Survival Analysis of Post-Myocardial Infarction Patients

Research: Kaelin Alvein
Parametric Methods: Patricia Orr
Non-Parametric Methods: Pete Pham

6/12/2020

Abstract

Background: Myocardial infarction episodes have become an increasing common occurrence in the United States. Rapid development of medical technology and knowledge have led to an decline in myocardial infarction fatalities[6]. 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[8].

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. 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: Non-parametric Kaplan-Meier curve shows a median survival time of ~30 months for all age groups with the exception of pericardial effusion presence. Patients with pericardial effusion have a slightly lower. Multiple parametric models were assessed for potential distributional fit of the survival data and the Weibull distribution was found to model the overall survival behavior well. Reg

Conclusion: Thus, for predictive model we found the Weibull regression fit to be the most ideal candidate for modeling overall 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. Finally, we identified a regression model that predicts survival given a selection of prognostic factors. 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[2,7]. In 2009 cardiovascular disease represented nearly 64% of all cardiac related deaths[4]. 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[2].

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%[15]. Decreases in mortality was common through the world better medical intervention techniques and increase awareness of healthier lifestyle choices became more prevalent[5].

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[7]. 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, fractional 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 two groups with 0 denoting younger than 63 years, 1 denoting older or equal to 63 years. Pericardial effusion is already grouped into binary values with 0 denoting the absence of fluid while 1 denotes the presence of the effusion. Wall motion score is divided into two groups: 0 denoting scores less than 14, 1 denoting scores greater than or equal to 14. Finally, we divided fractional shortening into two strata as well with 0 being lengths being less than 0.2 seconds and 1 being greater than or equal to 0.2 seconds.

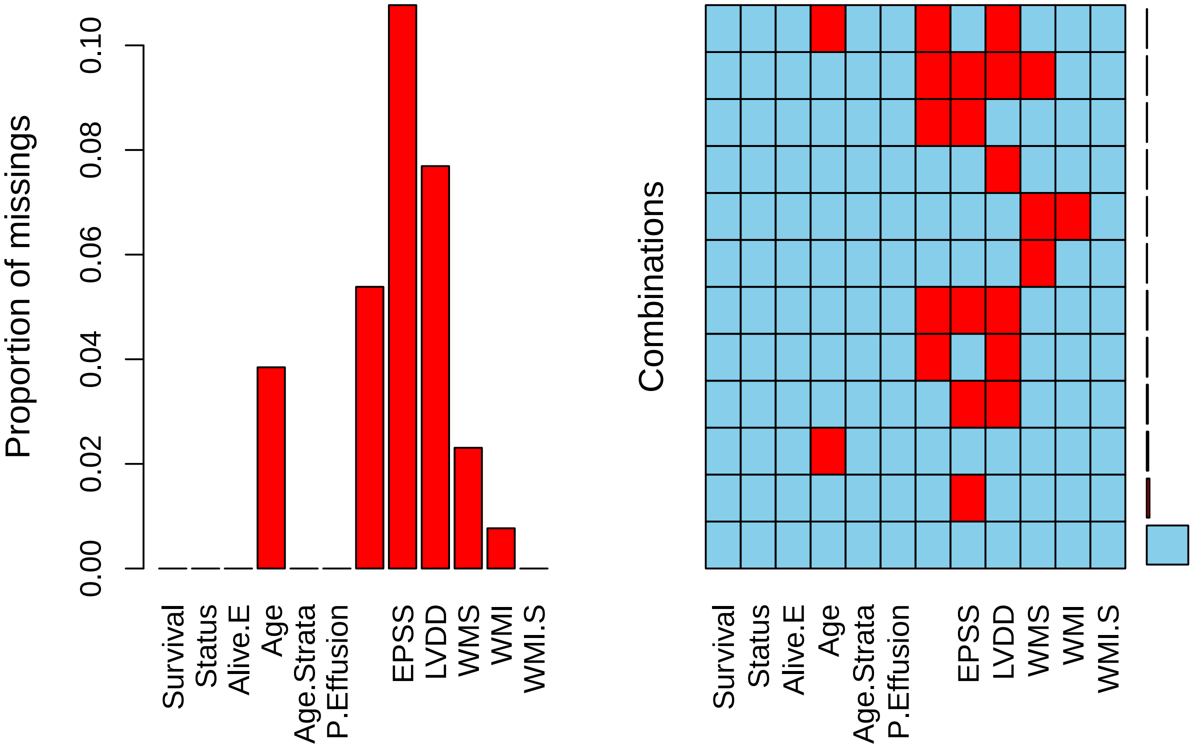
Table 1: Stratification Groupings

Indicator	Age	Effusion	WMS	FS
0	< 63 Years	Fluid is absent	< 11	< 0.2
1	>= 63 Years	Fluid is present	>= 11	>= 0.2

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. The graphic below describes the number of missing values per variable:



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[12]. 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)[8]. The general performance of our imputed dataset can be expressed by:

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

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[14]. 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. Additionally, as we stratify groups by covariates, we use the Mantel-Haenszel/log-rank test. The following equation is used to calculate the test statistic in order to compare two strata[14]:

$$Mantel - HaenszelStatistic = \frac{\sum_{i=1}^{k}(a_i - E_0(A_i))}{\sqrt{\sum_{i=1}^{k} Var_0(A_i)}}$$

Where,

$$E_0(A) = \frac{m_1 n_1}{n} \quad and \quad Var_0(A) = \frac{m_1(n - m_1)}{n-1} * \frac{n_1}{n}(1 - \frac{n_1}{n})$$

We then use the Mantel-Haenzel statistics to perform a standard chi-square test to examine the differences between our strata.

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 Modeling of Survival Data

Another technique for characterizing the survival function is to assume a distributional model for the data. Compared with the Kaplan-Meier approach, this method has certain advantages; it enables construction of a continuous survival curve and allows simplicity of estimation and prediction. If the selected model accurately describes the data, it may also lend insight into the underlying mechanism for the survival behavior. This method is only applicable if a distributional model can be identified that adequately fits the survival data.

For the post-myocardial infarction dataset, we fit three commonly employed distributional models to the survival data and evaluating goodness of fit of the three models. This is accomplished by comparing the modeled survival curves to the Kaplan-Meier curve and by comparing point estimates for each model.

The three models chosen for comparison are the Weibull, log-normal, and log-logistic distributions.

The Weibull hazard function is given below, where λ and α are the scale and shape parameters. Weibull hazard is rising if$\alpha > 1$, constant if $\alpha = 1$, and declining if $\alpha < 1$.

$$h(t) = \lambda^{-1}(-log(1-p))^{1/\alpha}$$

The log-normal distribution can be defined relative to the standard normal distribution; a random variable Y may be said to have the log-normal distribution if for some random variable T that has standard normal distribution:

$$log(Y) = \alpha + \sigma T$$

The hazard function of the log-normal distribution increases with time from 0 until it reaches a maximum and then decreases, approaching 0 as time approaches infinity.

The log-logistic distribution can be defined relative to the standard logistic distribution; a random variable X may be said to have the log-logistic distribution if for some random variable S that has standard logistic distribution:

$$log(X) = \alpha + \sigma S$$

the hazard function of the log-logistic distribution decreases with time from ∞ if $\alpha < 1$, decreases from λ if $\alpha = 1$, and if $\alpha > 1$ resembles the log-normal distribution.

Semi-Parametric Modeling of Survival Data

Where fully parametric models offer flexibility and the efficient, relatively simple estimation of overall survival function parameters, semi-parametric models offer the advantage of being well-suited to the estimation of covariate effects. Semi-parametric models decompose risk into a baseline hazard component and a relative risk component that is dependent on the covariates.

The semi-parametric Cox proportional hazards model is employed here to explore the relationship between predictor variables and survival behavior.

The Cox PH hazard function is defined as follows:

$$h(t) = h_0(t)exp(b_1x_1 + b_2x_2 + ... + b_nx_n)$$

where t represents the survival time, $x_1, x_2, ..., x_n$ are the set of prognostic factors or covariates, and the coefficients $b_1, b_2, ..., b_n$ measure the effect of the covariates on survival time. The baseline hazard $h_0(t)$ corresponds to the value of the hazard if all covariates are equal to zero.

Use of the Cox PH model requires that the baseline hazard is not dependent on the covariates, and that the covariate terms do not depend on time - that is, that the slope coefficients are constant. Hazard functions stratified on covariate group may be used to assess the proportional hazard assumption. If the hazards functions for covariate groups cross over time, the proportionality assumption is not met and alternate analysis methods should be employed. One such method involves sub-setting the survival data and covariates based on hazard cross-over time. In this approach, a separate model is fit to each subset where it has been determined that the proportionality assumption holds. Alternately one may apply alternate modeling techniques that are suitable for time-varying effects, or simply investigate the impact of covariate by inspection of stratified Kaplan-Meier survival curves [14].

Results

Non-Parametric: Kaplan-Meier Survival Estimates

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

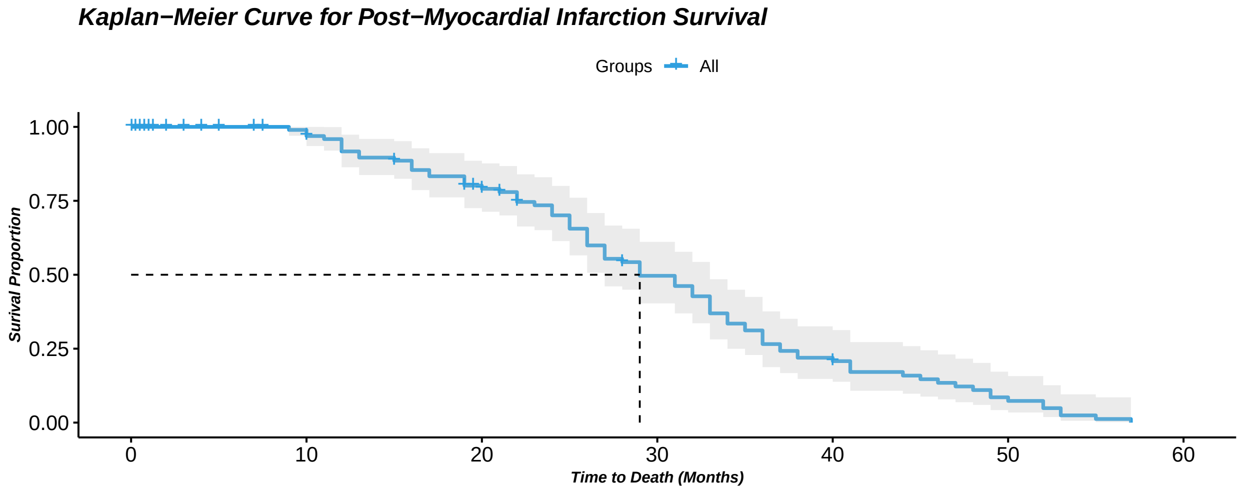
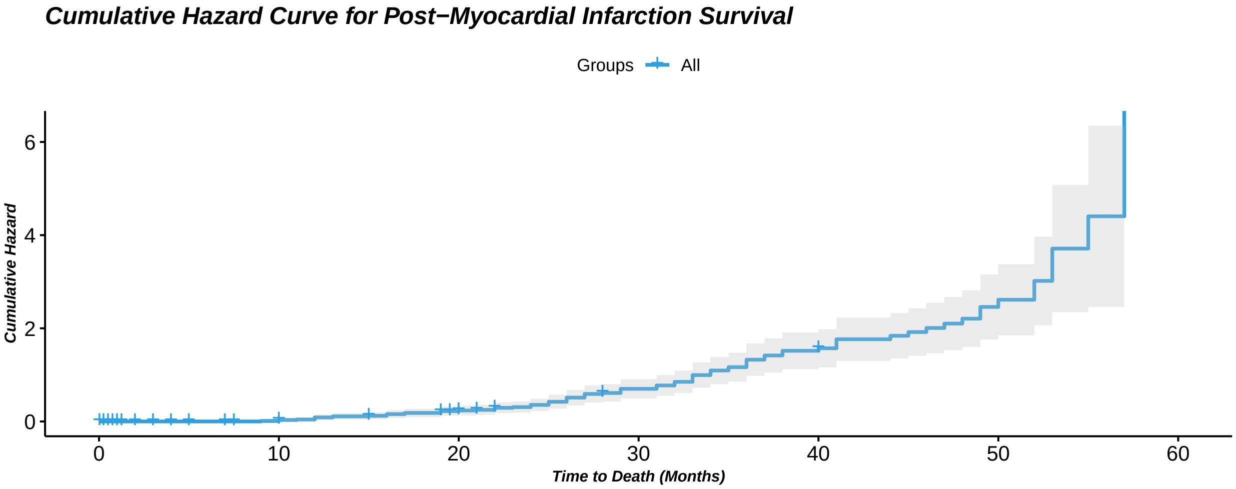


Table 2: Kaplan-Meier Estimates for All Groups

	Records	Events	Mean	Median	Median 0.95 LCL	Median 0.95 UCL
All Groups	130	88	30.53	29	27	33

The Kaplan-Meier estimates for for all groups within our dataset is shown above. The curve follows a general pattern of decreasing survivability over time. 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 29 months with 95% confidence limits between 27 and 33 months.

