Analysis of Survival Post-Myocardial Infarction

Samantha Maillie

Abstract

Through the use of survival analysis this study aims to develop an understanding of a patient's survival after the experiencing a myocardial infarction. The study uses R software to develop a model that best incorporates relevant covariates into predicting the survival of a patient. One covariate that proved to be particularly useful as a covariate is a patient's wall-motion score index (WMSI). The WMSI is a measure of the heart's ability to function properly (9). The investigation of how covariates are related to a patient's survival is essential to developing ways to extend a patient's lifetime.

Introduction

Myocardial infarction, (MI), informally known as heart attack, patient data was collected in 1989 to observe their survival after experiencing an MI (1). The data set contains the survival time in months and a binary variable that indicates whether they were still alive at the end of the corresponding survival interval. There were also a number of additional variables that could be evaluated for their usefulness as covariates.

Omitted Variables

There was some data that was collected that was known to provide no added value to the model before any analyses were conducted. The patients' names are redacted so this column was taken out. The group and variable referred to as "mult" were described as meaningless in the documentation, (1). The wall-motion score and the wall-motion score index (WMSI) contain the same information. It was advised to use WMSI so the wall-motion scores were also removed. The data also contained a binary variable indicating if a patient passed away within the first year or not. Since this information is contained within the survival times it was also an omitted variable.

Categorical Classification of Covariates

The variables that were considered as covariates consisted of age, meaning the patients age at the time of the MI, pericardial effusion, fractional shortening, E-point septal separation, (EPSS), left ventricular end-diastolic dimension, (LVDD), and the wall-motion score index (WMSI). All of the variables, except pericardial effusion, were in a continuous format. In order to best convert them to a categorical format information not provided within the documentation was required.

Based on the American Heart Association's publications being over the age of 65 years old is considered a risk factor for having an MI (2). A patient under years old would be considered a young patient. While there was a young patient at the age of 35 the majority of patients were 50 year old or older, with a first quartile equal to 57 years old. Therefore, the data was split into young patients and old patients with the cut-off being 65 years old based on the American Heart Association's provided information, (2).

Fractional shortening is by definition a way to measure a heart's muscular contractility, (3). The collected measurements can be interpreted as follows: if the fractional shortening is less than 28% there is reason to believe that the efficiency of the heart ejecting blood is impaired (3). Therefore, this will be split into patients that shorten by at least 28% and those that fail to do so.

EPSS is a way of assessing the heart's ejection fraction or in other words the percentage of blood exiting the heart at each contraction (4). A defining feature of systolic heart failure is a lower ejection fraction (5). An ejection fraction (EF) less than 30% is considered to be a problem (5). There is a negative linear relationship between EF and EPSS (5). An EPSS of 7mm or above corresponds to an EF of 30% or less (5). Therefore the data is split into patients with an EPSS less than 7mm, a score indicating a good EF, and values greater than or equal to 7mm.

It was noted in the documentation for the data set that 'large hearts tend to be sick hearts' (1). An average range for LVDD, a way of measuring the heart, is 47.86mm +/ 4.3mm (6). The dataset documentation failed to state that the collected data was measured in cm however, it appears clear that this is the case based on the data's range of values and mean. The mean is 4.765 with a range of 2.32 to 6.78. Assuming the measurements are in cm this data makes sense given the documented baseline information. Therefore the average heart should be expected to have an LVDD between 4.74cm and 4.826cm (6). Originally the measurements were split into small, normal and large. It was thought that given this particular application the small and normal categories would later become combined so that it was an indicator of whether the heart is large or not. The main analyses did not support that the small category was contributing significant information to this model so it was decided that it would be most beneficial to go back and just split LVDD measurements into large and not large.

The wall-motion score index is an average of scores 16 different segments within the heart, (7). The main take away is that a normal score is between 1 and 1.3 (8,9). A score between 1.4 and 1.9 is indicative of potential problems however; the patient may not be experiencing symptoms yet, (8,9). A score between 2 and 2.1 is when complications start and anything 2.2 and above is very problematic (8,9). Given the scores in this data have a 3rd quartile equal to 1.51 it was decided to group them into three groups 1.1 to 1.3 for a normal WMSI, 1.4 to 1.9 as an WMSI of mild concern and any score greater than or equal to 2 as a problematic score. The sample

size wouldn't be large enough to warrant scores between 2 and 2.1 to have their own factor level.

