Empirical Study: Survival Analysis Based on PTSD and Heart Failure

Haoxi Ma 09/01/2021

1. BACKGROUND

To investigate the association between post-traumatic stress (PTSD) and heart failure, a (hypothetical) sample of 850 veterans were enrolled and followed for approximately 8 years. All participants were free of heart disease and enrollment, and our outcome of interest is the incidence of heart failure.

Variables in the data set are:

Factors	Description
Idnum	Study ID
sexM	Gender, coded 1 for males, 0 for females
age	Age in years
bmicat	Coded 1 for underweight or normal weight, 2 for those overweight, and 3 for those obese
combat	Coded 1 for those who served in active combat, 0 otherwise
ptsd	Coded 1 for those with PTSD, 0 otherwise
SurvWeeks	Follow-up time, in weeks, to either heart failure or censoring
hfailure	Coded 1 for those developing heart failure, 0 otherwise

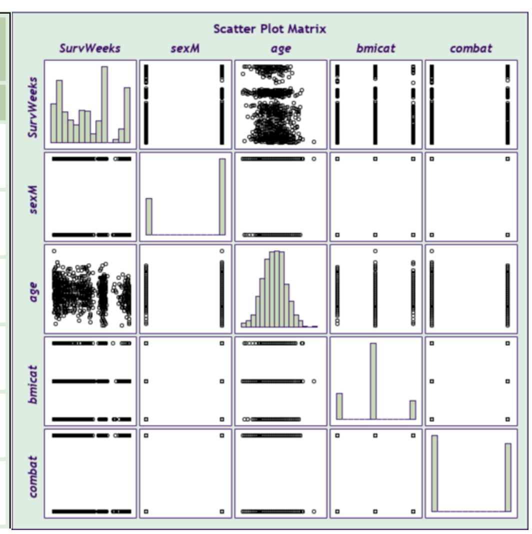
2. DATA PREPROCESSING

2.1. Pearson Correlation Coefficients and Scatter Matrix

Here, we generate the Pearson correlation coefficients and scatter matrix:

Pearson Correlation Coefficients, N = 850
Prob > |r| under H0: Rho=0

	SurvWeeks	sexM	age	bmicat	combat	ptsd
SurvWeeks SurvWeeks	1.00000	-0.12076 0.0004	-0.20318 <.0001	-0.09233 0.0071	-0.09343 0.0064	-0.16988 <.0001
sexM sexM	-0.12076 0.0004	1.00000	-0.04014 0.2424	0.04477 0.1922	-0.02248 0.5127	0.02993 0.3834
age age	-0.20318 <.0001	-0.04014 0.2424	1.00000	0.00183 0.9574	-0.02597 0.4495	0.03173 0.3556
bmicat bmicat	-0.09233 0.0071	0.04477 0.1922	0.00183 0.9574	1.00000	0.04695 0.1714	-0.01234 0.7194
combat combat	-0.09343 0.0064	-0.02248 0.5127	-0.02597 0.4495	0.04695 0.1714	1.00000	-0.01364 0.6913
ptsd ptsd	-0.16988 <.0001	0.02993 0.3834	0.03173 0.3556	-0.01234 0.7194	-0.01364 0.6913	1.00000

