

# Survival Analysis on Cirrhosis Clinical Trial Data

Group 1

## Contents

<b>1</b>	<b>Objective</b>	<b>3</b>
<b>2</b>	<b>Background</b>	<b>3</b>
<b>3</b>	<b>Dataset Overview</b>	<b>3</b>
<b>4</b>	<b>Methods</b>	<b>3</b>
4.1	Data Cleaning . . . . .	3
4.2	Exploratory Data Analysis (EDA) . . . . .	4
4.2.1	Distributions of Key Variables by Survival Status . . . . .	4
4.2.2	Continuous Baseline Variables . . . . .	4
4.2.3	Categorical Baseline Variables . . . . .	4
4.3	Kaplan-Meier Survival Analysis and Log-rank Tests . . . . .	5
4.4	Cox Proportional Hazards Model . . . . .	5
4.5	Stepwise and Lasso Cox Proportional Hazard Regression Model . . . . .	5
<b>5</b>	<b>Results</b>	<b>6</b>
5.1	Descriptive and Exploratory Analysis . . . . .	6
5.1.1	Distributions of Key Variables by Survival Status . . . . .	6
5.1.2	Continuous Baseline Variables . . . . .	6
5.1.3	Categorical Baseline Variables . . . . .	6
5.2	Kaplan-Meier Survival Curves . . . . .	6
5.3	Log-rank Tests . . . . .	7
5.4	Cox Proportional Hazards Model . . . . .	7
5.5	Stepwise and Lasso Cox Proportional Hazard Regression Model . . . . .	9
5.6	Nomogram . . . . .	10
<b>6</b>	<b>Discussion</b>	<b>10</b>
6.1	Descriptive and Exploratory Analysis . . . . .	10
6.2	Kaplan-Meier Summary and Discussion . . . . .	11
6.3	Cox Proportional Hazards Model . . . . .	11
6.4	Stepwise and Lasso Cox Proportional Hazard Regression Model . . . . .	11
6.5	Nomogram . . . . .	12
<b>7</b>	<b>Conclusion</b>	<b>12</b>
<b>8</b>	<b>Reference</b>	<b>13</b>
<b>9</b>	<b>Team members and Roles</b>	<b>13</b>



# 1 Objective

The main goal of this project is to determine whether D-penicillamine treatment improves survival compared with placebo in patients with primary biliary cirrhosis (PBC). The secondary goal is to identify and quantify the extent to which baseline clinical and biochemical factors influence the risk of death. Together, these goals provide a clear picture of how effective the treatment is and of the factors that impact survival in PBC.

## 2 Background

Liver cirrhosis is a chronic condition marked by progressive liver damage, leading to complications such as ascites, hepatic encephalopathy, and eventually death. Understanding factors that influence survival can help guide treatment strategies and improve patient outcomes. In this project, we apply survival analysis techniques to a clinical trial dataset of patients with cirrhosis to identify key predictors of survival time and assess the effectiveness of treatment.

## 3 Dataset Overview

The dataset comes from a publicly available clinical trial on cirrhosis patients, originally hosted on Kaggle and was collected from a liver cirrhosis clinical trial dataset referred to Mayo Clinic during a ten-year interval. It contains **418 observations** with demographic, clinical, and biochemical variables. Each row represents a patient, and the primary outcome is survival status:

- **Status “D”**: Death (event occurred)
- **Status “C”**: Censored (patient still alive at last follow-up)
- **Status “CL”**: Censored but received a liver transplant

Key variables include age (in days), treatment group (**drug**), albumin, bilirubin, prothrombin time, ascites, hepatomegaly, edema, and other lab values such as cholesterol, copper, and SGOT.

## 4 Methods

### 4.1 Data Cleaning

To ensure analysis validity, we applied several preprocessing steps:

1. **Filtered to randomized trial participants**: patients with missing **drug** assignment were excluded.
2. **Created event indicator**: **status** values were mapped to a binary variable, where “D” = 1 and “C”/“CL” = 0.
3. **Converted units**: Age was transformed from days to years for easier interpretation.
4. **Categorical variables formatted**: Factors like **sex**, **ascites**, **hepatomegaly**, **edema** were properly converted into R factors, with **edema** treated as an ordered factor.
5. **Constructed complete-case dataset**: We kept only observations with complete data on key baseline variables for consistent analysis.

## 4.2 Exploratory Data Analysis (EDA)

We conducted exploratory analysis to better understand survival status distribution, how key variables relate to patient outcomes, and assess baseline balance between the Placebo and D-penicillamine groups using summaries of key demographic and clinical variables

### 4.2.1 Distributions of Key Variables by Survival Status

Age shows an approximately bell-shaped distribution centered in the mid-50s, with most patients between 35 and 70 years old, as shown in Figure 1 and Figure 2. This suggests that the trial population consists largely of middle-aged to older adults, with relatively few very young or very old participants.

Albumin, a marker of liver synthetic function, is also roughly symmetric, with values concentrated between 3 and 4 g/dL. In contrast, bilirubin exhibits a highly right-skewed distribution: the majority of patients have low to moderate levels, but a non-trivial subset has markedly elevated bilirubin, reflecting substantial variability in disease severity. (see Figure 3) Prothrombin time shows a unimodal distribution with a slight right tail, indicating that most patients have only mild prolongation, whereas a smaller number exhibit more serious coagulation abnormalities.

Together, these histograms highlight substantial heterogeneity in liver function and disease severity across the cohort. In particular, the strong skewness in bilirubin and the variability in prothrombin suggest that transformations or robust methods may be needed in downstream models, and that these biomarkers are plausible predictors of survival outcomes to be investigated in subsequent analyses.

### 4.2.2 Continuous Baseline Variables

For continuous baseline variables, we compared the two treatment groups using both: Two-sample t-tests (for differences in mean values), and Wilcoxon rank-sum tests (nonparametric robustness checks). The corresponding hypotheses were:

$$H_0 : \mu_{\text{Placebo}} = \mu_{\text{D-penicillamine}} \quad H_1 : \mu_{\text{Placebo}} \neq \mu_{\text{D-penicillamine}}$$

Only clinically relevant continuous baseline variables with minimal missingness were chosen ("age", "bilirubin", "albumin", "prothrombin"). Inclusion of variables such as copper or cholesterol would greatly reduce the complete-case sample size. So, following the Mayo PBC literature by Dickson, only key prognostic markers are summarized.