Objectives

The primary aim of this study was to find a suitable predictive model for a patient's survival time following an MI. A secondary goal of the study is to develop a general understanding of the determinants involved in how long a patient will survive. The more that is understood about how to predict survival time the more doctors will be able to future advise patients on how to modify their lifestyles to decrease the risk of an MI. It is relevant to note at this point that the data was collected in 1989. However, based on the nature of the variables collected in this study it is safe to assume that they are still relevant today in monitoring heart health.

Preliminary Analysis

Overview
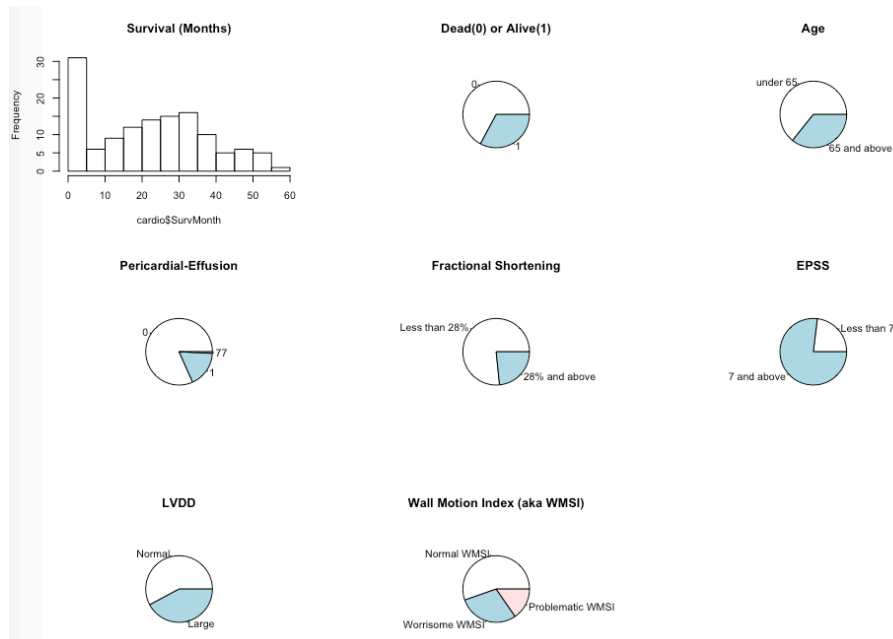
Once the variables were converted from continuous to factor data histograms and pie charts were examined. The survival months showed a high frequency of patients passing away within the first year, this is unsurprising given the original purpose of this data set's collection was to attempt to predict which patients would live past a year. After the first year the survival month's approximately distributed normally. The pie charts show within this study the number of patients that have passed out number those that are still alive. Other than that they are mostly used to make a quick overview of the collected data and ensure there is nothing unusual going on that must be dealt with.