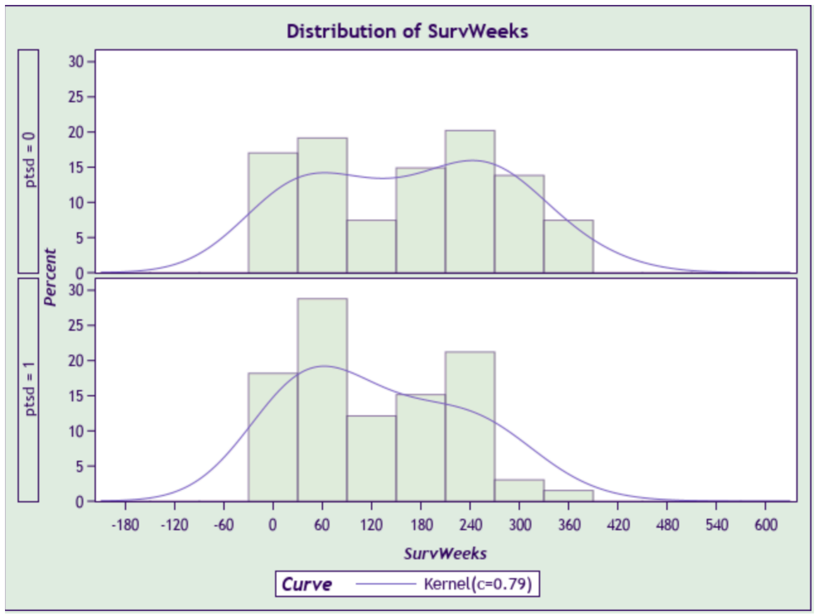


Combing the figure and the SAS output, all factors have some association with the follow-up time while the linear relationship between pair of them might not be significant.

2.2. Data Visualization

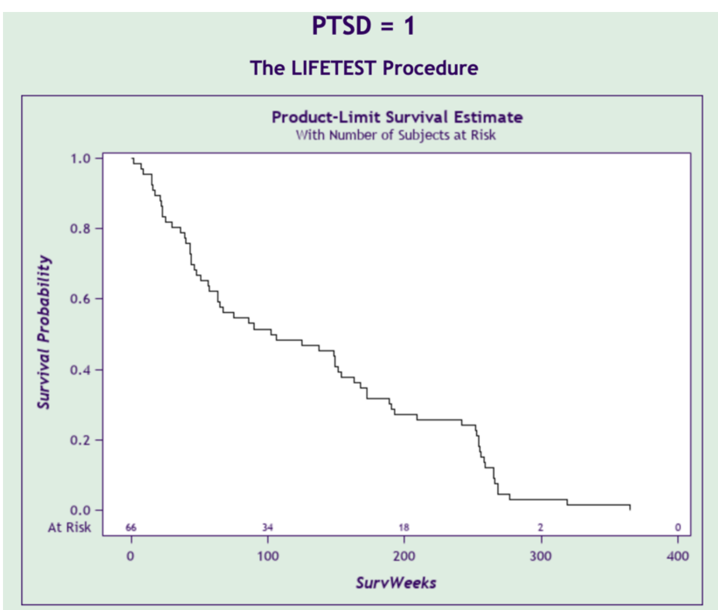
Here, we construct some plots to have an intuitive understanding of the data based on the people who had a heart failure finally.

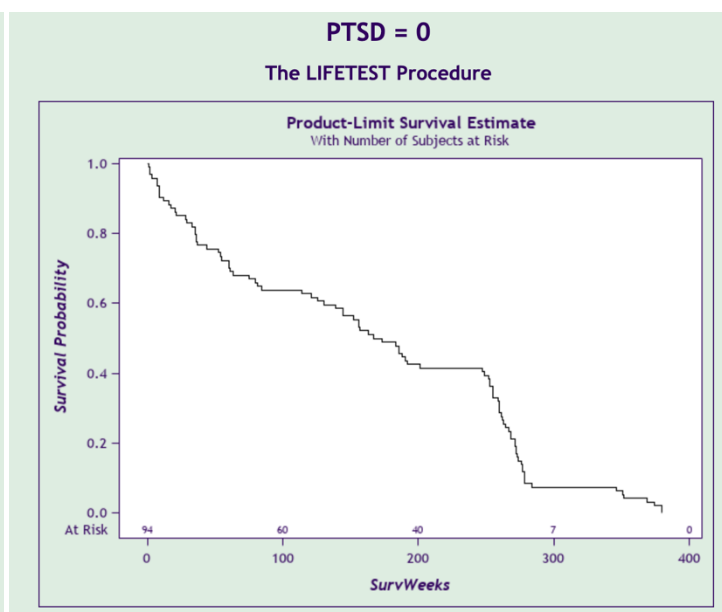
First, we draw the PDF of the survey time between people getting PTSD and people not:



According to the plot, compared to the people without PTSD, the shorter survival time of the people with PTSD are more probable, indicating that the risk of getting heart failure is higher.

