Empirical Study: Survival Analysis Based on PTSD and Heart Failure

**Haoxi Ma 09/01/2021**

# BACKGROUND

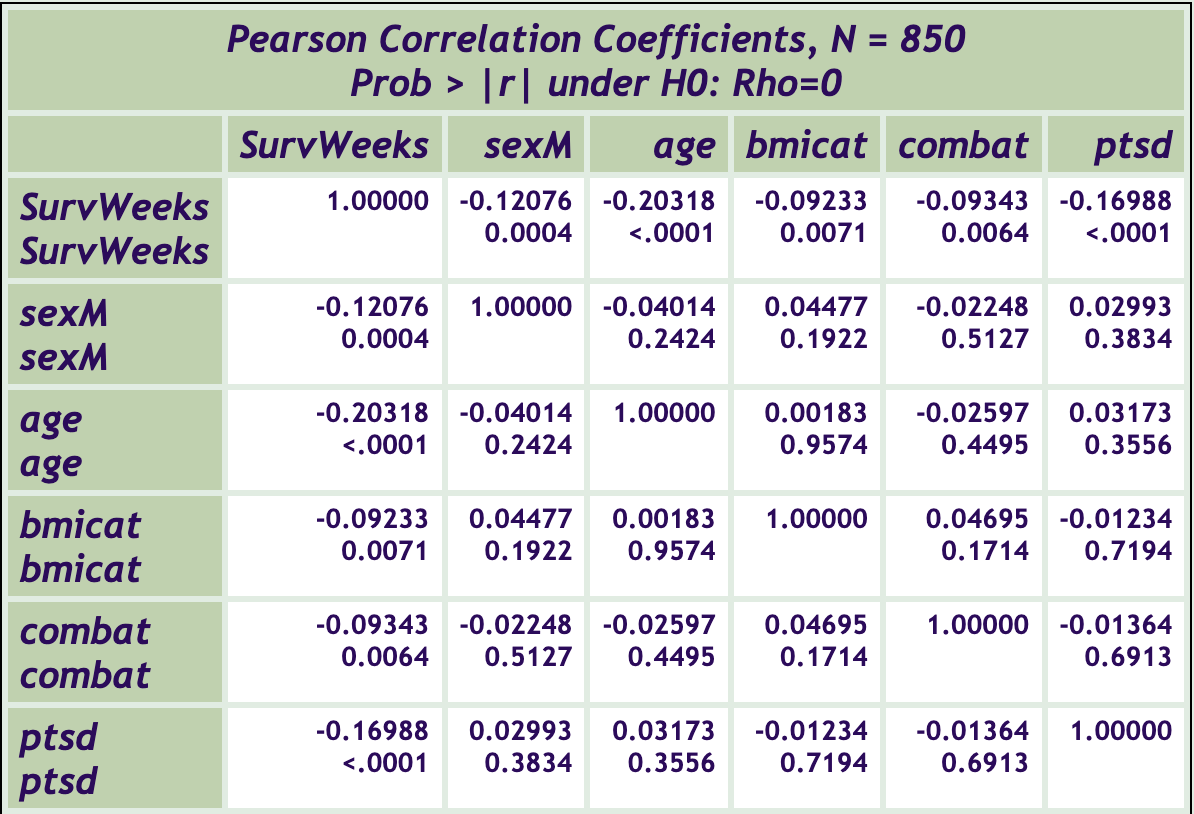
To investigate the association between post-traumatic stress (PTSD) and heart failure, a (hypothetical) sample of 850 veterans were enrolled and followed for approximately 8 years. All participants were free of heart disease and enrollment, and our outcome of interest is the incidence of heart failure.

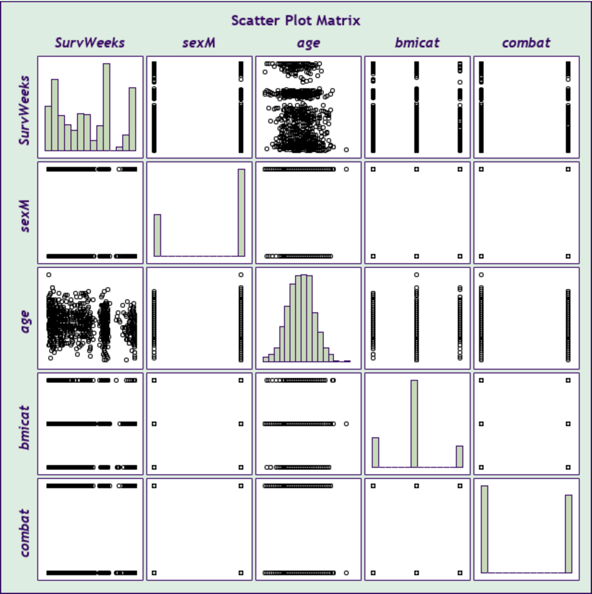
Variables in the data set are:

|  |  |
| --- | --- |
| 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 |

# DATA PREPROCESSING

## Pearson Correlation Coefficients and Scatter Matrix

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

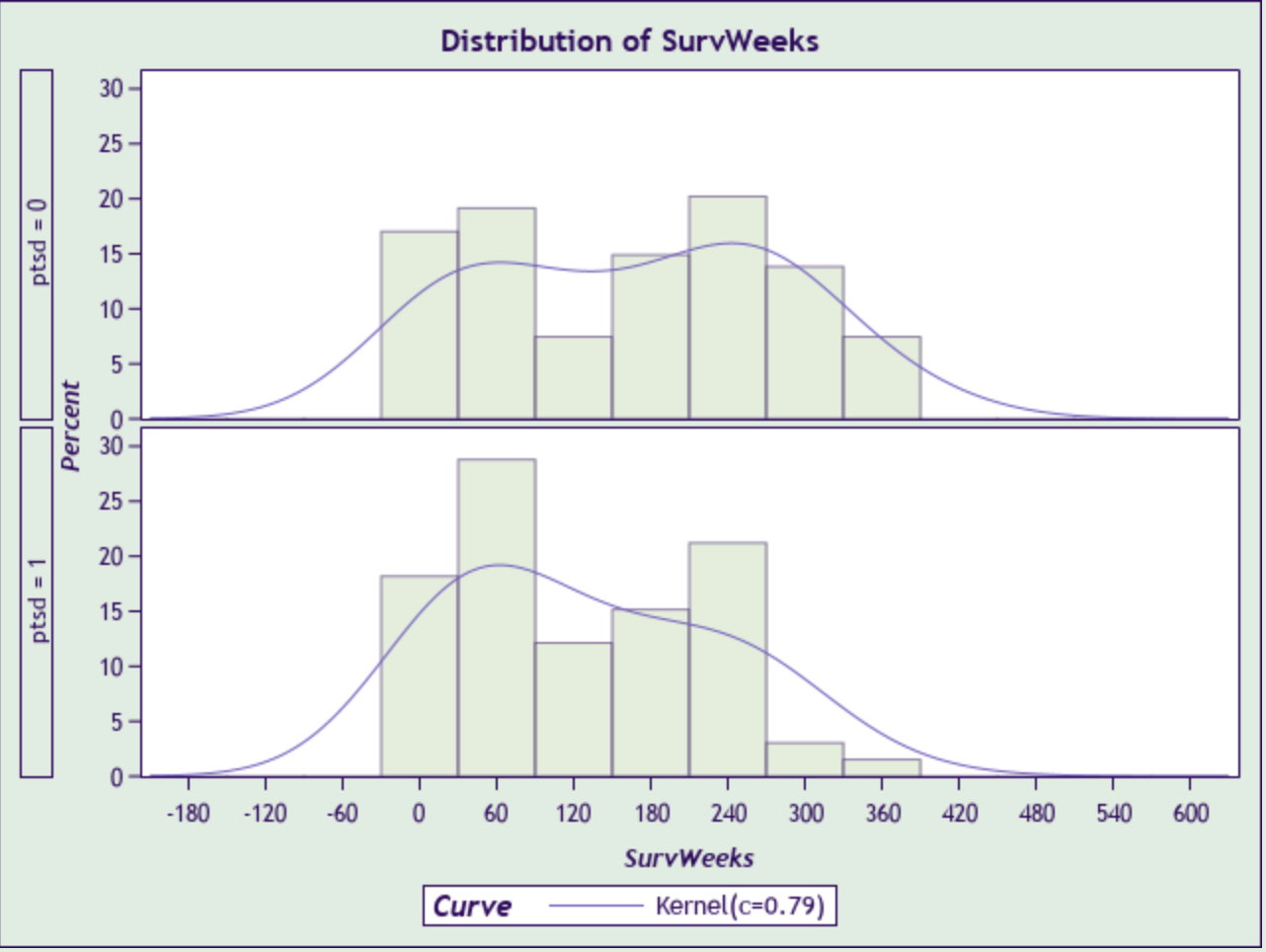


Combing the figure and the SAS output, all factors have some association with the follow-up time while the linear relationship between pair of them might not be significant.

