UCLA

Factors Related to Time to Death of Heart Attack Survivors

Yilan Huang
Leyan Li
Nan Liu
Wenlan Pan

Department of Biostatistics

Reviewer: Gang Li

Los Angeles, December 18, 2020

Contents

1 Contribution

All team members contribute great efforts to the design and development of the entire research project. In the initial stage, Wenlan found the data set provided by Dua and Graff (2017), and we discussed it together to determine the research problem in this project. Each group member performed part of the statistical methods applied in the data set and is responsible for drafting the manuscript of that part in the report. Detailed responsibility of each group member are:

Yilan Huang: - Introduction - Data preparation - Descriptive statistics

Leyan Li: - Abstract - Introduction - Log-rank test - AFT model - Discussion

Nan Liu: - Kaplan-Meier estimate - Cox model (selection & prediction) - Conclusion

Wenlan Pan: - Introduction - Cox & AFT model assessments - Conclusion - Reference

2 Abstract

Heart attack is identified as the leading cause of death in the world and the United States. An echocardiogram is a clinical test that helps to identify human damages from a heart attack. Learning from echocardiogram data on how to reduce the risk of death from heart attack is significant. The purpose of this study is to explore factors associated with the risk of death due to heart attack and to compare the survival function, the 1-year and 2-year survival probability among different levels of risk factors. The effects of age and five echocardiographic assessment variables on the death risk of heart attack survivors were discussed. Based on Cox's model findings, we conclude that the heart attack survivors with normal fractional-shortening value had increased survival probability with a slight significant (hazard ratio = 0.72, p-value = 0.14). The prediction performance of the final model is limited due to the limited sample size. The results from the log-normal and log-logistic accelerated failure time model support that the fractional shortening is a potential risk factor with relatively smaller p-values.

3 Introduction

Heart attack, medically known as myocardial infarction (MI), is defined as myocardial necrosis caused by a short supply of oxygen-rich blood for part of the heart muscles (Goldstein et al., 2000). It is responsible for more than 15% of global mortality each year and led to over 110,000 deaths in 2017 in the United States (Virani et al., 2020). Although the mortality rate has reduced in some high-income areas due to improved medical conditions, it has an even heavier health-care and financial strain caused by the increasing population of aging and long-term survivors (Smolina et al., 2012). Moreover, since survivors of their first heart attack are typically followed by other cardiovascular events, such as recurrent MI, heart failure, and stroke, approximately 14% of them will die within one year after the first MI, according to the American Heart Association (Benjamin et al., 2018). The probability of death after the first MI, however, is extremely variable and depends on several factors, which mainly include demographic characteristics (age, sex, and family history) and cardiac-related clinical indicators (Virani et al., 2020). The prediction and prevention based on these factors are needed to reduce the high mortality rate of heart attack and the burden it causes.

Heart disease refers to various types of conditions that can affect heart function, and MI has been a leading cause of human heart disease and death. The risk of heart disease is associated with several demographic factors, such as age, sex, and ethnicity, as suggested by the Centers for Disease Control and Prevention. Age and sex could interact to affect MI risks, as it is reported that men older than 45 and women more aged than 55 are more likely to have heart attacks. Some other behavioral factors also vary between the young and elderly. A higher prevalence of other lifestyle risk factors, including smoking and family history of heart disease, among the young MI patients was reported by Hoit et al. (1986). In addition, there is a more significant gender bias in young MI patients. The proportion of men (90%) in young MI patients is much higher relative to the older population, which is one of the possible causes of the risk difference in young and old MI

patients (Shah et al., 2016). Family history is another risk factor closely related to MI. Ranthe et al. (2015) concluded that family history would double the MI risk from first degree relatives. Although family history is commonly accessible to physicians, it has not been included in major risk scores as a criterion for determining risk classification. Several studies have been conducted to prove its efficiency in optimize risk algorithm predictions. Echocardiography provides a rapid assessment of regional systolic activity and is essential for the diagnosis of coronary artery disease and acute MI localization (Esmaeilzadeh et al., 2013). An effective diagnosis method could improve both the health status of patients and the efficiency of hospitals. Therefore, the relationship between echocardiogram data and death after heart attack is meaningful to study. Moreover, since the echocardiogram shows distinct and characteristic abnormalities at the onset, it helps identify the physiological risk of a heart attack. Several echocardiogram examinations are available below, covering two major categories - cardiac structure and function.

Two variables relating to the cardiac structure are pericardial effusion (PeEf) and left ventricular end-diastolic dimension (LVDD). The former PeEf reflects the excess fluid around the heart, puts extra pressure, and further influences the heart functions. PeEF with fluid has been confirmed to accompanied by MI (Galve et al., 1986). The latter LVDD measures the size of the heart at end-diastole. In older patients, isolated left ventricular diastolic dysfunction is the most common form of heart failure (Yu et al., 2002). Thus, its value is useful in predicting the risk of death from heart failure.

The determinants of cardiac function are three variables: fractional shortening (FS), e-point septal separation (EPSS) and wall motion score index (WMSI). FS is used as a systolic function index. It is calculated by the percentage change in left ventricular diameter during systole. Researchers found that, based on Cox stepwise regression analysis, lower midwall FS is associated with a higher risk of heart disease or death. Therefore, it is a recommended stratification factor for the risk group of heart failure-related death. EPSS is another evaluation measure of left ventricular systolic function. It tests the interval

between the most apical point of the anterior leaflet of the mitral valve and the septum at diastole (Boon et al., 2020). It could be used to identify a severely reduced left ventricular ejection fraction (LVEF), which is of great significance for clinical management. Finally, assessment of WMSI is also essential in the echocardiographic examination. LVEF could also be quantified by it, and a higher WMSI value is recognized to be associated with the occurrence of acute MI (Møller et al., 2006).

The purpose of our study is to identify risk factors for heart attack-related death and thus find potential variables for prediction in the future. The effect of both demographic variables (age) and echocardiographic assessment variables (PeEf, FS, EPSS, LVDD, WMSI) on the risk of death after the first heart attack was examined. Non-parametric, semi-parametric, and parametric approaches were applied in modeling the survival probability of patients with different values of the covariates. The overall fit of the semi-parametric Cox model and parametric accelerated failure time (AFT) models were evaluated using the Cox-Snell residuals plot. A comparison to previous studies was made, and the research strength, as well as its limitations, were discussed.

4 Methods

4.1 Data preparation and summary statistics

The Echocardiogram data set contains records of 132 patients who suffered heart attacks at some point in the past. Some patients are still alive at the end of the survival period and some are not. Variables in the data set mainly focus on clinical assessment, to describe the patient's heart attack and condition, which include age in years when heart attack occurred, PeEf, FS, EPSS, LVDD and WMSI. Two patients without survival and/or alive information were excluded, so our analyses are limited to the 130 patients with known time-to-event data. Since using only data with complete variables information could greatly reduce the sample size, we performed imputation for continuous variables using the mean.

In our preliminary analysis, we used the original continuous variables to fit the models. However, the results indicated low significance in the Cox survival models. To help interpretation, we dichotomized the continuous variables (FS, EPSS, LVDD and WMSI) based on clinical diagnostic criteria (1 = normal, 0 = abnormal). The dichotomization grouped patients with similar performance in variables of interest together, and clearly presented the clinical interpretation of each value. It also enabled us to calculate the hazard ratios between different groups. Normal FS was defined as a value $\geqslant 25$ % (Stöllberger et al., 2019), while normal EPSS were defined as a value $\leqslant 7$ mm (Erbel et al., 1985). A LVDD $\leqslant 5.6$ cm was considered normal (Gaasch et al., 1983). WMSI $\geqslant 1.7$ is typically associated with heart failure, so we used 1.7 as the cut point between WMSI normal and abnormal groups (Boiten et al., 2016).

