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. As medical knowledge and techniques improve to meet need of infarction episode, so does the need to understand the survivbility of patients who have survived such episodes.

Objectives

Our goal is to provide detailed survival statistics of post-myocardial infarction patients as well as provide an accurate regression model to best prediction of survival outcomes of a single year following an infarction episode.

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 (Weibull/Cox PH) to determine estimates of survival among gender and myocardial strata (contraction depth, muscular activity,). We consider a slew of statistical and graphical results before determining the most appropriate method of modeling.

Results

Out of all of our methods, we have determined that [] is the most appropriate model for prediction of patient survival. We have AIC values of. We have BIC values. Thus, this model is the best.

Conclusion

[summary statistics] [review of our model + specific surval rates]

Introduction

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.

Table 1: Dataset Variable Labels

Variable	Label	Definition
Survival	Survival	The number of months the patints survived, post-myocardial i
Status	Status	Censorship status. 0 denotes that a patient is a censored while
Alive at the end of Survival Period	Alive.E	Binary variable. 0 denotes that patient is alive at the end of th
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 o
Pericardial Effusion	P.Effusion	Binary variable. Pericardial effusion is excess fluid surrounding
Fractional Shortening	F.Shortening	Fractional shortening is a measure of contractility around the
E-Point Septal Separation	EPSS	E-point septal separation is an addition measure of heart cont
Left Ventricular End-Diastolic Dimension	LVDD	Left ventricular end-diastolic dimension is the measure of the
Wall Motion Score	WMS	Wall motion score is a measure of how the segments of the left
Wall Motion Index	WMI	Wall motion index is the wall motion score divided by the nun

Dataset

We have obtained our data set from Kaggle. The data set contains 131 of observations and records 8 variables. It should be noted that two patients data were removed as there was no recorded months survival date.

Below is a table summary of each table header variable label and a definition of that variable.

Since the time of infartion varies, some patients were followed for less than a year. This provides a clear censoring and truncation provide. We will address this concern in detail in the later in this section. It should be noted that we have a slew of missing values. Since a single patient has shown as missing, we have opted to impute the values for this data row. With this in mind, our predictive and summary models will have less than ideal accuracy.

The full dataset is available in the appendix of this paper.

Methodology

Censoring

Right censored: this study has a fixed start and end date. Some patients start the study midway through the study and thus, we cannot measure the total little length.

Imputation

There are several missing values in the data set. The variables we chose to more ## 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.

