Survival Analysis of Post-Myocardial Infarction Patients

Alvein, Orr, Pham

5/22/2020

Abstract

Background

The rates of myocardial infarction is becoming an increasing common occurence 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).

Objectives

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 cardic 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 surviability.

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 nonparametric (Kaplan-Meier) and parametric methods (Weilbull, Log-Normal, Log-Logistic, Cox PH) to determine estimates of survival among gender and physiological cardic 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 (tenative) 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-mycardial episode survival.

Introduction

Heart disease has become the leader cause of death among the US population among a majority of all racial and ethnic groups (Heron 2019). Myocardial infarctions are becoming largely common among U.S. populations. The number of myocardial infarctions are remarkedly increased from [start date] to [end date] by [x] amount (add citation). In 2015, approximately 23% of all fatalities in the United States was related to some degree of heart disease (cdc. cite please). Unsurprisingly, clinical studies have shown harmful symptoms in post-infarction survival patients. Our obtained dataset to examine the tangible difference in survability rates from the course of year following an infarction episode.

By applying survival analysis techniques to this data set, we seek to achieve improved understanding of the characteristics exhibited by patients in a one year post-infarction interval.We also propose a model to better predict the probability of a survival of patients based on these variable characteristics.

Dataset

We have obtained our data set from Kaggle. The data set contains 133 total patient observations and records 8 variables. Two 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 starat), some patients were followed for less than a year. This provides a clear censoring and truncation provide. We will address this specifically in the following section.

It should be noted the the dataset have 40 missing values after the removal of the two patients. To develop estimates for the missing dated, we employed a k-nearest Neighbor algorithmic approach to predict the values for the missing values. With this in mind, our predictive and summary models will have less than ideal accuracy.

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

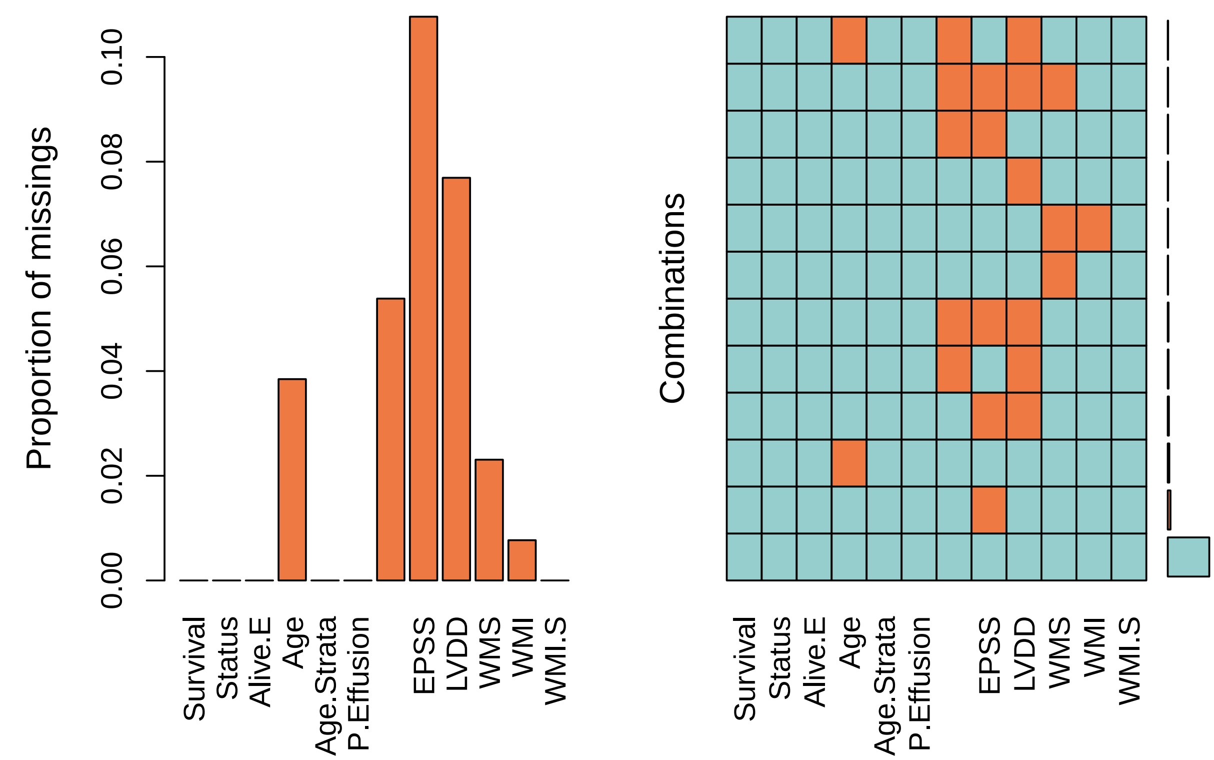
Methodology

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:

Table 1: Missing Values in Original Dataset

	Variable	Count
Survival	Survival	0
Status	Status	0
Alive.E	Alive.E	0
Age	Age	5
Age.Strata	Age.Strata	0
P.Effusion	P.Effusion	0
F.Shortening	F.Shortening	7
EPSS	EPSS	14
LVDD	LVDD	10
WMS	WMS	3
WMI	WMI	1
WMI.S	WMI.S	0



The algorithmic process used here uses a modified k-nearest neighbot (KNN) approach. Using a training data set, the routines of the missForest algorithm predicts the missing values trained on the observed parts of the dataset (Stekhoven 2012). Refer to Stekhoven, et. al 2012 for more detail.

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 imputated dataset can be expressed by:

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

Table 2: Normal Root Mean Square Error

NRMSE
0.2414063

Where X is a matrix of our dataset. Our calculated NRMSE is as follows:

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

Being a random forest iterative process, each imputed dataset will be different from each other. Out of 500 iterations, we have an averaged NRMSE value of [] - that is our inputted values have an estimate []% deviation from true accuracy.

The full imputed dataset may be found in the appendix of this paper.

Censoring

Fixed study start time and fixed study end time.

Patients can start late.

Left censored: Some values, we are unsure how long ago was their myocardial infarction episode. We only have recorded data from the moment patients entered the study and when the study ends.

Right censored: Additionally, some patients live far beyond the scope of the study.

Summary statistics

Over the course of the study, there is a ## Kaplan-Meier We conduct nonparametric Kaplan-Meier fit on our data over multiple strata. We first conduct a fit over all patients in regards to censoring. Then we work on gender groups, age, and general ventriculular condition. A survival function can be defined as:

[insert latex equation for survival function, with censored values]

Weibull Fits (+Cox PH)

