Survival Analysis of the Mayo Clinic Primary Biliary Cirrhosis Data

Project Carried out by

Chukwuma Franklin Ofoegbu

S211235

Course: Event History Analysis

Submitted to Professor Dr. Joanna Landmesser

Warsaw University of Life Sciences

Warsaw, May 31, 2023.

Contents

INTRODUCTION

The liver is an organ that is located on the upper right belly. It is a crucial organ, which supports metabolism and engages in secretory functions. The liver is reddish-brown, and is about the size of a football, below the ribcage.

The liver supports immunity, digestion, detoxification, metabolism, vitamin storage, bile production and a host of other functions. The liver is about 2% of an adult body's weight.

The liver is sometimes plagued with some diseases. Some common diseases that the liver is prone to are hepatitis, non-alcoholic fatty liver disease, liver cancer, primary biliary cirrhosis (pbc) etc. The focus of this analysis involves the survival analysis of patients with primary biliary cirrhosis (pbc).

Primary biliary cirrhosis (pbc) also known as primary biliary cholangitis is a liver disease that primarily affects the bile ducts. Biliary means bile ducts, cholangitis means inflammation in the bile ducts. This disease is a chronic and progressive one and can get worse over time. Liver transplant can fix this, but more so, there are drugs which can be used to slow down the progression of the disease.

This study explores the effects of the drug D-penicillmain, in PBC patients. The survival analysis of these patients.

<u>Dataset Description</u>

This data is from the Mayo Clinic trial in primary biliary cirrhosis (PBC) of the liver conducted between 1974 and 1984. A total of 424 PBC patients, referred to Mayo Clinic during that ten-year interval, met eligibility criteria for the randomized placebo-controlled trial of the drug D-penicillamine. The first 312 cases in the data set participated in the randomized trial and contain largely complete data. The additional 112 cases did not participate in the clinical trial but consented to have basic measurements recorded and to be followed for survival. Six of those cases were lost to follow-up shortly after diagnosis, so the data here are on an additional 106 cases as well as the 312 randomized participants.

Table 1 Dataset parameters and description

Parameter	Description

age:	in years
albumin:	serum albumin (g/dl)
alk.phos:	alkaline phosphotase (U/liter)
ascites:	presence of ascites
ast:	aspartate aminotransferase, once called SGOT (U/ml)
bili:	serum bilirunbin (mg/dl)
chol:	serum cholesterol (mg/dl)
copper:	urine copper (ug/day)
edema:	0 no edema, 0.5 untreated or successfully treated
	1 edema despite diuretic therapy
hepato:	presence of hepatomegaly or enlarged liver
id:	case number
platelet:	platelet count
protime:	standardised blood clotting time
sex:	m/f
spiders:	blood vessel malformations in the skin
stage:	histologic stage of disease (needs biopsy)
status:	status at endpoint, 0/1/2 for censored, transplant, dead
	0 = censored;transplant
	1 = death
time:	number of days between registration and the earlier of death,
	transplantion, or study analysis in July, 1986
trt:	1/2/NA for D-penicillmain, placebo, not randomised
trig:	triglycerides (mg/dl)

A new variable named AG is introduced. This is for the age. If age is less than 50 years, then AG = LT50, else age = OV50.

<u>Descriptive Statistics</u>

The picture below shows the descriptive Statistics of the dataset, and they

```
'data.frame':   418 obs. of  20 variables:
 $ id      : int  1 2 3 4 5 6 7 8 9 10 ...
 $ time    : int  400 4500 1012 1925 1504 2503 1832 2466 2400 51 ...
 $ status  : int  2 0 2 2 1 2 0 2 2 2 ...
 $ trt     : int  1 1 1 1 2 2 2 2 1 2 ...
 $ age     : num  58.8 56.4 70.1 54.7 38.1 ...
 $ sex     : Factor w/ 2 levels "m","f": 2 2 1 2 2 2 2 2 2 2 ...
 $ ascites : int  1 0 0 0 0 0 0 0 0 1 ...
 $ hepato  : int  1 1 0 1 1 1 1 0 0 0 ...
 $ spiders : int  1 1 0 1 1 0 0 0 1 1 ...
 $ edema   : num  1 0 0.5 0.5 0 0 0 0 0 1 ...
 $ bili    : num  14.5 1.1 1.4 1.8 3.4 0.8 1 0.3 3.2 12.6 ...
 $ chol    : int  261 302 176 244 279 248 322 280 562 200 ...
 $ albumin : num  2.6 4.14 3.48 2.54 3.53 3.98 4.09 4 3.08 2.74 ...
 $ copper  : int  156 54 210 64 143 50 52 52 79 140 ...
 $ alk.phos: num  1718 7395 516 6122 671 ...
 $ ast     : num  137.9 113.5 96.1 60.6 113.2 ...
 $ trig    : int  172 88 55 92 72 63 213 189 88 143 ...
 $ platelet: int  190 221 151 183 136 NA 204 373 251 302 ...
 $ protime : num  12.2 10.6 12 10.3 10.9 11 9.7 11 11 11.5 ...
 $ stage   : int  4 3 4 4 3 3 3 2 4 ...
```

```
       id             time          status            trt             age            sex          ascites            hepato          spiders
 Min.   :  1.0   Min.   :  41   Min.   :0.0000   Min.   :1.000   Min.   :26.28   m: 44   Min.   :0.00000   Min.   :0.0000   Min.   :0.0000
 1st Qu.:105.2   1st Qu.:1093   1st Qu.:0.0000   1st Qu.:1.000   1st Qu.:42.83   f:374   1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.0000
 Median :209.5   Median :1730   Median :0.0000   Median :1.000   Median :51.00           Median :0.00000   Median :1.0000   Median :0.0000
 Mean   :209.5   Mean   :1918   Mean   :0.8301   Mean   :1.494   Mean   :50.74           Mean   :0.07692   Mean   :0.5128   Mean   :0.2885
 3rd Qu.:313.8   3rd Qu.:2614   3rd Qu.:2.0000   3rd Qu.:2.000   3rd Qu.:58.24           3rd Qu.:0.00000   3rd Qu.:1.0000   3rd Qu.:1.0000
 Max.   :418.0   Max.   :4795   Max.   :2.0000   Max.   :2.000   Max.   :78.44           Max.   :1.00000   Max.   :1.0000   Max.   :1.0000
                                                                                 NA's   :106       NA's   :106       NA's   :106
     edema              bili             chol           albumin          copper          alk.phos           ast              trig           platelet
 Min.   :0.0000   Min.   : 0.300   Min.   : 120.0   Min.   :1.960   Min.   :  4.00   Min.   :  289.0   Min.   : 26.35   Min.   : 33.00   Min.   : 62.0
 1st Qu.:0.0000   1st Qu.: 0.800   1st Qu.: 249.5   1st Qu.:3.243   1st Qu.: 41.25   1st Qu.:  871.5   1st Qu.: 80.60   1st Qu.: 84.25   1st Qu.:188.5
 Median :0.0000   Median : 1.400   Median : 309.5   Median :3.530   Median : 73.00   Median : 1259.0   Median :114.70   Median :108.00   Median :251.0
 Mean   :0.1005   Mean   : 3.221   Mean   : 369.5   Mean   :3.497   Mean   : 97.65   Mean   : 1982.7   Mean   :122.56   Mean   :124.70   Mean   :257.0
 3rd Qu.:0.0000   3rd Qu.: 3.400   3rd Qu.: 400.0   3rd Qu.:3.770   3rd Qu.:123.00   3rd Qu.: 1980.0   3rd Qu.:151.90   3rd Qu.:151.00   3rd Qu.:318.0
 Max.   :1.0000   Max.   :28.000   Max.   :1775.0   Max.   :4.640   Max.   :588.00   Max.   :13862.4   Max.   :457.25   Max.   :598.00   Max.   :721.0
                                   NA's   :134                      NA's   :108       NA's   :106       NA's   :106       NA's   :136       NA's   : 11
    protime          stage
 Min.   : 9.00   Min.   :1.000
 1st Qu.:10.00   1st Qu.:2.000
 Median :10.60   Median :3.000
 Mean   :10.73   Mean   :3.024
 3rd Qu.:11.10   3rd Qu.:4.000
 Max.   :18.00   Max.   :4.000
 NA's   :2       NA's   :6
```