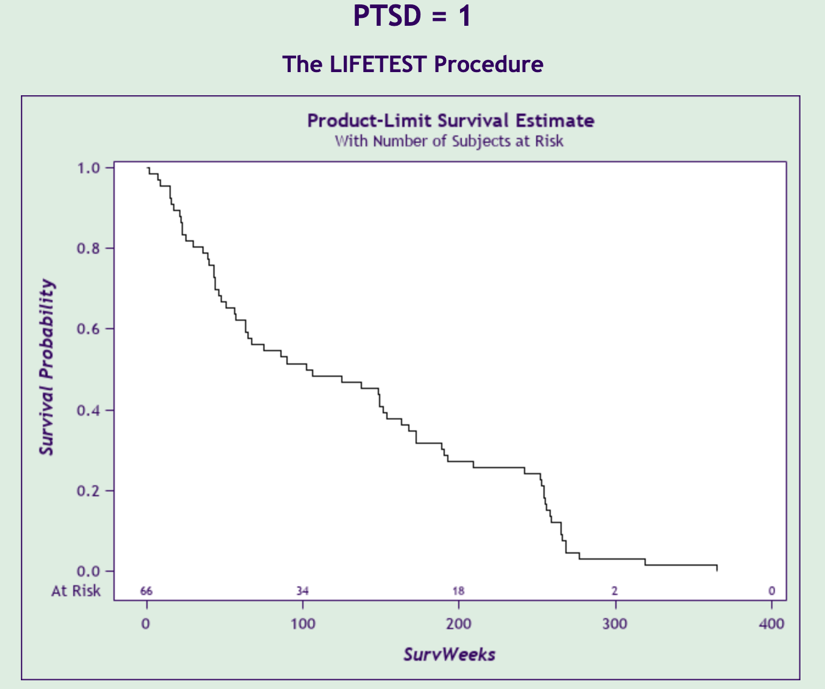
## Data Visualization

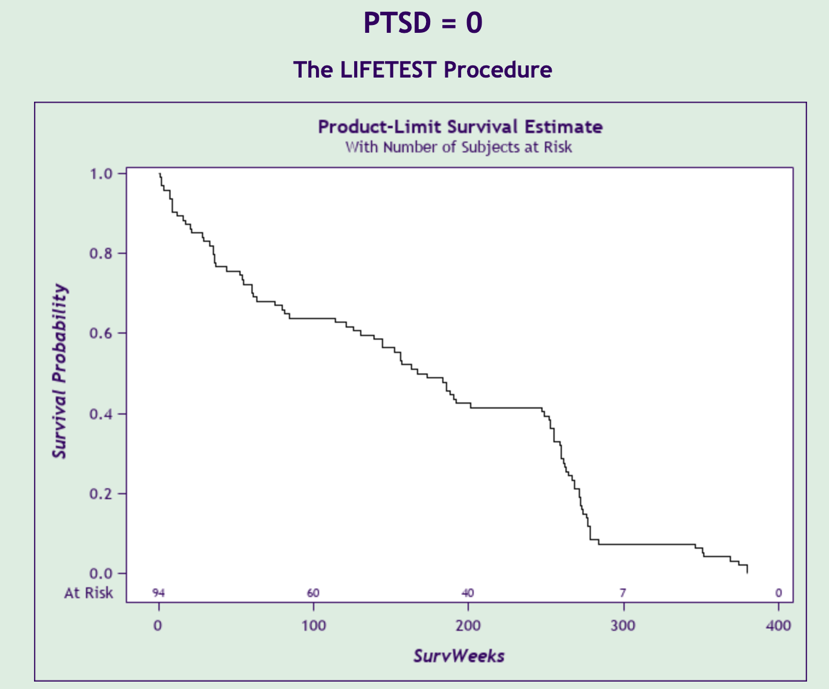
Here, we construct some plots to have an intuitive understanding of the data based on the people who had a heart failure finally.

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



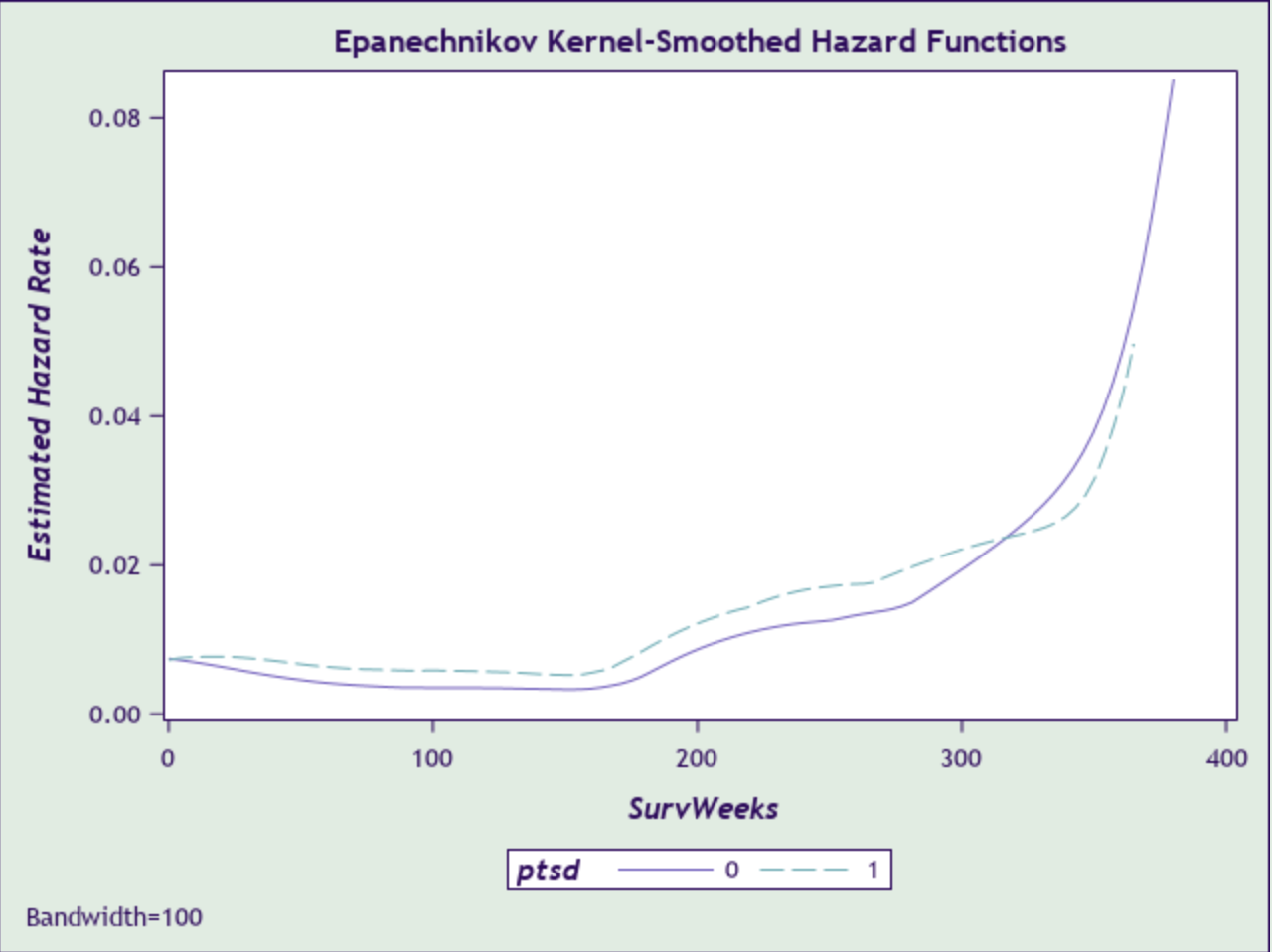
According to the plot, compared to the people without PTSD, the shorter survival time of the people with PTSD are more probable, indicating that the risk of getting heart failure is higher.