Product Estimate Fit without Covariates
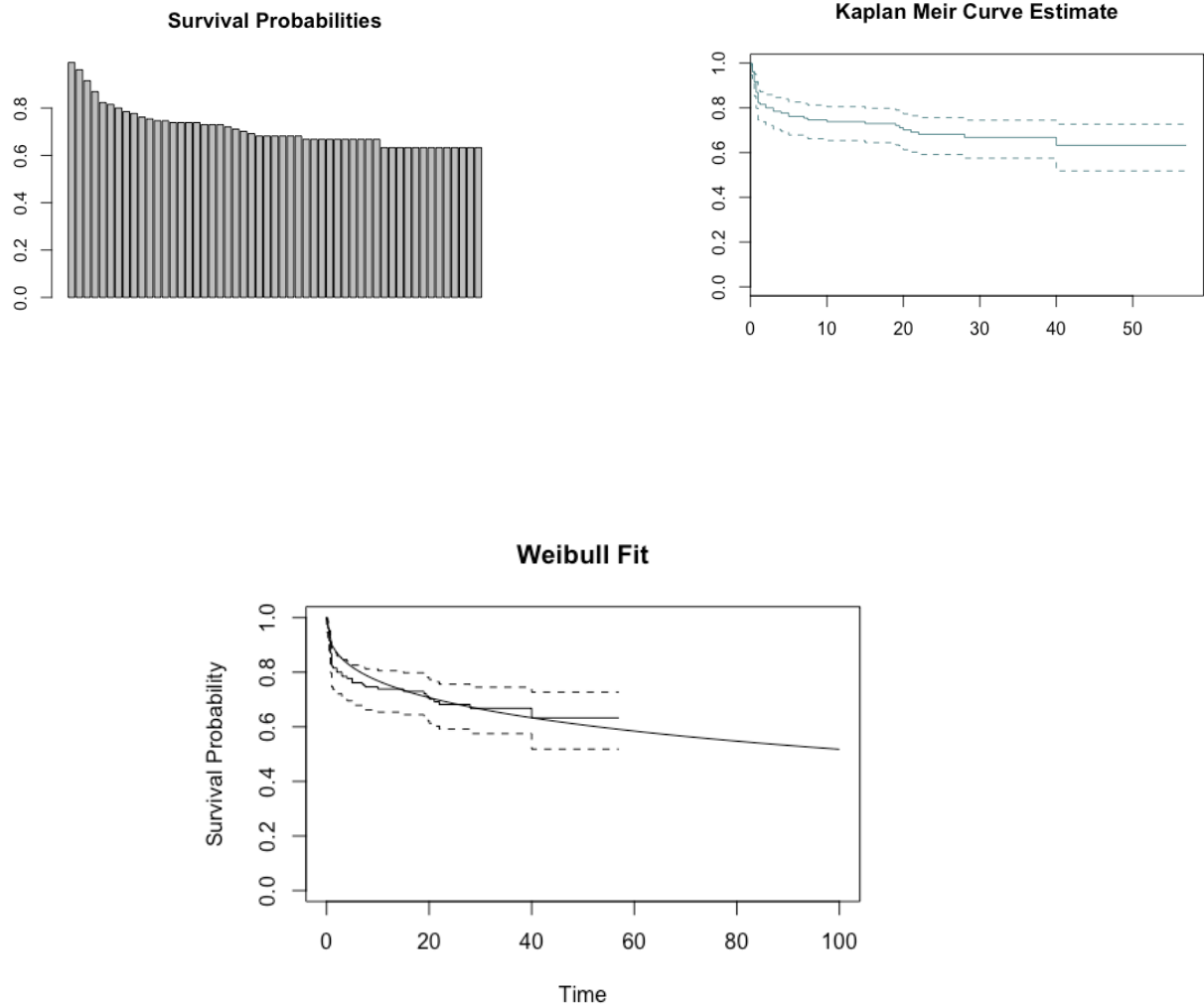
A model with no covariates using the Kaplan-Meir produce estimate method via the survival package's survfit() function. No median is produced because the median occurs when the survival probability reaches .5 and this data never gets lower than .633. The data suggests that it may be modeled well by a Weibull fit. Utilizing Dirk F. Moore's Weibull survival function a model is fit to the data and maximum likelihood estimates for the Weibull distribution parameters are extracted to make a smoothed curve. The Kaplan-Meir product estimate curve and the Weibull estimate curve match each other and the data very well. Suggesting a Weibull model should be fit to the data for further analysis.