Full Dataset (418 Patients)

```
'data.frame':   312 obs. of  20 variables:
 $ id      : int  1 2 3 4 5 6 7 8 9 10 ...
 $ time    : int  400 4500 1012 1925 1504 2503 1832 2466 2400 51 ...
 $ status  : int  2 0 2 2 1 2 0 2 2 2 ...
 $ trt     : int  1 1 1 1 2 2 2 2 1 2 ...
 $ age     : num  58.8 56.4 70.1 54.7 38.1 ...
 $ sex     : Factor w/ 2 levels "m","f": 2 2 1 2 2 2 2 2 2 2 ...
 $ ascites : int  1 0 0 0 0 0 0 0 0 1 ...
 $ hepato  : int  1 1 0 1 1 1 1 0 0 0 ...
 $ spiders : int  1 1 0 1 1 0 0 0 1 1 ...
 $ edema   : num  1 0 0.5 0.5 0 0 0 0 0 1 ...
 $ bili    : num  14.5 1.1 1.4 1.8 3.4 0.8 1 0.3 3.2 12.6 ...
 $ chol    : int  261 302 176 244 279 248 322 280 562 200 ...
 $ albumin : num  2.6 4.14 3.48 2.54 3.53 3.98 4.09 4 3.08 2.74 ...
 $ copper  : int  156 54 210 64 143 50 52 52 79 140 ...
 $ alk.phos: num  1718 7395 516 6122 671 ...
 $ ast     : num  137.9 113.5 96.1 60.6 113.2 ...
 $ trig    : int  172 88 55 92 72 63 213 189 88 143 ...
 $ platelet: int  190 221 151 183 136 NA 204 373 251 302 ...
 $ protime : num  12.2 10.6 12 10.3 10.9 11 9.7 11 11 11.5 ...
 $ stage   : int  4 3 4 4 3 3 3 3 2 4 ...
```

```
       id             time          status            trt             age            sex          ascites            hepato          spiders
 Min.   :  1.00   Min.   :  41   Min.   :0.0000   Min.   :1.000   Min.   :26.28   m: 36   Min.   :0.00000   Min.   :0.0000   Min.   :0.0000
 1st Qu.: 78.75   1st Qu.:1191   1st Qu.:0.0000   1st Qu.:1.000   1st Qu.:42.24   f:276   1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.0000
 Median :156.50   Median :1840   Median :0.0000   Median :1.000   Median :49.79           Median :0.00000   Median :1.0000   Median :0.0000
 Mean   :156.50   Mean   :2006   Mean   :0.8622   Mean   :1.494   Mean   :50.02           Mean   :0.07692   Mean   :0.5128   Mean   :0.2885
 3rd Qu.:234.25   3rd Qu.:2697   3rd Qu.:2.0000   3rd Qu.:2.000   3rd Qu.:56.71           3rd Qu.:0.00000   3rd Qu.:1.0000   3rd Qu.:1.0000
 Max.   :312.00   Max.   :4556   Max.   :2.0000   Max.   :2.000   Max.   :78.44           Max.   :1.00000   Max.   :1.0000   Max.   :1.0000
                                                                                         NA's   :106       NA's   :106       NA's   :106
     edema              bili             chol           albumin          copper          alk.phos           ast              trig           platelet
 Min.   :0.0000   Min.   : 0.300   Min.   : 120.0   Min.   :1.96   Min.   :  4.00   Min.   :  289.0   Min.   : 26.35   Min.   : 33.00   Min.   : 62.0
 1st Qu.:0.0000   1st Qu.: 0.800   1st Qu.: 249.5   1st Qu.:3.31   1st Qu.: 41.25   1st Qu.:  871.5   1st Qu.: 80.60   1st Qu.: 84.25   1st Qu.:199.8
 Median :0.0000   Median : 1.350   Median : 309.5   Median :3.55   Median : 73.00   Median : 1259.0   Median :114.70   Median :108.00   Median :257.0
 Mean   :0.1106   Mean   : 3.256   Mean   : 369.5   Mean   :3.52   Mean   : 97.65   Mean   : 1982.7   Mean   :124.70   Mean   :261.9
 3rd Qu.:0.0000   3rd Qu.: 3.425   3rd Qu.: 400.0   3rd Qu.:3.80   3rd Qu.:123.00   3rd Qu.: 1980.0   3rd Qu.:151.90   3rd Qu.:151.00   3rd Qu.:322.5
 Max.   :1.0000   Max.   :28.000   Max.   :1775.0   Max.   :4.64   Max.   :588.00   Max.   :13862.4   Max.   :457.25   Max.   :598.00   Max.   :563.0
                                   NA's   :28                      NA's   :2                                           NA's   :30       NA's   :4
    protime          stage
 Min.   : 9.00   Min.   :1.000
 1st Qu.:10.00   1st Qu.:2.000
 Median :10.60   Median :3.000
 Mean   :10.73   Mean   :3.032
 3rd Qu.:11.10   3rd Qu.:4.000
 Max.   :17.10   Max.   :4.000
```