Weibull Fits

Log-Normal

Cox PH

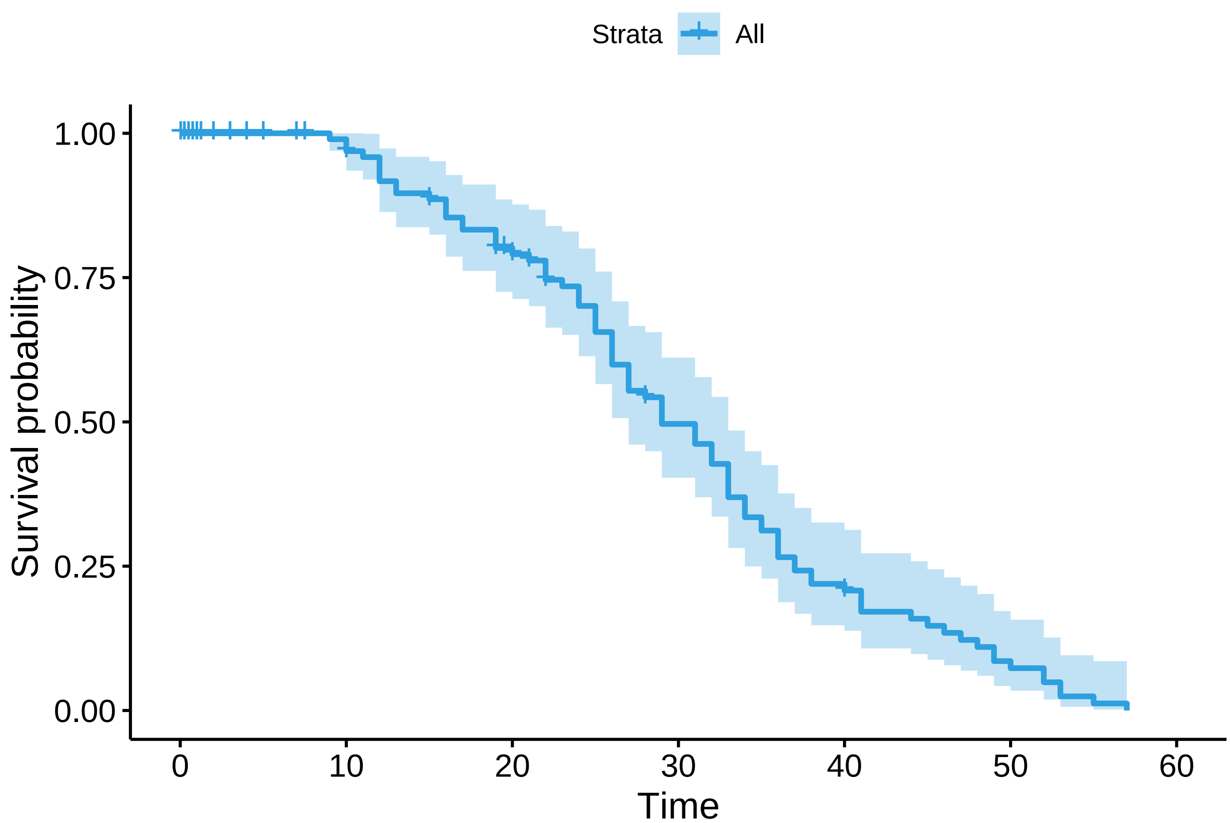
Results

Table of Summary Statistics

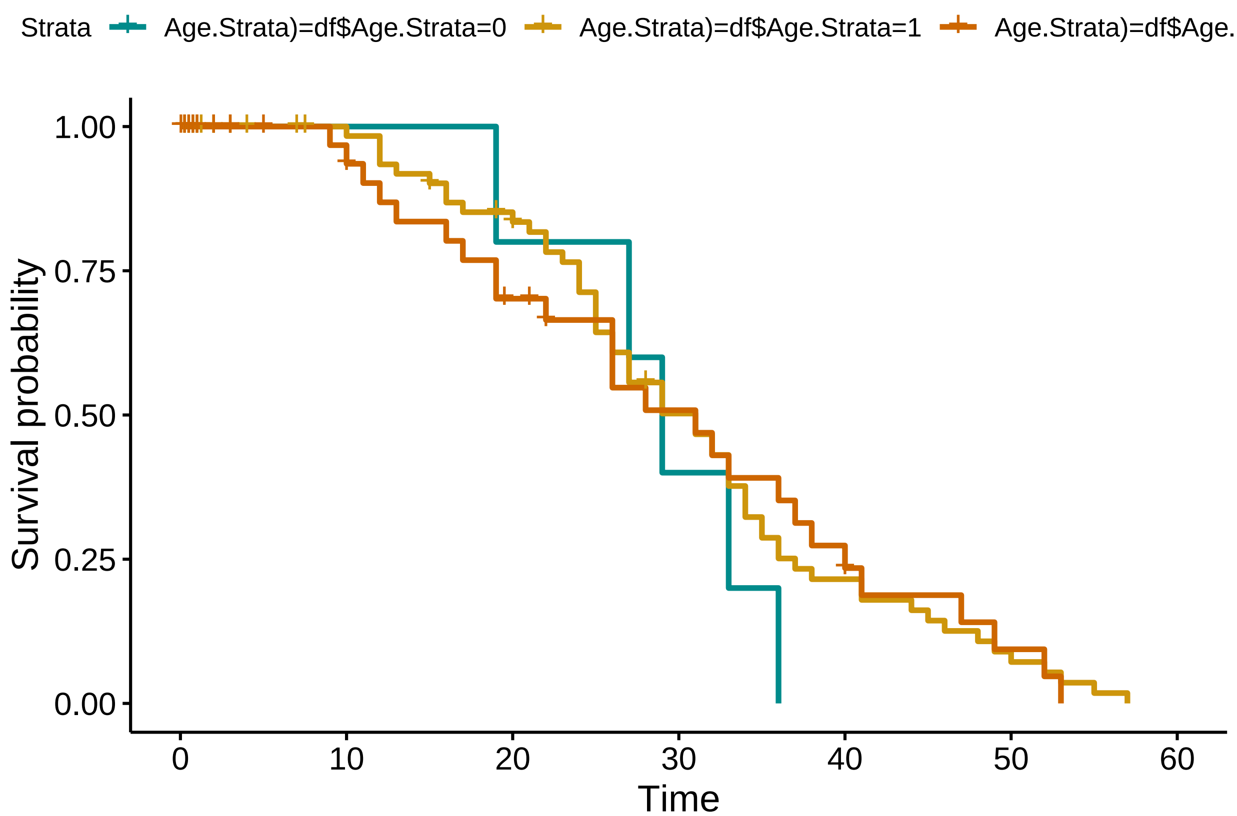
```
##      speed           dist
```

```
##  Min.   : 4.0   Min.   :  2.00
##  1st Qu.:12.0   1st Qu.: 26.00
##  Median :15.0   Median : 36.00
##  Mean   :15.4   Mean   : 42.98
##  3rd Qu.:19.0   3rd Qu.: 56.00
##  Max.   :25.0   Max.   :120.00
```