When testing for significant difference between strata groups we use the log-ran



To explore differences among groups, we stratify among age, pericardial effusion presence, wall motion score, and fractional shortening. We first begin exploring the effects of age and pericardial effusion presence:

Kaplan-Meier Stratified by Age and Pericardial Effusion Presence

The results of a Kaplan-Meier estimate for age and pericardial effusion stratification can be seen below:

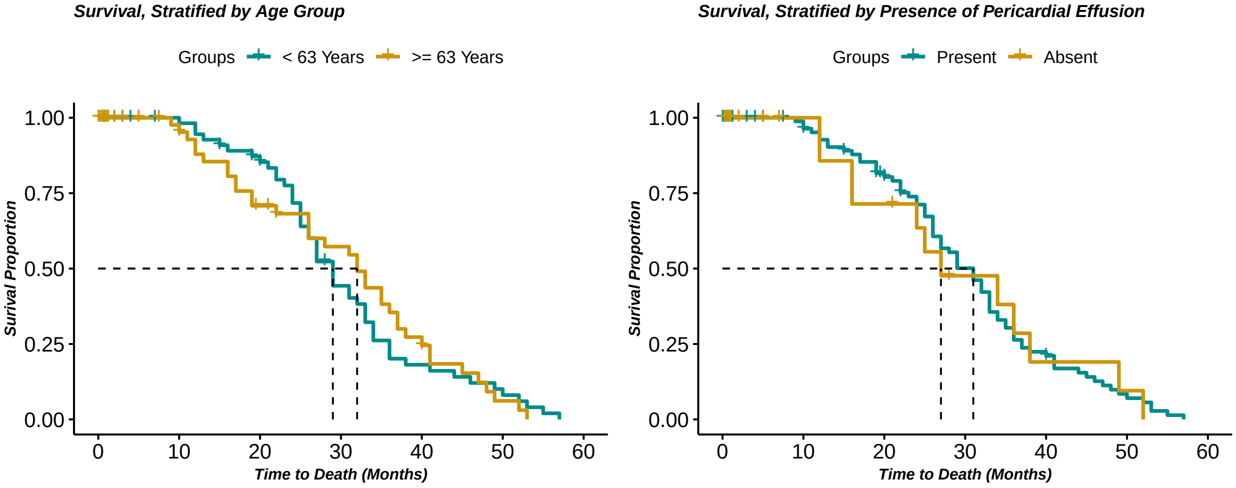


Table 3: Kaplan-Meier Estimates Stratified by Age and Pericardial Effusion Presence

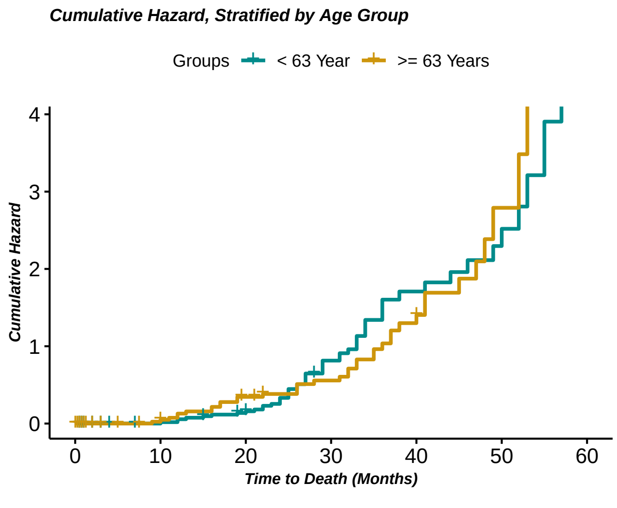
	Records	Events	Mean	Median	Median 0.95 LCL	Median 0.95 UCL
Age < 63	66	51	30.47	29	26	33
Age >= 63	64	37	30.60	32	26	37
Absent	106	76	30.63	31	27	33
Present	24	12	29.94	27	24	NA

Table 4: Summary of Differences Between Strata

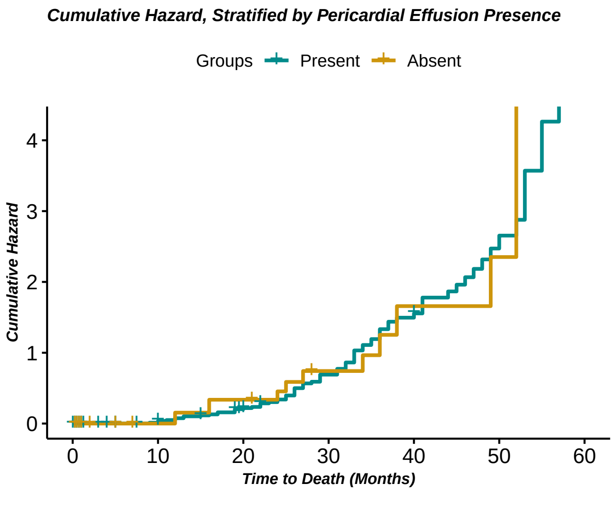
	N	Observed	Expected
Age < 63	66	51	50.58084
Age >= 63	64	37	37.41916
Absent	106	76	76.42260
Present	24	12	11.57740

When stratified by age, we find a slight difference between the curves. The age group younger than 63 has a mean survival time of 30.47 months with a median survival time of 29 months. The older group - ages greater than 63 - has a similar mean survival time of 30.6 months and a slightly longer median survival time of 32 months. When comparing the presence of pericardial effusion, there are 106 cases where the effusion is absent, while 24 cases have the effusion present. The mean survival time when pericardial effusion is absent is 30.63 months with a median survival time of 31 months. For converse case, the mean survival time is 29.94 months while the median is lower at 27 months.

Log-rank tests between both stratification groups returns a p-value of 0.9 for both age strata and effusion strata. When testing at the 95% significance level, we do not have significant differences between groups.



Cumulative Hazard, Stratified by Age Group

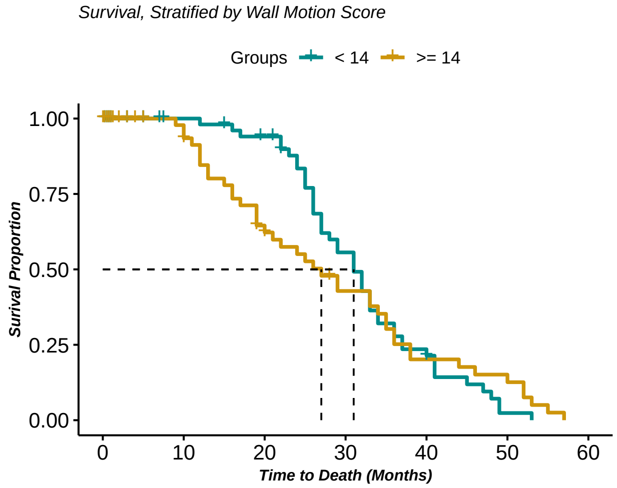


Cumulative Hazard, Stratified by Pericardial Effusion Presence

For both groups, there does not seem to be a large departure from cumulative hazard. When stratifying by age, we see a slight increase in cumulative hazard of the younger group between 30 and 50 months. After that mark, the older group experiences a relative increase in cumulative hazard. When stratified by pericardial effusion presence, very little difference can be observed with any difference being the result of sample size differences.

Kaplan-Meier Stratified by Wall Motion Score and Fractional Shortening Length:

We then explore the effects of wall motion score and fractional shortening:



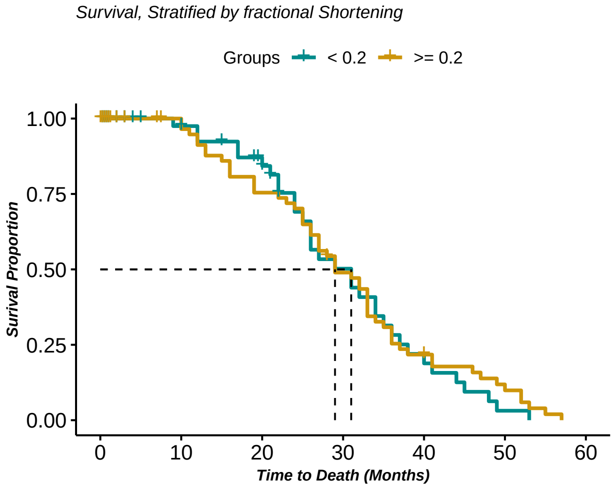


Table 5: Kaplan-Meier Estimates Stratified by Wall Motion Score and fractional Shortening

	Records	Events	Mean	Median	Median 0.95 LCL	Median 0.95 UCL
Score < 14	62	46	32.17	31	27	34
Score >= 14	68	42	28.61	27	20	35
Length < 0.2	61	33	30.45	31	26	36
Length >= 0.2	69	55	30.44	29	26	33

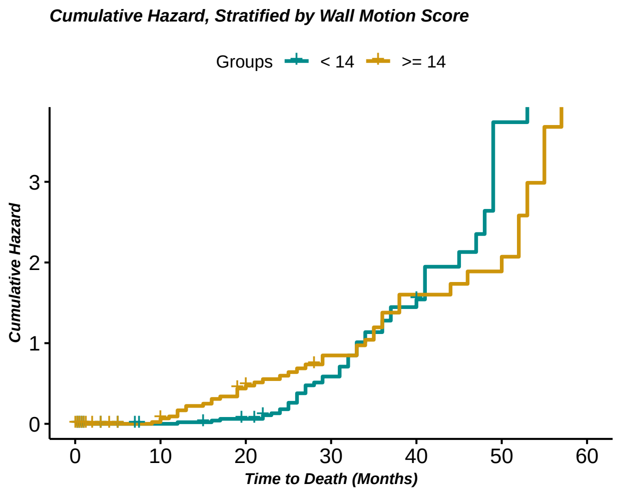
Table 6: Summary of Differences Between Strata

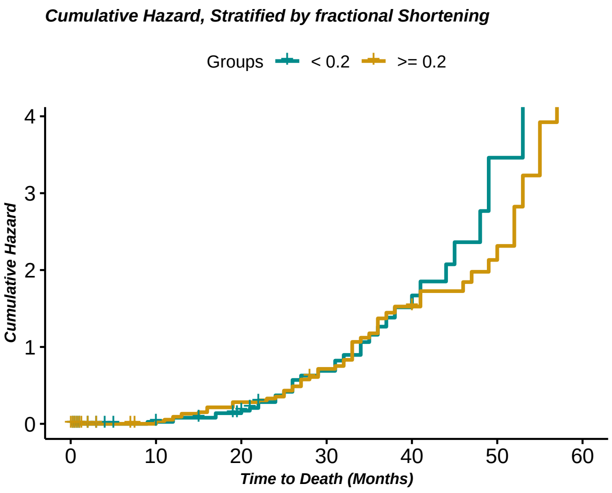
	N	Observed	Expected
Wall Motion Score < 14	62	46	46.76057
Wall Motion Score >= 14	68	42	41.23943
Fractional Shortening < 0.2	61	33	30.98374
Fractional Shortening >= 0.2	69	55	57.01626

When stratified by wall motion score, we find a difference between the curves. Wall motion scores less than 14 have a mean survival time of 32.17 months with a median survival time of 31 months. Wall motion scores greater or equal to 14 have a lower mean survival time of 28.61 months and a median survival time of 27 months.

When stratified by fractional shortening length, both groups have a similar mean at approximately 30.4 months. When fractional shortening is less than 0.2, the median survival time is 31 months while having a fractional shortening length that is greater than 0.2, we have a slightly lower median survival time of 29 months.

Log-rank tests to test differences between wall motion score strata show p-value of 0.9 while the same test produces a p-value of 0.6 for fractional shortening. Both of these failures fail to reject the null hypothesis at the 95% significance level. As such, there is no significant difference between either strata groupings.



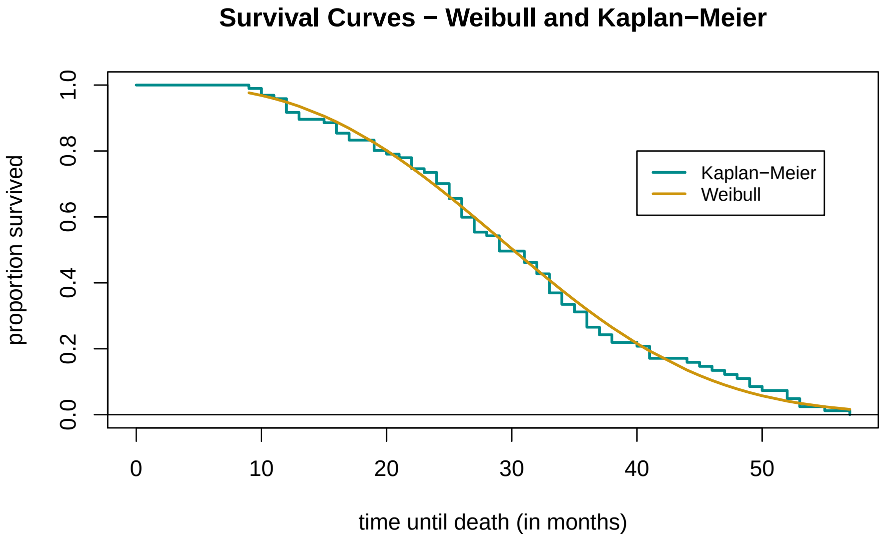


Here, we see some minute differences between the hazard curves. When stratified by Wall Motion Score, we see some overlap in the initial stages of the study as well as around approximately 35 months. The exceptions are seen with higher wall motion scores seeing increased risk before the median and decreased relative risk after the median. The converse is seen for the lower wall motion scores.