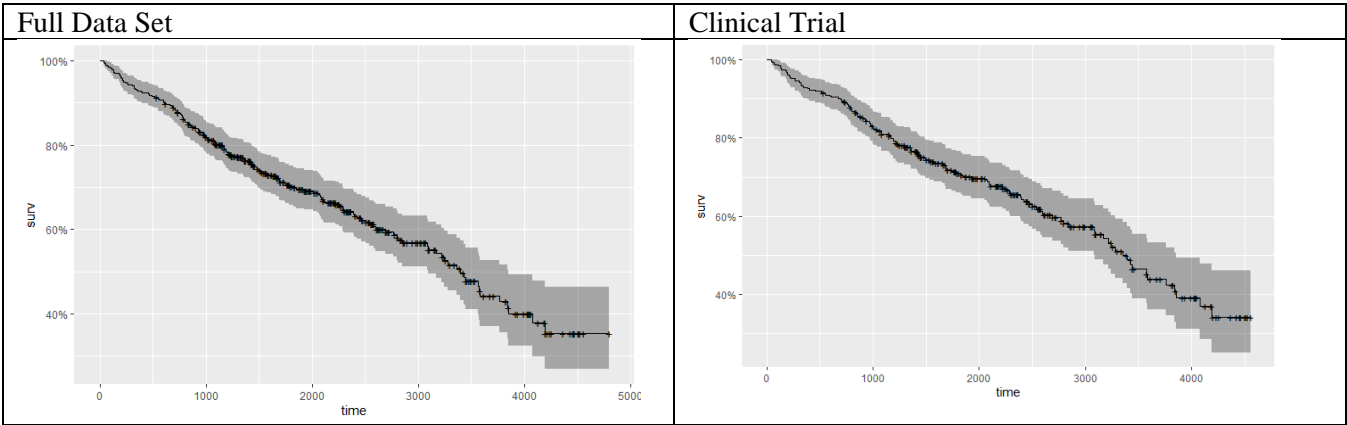Partial Dataset (312 Patients – clinical trial)

Full Data Set	Clinical Trial
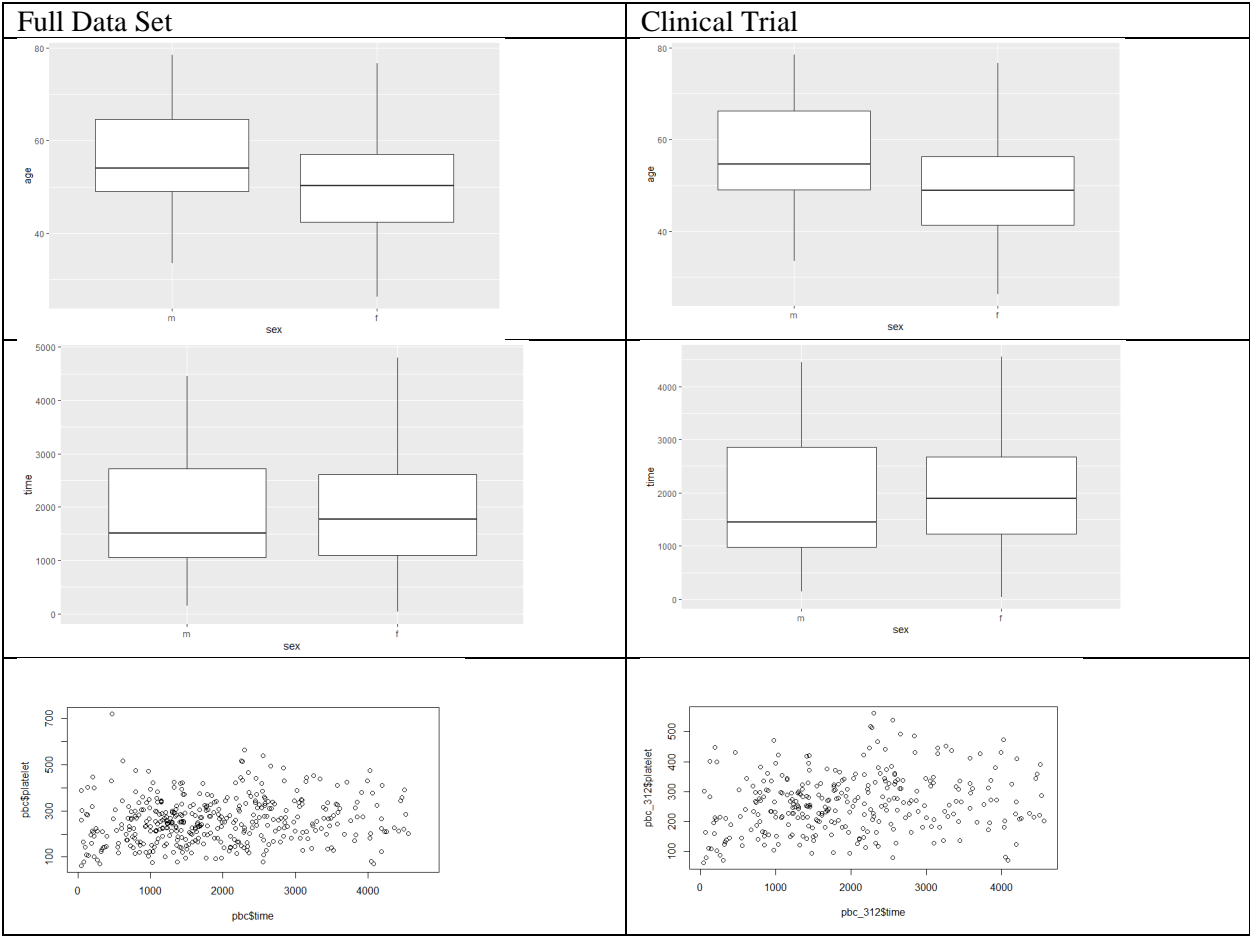
Data Pre-processing

Some missing data was found. In the Full dataset, trt = NA is encoded to be 3 (non-randomised)

| Full Data Set	Clinical Trial
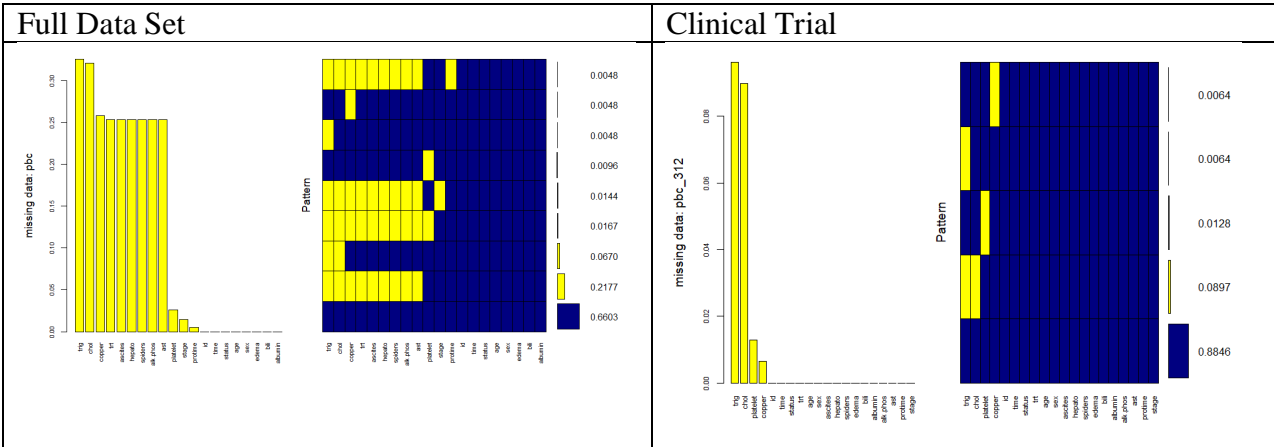
Mice package was used to input missing values using the predictive matching method, so as not to lose vital information via NAs.