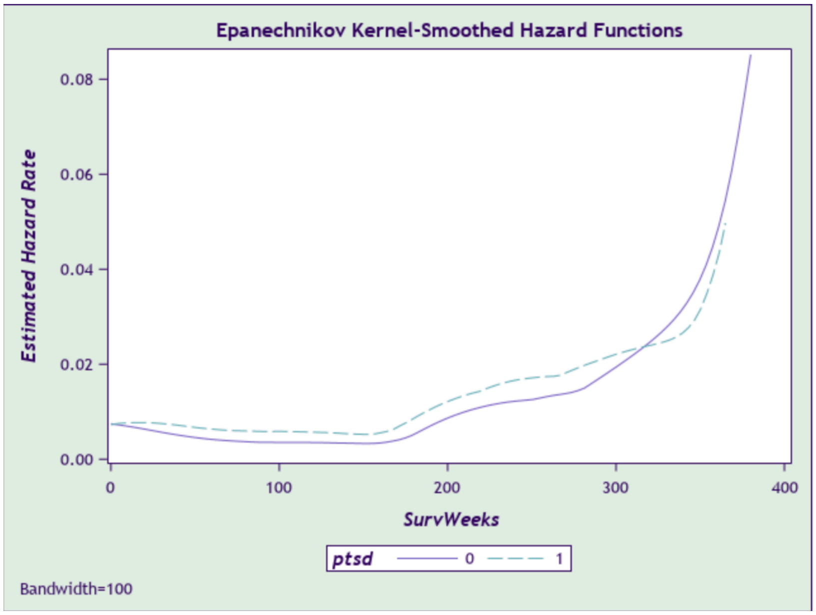
Then, we draw the survival plots within these two groups:





From the two plots above, we can infer that having PTSD might increase the risk of getting heart failure.

At last, we draw the hazard plot for two groups:



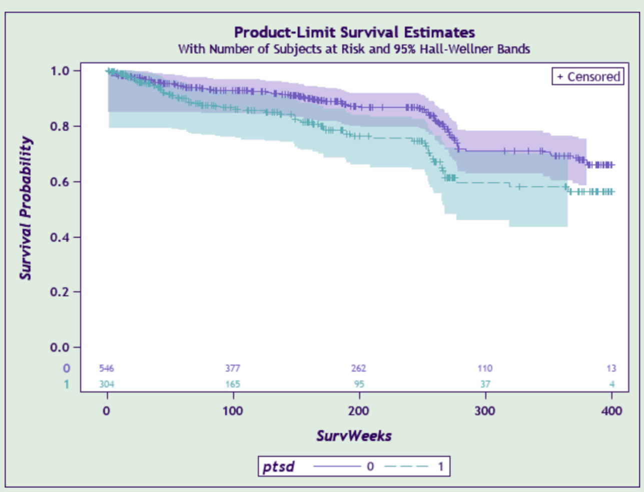
This plot comes to the same conclusion as before.

3. Non-parametric Method for Testing the Effect of PTSD

We use Kaplan-Meier estimation addition with the confidence bands to compare the survival functions of PTSD = 1 an PTSD = 0:

Test of Equality over Strata

Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	11.1383	1	0.0008
Wilcoxon	11.1013	1	0.0009
-2Log(LR)	9.5171	1	0.0020