The mean and standard deviation were calculated for continuous variables in the data set, while the frequency and percentage of each group were calculated for binary variables. We also computed the correlation matrix and created scatter plots to describe bivariate distribution relation of the variables.

4.2 Non-parametric method

To analyze the survival time to death for heart attack patients, we started with a non-parametric method, the Kaplan-Meier Product-Limit estimate to find an estimate of survival function. The Kaplan-Meier estimate was visualized by the plot of product-limit estimates. Also, we graphically represented the estimated cumulative hazard for heart attack patients along time. For each categorical variable, we compared the survival process on different class by creating the stratified Kaplan-Meier curves. More specifically, we compared the time to death probability based on different levels of PeEf, FS, EPSS, LVDD and WMSI.

To formally examine the difference of survival among different groups, we conducted log-rank test with equal and Peto-Peto weights, respectively, on the survival variable with the five categorical covariates (PeEf, FS, EPSS, LVDD, WMSI). The Peto-Peto weights allow us to focus on detecting the difference on the early stage after heart attack. We reported the one degree Chi-Square test p-values on Table 4.

4.3 Semi-parametric method - Cox model

We attempted to use the Cox proportional hazard model to access the effect of the PeEf, FS, EPSS, LVDD, WMSI and age on survival probability for heart attack patients. We started with fitting all these variables in the Cox regression model. The models for the five categorical variables were assessed independently in separate Cox regressions, and then the selections were combined in a final Cox proportional hazards model. Before further inference, we conducted variable selection by step-wise method using AIC (Akaike information criterion) and checked the proportional assumption again with the final model after selection.

After identifying risk factor by variable selection, the final Cox model was fitted to analyze the effects of risk factors when adjusting for other variables. Further, the 1-year survival probability and 2-year survival probability within each strata of important variables were

estimated from Cox Proportional hazard model. The baseline cumulative hazard was approximated by the Breslow's estimator. Moreover, the survival process for different groups of important variables was visualized in the estimated survival function plot.
Martingale residuals for continuous variables were plotted to determine their functional form. The proportional hazard assumption was both formally validated by artificially generating time-dependent covariates and graphically checked by plotting the log estimated baseline cumulative hazard rates for each stratum against time. We also graphically assessed the Cox model fit by plotting the Cox-Snell residuals for all related models.

4.4 Parametric method - AFT model

We applied accelerated failure time model with four different distribution assumptions to model the survival data. AFT models assume that the survival function of patient at time t with covariate Z is the baseline survival function at time $t\exp(\theta' Z)$. This is expressed as $S(t|Z) = S_0(t\exp(\theta' Z))$. The corresponding linear representation is $Y = \log(t) = \mu + \gamma' Z + \sigma W$. In our model, the covariate Z includes a continuous variable (age) and five categorical variables (PeEf, FS, EPSS, LVDD, WMSI). W is assumed to be Weibull, exponential, log-logistic, log-normal distributed in four models, respectively. The Weibull model assumes a flexible hazard rate along time, while the log-logistic model assumes that the hazard rate is hump shape. Estimation of the model parameters and corresponding p-values are done using maximum likelihood estimation.
Model diagnostics were assessed using an analog of Cox-Snell residual plot. Overall fit is evaluated by comparing the plot $H_r(t)$ v.s. r_i with the theoretical 45 degree line passing origin (or vertical line at 0 for Weibull regression).

5 Results

5.1 Summary statistics

Descriptive statistics of each variable used in the analysis are reported in Table 1 and Table 2. The average survival time for all patients in the data set is 22 months, and the survival time varies greatly from 0.03 month to 57 months. The age in years when heart attack occurred ranged from 35 to 86. Among the 130 patients, 88 were known to have died during the survival period, and the remaining 42 were still alive at the end of survival period. 81.5 % of patients (n = 106) had no fluid pericardial effusion around the heart, while 67.7 % of patients (n = 88) were with abnormal FS and 73.8 % of patients (n = 96) were with abnormal EPSS. The correlation matrix (Table 3) does not show obvious relations between the covariates.

5.2 Non-parametric method

From the Kaplan-Meier estimator plot (Figure 1), we saw the survival probability of heart attack patients remained near 1 in the first 10 months after the patient experienced a heart attack. After 10 months, the survival probability decreased almost constantly, and after about 30 months, the survival probability decreased to 50% (Figure 2). The cumulative hazard estimated by Nelson-Aalen Estimate and Kaplan-Meier Estimate (Figure 3) aligned with each other and they identically implied the low cumulative hazard in the first 10 months after a heart attack.

The stratified Kaplan-Meier curves (Figure 4) did not show significant difference between categories for each of these five categorical variables (ReEf, FS, EPSS, LVDD, WMSI). Whether the patients had fluid, abnormal contracility around the heart, abnormal E-point septal separation value, abnormal size of the heart at end-diastole and abnormal movement of the segments of the left ventricle did not show big differences in the survival probability separately. The curves of Kaplan-Meier Estimates of the two levels for each of these

variables were almost coincide. However, the survival probabilities of normal fractional-shortening group were always higher than those of the abnormal fractional-shortening group at each time point, although the gaps were minor.

Log-rank test with both equal weights and Peto-Peto weights on all the five categorical variables gave p-values greater than 0.05 (Table 4). We concluded that none of the 5 categorical variables have significant effect on the survival probability for heart attack patients, as suggested by our data. Among all the five categorical variables, normal and abnormal FS groups gave the most significant difference with smallest p-values (0.1 and 0.3).

5.3 Semi-parametric method - Cox model

The Cox proportional model fitting all the variables indicated none of the variable was significant (Table 5). Since all the variables had p-values greater than 0.05, none of them had significant affect on the survival probability of heart attack patients. After conducting the step-wise variable selection, only one variable left in the result model which was FS with a p-value of 0.14.(Table 6) The final model was the Cox proportional model with the predictor FS. This aligned with our findings using the non-parametric method. The survival probability for normal FS group always had slightly higher survival probabilities than those of the abnormal FS group at each time point.

Since FS was the only influence factor in the final Cox proportional hazard model, we compared and predicted the survival probability for normal FS group and abnormal FS group. The survival function plot (Figure 5) indicated the difference of survival probability increased along the time. The 1-year survival probability for the normal FS group is 0.933 and that for the abnormal FS group is 0.908, while the 2-year survival probability for the normal FS group is 0.75 and that for the abnormal FS group is 0.67.

According to the Martingale residual plots for continuous variables, all of them followed the linearity assumption, displaying straight lines around 0 (Figure 6). This suggested that a

transformation was not necessary. In addition, the Cox-Snell residual plots for the models with each covariate individually or with all covariates were roughly followed a 45-degree line, indicating acceptable model fitting (Figure 7).

However, non-proportional hazards might exist. Since all p-values for the time-dependent covariates were greater than 0.05, we could infer that they were all proportional hazards (Table 7). Nevertheless, the p-value for the time-dependent covariate of WMSI was 0.1, which indicated a slight significance and could be a potential assumption violation. Therefore, the time-dependent covariate of WMSI was then proposed to be included in the Cox model. This was also confirmed by the graphical diagnosis implemented by stratifying the covariates and plotting their log estimated baseline cumulative hazard rates for each stratum against time (Figure 8). For each stratum, the baseline hazards of LVDD and WMSI were far away from parallel, and stratification on LVDD or/and WMSI would thus be more appropriate. But for the model with all significant variables (p-values all smaller than 0.05) including FS, WMSI and time-dependent variable created for WMSI (WMSI $\times \log(t)$), the plot of Cox-Snell residual did not show a better fit and even more deviated from the 45-degree line (Figure 7).

5.4 Parametric method - AFT model

We summarized the results of exponential, log-logistic and log-normal AFT models in Table 8. We did not include the results from Weibull model because the algorithm did not converge.