METHODS AND RESULTS

Kaplan-Meier: The Kaplan-Meier estimates was carried out for the full and clinical trial datasets and the results are seen as follows:

Full Data Set	Clinical Trial

The figure above depicts the survival probabilities of the Full data set and the clinical trial patients. It is seen that the probability that a patient who participated in the clinical study will survive past 4750 days is 0%. This shows that patients who did not participate in the study have higher times.

Full Data Set	Clinical Trial
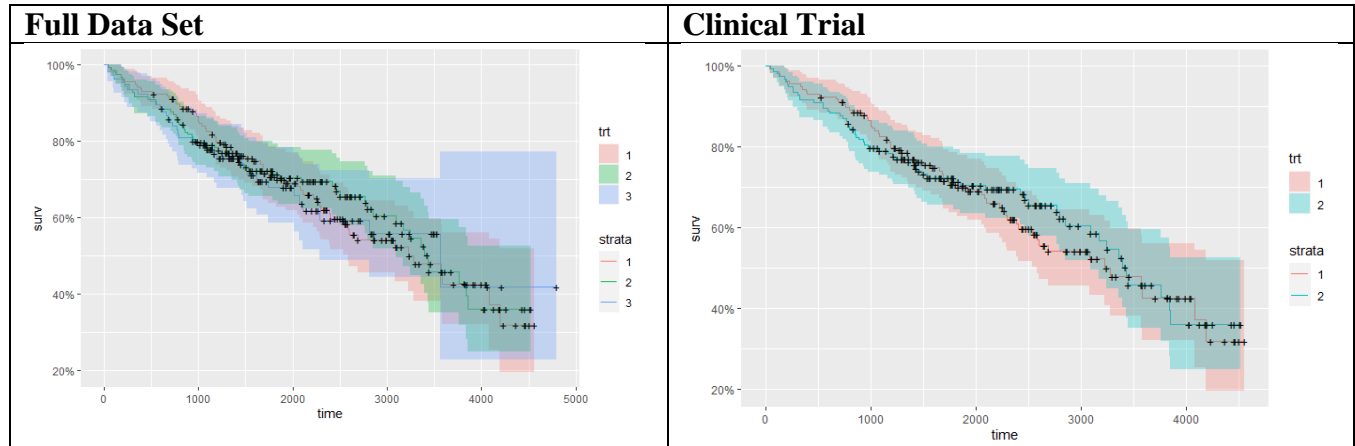
It is seen that the survival probability for PBC patients that did not participate in the clinical trials (strata 3) is much higher than the survival times for the patients who participated in the trials (strata 1 and strata 2). Strata 3 PBC patients will survive past 4600 days at a probability of 40%, while none of the patients in the clinical trial survived past 4600 days.

Furthermore, there seem to be almost no effect of the clinical drug (D-penicillamine), as the survival curves of both the treatment (strata 1) and placebo (strata 2), are interchanging. Although, noticeably, from day 2000 of the trial, placebo pbc patients begin to have higher probabilities of survival than the treatment patients, almost up until day 3000. For instance, the probability of survival of a clinical trial patient to survive past 2500 days is 60% for the treatment patient, and approximately 67% for the placebo patient.

Towards the end of the curve, the probability of survival of a clinical trial patient to survive past 4500 days is approximately 32% for the treatment patient, and approximately 36% for the placebo patient. There is almost no effect of the drug, or it may have negative effects on PBC patients.
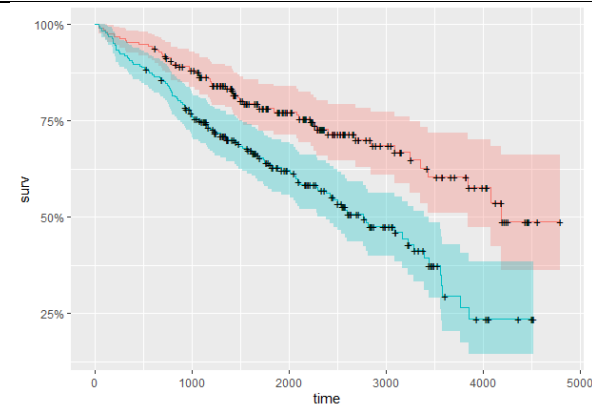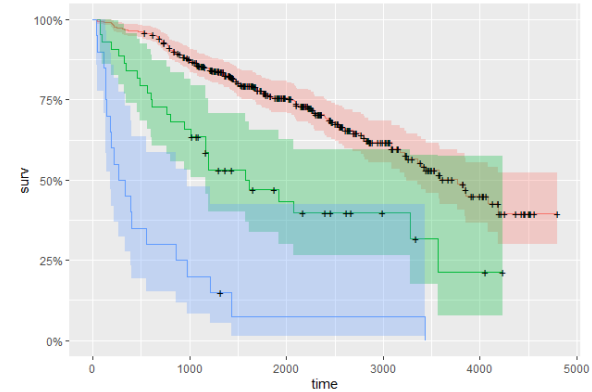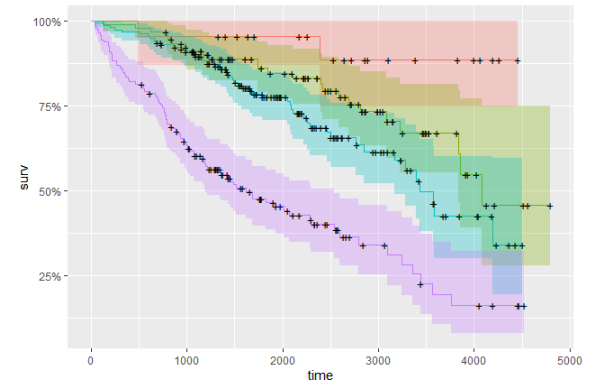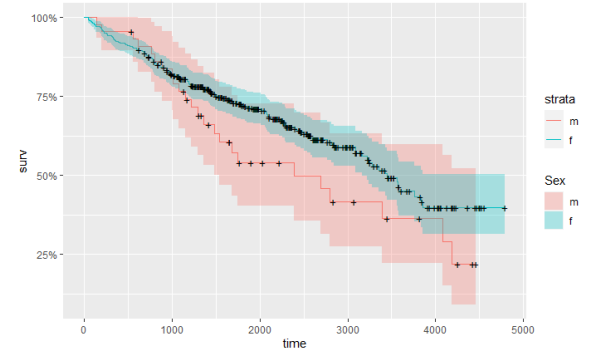
Full Data Set	Clinical Trial
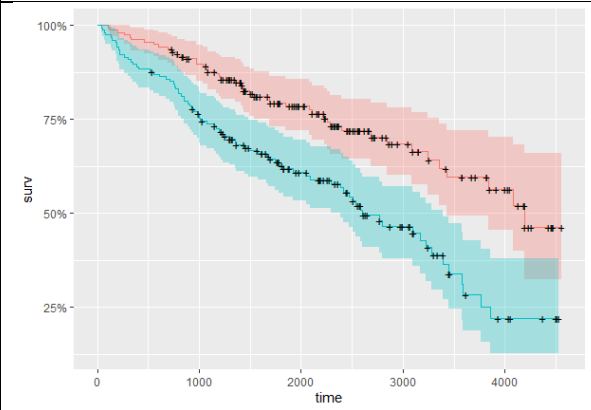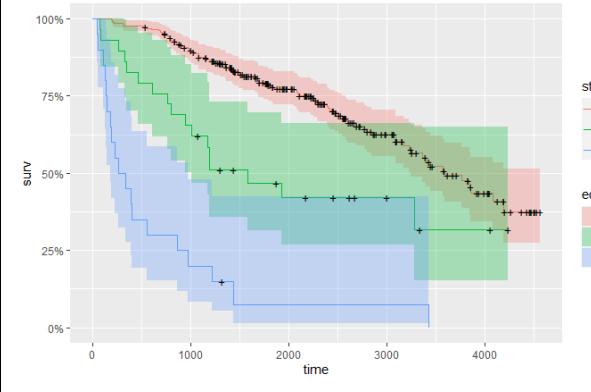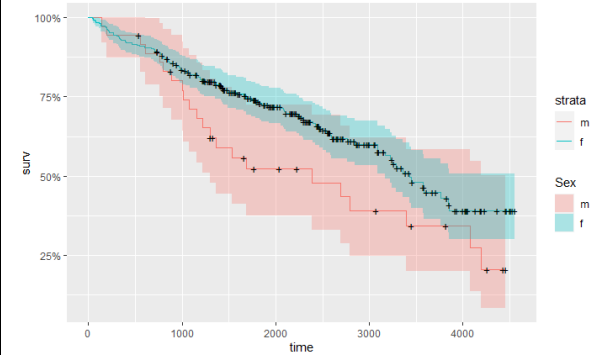
In the first row, it is seen that female pbc patients have a higher survival probability than male pbc patients. The curves for the male pbc patients drop faster, when compared to the females pbc curves.