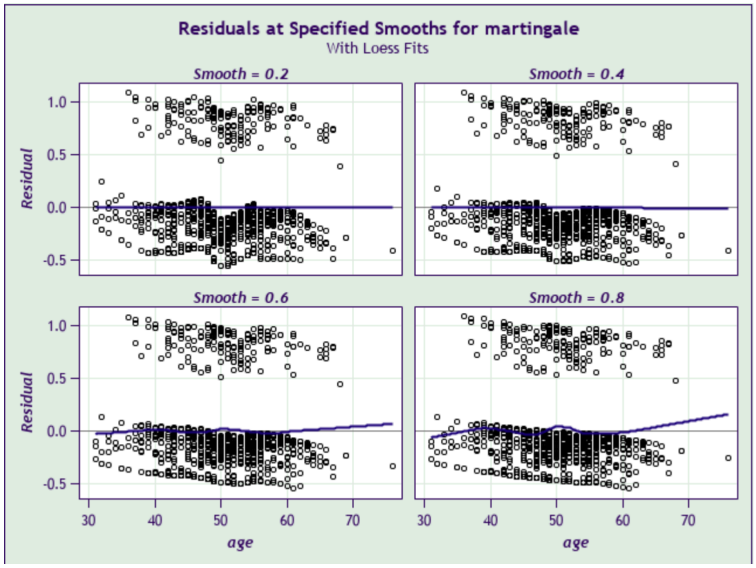


From the plot, we can infer that the hazard rate of people with PTSD are significantly higher than those without PTSD. The three tests also proof this conclusion. But this method doesn't take other factors into consideration. Therefore, we need use parametric method to control the other factors' effect.

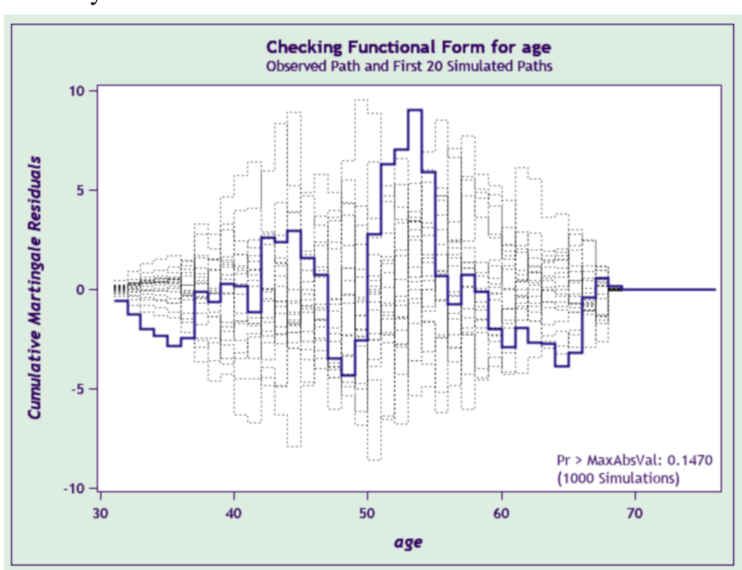
4. Parametric Method for Testing the Effect of PTSD

4.1. Exploring Functional Form of Continuous Covariates

In this dataset, there is only one continuous covariate—age. In order to identify its optimal form in the model, we construct the cumulative martingale residuals plot:

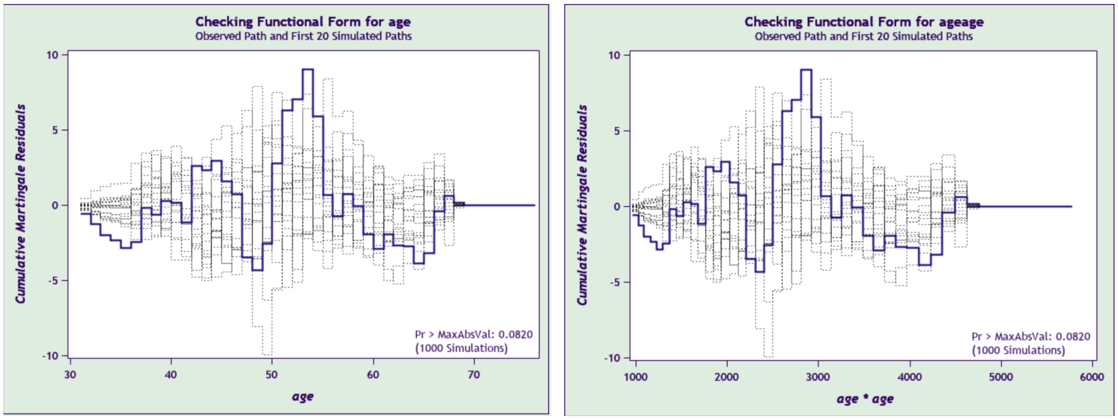


From the panel above, we can infer that the linear form is adequate. But, we can still check whether the quadratic form is useful by statistics.



