"),
           font.x = c(9, "bold.italic"),
           font.y = c(9, "bold.italic"),
```

```
          ylab="Surival Proportion",
          xlab="Time to Death (Months)",
          legend.title = "Groups",
          legend.labs = c("< 0.2","\u2265 0.2"))

arrange_ggsurvplots(haz.plots, print=TRUE, ncol=2, nrow=3)
```

```
df.sum = data.frame(read_excel("df.sum.xlsx"))

kable(df.sum, "latex",
      booktabs = TRUE,
      longtable = TRUE,
      linesep = "\\addlinespace",
      caption = "Summary of Dataset Covariates") %>%
  kable_styling(latex_options = c("hold_position","repeat_header"),
                full_width = TRUE)


kable(df, "latex",
      booktabs = TRUE,
      longtable = TRUE,
      caption = "Original Dataset") %>%
  kable_styling(latex_options = c("hold_position","repeat_header"),
                full_width = TRUE)
kable(df.new, "latex",
      booktabs = TRUE,
      longtable = TRUE,
      caption = "Imputed Dataset") %>%
  kable_styling(latex_options = c("hold_position","repeat_header"),
                full_width = TRUE)


km.sum = data.frame(km.all$n.risk, km.all$n.event, km.all$n.censor, km.all$surv, km.all$std.err, km.all

kable(km.sum,
      caption="Kaplan-Meier Estimate Summary",
      col.names = c("Ni","Di","Ci","Survival","Std. Err", "95% LCL", "95% UCL")) %>%
  kable_styling(latex_options = "hold_position")
```

Table 6: Kaplan-Meier Estimate Summary

Ni	Di	Ci	Survival	Std. Err	95% LCL	95% UCL
130	0	1	1.0000000	0.0000000	1.0000000	1.0000000
129	0	4	1.0000000	0.0000000	1.0000000	1.0000000
125	0	6	1.0000000	0.0000000	1.0000000	1.0000000
119	0	6	1.0000000	0.0000000	1.0000000	1.0000000
113	0	6	1.0000000	0.0000000	1.0000000	1.0000000
107	0	1	1.0000000	0.0000000	1.0000000	1.0000000
106	0	2	1.0000000	0.0000000	1.0000000	1.0000000
104	0	2	1.0000000	0.0000000	1.0000000	1.0000000
102	0	1	1.0000000	0.0000000	1.0000000	1.0000000
101	0	2	1.0000000	0.0000000	1.0000000	1.0000000
99	0	1	1.0000000	0.0000000	1.0000000	1.0000000
98	0	1	1.0000000	0.0000000	1.0000000	1.0000000
97	1	0	0.9896907	0.0103628	0.9697921	1.0000000
96	2	1	0.9690722	0.0181389	0.9352253	1.0000000
93	1	0	0.9586520	0.0211163	0.9197860	0.9991604
92	4	0	0.9169715	0.0306589	0.8634932	0.9737618
88	2	0	0.8961312	0.0347021	0.8372075	0.9592021
86	1	1	0.8857111	0.0366202	0.8243677	0.9516193
84	3	0	0.8540786	0.0422132	0.7862595	0.9277475
81	2	0	0.8329902	0.0457657	0.7615248	0.9111624
79	3	1	0.8013577	0.0509330	0.7252240	0.8854839
75	0	1	0.8013577	0.0509330	0.7252240	0.8854839
74	1	1	0.7905285	0.0527189	0.7129238	0.8765809
72	1	1	0.7795490	0.0545427	0.7005136	0.8675015
70	3	1	0.7461397	0.0601212	0.6632005	0.8394512
66	1	0	0.7348346	0.0620295	0.6507137	0.8298302
65	3	0	0.7009191	0.0677649	0.6137427	0.8004782
62	4	0	0.6556985	0.0755277	0.5654770	0.7603149
58	5	0	0.5991728	0.0856211	0.5066071	0.7086518
53	4	0	0.5539522	0.0941871	0.4605747	0.6662612
49	1	1	0.5426471	0.0964177	0.4492070	0.6555237
47	4	0	0.4964643	0.1061866	0.4031827	0.6113280
43	3	0	0.4618273	0.1141043	0.3692784	0.5775709
40	3	0	0.4271902	0.1226655	0.3358987	0.5432933
37	5	0	0.3694618	0.1388157	0.2814553	0.4849865
32	3	0	0.3348248	0.1500085	0.2495342	0.4492676
29	2	0	0.3117334	0.1582935	0.2285829	0.4251312
27	4	0	0.2655507	0.1774769	0.1875335	0.3760244
23	2	0	0.2424593	0.1887825	0.1674738	0.3510192
21	2	0	0.2193680	0.2016218	0.1477585	0.3256822
19	1	1	0.2078223	0.2087471	0.1380404	0.3128801
17	3	0	0.1711478	0.2370240	0.1075514	0.2723494
14	1	0	0.1589229	0.2483443	0.0976777	0.2585697
13	1	0	0.1466981	0.2609313	0.0879669	0.2446412
12	1	0	0.1344732	0.2750653	0.0784332	0.2305536
11	1	0	0.1222484	0.2911216	0.0690939	0.2162949
10	1	0	0.1100236	0.3096174	0.0599707	0.2018517
9	2	0	0.0855739	0.3572240	0.0424885	0.1723498
7	1	0	0.0733490	0.3891253	0.0342114	0.1572601
6	2	0	0.0488994	0.4845119	0.0189185	0.1263923
4	2	0	0.0244497	0.6962412	0.0062464	0.0957006
2	1	0	0.0122248	0.9923466	0.0017481	0.0854929
1	1	0	0.0000000	Inf	NA	NA