In the second row, it is seen that stage 4 pbc patients have the least survival probabilities. In terms of their survival probabilities and times, Stage 1 > Stage 2 > Stage 3 > Stage 4. Stage 1 pbc patients have the best survival probabilities.

Edema is fluid retention/swelling by the body tissues. It is a phenomenon seen in pbc patients. Patients with no edema survive better than the patients (strata 0) with edema (strata 1, and 2). Strata 2 (edema, despite **diuretic therapy**[1]) patients have the lowest survival rate.

There is a huge difference in the survival times for patients aged over 50, and those less than 50. PBC patients that are aged less than 50 have a higher chance of survival than those below 50 years.

Log Rank Test for some selected groups

Hypothesis:

$$H_0: There\ is\ no\ difference\ in\ the\ survival\ curves\ (for\ all\ time)$$

$$H_1: There\ is\ difference\ in\ the\ survival\ curves\ (for\ all\ time)$$

decision rule is to retain H_0 if $p > 0.05$, else H_0 is rejected.

Full Data Set (Log rank test)	Clinical Trial (log rank test)	Conclusion

Call:
survdiff(formula = Surv(time, status) ~ trt, data = data)
 	
    N Observed Expected (O-E)^2/E (O-E)^2/V
trt=1 158   65   64.0   0.0141   0.0235
trt=2 154   60   62.4   0.0892   0.1461
trt=3 106   36   34.6   0.0573   0.0737
 	
 Chisq= 0.2 on 2 degrees of freedom, p= 0.9
	Call:
	survdiff(formula = Surv(time, status) ~ trt, data = data)
	
	    N Observed Expected (O-E)^2/E (O-E)^2/V
	trt=1 158 65 63.2 0.0502 0.102
	trt=2 154 60 61.8 0.0513 0.102
	
	 Chisq= 0.1 on 1 degrees of freedom, p= 0.7
		$p > 0.05$, Ho is retained. No difference between the curves for **trt.**

Call:
survdiff(formula = Surv(time, status) ~ sex, data = data)
 	
    N Observed Expected (O-E)^2/E (O-E)^2/V
sex=m 44   24   17.3   2.640   2.98
sex=f 374 137 143.7   0.317   2.98
 	
 Chisq= 3 on 1 degrees of freedom, p= 0.08
	Call:
	survdiff(formula = Surv(time, status) ~ sex, data = data)
	
	    N Observed Expected (O-E)^2/E (O-E)^2/V
	sex=m 36 22 14.6 3.728 4.27
	sex=f 276 103 110.4 0.494 4.27
	
	 Chisq= 4.3 on 1 degrees of freedom, p= 0.04
		$p < 0.05$, Ho is rejected. There is difference between the curves for **sex**.

survdiff(formula = Surv(time, status) ~ stage, data = data)
 	
    N Observed Expected (O-E)^2/E (O-E)^2/V
stage=1 22   2   12.4   8.76   9.60
stage=2 93   24   44.7   9.56 13.39
stage=3 157   49   62.4   2.89   4.74
stage=4 146   86   41.5 47.81 65.48
 	
 Chisq= 70.6 on 3 degrees of freedom, p= 3e-15
	Call:
	survdiff(formula = Surv(time, status) ~ stage, data = data)
	
	    N Observed Expected (O-E)^2/E (O-E)^2/V
	stage=1 16 1 9.9 8.00 8.78
	stage=2 67 16 32.3 8.22 11.17
	stage=3 120 43 51.2 1.30 2.22
	stage=4 109 65 31.6 35.17 47.84
	
	 Chisq= 53.8 on 3 degrees of freedom, p= 1e-11
		$p < 0.05$, Ho is rejected. There is difference between the curves for **stage**.

---

[1] Diuretic drugs are widely used for the treatment of patients with edema. (Source: https://www.nejm.org/doi/full/10.1056/nejm199808063390607)

```Call:``` ```survdiff(formula = Surv(time, status) ~ edema, data = data)```
 	
 ``` N Observed Expected (O-E)^2/E (O-E)^2/V```
 ```edema=0   354   116   145.47   5.97   62.3```
 ```edema=0.5   44   26   13.05   12.84   14.0```
 ```edema=1   20   19   2.47 110.44   113.1```
 	
 ``` Chisq= 131 on 2 degrees of freedom, p= <2e-16```
	```Call:``` ```survdiff(formula = Surv(time, status) ~ edema, data = data)```
	
	 ``` N Observed Expected (O-E)^2/E (O-E)^2/V```
	 ```edema=0 263 89 113.38 5.24 56.76```
	 ```edema=0.5 29 17 9.20 6.61 7.15```
	 ```edema=1 20 19 2.42 113.48 117.00```
	
	 ``` Chisq= 127 on 2 degrees of freedom, p= <2e-16```
		$p < 0.05$, Ho is rejected. There is difference between the curves for sex **edema**

```Call:``` ```survdiff(formula = Surv(time, status) ~ AG, data = vet)```
 	
 ``` N Observed Expected (O-E)^2/E (O-E)^2/V```
 ```AG=LT50 195   55   81.4   8.56   17.5```
 ```AG=OV50 223   106   79.6   8.75   17.5```
 	
