Primary Biliary Cirrhosis Clinical Data: Exploratory Data and Survival Analyses

presented by AlNasiri Y. , Burcak Y. , Duzen N. , Ozsut O.

The complete analysis, including the detailed code and results, is available in the corresponding Jupyter Notebook at this link.

workflow

1 Exploratory Data Analysis (EDA)

- Data Preprocessing
- Sample Size Comparisons
- Selection Bias (Trial vs. Non-Trial)

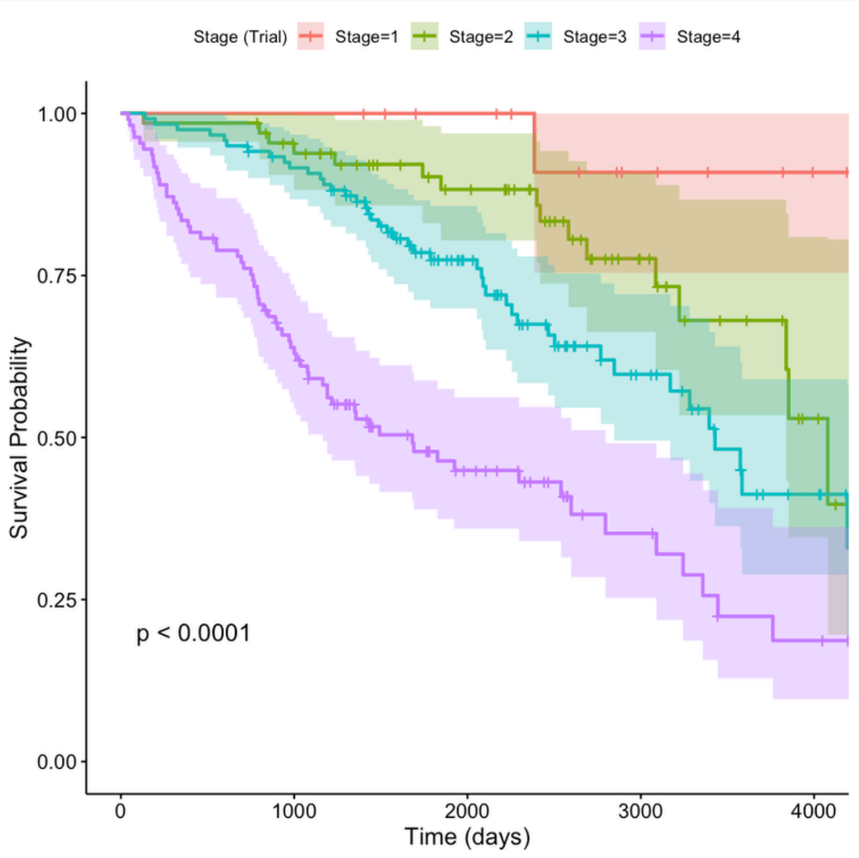
2 Univariate Survival Analysis

- Stratified Kaplan Meier Curves and Log Rank Tests for Categorical Variables
- Cox Regression Model for Continuous Variables

