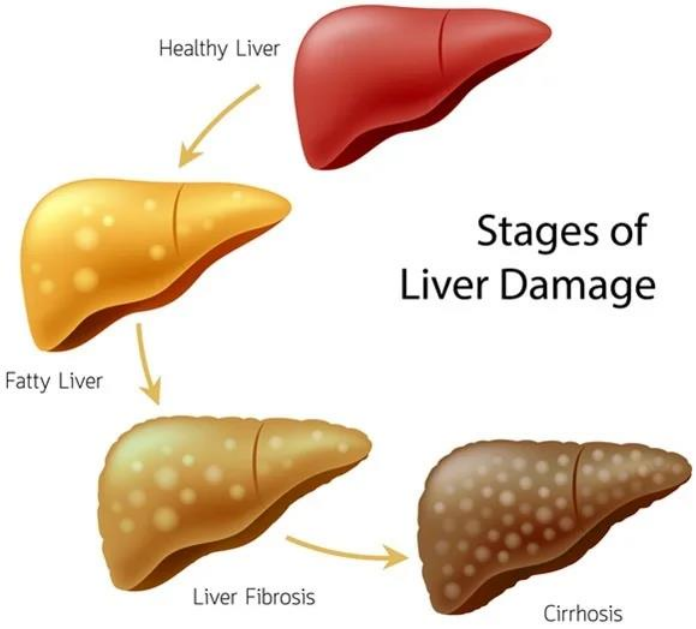
Cirrhosis Patient Survival Prediction

P8108 Survival Analysis Final Project

Chen Liang, Jessie Li, Xinyi Shang, Ixtaccihuatl Obregon, Jane Ma



Cirrhosis

- A condition in which the liver is scarred and permanently damaged.
- Scar tissue replaces healthy liver tissue and prevents your liver from working normally. As cirrhosis gets worse, your liver begins to fail.

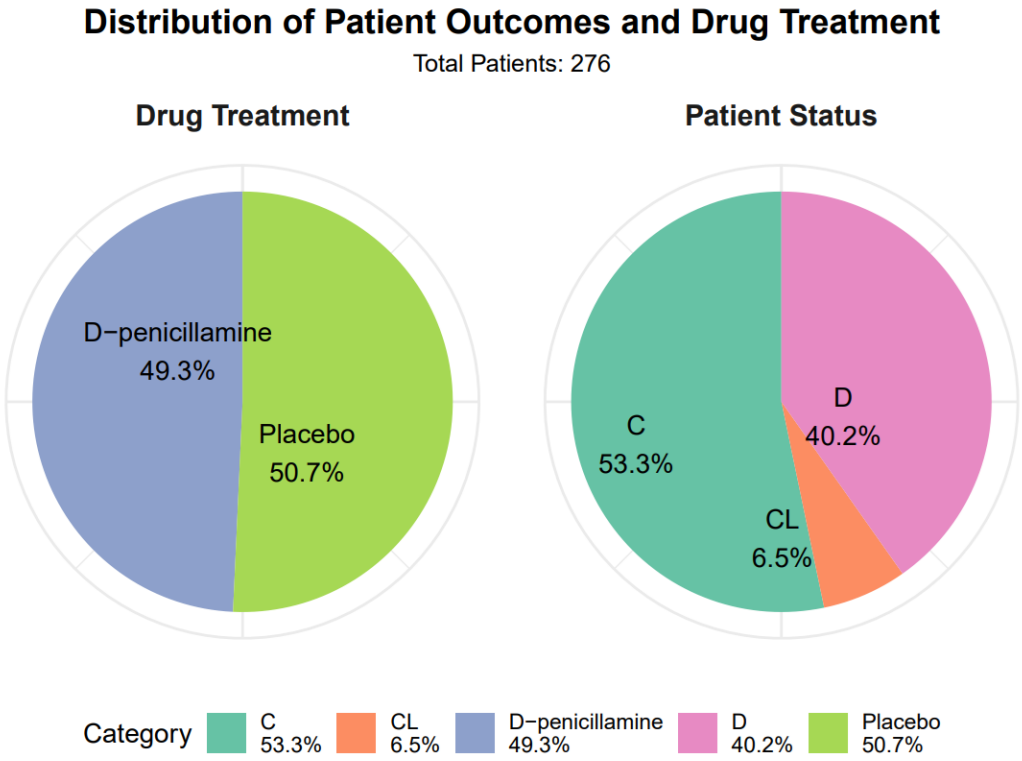
https://www.niddk.nih.gov/health-information/liver-disease/cirrhosis