Factor Level Comparisons

A general investigation into significant survival differences between each individual covariate's factor levels was conducted to see which covariates might add beneficial information to the model. There is a function called survdiff() within the survival package in R that allows for a quick computation of a stratified log rank (SLR) test that compares its test statistic to a chi-square with one degree of freedom. The log rank tests assume that the difference is in location as opposed to location and scale. This test along with the Kaplan-Meir estimate survival curves corresponding with each factor level are utilized in this study to determine which covariates warrant further investigation. The null hypothesis for these log rank tests is that there is no difference in survival time between the factor levels. The alternative hypothesis is that there is at least one factor level's survival time that is different.
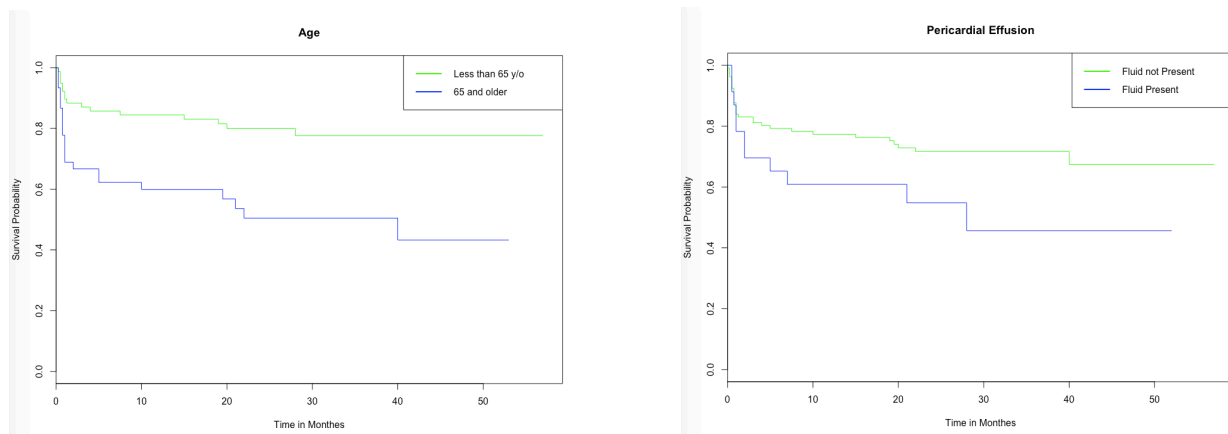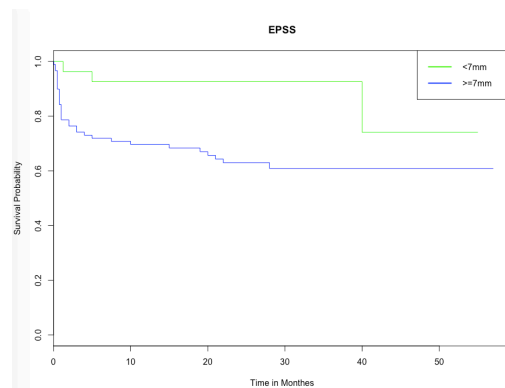

The two factor levels for age are less than 65 and at least 65 years old. The SLR test produces a p-value of .00128 which supports rejecting the null at alpha level .01. The graph shows a clear difference with older patients passing away at a faster rate. Therefore, there is evidence to support that there is a statistically significant difference in survival between the two patient age groups.

The two factor levels for pericardial effusion (fluid around the heart) are pericardial effusion is present or it is not.. The SLR test produces a p-value of .0659 which supports rejecting the null at alpha level .1. However, there would not be enough evidence to reject at an alpha of .05 or .01. The Kaplan-Meir curves appear very similar in the early months and then begin to diverge with patients where fluid is present around the heart passing away at an accelerated rate. Based on the p-value and the similarities this covariate may prove to be of less use when the other covariates are already being considered.
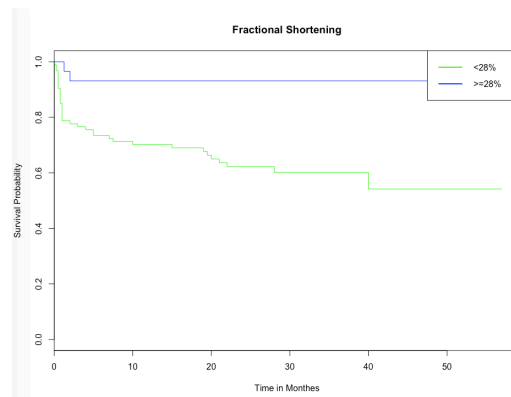
The two factor levels for fractional shortening are patients whose hearts shorten at least 28% and those that do not. The SLR test produces a p-value of .00173 which supports rejecting the null at alpha level .01. The graph shows a clear difference with patients failing to shorten by at least 28% passing away at a faster rate. The patients shortening at least 28% have a very high probability of survival based on this graph. It is important to note that this group has only 29 patients compared to 94 in the other. There is evidence to support that there is a statistically significant difference in survival between the two patient fractional shortening groups.

