Colin Busby
BSTA 662
3/10/23

Introduction

The data used in this project was collected from 500 patients who participated in the Worcester Heart Attack Study by Dr. Robert J. Goldberg of the Department of Cardiology at the University of Massachusetts Medical School. While there are 22 variables in the data set, only those used in this project will be mentioned. Of significant note are the variables *fstat* (vital status at last follow-up, 0 = Alive, 1 = Dead) and *lenfol* (total length of follow-up, Days between date of last follow-up and hospital admission date), which are used to create the statistical analysis below, the goal of which is to describe the effects of significant factors/covariates associated with trends over time in the incidence (death) and survival rates following hospital admission for heart attack (myocardial infarction). The final SAS code used in this project is found in the appendix at the end of this document.

A) Conduct model selection and summarize the selection procedure. Check the proportional hazard (PH) assumption of your final model.

The selection procedure consists of using the arguments `ties=exact selection=backwards` into the model section of the **`PROC PHREG`** code block. This model selection gives us the factors *age* (age in years at hospital admission), *gender* (0 = Male, 1 = Female), *hr* (initial heart rate in bpm), *diasbp* (initial diastolic blood pressure in mmHg), *bmi* (body mass index, kg/m^2), *sho* (cardiogenic shock, 0 = No, 1 = Yes), *year* (cohort year, 1 = 1997, 2 = 1999, 3 = 2001), and *chf* (congestive heart complications, 0 = No, 1 = Yes), but checking this reduced model using the argument `assess ph/resample` reveals that only *sho* fails the proportional hazard assumption with a p-value of 0.0070 (Table 1) and an unusual-looking proportional hazard assumption graph (Figure 1). All other covariates show p-values well above the 0.05 significance level along with normal proportional hazard assumption graphs, meaning they will be used in my final model.

Supremum Test for Proportionals Hazards Assumption				
Variable	Maximum Absolute Value	Replications	Seed	Pr > MaxAbsVal
age	0.7688	1000	2115921505	0.5900
gender	0.6016	1000	2115921505	0.8380
hr	1.0157	1000	2115921505	0.2220
diasbp	0.9338	1000	2115921505	0.3630
bmi	1.1070	1000	2115921505	0.3610
sho	1.8111	1000	2115921505	0.0070
year	1.1912	1000	2115921505	0.1310
chf	0.8622	1000	2115921505	0.4360