Background

- Dataset Overview
  - Source: UCI Machine Learning Repository
  - Purpose: Predict survival states of patients with biliary cirrhosis (PBC)
  - Study Period: 1974–1984
  - Institution: Mayo Clinic
- Participants:
  - **424** patients diagnosed with Primary Biliary Cirrhosis (PBC)
    - **312** enrolled in a randomized, placebo-controlled trial testing D-penicillamine
    - **112** did not participate in the trial but provided basic metrics and survival data
      - **6** patients were lost to follow-up shortly after diagnosis
  - After drop missing values, there are **276** participants remaining in our analysis.

https://www.archive.ics.uci.edu/dataset/878/cirrhosis+patient+survival+prediction+dataset-1

- Total Instances: **276 patients (missing values are filtered out)**
- **Features:**
  - **Demographics**: Age, Sex
  - **Clinical Indicators**: Serum Bilirubin, Albumin, Alkaline Phosphatase, etc.
  - **Symptoms**: Presence of Ascites, Hepatomegaly, Spider angiomas, Edema
  - **Laboratory Results**: Cholesterol, Copper, SGOT, Triglycerides, Platelets, Prothrombin Time
  - **Disease Stage**: Histologic stage of disease (1 to 4)
- **Outcome:**
  - Event: D (Death)
  - Censor: C (Censored) & CL (Censored due to liver transplantation)