The two factor levels for EPSS for patients: less than 7mm, ideal, and greater than or equal to 7mm, considered problematic. The SLR test produces a p-value of .0125 which supports rejecting the null at alpha level .05. The graph shows a clear difference between patient groups. The patients in the group corresponding to 7mm or greater are passing away at a faster rate than the group in the ideal EPSS range. There is evidence to support that there is a statistically significant difference in survival between the two patient groups.
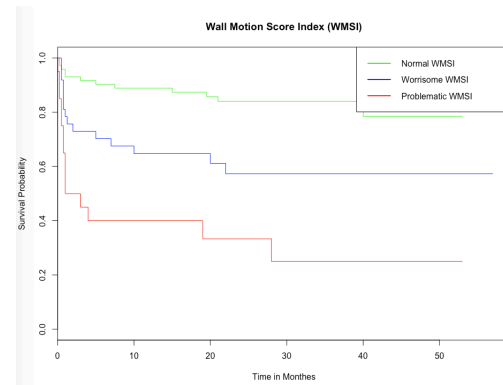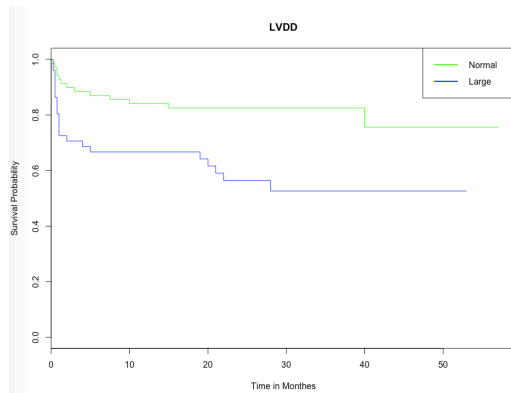
The LVDD groups were divided into patients with a heart that is considered large and those without a large heart. The SLR test produces a p-value of .00262 which supports rejecting the null at alpha level .01. The graph shows a clear difference between patient groups. The patients with what is considered to be an abnormally large heart are passing away at a faster rate. There is evidence to support that there is a statistically significant difference in survival between the two patient groups.

The three factor levels for WMSI are normal, worrisome, and problematic. The SLR test produces a p-value of 2.97e-7 which supports rejecting the null at alpha level .001. Out of all of the covariates this one is perhaps most indicative of a clear relationship between factor levels and survival. As the score gets worse the survival significantly drops. The first two levels never make it to a probability of .5 and therefore do not have a median. The problematic WMSI score's survival probability drops so rapidly that its median is just 2 months. It appears to be clear that this will be a beneficial covariate for the model. This is not surprising since the WMSI is a way of measuring the heart's ability to function properly.

Analysis

Survival Time Descriptive Statistics

| Min	1st Q	Median	Mean	3rd Q	Max
.030	7.875	23.5	22.18	33.00	57.00

Model Selection: Weibull Attempt

Methods

Based on preliminary analyses it was deduced that a Weibull model would be the best fit for the data. The Weibull function can be fit to survival data using the survreg() function which bases its calculations on the accelerated failure time (AFT) model. The first step in searching for the best model was using stepwise selection based on AIC. The method decided upon fractional shortening, LVDD, EPSS, age, and WMSI as covariates. The corresponding AIC is 282.26.

The weibull regression function will produce event time ratios and also runs a goodness of fit test. Many of the event time ratios(ETR's) contained 1 in their corresponding 95% confidence interval indicating a lack of significant difference between factor levels. The stepwise selected model was run through this. Based on ETR's EPSS, LVDD, and fractional shortening were not significantly different amongst their factor levels. Also worth noting is the ETR confidence interval for fractional shortening was (.41, 241).

Only one predictor should be taken out at a time so a manual stepwise selection was used to remove covariates that were not contributing to the model. The model's AIC was calculated for three models, each with one of the insignificant predictors removed. The removal of fractional shortening resulted in the lowest penalty, in other words the smallest increase, in AIC for the model. Combined with its unevenly distributed sample size and therefore its very wide confidence interval fractional shortening was removed from the model. This left age, EPSS, LVDD and WMSI as predictors.

This process was repeated again and EPSS was selected to be removed from the model. At this stage, LVDD still contained 1 within its interval, however this was just barely the case, the goodness of fit test had a very low p-value and AIC was 294.7191. If LVDD was removed the AIC would jump up to 337.7526. The model was selected to utilize age, LVDD, and WMSI as covariates. Despite the model initially appearing as though it would fit well as a Weibull model, the diagnostics would raise concern in regards to the validity of the model.