Table 1 – Reduced model PH Assumption p-values

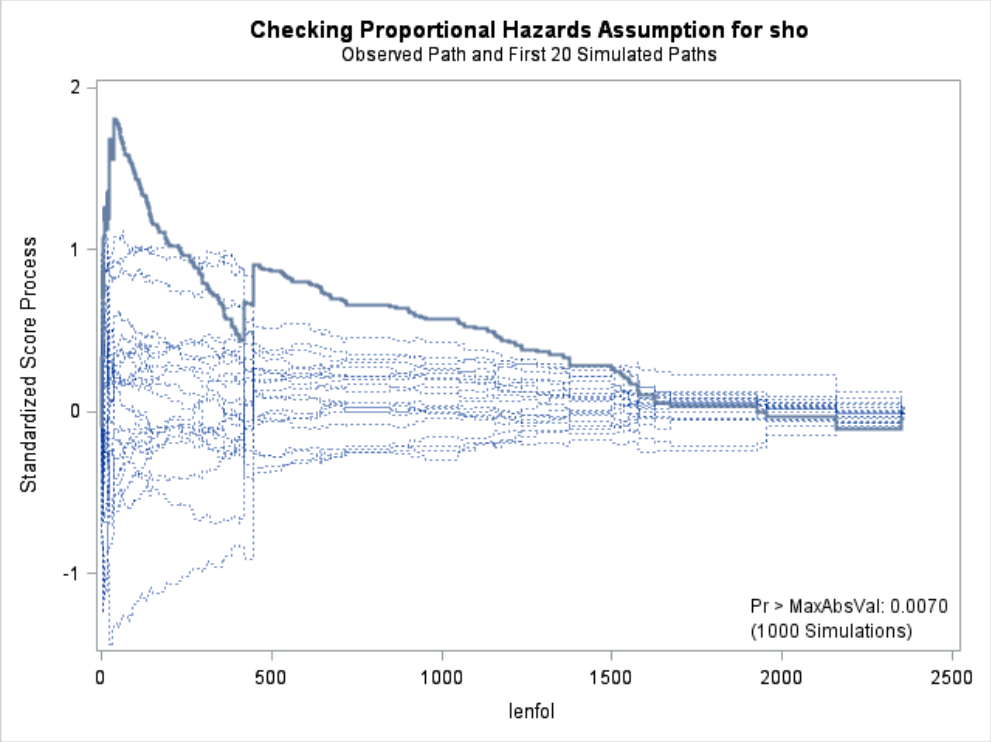

Figure 1 – Reduced model PH Assumption graph of *sho*

B) If the PH assumption does not hold for some factors in your final model. Revise/fix your final model. Describe the procedure of your model checking.

As only the factor *sho* failed proportionality, I attempted to fix it by stratifying *sho* in the `PROC PHREG` statement. This resulted in a final model with stratification of the sole non-proportional covariate. While this results in the sacrifice of some information, it allows for the analysis of different subpopulations; in this case, the different subpopulations are those with and without cardiogenic shock.

After stratifying *sho*, testing the proportional hazards assumption shows that all factors satisfy the assumption, as shown in Table 2.

Supremum Test for Proportionals Hazards Assumption				
Variable	Maximum Absolute Value	Replications	Seed	Pr > MaxAbsVal
age	0.6899	1000	1255418724	0.7120
gender	0.6717	1000	1255418724	0.7250
hr	0.9209	1000	1255418724	0.3020
diasbp	1.0912	1000	1255418724	0.1990
bmi	1.2090	1000	1255418724	0.2730
year	1.2285	1000	1255418724	0.1240
chf	0.9882	1000	1255418724	0.2850

Table 2 - All factors meet the PH assumption

This stratification shows that patients who have experienced cardiogenic shock have a much lower survival time compared to patients who have not experienced it, as shown in Figure 2.

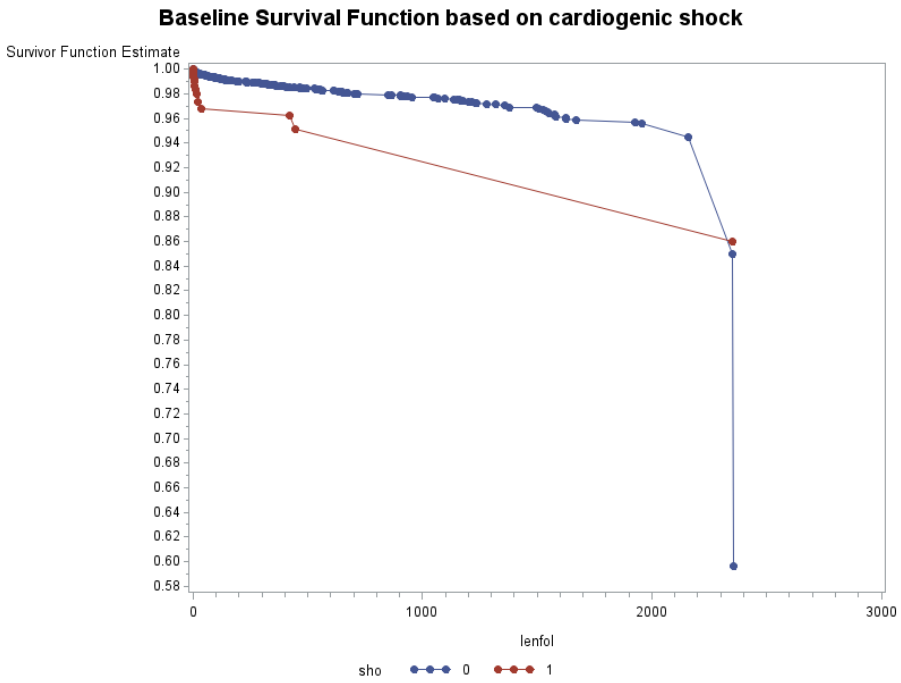

Figure 2 - Baseline survival function of final model

C) Write down the final fitted regression equation and explain the effects of (significant) covariates to the survivorship.