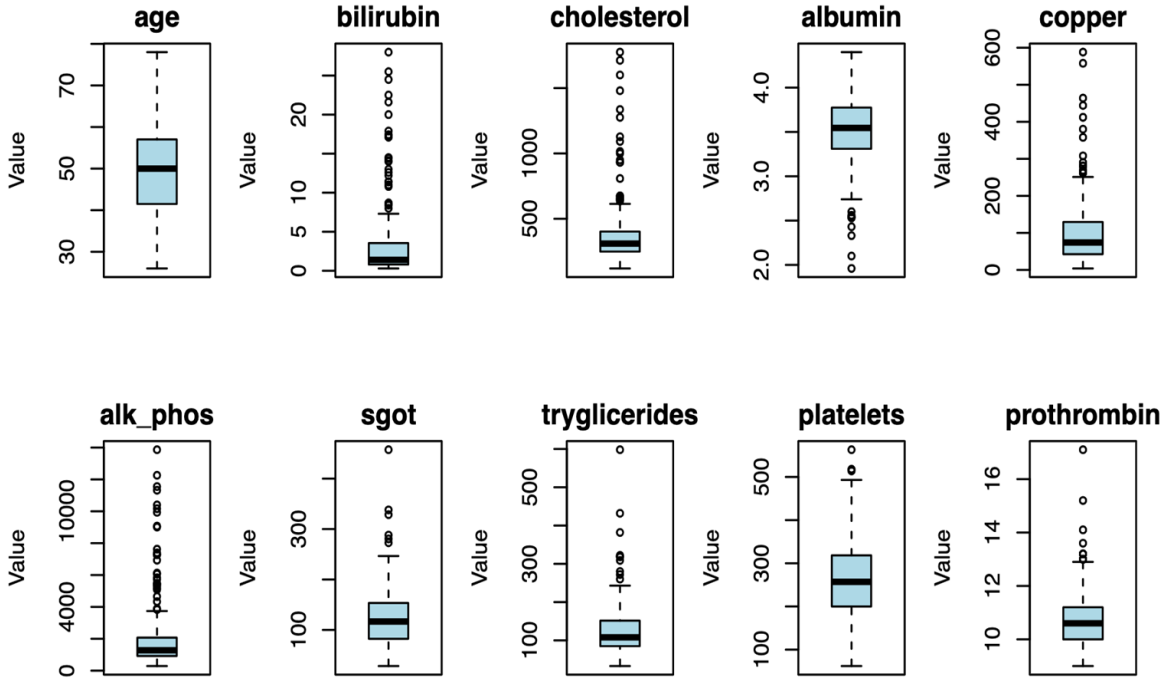


Baseline Statistics by Outcome Status

Characteristic	Censored, N = 147[1]	Censored due to liver tx, N = 18[1]	Death, N = 111[1]
N_days	2,391.8 / 2,224.0 (984.3)	1,511.6 / 1,368.0 (754.4)	1,508.5 / 1,191.0 (1,110.4)
Drug			
D-penicillamine	70 (48%)	9 (50%)	57 (51%)
Placebo	77 (52%)	9 (50%)	54 (49%)
Age	48.3 / 48.0 (10.3)	40.7 / 40.5 (6.0)	53.4 / 53.0 (10.0)
Sex			
Female	137 (93%)	15 (83%)	90 (81%)
Male	10 (6.8%)	3 (17%)	21 (19%)
Ascites	1 (0.7%)	0 (0%)	18 (16%)
Hepatomegaly	55 (37%)	12 (67%)	75 (68%)
Spiders	29 (20%)	5 (28%)	46 (41%)
Edema	8 (5.4%)	2 (11%)	32 (29%)
Bilirubin	1.6 / 0.9 (1.8)	3.2 / 3.3 (2.0)	5.7 / 3.3 (6.2)
Cholesterol	326.9 / 293.0 (168.1)	439.5 / 343.5 (335.5)	418.9 / 344.0 (277.9)
Albumin	3.6 / 3.6 (0.3)	3.6 / 3.6 (0.4)	3.4 / 3.4 (0.5)
Copper	68.1 / 52.0 (58.7)	123.3 / 101.0 (102.9)	140.3 / 121.0 (100.9)
Alk_phos	1,501.1 / 1,120.0 (1,376.8)	1,509.7 / 1,253.5 (854.4)	2,731.8 / 1,794.0 (2,765.3)
SGOT	110.2 / 97.0 (54.4)	130.2 / 123.5 (38.0)	141.5 / 134.9 (57.7)
Tryglicerides	111.1 / 103.0 (47.8)	133.9 / 124.0 (70.5)	141.8 / 124.0 (79.3)
Platelets	267.0 / 265.0 (86.4)	294.8 / 297.5 (79.9)	249.5 / 236.0 (102.1)
Prothrombin	10.4 / 10.2 (0.9)	10.4 / 10.2 (0.6)	11.2 / 11.0 (1.0)
Stage			
1	11 (7.5%)	0 (0%)	1 (0.9%)
2	42 (29%)	3 (17%)	14 (13%)
3	62 (42%)	8 (44%)	41 (37%)
4	32 (22%)	7 (39%)	55 (50%)

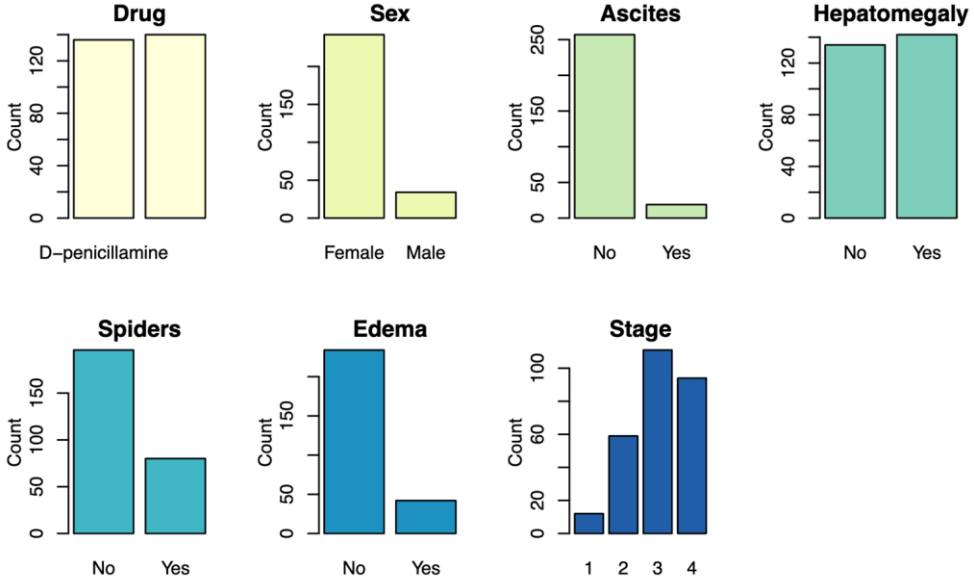
[1]Mean / Median (SD); n (%)