 ``` Chisq= 17.5 on 1 degrees of freedom, p= 3e-05```
	```Call:``` ```survdiff(formula = Surv(time, status) ~ AG, data = vet)```
	
	 ``` N Observed Expected (O-E)^2/E (O-E)^2/V```
	 ```AG=LT50 158 46 68.7 7.50 16.8```
	 ```AG=OV50 154 79 56.3 9.16 16.8```
	
	 ``` Chisq= 16.8 on 1 degrees of freedom, p= 4e-05```
		$p < 0.05$, Ho is rejected. There is difference between the curves for **AG**

Cox PH Model

Full Data Set	Clinical Trial

```coxph(formula = Surv(time, status) ~ trt + sex + ascites + hepa```
 ```   spiders + edema + bili + chol + albumin + copper + alk.phos```
 ```   ast + trig + platelet + protime + stage + AG, data = vet)```
 	
 ``` n= 418, number of events= 161```
 	
 ```   coef exp(coef) se(coef)   z Pr(>|z|)```
 ```trt   1.129e-02 1.011e+00 1.096e-01   0.103 0.917938```
 ```sexf   -2.897e-01 7.485e-01 2.450e-01 -1.183 0.236923```
 ```ascites   2.482e-01 1.282e+00 2.974e-01   0.835 0.403970```
 ```hepato   9.320e-02 1.098e+00 2.048e-01   0.455 0.649021```
 ```spiders   9.942e-02 1.105e+00 1.956e-01   0.508 0.611224```
 ```edema   9.061e-01 2.475e+00 3.032e-01   2.988 0.002808 **```
 ```bili   8.290e-02 1.086e+00 2.023e-02   4.098 4.17e-05 ***```
 ```chol   6.764e-04 1.001e+00 3.690e-04   1.833 0.066786 .```
 ```albumin -6.316e-01 5.318e-01 2.311e-01 -2.733 0.006282 **```
 ```copper   2.755e-03 1.003e+00 8.864e-04   3.109 0.001880 **```
 ```alk.phos -3.463e-05 1.000e+00 3.644e-05 -0.950 0.341872```
 ```ast   2.934e-03 1.003e+00 1.687e-03   1.740 0.081920 .```
 ```trig   -1.065e-03 9.989e-01 1.073e-03 -0.992 0.321149```
 ```platelet 5.887e-04 1.001e+00 9.532e-04   0.618 0.536806```
 ```protime   1.956e-01 1.216e+00 6.865e-02   2.850 0.004375 **```
 ```stage   4.548e-01 1.576e+00 1.309e-01   3.473 0.000514 ***```
 ```AGOV50   7.289e-01 2.073e+00 1.881e-01   3.875 0.000107 ***```
 ```---```
 ```Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1```
	```coxph(formula = Surv(time, status) ~ trt + sex + ascites + hepato +```
	 ``` spiders + edema + bili + chol + albumin + copper + alk.phos +```
	 ``` ast + trig + platelet + protime + stage + AG, data = vet)```
	
	 ``` n= 312, number of events= 125```
	
	 ``` coef exp(coef) se(coef) z Pr(>|z|)```
	 ```trt -1.069e-01 8.986e-01 1.920e-01 -0.557 0.577544```
	 ```sexf -3.800e-01 6.839e-01 2.733e-01 -1.390 0.164466```
	 ```ascites 3.149e-01 1.370e+00 3.350e-01 0.940 0.347284```
	 ```hepato 1.936e-01 1.214e+00 2.420e-01 0.800 0.423662```
	 ```spiders 9.815e-02 1.103e+00 2.286e-01 0.429 0.667650```
	 ```edema 8.531e-01 2.347e+00 3.380e-01 2.524 0.011606 *```
	 ```bili 7.706e-02 1.080e+00 2.335e-02 3.300 0.000965 ***```
	 ```chol 4.804e-04 1.000e+00 4.200e-04 1.144 0.252759```
	 ```albumin -7.597e-01 4.678e-01 2.737e-01 -2.776 0.005506 **```
	 ```copper 2.591e-03 1.003e+00 1.076e-03 2.409 0.015991 *```
	 ```alk.phos -2.121e-05 1.000e+00 3.859e-05 -0.550 0.582470```
	 ```ast 4.718e-03 1.005e+00 1.770e-03 2.666 0.007672 **```
	 ```trig -7.148e-04 9.993e-01 1.230e-03 -0.581 0.561079```
	 ```platelet 3.436e-04 1.000e+00 1.166e-03 0.295 0.767937```
	 ```protime 2.686e-01 1.308e+00 9.379e-02 2.863 0.004191 **```
	 ```stage 3.578e-01 1.430e+00 1.546e-01 2.314 0.020669 *```
	 ```AGOV50 7.316e-01 2.078e+00 2.101e-01 3.482 0.000498 ***```
	 ```---```
	 ```Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1```

```
          exp(coef) exp(-coef) lower .95 upper .95                    exp(coef) exp(-coef) lower .95 upper .95
trt         1.0114     0.9888    0.8159    1.2536       trt            0.8986     1.1128    0.6168    1.3091
sexf        0.7485     1.3360    0.4631    1.2097       sexf           0.6839     1.4623    0.4002    1.1685
ascites     1.2817     0.7802    0.7156    2.2956       ascites        1.3701     0.7299    0.7105    2.6420
hepato      1.0977     0.9110    0.7348    1.6398       hepato         1.2136     0.8240    0.7553    1.9503
spiders     1.1045     0.9054    0.7528    1.6205       spiders        1.1031     0.9065    0.7048    1.7266
edema       2.4747     0.4041    1.3658    4.4837       edema          2.3468     0.4261    1.2100    4.5517
bili        1.0864     0.9204    1.0442    1.1304       bili           1.0801     0.9258    1.0318    1.1307
chol        1.0007     0.9993    1.0000    1.0014       chol           1.0005     0.9995    0.9997    1.0013
albumin     0.5318     1.8805    0.3381    0.8365       albumin        0.4678     2.1377    0.2736    0.7999
copper      1.0028     0.9972    1.0010    1.0045       copper         1.0026     0.9974    1.0005    1.0047
alk.phos    1.0000     1.0000    0.9999    1.0000       alk.phos       1.0000     1.0000    0.9999    1.0001
ast         1.0029     0.9971    0.9996    1.0063       ast            1.0047     0.9953    1.0013    1.0082
trig        0.9989     1.0011    0.9968    1.0010       trig           0.9993     1.0007    0.9969    1.0017
platelet    1.0006     0.9994    0.9987    1.0025       platelet       1.0003     0.9997    0.9981    1.0026
protime     1.2161     0.8223    1.0630    1.3912       protime        1.3081     0.7645    1.0884    1.5720
stage       1.5759     0.6346    1.2192    2.0369       stage          1.4301     0.6992    1.0563    1.9363
AGOV50      2.0727     0.4825    1.4337    2.9967       AGOV50         2.0783     0.4812    1.3768    3.1373

Concordance= 0.838  (se = 0.015 )                      Concordance= 0.858  (se = 0.016 )
Likelihood ratio test= 231.3  on 17 df,   p=<2e-16     Likelihood ratio test= 203.7  on 17 df,   p=<2e-16
wald test          = 245.7  on 17 df,   p=<2e-16       wald test          = 209.7  on 17 df,   p=<2e-16
Score (logrank) test = 359.4  on 17 df,   p=<2e-16     Score (logrank) test = 335.4  on 17 df,   p=<2e-16
```

Hazard increasing over time: AGOV50, ascites, bili, edema, platelet, protime, stage.