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.

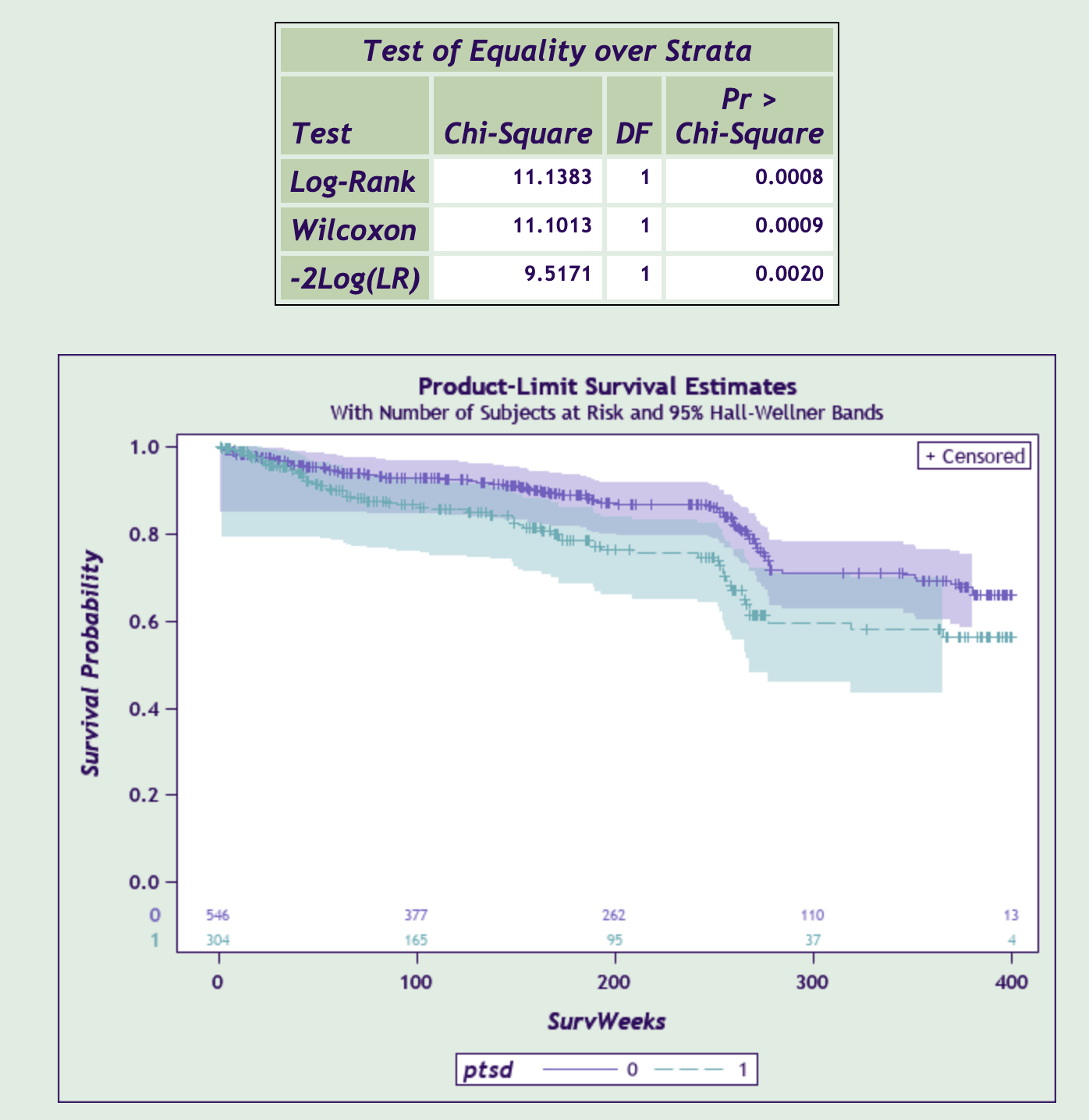
At last, we draw the hazard plot for two groups:



This plot comes to the same conclusion as before.

# 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:

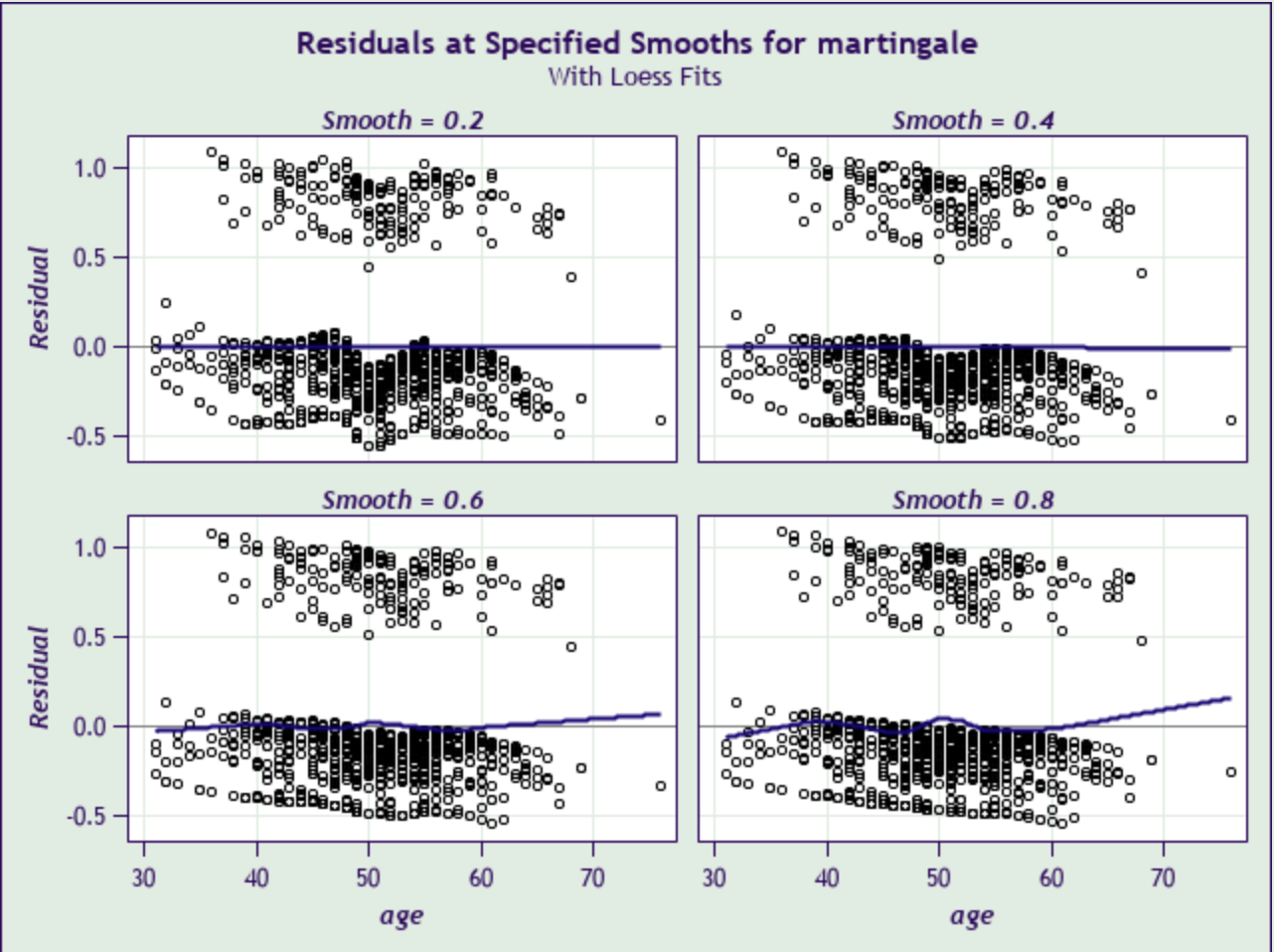


From the plot, we can infer that the hazard rate of people with PTSD are significantly higher than those without PTSD. The three tests also proof this conclusion. But this method doesn’t take other factors into consideration. Therefore, we need use parametric method to control the other factors’ effect.