Exploratory Data Analysis

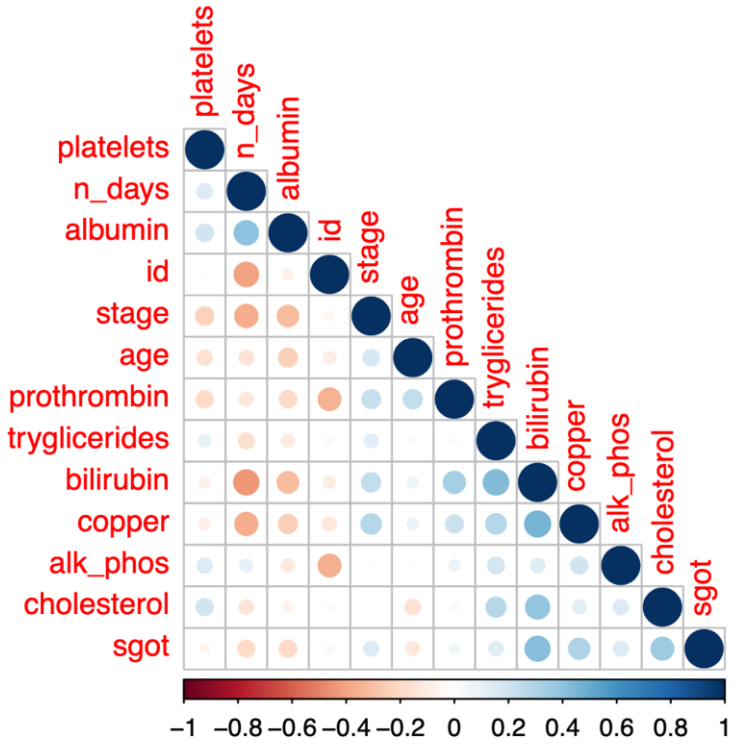


Box plots for continuous variables

Exploratory Data Analysis



Bar plots for categorical variables

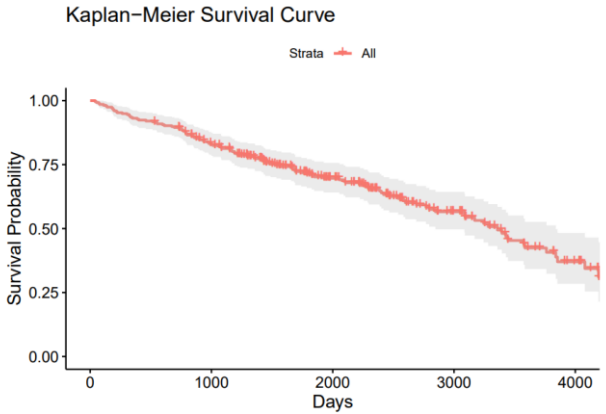


Correlation plot

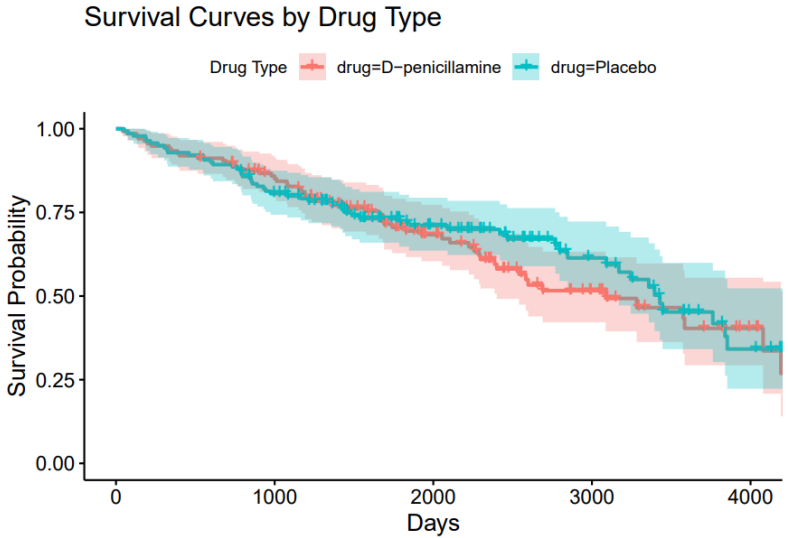
Kaplan Meier Survival Estimate

- **75% Survival time:** ~4 years
- **50% Survival time:** ~9 years

Time Interval (Years)	At Risk	Events	Censored	Survival Probability	Lower CI	Upper CI
[0, 1)	276	19	0	0.93	0.90	0.96
[1, 2)	257	10	1	0.89	0.86	0.93