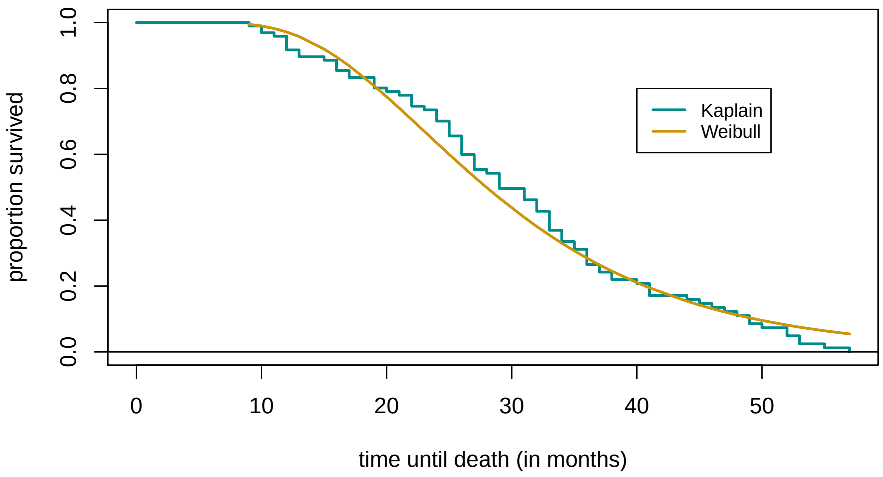
When stratified by fractral shortening, the cumulative hazard curves are approximately similar with higher fractional shortening lengths have less risk after the median.

Parametric Analysis and Estimation

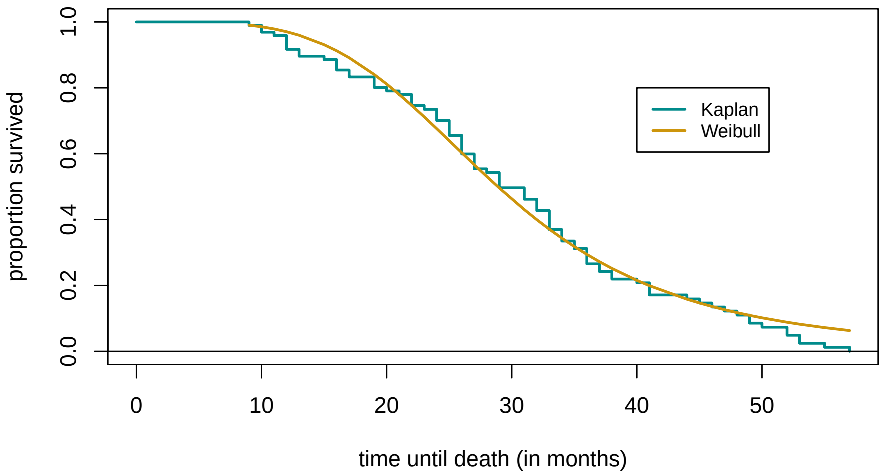
The estimated distributional model curves are overlaid on the K-M curve for the post-myocardial infarction data in the figures below.



Survival Curves − Log−normal and Kaplan−Meier



Survival Curves − Log−logistic and Kaplan−Meier



The table below summarizes the parameter point estimates and corresponding 95% confidence intervals.

Table 7: Parametric Model Estimates

Quantile	Point Estimate	95% LCL	95% UCL	Interval Length
Weibull Model Fit				
0.25	22.42	19.96	25.19	5.23
0.50	30.32	27.86	33.01	5.15
0.75	38.47	35.71	41.44	5.72
Log-normal Model Fit				
0.25	20.73	18.77	22.89	4.11
0.50	27.97	25.54	30.63	5.09
0.75	37.74	34.06	41.83	7.77
Log-logistic Model Fit				
0.25	21.90	19.77	24.27	4.50
0.50	28.88	26.41	31.59	5.18
0.75	38.09	34.45	42.12	7.67

Q-Q plots were also prepared for each distribution, and are provided in the Appendix. Based on inspection of the Q-Q plots and the K-M overlay plots, we found that the Weibull model appears to provide the best fit. In addition, it was noted that the confidence intervals for point estimation were overall more narrow for the Weibull model compared to log-normal and log-logistic. It was also observed that the point estimate for median is close to that identified by the Kaplan-Meier approach (29 months in K-M estimation, 30 months when estimating with Weibull fit). We propose the Weibull model as descriptive of the post-myocardial infarction survival data.

Regression Analysis - Cox Proportional Hazard Modeling

Regression analysis was conducted to identify the relationship between potential predictor variables and survival. As a first step, the covariate pool was screened for potential multicollinearity (see Appendix for plots of the covariates). Potential dependency was observed between Fractional Shortening, E-Point Septal Separation, and Left Ventrical Diastolic Dysfunction. Mechanistically, this is intuitive, since all three variables measure ventrical diastolic behavior. Models were fit with each of these collinear variables, and based on this assessment of Fractional Shortening was selected for inclusion in the final covariate pool of potential prognostic factors for regression. These covariates are:

- Age
- Pericardial Effusion
- Wall Motion Score
- Fractional Shortening

A model employing all of the covariates listed above was created. In order to test validity of the proportional hazard assumption, survival curves stratified by group were assessed for each of the four selected covariates. It was found that the proportional hazard assumption does not hold, as hazards for each group cross over with time for each covariate.

This limitation was addressed through the identification of survival time subsets in which covariates met the proportional hazards assumption, and could thus be employed as predictor variables. These subsets were identified by inspecting hazard function data stratified on all covariates and identifying regions of crossover. (See Appendix for plots of stratified hazard functions with the approximate crossover regions marked).

The selected time subsets were obtained as follows:

$Subset\,1 : t \leq 26\,months$

$Subset\,2 : 26\,months < t < 46\,months$

$Subset\,3 : 46\,months \leq t$

Within each time subset, the proportional hazard assumption was tested for each of the four covariates and found to hold. A Cox PH model was fitted to each subset of the survival data, reduced as far as possible using a Step AIC procedure up to second order interactions with Likelihood Ratio Test selection of the final model, and evaluated for adequacy using standard diagnostic techniques.

The three models are summarized below:

Table 8: Summary of CoxPH Regression Models

	Model 1			Model 2			Model 3		
	LRT = 10.34 on 3 df; p = 0.02			LRT = 13.9 on 1 df; p = 0.05			LRT = 10.34 on 1 df; p = 0.1		
	n = 77; #events = 37			n = 42; #events = 40			n = 11; #events = 11		
Covariate	HR	95% CI	p-val	HR	95% CI	p-val	HR	95% CI	p-val
FS	0.85	(0.69, 1.05)	0.13				0.9	(0.79, 1.02)	0.11
WMS	0.00	(0.00, 0.01)	9						
WMS*FS	4.31	(1.51, 12.4)	006						
Age				0.96	(0.92, 1.00)	0.051			

For Model 1, the significant predictors of survival time are identified as Fractional Shortening and its interaction with Wall Motion Score. Although Wall Motion Score does not have a significant effect on survival as a main effect, it is retained due to is presence in the interaction term.

For Model 2, the significant predictor of survival time is identified as patient age. Age is found to have a very weak effect on survival; none of the echocardiographic covariates had an impact on survival in this time region. Both these findings are supported by inspection of the stratified hazard function plots.

For subset 3, the sample size of events in the survival data subset was small (n=11) and a significant model was not obtained. The model is presented for context but is poorly descriptive of the relationship between prognostic factors and survival.

Model residual diagnostics were assessed for all three models. Models 1 and 2 showed overall good fit to the data and demonstrated proportional hazard assumption compliance. As expected given the discussion above, Model 3 did not result in a good overall fit, although proportional hazard assumption was met. Model diagnostic residual plots and further discussion may be found in the Appendix.

Discussion

Explanation of results

In our non-parametric analysis, median survival times for nearly all of the stratification elements show relatively similar results with the exception of high wall motion scores. When stratified by age group, we see relatively similar survival proportions with some distinction before and after the median. Before the median, the older group has a slightly lower survival proportion. However, after the median, the older group has a slightly higher survival proportion before 45 months. This slight discrepancy could be due to the externalities attached to young heart attack patients. If patients are experiencing heart attacks at a younger age, there could be higher chances that those patients already have other health concerns. Patients experiencing and surviving a heart attack at older ages could be in physically superior condition than their younger counterparts. All other stratification groups show relatively similar median survival proportions.

An interesting observation can be seen when we arbitrarily divide the age strata into three groups. We see a much greater risk associated with patients younger than 55 while all other age groups perform similarly.

In the parametric analysis, it was found that the survivor data is well-described by the Weibull distribution. This finding enables the construction of a continuous survivor function and provides the opportunity for flexible point estimation as well as estimation of survival probability over a given input time. Semi-parametric modeling was also performed. Three Cox Proportional Hazards models were developed that described the relationship between prognostic factors and survival; one for short survival times; one for intermediate survival times; and another for long survival times. Given a small sample size for long survival times, only the first two models were significant. A key finding from this analysis is that when survival time is shorter, abnormalities in the echocardiographic profile of the patient's heart are what predict survival, rather than age. Conversely, when the survival times are longer, it is the age of the patient that predicts survival (albeit weakly), rather than any of the echocardiographic properties do not. This is consistent with the behavior of the stratified hazard curves for these variables over time. This result seems to imply that in the acute recovery period following a myocardial infarction, patient prognosis is closely tied to heart health; but as time progresses, patient age becomes the key driver of prognosis. This finding is intuitive, since as treatment and recovery progress over time, risk related to the myocardial event would decrease and other pre-existing risks would play a larger role in survival.

Summary of Limitations

Very clearly, our data is smaller than we hoped for, both in the number of observations and in the availability of meaningful prognostic factors. Development of a single large regression fit that reliably predicts covariates would require a much larger dataset. In particular, the lack of gender as an available covariate is a potential limitation of our data, as the effect of gender on heart-related survival is well established in the literature. Longer collection period over a greater number of patients would yield stronger regression model development. As-is, there is not enough data to produce a significant model across the entire survival time window.

A particular limitation related to subgroup dimension was identified through the stratification analysis. Examination of stratified data showed relative unequal distributions among groups. For example, when examining pericardial effusion, we compare 106 records of with effusion absent to 24 records of effusion present. Given that the collection methods are unclear and the groups sizes are so much different form each other, we cannot fully be confident in our results. Such comparisons are unequal comparisons.

Finally, the given dataset did not have clear statements as to what data collection methods were used. Given our smaller sample size and these unknown methods, we cannot be entirely confident that our results reflect the survival behavior of the population as a whole. Future studies would benefit greatly

Conclusion

This study identified a variety of properties of the post-myocardial infarction survival data. The overall survival function was characterized by both non-parametric Kaplan-Meier and Weibull models, and point estimates determined. Analysis of the effect of echocardiographic data and of age on survival was performed both by inspection of stratified survival curves and by regression analysis. Findings included that age group has an impact on survival, but that this effect potentially varies over time, and that certain heart-related prognostic factors (in particular wall motion score and its interaction with fractional shortening) also have an impact on survival that varies over time. We conclude that future studies should expand on collection of variables that could potentially influence survivability, as well as the cohort size and length of follow-up period. Interesting effects to include may include poverty, diet, ethnicity, race, sex, and even work place stress/effects. Finally, further investigation with a larger dataset on the relationship between acute survival time and echocardiographic data profile and between the relationship of age group on survival data may yield useful insights into specific risk profiles for post-myocardial infarction patients.

Appendix

References

1. 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

2. 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.

3. 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

4. Dalen JE, Alpert JS, Goldberg RJ, Weinstein RS. The epidemic of the 20(th) century: coronary heart disease. Am J Med. 2014;127(9):807-812. doi:10.1016/j.amjmed.2014.04.015

5. 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

6. 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

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

8. 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.

9. 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

10. 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

11. 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).