### 4.2.3 Categorical Baseline Variables

For categorical baseline variables, Group differences are evaluated using:

- Pearson's chi-squared test, when all expected cell counts were adequate
- Fisher's exact test, when sparse cells were present.

the corresponding hypotheses were:

$$H_0 : P(X = k \mid \text{Placebo}) = P(X = k \mid \text{D-penicillamine}) \text{ for all } k$$

$$H_1 : \exists k \text{ such that } P(X = k \mid \text{Placebo}) \neq P(X = k \mid \text{D-penicillamine})$$

### 4.3 Kaplan-Meier Survival Analysis and Log-rank Tests

Kaplan–Meier (KM) methods were used to estimate the survival function for the two treatment groups (D-penicillamine vs. placebo). Survival curves were constructed for the overall cohort as well as for clinically relevant subgroups, including histologic stage (1–4), age tertiles, bilirubin quartiles, and albumin tertiles. For each analysis, the time-to-event outcome was defined as days from study entry until death, with censoring applied for patients who were still alive at the end of follow-up.

Confidence bands were generated using the Greenwood estimator for variance. Visual comparison of survival curves was supplemented by formal hypothesis testing using the log-rank test. The overall log-rank test was used to compare survival distributions between treatment groups in the unstratified setting. To assess whether the treatment effect varied across baseline characteristics, stratified log-rank tests were performed using histologic stage, age tertile, bilirubin quartile, and albumin tertile as stratification factors.

All KM estimates and log-rank tests were computed using the `survival` package in R. Results of the stratified and unstratified log-rank tests are reported in Table 1, and corresponding survival curves are displayed in Figures 7–11.

### 4.4 Cox Proportional Hazards Model

We fitted a Cox proportional hazards regression model to evaluate the treatment effect of D-penicillamine versus placebo on survival, adjusting for baseline clinical and laboratory covariates. Before model fitting, we created clinically meaningful transformed variables: log-bilirubin was used to reduce skewness and stabilize hazard ratio estimates, and age was converted from days to years for easier interpretation. Edema was collapsed into a binary indicator (0 = no edema, 1 = edema present with or without diuretics). Due to large loss in triglycerides (9.6%) and cholesterol (9.0%), these variables were excluded from the analysis. The remaining covariates had minimal missingness (platelets: 1.3%, copper: 0.6%), and a complete-case approach was employed for model fitting.

The initial main Cox proportional hazards regression model included treatment (drug), age, log-bilirubin, albumin, prothrombin time, sex, ascites, edema, platelets, and copper, and histologic stage is treated as a stratification factor. The proportional hazards assumption was evaluated using scaled Schoenfeld residuals and the global test. After fitting the main model, we seek to simplify the model using variable selection, and we used purposeful backward elimination strategy: at each step, we removed one candidate variable and assessed whether the likelihood ratio test indicated significantly worse fit ( $p < 0.05$ ) and whether the treatment hazard ratio or key prognostic hazard ratios changed substantially ( $>10\%$ ). Variables that met neither criterion were dropped for parsimony. Also, the candidate variable is selected from the variable that is not strong prognostic factors.

### 4.5 Stepwise and Lasso Cox Proportional Hazard Regression Model

Two model choosing methods are used to confirm our variable selections. A Stepwise Cox Proportional Hazard regression was performed to identify the most parsimonious model that estimates subjects' hazard of Cirrhosis. The model selection was based on Akaike's Information Criterion (AIC). Another LASSO Cox Proportional Hazard regression was performed to identify the most parsimonious model that estimates subjects' hazard of Cirrhosis. The model selection was based on the penalty parameter  $\lambda$ . Cross-validation further selects the  $\lambda$  that generalizes the best and determines variables to develop the final model.

## 5 Results

### 5.1 Descriptive and Exploratory Analysis

#### 5.1.1 Distributions of Key Variables by Survival Status

- **Age Distribution Across Survival Groups:** Figure 1 demonstrates that individuals who experienced the event (status = D) tend to be older than those in censored groups.
- **Distribution of Key Numeric Variables:** Histograms in Figure 2 illustrate the substantial variability present in biochemical and demographic measures.
- **Biomarker Differences by Survival Status:** Boxplots in Figure 3 show clear separation in clinical biomarkers:
  - Patients who died generally have *higher bilirubin levels*.
  - The same group shows *lower albumin levels*.
- **Correlation Structure Among Continuous Variables:** The correlation heatmap (Figure 4) reveals moderate associations among laboratory variables. In particular, albumin and bilirubin demonstrate a notable negative correlation.

#### 5.1.2 Continuous Baseline Variables

Figure 5 shows that, for these 4 key continuous baseline variables, only age has a statistically significant difference between treatment groups (t-test  $p \approx 0.018$ ; Wilcoxon  $p \approx 0.020$ ). No statistically significant differences were observed for bilirubin, albumin, or prothrombin time. In general, continuous variables are well balanced between treatment groups, except for age.

#### 5.1.3 Categorical Baseline Variables

Figure 6 shows that all categorical variables (sex, ascites, spider angiomas, edema, and histologic stage) are well balanced between treatment groups (all  $p > 0.20$ ). Therefore, we fail to reject the null hypothesis, indicating that there is no evidence of imbalance in the distribution of any categorical baseline variables.

### 5.2 Kaplan-Meier Survival Curves

We first examined overall survival patterns using Kaplan–Meier (KM) estimates. The unstratified KM curves for D-penicillamine and placebo overlapped almost completely across the 4,000-day follow-up period (Figure 7), indicating no apparent survival benefit associated with treatment. To further explore potential heterogeneity of survival patterns, KM curves were stratified across several clinically important baseline factors:

- **Histologic stage:** Survival decreased markedly with increasing disease severity, yet within each stage, treatment curves remained nearly identical (Figure 8).
- **Age group:** Older patients experienced worse survival overall, but the D-penicillamine and placebo curves within each age tertile showed minimal separation (Figure 9).
- **Bilirubin quartiles:** Patients in higher bilirubin quartiles had poorer outcomes; however, treatment curves again overlapped within each subgroup (Figure 10).

- **Albumin tertiles:** Higher albumin levels were associated with improved survival, but treatment effects remained similar across all albumin subgroups (Figure 11).

Across all stratified KM analyses, no visual evidence suggested any differential treatment effect.

### 5.3 Log-rank Tests

To formally evaluate whether D-penicillamine improved survival, we conducted both overall and stratified log-rank tests. The unstratified log-rank test produced a p-value of **0.70**, indicating no significant difference in survival between the two treatment arms.

Stratified log-rank tests adjusting for stage, age tertile, bilirubin quartile, and albumin tertile similarly produced non-significant results. In every subgroup, p-values were far above the 0.05 significance threshold, suggesting that the treatment effect does not vary across baseline characteristics.

Table 1 summarizes all log-rank test results.