As shown in Figure 9, the model assessment utilizing Cox-Snell residual supported the lack of fit for the Weibull regression, and also the exponential regression models. Thus, we only considered results from the log-logistic and log-normal models. P-values for FS was the smallest in both models (0.23, 0.32). This was consistent with the variable selection results from semi-parametric model containing only the FS variable. From AFT models, the survival time after heart attack was longer for normal FS group than abnormal FS

group ($\hat{\theta} = 0.10, 0.11 > 0$). However, the difference was not statistically significant with p-values (0.23 or 0.32) greater than 0.05. The conclusions were considered as a weak evidence supporting the effect of normal and abnormal FS because the Cox-Snell residual plots showed strong deviations at the tails of these two models (Figure 9).

6 Discussion

After variable selection, the Cox proportional hazard model indicates that fractional shorting is a risk factor with a relatively small p-value of 0.14. Although in the log-normal and log-logistic AFT model, it is still not significant (p-values equal to 0.23 and 0.32, respectively), the results from these two models confirmed our inference from the Cox proportional hazard model as abnormal fractional shorting is related to an increased hazard ratio.

Some researchers also analyzed the same echocardiogram data set to predict whether the patient would survive for at least a year after a heart attack or not. For instance, with this data set, Kitayama (2018) performed conventional survival analysis. Little significant relationships were identified between survival time and all the covariates. However, he observed that the wall motion score changed from 24 months (two years), and thus applied a transformation from the survival time and death variable to a 2-year survival variable. Log-rank tests suggested a significant difference between survival time in different wall motion score categories for the first two years after a heart attack. He then concluded that people with a high wall motion score would have a higher risk of death for the first two years after he/she experienced a heart attack. As there is no identification of any significant association, this study is consistent with our conclusions. Our study also agreed that the wall motion score could be a time-dependent covariate, while we used the WMSI instead, which was derived from the wall motion score. However, this model was not be adopted after we found the Cox-Snell residual plot revealed a worse model fit with a very large estimator. Some data issues could cause this during this model fitting. Moreover, since the data set's donor worked in computer science, most of the researchers applied machine learning approaches with it, including one who used a tree classifier to distinguish between patients who died or not (Kaynar, 2019). His cross-validation results showed success in prediction (CV score $\geqslant 0.916$), which suggested that other methods could be more pursued, such as different machine learning algorithms that could partially solve the issue caused by limited data.

Our study has some achievements. Comprehensive examinations are carried out on all the information extracted from the data set. For data preparation, preliminary biological knowledge was taken into account to create survival models with consensus biological cut-off values. Such discretization of continuous variables makes the results more interpretable and also gives us some hazard ratio estimations between groups by each categorical variable. Additionally, from the statistical perspective, non-parametric, semi-parametric, and parametric models were all introduced to compare survival experiences among groups with different covariates. This ensures the validation of our conclusion required different assumptions, which have their benefits and drawbacks. The advantages of these three strategies are fully used in this manner. Besides, the gap among medical test results, diagnosis, and prediction caused by the lack of previous direct research on echocardiogram data survival analysis is filled by our investigation. While a few variables are significant, it promotes future studies on this area with updated data sources. The identified survival risks and prediction of survival time may also be a reference to doctors' diagnoses and a focus of survivors from their first heart attack.

However, there are some limitations to this study, especially some data issues. First of all, the data were donated in 1989, but when and where the data were obtained were unknown. Besides, little information is available online because of the really early donated time. Secondly, since the data were gathered prior to 1989, it may happen that the inference drawn from such data does not comply with the present situation. For example, updates have been introduced to the echocardiogram measurement, and new techniques have been used to diagnose the heart attack (Chaulin and Duplyakov, 2020). Moreover, as the data set contributor mentioned on the website, the small sample size (130) could make it difficult to predict survival, and this was confirmed in our analysis. Finally, the variables are redundant and limited. While the original data source included twelve variables, only six of them could be used in the actual analysis as they were irrelevant and repetitive variables (same information and dropping suggested by the data donor). In addition, there is only

one demographic variable (age at the heart attack) in the data set, while several widely used demographic factors were not included, such as sex and family history, which commonly affect the potential risk of death and, thus, physician's diagnosis. Further analysis, such as utilizing an updated data set of echocardiogram test results containing more relevant variables, notably demographic variables, will be recommended.

7 Conclusion

The survival probability of heart attack patients remained around 100% in the first ten months after they experienced a heart attack. However, the survival probability for the patients would be halved after 30 months of a heart attack. The fractional shortening is a potential factor of survival probability for the heart attack patients. The higher contractility around the heart, the higher the survival probability a patient will have. The hazard ratio between a heart attack patient with normal fractional shortening value and one with abnormal fractional shortening value is 0.72. The 1-year survival probability for the patient has normal fractional shortening value is 0.933 and that for the patient has the abnormal fractional shortening value is 0.908, while the 2-year survival probability for the normal fractional shortening group is 0.75 and that for the abnormal fractional-shortening group is 0.67. Thus, for the first two years after each patient experiences heart attack, people with lower fractional shortening value, especially lower than 25%, would have a higher risk of death. This result suggests more attention should be paid to this group of patients.

8 Tables

Table 1: Descriptive Statistics (Continuous Variables, N = 130)

	Mean	S.D.	Range
Survival	22.18	15.86	0.03 - 57.00
Age	62.83	8.21	35.00 - 86.00
FS	0.22	0.11	0.01 - 0.61
EPSS	12.19	6.99	0.00 - 40.00
LVDD	4.77	0.78	2.32 - 6.78
WMSI	1.38	0.45	1.00 - 3.00

Table 2: Descriptive Statistics (Discrete Variables, N = 130)

		N	%
Outcome	Alive	42	32.3
	Dead	88	67.7
PeEf	Fluid	24	18.5
	No fluid	106	81.5
FS	Abnormal	88	67.7
	Normal	42	32.3
EPSS	Abnormal	96	73.8
	Normal	43	26.2
LVDD	Abnormal	17	13.1
	Normal	113	86.9
WMSI	Abnormal	24	18.5
	Normal	106	81.5

Table 3: Correlation Matrix

	Survival	PeEf	Age	FS	EPSS	WMSI	LVDD
Survival	1.00	-0.15	-0.20	0.30	0.22	0.26	0.07
PeEf	-0.15	1.00	0.01	-0.03	-0.13	-0.13	0.07
Age	-0.20	0.01	1.00	-0.16	-0.06	-0.02	-0.05
FS	0.30	-0.03	-0.16	1.00	0.26	0.23	0.20
EPSS	0.22	-0.13	-0.06	0.26	1.00	0.17	0.22
WMSI	0.26	-0.13	-0.02	0.23	0.17	1.00	0.23
LVDD	0.07	0.07	-0.05	0.20	0.22	0.23	1.00

Table 4: Log-rank Test Results on Categorical Variables

p-values	PeEf	FS	EPSS	LVDD	WMSI
Equal Weights	0.9	0.1	0.8	0.9	0.4
Peto-Peto Weights	0.8	0.3	1	0.6	0.9

Table 5: Table for Cox Proportional Hazard Model Fitted All Variables

Variable	coef (se)	z-statisitc	p-value
Age	-0.013 (0.0142)	-0.923	0.356
PeEf	0.0492(0.313)	0.157	0.875
FS	-0.410 (0.229)	-1.789	0.074
EPSS	0.092 (0.238)	0.387	0.699
LVDD	0.037 (0.396)	0.094	0.926
WMSI	0.387 (0.375)	1.031	0.302

Table 6: Table for Cox Proportional Hazard Model After Step-wise Variable Selection

Variable	coef (se)	z-statisitc	p-value
FS	-0.328 (0.222)	-1.476	0.14

Table 7: Proportional Hazard Assumption Tested by Creating Time-dependent Covariates

Time-dependent covariate	Age	PeEf	FS	EPSS	LVDD	WMSI
z-statistic	-1.17	-0.46	-0.44	0.78	-0.20	1.65
p-value	0.24	0.65	0.66	0.44	0.84	0.10