[2, 3)	246	22	12	0.81	0.77	0.86
[3, 4)	212	14	29	0.76	0.71	0.81
[4, 5)	169	9	24	0.71	0.66	0.77
[5, 6)	136	6	18	0.68	0.62	0.74
[6, 7)	112	9	23	0.62	0.55	0.69
[7, 8)	80	6	15	0.57	0.50	0.64
[8, 9)	59	5	13	0.51	0.43	0.60
[9, 10)	41	6	8	0.42	0.34	0.53
[10, 11)	27	3	7	0.37	0.28	0.48
[11, 12)	17	2	10	0.31	0.21	0.45



Kaplan Meier Survival Probability By **Drug**

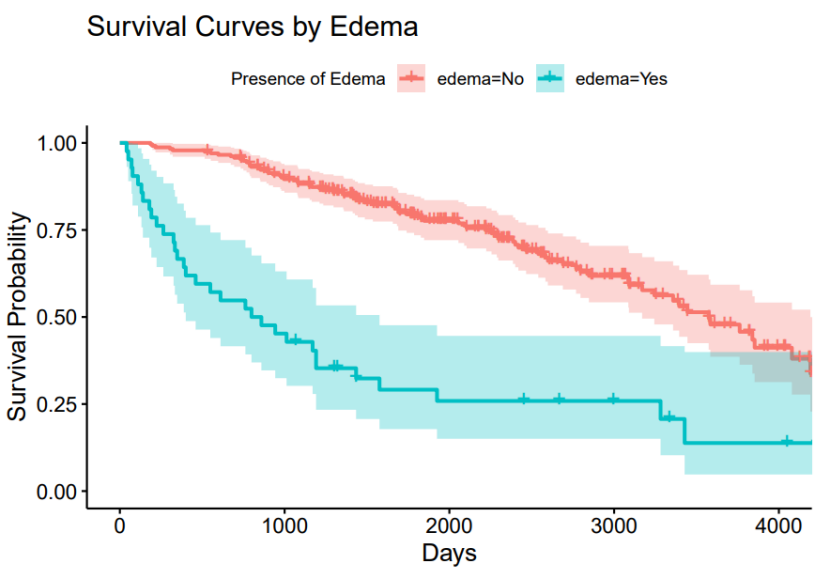


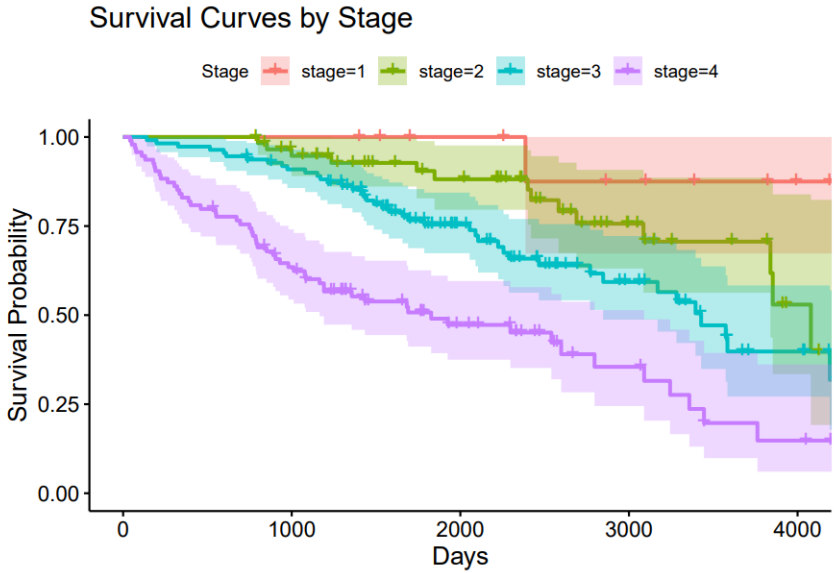
Chi-Squared Statistic	Degrees of Freedom	P-Value
0.4049	1	0.5246