The final fitted regression equation (Figure 3) is shown below, with $h_{0(i)}(y)$ showing the baseline hazard depending on a patient's cardiogenic shock status (Figure 4):

$$h(y|x) = h_{0(i)}(y) * \exp((0.04818 * age) + (-0.30268 * gender) + (0.01182 * hr) + (-0.01161 * disabp) + (-0.05053 * bmi) + (0.24809 * year) + (0.69842 * chf))$$

Figure 3 - Final fitted regression equation

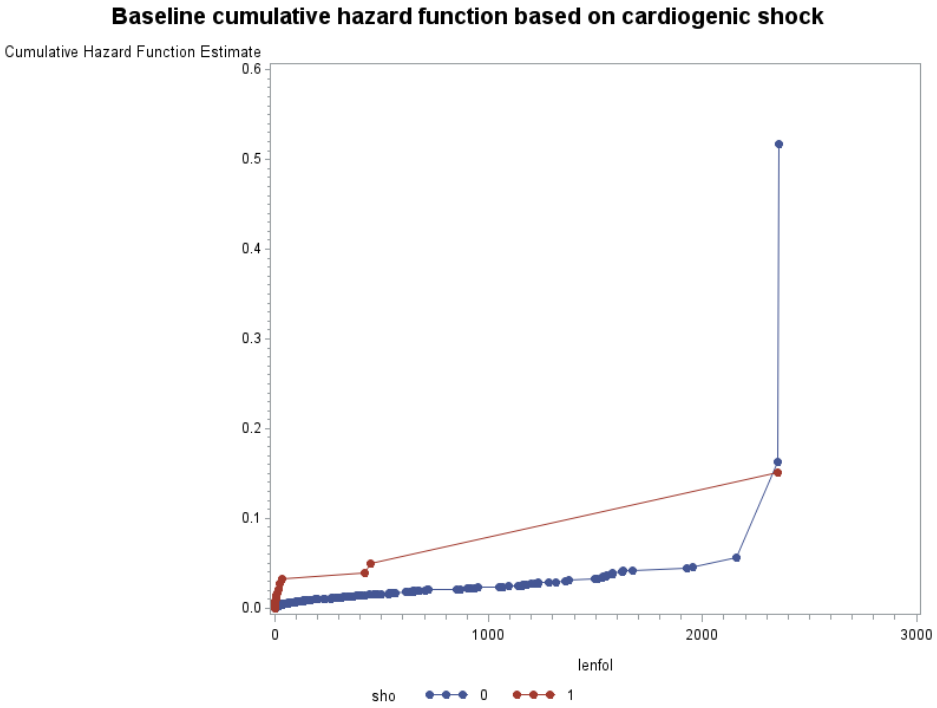

Figure 4 - Baseline cumulative hazard function

The hazard ratio, also known as relative risk, and the parameter estimates used in the final fitted regression equation, are found in Table 3 below:

Analysis of Maximum Likelihood Estimates						
Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
age	1	0.04818	0.00661	53.1972	<.0001	1.049
gender	1	-0.30268	0.14516	4.3475	0.0371	0.739
hr	1	0.01182	0.00290	16.6125	<.0001	1.012
diasbp	1	-0.01161	0.00351	10.9210	0.0010	0.988
bmi	1	-0.05053	0.01665	9.2060	0.0024	0.951
sho	0	0	.	.	.	.
year	1	0.24809	0.09998	6.1571	0.0131	1.282
chf	1	0.69842	0.14989	21.7112	<.0001	2.011

Table 3 – Hazard ratios and parameter estimates

Controlling for the other covariates, the effect of *age* has a relative risk of 1.049, indicating that for a patient of age *a*, every year increase in age increases the risk of death by a factor of 1.049 compared to the previous year. In other words, the risk of death for a patient who is 1 year older is 104.9% that of a patient of age *a*.

Controlling for the other covariates, the effect of *gender* (male=0, female=1) has a relative risk of 0.739, indicating that a female patient has a decreased risk of death by a factor of 0.739 compared to male patients.