12. 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

13. 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

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

15. 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 9: 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 9: 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 10: 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		7.000		2.00	1.000	0
0.25	0	1	75.00	2	0					1.000	0
36.00	1	0	55.00	1	1	0.210	4.200	4.160	14.00	1.560	1

Table 10: 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		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		2	0		12.000		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		10.000		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			5.310	5.00	1.000	0
1.00	0	1	65.00	2	0	0.060	23.600		21.50	2.150	1
19.00	1	0	68.00	2	0	0.510		3.880	15.00	1.670	1
31.00	1	0	80.00	2	0	0.410	5.400	4.360		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		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			12.00	2.000	1
19.50	0	1	81.00	2	0	0.120			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				23.00	2.300	1
2.00	0	1	56.00	1	1	0.040	14.000	5.000			1
7.00	0	1	61.00	1	1	0.270			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		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		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 10: Original Dataset *(continued)*

Survival	Status	Alive.E	Age	Age.Strata	P.Effusion	F.Shortening	EPSS	LVDD	WMS	WMI	WMI.S
0.03	0	1		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				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		4.690	10.00	1.000	0
33.00	1	0	66.00	2	0	0.200		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		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		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 10: 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		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 11: 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

Table 11: 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
2.00	0	1	67.00	1	0.44	9.00	3.96	17.50	1.45	1	1	1	0	0
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

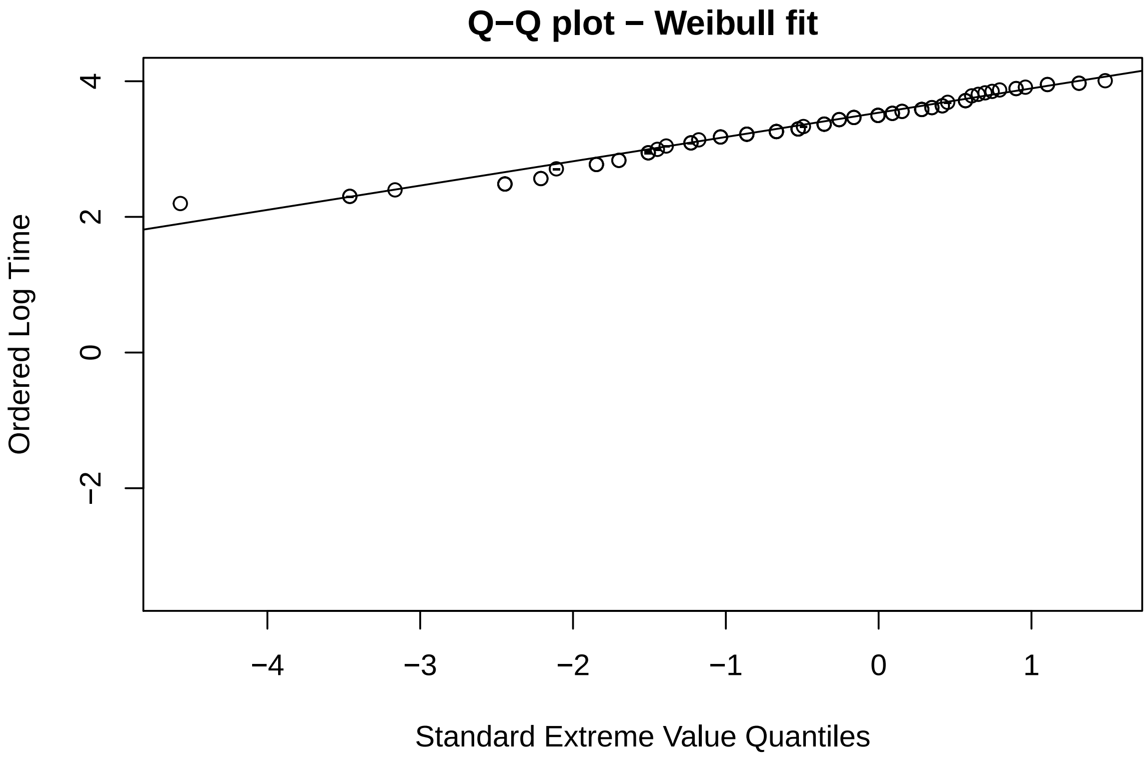
Table 11: Imputed Dataset *(continued)*

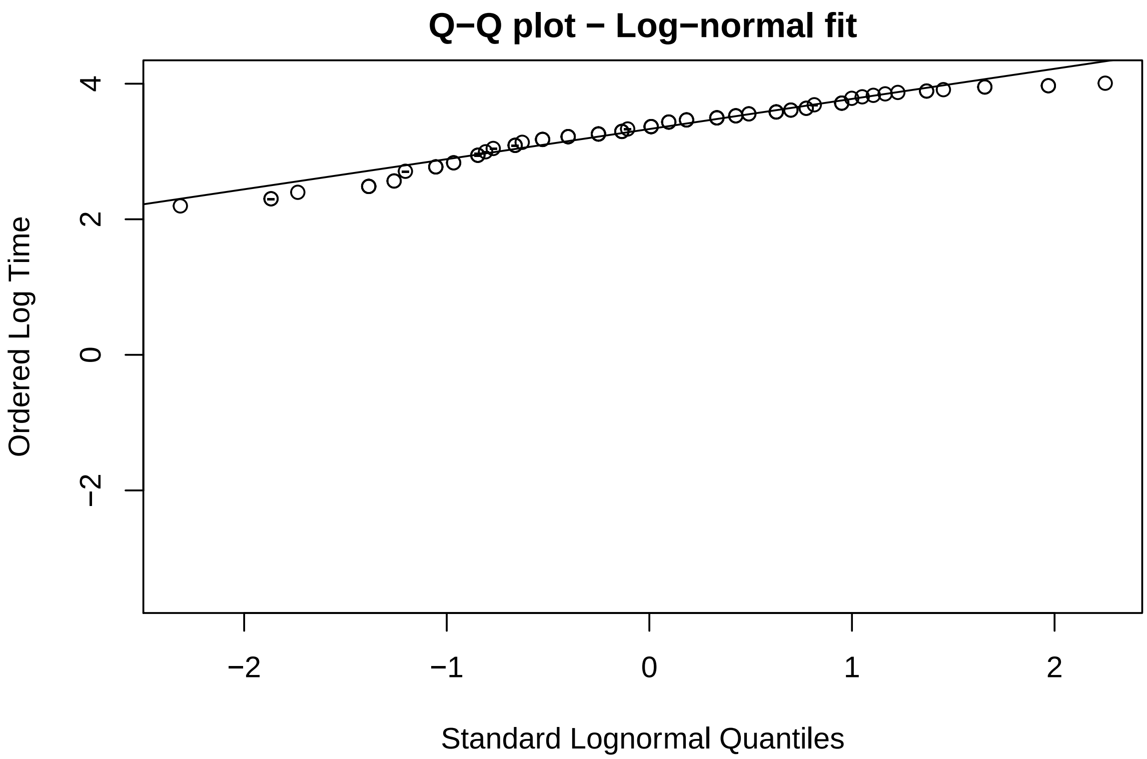
Survival	Status	Alive.E	Age	P.Effusion	F.Shortening	EPSS	LVDD	WMS	WMI	Age.s	WMS.s	F.Short.s	LVDD.s	EPSS.s
35.00	1	0	64.00	0	0.30	6.60	4.36	14.00	1.27	1	1	1	0	0
53.00	1	0	59.00	0	0.34	9.10	4.04	9.00	1.00	0	0	1	0	0
33.00	1	0	46.00	0	0.27	16.50	5.36	12.67	1.06	0	0	1	1	1
33.00	1	0	63.00	0	0.25	5.60	3.87	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

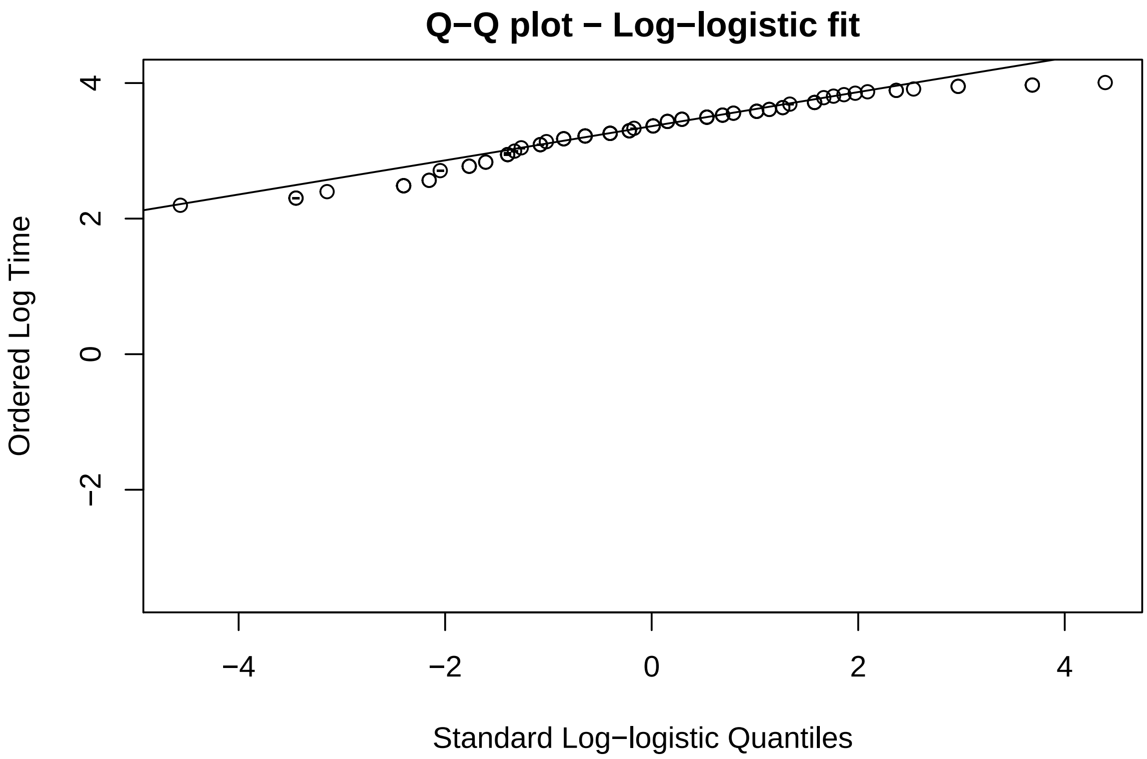
Table of Kaplan-Meier Estimators

Parametric Model Q-Q Plots

Q-Q plots for each of the three assessed parametric models are provided below. We see that the Weibull model has the best fit, with no large departures from the straight line.