Log-Normal

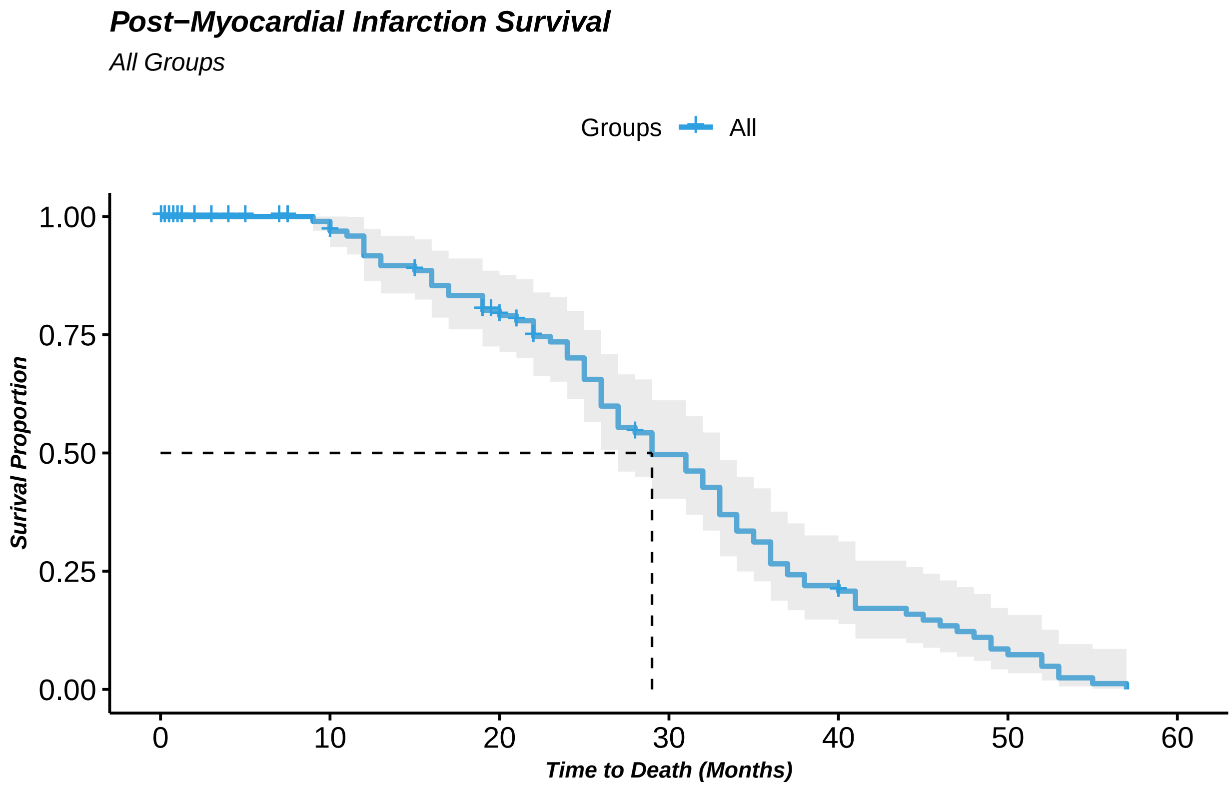
Results

Table of Summary Statistics

KM Curve & Hazard

Survival Curves

KM Overall



Here is the survival curve for all groups within our dataset. We see that a majority of our censored values have very short survival times. This is very intuitive given to our limited interval study time of a single year. We clearly see a median survival time of approximately 29 weeks.

records	n.max	n.start	events	*rmean	*se(rmean)	median	0.95LCL	0.95UCL
130	130	130	88	30.53008	1.249886	29	27	33

Hazard Plots

Weibull Curve

Cox Proportional Hazard

Model Diagnostics

AIC, BIC, and Confidence Intervals

Residual Analysis/QQ Plot

Discussion

References

Introduction

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.

Eric B. Rimm, Meir J. Stampfer, Edward Giovannucci, Alberto Ascherio, Donna Spiegelman, Graham A. Colditz, Walter C. Willett, Body Size and Fat Distribution as Predictors of Coronary Heart Disease among Middle-aged and Older US Men, American Journal of Epidemiology, Volume 141, Issue 12, 15 June 1995, Pages 1117–1127, https://doi.org/10.1093/oxfordjournals.aje.a117385

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.

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.

Methodology

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

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).

Imputation

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

Daniel J. Stekhoven, Peter Bühlmann, MissForest—non-parametric missing value imputation for mixed-type data, Bioinformatics, Volume 28, Issue 1, 1 January 2012, Pages 112–118,

Discussion

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

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

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

Appendix

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.Strata	0 denotes 49 or younger. 1 denotes 50 or older. 2 denotes 65 or older.
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	WMI.S	Binary Variable. 0 denotes that WMI is less than or equal to 1.28. 1 denotes that WMI is greater than or equal to 1.28.

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

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	65.00	2	0	0.150	NA	5.050	10.00	1.000	0
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

Table 4: Original Dataset *(continued)*

Survival	Status	Alive.E	Age	Age.Strata	P.Effusion	F.Shortening	EPSS	LVDD	WMS	WMI	WMI.S
0.03	0	1	NA	2	0	0.260	19.400	4.770	21.00	2.100	1
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