Table 8: AFT Model Summaries

	Exponential		Log logistic		Log normal	
Variable	Parameter*	p-value	Parameter*	p-value	Parameter*	p-value
Age	0.01(0.01)	0.59	0.00 (0.01)	0.77	-0.00(0.01)	0.95
PeEf	0.03(0.31)	0.92	-0.04 (0.14)	0.77	-0.04(0.13)	0.78
FS	0.04 (0.22)	0.85	0.11 (0.09)	0.23	0.10(0.10)	0.32
EPSS	-0.02 (0.24)	0.92	-0.01 (0.10)	0.91	0.03(0.10)	0.80
LVDD	-0.23(0.39)	0.56	-0.18 (0.16)	0.27	-0.06(0.16)	0.72
WMSI	-0.07(0.37)	0.86	0.12 (0.19)	0.53	0.05(0.15)	0.74

Parameter*: Estimator (Standard Error)

9 Figures

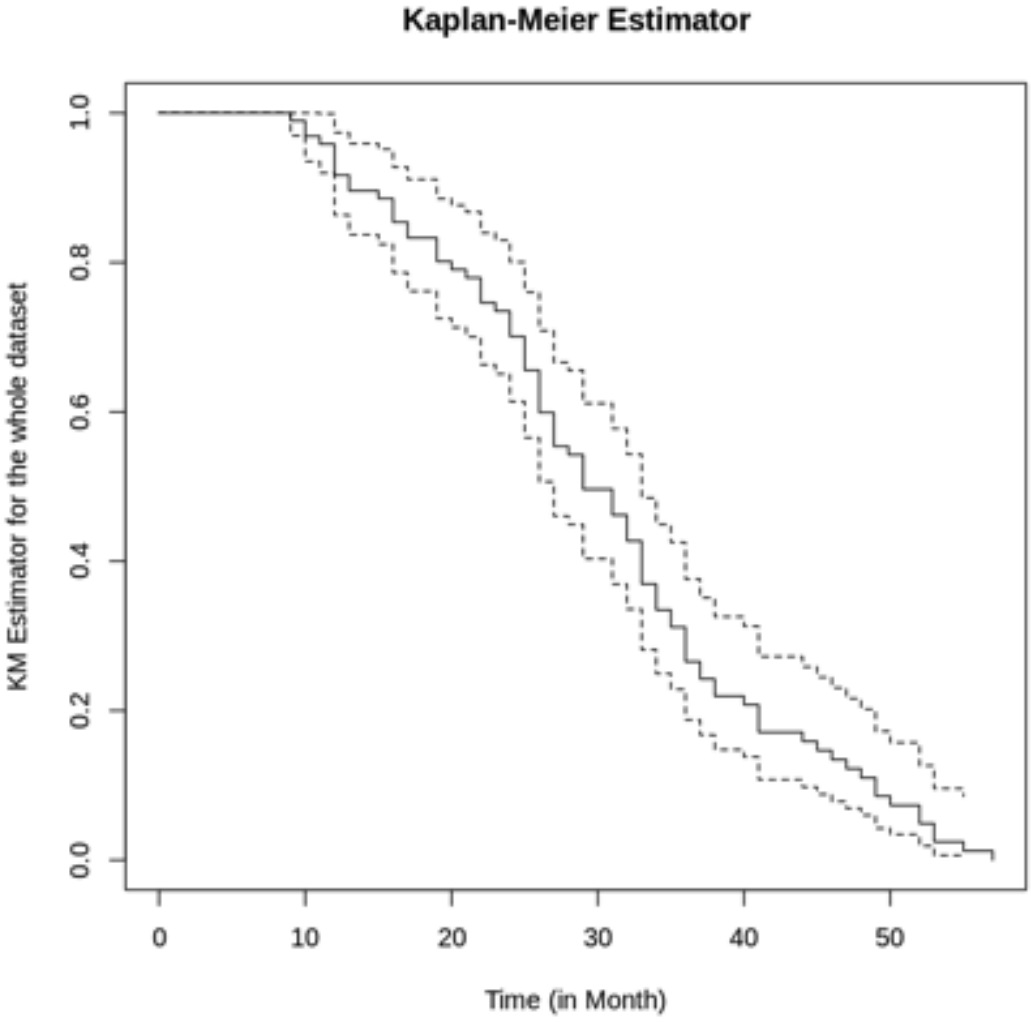

Figure 1: Kaplan-Meier Estimates for Heart Attack Patients

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
9	97	1	0.9897	0.0103	0.96979	1.0000
10	96	2	0.9691	0.0176	0.93523	1.0000
11	93	1	0.9587	0.0202	0.91979	0.9992
12	92	4	0.9170	0.0281	0.86349	0.9738
13	88	2	0.8961	0.0311	0.83721	0.9592
15	86	1	0.8857	0.0324	0.82437	0.9516
16	84	3	0.8541	0.0361	0.78626	0.9277
17	81	2	0.8330	0.0381	0.76152	0.9112
19	79	3	0.8014	0.0408	0.72522	0.8855
20	74	1	0.7905	0.0417	0.71292	0.8766
21	72	1	0.7795	0.0425	0.70051	0.8675
22	70	3	0.7461	0.0449	0.66320	0.8395
23	66	1	0.7348	0.0456	0.65071	0.8298
24	65	3	0.7009	0.0475	0.61374	0.8005
25	62	4	0.6557	0.0495	0.56548	0.7603
26	58	5	0.5992	0.0513	0.50661	0.7087
27	53	4	0.5540	0.0522	0.46057	0.6663
28	49	1	0.5426	0.0523	0.44921	0.6555
29	47	4	0.4965	0.0527	0.40318	0.6113
31	43	3	0.4618	0.0527	0.36928	0.5776
32	40	3	0.4272	0.0524	0.33590	0.5433
33	37	5	0.3695	0.0513	0.28146	0.4850
34	32	3	0.3348	0.0502	0.24953	0.4493
35	29	2	0.3117	0.0493	0.22858	0.4251
36	27	4	0.2656	0.0471	0.18753	0.3760
37	23	2	0.2425	0.0458	0.16747	0.3510
38	21	2	0.2194	0.0442	0.14776	0.3257
40	19	1	0.2078	0.0434	0.13804	0.3129
41	17	3	0.1711	0.0406	0.10755	0.2723
44	14	1	0.1589	0.0395	0.09768	0.2586
45	13	1	0.1467	0.0383	0.08797	0.2446
46	12	1	0.1345	0.0370	0.07843	0.2306
47	11	1	0.1222	0.0356	0.06909	0.2163
48	10	1	0.1100	0.0341	0.05997	0.2019
49	9	2	0.0856	0.0306	0.04249	0.1723
50	7	1	0.0733	0.0285	0.03421	0.1573
52	6	2	0.0489	0.0237	0.01892	0.1264
53	4	2	0.0244	0.0170	0.00625	0.0957
55	2	1	0.0122	0.0121	0.00175	0.0855
57	1	1	0.0000	NaN	NA	NA