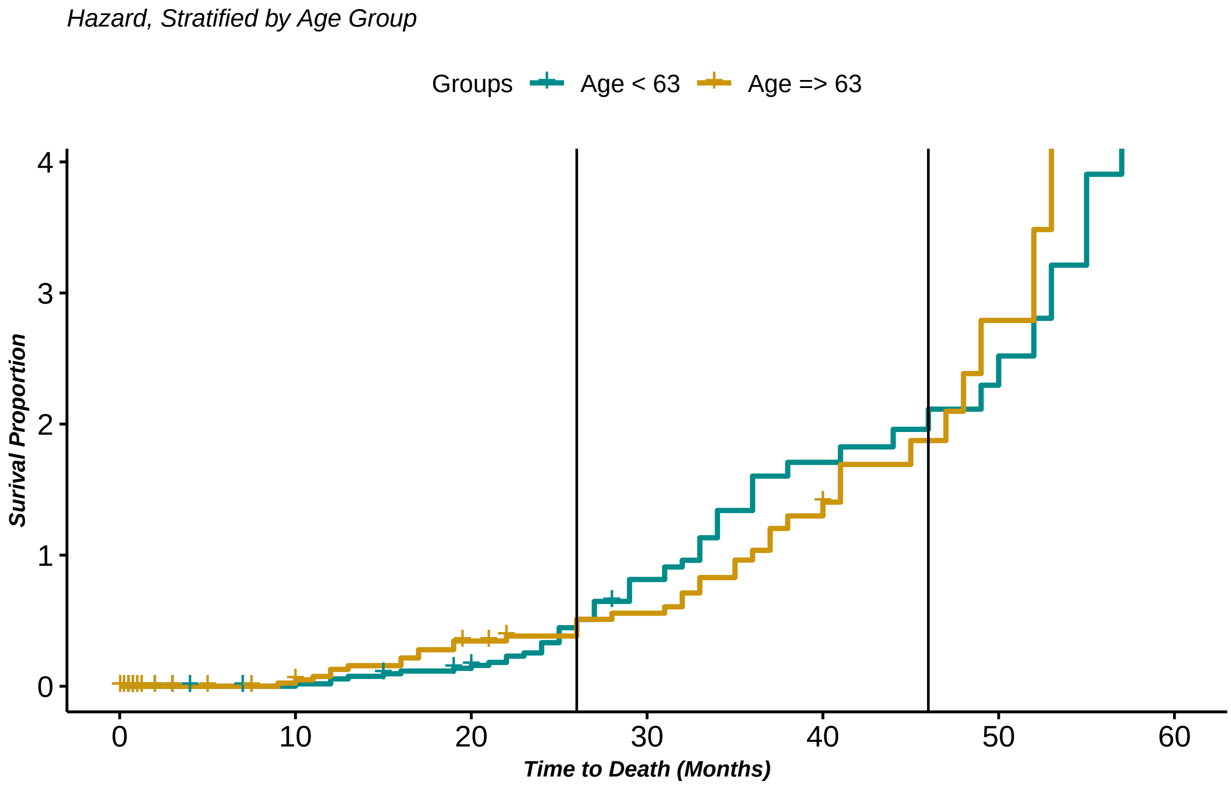


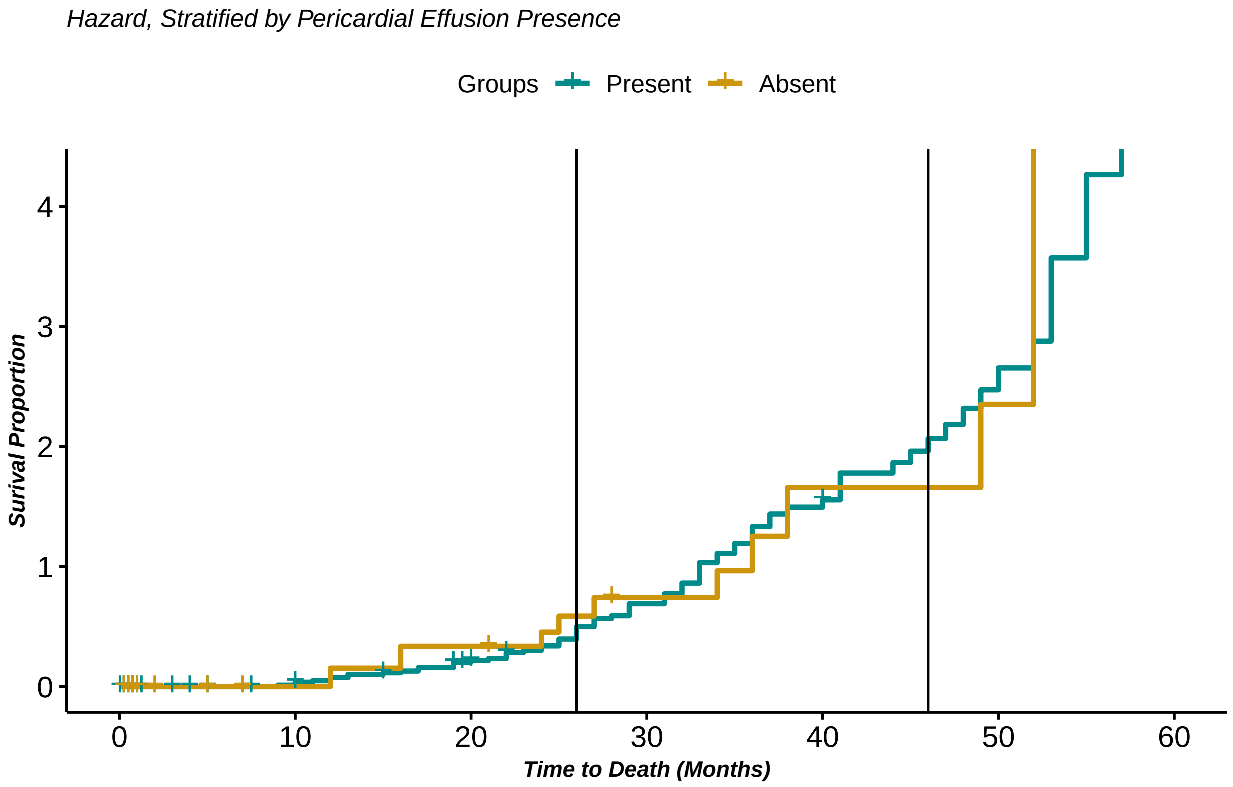

Q−Q plot − Log−normal fit
Ordered Log Time
4
2
0
−2
−2
−1
0
1
2
Standard Lognormal Quantiles


Q−Q plot − Log−logistic fit
Ordered Log Time
4
2
0
−2
−4
−2
0
2
4
Standard Log−logistic Quantiles

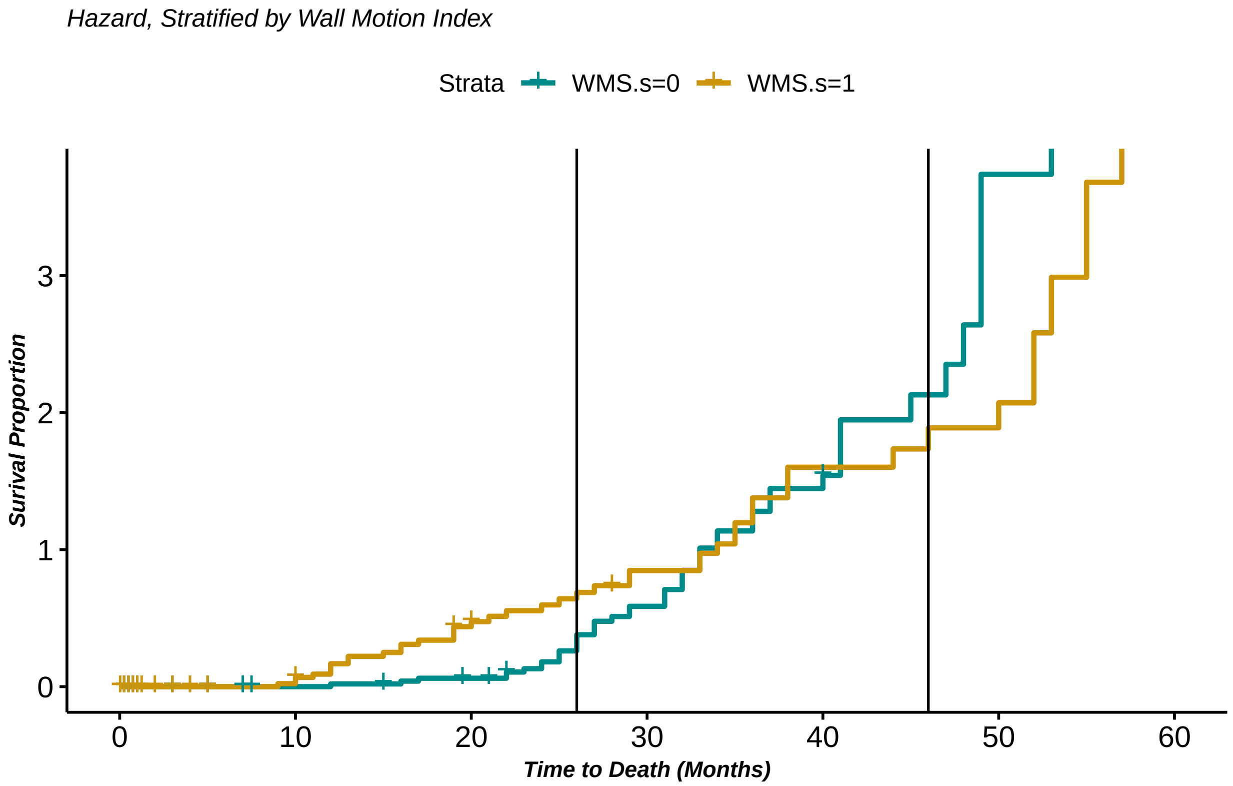
Hazard Plots for Regression Covariates - Regression Time Interval Identification

Hazard curves stratified by group for each regression covariate are provided below, with the three time interval for the cox PH modeling marked on each plot. It can be observed that these time intervals approximately correspond to hazard crossover behavior for the covariates, with the exception of pericardial infusion.

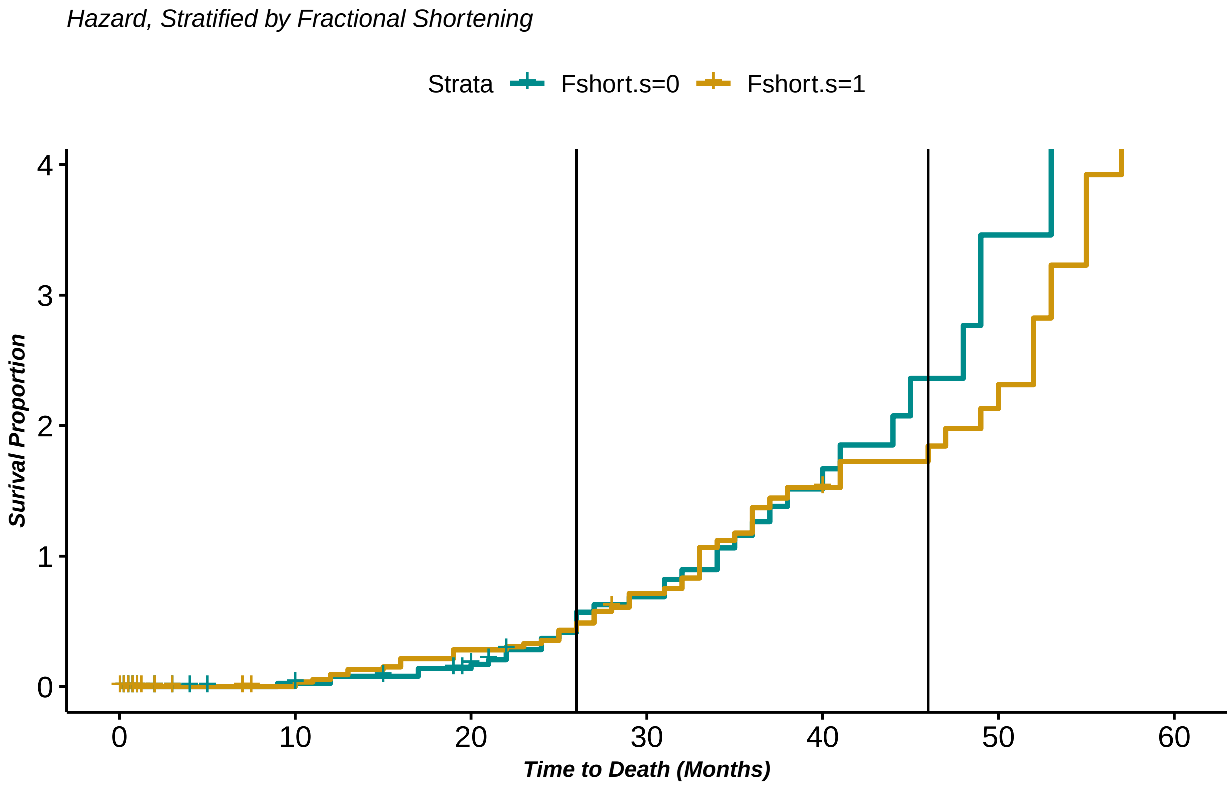




Hazard, Stratified by Pericardial Effusion Presence



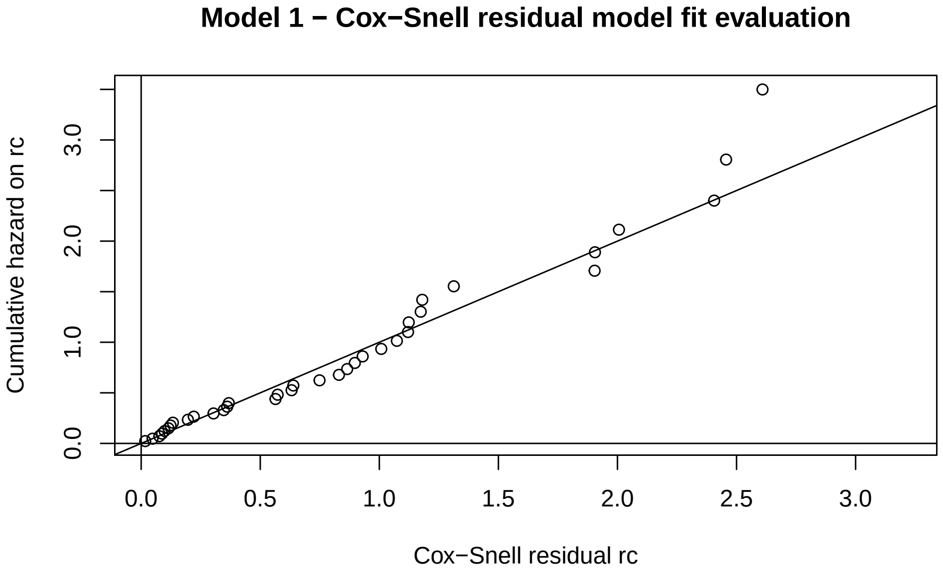
Hazard, Stratified by Wall Motion Index

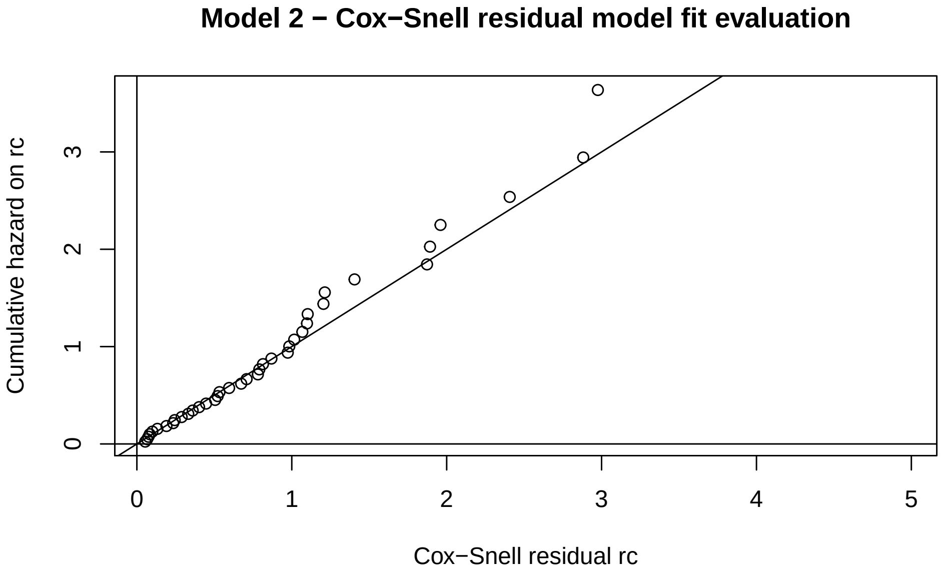


Hazard, Stratified by Fractional Shortening
Strata
Fshort.s=0
Fshort.s=1
Surival Proportion
0
1
2
3
4
0
10
20
30
40
50
60
Time to Death (Months)

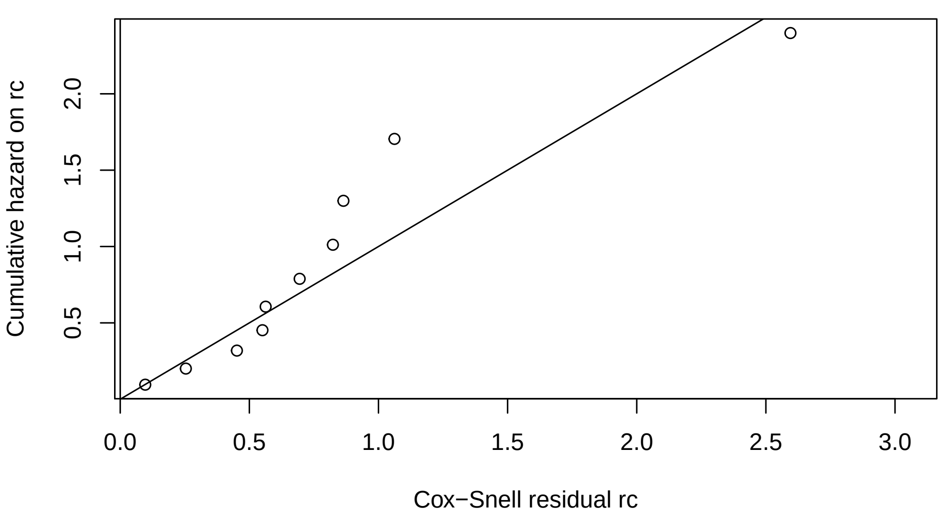
Regression Model Diagnostics

Cox-Snell residual plot for assessment of overall model fit and Schoenfeld residual plots for evaluation of constant coefficients are presented below for each of the three models. It can be seen that the Models 1 and 2 fit the data well overall given that the Cox-Snell residuals fall closely along the straight line. and that Schoenfeld residuals are symmetric about 0 with large p-values. However, these assumptions are not met for Model 3, which is not unexpected due to the small sample size.