Controlling for the other covariates, the effect of *hr* (heart rate) has a relative risk of 1.012, indicating that a patient with a higher number of beats per minute for initial heart rate has an increased risk of death compared to a patient with a slower heart rate.

Controlling for the other covariates, the effect of *diasbp* has a relative risk of 0.988, indicating that the higher a patient's initial diastolic blood pressure (mmHg), the lower their risk of death compared to a patient with a lower initial diastolic blood pressure.

Controlling for the other covariates, the effect of *bmi* has a relative risk of 0.951, indicating that the higher a patient's body mass index, the lower their risk of death.

Controlling for the other covariates, the effect of *year* has a relative risk of 1.282, indicating that for a patient of the 1999 cohort, their risk of death is 128.2% that of a patient from the 1997 patient cohort, and that the risk of death for a patient of the 2001 cohort is 128.2% of that of the 1999 cohort.

Controlling for the other covariates, the effect of *chf* has a relative risk of 2.011, indicating that for a patient with congestive heart complications, their risk of death is 201.1% of that of a patient without congestive heart complications.

D) Draw the predicted survival function for a 40 year old woman from the 1997 cohort with 70 bpm HR, 80 mmHg diastolic bp, 24 kg/m2 bmi, cardiogenic shock, and congestive heart failure. What is her 5th percentile of survival time?

For a woman with the stated attributes, the estimated survival function is shown below in Figure 5, which plots the survivor function estimate over the *lenfol* factor.