Figure 2: Kaplan-Meier Estimates

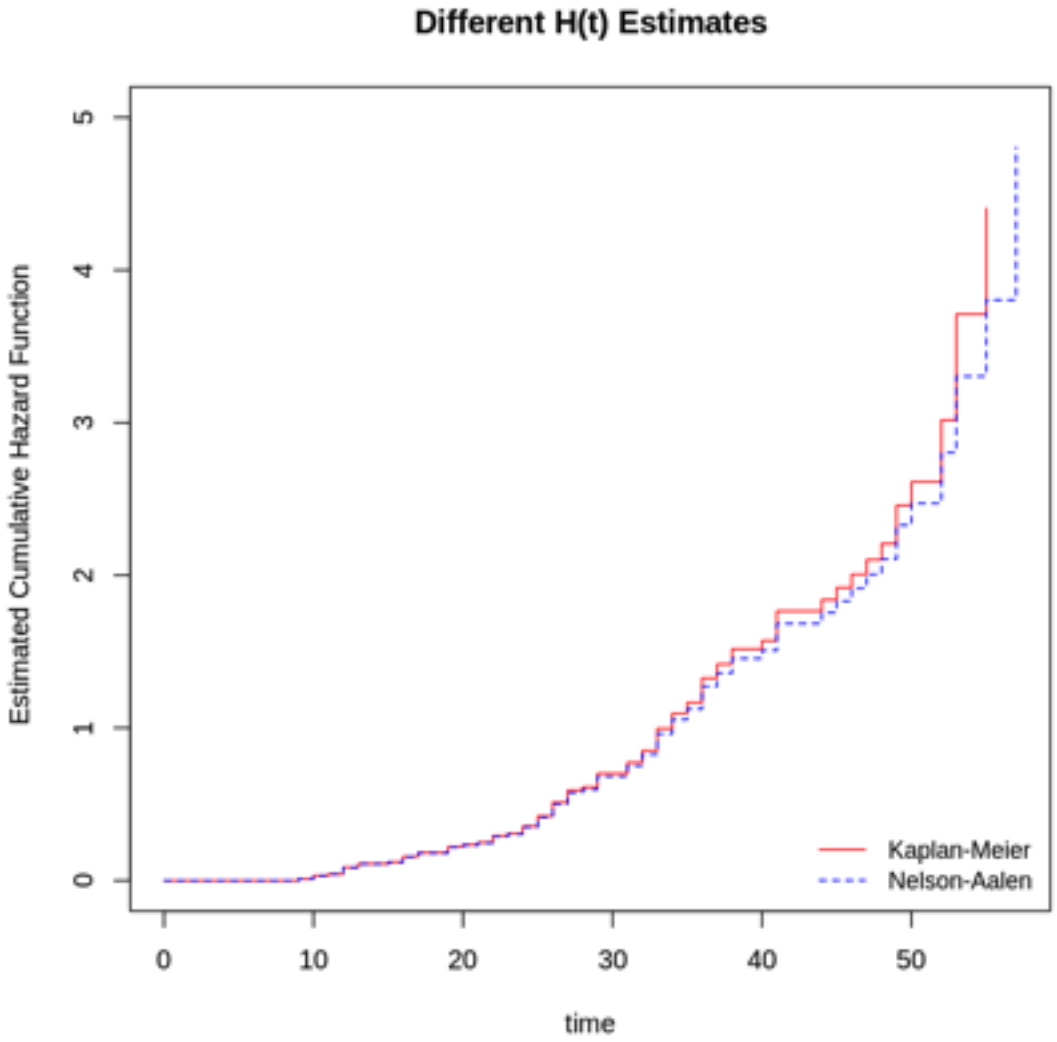

Figure 3: Comparing the Nelson-Aalen Estimate to the Kaplan-Meier Estimate

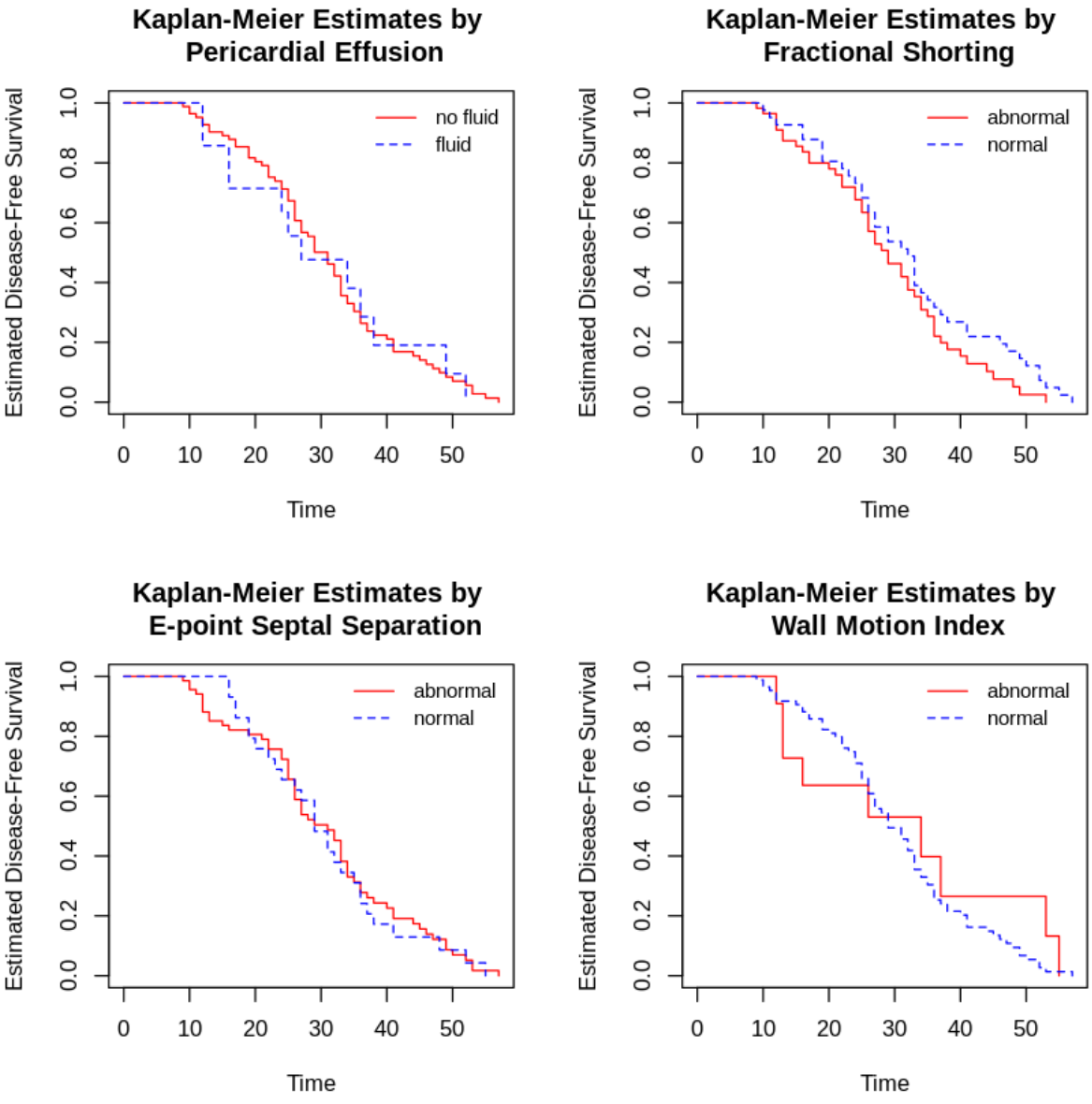

Figure 4: Kaplan-Meier Estimates for Each Covariate

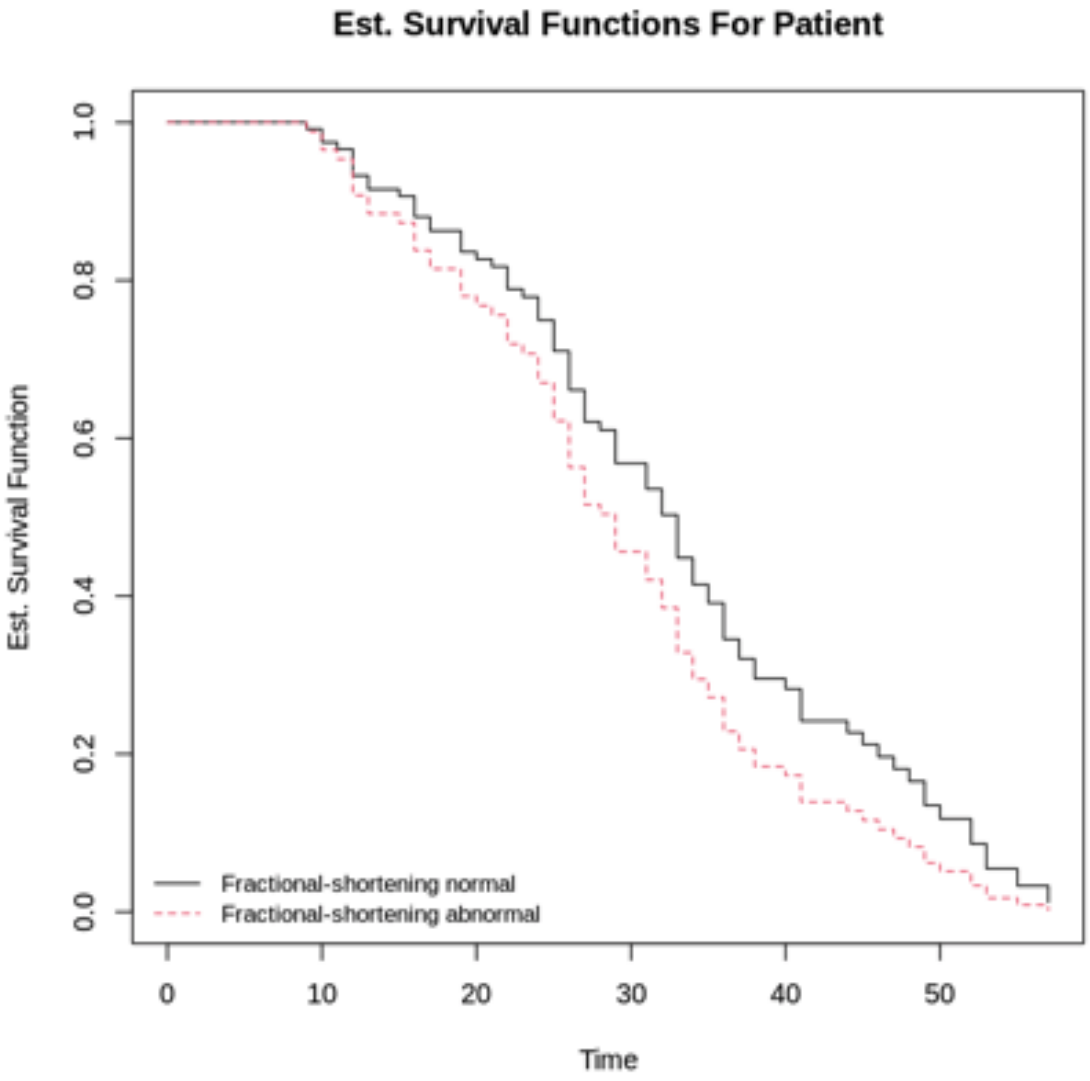

Figure 5: Survival probability by Fractional Shorting

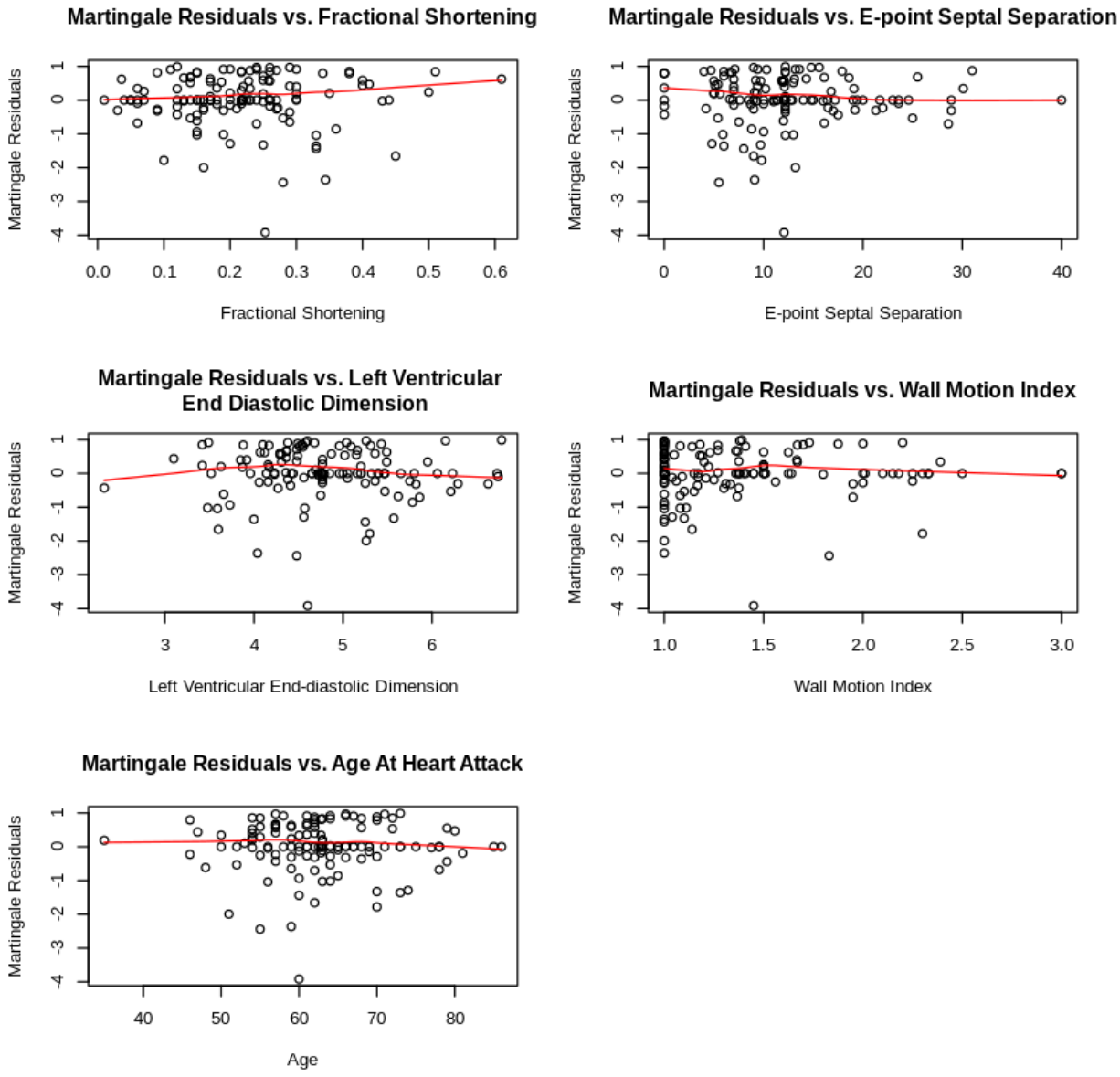

Figure 6: Martingale Residuals for All Continuous Variables

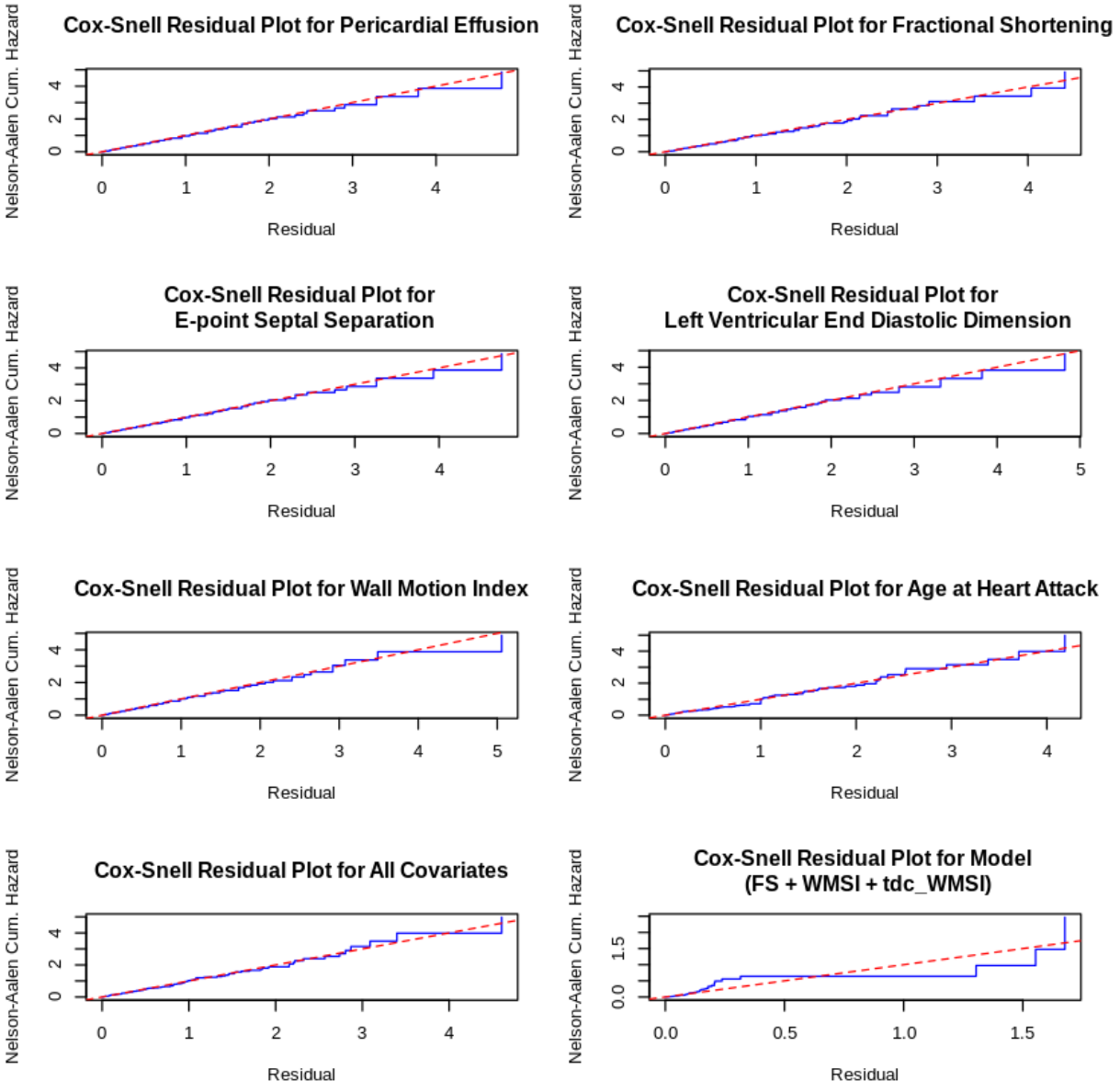

Figure 7: Fit Assessment for Cox Models

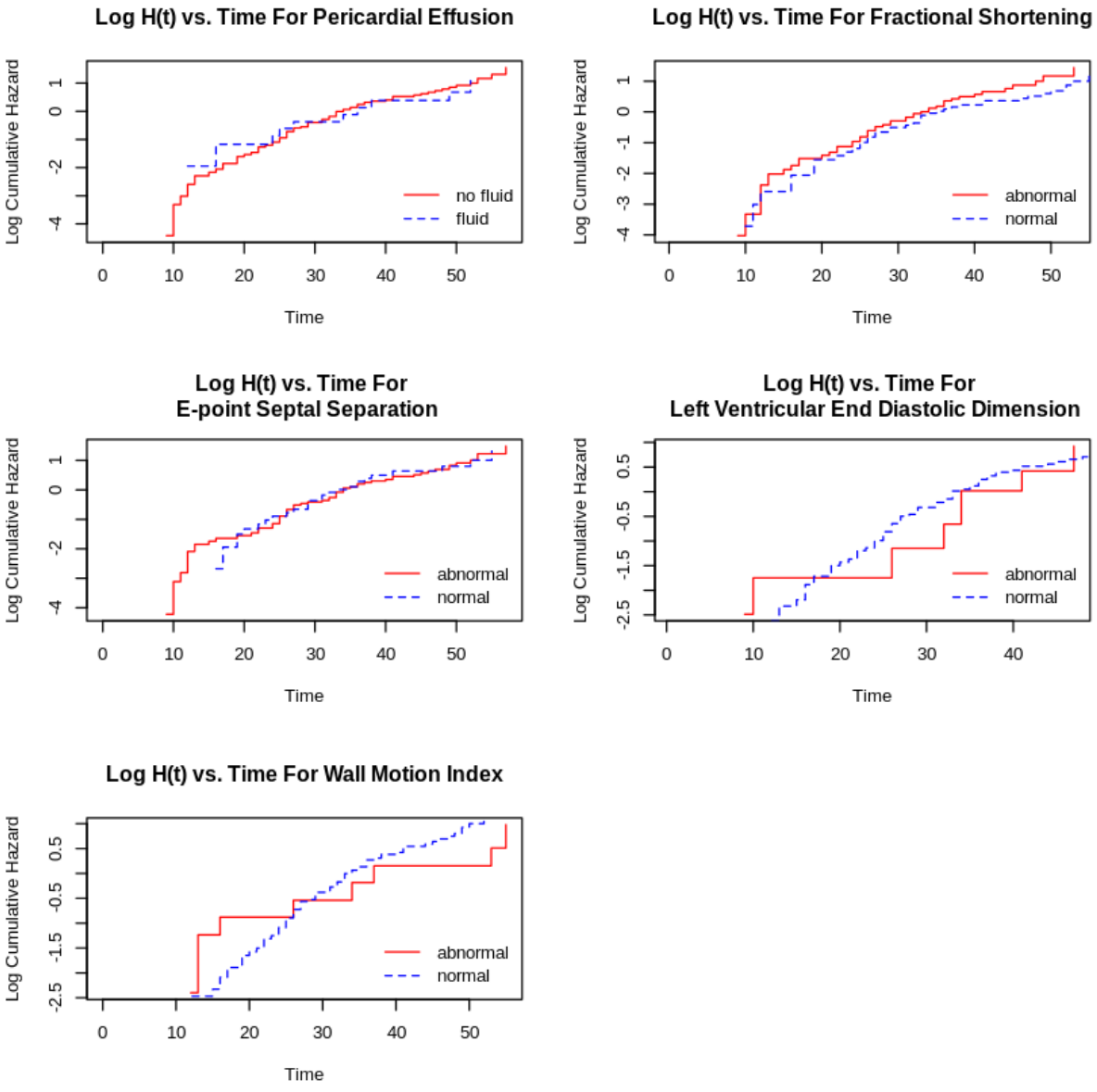

Figure 8: Model Diagnosis for Proportional Hazard Assumption

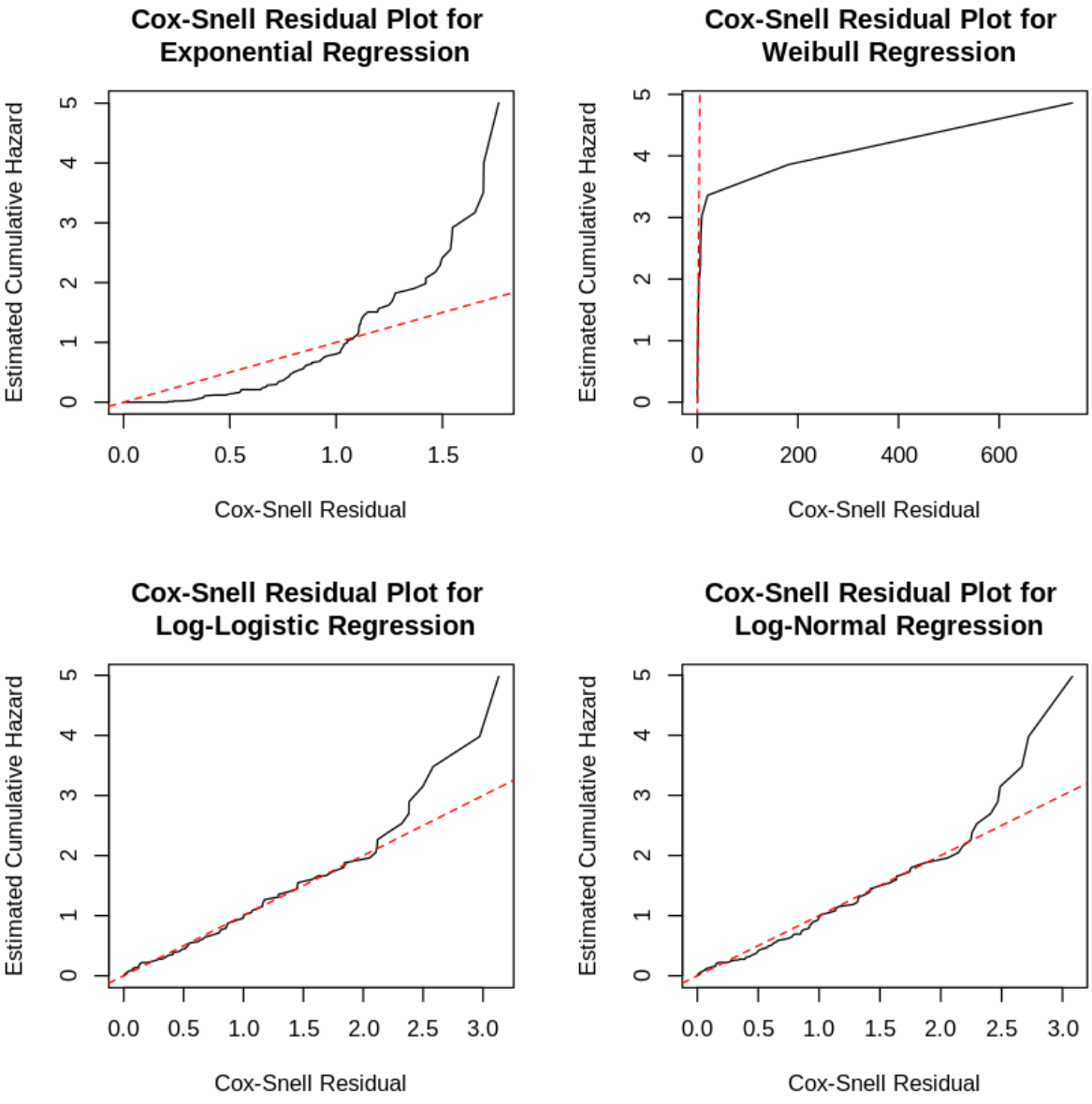

Figure 9: Model Diagnosis for AFT Models

References

Benjamin, E. J., Virani, S. S., Callaway, C. W., Chamberlain, A. M., Chang, A. R., Cheng, S., Chiuve, S. E., Cushman, M., Delling, F. N., Deo, R., et al. (2018). Heart disease and stroke statistics—2018 update: a report from the american heart association. *Circulation*.

Boiten, H. J., Ekmen, H., Zijlstra, F., van Domburg, R. T., and Schinkel, A. F. (2016). Impact of early coronary revascularization on long-term outcomes in patients with myocardial ischemia on dobutamine stress echocardiography. *The American journal of cardiology*, 118(5):635–640.