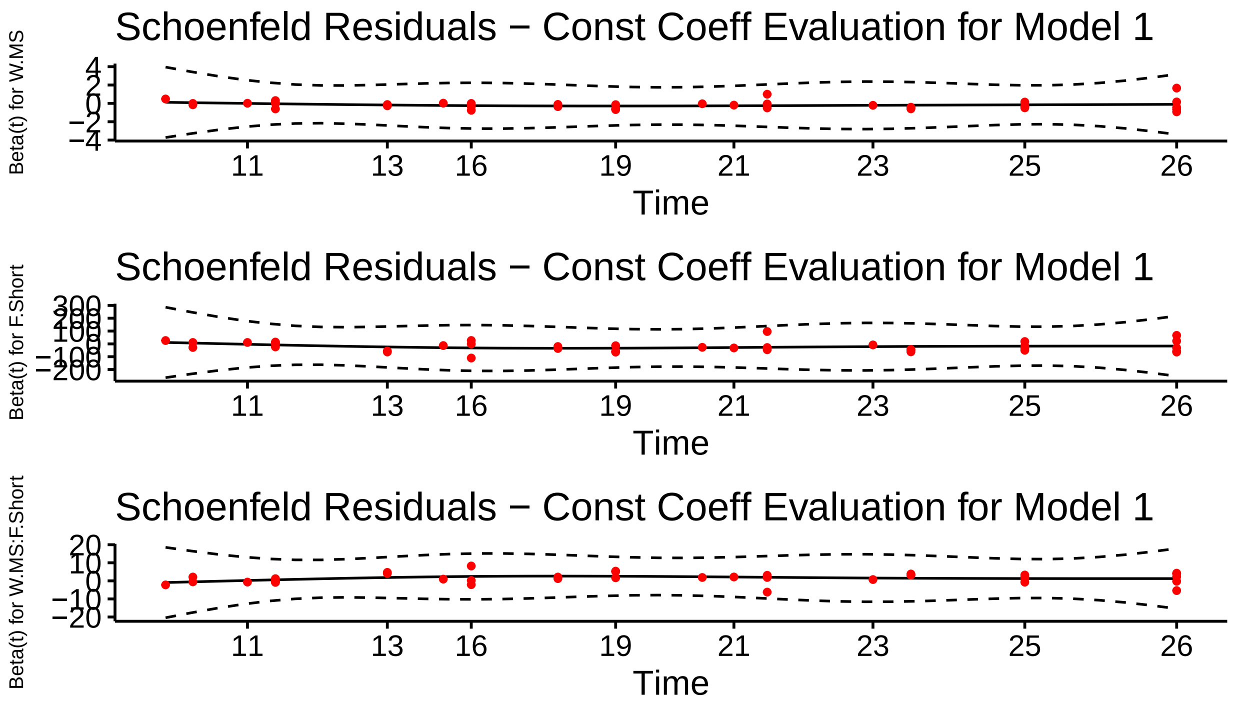




Model 3 − Cox−Snell residual model fit evaluation

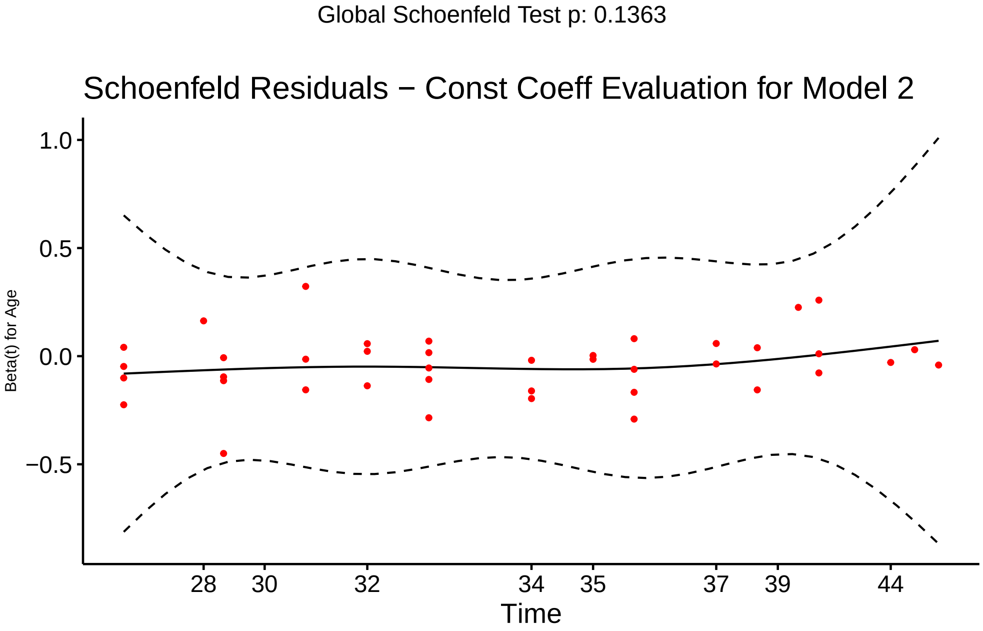


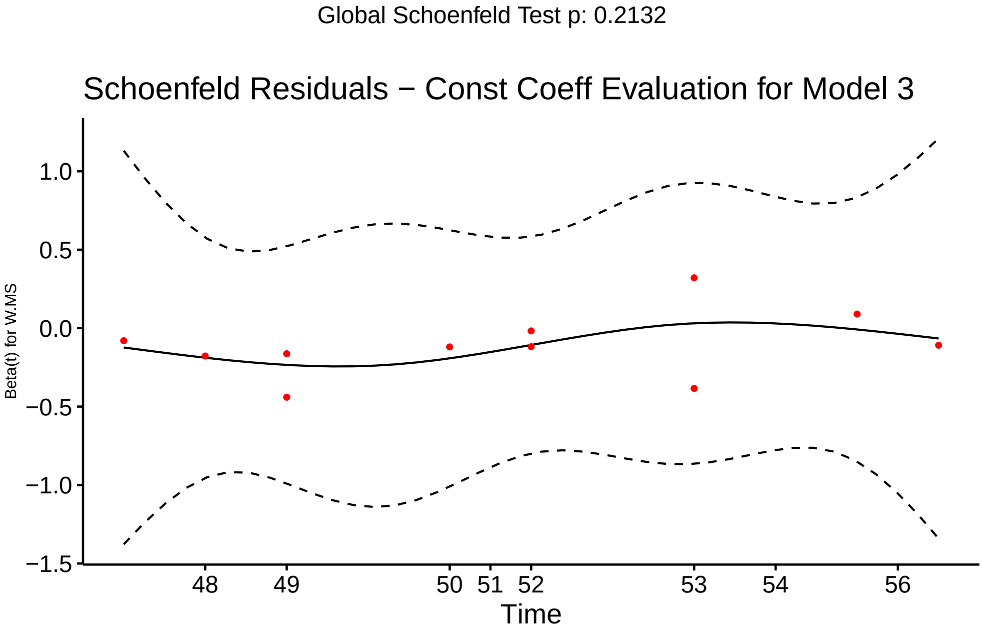
Global Schoenfeld Test p: 0.8463


Schoenfeld Residuals − Const Coeff Evaluation for Model 1

Schoenfeld Residuals − Const Coeff Evaluation for Model 1

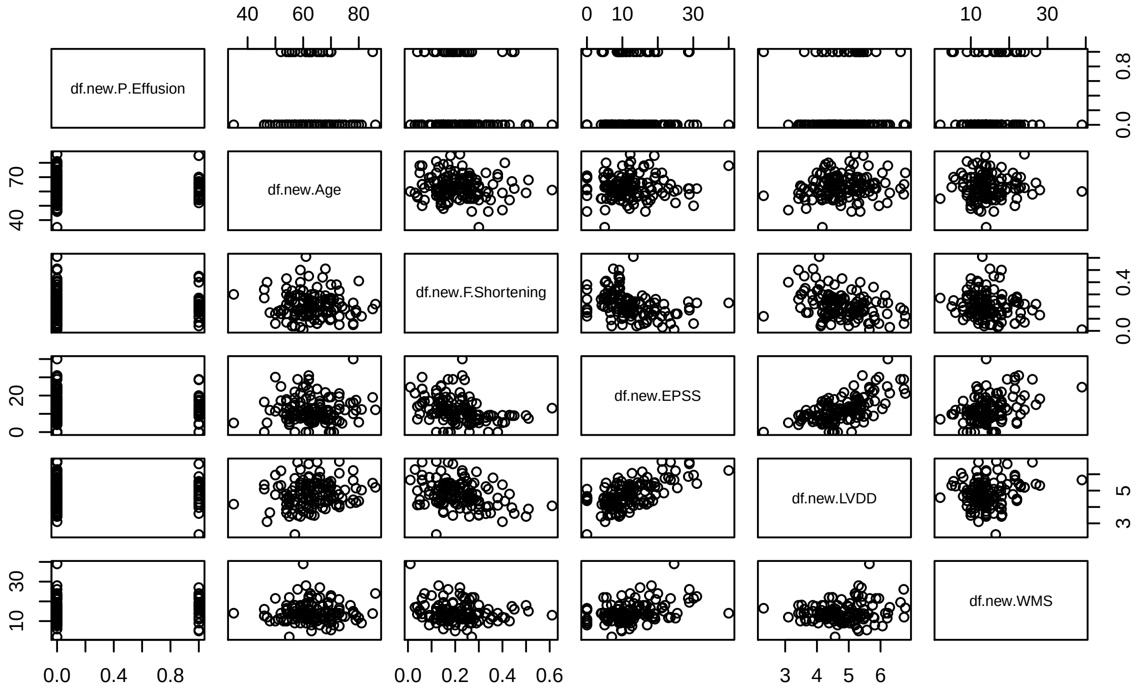
Schoenfeld Residuals − Const Coeff Evaluation for Model 1


Global Schoenfeld Test p: 0.1363
Schoenfeld Residuals − Const Coeff Evaluation for Model 2
Beta(t) for Age
1.0
0.5
0.0
−0.5
28
30
32
34
35
37
39
44
Time


Global Schoenfeld Test p: 0.2132
Schoenfeld Residuals − Const Coeff Evaluation for Model 3
Beta(t) for W.MS
1.0
0.5
0.0
−0.5
−1.0
−1.5
48
49
50
51
52
53
54
56
Time

Pairs Plot of Covariates

Pairs plots of the prognistic factors in the dataset are provided below. These plots were investigated as part of the covariate selection process for identification of possible multicollinearity.