Test	Group	N	Observed	Expected	$(O - E)^2/E$	$(O - E)^2/V$	p-value
Unstratified	D-penicillamine	158	65	63.2	0.050	0.102	0.70
	Placebo	154	60	61.8	0.051	0.102	
Stratified by Stage	D-penicillamine	158	65	61.8	0.169	0.343	0.60
	Placebo	154	60	63.2	0.165	0.343	
Stratified by Age Group	D-penicillamine	158	65	65.2	0.0005	0.0011	1.00
	Placebo	154	60	59.8	0.0006	0.0011	
Stratified by Bilirubin Quartiles	D-penicillamine	158	65	64.4	0.0063	0.0137	0.90
	Placebo	154	60	60.6	0.0067	0.0137	
Stratified by Albumin Tertiles	D-penicillamine	158	65	65.1	0.0026	0.0006	1.00
	Placebo	154	60	59.9	0.0028	0.0006	

Table 1: Log-rank and stratified log-rank tests comparing treatment groups

### 5.4 Cox Proportional Hazards Model

The main Cox model was fitted on 306 patients with 123 observed deaths. In this fully adjusted model, D-penicillamine showed no significant survival benefit compared with placebo (HR = 1.19, 95% CI: 0.81–1.73, p = 0.375). Several baseline factors emerged as strong prognostic predictors: older age (HR = 1.03 per year, p = 0.009), higher log-bilirubin (HR = 2.25, p < 0.001), lower albumin (HR = 0.46, p = 0.003), and longer prothrombin time (HR = 1.31, p = 0.009). In contrast, sex, ascites, platelets, and copper showed weak or null effects.

The proportional hazards assumption was evaluated using the `cox.zph` test.

The global test (p = 0.123) did not indicate strong overall violation of the PH assumption. Most individual covariates satisfied the assumption well (p > 0.10), though prothrombin (p = 0.043) and edema (p = 0.010) showed borderline departures. Schoenfeld residual plots for these two variables revealed mild curvature over time (Figures 12 and 13), it shows their effects may vary slightly during follow-up, but the deviations is slightly.

For purposeful backward elimination, we removed copper (LRT p = 0.18), platelets (LRT p = 0.59), sex (LRT p = 0.12), and ascites (LRT p = 0.59), no of them improved model fit or altered the treatment effect. The final reduced model retained drug, age, log-bilirubin, albumin, prothrombin, and edema, and stage is a stratification factor.

Table 2: Cox proportional hazards model for the main model with all covariates (n = 306, events = 123).

Variable	Coef	HR	95% CI	z	p-value
Drug (Placebo vs D-pen)	0.171	1.19	0.81 – 1.73	0.888	0.375
Age (per year)	0.025	1.03	1.01 – 1.05	2.615	0.009 **
log(Bilirubin)	0.810	2.25	1.78 – 2.83	6.837	<0.001 ***
Albumin	−0.776	0.46	0.27 – 0.77	−2.922	0.003 **
Prothrombin	0.273	1.31	1.07 – 1.61	2.632	0.009 **
Sex (Male vs Female)	0.246	1.28	0.71 – 2.30	0.818	0.413
Ascites (Yes vs No)	0.020	1.02	0.55 – 1.89	0.063	0.950
Edema (Yes vs No)	0.455	1.58	0.97 – 2.57	1.825	0.068 .
Platelets (per unit)	−0.001	1.00	1.00 – 1.00	−0.555	0.579
Copper (per unit)	0.002	1.00	1.00 – 1.00	1.370	0.171

*Note:* Concordance = 0.803 (SE = 0.025). Stratified by histologic stage.

Signif. codes: \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ , .  $p < 0.1$

Table 3: Proportional hazards test for the main Cox model.

Variable	Chi-square	df	p-value
Drug	0.759	1	0.384
Age	0.203	1	0.652
log(Bilirubin)	1.820	1	0.177
Albumin	0.007	1	0.935
Prothrombin	4.105	1	0.043
Sex	0.064	1	0.800
Ascites	0.011	1	0.916
Edema	6.623	1	0.010
Platelets	0.255	1	0.614
Copper	0.419	1	0.517
<b>GLOBAL</b>	<b>15.249</b>	<b>10</b>	<b>0.123</b>

In this parsimonious model, all retained prognostic factors remained statistically significant: age (HR = 1.03,  $p < 0.001$ ), log-bilirubin (HR = 2.39,  $p < 0.001$ ), albumin (HR = 0.46,  $p = 0.002$ ), prothrombin (HR = 1.30,  $p = 0.008$ ), and edema (HR = 1.58,  $p = 0.047$ ). The treatment effect remained non-significant (HR = 1.23, 95% CI: 0.85–1.80,  $p = 0.271$ ). The model has a concordance index of 0.793, which shows good discriminative ability.

But from the results of proportional hazards test for this model, the proportional hazards



Table 4: Cox proportional hazards model after variable selection (n = 306, events = 123).

Variable	Coef	HR	95% CI	z	p-value
Drug (Placebo vs D-pen)	0.211	1.23	0.85 – 1.80	1.100	0.271
Age (per year)	0.032	1.03	1.01 – 1.05	3.454	<0.001 ***
log(Bilirubin)	0.873	2.39	1.94 – 2.95	8.214	<0.001 ***
Albumin	−0.780	0.46	0.28 – 0.74	−3.165	0.002 **
Prothrombin	0.263	1.30	1.07 – 1.58	2.634	0.008 **
Edema (Yes vs No)	0.460	1.58	1.01 – 2.50	1.982	0.047 *

*Note:* Concordance = 0.793 (SE = 0.025). Stratified by histologic stage.

Signif. codes: \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$

Table 5: Proportional hazards test for the reduced Cox model.

Variable	Chi-square	df	p-value
Drug	0.603	1	0.437
Age	0.716	1	0.398
log(Bilirubin)	2.080	1	0.149
Albumin	0.000	1	0.994
Prothrombin	3.840	1	0.050
Edema	6.390	1	0.011
<b>GLOBAL</b>	<b>14.10</b>	<b>6</b>	<b>0.029</b>

assumption is violated, non-PH signal mainly comes from prothrombin and edema, dropping the weak variables does not solve this.

## 5.5 Stepwise and Lasso Cox Proportional Hazard Regression Model

As summarized in Table 6, both the stepwise Cox proportional hazards model and the LASSO-penalized Cox model were fitted to identify important predictors of time to death.

The final stepwise Cox model achieved the lowest Akaike Information Criterion ( $AIC = 1068.622$ ), indicating the best overall model fit among the candidate models. This model retained edema presence, serum albumin, urine copper, SGOT, prothrombin time, histologic stage, age, and log-transformed serum bilirubin as covariates. Several of these variables demonstrated statistically significant associations with the hazard of death ( $p < 0.05$ ), while histologic stage indicators and SGOT showed weaker contributions (all  $p$ -values  $> 0.05$ ). The model exhibited strong discriminative ability, with a concordance statistic of 0.858, and global testing confirmed an overall good fit (Wald test,  $p < 2 \times 10^{-16}$ ). Treatment assignment was not selected in the stepwise model.

