**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

RESULTS AND METHODS

CONCLUSION

# 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.

Dataset Description

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.

Descriptive Statistics

The picture below shows the descriptive Statistics of the dataset.

|  |
| --- |
|  |
|  |

Full Dataset (418 Patients)

|  |
| --- |
|  |
|  |

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 |
| A picture containing text, screenshot, diagram, plot  Description automatically generated | A picture containing text, diagram, screenshot, plot  Description automatically generated |

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]](#footnote-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:

 decision rule is to retain H0 if  p > 0.05, else H0 is rejected.

|  |  |  |
| --- | --- | --- |
| Full Data Set (Log rank test) | Clinical Trial (log rank test) | Conclusion |
|  |  | p > 0.05, Ho is retained. No difference between the curves for **trt.** |
|  |  | p < 0.05, Ho is rejected. There is difference between the curves for **sex**. |
|  |  | p < 0.05, Ho is rejected. There is difference between the curves for **stage**. |
|  |  | p < 0.05, Ho is rejected. There is difference between the curves for sex **edema** |
|  |  | p < 0.05, Ho is rejected. There is difference between the curves for **AG** |

Cox PH Model

|  |  |
| --- | --- |
| Full Data Set | Clinical Trial |
|  |  |
|  |  |
|  |  |

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.

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] = [β1, …, β17]

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] = [β1, …, β17]

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | chisq | df | p |  |
| trt | 0.6240 | 1 | 0.4297 |  |
| sex | 0.1730 | 1 | 0.6775 |  |
| ascites | 0.0338 | 1 | 0.8542 |  |
| hepato | 0.1140 | 1 | 0.7359 |  |
| spiders | 0.0893 | 1 | 0.7651 |  |
| edema | 3.9100 | 1 | 0.0479 | violates PH |
| bili | 9.8700 | 1 | 0.0017 | violates PH |
| chol | 9.2300 | 1 | 0.0024 | violates PH |
| albumin | 0.0718 | 1 | 0.7887 |  |
| copper | 0.2080 | 1 | 0.6482 |  |
| alk.phos | 1.1700 | 1 | 0.2803 |  |
| ast | 1.6800 | 1 | 0.1945 |  |
| trig | 5.0400 | 1 | 0.0247 | violates PH |
| platelet | 1.1200 | 1 | 0.2893 |  |
| protime | 5.5300 | 1 | 0.0187 | violates PH |
| stage | 3.3700 | 1 | 0.0665 |  |
| AG | 0.0000 | 1 | 0.9996 |  |
| GLOBAL | 26.4000 | 17 | 0.0673 |  |

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

1. Log-log plot

proportional hazards assumption is violated when the log-log survival curves are not parallel.

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

Parametric Models

|  |  |
| --- | --- |
| Weibull Model | Exponential Model |
| Lognormal Model | Loglogistic Model |

Best Model Using the AIC and LLC

A picture containing text, screenshot, font, line

Description automatically generated

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 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.

# CONCLUSION

In conclusion, it is seen that the treatment drug (d-penicillamine) has almost no visible effects on PBC patients, further supported by these articles (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8846335/> and <https://pubmed.ncbi.nlm.nih.gov/3905561/> ) which says that PBC has negative and no effects respectively. As the stage of PBC worsens, the survival rate becomes also worse. The age of PBC patients (>50 years), as the PBC patient becomes older, his survival rate decreases (Older PBC patients die quicker than younger ones).

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