R Code

```r
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)
library(MASS)
library(SurvCorr)
library(broom)

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")
Age = c("< 63 Years", ">= 63 Years")
Effusion = c("Fluid is absent", "Fluid is present")
WMS = c("< 11", ">= 11")
FS = c("< 0.2",">= 0.2")

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

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$WMS < 14,0,1) #new WMS strata based on imputed data
Fshort.s = ifelse(df.impute$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.all = survfit(s.df~1,type="kaplan-meier", data=df.new)
km.allp = ggsurvplot(km.all,
                           palette = "#2E9FDF",
                           conf.int = TRUE,
                           title="Kaplan-Meier Curve for Post-Myocardial Infarction Survival",
                           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.allp

ks1 = data.frame(t(summary(km.all)$table))
ks1 = ks1[,c(1,4,5,7,8,9)]
colnames(ks1) = c("Records","Events","Mean","Median","Median 0.95 LCL","Median 0.95 UCL")
rownames(ks1) = c("All Groups")

kable(ks1, caption="Kaplan-Meier Estimates for All Groups",align="c", digits=2) %>%
  kable_styling(position = "center", latex_options="hold_position")
haz.all = ggsurvplot(km.all,
           fun = "cumhaz",
           palette = "#2E9FDF",
           conf.int = TRUE,
           title="Cumulative Hazard Curve for Post-Myocardial Infarction Survival",
           font.title=c(14,"bold.italic"),
           font.subtitle = c(10,"italic"),
           font.x = c(9, "bold.italic"),
           font.y = c(9, "bold.italic"),
           ylab="Cumulative Hazard",
           xlab="Time to Death (Months)",
           legend.title = "Groups",
           legend.labs = "All")

haz.all
km.p1 = list()

km.age = survfit(s.df~Age.s, type="kaplan-meier", data = df.new)
km.p1[[1]] = ggsurvplot(km.age,
                           palette = c("darkcyan","darkgoldenrod3","darkorange3"),
                           subtitle="Survival, Stratified by Age Group",
                           font.subtitle = c(10,"bold.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",">= 63 Years"))

km.effusion = survfit(s.df~P.Effusion, type="kaplan-meier", data = df.new)
km.p1[[2]] = ggsurvplot(km.effusion,
                           palette = c("darkcyan","darkgoldenrod3"),
                           subtitle="Survival, Stratified by Presence of Pericardial Effusion",
                           font.subtitle = c(10,"bold.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("Present","Absent"))

arrange_ggsurvplots(km.p1, print=TRUE, ncol=2, nrow=1)
ks2 = data.frame(summary(km.age)$table)
ks3 = data.frame(summary(km.effusion)$table)

ks2.3 = rbind(ks2, ks3)
ks2.3 = ks2.3[,c(1,4,5,7,8,9)]
colnames(ks2.3) = c("Records","Events","Mean","Median","Median 0.95 LCL","Median 0.95 UCL")
rownames(ks2.3) = c("Age < 63", "Age >= 63","Absent","Present")

kable(ks2.3, caption="Kaplan-Meier Estimates Stratified by Age and Pericardial Effusion Presence",align=
  kable_styling(position = "center", latex_options="hold_position")
km.agediff = survdiff(s.df~Age.s, data = df.new)
km.effusiondiff = survdiff(s.df~P.Effusion, data = df.new)

ks2.diff = data.frame(tidy(km.agediff))
ks3.diff = data.frame(tidy(km.effusiondiff))

ks2.diff = ks2.diff[,c(2,3,4)]
ks3.diff = ks3.diff[,c(2,3,4)]
ks2.3diff = rbind(ks2.diff,ks3.diff)
colnames(ks2.3diff) = c("N","Observed","Expected")
rownames(ks2.3diff) = c("Age < 63", "Age >= 63","Absent","Present")

kable(ks2.3diff, caption = "Summary of Differences Between Strata")%>%
  kable_styling(position = "center", latex_options="hold_position")
haz.p1 = list()

haz.p1[[1]] = ggsurvplot(km.age,
           fun = "cumhaz",
           palette = c("darkcyan","darkgoldenrod3"),
           title="Cumulative Hazard, Stratified by Age Group",
           font.title = c(10,"bold.italic"),
           font.x = c(9, "bold.italic"),
           font.y = c(9, "bold.italic"),
           ylab="Cumulative Hazard",
           xlab="Time to Death (Months)",
           legend.title = "Groups",
           legend.labs = c("< 63 Year",">= 63 Years"))

haz.p1[[2]] = ggsurvplot(km.effusion,
           fun = "cumhaz",
           palette = c("darkcyan","darkgoldenrod3"),
           title="Cumulative Hazard, Stratified by Pericardial Effusion Presence",
           font.title = c(10,"bold.italic"),
           font.x = c(9, "bold.italic"),
           font.y = c(9, "bold.italic"),
           ylab="Cumulative Hazard",
           xlab="Time to Death (Months)",
           legend.title = "Groups",
           legend.labs = c("Present","Absent"))

arrange_ggsurvplots(haz.p1, print=TRUE, ncol=2, nrow=1)
km.p2 = list()

km.wms = survfit(s.df~WMS.s, type="kaplan-meier", data = df.new)
km.p2[[1]] = 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",">= 14"))

km.fshort = survfit(s.df~Fshort.s, type="kaplan-meier", data = df.new)
km.p2[[2]] = ggsurvplot(km.fshort,
                           palette = c("darkcyan","darkgoldenrod3"),
                           subtitle="Survival, Stratified by fractional 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",">= 0.2"))

arrange_ggsurvplots(km.p2, print=TRUE, ncol=2, nrow=1)
ks4 = data.frame(summary(km.wms)$table)
ks5 = data.frame(summary(km.fshort)$table)
```

```r
ks4.5 = rbind(ks4, ks5)
ks4.5 = ks4.5[,c(1,4,5,7,8,9)]
colnames(ks4.5) = c("Records","Events","Mean","Median","Median 0.95 LCL","Median 0.95 UCL")
rownames(ks4.5) = c("Score < 14", "Score >= 14","Length < 0.2", "Length >= 0.2")

kable(ks4.5, caption="Kaplan-Meier Estimates Stratified by Wall Motion Score and fractional Shortening"
  kable_styling(position = "center", latex_options="hold_position")
km.wmsdiff = survdiff(s.df~WMS.s, data = df.new)
km.f.shortdiff = survdiff(s.df~Fshort.s, data = df.new)

ks4.diff = data.frame(tidy(km.wmsdiff))
ks5.diff = data.frame(tidy(km.f.shortdiff))

ks4.diff = ks4.diff[,c(2,3,4)]
ks5.diff = ks5.diff[,c(2,3,4)]
ks4.5diff = rbind(ks4.diff,ks5.diff)
colnames(ks4.5diff) = c("N","Observed","Expected")
rownames(ks4.5diff) = c("Wall Motion Score < 14", "Wall Motion Score >= 14","Fractional Shortening < 0.

kable(ks4.5diff, caption = "Summary of Differences Between Strata")%>%
  kable_styling(position = "center", latex_options="hold_position")
haz.p2 = list()

haz.p2[[1]] = ggsurvplot(km.wms,
           fun = "cumhaz",
           palette = c("darkcyan","darkgoldenrod3"),
           title="Cumulative Hazard, Stratified by Wall Motion Score",
           font.title = c(10,"bold.italic"),
           font.x = c(9, "bold.italic"),
           font.y = c(9, "bold.italic"),
           ylab="Cumulative Hazard",
           xlab="Time to Death (Months)",
           legend.title = "Groups",
           legend.labs = c("< 14",">= 14"))

haz.p2[[2]] = ggsurvplot(km.fshort,
           fun = "cumhaz",
           palette = c("darkcyan","darkgoldenrod3"),
           title="Cumulative Hazard, Stratified by fractional Shortening",
           font.title = c(10,"bold.italic"),
           font.x = c(9, "bold.italic"),
           font.y = c(9, "bold.italic"),
           ylab="Cumulative Hazard",
           xlab="Time to Death (Months)",
           legend.title = "Groups",
           legend.labs = c("< 0.2",">= 0.2"))

arrange_ggsurvplots(haz.p2, print=TRUE, ncol=2, nrow=1)

months=df.new$Survival
status=df.new$Status
months.u=months[status == 1]
months.u = sort(months.u)
```

```r
nu = length(months.u)

#Weibull model plot

weib.fit=survreg(Surv(months,status)~1,dist="weib")
alphahat=1/weib.fit$scale
scalehat=exp(weib.fit$coefficients)
Shat.w = 1- pweibull(months.u,alphahat,scalehat)
plot(km.all,conf.int=F,xlab="time until death (in months)",
     ylab="proportion survived",
     main= "Survival Curves - Weibull and Kaplan-Meier",
     lwd=2,
     col = "darkcyan")
lines(months.u, Shat.w, col="darkgoldenrod3",lwd=2)
legend(40, 0.8, legend=c("Kaplan-Meier", "Weibull"),
       col=c("darkcyan","darkgoldenrod3"), lty=1:1, cex=0.8,lwd=2)
abline(h=0)

#log-normal model plot

lognorm.fit=survreg(Surv(months,status)~1,dist="lognormal")
muhat=lognorm.fit$coefficients
sigmahat=lognorm.fit$scale
Shat.l = 1- pnorm(log(months.u),muhat,sigmahat)
plot(km.all,conf.int=F,xlab="time until death (in months)",
     ylab="proportion survived",
     main="Survival Curves - Log-normal and Kaplan-Meier",
     lwd=2,
     col="darkcyan")
lines(months.u, Shat.l, col="darkgoldenrod3",lwd=2)
legend(40, 0.8, legend=c("Kaplain", "Weibull"), lwd=2,
       col=c("darkcyan","darkgoldenrod3"), lty=1:1, cex=0.8)
abline(h=0)

#log-logistic model plot

loglog.fit=survreg(Surv(months,status)~1,dist="loglogistic")
muhat=loglog.fit$coefficients
sigmahat=loglog.fit$scale
Shat.ll = 1- plogis(log(months.u),muhat,sigmahat)
plot(km.all,conf.int=F,xlab="time until death (in months)",
     ylab="proportion survived",
     main="Survival Curves - Log-logistic and Kaplan-Meier",
     lwd=2,
     col="darkcyan")
lines(months.u, Shat.ll, col="darkgoldenrod3",lwd=2,)
legend(40, 0.8, legend=c("Kaplan", "Weibull"), lwd=2,
       col=c("darkcyan","darkgoldenrod3"), lty=1:1, cex=0.8)
abline(h=0)

library(kableExtra)
param.est = data.frame(read_excel("param.est.xlsx"))
kable(param.est,
      col.names = c("Quantile", "Point Estimate", "95% LCL", "95% UCL",
      "Interval Length"),
       caption="Parametric Model Estimates",align="c", digits=2)%>%
    group_rows("Weibull Model Fit", 1, 3) %>%
    group_rows("Log-normal Model Fit", 4,6)%>%
    group_rows("Log-logistic Model Fit", 7,9)%>%
    kable_styling(position = "center", latex_options="HOLD_position")

options(knitr.kable.NA = '')
library(kableExtra)
models = data.frame(read_excel("models.xlsx"))
kable(models,
     col.names =c("Covariate","HR","95% CI","p-val", "HR","95% CI","p-val","HR","95% CI","p-val"),
      caption="Summary of CoxPH Regression Models",align="c", digits=2)%>%
      add_header_above(c(" " = 1,"n = 77; #events = 37" = 3, "n = 42; #events = 40" = 3,
                         "n = 11; #events = 11" = 3),font_size = 8)%>%
      add_header_above(c(" " = 1,"LRT = 10.34 on 3 df; p = 0.02" = 3, "LRT = 13.9 on 1 df; p = 0.05" =
                         "LRT = 10.34 on 1 df; p = 0.1" = 3),font_size = 8)%>%
      add_header_above(c(" " = 1,"Model 1" = 3, "Model 2" = 3, "Model 3" = 3)) %>%
     kable_styling(position = "center", latex_options="HOLD_position")

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")
```

```
#Q-Q Plots - Weibull, Log-lognormal, Log-logistic
#qq.surv function:
#Author: Jong Sung Kim, Date: 8/10/2004
# Edited by D. Leif Rustvold, Date: 6/7/2006

qq.surv <- function(time, status, pdgy = 0, distribution = "weibull", scale = 0, adjpb =
                      0.025, ...)
{
  ## Purpose: qqplot for distributions that satisfy a log-linear form
  ## for one sample.  It fits each sample with own intercept and slope
  ## (location and scale).
  ##-----------------------------------------------------------------
  ## Arguments
  ## =========
  ## time:   observed time
  ## status: censoring indicator
  ##
  ## Options
  ## =======
  ## pdgy:   Flag to generate for pedagogical purposes additional lines
  ##         incorporating the effect of how we treat censored
  ##         observations on the MLE's (equivalently estimated line).
  ##         pdgy=0 is the default, for no additional lines.
  ##         pdgy=1 generates additional lines.
  ## distribution:  Distribution for fit.
  ##         May take values "weibull", "loglogistic", or "lognormal".
  ##         The default is "weibull" distribution (exponential model with
  ##         scale=1).  Enter "loglogistic" to fit loglogistic distribution;
  ##         Enter "lognormal" to fit lognormal distribution.
  ## scale:  Scale parameter.  scale=0 is the default. This estimates
  ##         the scale. With distribution "weibull", scale=1 fits the
  ##         exponential model.
  ## adjpb:  Replaces the zero survival probability when the max is exact.
  ##         Or when the min is censored, it replaces the survival
  ##         probability by 1 - adjpb.  Default is 0.025.
  ##         This has nothing to do with the MLE line, but is solely for
  ##         plotting the point on the graph.
  ##-----------------------------------------------------------------