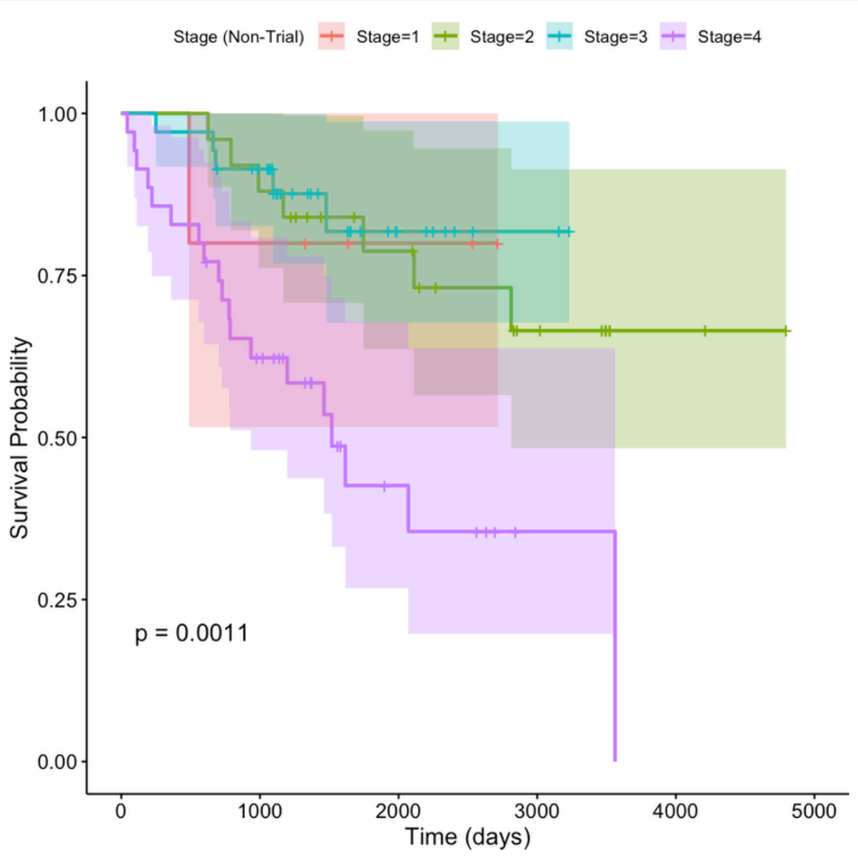
3 Multivariate Survival Analysis

- Multivariable Cox Model
- Forest Plots
- Interaction Analysis (Sex * Platelets)