The LASSO Cox model selected a similar but slightly more parsimonious set of predictors. At the optimal penalty parameter ( $\lambda = 0.071$ ), the model included ascites presence, edema presence, serum albumin, urine copper, prothrombin time, histologic stage 4, age, and log-transformed

serum bilirubin. A more restrictive model based on the one-standard-error rule retained only log-transformed serum bilirubin as a predictor; however, the model minimizing cross-validated mean squared error was preferred due to superior estimation performance. Consistent with both the unpenalized and stepwise Cox models, treatment assignment was not selected by the LASSO procedure.

Table 6: Stepwise Cox Proportional Hazards Regression Model

Term	Estimate	Std. Error	Wald $z$	$p$ -value
edema_bin	0.476	0.231	2.058	0.040*
albumin	-0.763	0.255	-2.992	0.003**
copper	0.002	0.001	1.995	0.046*
sgot	0.003	0.002	1.644	0.100
prothrombin	0.289	0.103	2.806	0.005**
stage 2	1.346	1.061	1.269	0.204
stage 3	1.479	1.034	1.430	0.153
stage 4	1.763	1.033	1.707	0.088
age_years	0.030	0.009	3.231	0.001**
log(bilirubin)	0.715	0.121	5.908	< 0.001***
<b>Model summary</b>				
$n$	306			
Events	123			
Wald statistic	195.21			
AIC	1068.62			
BIC	1096.74			

\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ . Hazard ratios are obtained by exponentiating the coefficient estimates.

## 5.6 Nomogram

Figure 14 presents the nomogram constructed from the final Cox proportional hazards (PH) model to estimate individualized 1-, 3-, and 5-year survival probabilities for patients with cirrhosis. The nomogram translates the fitted Cox regression coefficients into a point-based scoring system: each predictor (drug assignment, age, log-transformed bilirubin, albumin, prothrombin time, edema status, and histologic stage) corresponds to a points scale, and the total points are obtained by summing points across predictors. This total score is then mapped to predicted survival probabilities at 1, 3, and 5 years.

## 6 Discussion

### 6.1 Descriptive and Exploratory Analysis

The observed age differences across survival groups suggest that age may be an important predictor of mortality in cirrhosis patients. The substantial variability in biochemical and demographic measures highlights the heterogeneous nature of the study population and the need for potential transformation or normalization of skewed markers such as bilirubin.

Differences in bilirubin and albumin levels by survival status align with established clinical expectations in cirrhosis prognosis, reflecting impaired hepatic function and more severe cholestasis

among patients with poorer outcomes.

The moderate correlations shown among laboratory variables, especially the negative association between albumin and bilirubin, are consistent with liver disease physiology. No pairs exhibited correlations strong enough to indicate harmful multicollinearity, though relationships will be monitored in subsequent multivariable modeling.

In general, continuous variables are well balanced between treatment groups besides age, with patients in the D-penicillamine group being slightly older at baseline. This suggests that age might need to be adjusted in subsequent analyses. Otherwise, baseline characteristics are comparable between the two randomized groups.

## 6.2 Kaplan-Meier Summary and Discussion

The Kaplan-Meier analyses showed that the survival curves for the D-penicillamine and placebo groups were almost identical. This pattern was consistent across all stratified analyses, including stage, age, bilirubin, and albumin, and none of the corresponding log-rank tests were significant.

These findings suggest that D-penicillamine does not improve survival in this study population. Although baseline factors such as disease stage and liver function markers strongly affected prognosis, they did not modify the treatment effect. The KM results therefore support the overall conclusion that the lack of a treatment benefit is robust and consistent with the Cox regression findings.

## 6.3 Cox Proportional Hazards Model

The Cox regression analysis shows that D-penicillamine does not improve survival in patients with primary biliary cirrhosis. The treatment hazard ratio of approximately 1.2 (95 CI: 0.85–1.80) was consistent across both the full and reduced models, it means that this null finding is robust to different model specifications.

In contrast, several baseline markers of disease severity were strong prognostic factors. For example: elevated bilirubin, lower serum albumin, older age and prolonged prothrombin time.

For the proportional hazards diagnostics, the global test was not significant for the main model ( $p = 0.123$ ), the reduced model showed PH violation ( $p = 0.029$ ), it's because the effect of prothrombin and edema. And the treatment effect showed no time-varying behavior, shows the validity of the null effect estimate. Overall, these findings indicate that D-penicillamine does not improve survival in this population, while baseline liver function markers are the main predictors of prognosis.

## 6.4 Stepwise and Lasso Cox Proportional Hazard Regression Model

Consistent with earlier Cox proportional hazards analyses, treatment assignment (drug) was not retained in both final stepwise and lasso model, suggesting no significant treatment effect on survival. The close agreement between the stepwise and LASSO Cox models reinforces the robustness of the identified risk factors and supports the conclusion that biochemical markers of liver dysfunction and patient age play dominant roles in determining survival, whereas treatment assignment does not appear to significantly affect mortality risk in this cohort. Although ascites presence and urine copper were retained in the LASSO-selected model, they were ultimately excluded from the selected model to improve interpretability while maintaining predictive performance.

## 6.5 Nomogram

Consistent with the multivariable Cox model results, the treatment assignment contributes relatively little to the overall risk score compared with key clinical markers (e.g., bilirubin, albumin, and prothrombin time), indicating that survival prediction is primarily driven by baseline disease severity rather than treatment group in this dataset. Overall, the nomogram provides an interpretable and clinically convenient tool for rapid risk identification and individualized prognosis.

## 7 Conclusion

Looking at both non-parametric tests, such as the Kaplan-Meier (KM) and log-rank tests, and regression methods, including Cox analysis, there is no evidence that D-penicillamine improves survival in this primary biliary cholangitis (PBC) trial group. The results show that baseline disease severity markers are key for predicting outcomes and should remain at the center of risk assessment and clinical monitoring.

When modeling, we log-transformed skewed biomarkers and tried different ways to select variables. These steps led to similar sets of prognostic factors, which supports the reliability of our main clinical findings. Still, model checks showed that some covariates did not fully meet the proportional hazards assumption. This means future studies could use time-varying effects or other modeling methods to improve prediction, instead of relying on a single summary hazard ratio.

In summary, this project demonstrates how survival methods can be employed, in a manner consistent with standard survival analysis workflows, to (i) estimate a randomized treatment effect and, based on baseline clinical data, (ii) build interpretable prognostic tools.