KM Curve



KM Overall



KM Age

KM Ventricular Activity

```
##      speed           dist
##  Min.   : 4.0   Min.   :  2.00
##  1st Qu.:12.0   1st Qu.: 26.00
##  Median :15.0   Median : 36.00
##  Mean   :15.4   Mean   : 42.98
##  3rd Qu.:19.0   3rd Qu.: 56.00
##  Max.   :25.0   Max.   :120.00
```

Weibull Curve

```
##      speed           dist
##  Min.   : 4.0   Min.   :  2.00
##  1st Qu.:12.0   1st Qu.: 26.00
##  Median :15.0   Median : 36.00
##  Mean   :15.4   Mean   : 42.98
##  3rd Qu.:19.0   3rd Qu.: 56.00
##  Max.   :25.0   Max.   :120.00
```

Cox Proportional Hazard

```
##      speed           dist
##  Min.   : 4.0   Min.   :  2.00
##  1st Qu.:12.0   1st Qu.: 26.00
##  Median :15.0   Median : 36.00
##  Mean   :15.4   Mean   : 42.98
```

```
## 3rd Qu.:19.0   3rd Qu.: 56.00
## Max.   :25.0   Max.   :120.00
```

Model Diagnostics

AIC, BIC, and Confidence Intervals

```
##      speed           dist
## Min.   : 4.0   Min.   :  2.00
## 1st Qu.:12.0   1st Qu.: 26.00
## Median :15.0   Median : 36.00
## Mean   :15.4   Mean   : 42.98
## 3rd Qu.:19.0   3rd Qu.: 56.00
## Max.   :25.0   Max.   :120.00
```

Residual Analysis/QQ Plot

```
##      speed           dist
## Min.   : 4.0   Min.   :  2.00
## 1st Qu.:12.0   1st Qu.: 26.00
## Median :15.0   Median : 36.00
## Mean   :15.4   Mean   : 42.98
## 3rd Qu.:19.0   3rd Qu.: 56.00
## Max.   :25.0   Max.   :120.00
```

```
##      speed           dist
## Min.   : 4.0   Min.   :  2.00
## 1st Qu.:12.0   1st Qu.: 26.00
## Median :15.0   Median : 36.00
## Mean   :15.4   Mean   : 42.98
## 3rd Qu.:19.0   3rd Qu.: 56.00
## Max.   :25.0   Max.   :120.00
```

Discussion

References

Heron, M. Deaths: Leading causes for 2017 pdf icon[PDF – 3 M]. National Vital Statistics Reports;68(6). Accessed November 19, 2019.