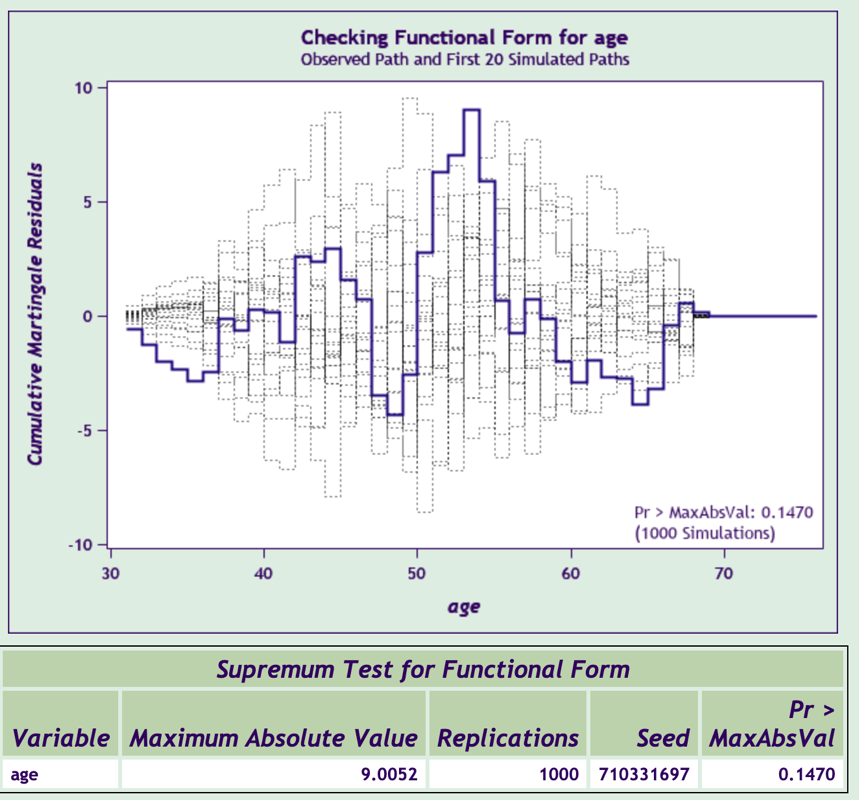
# Parametric Method for Testing the Effect of PTSD

## Exploring Functional Form of Continuous Covariates

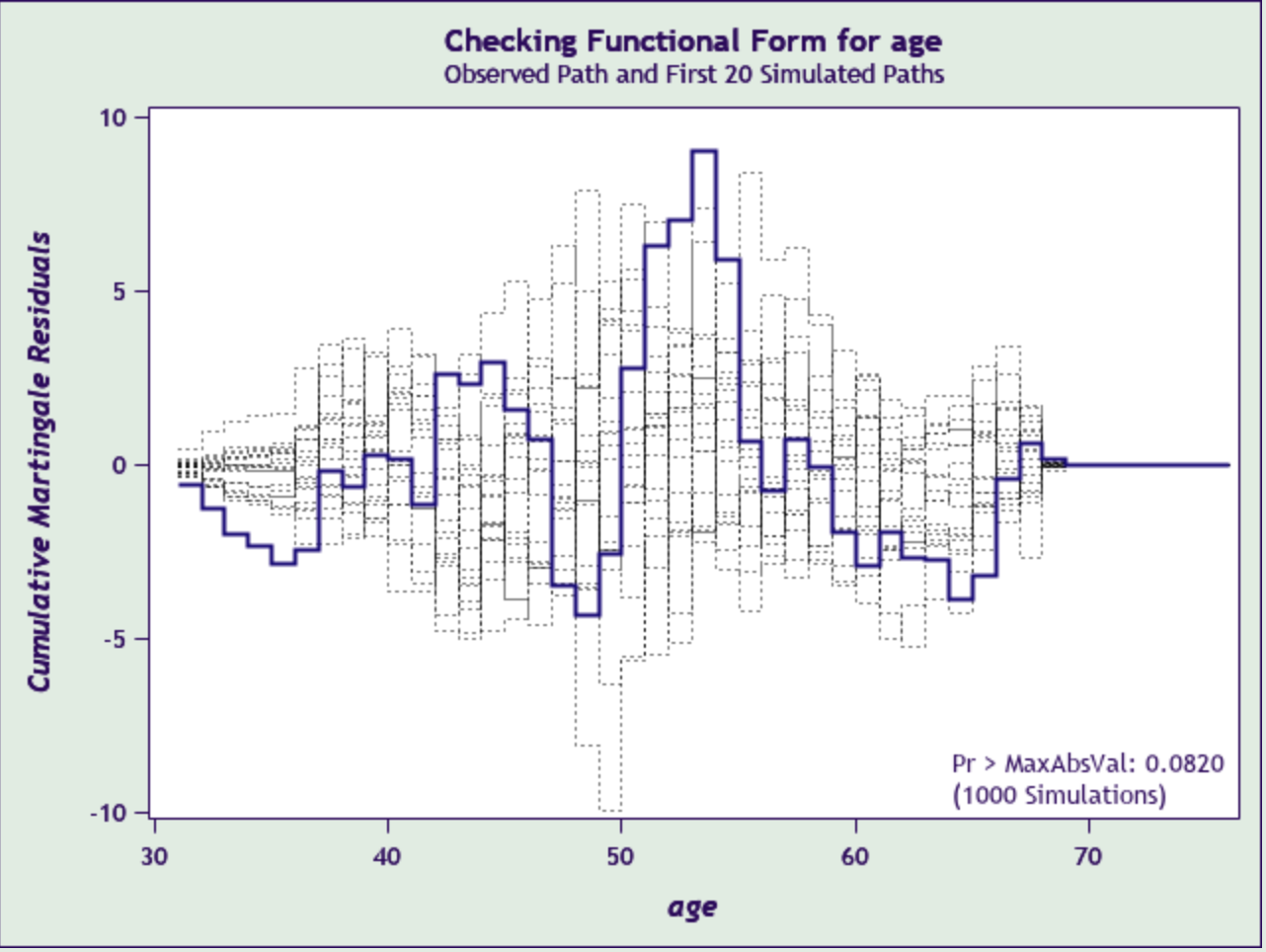
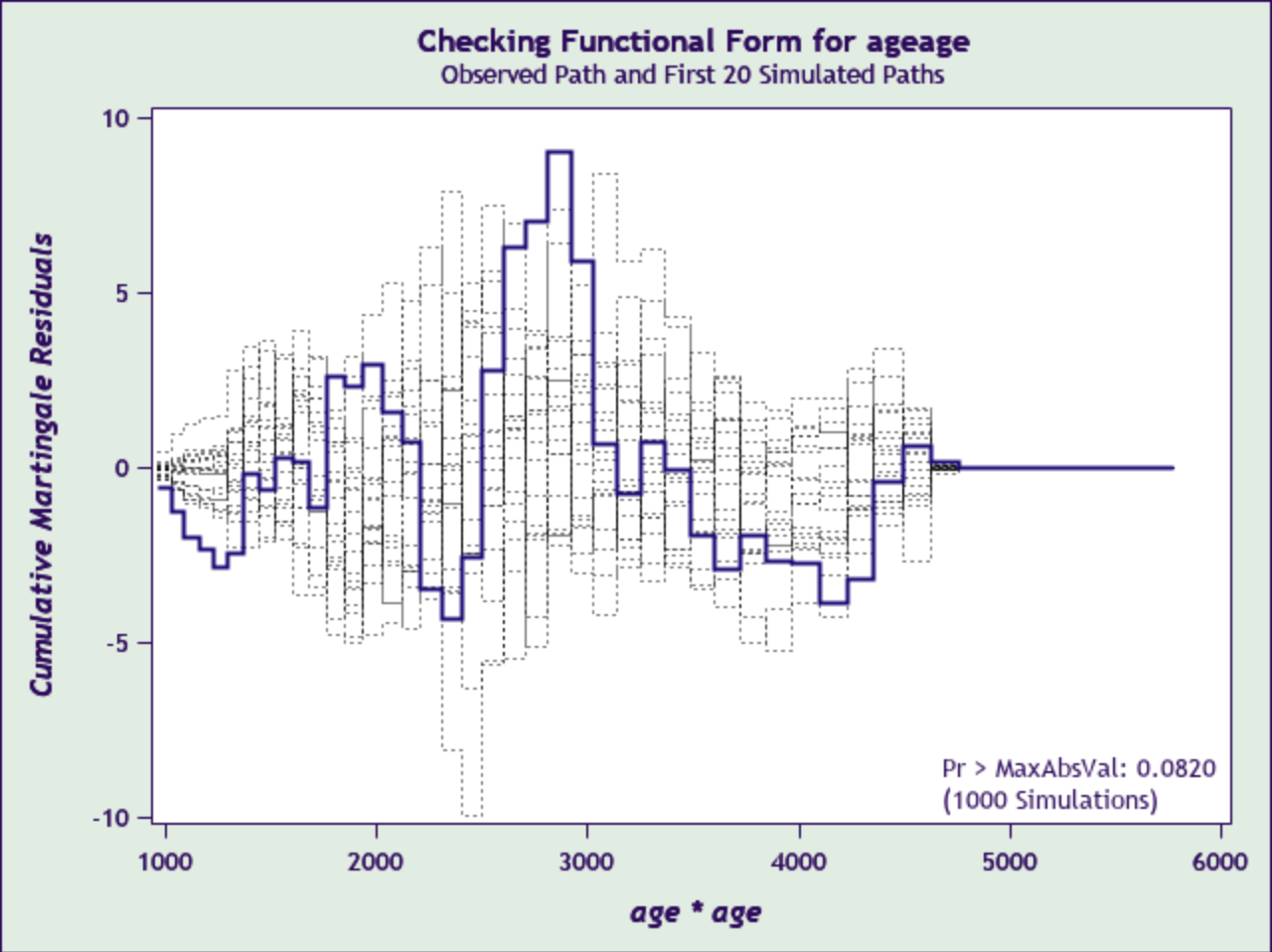
In this dataset, there is only one continuous covariate—age. In order to identify its optimal form in the model, we construct the cumulative martingale residuals plot:

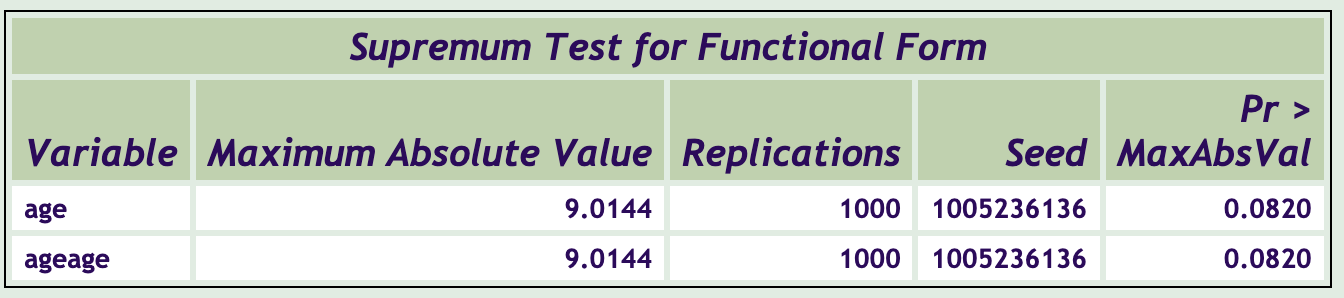


From the panel above, we can infer that the linear form is adequate. But, we can still check whether the quadratic form is useful by statistics.



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:

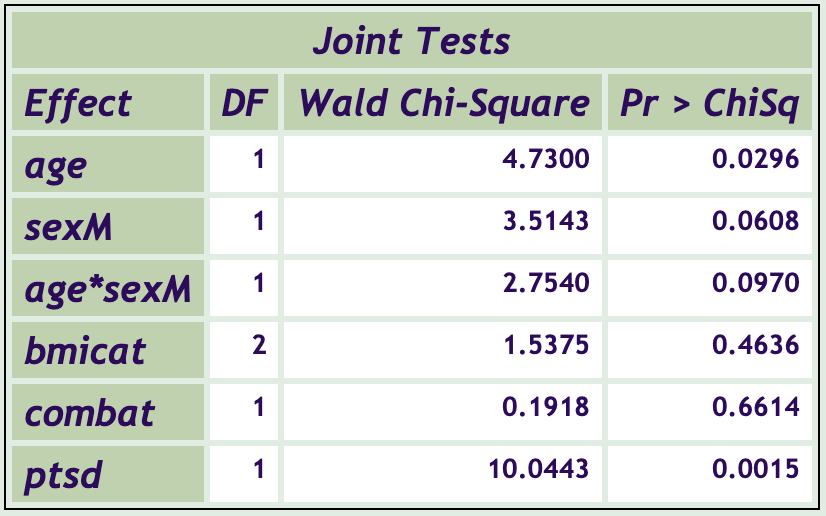


From the results, cumulative martingale residuals plots don’t change much and the p-value becomes 0.082, indicating less adequate. Therefore, only using linear form of age covariate is enough for our model.

* 1. 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:

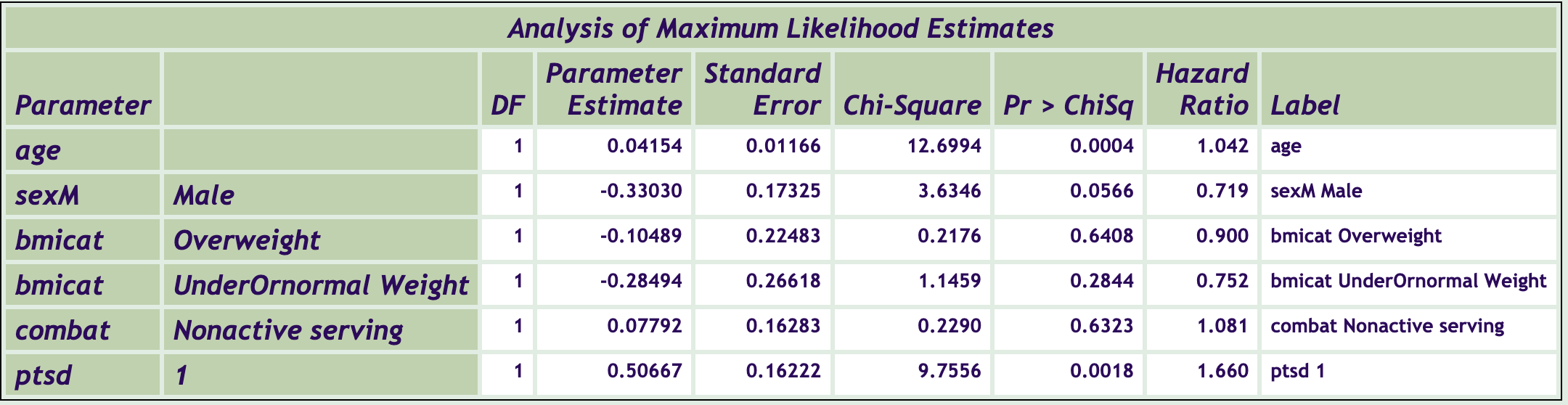


According to output, the interaction term is not significant, so we don’t need to add them into the model.

