Primary Biliary Cholangitis:

An Analysis Using Statistical Methods

Abraham Hussein, Abraham Matur Achuil, Simran Bhattarai, Thomas Wunsch

April 15, 2019

Contents

1 Introduction

1.1 Context of Primary Biliary Cholangitis and Study

This replication study is on the survival analysis of Primary Biliary Cholangitis (PBC). PBC is a chronic, autoimmune disease in which your immune system causes progressive damage to the bile ducts in the liver and eventually destroys them. When the bile ducts are compromised, bile can back up in the liver and may lead to irreversible scarring of liver tissue called Cirrhosis. There is no cure for PBC, only programs to slow the progression of the disease. Liver transplantation is considered to be potentially life-saving for patients with advanced or end-stage PBC. A model that predicts the survival probability of a patient can aid in the liver matching and selection process. Such a model can help predict which patient is in more urgent need of a transplant and their position in the transplantation waiting list can be altered accordingly. In 1989 the authors of the paper "Prognosis in Primary Biliary Cirrhosis: Model for Decision Making"[1] created such a model. The goal of this investigation is to replicate the analysis done in this paper.

1.2 Data-Set used in Model

The data set used in this study was initially gathered from either of two double-blind, placebo controlled, randomized clinical trials performed by the Mayo clinic from 1974 to 1986[2]. The original goal of the study was to test the effectiveness of the drug D-penicillamine in treating PBC. Patients were accrued over the first 10 years of the study and the study concluded in 1989. Over that time period 424 patients with PBC were eligible to enter the study, but only 312 patients volunteered to enter. Patients that underwent transplants were considered censored at the time of transplant. All participant deaths during the study were considered as failure for this the purpose of this analysis. At the conclusion of the trial in 1986, 125 participants had died, 19 had undergone transplants, 8 had been lost to follow up, and 160 were still alive.

When a patient entered the study they had 45 biomarkers for PBC measured and measurements were repeated yearly. The authors of the original paper wished to create a predictive survival model for PBC based on biomarkers that could be measured non-invasively, so they chose variables for their model from a reduced set of 12 biomarkers. These non invasive factors used in their analysis included: age, serum levels of albumin in gm/dl, bilirubin mg/dl, alkaline phosphatase in U/L, cholesterol in mg/dl, aspartate aminotransferase mg/dl, the prothrombin time in seconds, the platelet count per ml , the presence of vascular spiders, presence of hepatomegaly, presence ascites and presence of edema.

[1] https://www.ncbi.nlm.nih.gov/pubmed/2737595
[2] https://www.ncbi.nlm.nih.gov/pubmed/15495127

The additional 112 patients that were eligible for the drug trials but did not enter were used for model validation. The necessary biological data was taken from patient histories. At the conclusion of the study 36 patients had died, 6 had undergone transplantation, 64 were still alive, and 6 were lost to followup to quickly to provide any information.

There were some patients that had missing data values. The Researchers of "Prognosis in Primary Biliary Cirrhosis: Model for Decision Making" used mean and modal imputations to estimate the missing data points. For the purposes of this study, patients with missing values were removed from the model creation process. This resulted in data set with 276 patients, as compared to the study authors, who had the full 312 patient set.

1.3 Methodology

The researchers of the original study used Cox Proportional Hazard Regression for the creation of their survival models. This method would allow a patient to be assigned a risk score of the form:

$$R := \beta_1 Z_1 + \beta_2 Z_2 + ... + \beta_k Z_k$$

Where Z_i is the value of the ith prognostic covariate measured in the study and the β_i is the value of the ith regression coefficient calculated from the Cox proportional hazard regression. In general, when a patients risk score increases their probability of survival goes down; the opposite is true as well. Furthermore, since the risk score is calculated using the Cox proportional hazard assumption, a patient whos risk score differs from another patients by an amount d, then their risk of dying will differ by a multiplicative factor of $\exp(d)$ at any time t. To make survival predictions using the risk score, let $S(t|R)$ be the survival probability that for an individual with risk score R at time t. Their survival probability can be calculated using the formula

$$S(t|R) = S_0(t) \exp(R - R_0)$$

where $S_0(t)$ is the survival curve for a patient with risk score R_0. Ideally, $S_0(t)$ would be the population survival curve for patients with PBC and R_0 would be the average population risk score. The Mayo Clinic study obtained an estimate for $S_0(t)$ from a separate study performed in Europe and took R_0 to be the average risk score calculated from their study using the model they created. In this analysis, $S_0(t)$ will be assumed to be the Kaplan-Meier estimate for the survival curve of the 276 patients with no missing data fields, and R_0 will be the average risk score for these patients calculated using the cox proportional hazard model that will be created in the following analysis.