Log Rank Test Result:

- **P-value:** 0.5246 (> 0.05)
- **Interpretation:** No statistically significant difference in survival outcomes between the drug groups.
- **Conclusion:** D-penicillamine does not demonstrate a statistically significant difference from the placebo group.

Kaplan Meier Survival Probability By **Edema** and **Stage**





Chi-Squared Statistic	Degrees of Freedom	P-Value
53.0933	1	<0.0001

Chi-Squared Statistic	Degrees of Freedom	P-Value
44.6499	3	<0.0001

Feature Selection

- Drug(primary variable of interest) is forced in final model
- Used AIC as the criterion

Model	Log_Lik	AIC	Kept_Variable
Forward	-467.8089	969.6179	drug, age, sex, ascites, hepatomegaly, spiders, edema, bilirubin, cholesterol, albumin, copper, alk_phos, sgot, tryglicerides, platelets, prothrombin
Forward with Interaction	-469.1367	968.2734	drug:bilirubin, drug:stage, drug:albumin, drug:copper, drug:prothrombin, drug:age, drug:sgot, drug
Backward	-469.5553	957.1105	age, edema, bilirubin, albumin, copper, sgot, prothrombin, stage, drug
Backward with Interaction	-463.5469	957.0937	drug, age, edema, bilirubin, cholesterol, albumin, copper, sgot, tryglicerides, platelets, prothrombin, stage, drug:sgot, drug:tryglicerides, drug:platelets
Stepwise	-469.5553	957.1105	age, edema, bilirubin, albumin, copper, sgot, prothrombin, stage, drug
Stepwise with Interaction	-463.5469	957.0937	drug, age, edema, bilirubin, cholesterol, albumin, copper, sgot, tryglicerides, platelets, prothrombin, stage, drug:sgot, drug:tryglicerides, drug:platelets
LASSO	-469.4033	960.8066	age, ascites, spiders, edema, bilirubin, albumin, copper, sgot, prothrombin, stage, drug

missing stage

Cox Model

PH Assumption Violations:

- Edema
- Bilirubin
- Prothrombin

If PH Assumptions are not met conduct:

- A stratified analysis
- Include a time-varying covariate to allow changing hazard ratios over time
- Include interactions with time

Table 1: Multivariate Cox Proportional Hazards Analysis

Characteristic	HR[1]	95% CI[1]	p-value
drug			
D-penicillamine	—	—	
Placebo	0.94	0.63, 1.40	0.7
age	1.03	1.01, 1.05	0.004
edema			
No	—	—	
Yes	1.47	0.88, 2.47	0.14
bilirubin	1.09	1.05, 1.13	<0.001
albumin	0.47	0.28, 0.82	0.007
copper	1.00	1.00, 1.00	0.002
sgot	1.00	1.00, 1.01	0.015
prothrombin	1.33	1.07, 1.64	0.010
stage			
1	—	—	
2	3.88	0.47, 32.1	0.2
3	5.29	0.68, 41.1	0.11
4	8.02	1.04, 61.8	0.046

[1]HR = Hazard Ratio, CI = Confidence Interval

Table 2: Proportional Hazards Assumption Test for Cox PH Model Unadjusted

	chisq	df	p
drug	0.1600772	1	0.6890854
age	2.6909476	1	0.1009198
edema	6.1134319	1	0.0134158
bilirubin	8.3071868	1	0.0039489
albumin	0.6258766	1	0.4288719
copper	0.1021024	1	0.7493211
sgot	1.3384725	1	0.2473035
prothrombin	5.0189196	1	0.0250718
stage	4.5185052	3	0.2106456
GLOBAL	24.6087203	11	0.0103973

Stratification and Time-Varying

Stratification of Edema

PH Assumption Violations:

- Bilirubin
- Overall Model

Time Varying

Dataset did not contain observations of bilirubin or prothrombin at multiple time points.