Table 4: Original Dataset *(continued)*

Survival	Status	Alive.E	Age	Age.Strata	P.Effusion	F.Shortening	EPSS	LVDD	WMS	WMI	WMI.S
36.00	1	0	48.00	0	0	0.150	12.000	3.660	10.00	1.000	0
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	Age.Strata	P.Effusion	F.Shortening	EPSS	LVDD	WMS	WMI	WMI.S
27.42	1.00	0.04	71.00	2.00	0.00	0.30	9.00	4.60	14.00	1.00	0.00
19.00	1.00	0.00	68.25	2.00	0.00	0.39	6.85	4.10	14.00	1.70	1.00
32.20	1.00	0.00	55.00	1.00	0.00	0.26	4.00	3.42	14.00	1.00	0.00
57.00	1.00	0.00	60.00	1.00	0.00	0.25	12.06	4.60	16.00	1.45	1.00
2.56	0.02	1.00	57.00	1.00	0.00	0.16	21.58	5.75	18.00	2.25	1.00
26.00	1.00	0.00	54.36	0.94	0.00	0.26	5.00	4.31	12.52	1.00	0.00
28.83	1.00	0.00	57.83	1.01	0.00	0.23	31.00	5.43	22.50	1.88	1.00
50.00	1.00	0.00	60.00	1.00	0.00	0.33	8.00	5.25	14.00	1.06	0.00
19.00	1.00	0.00	46.00	0.00	0.00	0.26	0.00	5.09	16.00	1.14	0.00
25.00	0.99	0.00	54.00	1.03	0.00	0.14	13.00	4.49	15.50	1.19	0.00
10.00	0.77	0.29	77.00	1.90	0.00	0.13	16.00	4.23	18.00	1.80	1.00
52.00	1.00	0.00	62.00	1.00	1.00	0.45	9.00	3.60	16.00	1.14	0.00
52.00	1.00	0.01	64.68	2.00	0.00	0.30	6.00	4.00	14.00	1.00	0.02
44.00	1.00	0.00	60.00	1.04	0.00	0.15	10.00	3.73	14.00	1.00	0.00
0.50	0.00	1.00	60.39	1.00	0.00	0.12	23.00	5.80	11.67	2.33	1.00
24.00	1.00	0.00	55.00	1.00	1.00	0.25	12.06	4.29	14.00	1.00	0.00
0.50	0.00	1.00	69.00	1.91	1.00	0.17	11.00	5.16	18.00	1.64	1.00
5.34	0.00	1.00	62.53	1.00	1.00	0.07	16.91	5.20	24.00	2.00	1.00
3.11	0.00	1.00	71.13	2.00	0.21	0.09	17.00	5.82	8.00	1.73	1.00
1.00	0.00	1.00	66.00	2.00	0.27	0.22	15.00	5.40	13.95	1.54	1.00
0.75	0.00	1.00	72.24	2.00	0.00	0.16	12.00	5.39	12.50	1.62	0.94
0.75	0.00	1.00	85.00	2.00	1.00	0.17	19.00	5.46	13.83	1.38	1.00
0.50	0.00	0.96	73.00	2.00	0.00	0.23	24.30	6.06	7.50	1.66	1.00
5.00	0.00	1.00	59.66	1.14	0.00	0.17	0.00	4.65	8.00	1.00	0.00
27.99	0.99	0.00	64.00	1.00	0.00	0.30	5.90	3.48	10.00	1.11	0.07
29.00	1.00	0.00	61.20	1.00	0.00	0.30	6.11	3.85	10.00	1.67	1.00
29.00	1.00	0.00	35.00	0.00	0.04	0.30	5.00	3.95	14.00	1.00	0.00
31.49	1.00	0.00	55.00	1.00	0.00	0.27	7.00	4.07	12.93	1.00	0.03
0.25	0.00	1.00	75.00	1.88	0.00	0.19	13.11	4.90	12.56	1.00	0.00
36.00	1.00	0.00	55.00	1.03	0.16	0.21	13.02	4.16	14.00	1.56	1.00
1.00	0.00	1.00	65.00	2.00	0.00	0.15	12.54	5.05	12.16	1.00	0.00
1.00	0.00	1.00	52.00	1.00	1.00	0.17	17.20	5.32	14.00	1.17	0.32