Model Diagnostics
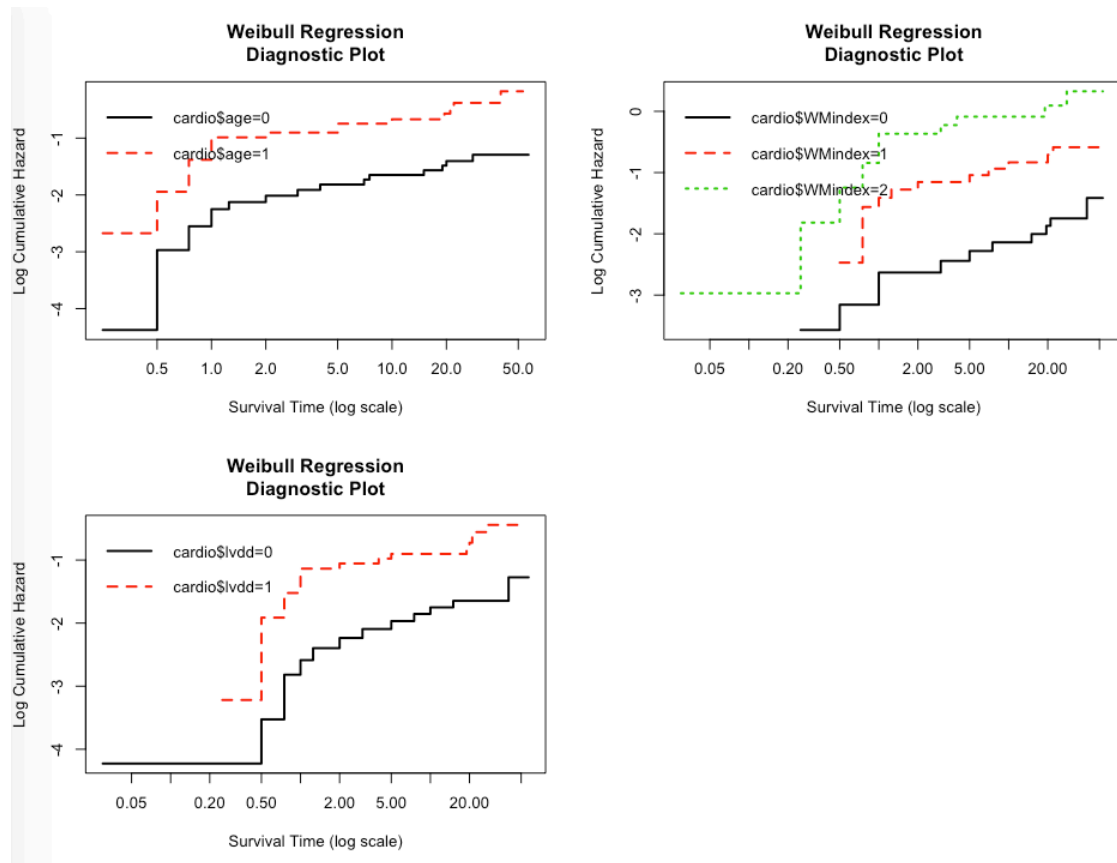
If the weibull model was a good fit for this data than we would see this lines as linear and parallel to one another. While they do appear to maintain the same shape they do not appear to be linear at all. Suggesting that the assumptions of the weibull model have been violated.

Fitted Model with Weibull using AFT

formula = Surv(cardio$SurvMonth, cardio$Status0Dead) ~ cardio$age + cardio$lvdd + cardio$WMindex

	Value	Std. Error	z	p
(Intercept)	8.951	1.223	7.32	2.45e-13
cardio$age	-2.173	0.819	-2.65	7.95e-03
cardio$lvdd	-1.650	0.844	-1.96	5.05e-02
cardio$WMindex1	-2.677	0.944	-2.84	4.55e-03
cardio$WMindex2	-3.466	1.116	-3.10	1.90e-03

Log(scale)   0.768   0.153   5.02   5.26e-07

Scale= 2.16
Weibull distribution
Loglik(model)= -141.4   Loglik(intercept only)= -157.3
        Chisq= 31.81 on 4 degrees of freedom, p= 2.1e-06
Number of Newton-Raphson Iterations: 6
n=115 (15 observations deleted due to missingness)

The goodness of fit has a p-value well below indicating that this model is significant. All of the predictors are also appearing to be significant. The LVDD does contain 1 within its ETR indicating that differentiating between levels of LVDD is not necessarily benefiting the model. Overall the diagnostic plots for this model are not horrible but they are not great. This is to be expected when using real life data, parametric models rarely provide sufficient models.