stage stratified kaplan-meier

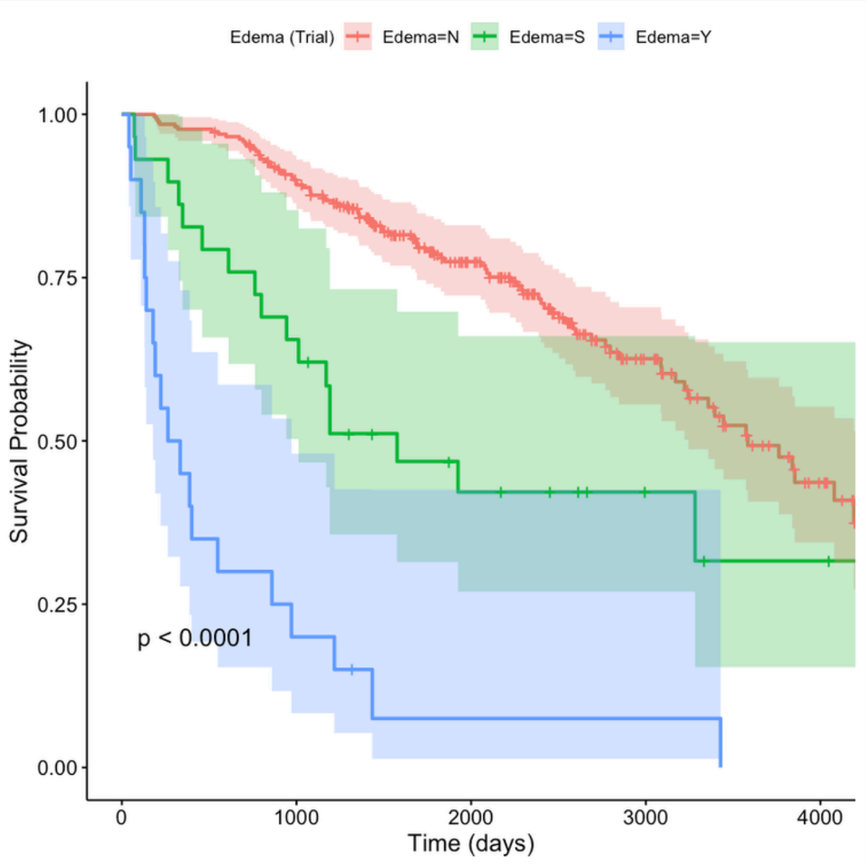


Log rank: Chisq= 53.8 on 3 degrees of freedom, p= 1e-11

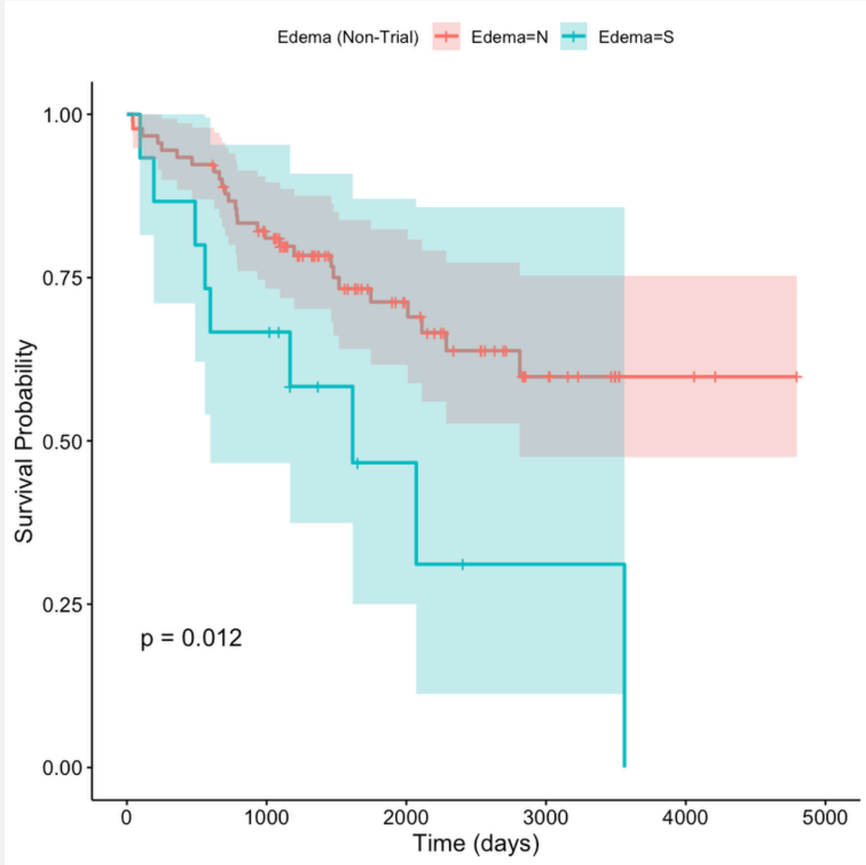


Log rank: Chisq= 16.2 on 3 degrees of freedom, p= 0.001

edema stratified kaplan-meier

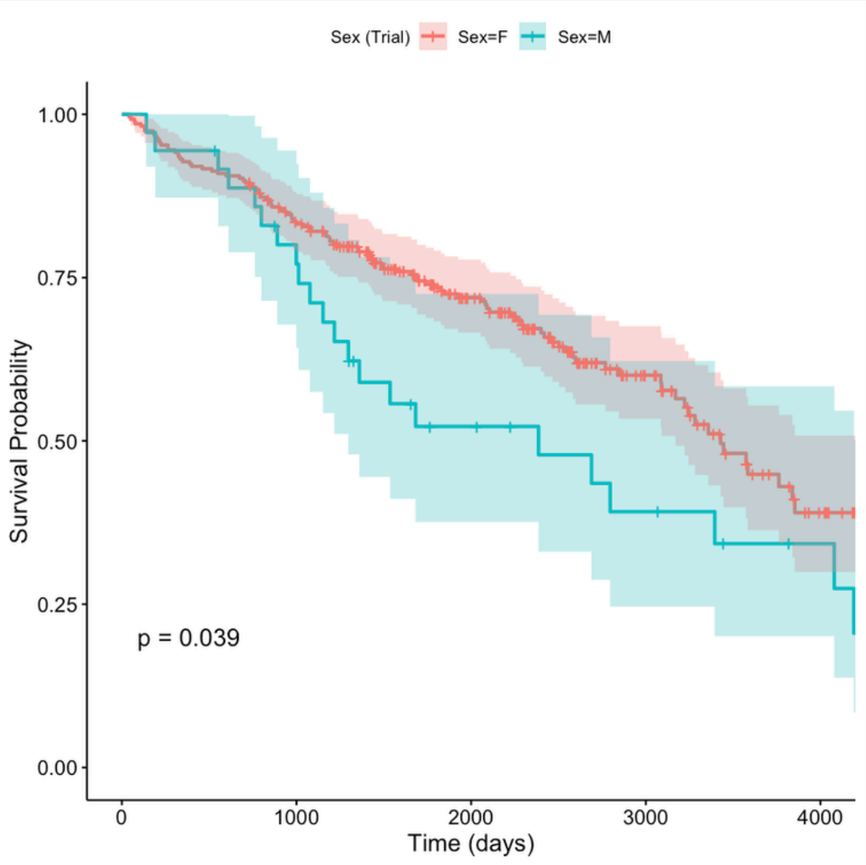


Log rank: Chisq= 127 on 2 degrees of freedom, p= <2e-16