## 8 Reference

- [1] Dickson, E. R., Grambsch, P. M., Fleming, T. R., Fisher, L. D., & Langworthy, A. (1989). Prognosis in primary biliary cirrhosis: Model for decision making. *Hepatology*, 10(1), 1–7. <https://doi.org/10.1002/hep.1840100102>
- [2] Rowe, I. A. (2017). Lessons from epidemiology: The burden of liver disease. *Digestive Diseases*, 35(4), 304–309. <https://doi.org/10.1159/000456580>
- [3] Sharma, B., & John, S. (2025). Hepatic cirrhosis. In *StatPearls*. StatPearls Publishing. <http://www.ncbi.nlm.nih.gov/books/NBK482419/>

## 9 Team members and Roles

Each member owns the methods, results, discussions for their assigned analysis.

- Yixin Zheng yz4993 : Title, Objective, Conclusion, Descriptive and Exploratory Analysis(Baseline Table, T-test/Wilcoxon/X2), final report structure and content wrap up
- Zhaokun Lin zl3544 : Statistical analysis (part), Cox Model methods, HR, CI, and PH assumptions check
- Wenjie Wu ww2744 : Data description, statistical analysis(part), Data Cleaning, part of EDA, github build, final report format and content wrap up
- Puyuan Zhang pz2334 : Timeline, statistical analysis(part), primary analysis (Kaplan-Meier and log-rank analysis)
- Chuyuan Xu cx2347 : Background, Reference, Exploratory multivariable modeling (stepwise Cox/Lasso, nomograms)

## 10 Appendix: Figures and Code Results

- Figure 1 — Age distribution by survival status
- Figure 2 — Distribution of numeric variables
- Figure 3 — Albumin and Bilirubin by survival status
- Figure 4 — Correlation heatmap
- Figure 5 — Baseline continuous variables by treatment group (randomized complete cases)
- Figure 6 — Baseline categorical variables by treatment group (randomized patients)
- Figure 7 – Overall Kaplan–Meier survival curves
- Figure 8 – KM curves stratified by histologic stage
- Figure 9 – KM curves stratified by age groups

- Figure 10 – KM curves stratified by bilirubin quartiles
- Figure 11 – KM curves stratified by albumin tertiles