To select the variables in the models, the original authors used SAS procedure PHGLM. These procedures included both forwards and backwards stepwise variable selections. The authors did not state the specific criterions used for the variables selections, hence, in this study, Backwards Stepwise AIC and Forward Stepwise p-value selection approach will be used.

The Backwards Akaikes Information Criterion approach is a method of variable selection that eliminates a candidate variable and tests if the elimination of this variable lowers the AIC value, which is an estimator of model quality. This process of eliminating variables happens until a model is found where removing any more variables increases the AIC or significantly impacts the model of the fit. However, in the forward p-value approach, variables that have a p-value< 0.05, are added to the model one step at a time. This leads to the creation of model with only variables that are considered to have significant impact on the responding variable. In the following section, the model creation and variable selections done using these two process will be described and compared. After the creation of the two models, a diagnostics test will be done using the Cox-Snell residuals to check whether the proportionality assumption is valid for the two models. The graph plots the estimated cumulative hazard rate against time. If the model is a valid Cox proportional hazard model, it will follow the 45 degree line.

Based on the significance of the covariates, the AIC values of the model, analysis of the residuals, and the coefficient of determination, this study will determine which of the two models is better and compare it to the one created in the original study. Finally, a cross validation using the 106 patients that were not included in the original study will be done to validate the chosen model and confirm that no overfitting or underfitting has taken place. To cross validate the model, a survival curve will be plotted from the dataset and the a prediction curve will be made from the model using the dataset. Then a log rank test will be done to test the following hypothesis:

$$H_0 : S_{Cross-val-dataset}(t) = S_{Predicted}(t) \qquad vs. \qquad H_A : S_{Cross-val-dataset}(t) \neq S_{Predicted}(t)$$

If the p-value is less than 0.05, H_0 will be rejected. In the upcoming section, the results of the methods described in this section will be discussed.

2 Analysis

2.1 Models Constructed

During our variable selection process some common transformations of variables were included in the set of variables that could be chosen. These transformations included the logarithmic transformation.

The model produced by backward elimination based upon the AIC had five risk factors significant at the 95% confidence level. These factors were: age, log of serum bilirubin measured in mg/dl, log of prothrombin time in seconds, log of albumin in gm/dl, and a factor indicating the presence of edema. The summary of the backwards AIC model[3] is displayed on the table below (Figure 1):

```
Call:
coxph(formula = Surv(workdata$time, nstatus) ~ age + edema +
    log(bili) + log(albumin) + log(protime), data = workdata)

  n= 276, number of events= 111

                  coef exp(coef)  se(coef)      z Pr(>|z|)
age           0.032845  1.033390  0.009367  3.506 0.000454 ***
edema         0.807980  2.243372  0.329410  2.453 0.014175 *
log(bili)     0.872992  2.394064  0.105802  8.251  < 2e-16 ***
log(albumin) -2.776956  0.062228  0.789111 -3.519 0.000433 ***
log(protime)  2.598266 13.440418  1.106657  2.348 0.018882 *
---
```

Figure 1: Summary of Backwards AIC Model

The model produced by forward p-value selection had 5 risk factors significant at the 95% confidence level. These factors were: age, aspartate aminotrans-ferase in mg/dl, log of serum bilirubin measured in mg/dl, log of albumin in gm/dl, and a factor indicating the presence of edema. The forwards p-value model[4] can be summarized by the table below (Figure 2):

A Cox-Snell residual plot was created to test the validity of the Cox proportional hazard assumption for each model. The residual plots[5] are as follows (Figure 3):
It is evident from the plots above that the residuals in both cases pass through the origin and follow the 45-degree line closely. There is slightly more variation at the end, which is to be expected due to the reduced sample size of survivors leading to increasing variance over time. Thus, it can be concluded that both the models are valid Cox Proportional Hazard models.