  ## Author: Jong Sung Kim, Date: 8/10/2004
  ## Edited by D. Leif Rustvold, Date: 6/7/2006
  d <- data.frame(time, status)
  # data frame
  d <- na.exclude(d)
  # Missing observations excluded
  d <- d[order(d$time),  ]
  # Rearranging the observed times into a nondecreasing order
  # Unordered times sometimes mess up QQ-plots.
  time <- d$time
  # sorted time
  status <- d$status
  # status corresponding to sorted time
  data <- Surv(time, status)
  # Surv object
```

```
t.c <- class(data)
if((!is.null(t.c)) && t.c == "Surv")
  data <- list(data)
t.s <- summary(survfit(Surv(time, status)~1, type = "kaplan-meier",
                       na.action = na.exclude))
survp <- t.s$surv
survtime <- t.s$time
rare <- F
# rare = T indicates that the smallest observation is censored
if(time[1] < survtime[1]) {
  #print("Smallest observation is censored!")
  survp <- c(1 - adjpb, survp)
  survtime <- c(time[1], survtime)
  rare <- T
}
############
############
xlabs <- ifelse(distribution == "weibull",
                "Standard Extreme Value Quantiles", ifelse(distribution ==
                                                             "loglogistic", "Standard Log-logistic Qua
                                                               distribution == "lognormal", "Standard
                                                               "")))
if(pdgy == 1) {
  ###############
  t.s.exactall <- summary(survfit(Surv(time, status >= 0)~1, type
                                  = "kaplan-meier", na.action = na.exclude))
  exactall.survp <- t.s.exactall$surv
  exactall.survtime <- t.s.exactall$time
  exactall.length <- length(exactall.survtime)
  exactall.survp[exactall.length] <- adjpb
  t.ss.exactall <- exactall.survp
  #quant.exactall <- qweibull(1 - t.ss.exactall, 1)
  quant.exactall <- switch(distribution,
                           weibull = qweibull(1 - t.ss.exactall, 1),
                           lognormal = qlnorm(1 - t.ss.exactall),
                           loglogistic = exp(logis((1 - t.ss.exactall))))
  exactall.sevq <- log(quant.exactall)
  # standard extreme value quantile
  exactall.logtime <- log(exactall.survtime)
  #print(data.frame(exactall.logtime, exactall.sevq))
  ###############
  ok <- status == 1
  t.s.exact <- summary(survfit(Surv(time[ok], status[ok])~1, type
                               = "kaplan-meier", na.action = na.exclude))
  exact.survp <- t.s.exact$surv
  exact.survtime <- t.s.exact$time
  exact.length <- length(exact.survtime)
  exact.survp[exact.length] <- adjpb
  t.ss.exact <- exact.survp
  #quant.exact <- qweibull(1 - t.ss.exact, 1)
  quant.exact <- switch(distribution,
                        weibull = qweibull(1 - t.ss.exact, 1),
                        lognormal = qlnorm(1 - t.ss.exact),
                        loglogistic = exp(qlogis(1 - t.ss.exact)))
  exact.sevq <- log(quant.exact)
  # standard extreme value quantile
  exact.logtime <- log(exact.survtime)
  #print(data.frame(exact.logtime, exact.sevq))
  ###############
  n <- length(time)
  t.ss <- rep(0, n)
  for(i in 1:n) {
    # This loop assigns probabilities to censored time points,
    # and takes care of tied observations as well
    idx <- time[i] >= survtime
    t.ss[i] <- min(survp[idx], na.rm = T)
  }
  #sevq <- log(qweibull(1 - t.ss, 1))
  sevq <- log(switch(distribution,
                     weibull = qweibull(1 - t.ss, 1),
                     lognormal = qlnorm(1 - t.ss),
                     loglogistic = exp(qlogis(1 - t.ss))))
  # standard extreme value quantile
  logtime <- log(time)
  #print(data.frame(logtime, sevq))
  ######## Multiple Plot starts ##########
  xrange <- range(c(exactall.sevq, exact.sevq, sevq))
  yrange <- range(c(exactall.logtime, exact.logtime, logtime))
  par(mar = c(5, 5, 2, 2))
  plot(sevq, logtime, type = "n", lty = 1, xlim = xrange, ylim
       = yrange, xlab = xlabs, ylab = "Ordered Log Time",
       ...)
  points(sevq[ok], logtime[ok], pch = 1)
  # exact points portion
  points(sevq[!ok], logtime[!ok], pch = "\255", font = 8)
  # censored points portion
  points(exactall.sevq, exactall.logtime, pch = 3, col = 6)
  # exactall
  exactallfit <- survreg(Surv(time, status >= 0) ~ 1, dist =
                           distribution, scale = scale)
  # treating censored as exac
  t
  abline(exactallfit$coef, exactallfit$scale, lty = 3, col = 6)
  points(exact.sevq, exact.logtime, pch = 5, col = 5)
  # exact points only
  exactonlyfit <- survreg(Surv(time[ok], status[ok]) ~ 1, dist
                          = distribution, scale = scale)
  # deleting censored
  abline(exactonlyfit$coef, exactonlyfit$scale, lty = 2, col = 5
  )
  fit <- survreg(Surv(time, status) ~ 1, dist = "weibull", scale
                 = scale)
  # censoring taken into account
  abline(fit$coef, fit$scale, lty = 1, col = 1)
}
else {
```

```
  n <- length(time)
  t.ss <- rep(0, n)
  for(i in 1:n) {
    # This loop assigns probabilities to censored time points,
    # and takes care of tied observations as well
    idx <- time[i] >= survtime
    t.ss[i] <- min(survp[idx], na.rm = T)
  }
  #sevq <- log(qweibull(1 - t.ss, 1))
  sevq <- log(switch(distribution,
                     weibull = qweibull(1 - t.ss, 1),
                     lognormal = qlnorm(1 - t.ss),
                     loglogistic = exp(qlogis(1 - t.ss))))
  # standard extreme value quantile
  logtime <- log(time)
  #print(data.frame(logtime, sevq))
  par(mar = c(5, 5, 2, 2))
  plot(sevq, logtime, type = "n", xlab = xlabs, ylab =
         "Ordered Log Time", ...)
  ok <- status == 1
  # exact status only
  points(sevq[ok], logtime[ok], pch = 1)
  # exact points only
  points(sevq[!ok], logtime[!ok], pch = "\255", font = 8)
  # censored points only
  fit <- survreg(Surv(time, status) ~ 1, dist = distribution,
                 scale = scale)
  # censoring taken into account
  abline(fit$coef, fit$scale, lty = 1, col = 1)
}
ymax <- max(logtime)
yrange <- diff(range(logtime))
yn <- ymax - yrange * seq(0, by = 0.05, length = 5)
if(pdgy == 1) {
  xmin <- min(c(sevq, exact.sevq, exactall.sevq))
  xrange <- diff(range(c(sevq, exact.sevq, exactall.sevq)))
}
else {
  xmin <- min(sevq)
  xrange <- diff(range(sevq))
}
x1 <- xmin + 0.05 * xrange
x2 <- xmin + 0.1 * xrange
x3 <- xmin + 0.15 * xrange
points(x1, yn[1], pch = "\255", font = 8)
text(x3, yn[1], "censored", adj = 0)
points(x1, yn[2], pch = 1)
text(x3, yn[2], "exact", adj = 0)
if(pdgy == 1) {
  lines(c(x1, x2), rep(yn[3], 2), lty = 1, col = 1, lwd = 3)
  text(x3, yn[3], "censoring taken into account", adj = 0)
  lines(c(x1, x2), rep(yn[4], 2), lty = 3, col = 6, lwd = 3)
  text(x3, yn[4], "treating censored as exact", adj = 0)
```

```
    lines(c(x1, x2), rep(yn[5], 2), lty = 2, col = 5, lwd = 3)
    text(x3, yn[5], "deleting censored", adj = 0)
  }
  on.exit()
  #paste("Q-Q plot for", distribution, "done")
}

months=df.new$Survival
status=df.new$Status
qq.surv(months, status, distribution = "weibull",
        adjpb=0,
        main="Q-Q plot - Weibull fit")
qq.surv(months, status, distribution = "lognormal",
        adjpb=0,
        main="Q-Q plot - Log-normal fit")
qq.surv(months, status, distribution = "loglogistic",
        adjpb=0,
        main="Q-Q plot - Log-logistic fit")


p= ggsurvplot(km.age,
              fun = "cumhaz",
              palette = c("darkcyan","darkgoldenrod3","darkorange3"),
              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("Age < 63","Age => 63")
)
p$plot + geom_vline(xintercept=26)+
  geom_vline(xintercept=46)

p = 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"))
p$plot + geom_vline(xintercept=26)+
  geom_vline(xintercept=46)

p = ggsurvplot(km.wms,
               fun = "cumhaz",
               palette = c("darkcyan","darkgoldenrod3","darkorange3"),
               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)")
p$plot + geom_vline(xintercept=26)+
  geom_vline(xintercept=46)


p = ggsurvplot(km.fshort,
               fun = "cumhaz",
               palette = c("darkcyan","darkgoldenrod3","darkorange3"),
               subtitle="Hazard, Stratified by Fractional 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)"
)
p$plot + geom_vline(xintercept=26)+
  geom_vline(xintercept=46)

#Cox-Snell residual analysis for overall model fit
LL=0.0
UL=26.0

#Subset the data based on the time region

months=df.new$Survival[df.new$Survival>=LL & df.new$Survival<=UL]
status=df.new$Status[df.new$Survival>=LL & df.new$Survival<=UL]
Age=df.new$Age[df.new$Survival>=LL & df.new$Survival<=UL]
P.Eff=df.new$P.Effusion[df.new$Survival>=LL & df.new$Survival<=UL]
W.MS=df.new$WMS[df.new$Survival>=LL & df.new$Survival<=UL]
F.Short=df.new$F.Shortening[df.new$Survival>=LL & df.new$Survival<=UL]

#Create initial model fit

cph.fit1=coxph(Surv(months,status)~Age+P.Eff+W.MS+F.Short,x=T)


#Reduce with StepAIC procedure

cph.fit2=stepAIC(cph.fit1,~.^2,direction="both",trace=FALSE)
mod1=cph.fit2

lrt=-2*cph.fit2$loglik[2]+2*cph.fit1$loglik[2]
varstartcount=10
varendcount=3
vars=varstartcount - varendcount
pval=1-pchisq(lrt,vars)

#Note: selects the reduced model (p-val>0.05)
```

```
LL=26
UL=46

#Subset the data based on the time region

months=df.new$Survival[df.new$Survival>LL & df.new$Survival<=UL]
status=df.new$Status[df.new$Survival>LL & df.new$Survival<=UL]
Age=df.new$Age[df.new$Survival>LL & df.new$Survival<=UL]
P.Eff=df.new$P.Effusion[df.new$Survival>LL & df.new$Survival<=UL]
W.MS=df.new$WMS[df.new$Survival>LL & df.new$Survival<=UL]
F.Short=df.new$F.Shortening[df.new$Survival>LL & df.new$Survival<=UL]

#Create initial model fit

cph.fit1=coxph(Surv(months,status)~Age+P.Eff+W.MS+F.Short,x=T)

#Reduce with StepAIC procedure

cph.fit2=stepAIC(cph.fit1,~.^2,direction="both",trace=FALSE)
mod2=cph.fit2

lrt=-2*cph.fit2$loglik[2]+2*cph.fit1$loglik[2]
varstartcount=10
varendcount=1
vars=varstartcount - varendcount
pval=1-pchisq(lrt,vars)

#Note: selects the reduced model (p-val>0.05)

#Subset the data based on the time region

#Subset the data based on the time region

months=df.new$Survival[df.new$Survival>UL]
status=df.new$Status[df.new$Survival>UL]
Age=df.new$Age[df.new$Survival>UL]
P.Eff=df.new$P.Effusion[df.new$Survival>UL]
W.MS=df.new$WMS[df.new$Survival>UL]
F.Short=df.new$F.Shortening[df.new$Survival>UL]

#Create initial model fit

cph.fit1=coxph(Surv(months,status)~Age+P.Eff+W.MS+F.Short,x=T)

#Reduce with StepAIC procedure

cph.fit2=stepAIC(cph.fit1,~.^2,direction="backward",trace=FALSE)
mod3=cph.fit2

status=df.new$Status[df.new$Survival>=0 & df.new$Survival<=26]
rc=abs(status - mod1$residuals)
km.rc = survfit(Surv(rc,status)~1)
summary.km.rc=summary(km.rc)
```

```r
rcu=summary.km.rc$time
surv.rc = summary.km.rc$surv
plot(rcu,-log(surv.rc),type="p",
     xlab="Cox-Snell residual rc",ylab="Cumulative hazard on rc",
     main="Model 1 - Cox-Snell residual model fit evaluation")
abline(a=0,b=1); abline(v=0); abline(h=0)

status=df.new$Status[df.new$Survival>26 & df.new$Survival<=46]
rc=abs(status - mod2$residuals)
km.rc = survfit(Surv(rc,status)~1)
summary.km.rc=summary(km.rc)
rcu=summary.km.rc$time
surv.rc = summary.km.rc$surv
plot(rcu,-log(surv.rc),type="p",
     xlab="Cox-Snell residual rc",ylab="Cumulative hazard on rc",
     main="Model 2 - Cox-Snell residual model fit evaluation")
abline(a=0,b=1); abline(v=0); abline(h=0)

status=df.new$Status[df.new$Survival>46]
rc=abs(status - mod3$residuals)
km.rc = survfit(Surv(rc,status)~1)
summary.km.rc=summary(km.rc)
rcu=summary.km.rc$time
surv.rc = summary.km.rc$surv
plot(rcu,-log(surv.rc),type="p",
     xlab="Cox-Snell residual rc",ylab="Cumulative hazard on rc",
     main="Model 3 - Cox-Snell residual model fit evaluation")
abline(a=0,b=1); abline(v=0); abline(h=0)

#Schoenfeld residuals; test for constant coefficients
test.ph <- cox.zph(mod1)
ggcoxzph(test.ph, main="Schoenfeld Residuals - Const Coeff Evaluation for Model 1",font.y=8)



test.ph <- cox.zph(mod2)
ggcoxzph(test.ph, main="Schoenfeld Residuals - Const Coeff Evaluation for Model 2",font.y=8)

test.ph <- cox.zph(mod3)
ggcoxzph(test.ph, main="Schoenfeld Residuals - Const Coeff Evaluation for Model 3",font.y=8)
df=data.frame(df.new$P.Effusion,df.new$Age,df.new$F.Shortening,df.new$EPSS,df.new$LVDD,df.new$WMS)
pairs(df)
```

Table 12: 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		