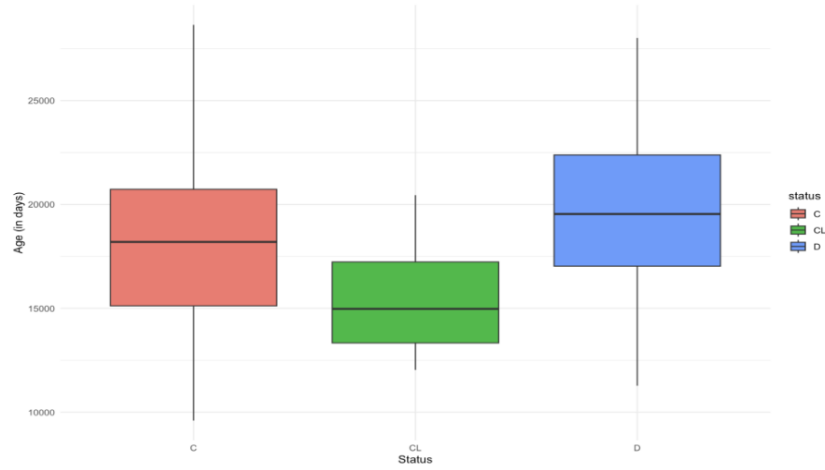


Figure 1: Age distribution by survival status

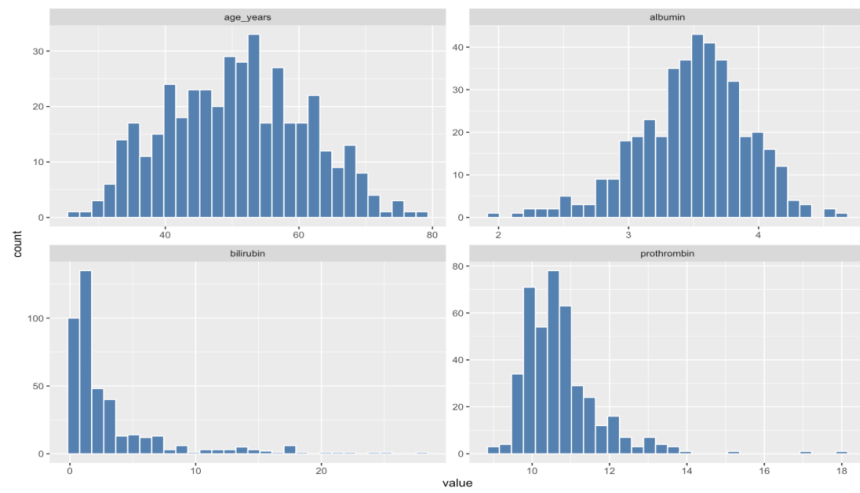


Figure 2: Distribution of numeric variables across patients

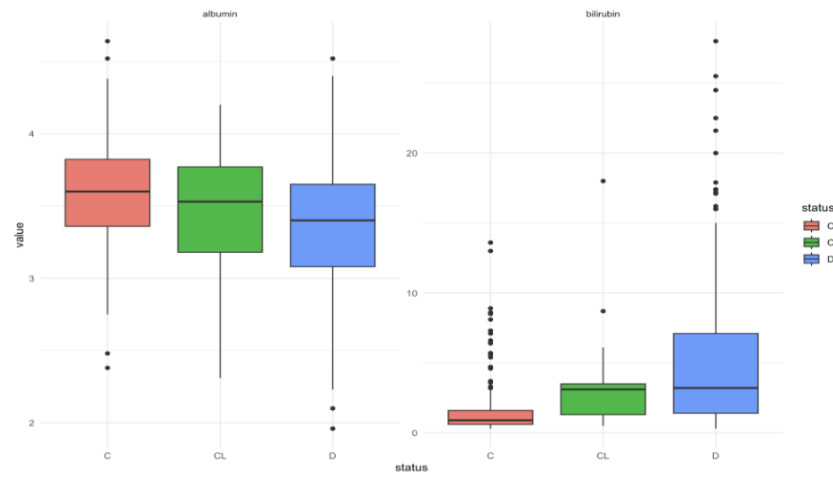


Figure 3: Key lab markers (albumin and bilirubin) by survival status

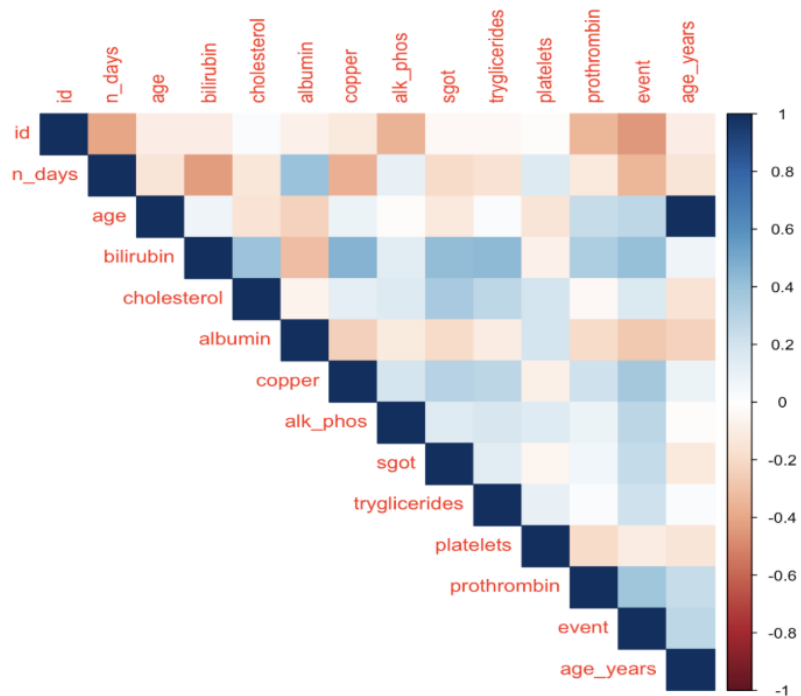


Figure 4: Correlation heatmap among continuous variables

Baseline continuous variables by treatment group (randomized complete cases)

Baseline variable	Placebo mean (SD)	D-penicillamine mean (SD)	Placebo median (IQR)	D-penicillamine median (IQR)	t-test p-value	Wilcoxon p-value
Age (years)	48.58 (9.96)	51.42 (11.01)	48.11 (41.43, 55.80)	51.93 (42.98, 58.90)	0.018	0.020
Serum bilirubin (mg/dL)	3.65 (5.28)	2.87 (3.63)	1.30 (0.72, 3.60)	1.40 (0.80, 3.20)	0.133	0.842
Albumin (g/dL)	3.52 (0.40)	3.52 (0.44)	3.54 (3.34, 3.78)	3.56 (3.21, 3.83)	0.874	0.951
Prothrombin time (seconds)	10.80 (1.14)	10.65 (0.85)	10.60 (10.00, 11.40)	10.60 (10.03, 11.00)	0.199	0.588

Figure 5: Baseline continuous variables by treatment group (randomized complete cases)

Baseline categorical variables by treatment group (randomized patients)

Baseline variable	Level	Placebo n (%)	D-penicillamine n (%)	Test type	p-value
Sex	F	139 (90.3%)	137 (86.7%)	Chi-squared	0.326
Sex	M	15 (9.7%)	21 (13.3%)	Chi-squared	0.326
Ascites	N	144 (93.5%)	144 (91.1%)	Chi-squared	0.433
Ascites	Y	10 (6.5%)	14 (8.9%)	Chi-squared	0.433
Hepatomegaly	N	67 (43.5%)	85 (53.8%)	Chi-squared	0.069
Hepatomegaly	Y	87 (56.5%)	73 (46.2%)	Chi-squared	0.069
Spider angiomas	N	109 (70.8%)	113 (71.5%)	Chi-squared	0.885
Spider angiomas	Y	45 (29.2%)	45 (28.5%)	Chi-squared	0.885
Edema	N	131 (85.1%)	132 (83.5%)	Chi-squared	0.877
Edema	S	13 (8.4%)	16 (10.1%)	Chi-squared	0.877
Edema	Y	10 (6.5%)	10 (6.3%)	Chi-squared	0.877
Histologic stage	1	4 (2.6%)	12 (7.6%)	Fisher's exact	0.205
Histologic stage	2	32 (20.8%)	35 (22.2%)	Fisher's exact	0.205
Histologic stage	3	64 (41.6%)	56 (35.4%)	Fisher's exact	0.205
Histologic stage	4	54 (35.1%)	55 (34.8%)	Fisher's exact	0.205

Figure 6: Baseline categorical variables by treatment group (randomized patients)



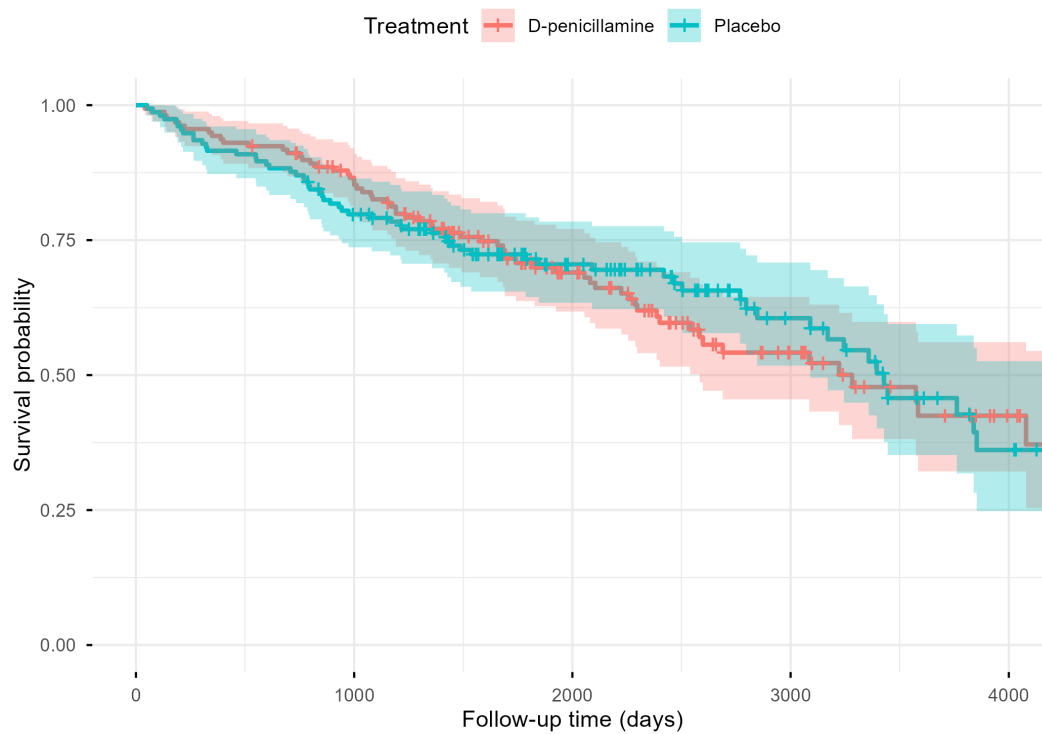


Figure 7: Kaplan–Meier survival curves comparing D-penicillamine and placebo.