Model Selection: Cox Proportional Hazards Model

A similar method of comparing AICs was used to determine which predictors should remain in this model. The coxph() function fits survival data to the Cox proportional hazrds model. This method decided on age, fractional shortening, EPSS, LVDD and WMSI as covariates. The AIC is 259.9925, BIC is 268.777 and the goodness of fit test, via a likelihood ratio test, p-value is 5.15e-06. Relative to the previous search for a model these values are decent.

A new model was fit where age was adjusted to become a stratified variable. There wasn't any reason for this decision other than it made sense to at least try it. The resulting model had an AIC of 216.5862 and BIC of 223.9148 both of which were quite a bit lower than when age was not stratified. The downfall being that the goodness of fit p-value became .00021. In both of these models multiple predictors had insignificant p-values.

Through a systematic procedure of evaluating the effects on the model of removing one variable at a time it was discovered that using stratified age, EPSS and WMSI is the best possible model. The AIC is 222.9803, BIC is 227.4698, the goodness of fit p-value is 6.8e-05 and now all of the predictors are significant at the .1 alpha level. The WMSI indicators are both significant at the .01 alpha level.

Model Diagnostics
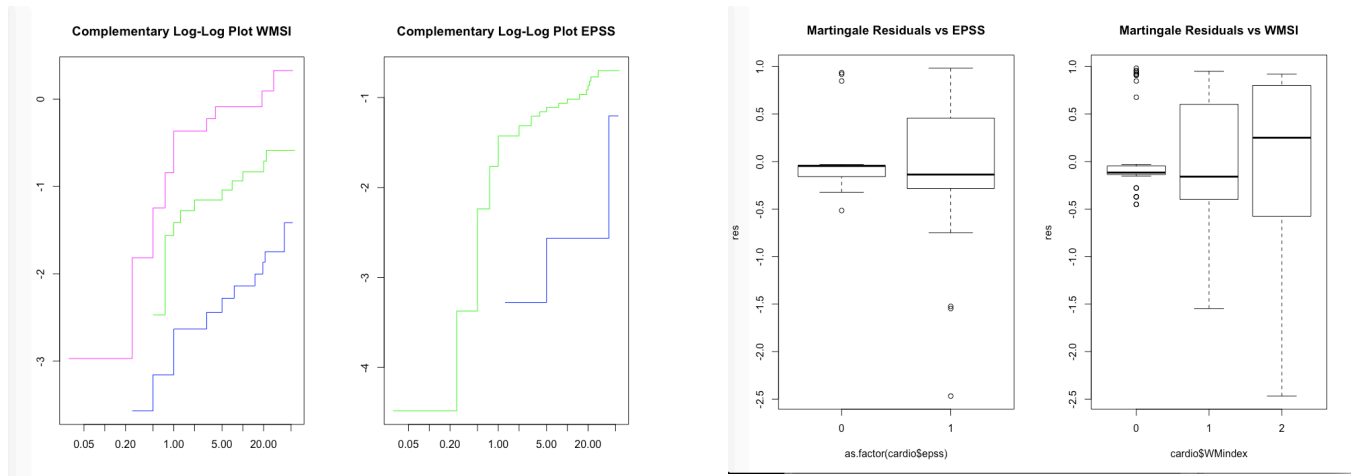
The complementary log-log plot can be utilized to test if the assumptions of the Cox proportional hazard models have been violated. If they are all satisfied these lines will appear parallel to each other. They are not perfectly parallel here but they appear to be approximately parallel to each other. There is a little concern in the middle WMSI line. The EPSS line in blue is clearly suffering from a small sample size, as evident by the very evident step plot appearance of the line. While these are not perfect, this model seems to be doing better than the Weibull model did relative to violating its assumptions.

The residuals are also somewhat problematic. They are clearly not evenly distributed in either of the two cases. This again may be something that could become better, or at least less drastic with an increase in sample size. Out of the two models the Cox proportional hazard model is the best pick because it will be more robust than a parametric model like the weibull model to assumption violations since it is only semi-parametric. Based on all of the analyses this data set could surely benefit from a higher sample size hopefully resulting in factor level sample sizes being closer to equal.