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).

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.

Appendix

Dataset

R Code

Table 2: Dataset

Survival	Status	Alive.E	Age	Age.Strata	P.Effusion	F.Shortening	EPSS	LVDD
11.00	1	0	71	2	0	0.26	9	4.599999
19.00	1	0	72	2	0	0.38	6	4.099999
16.00	1	0	55	1	0	0.26	4	3.42
57.00	1	0	60	1	0	0.253	12.061999999999999	4.602999
19.00	0	1	57	1	0	0.16	22	5.75
26.00	1	0	68	2	0	0.26	5	4.309999
13.00	1	0	62	1	0	0.23	31	5.43
50.00	1	0	60	1	0	0.33	8	5.25
19.00	1	0	46	0	0	0.34	0	5.09
25.00	1	0	54	1	0	0.14000000000000001	13	4.49
10.00	0	1	77	2	0	0.13	16	4.230000
52.00	1	0	62	1	1	0.45	9	3.6
52.00	1	0	73	2	0	0.33	6	4
44.00	1	0	60	1	0	0.15	10	3.73
0.50	0	1	62	1	0	0.12	23	5.8
24.00	1	0	55	1	1	0.25	12.063000000000001	4.29
0.50	0	1	69	2	1	0.26	11	4.650000
0.50	0	1	62.53	1	1	7.0000000000000007E-2	20	5.2
22.00	0	1	66	2	0	0.09	17	5.819
1.00	0	1	66	2	1	0.22	15	5.4
0.75	0	1	69	2	0	0.15	12	5.39
0.75	0	1	85	2	1	0.18	19	5.46
0.50	0	1	73	2	0	0.23	12.733000000000001	6.06
5.00	0	1	71	2	0	0.17	0	4.650000
48.00	1	0	64	1	0	0.19	5.9	3.48
29.00	1	0	54	1	0	0.3	7	3.85
29.00	1	0	35	0	0	0.3	5	4.17
29.00	1	0	55	1	0	?	7	?
0.25	0	1	75	2	0	?	?	?
36.00	1	0	55	1	1	0.21	4.2	4.16
1.00	0	1	65	2	0	0.15	?	5.05
1.00	0	1	52	1	1	0.17	17.2	5.32
3.00	0	1	?	2	0	?	12	?
27.00	1	0	47	0	0	0.4	5.12	3.1
35.00	1	0	63	1	0	?	10	?
26.00	1	0	61	1	0	0.61	13.1	4.07
16.00	1	0	63	1	1	?	?	5.31
1.00	0	1	65	2	0	0.06	23.6	?
19.00	1	0	68	2	0	0.51	?	3.88
31.00	1	0	80	2	0	0.41	5.4	4.360000
32.00	1	0	54	1	0	0.35	9.3000000000000007	3.63
16.00	1	0	70	2	1	0.27	4.7	4.49
40.00	1	0	79	2	0	0.15	17.5	4.269999
46.00	1	0	56	1	0	0.33	?	3.59
2.00	0	1	67	2	1	0.44	9	3.96
37.00	1	0	64	1	0	0.09	?	?
19.50	0	1	81	2	0	0.12	?	?
20.00	0	1	59	1	0	0.03	21.3	6.29
0.25	0	1	63	1	1	?	?	?
2.00	0	1	56	1	1	0.04	14	5
7.00	0	1	61	1	1	0.27	?	?
10.00	1	0	57	1	0	0.24	14.8	5.26
12.00	1	0	58	1	0	0.3	9.4	3.49
1.00	0	1	60	1	0	0.01	24.6	5.65
10.00	1	0	66	2	0	0.28999999999999998	15.6	6.15