[3] Appendix A.2
[4] Appendix A.3
[5] Appendix B.1

```
Call:
coxph(formula = Surv(workdata$time, nstatus) ~ age + factor(edema) +
    log(bili) + log(albumin) + ast, data = workdata)

  n= 276, number of events= 111

                      coef exp(coef)  se(coef)      z Pr(>|z|)
age                0.039224  1.040003  0.009712  4.039 5.37e-05 ***
factor(edema)0.5   0.282930  1.327013  0.289844  0.976 0.328992
factor(edema)1     1.182112  3.261255  0.333340  3.546 0.000391 ***
log(bili)          0.865328  2.375786  0.117946  7.337 2.19e-13 ***
log(albumin)      -2.697890  0.067347  0.807394 -3.341 0.000833 ***
ast                0.001394  1.001395  0.001691  0.824 0.409764
---
```

Figure 2: Summary of Forwards P-value Model

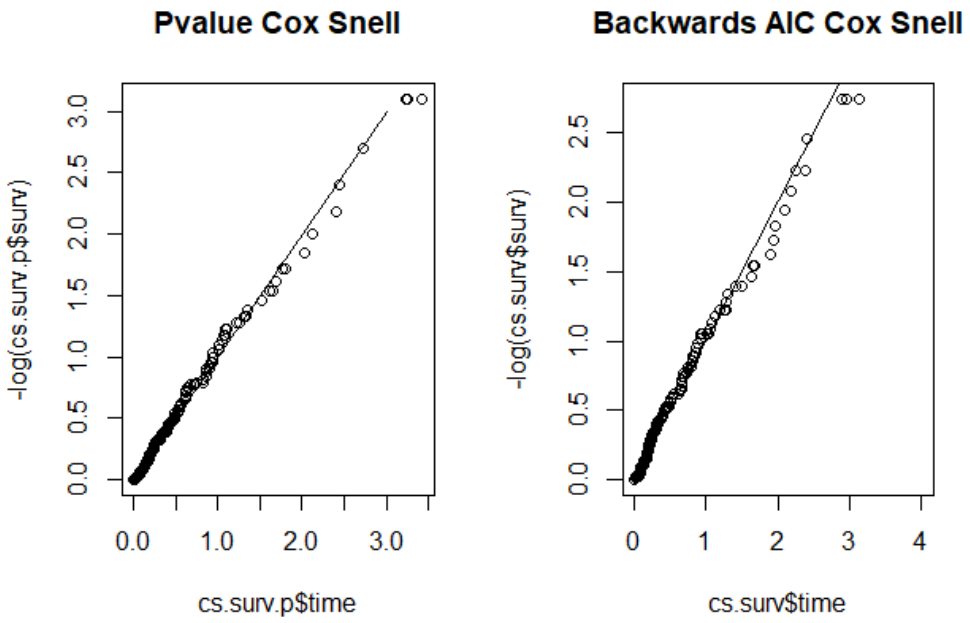

Figure 3: Cox-Snell Residuals

Following the determination of validity, it was decided that the model with the higher number of significant factors would be chosen as the final model. Thus, the p-values of the factors in the backwards AIC model and forwards p-value model were both examined. It was discovered that aspartate aminotransferase in the p-value model had a p-value of ≈ 0.5, while all the factors in the AIC model had a p-value ≤ 0.05. Thus, the AIC model was determined to be superior.