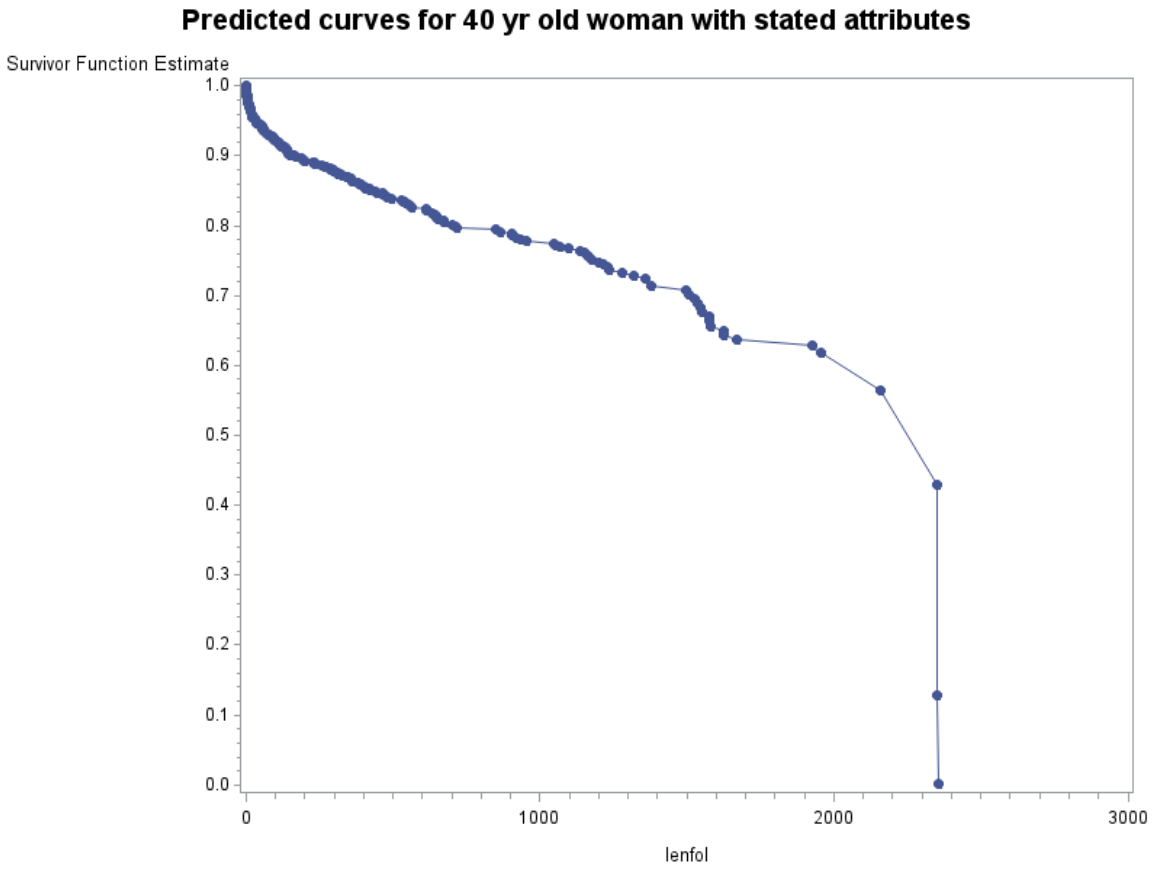

Figure 5 – Predicted Survival Function

The 5th percentile of survival time for the woman is the 33rd day of her follow-up, where her survival estimate falls below 0.95 for the first time (Table 4).

Predicted curves for 40 yr old woman with stated attributes

Obs	age	gender	hr	diasbp	bmi	sho	year	chf	lenfol	s	lcl	ucl	H
1	40	1	70	80	24	1	1	1	31	0.95371	0.91812	0.99067	0.047400
2	40	1	70	80	24	1	1	1	32	0.95176	0.91480	0.99022	0.049439
3	40	1	70	80	24	1	1	1	33	0.94884	0.90981	0.98954	0.052517
4	40	1	70	80	24	1	1	1	34	0.94786	0.90813	0.98932	0.053553

Table 4 – Selected observations showing 5th percentile cutoff

Appendix – SAS code

```
PROC IMPORT OUT= WORK.whas500
            DATAFILE= "C:\Users\Home\Documents\STAT 631\whas500.csv"
            DBMS=CSV REPLACE;
     GETNAMES=YES;
     DATAROW=2;
RUN;

ods graphics on;

/*Full model with all covariates*/

proc phreg data=whas500;
     title;
     model lenfol*fstat(0)=age gender hr sysbp diasbp bmi cvd afb
sho year chf av3 miord mitype los/ties=exact selection=backwards;
run;

proc phreg data=whas500;
     title;
     model lenfol*fstat(0)= age gender hr diasbp bmi sho year
chf/ties=exact;
     assess ph/resample;
run;

proc phreg data=whas500;
title "Final model with sho (Cardiogenic Shock) strata";
     model lenfol*fstat(0)=age gender hr sysbp diasbp bmi cvd afb
sho year chf av3 miord mitype los/ties=exact selection=backwards;
     strata sho;
     assess ph/resample;
run;

/*Baseline survival function*/
data null;
     title;
     input age gender hr diasbp bmi sho year chf;
     cards;
0 0 0 0 0 0 0 0
```

```
run;

proc phreg data=whas500;
title "Final model with sho (Cardiogenic Shock) strata";
     model lenfol*fstat(0)= age gender hr diasbp bmi sho year
chf/ties=exact;
     strata sho;
     assess ph/resample;
     baseline out=final covariates=null survival=s lower=lcl
upper=ucl cumhaz=H lowercumhaz=lH uppercumhaz=uH;
run;


proc gplot data=final;
     title "Baseline Survival Function based on cardiogenic shock";
     symbol1 value=dot i=join;
     plot s*lenfol=sho;
run;

proc gplot data=final;
     title "Baseline cumulative hazard function based on cardiogenic
shock";
     plot H*lenfol=sho;
     run;
quit;

data covals;
     title;
     input age gender hr diasbp bmi sho year chf;
     cards;
40 1 70 80 24 1 1 1
run;

proc phreg data=whas500;
title;
     model lenfol*fstat(0)= age gender hr diasbp bmi sho year
chf/ties=exact;
     baseline out=pred covariates=covals survival=s lower=lcl
upper=ucl cumhaz=H /nomean;
```

```
run;

proc gplot data=pred;
     title "Predicted curves for 40 yr old woman with stated
attributes";
     plot s*lenfol H*lenfol;
run;

proc print data=pred;
run;

data subset;
   set pred;
   if _n_ in (19,20,21,22);
run;

proc print data=subset;
run;

ods graphics off;
```