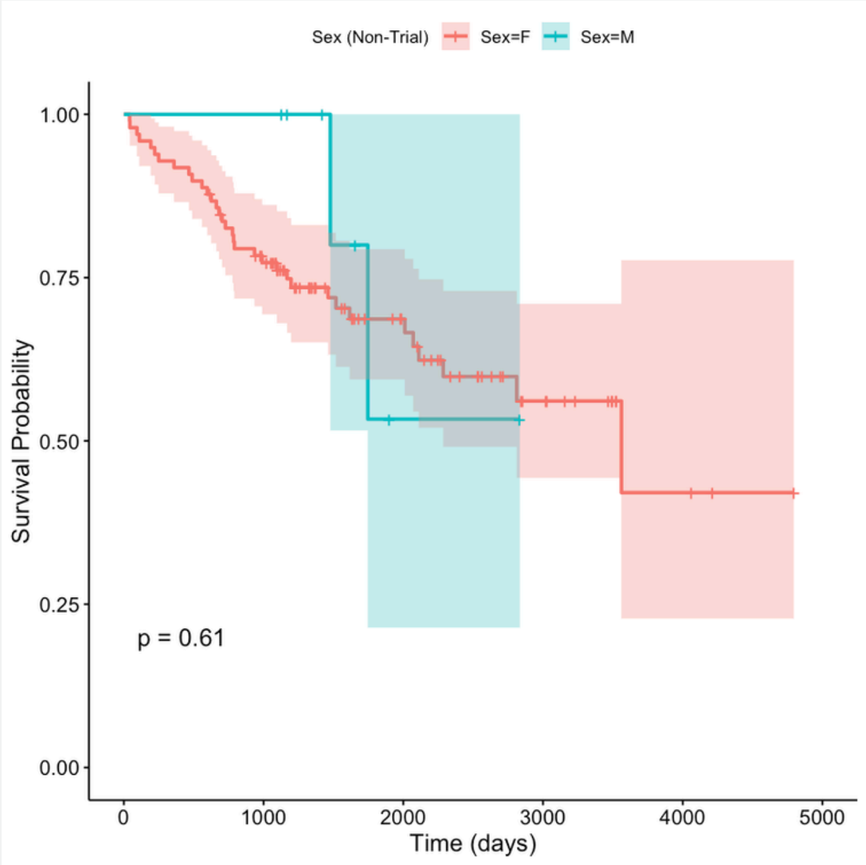


Log rank: Chisq= 6.3 on 1 degrees of freedom, p= 0.01

sex stratified kaplan-meier



Log rank: Chisq= 4.3 on 1 degrees of freedom, p= 0.04



Log rank: Chisq= 0.3 on 1 degrees of freedom, p= 0.6

forest plot for multivariable cox model

Significant Predictors of Survival:

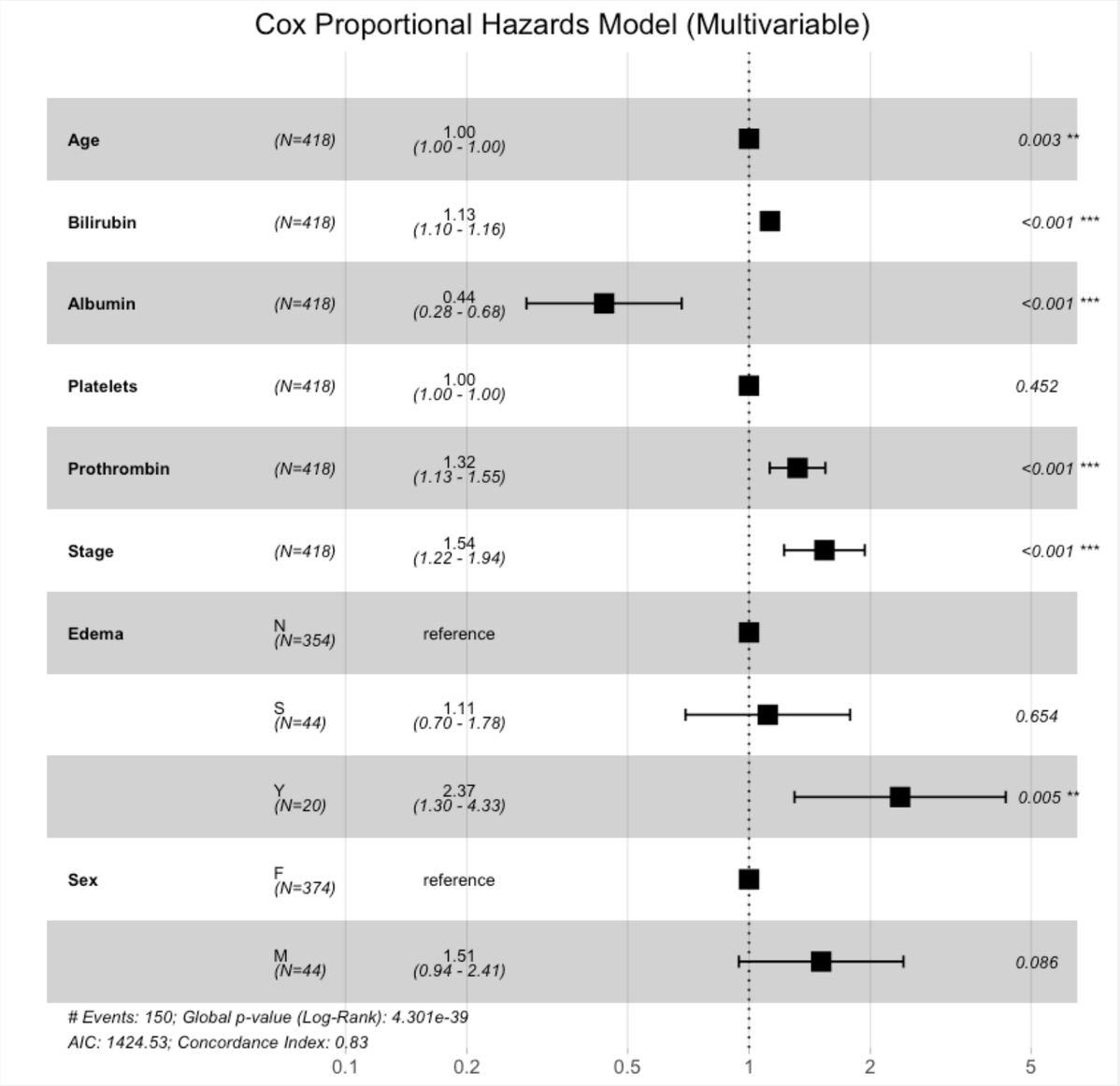
For every one-unit increase in **bilirubin**, the hazard increases by 12.7%.

For every one-unit increase in **albumin**, the hazard decreases by 56.3%.

For every one-unit increase in **prothrombin**, the hazard increases by 31.9%.

For every one-unit increase in **stage**, the hazard increases by 53.8%.

Compared to patients without edema (N), those with **persistent edema (Y)** have a 136.9% increase in hazard.



Concordance Index: 0.827 (high predictive accuracy).

interaction analysis on platelets and sex

	Term	Coefficient	Hazard_Ratio	Lower_95_CI	Upper_95_CI	P_Value
coef	Platelets:SexM	0.005866475	1.005884	1.000717	1.011077	0.02555408

for each unit increase in platelet count, the hazard of the event (death) increases by approximately **0.59% in males relative to females**

A p-value less than 0.05 indicates that the interaction between platelets and sex is statistically significant

- **In females (Sex=F)**, the effect of platelet counts on survival is determined by the main effect of platelets. This effect was not statistically significant in the model
- **In males (Sex=M)**, the effect of platelets on survival is modified by the interaction term. The total effect for males is given by:

Effect in Males = Main Effect of Platelets + Interaction Term (Platelets:SexM)

The interaction term is statistically significant -> platelet counts influence survival in males more compared to females.

conclusions

indicator of survival risk

Kaplan-Meier curves and log-rank tests showed that patients with severe edema (Y: Edema despite diuretics) had the worst survival outcomes.
This remained significant in the multivariable Cox model, meaning that edema is a clinically relevant marker for disease severity and mortality risk.

platelets show a sex-based interaction with survival

While platelets alone were not significantly associated with survival, the interaction term showed statistical significance. In males, platelet count plays a stronger role in survival outcomes. In females, platelet count appears to have a weaker effect.

strongest predictors of survival

Disease stage and liver function markers are the strongest predictors of survival:

- As expected, higher disease stages (Stage 3 & 4) are strongly associated with worse survival outcomes.
- Higher Bilirubin significantly increases mortality risk, confirming its role as a key marker of liver dysfunction.
- Lower Albumin is also a strong predictor of poor prognosis, reinforcing the idea that liver synthetic function is crucial for survival.