Table 5: Imputed Dataset *(continued)*

Survival	Status	Alive.E	Age	Age.Strata	P.Effusion	F.Shortening	EPSS	LVDD	WMS	WMI	WMI.S
2.48	0.00	1.00	72.75	2.00	0.00	0.17	14.18	5.02	6.00	1.57	1.00
27.00	1.00	0.00	47.00	0.00	0.00	0.40	7.12	3.10	12.00	1.00	0.00
30.81	1.00	0.00	63.00	1.00	0.00	0.21	10.00	4.43	14.00	1.17	0.00
26.00	1.00	0.00	61.00	1.00	0.30	0.61	13.10	4.07	13.00	1.53	1.00
16.00	1.00	0.00	63.00	1.00	1.00	0.28	7.62	5.31	5.00	1.00	0.00
1.00	0.00	1.00	65.00	2.00	0.00	0.06	23.60	5.74	21.50	2.15	0.98
19.00	1.00	0.00	68.00	2.00	0.00	0.51	7.01	4.20	15.00	1.67	1.00
28.29	1.00	0.09	80.00	2.00	0.00	0.41	5.40	4.26	13.10	1.17	0.10
32.00	0.99	0.00	60.39	1.00	0.00	0.35	9.30	3.63	12.83	1.22	0.05
16.00	1.00	0.00	70.00	2.00	1.00	0.27	4.70	4.49	22.00	2.00	1.00
29.95	1.00	0.00	79.00	2.00	0.00	0.15	11.26	4.27	13.00	1.30	1.00
46.00	1.00	0.00	56.00	1.00	0.00	0.30	7.67	3.59	14.00	1.00	0.00
2.00	0.00	1.00	67.00	2.00	1.00	0.44	9.00	3.96	17.50	1.45	1.00
37.00	1.00	0.00	64.00	1.00	0.00	0.09	16.16	4.55	12.00	2.00	1.00
19.50	0.00	1.00	81.00	2.00	0.00	0.16	12.15	4.94	9.00	1.42	0.81
20.00	0.22	1.00	59.00	1.18	0.00	0.03	21.30	6.29	20.49	1.31	1.00
0.25	0.02	1.00	63.00	1.00	0.35	0.10	19.38	5.41	23.00	2.30	1.00
2.00	0.00	1.00	56.00	1.00	1.00	0.04	14.00	5.00	18.29	1.73	1.00
7.00	0.00	0.98	61.00	1.19	1.00	0.18	11.92	5.09	9.00	1.42	0.84
31.34	1.00	0.00	57.00	1.00	0.00	0.24	14.80	4.25	18.00	1.38	1.00
12.00	1.00	0.00	58.00	1.00	0.00	0.30	9.40	3.49	12.57	1.00	0.00
8.74	0.15	1.00	60.00	1.10	0.00	0.01	24.60	5.86	39.00	2.00	1.00
10.00	1.00	0.00	66.00	2.00	0.00	0.29	15.60	6.15	14.00	1.00	0.19
45.00	1.00	0.00	63.00	1.00	0.02	0.23	13.00	4.56	13.00	1.08	0.00
33.28	1.00	0.00	59.92	1.00	0.00	0.13	18.60	4.37	12.33	1.37	0.92
20.44	1.00	0.00	70.00	2.00	0.48	0.10	9.80	5.30	23.00	1.91	1.00
32.58	1.00	0.00	68.00	2.00	0.00	0.29	8.16	4.41	14.00	1.17	0.00
32.43	1.00	0.00	66.14	2.00	0.00	0.17	11.90	4.66	10.50	1.05	0.00
25.41	1.00	0.00	73.00	2.00	0.00	0.16	20.78	6.78	16.67	1.39	0.93
26.00	0.93	0.00	72.00	2.00	0.11	0.19	12.17	5.05	13.00	1.18	0.09
0.50	0.00	1.00	59.00	1.13	0.00	0.13	16.40	4.96	14.57	1.37	1.00
12.00	1.00	0.00	67.00	2.00	1.00	0.11	11.38	4.68	11.00	1.00	0.14
49.00	1.00	0.00	51.00	1.00	0.00	0.16	13.20	5.26	11.00	1.13	0.00
0.75	0.00	1.00	50.00	1.00	0.00	0.14	11.40	5.19	10.00	2.50	1.00
23.02	0.88	0.00	70.00	1.91	1.00	0.25	12.41	5.57	5.50	1.10	0.00
47.00	1.00	0.00	65.00	2.00	0.14	0.36	11.69	5.78	12.00	1.00	0.00
41.00	1.00	0.00	78.00	2.00	0.00	0.06	16.10	4.87	13.67	1.37	1.00
0.25	0.00	1.00	86.00	2.00	0.51	0.13	12.20	5.20	24.00	2.18	1.00
33.00	1.00	0.00	56.00	1.00	0.00	0.23	11.00	4.05	12.41	1.00	0.00
29.00	1.00	0.00	58.90	1.03	0.00	0.12	15.39	4.31	15.00	1.67	1.00
41.00	0.97	0.01	58.43	1.08	0.00	0.29	7.50	4.75	12.13	1.03	0.00
25.91	1.00	0.00	50.00	1.00	0.00	0.06	30.10	5.62	21.50	2.39	1.00
15.00	1.00	0.00	54.00	1.00	0.00	0.22	17.90	4.54	16.50	1.18	0.00
0.25	0.05	1.00	68.00	2.00	0.00	0.22	21.70	4.85	15.00	1.15	0.00
3.98	0.00	1.00	72.92	2.00	0.00	0.14	19.40	4.77	11.94	1.58	1.00
12.00	1.00	0.02	64.00	1.00	0.17	0.20	10.46	4.56	14.00	1.00	0.00