Boon, S., Matta, J. L., Kraemer, C. E., Tuinman, P., and van Westerloo, D. (2020). Pocus series: E-point septal separation, a quick assessment of reduced left ventricular ejection fraction in a pocus setting. *Neth J Crit Care*, 28:139–141.

Chaulin, A. M. and Duplyakov, D. V. (2020). Biomarkers of acute myocardial infarction: diagnostic and prognostic value. part 1. *Journal of Clinical Practice*, 11(3):75–84.

Dua, D. and Graff, C. (2017). UCI machine learning repository.

Erbel, R., Schweizer, P., Meyer, J., Krebs, W., Yalkinoglu, Ö., and Effert, S. (1985). Sensitivity of cross-sectional echocardiography in detection of impaired global and regional left ventricular function: prospective study. *International journal of cardiology*, 7(4):375–389.

Esmaeilzadeh, M., Parsaee, M., and Maleki, M. (2013). The role of echocardiography in coronary artery disease and acute myocardial infarction. *The journal of Tehran university heart center*, 8(1):1.

Gaasch, W. H., Carroll, J. D., Levine, H. J., and Criscitiello, M. G. (1983). Chronic aortic regurgitation: prognostic value of left ventricular end-systolic dimension and end-

diastolic radius/thickness ratio. *Journal of the American College of Cardiology*, 1(3):775–782.