Supremum Test for Functional Form				
Variable	Maximum Absolute Value	Replications	Seed	Pr > MaxAbsVal
age	9.0052	1000	710331697	0.1470

When using linear form in the model, the p-value is 0.1470 and the cumulative martingale residuals plot is good enough. Then, we put the quadratic form into the model:



Supremum Test for Functional Form				
Variable	Maximum Absolute Value	Replications	Seed	Pr > MaxAbsVal
age	9.0144	1000	1005236136	0.0820
ageage	9.0144	1000	1005236136	0.0820

From the results, cumulative martingale residuals plots don't change much and the p-value becomes 0.082, indicating less adequate. Therefore, only using linear form of age covariate is enough for our model.

4.2. Check the Interaction between Age and Gender

We can dig deeper into the interaction within covariates based on the articles and common sense. Usually, when we focus on a healthcare data, checking the interaction between age and gender might give some useful information. Here, we construct the model by Cox regression:

Joint Tests			
Effect	DF	Wald Chi-Square	Pr > ChiSq
age	1	4.7300	0.0296
sexM	1	3.5143	0.0608
age*sexM	1	2.7540	0.0970
bmicat	2	1.5375	0.4636
combat	1	0.1918	0.6614
ptsd	1	10.0443	0.0015

According to output, the interaction term is not significant, so we don't need to add them into the model.

4.3. Cox Regression to Identify the PTSD's Effect

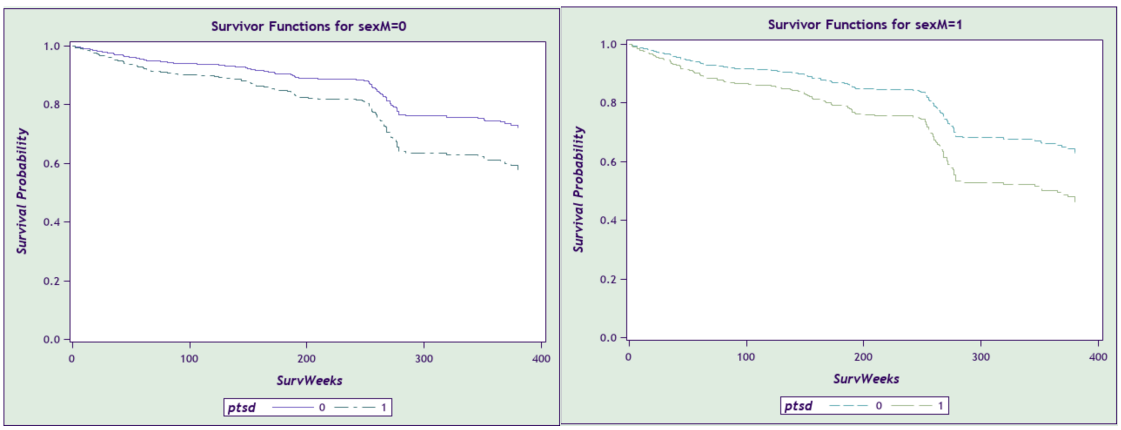
After the analysis above, we got our final model:

$$Survweeks * hfailure(0) = age + sexM + bmicat + combat + PTSD$$

Then, we fit the Cox regression and the result is:

Analysis of Maximum Likelihood Estimates								
Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Label
age		1	0.04154	0.01166	12.6994	0.0004	1.042	age
sexM	Male	1	-0.33030	0.17325	3.6346	0.0566	0.719	sexM Male
bmicat	Overweight	1	-0.10489	0.22483	0.2176	0.6408	0.900	bmicat Overweight
bmicat	UnderOrnormal Weight	1	-0.28494	0.26618	1.1459	0.2844	0.752	bmicat UnderOrnormal Weight
combat	Nonactive serving	1	0.07792	0.16283	0.2290	0.6323	1.081	combat Nonactive serving
ptsd	1	1	0.50667	0.16222	9.7556	0.0018	1.660	ptsd 1