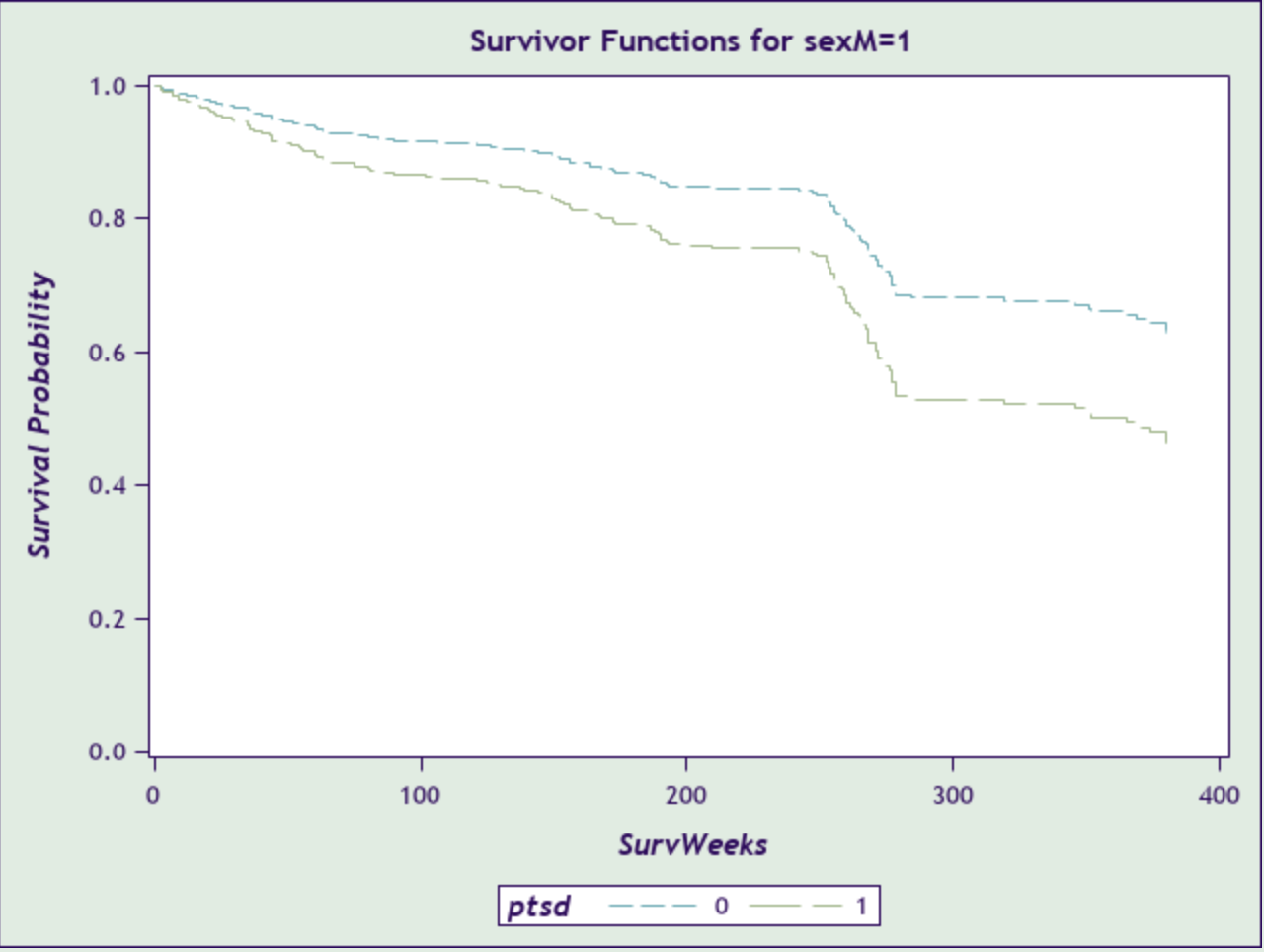
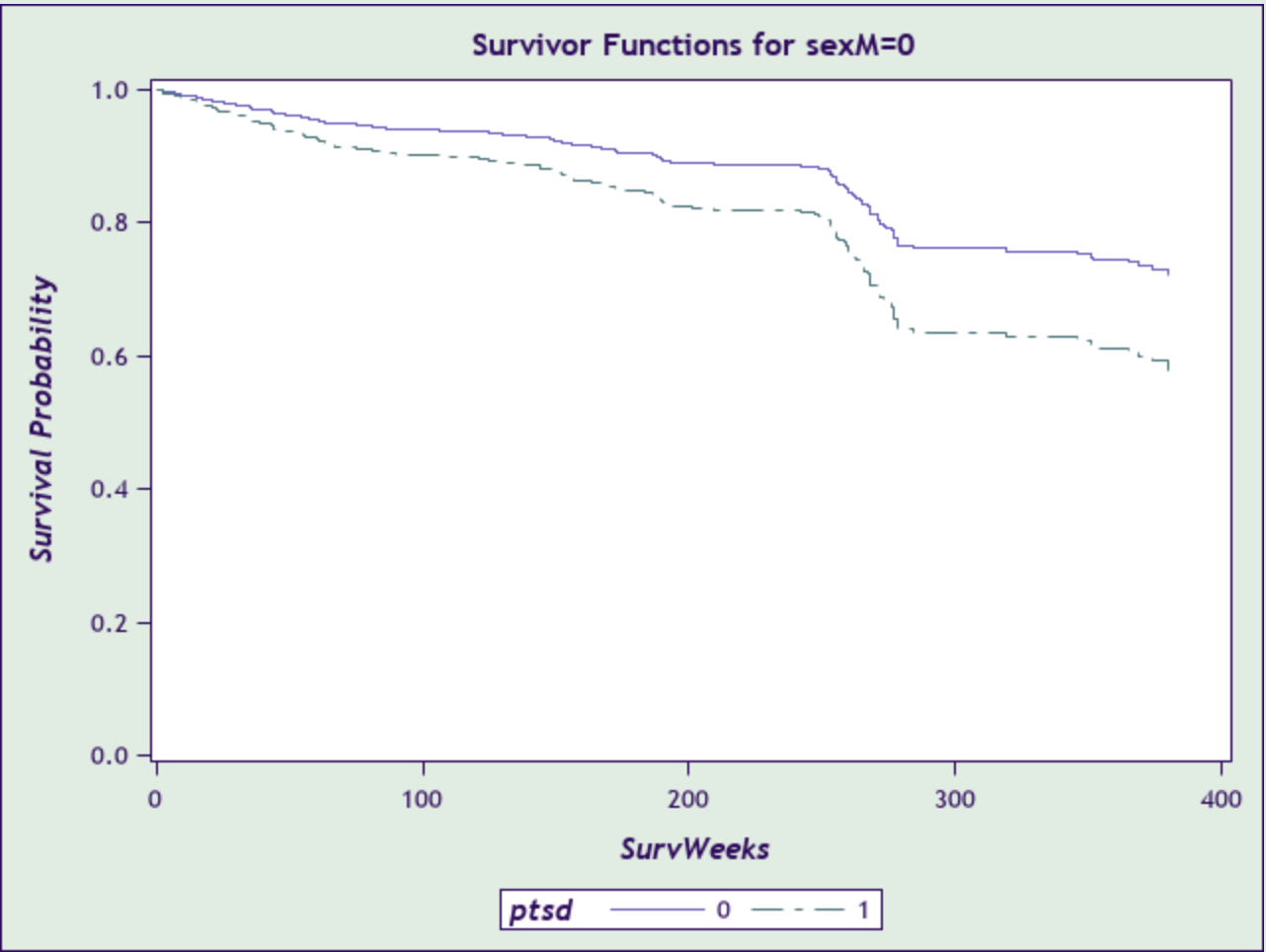
1. Cox Regression to Identify the PTSD’s Effect

After the analysis above, we got our final model:

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



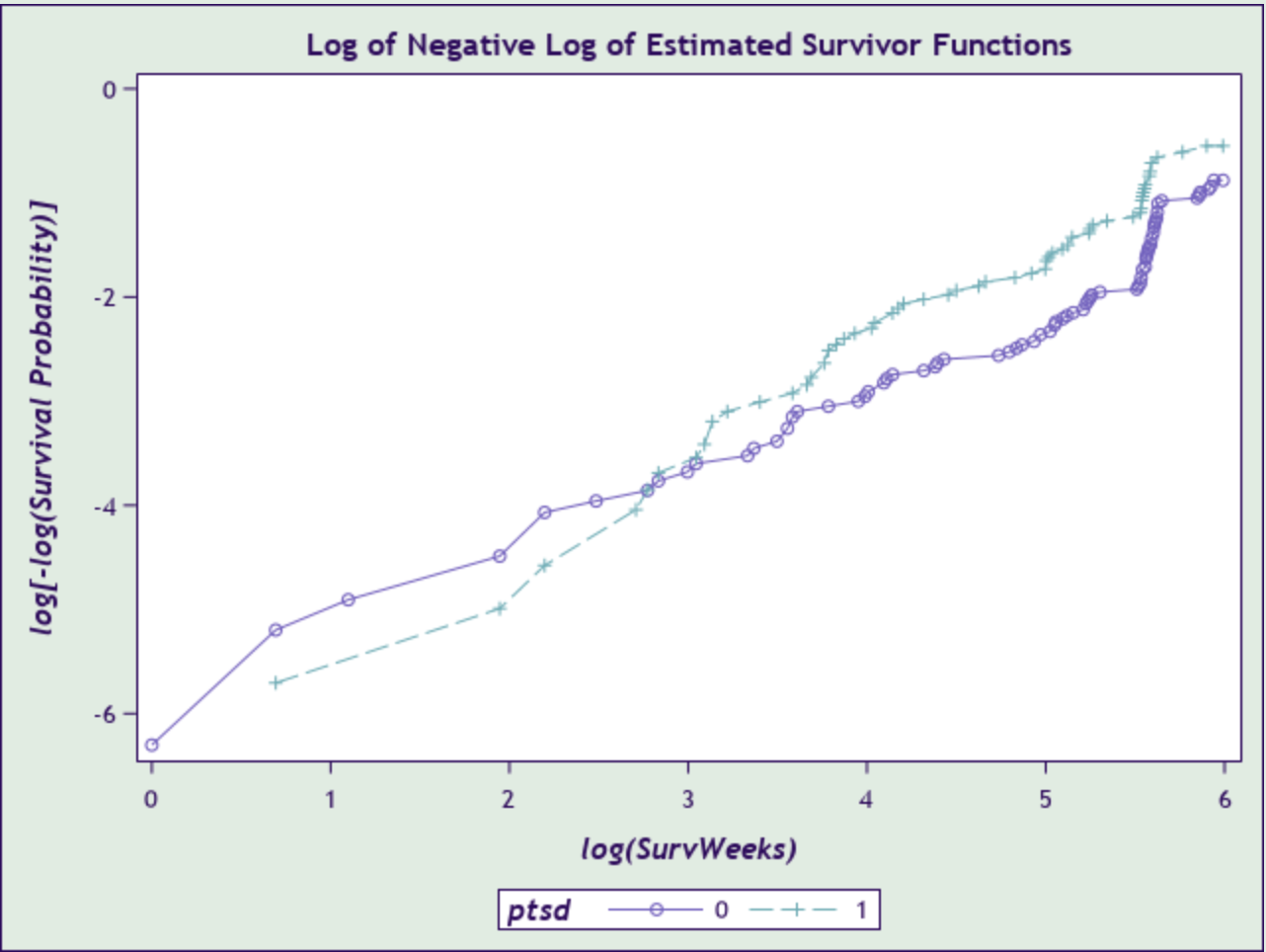
From the output, we find that regression coefficient of PTSD is 0.50667 and also significant (P-value<0.05). And the hazard ratio is 1.66 greater than 1, which means the veterans with PTSD are at higher risk of getting heart failure. Besides, we use graphs to interpret effects:



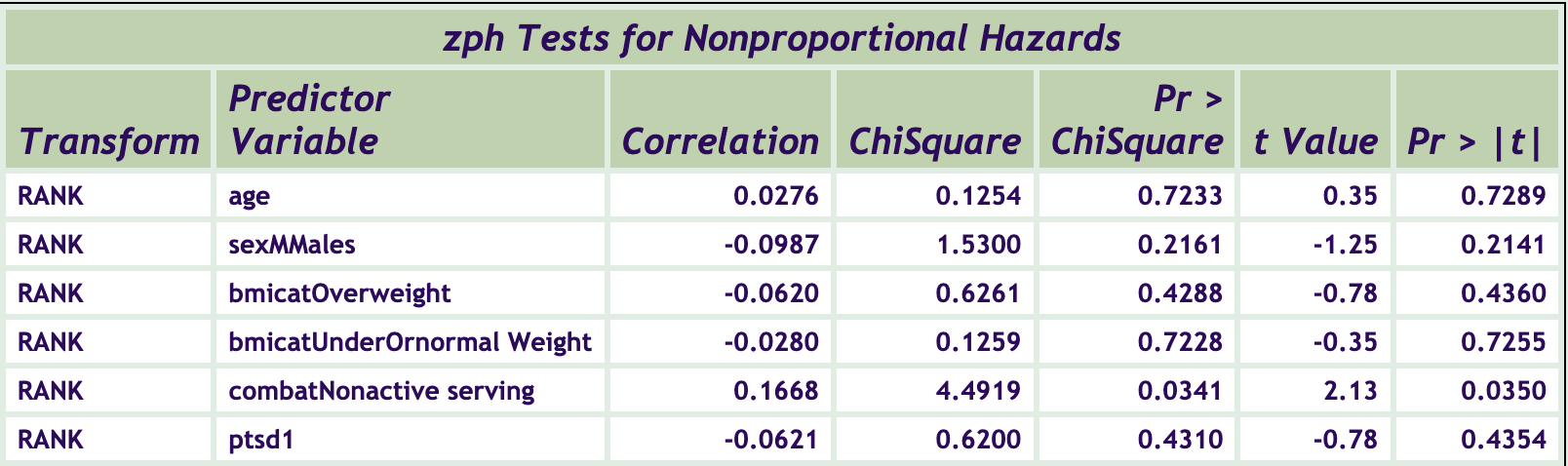
From the result, after controlling the age covariate, veterans with PTSD are at higher risk of getting heart failure regardless of the gender.

# 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:

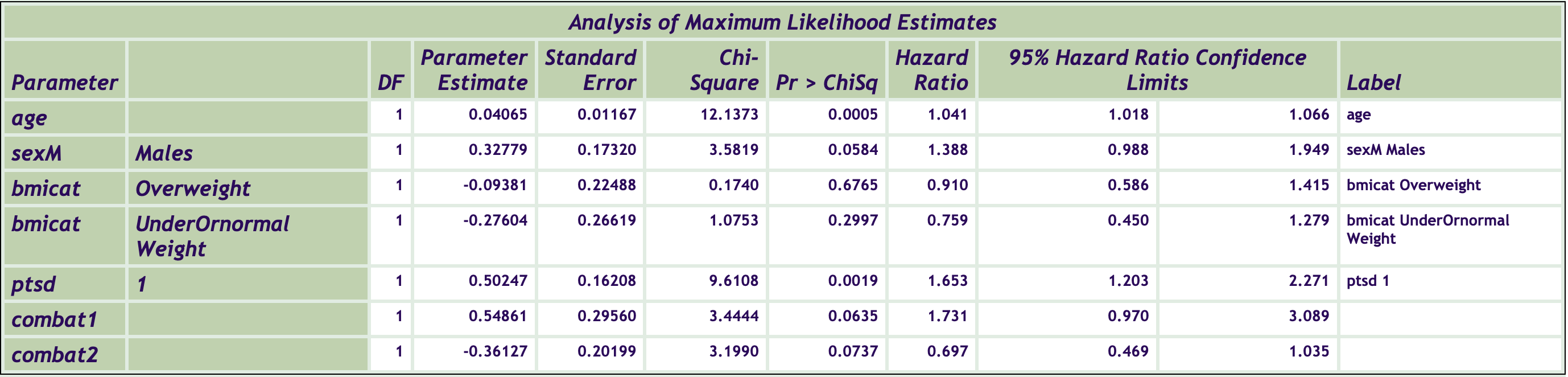


Because two lines intersect, we can conclude that there is a violation of PH assumption. Then, we need to identify which factor does not satisfy the PH assumption:

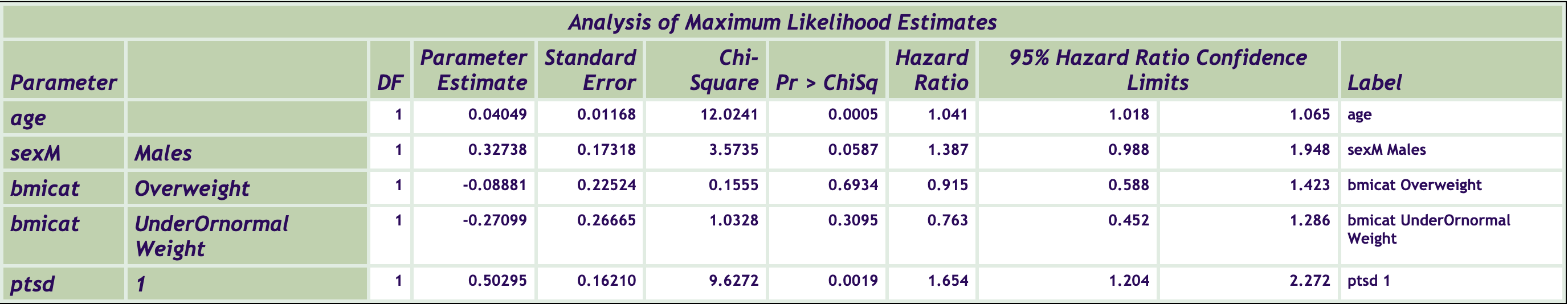


From the output, obviously, combat factor violates the assumption, so we need to do some remedies.