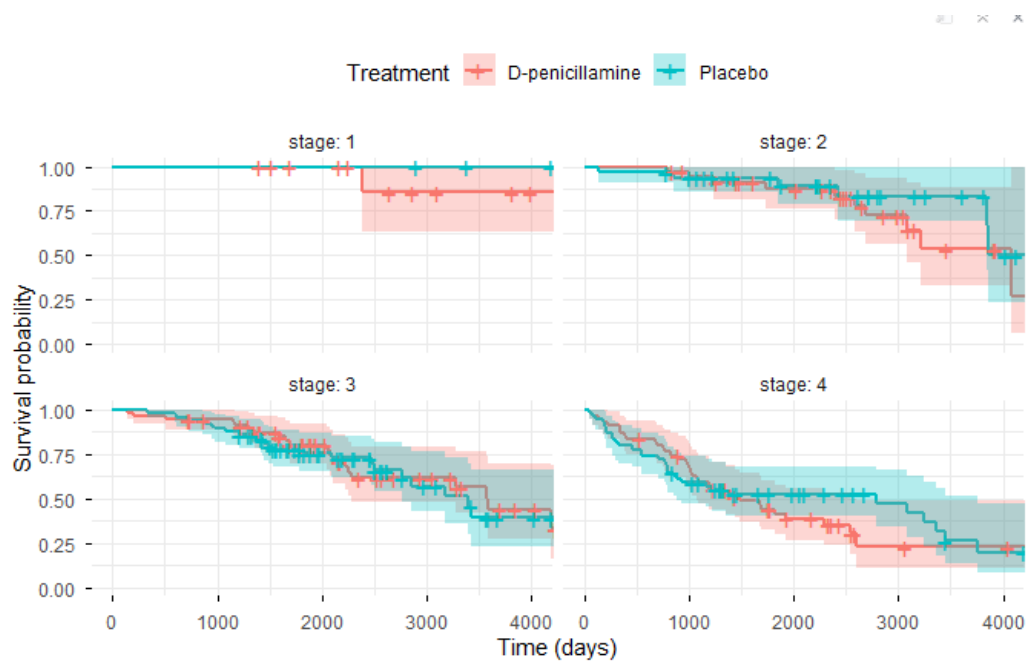


Figure 8: Kaplan–Meier survival curves stratified by histologic stage.

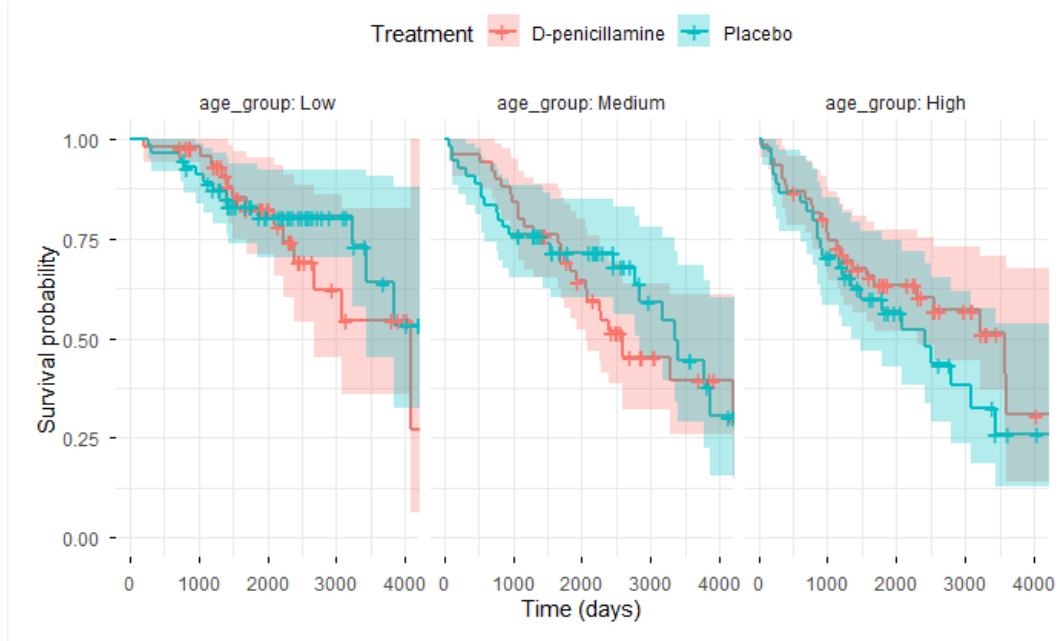


Figure 9: Kaplan–Meier survival curves stratified by age groups.

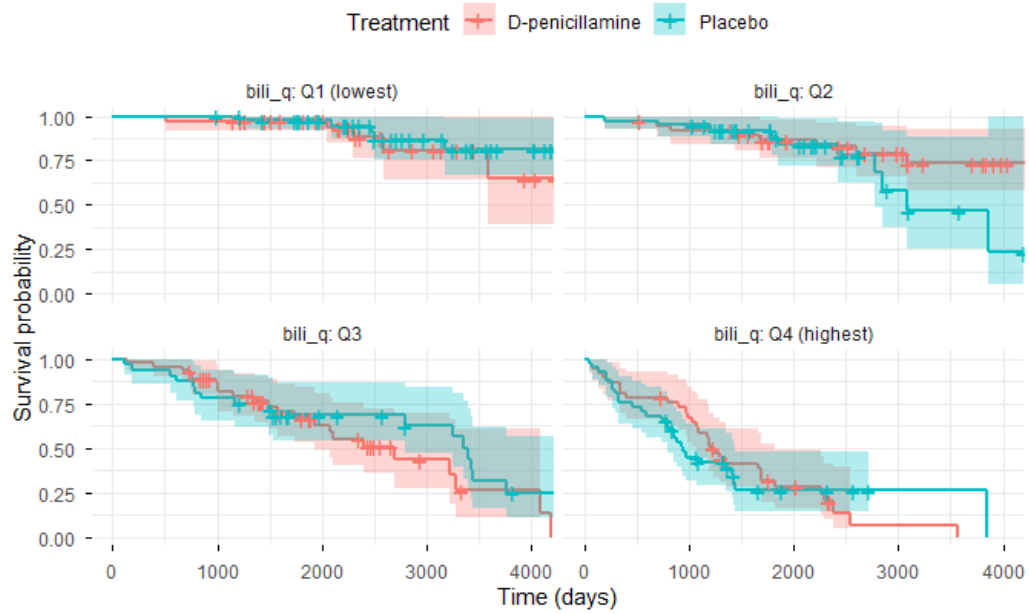


Figure 10: Kaplan–Meier survival curves stratified by bilirubin quartiles.