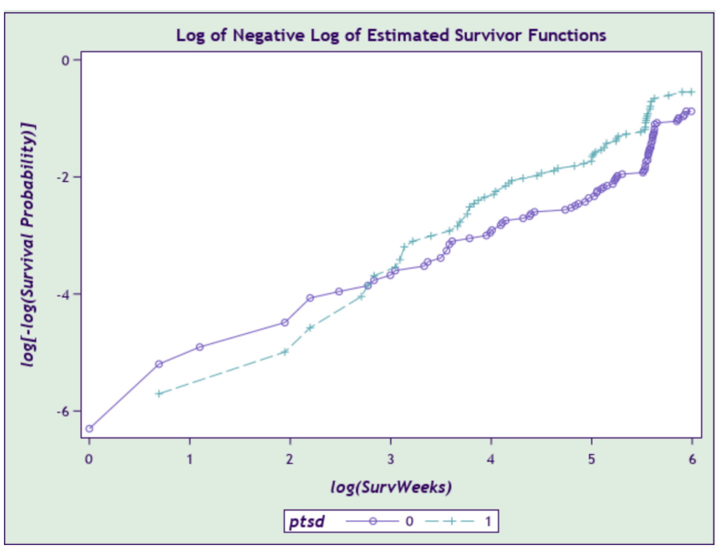
From the output, we find that regression coefficient of PTSD is 0.50667 and also significant (P-value<0.05). And the hazard ratio is 1.66 greater than 1, which means the veterans with PTSD are at higher risk of getting heart failure. Besides, we use graphs to interpret effects:



From the result, after controlling the age covariate, veterans with PTSD are at higher risk of getting heart failure regardless of the gender.

5. Check the Proportional Hazard Assumption

Here, we are going to check the PH assumption when fitting the cox regression. First, we draw Log-log plot:



Because two lines intersect, we can conclude that there is a violation of PH assumption. Then, we need to identify which factor does not satisfy the PH assumption:

zph Tests for Nonproportional Hazards						
Transform	Predictor Variable	Correlation	ChiSquare	Pr > ChiSquare	t Value	Pr > |t|
RANK	age	0.0276	0.1254	0.7233	0.35	0.7289
RANK	sexMMales	-0.0987	1.5300	0.2161	-1.25	0.2141
RANK	bmicatOverweight	-0.0620	0.6261	0.4288	-0.78	0.4360
RANK	bmicatUnderOrnormal Weight	-0.0280	0.1259	0.7228	-0.35	0.7255
RANK	combatNonactive serving	0.1668	4.4919	0.0341	2.13	0.0350
RANK	ptsd1	-0.0621	0.6200	0.4310	-0.78	0.4354

From the output, obviously, combat factor violates the assumption, so we need to do some remedies. We use two methods to deal with the nonproportionality. First, we add the time-related factor into the model: creates two 'combat' variables, the first turns on for the first year (52 weeks) of the study, and then turns off, while the second is turned off for the first year and then turns on, and the result is:

Analysis of Maximum Likelihood Estimates										
Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits		Label
age		1	0.04065	0.01167	12.1373	0.0005	1.041	1.018	1.066	age
sexM	Males	1	0.32779	0.17320	3.5819	0.0584	1.388	0.988	1.949	sexM Males
bmicat	Overweight	1	-0.09381	0.22488	0.1740	0.6765	0.910	0.586	1.415	bmicat Overweight
bmicat	UnderOrnormal Weight	1	-0.27604	0.26619	1.0753	0.2997	0.759	0.450	1.279	bmicat UnderOrnormal Weight
ptsd	1	1	0.50247	0.16208	9.6108	0.0019	1.653	1.203	2.271	ptsd 1
combat1		1	0.54861	0.29560	3.4444	0.0635	1.731	0.970	3.089	
combat2		1	-0.36127	0.20199	3.1990	0.0737	0.697	0.469	1.035	

According to the result, the interaction combat1 and combat2 are borderline significant, so our remedy works. Still, hazard ratio of PTSD is 1.653, significantly greater than 1 (Lower bound of CL > 1). Besides, we can allow non-proportional hazards for combat exposure by running a stratified Cox regression model, and the result is:

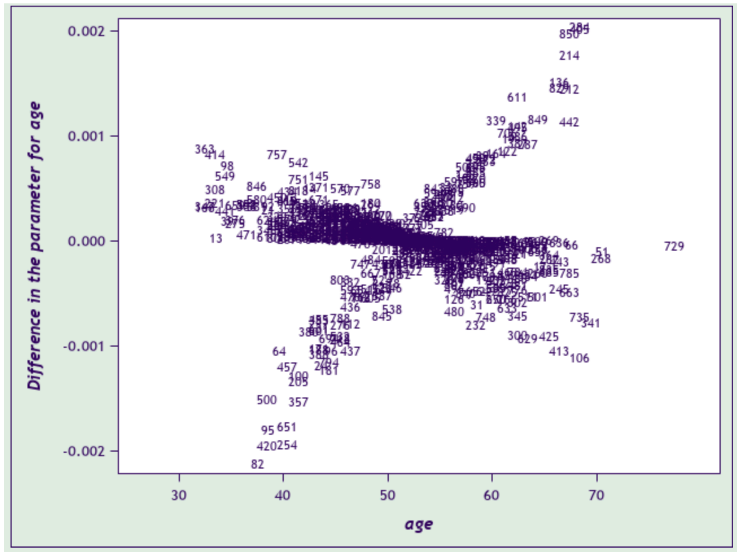
Analysis of Maximum Likelihood Estimates										
Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits		Label
age		1	0.04049	0.01168	12.0241	0.0005	1.041	1.018	1.065	age
sexM	Males	1	0.32738	0.17318	3.5735	0.0587	1.387	0.988	1.948	sexM Males
bmicat	Overweight	1	-0.08881	0.22524	0.1555	0.6934	0.915	0.588	1.423	bmicat Overweight
bmicat	UnderOrnormal Weight	1	-0.27099	0.26665	1.0328	0.3095	0.763	0.452	1.286	bmicat UnderOrnormal Weight
ptsd	1	1	0.50295	0.16210	9.6272	0.0019	1.654	1.204	2.272	ptsd 1

Again, the hazard ratio of PTSD is 1.654, indicating the veterans with PTSD are at higher risk of getting heart failure.

6. Check the Influential Case

6.1. Influence on Regression Coefficients

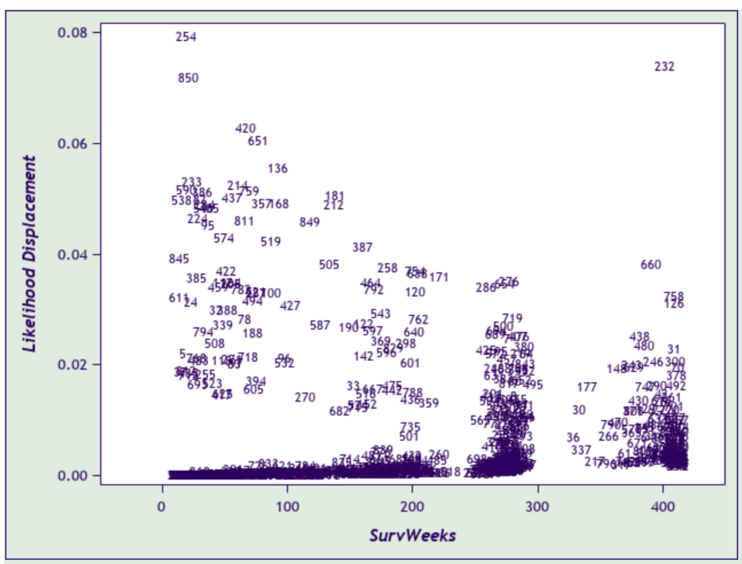
We check the influential case by DFBETAS statistics:



There is no influential case on regression coefficients.

6.2. Influence on Overall Model

Here, we check the influence on overall model by plotting likelihood displacement:



There is no influential case on overall model.

7. SUMMARY

Veterans with PTSD are at higher risk of getting heart failure controlling age, gender, BMI and whether active serving.