2.2 Cross-Validation of Model

A cross validation of the model was done using data on the 106 patients that were not part of the initial drug trial, to ensure that no overfitting or underfitting took place. By inspection of the plot below[6], it is evident that the prediction model closely follows the new observed data (Figure 4):

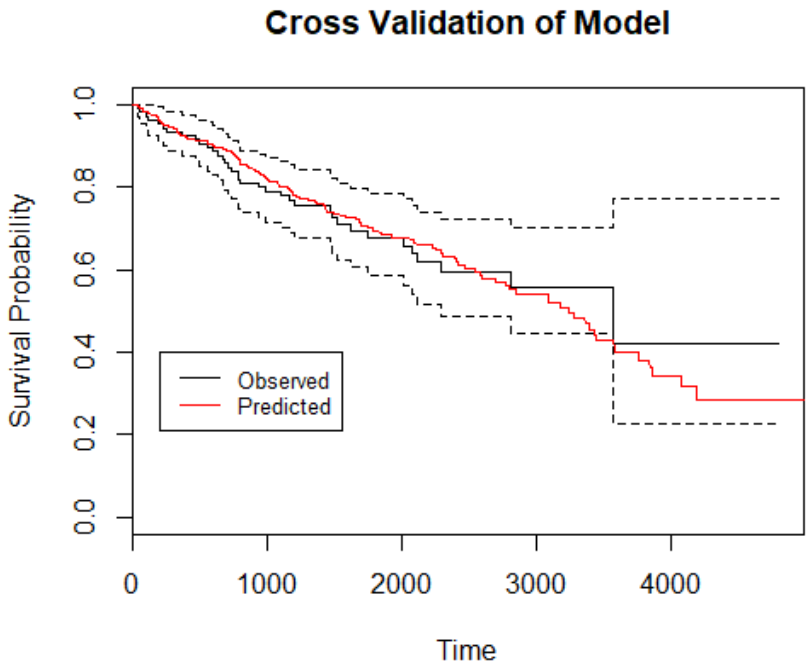

Figure 4: Visual Cross-Validation Check

A one sample log-rank test was performed to compare the predicted survival curve to the Kaplan-Meier estimate for the survival curve for the test set. The p-value for this test was 0.90 so it was concluded that the null hypothesis could not be rejected and it can be concluded that there is no statistically significant difference between the predicted survival curve and the actual survival curve. Thus it would appear that the model created in this analysis can accurately predict survival probabilities for out of sample patients.

3 Conclusion

The survival probability of patient having PBC can be predicted using the backwards AIC model above. This can be done since a risk factor, R, can be determined. It is made possible by using Cox Proportional

[6]Appendix B.3

```
Call:
coxph(formula = Surv(workdata$time, nstatus) ~ age + edema +
    log(bili) + log(albumin) + log(protime), data = workdata)

  n= 276, number of events= 111

                  coef exp(coef)  se(coef)      z Pr(>|z|)
age           0.032845  1.033390  0.009367  3.506 0.000454 ***
edema         0.807980  2.243372  0.329410  2.453 0.014175 *
log(bili)     0.872992  2.394064  0.105802  8.251  < 2e-16 ***
log(albumin) -2.776956  0.062228  0.789111 -3.519 0.000433 ***
log(protime)  2.598266 13.440418  1.106657  2.348 0.018882 *
---
```

Figure 5: Summary of Backwards AIC Model

Hazard model function in R given age, edema status, bilirubin, albumin and prothrombin time. R is formally defined as

$$R = \beta_0 + \beta_1\, Age + \beta_2\, Edema + \beta_3\, \log\left(Bilirubin\right) + \beta_0 \log\left(Albumin\right) + \beta_i \log\left(Prothrombin\right),$$

and,

$$S_T(t|x) = S_0(t)\exp\left(R - R_0\right)$$

where $R_0 = \beta_0$ assuming all the other factors are constant.

The first part of the model is the initial survival of a patient calculated from the initial stages, just at the point of entry into the study; R is the risk factor of the individual determined by using a Cox Proportional Hazard model. With the coefficients (β_i) expressed above, if the data about the covariates is available for a patient, their chances of survival can be determined. Patients with lower survival probabilities should be prioritized for liver transplants. Survival curves can be best used to determining a patient's chance of survival.

The prognosis of PBC is made easier by this model. The most important advantage of this model is that it does not require neither liver biopsy nor other complicated procedures. Even so, if more data is collected and more researches are done, the model could possibly be refined and improved to better predict the waiting time of patients with PBC until transplant or death.