Table 5: Imputed Dataset *(continued)*

Survival	Status	Alive.E	Age	Age.Strata	P.Effusion	F.Shortening	EPSS	LVDD	WMS	WMI	WMI.S
32.00	1.00	0.00	63.00	1.08	0.24	0.20	5.00	5.20	8.00	1.00	0.00
32.00	1.00	0.00	65.00	2.00	0.00	0.06	23.60	6.74	12.00	1.09	0.27
27.00	1.00	0.00	54.00	1.00	1.00	0.21	16.80	4.16	16.19	1.50	1.00
30.18	1.00	0.00	62.00	1.00	0.00	0.25	6.00	4.38	11.00	1.00	0.00
9.46	0.00	1.00	78.00	2.00	0.00	0.05	10.00	4.44	15.00	1.36	0.95
0.75	0.00	1.00	61.00	1.00	0.00	0.10	21.83	5.45	28.00	2.33	1.00
34.00	0.98	0.00	52.00	1.00	0.00	0.14	19.65	6.21	11.50	1.15	0.00
1.00	0.15	1.00	70.68	2.00	0.00	0.05	14.80	4.14	15.19	1.60	1.00
21.00	0.00	0.67	70.00	2.00	1.00	0.16	10.60	5.25	11.00	1.00	0.00
27.55	1.00	0.00	55.00	1.00	0.00	0.28	5.50	4.48	16.69	1.83	1.00
15.00	0.00	0.54	60.00	1.00	0.00	0.24	8.70	4.56	13.50	1.12	0.04
0.50	0.00	1.00	67.00	2.00	0.37	0.16	11.30	5.23	13.88	1.48	0.91
36.46	1.00	0.00	64.00	1.00	0.02	0.30	7.79	4.36	14.00	1.06	0.00
53.00	1.00	0.00	59.00	1.00	0.00	0.34	8.55	4.04	13.27	1.00	0.00
35.25	1.00	0.00	46.00	0.00	0.00	0.27	16.50	5.36	12.67	1.06	0.00
33.00	1.00	0.00	59.29	1.00	0.00	0.25	5.60	3.87	18.00	1.50	1.00
40.00	0.32	1.00	74.00	2.00	0.00	0.18	13.29	4.91	12.50	1.14	0.00
33.00	1.00	0.00	59.00	1.00	0.16	0.29	9.10	3.42	18.00	1.50	1.00
2.32	0.00	1.00	69.81	2.00	1.00	0.16	8.50	5.47	16.00	1.45	1.00
4.00	0.03	1.00	58.00	1.00	0.00	0.17	28.90	6.73	19.25	2.01	1.00
31.00	0.98	0.00	58.65	1.00	0.00	0.17	10.24	4.69	11.61	1.00	0.00
33.00	1.00	0.00	66.00	2.00	0.00	0.20	10.03	4.23	12.00	1.00	0.00
22.00	1.00	0.00	62.27	2.00	0.00	0.38	0.00	4.55	10.00	1.00	0.00
25.00	1.00	0.00	62.00	1.00	0.00	0.26	11.80	4.87	11.00	1.00	0.00
1.25	0.00	1.00	63.00	1.22	0.00	0.30	6.90	4.69	15.25	1.51	1.00
31.28	1.00	0.00	59.00	1.01	0.00	0.18	14.30	4.31	13.50	1.50	1.00
25.00	0.99	0.00	57.00	1.01	0.00	0.23	9.70	4.29	11.00	1.00	0.00
24.00	1.00	0.00	57.00	1.00	0.00	0.30	7.00	4.08	13.50	1.05	0.00
0.75	0.00	1.00	71.61	2.00	0.00	0.23	40.00	6.23	14.00	1.40	1.00
3.00	0.00	0.89	62.00	1.00	0.25	0.26	7.60	4.99	11.79	1.00	0.00
32.04	1.00	0.00	58.50	1.00	0.00	0.22	9.94	3.92	11.00	1.00	0.00
19.85	1.00	0.09	66.00	2.00	0.00	0.24	13.80	4.38	22.00	2.20	1.00
36.00	1.00	0.00	61.00	1.04	0.00	0.27	8.11	4.06	12.00	1.00	0.03
25.00	0.94	0.00	59.00	1.00	1.00	0.40	11.08	5.36	12.00	1.00	0.00
27.00	1.00	0.00	57.00	0.97	0.00	0.31	9.40	4.06	12.39	1.02	0.00
34.00	0.91	0.00	59.68	1.00	0.19	0.19	28.90	6.63	19.50	1.95	1.00
37.00	1.00	0.00	66.07	2.00	0.00	0.26	8.02	4.38	12.21	1.00	0.00
34.00	0.99	0.00	54.00	1.00	0.00	0.43	9.30	4.79	10.00	1.04	0.00
28.00	0.00	1.00	62.00	1.00	1.00	0.12	28.60	5.86	21.50	1.95	1.00
29.13	0.97	0.00	67.94	2.00	0.16	0.23	15.14	5.49	12.21	1.20	0.00
17.00	0.97	0.00	64.00	1.00	0.00	0.15	6.60	4.17	14.00	1.27	0.14
38.00	1.00	0.00	57.00	1.00	1.00	0.12	14.65	2.32	16.50	1.38	1.00
31.00	1.00	0.00	60.80	1.00	0.00	0.18	0.00	4.48	11.00	1.38	1.00
12.00	0.87	0.00	61.29	1.06	1.00	0.19	13.20	5.04	19.00	1.73	1.00
36.00	1.00	0.01	60.78	1.09	0.00	0.15	12.00	3.66	10.00	1.00	0.07
17.00	1.00	0.00	61.35	1.03	0.15	0.09	6.80	4.96	13.00	1.11	0.00