Hazard decreasing over time: albumin, sexf,

While the rest variables have little or no changes/impact on the survival over time.

Patients aged over 50, with edema and high ascites have the biggest NEGATIVE impact on the survival of patients.

$$h(t,X) = h_o(t)\exp\left[\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_{16} X_{16} + \beta_{17} X_{17}\right]$$

Coef (βi) : Full Data set

[1.011, 0.749,1.282, 1.098, 1.105, 2.475, 1.086, 1.001, 0.532, 1.003, 1.000, 1.003, 0.999, 1.001, 1.216, 1.576, 2.073] = [$\beta 1$, …, $\beta 17$]

Coef (βi) : Full Data set

[0.8986,0.6839,1.3701,1.2136,1.1031, 2.3468,1.0801,1.0005,0.4678,1.0026,1,1.0047,0.9993,1.0003,1.3081, 1.4301, 2.0783] = [$\beta 1$, …, $\beta 17$]

The PH assumption is that the Hazard Ratio (HR) is constant over time. To verify this, we use some methods:

PH is violated when $p < 0.05$

1. Goodness of fit

Full Data Set					Clinical Trial			
	chisq	df	p			chisq	df	p
trt	4.7431	1	0.0294		trt	0.67159	1	0.4125
sex	1.5224	1	0.2173		sex	0.14798	1	0.7005
ascites	0.2863	1	0.5926		ascites	0.07362	1	0.7861
hepato	0.0056	1	0.9404		hepato	0.11840	1	0.7308
spiders	0.1765	1	0.6744		spiders	0.07775	1	0.7804
edema	2.3962	1	0.1216		edema	4.06641	1	0.0437
bili	7.7792	1	0.0053		bili	9.32266	1	0.0023
chol	7.6655	1	0.0056		chol	8.07012	1	0.0045
albumin	1.4150	1	0.2342		albumin	0.04581	1	0.8305
copper	3.3978	1	0.0653		copper	0.20876	1	0.6477
alk.phos	3.1588	1	0.0755		alk.phos	1.06129	1	0.3029
ast	2.2378	1	0.1347		ast	1.59974	1	0.2059
trig	1.9113	1	0.1668		trig	6.32887	1	0.0119
platelet	1.9406	1	0.1636		platelet	1.42637	1	0.2324
protime	9.5152	1	0.0020		protime	5.31239	1	0.0212
stage	3.4860	1	0.0619		stage	3.32735	1	0.0681
AG	0.3541	1	0.5518		AG	0.00014	1	0.9906
GLOBAL	32.7547	17	0.0121		GLOBAL	25.78366	17	0.0785

Variables: edema (clinical trial), bili, chol, trig and protime all violate the PH assumption

2. Log-log plot
proportional hazards assumption is violated when the log-log survival curves are not parallel.
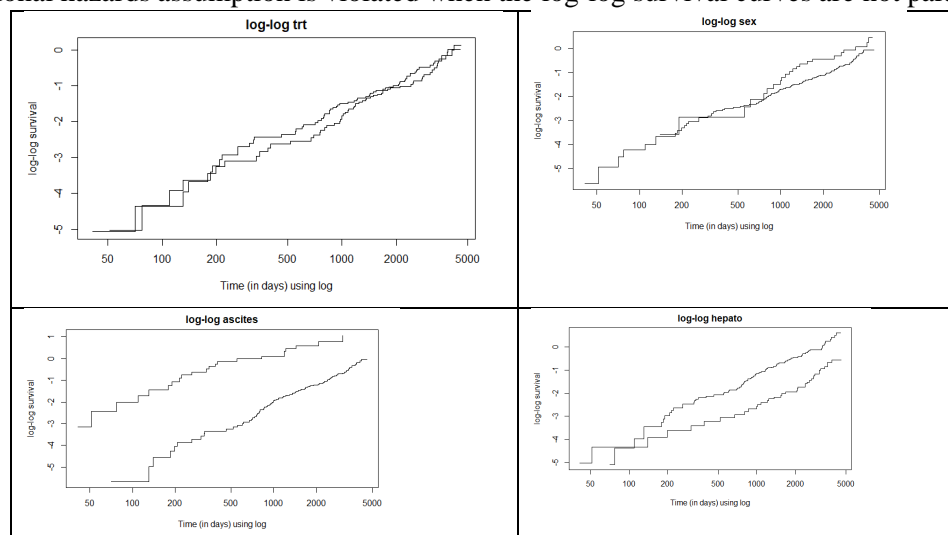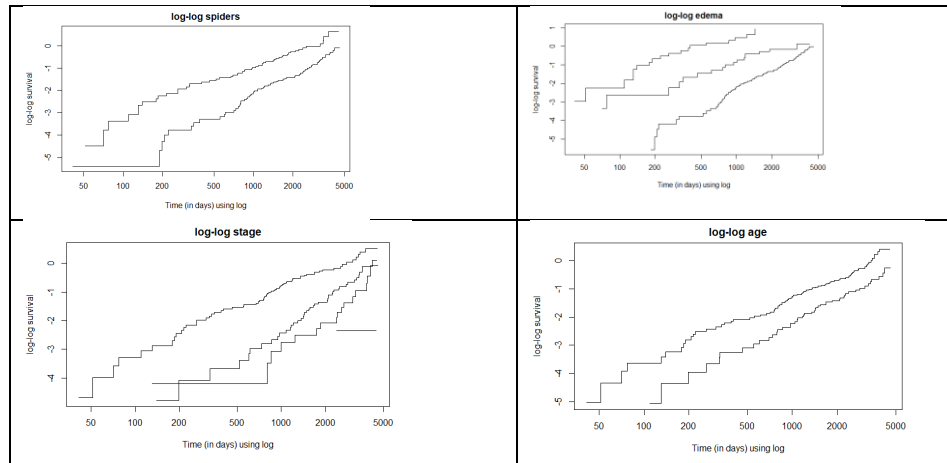
Parametric Models