Galve, E., Garcia-Del-Castillo, H., Evangelista, A., Batlle, J., Permanyer-Miralda, G., and Soler-Soler, J. (1986). Pericardial effusion in the course of myocardial infarction: incidence, natural history, and clinical relevance. *Circulation*, 73(2):294–299.

Goldstein, J. A., Demetriou, D., Grines, C. L., Pica, M., Shoukfeh, M., and O'Neill, W. W. (2000). Multiple complex coronary plaques in patients with acute myocardial infarction. *New England Journal of Medicine*, 343(13):915–922.

Hoit, B. D., Gilpin, E., Henning, H., Maisel, A., Dittrich, H., Carlisle, J., and Ross Jr, J. (1986). Myocardial infarction in young patients: an analysis by age subsets. *Circulation*, 74(4):712–721.

Kaynar, A. (2019). ecohartedaandclass — kaggle. `https://www.kaggle.com/anilkay/ecohartedaandclass/output`. (Accessed on 12/17/2020).

Kitayama, Y. (2018). Survival analysis — kaggle. `https://www.kaggle.com/yukikitayama/survival-analysis#Kaplan-Meier-estimates`. (Accessed on 12/17/2020).

Møller, J. E., Hillis, G. S., Oh, J. K., Reeder, G. S., Gersh, B. J., and Pellikka, P. A. (2006). Wall motion score index and ejection fraction for risk stratification after acute myocardial infarction. *American heart journal*, 151(2):419–425.