Table 5: Imputed Dataset *(continued)*

Survival	Status	Alive.E	Age	Age.Strata	P.Effusion	F.Shortening	EPSS	LVDD	WMS	WMI	WMI.S
31.18	1.00	0.00	61.00	1.01	0.00	0.14	25.50	5.16	14.00	1.27	0.00
7.50	0.00	1.00	64.00	1.00	0.00	0.24	12.90	4.72	12.00	1.00	0.00
41.00	1.00	0.00	57.10	1.00	0.12	0.28	5.40	5.47	11.00	1.05	0.00
36.00	1.00	0.00	68.14	2.00	0.00	0.28	7.00	4.49	14.50	1.21	0.00
22.00	0.98	0.03	57.00	1.00	0.00	0.14	13.17	4.36	15.00	1.36	1.00
20.00	1.00	0.00	62.00	1.00	0.00	0.27	0.00	4.51	15.50	1.56	1.00

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)

df = data.frame(read_excel("df.xlsx"))
df[df=="?"] = " "
s.df = Surv(df$Survival,df$Status)

missing.data = aggr(df, col=c("paleturquoise3","sienna2"))
  title(main="Graphical Represenation of Missing Values in Original Dataset", line = 5)

kable(missing.data$missings, caption="Missing Values in Original Dataset") %>%
  kable_styling(position = "center")

df.m = prodNA(df, noNA = 0.2) #seed 10% of the missing values
df.i = missForest(df.m)
kable(t(df.i$OOBerror), caption="Normal Root Mean Square Error") %>%
  kable_styling(position="center")


km.all = survfit(s.df~1,type="kaplan-meier", data = df)
ggsurvplot(km.all,
           palette = "#2E9FDF",
           conf.int = TRUE,
           title="Post-Myocardial Infarction Survival",
           subtitle="All Groups",
           font.title=c(12,"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")

kable(t(summary(km.all)$table)) %>%
  kable_styling(position = "center")
```

```
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_positon","repeat_header"),
                full_width = TRUE)


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

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.new = round_df(df.i$ximp,2)

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