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:



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:

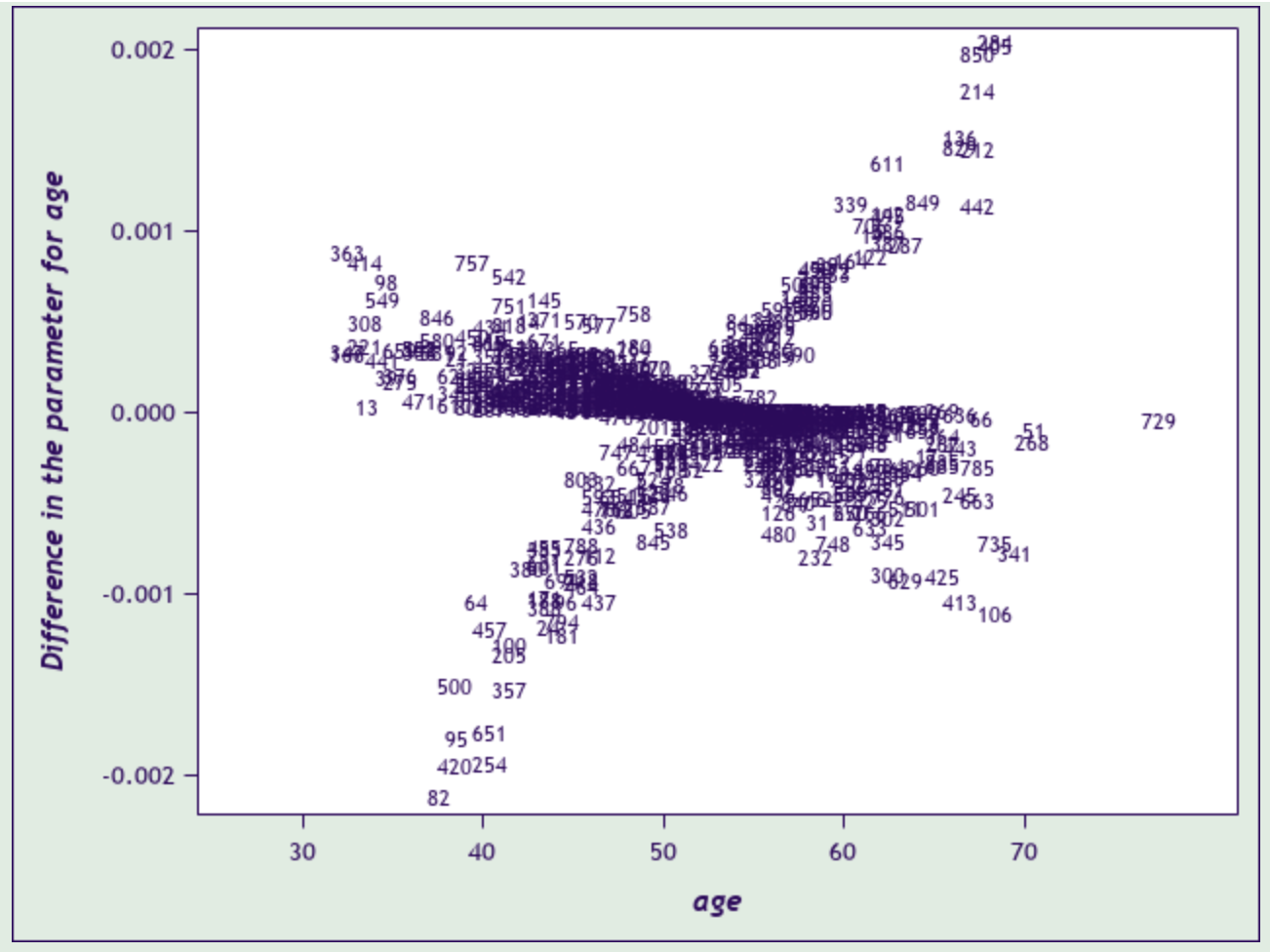


Again, the hazard ratio of PTSD is 1.654, indicating the veterans with PTSD are at higher risk of getting heart failure.

# Check the Influential Case

* 1. Influence on Regression Coefficients

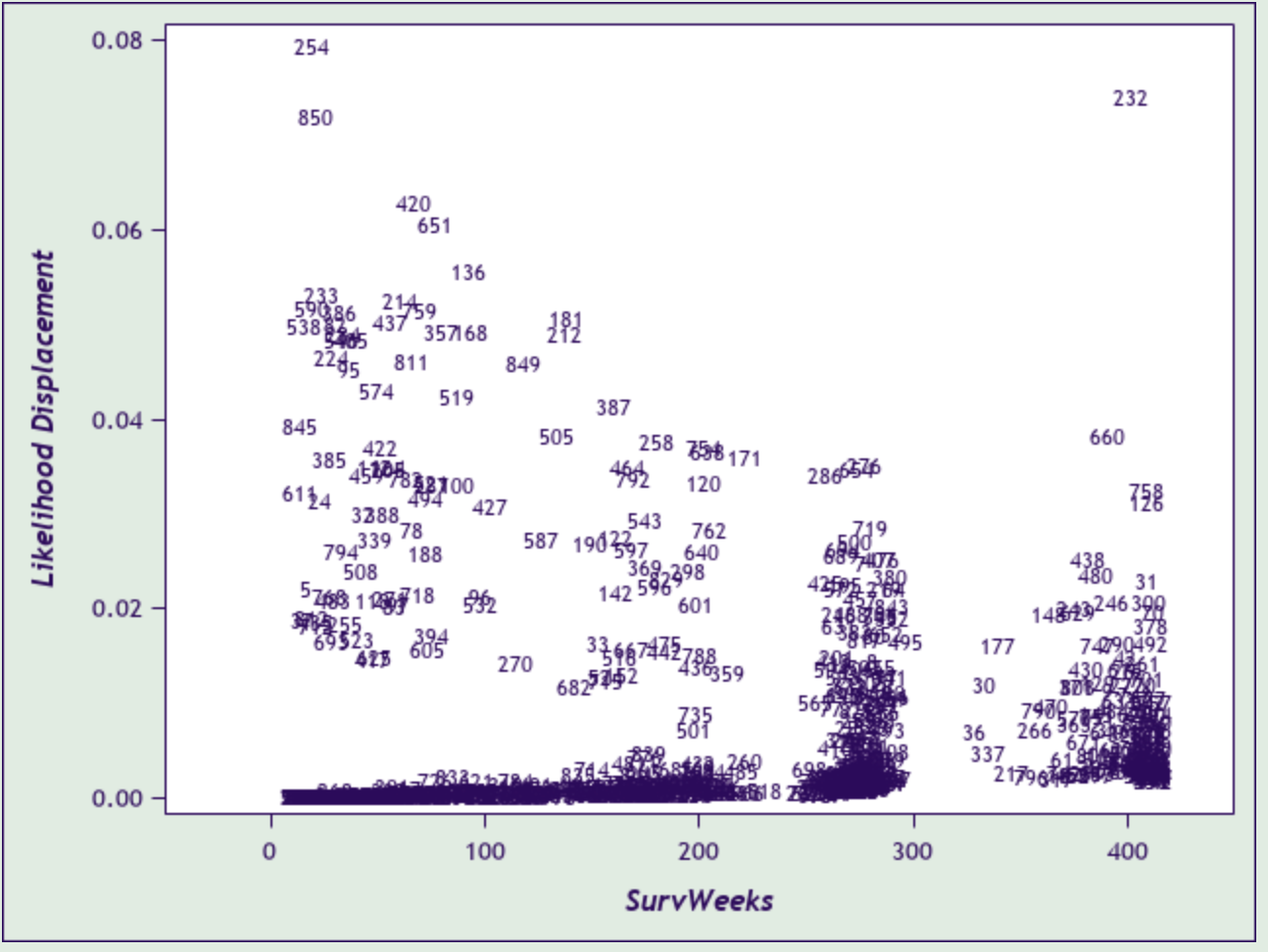
We check the influential case by DFBETAS statistics:



There is no influential case on regression coefficients.

* 1. Influence on Overall Model

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



There is no influential case on overall model.

# SUMMARY

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