Weibull Model

```
                Value Std. Error     z        p
(Intercept)   1.02e+01   9.87e-01  10.32 < 2e-16
trt           6.76e-02   1.15e-01   0.59 0.55563
sexf          2.04e-01   1.65e-01   1.24 0.21643
ascites      -2.73e-01   1.97e-01  -1.39 0.16445
hepato       -1.12e-01   1.44e-01  -0.78 0.43631
spiders      -4.94e-02   1.37e-01  -0.36 0.71849
edema        -5.44e-01   1.96e-01  -2.77 0.00564
bili         -4.75e-02   1.30e-02  -3.67 0.00024
chol         -2.84e-04   2.48e-04  -1.15 0.25178
albumin       4.09e-01   1.59e-01   2.57 0.01014
copper       -1.65e-03   6.38e-04  -2.58 0.00988
alk.phos      6.11e-06   2.29e-05   0.27 0.79012
ast          -2.90e-03   1.07e-03  -2.72 0.00662
trig          8.04e-04   7.69e-04   1.05 0.29576
platelet     -2.26e-04   6.99e-04  -0.32 0.74603
protime      -1.70e-01   5.67e-02  -3.00 0.00271
stage        -2.22e-01   9.26e-02  -2.40 0.01655
AGOV50       -4.38e-01   1.28e-01  -3.42 0.00062
Log(scale)   -5.10e-01   7.20e-02  -7.09 1.4e-12

Scale= 0.6

Weibull distribution
Loglik(model)= -1083.7   Loglik(intercept only)= -1188.8
        Chisq= 210.09 on 17 degrees of freedom, p= 2.7e-35
```

Exponential Model

```
                Value Std. Error     z        p
(Intercept)   1.20e+01   1.54e+00   7.76 8.4e-15
trt           6.44e-02   1.87e-01   0.34 0.7313
sexf          3.80e-01   2.63e-01   1.44 0.1491
ascites      -3.55e-01   3.19e-01  -1.11 0.2664
hepato       -2.04e-01   2.35e-01  -0.87 0.3848
spiders      -1.26e-01   2.23e-01  -0.57 0.5715
edema        -6.18e-01   3.30e-01  -1.87 0.0611
bili         -4.91e-02   2.29e-02  -2.15 0.0317
chol         -4.67e-04   4.08e-04  -1.15 0.2520
albumin       4.63e-01   2.55e-01   1.82 0.0693
copper       -1.74e-03   1.02e-03  -1.70 0.0888
alk.phos     -2.97e-05   3.71e-05  -0.80 0.4235
ast          -4.02e-03   1.77e-03  -2.28 0.0228
trig          7.39e-04   1.36e-03   0.54 0.5858
platelet      2.78e-04   1.14e-03   0.24 0.8067
protime      -2.77e-01   8.85e-02  -3.13 0.0017
stage        -2.97e-01   1.49e-01  -2.00 0.0455
AGOV50       -6.38e-01   2.08e-01  -3.07 0.0022

Scale fixed at 1

Exponential distribution
Loglik(model)= -1103.7   Loglik(intercept only)= -1189.8
        Chisq= 172.33 on 17 degrees of freedom, p= 9.7e-28
```

Lognormal Model

```
                Value Std. Error     z        p
(Intercept)   1.08e+01   1.12e+00   9.62 < 2e-16
trt          -4.68e-02   1.26e-01  -0.37 0.71030
sexf          2.56e-01   1.88e-01   1.36 0.17424
ascites      -4.97e-01   2.44e-01  -2.04 0.04132
hepato       -1.02e-01   1.48e-01  -0.69 0.49068
spiders      -2.18e-01   1.47e-01  -1.49 0.13691
edema        -6.51e-01   2.62e-01  -2.49 0.01285
bili         -4.19e-02   1.77e-02  -2.37 0.01794
chol         -2.18e-04   3.03e-04  -0.72 0.47182
albumin       3.44e-01   1.68e-01   2.05 0.04034
copper       -2.03e-03   7.79e-04  -2.60 0.00922
alk.phos     -5.38e-06   2.65e-05  -0.20 0.83927
ast          -3.31e-03   1.25e-03  -2.66 0.00791
trig          7.26e-04   1.06e-03   0.69 0.49173
platelet      1.80e-04   7.16e-04   0.25 0.80113
protime      -2.00e-01   6.76e-02  -2.96 0.00309
stage        -2.47e-01   9.35e-02  -2.64 0.00818
AGOV50       -4.62e-01   1.37e-01  -3.38 0.00073
Log(scale)   -1.78e-01   6.51e-02  -2.74 0.00615

Scale= 0.837

Log Normal distribution
Loglik(model)= -1081.1   Loglik(intercept only)= -1193.3
        Chisq= 224.44 on 17 degrees of freedom, p= 3.3e-38
```

Loglogistic Model

```
                Value Std. Error     z        p
(Intercept)   9.73e+00   1.05e+00   9.29 < 2e-16
trt          -4.11e-02   1.22e-01  -0.34 0.73677
sexf          2.84e-01   1.75e-01   1.62 0.10484
ascites      -4.81e-01   2.35e-01  -2.05 0.04044
hepato       -1.25e-01   1.41e-01  -0.89 0.37584
spiders      -1.60e-01   1.42e-01  -1.12 0.26062
edema        -5.05e-01   2.57e-01  -1.96 0.04965
bili         -3.86e-02   1.63e-02  -2.36 0.01819
chol         -2.18e-04   2.74e-04  -0.79 0.42702
albumin       4.87e-01   1.66e-01   2.94 0.00329
copper       -2.42e-03   7.16e-04  -3.38 0.00073
alk.phos     -2.44e-08   2.60e-05   0.00 0.99925
ast          -2.73e-03   1.15e-03  -2.37 0.01759
trig          7.54e-04   1.04e-03   0.72 0.46958
platelet      4.98e-04   6.97e-04   0.72 0.47451
protime      -1.88e-01   6.34e-02  -2.97 0.00300
stage        -1.93e-01   8.77e-02  -2.20 0.02754
AGOV50       -4.57e-01   1.29e-01  -3.54 0.00040
Log(scale)   -8.01e-01   7.38e-02 -10.85 < 2e-16

Scale= 0.449

Log logistic distribution
Loglik(model)= -1078.1   Loglik(intercept only)= -1190.6
        Chisq= 225.11 on 17 degrees of freedom, p= 2.4e-38
```

Best Model Using the AIC and LLC

```
Model selection based on AICc:

              K    AICc Delta_AICc AICcWt Cum.Wt       LL
Loglogistic  19 2196.76      0.00   0.95   0.95 -1078.08
Lognormal    19 2202.73      5.96   0.05   1.00 -1081.06
Weibull      19 2208.02     11.25   0.00   1.00 -1083.71
Exponential  18 2245.70     48.94   0.00   1.00 -1103.68
```

Model	K	AICc	Delta_AICc	AICcWt	Cum.Wt	LL
Loglogistic	19	2196.76	0	0.95	0.95	-1078.08
Lognormal	19	2202.73	5.96	0.05	1	-1081.06
Weibull	19	2208.02	11.25	0	1	-1083.71
Exponential	18	2245.7	48.94	0	1	-1103.68

The best model is the Loglogistic model.

The best model is the loglogistic model.