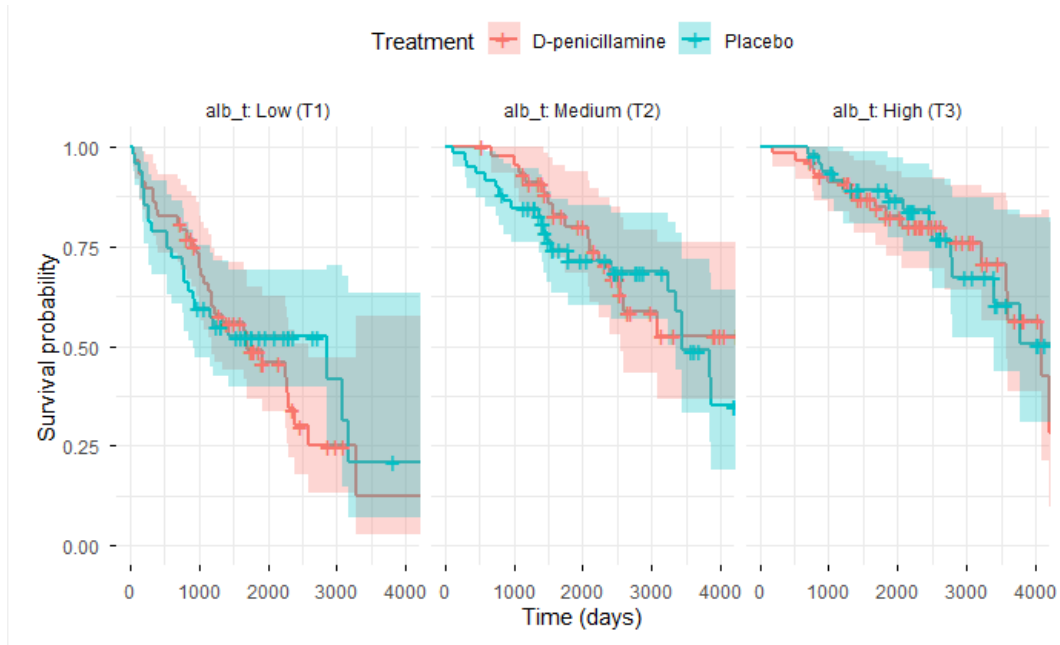


Figure 11: Kaplan–Meier survival curves stratified by albumin tertiles.

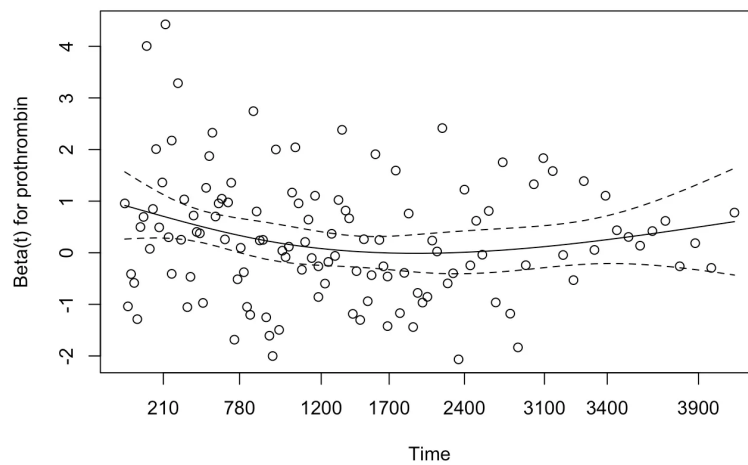


Figure 12: Schoenfeld residuals for prothrombin over time.

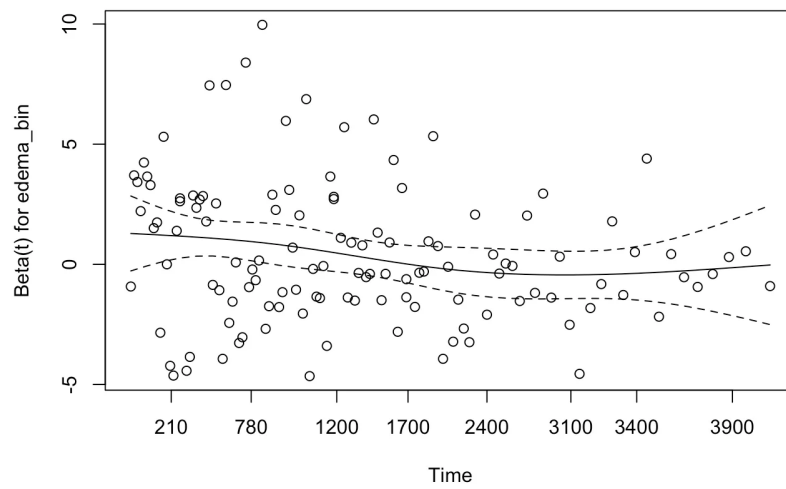
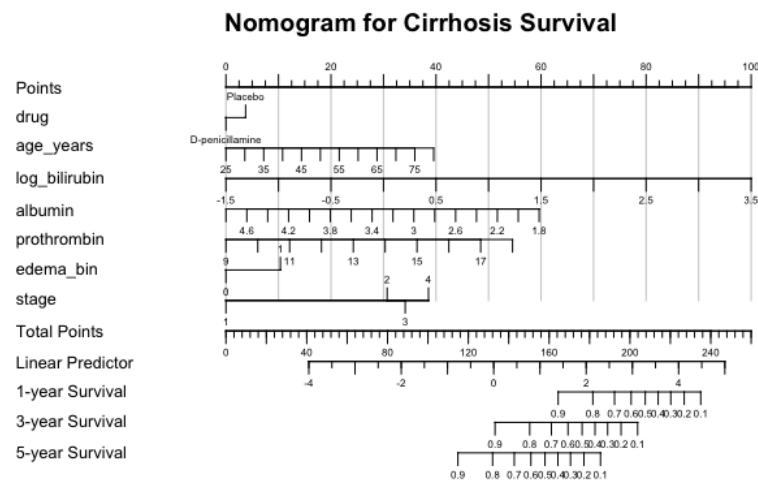


Figure 13: Schoenfeld residuals for edema over time.



1-, 3-, and 5-year survival predictions

Figure 14: Nomogram for Cirrhosis Survival: 1-, 3-, and 5-year survival predictions