Ranthe, M. F., Petersen, J. A., Bundgaard, H., Wohlfahrt, J., Melbye, M., and Boyd, H. A. (2015). A detailed family history of myocardial infarction and risk of myocardial infarction–a nationwide cohort study. *PloS one*, 10(5):e0125896.

Shah, N., Kelly, A.-M., Cox, N., Wong, C., and Soon, K. (2016). Myocardial infarction in the "young": risk factors, presentation, management and prognosis. *Heart, Lung and Circulation*, 25(10):955–960.

Smolina, K., Wright, F. L., Rayner, M., and Goldacre, M. J. (2012). Determinants of the decline in mortality from acute myocardial infarction in england between 2002 and 2010: linked national database study. *Bmj*, 344:d8059.

Stöllberger, C., Wegner, C., and Finsterer, J. (2019). Left ventricular hypertrabeculation/noncompaction, cardiac phenotype, and neuromuscular disorders. *Herz*, 44(7):659–665.

Virani, S. S., Alonso, A., Benjamin, E. J., Bittencourt, M. S., Callaway, C. W., Carson, A. P., Chamberlain, A. M., Chang, A. R., Cheng, S., Delling, F. N., et al. (2020). Heart disease and stroke statistics—2020 update: a report from the american heart association. *Circulation*, pages E139–E596.

Yu, C.-M., Lin, H., Yang, H., Kong, S.-L., Zhang, Q., and Lee, S. W.-L. (2002). Progression of systolic abnormalities in patients with "isolated" diastolic heart failure and diastolic dysfunction. *Circulation*, 105(10):1195–1201.

10 Appendix