**Fitted Model using Cox Proportional Hazards Model**

formula = Surv(cardio$SurvMonth, cardio$Status0Dead) ~ strata(cardio$age) + cardio$epss + cardio$WMindex

	coef	exp(coef)	se(coef)	z	p
cardio$epss	1.084	2.957	0.610	1.78	0.07564
cardio$WMindex1	1.234	3.434	0.432	2.85	0.00433
cardio$WMindex2	1.701	5.480	0.467	3.65	0.00027

Likelihood ratio test=21.9 on 3 df, p=6.8e-05

n= 110, number of events= 33

   (20 observations deleted due to missingness)
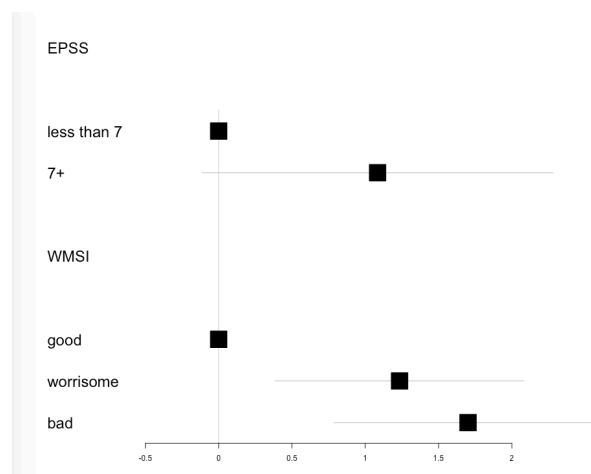
This forest plot can be used to better visualize the above table. This plot confirms what has been discovered before. EPSS is not quite significantly different between levels; it is possible a significant difference would be observed with a larger sample size. WMSI shows a clear significant difference between a good score and a worrisome or bad score. Worrisome and bad scores are more similar to each other.

Conclusions

The Cox proportional hazard with stratified age, WMSI and EPSS as covariates is the best model for this data. A large sample size would benefit future research. WMSI is a measure of the heart's ability to function and EPSS is a measure of how much blood the heart ejects at each pump. It is not surprising that these two are powerful covariates. The importance of understanding what covariates contribute to a patient's survival post-myocardial infarction cannot be overstated.

Acknowledgements

References

1. Salzber, Steven. Kinney, Evlin. (1989). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml/datasets/Echocardiogram]. Irvine, CA: University of California, School of Information and Computer Science

2. Understand Your Risks to Prevent a Heart Attack. (n.d.). Retrieved December 10, 2017, from http://www.heart.org/HEARTORG/Conditions/HeartAttack/UnderstandYourRiskstoPreventaHeartAttack/Understand-Your-Risks-to-Prevent-a-Heart-Attack_UCM_002040_Article.jsp#

3. fractional shortening. (n.d.) *Medical Dictionary*. (2009). Retrieved December 10 2017 from https://medical-dictionary.thefreedictionary.com/fractional+shortening

4. Ejection fraction: What does it measure? (2016, February 11). Retrieved December 10, 2017, from https://www.mayoclinic.org/ejection-fraction/expert-answers/faq-20058286

5. E-POINT SEPTAL SEPARATION. (2014, September 21). Retrieved December 10, 2017, from http://www.scancrit.com/2014/09/21/e-point-septal-separation/

6. Missiri, A. M., Meniawy, K. A., Sakr, S. A., & Mohamed, A. S. (2016). Normal reference values of echocardiographic measurements in young Egyptian adults. *The Egyptian Heart Journal,68*(4), 209-215. doi:10.1016/j.ehj.2016.01.002

7. Sharif, D., Sharif-Rasslan, A., Odeh, M., Shahla, C., Khalil, A., & Rosenschien, U. (2014). Effects of Left Ventricular Wall Motion Abnormality on Global and Regional Diastolic Function of the Left and Right Ventricles at Rest and After Stress. *Cardiology Research,5*(6), 176-182. doi:10.14740/cr366w

8. Ejection Fraction. (n.d.). Retrieved December 10, 2017, from https://my.clevelandclinic.org/health/articles/ejection-fraction

9. Lebeau, R., Serri, K., Morice, M., Hovasse, T., Unterseeh, T., Piéchaud, J., & Garot, J. (2012). Assessment of left ventricular ejection fraction using the wall motion score index in cardiac magnetic resonance imaging. *Archives of Cardiovascular Diseases,105*(2), 91-98. doi:10.1016/j.acvd.2012.01.002