Table 3: Proportional Hazards Assumption Test COX PH Model Stratified for Edema

	chisq	df	p
drug	1.4344763	1	0.2310353
age	1.8937409	1	0.1687806
bilirubin	11.5075689	1	0.0006931
albumin	0.0049180	1	0.9440915
copper	0.3914633	1	0.5315312
sgot	1.0484233	1	0.3058705
prothrombin	2.6117691	1	0.1060734
stage	3.0850010	3	0.3787045
GLOBAL	18.9514038	10	0.0408843

Interaction with Time

Time Interaction for Bilirubin and Prohrombin

PH Assumption Violations:

- Bilirubin

Time Interaction for Bilirubin

PH Assumption Violations:

- Bilirubin

Time Interaction for Prohrombin

PH Assumption Violations:

- Bilirubin
- Prothrombin

Interaction between Covariates

- Use **Likelihood Ratio Test**
- Procedure:
  - First step: add interaction between **albumin and copper**.
  - Second step: no interaction term is added.

Table 9: Siginificant Interaction term

Interaction Term	P Value
age * copper	0.031
albumin * copper	5e-04
copper * sgot	0.0082
copper * prothrombin	0.0238
copper * stage	0.0015

Final Cox Model

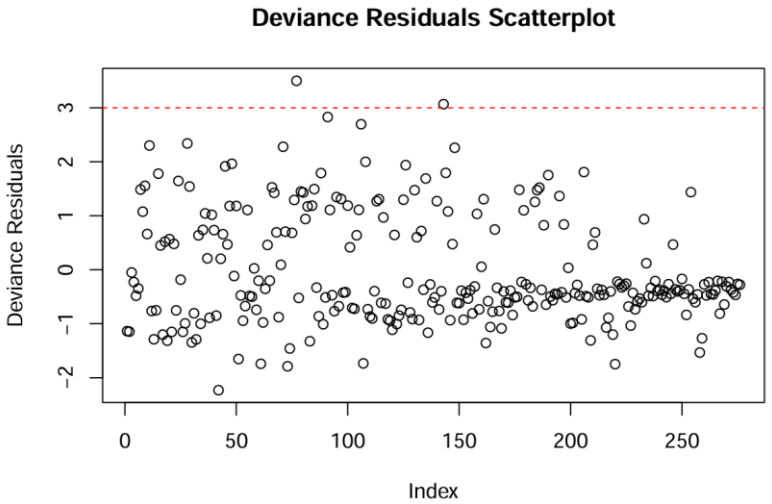
- **Drug:** insignificant negative survival impact
- **Significant Negative Impact** on Survival:
  - Age, Bilirubin, SGOT, Prothrombin
  - Stage 4 compared with Stage 1
  - Albumin-Copper **Interaction**
- **Significant Positive Impact** on Survival:
  - Albumin, Copper
  - Bilirubin-time **Interaction**

Table 10: Final Model Parameter Results

	Estimate	Hazard Ratio	p value	Sig.
drug.L	0.2406	1.2720	0.1124	
age	0.0337	1.0343	0.0025	**
bilirubin	0.2467	1.2798	0.0000	***
albumin	-1.4627	0.2316	0.0001	***
copper	-0.0224	0.9779	0.0020	**
sgot	0.0065	1.0065	0.0010	***
prothrombin	0.2819	1.3257	0.0168	*
stage2	1.3093	3.7034	0.2228	
stage3	1.6975	5.4604	0.1058	
stage4	2.0812	8.0139	0.0467	*
bilirubin:n_days	-0.0002	0.9998	0.0000	***
albumin:copper	0.0076	1.0076	0.0004	***

Model Evaluation

- Residuals:
  - **Deviance residuals**: two outliers – 77 and 143.
- Influence diagnostics:
  - **LD option**: 82, 100, 108, 129, 210
- After dropping out:
  - Signs of coefficients stay the same.
  - Estimates are slightly different.



Conclusion and Discussion

- **D-penicillamine** is proved to be inefficient.
- **Age & Stage**: Emphasize early detection and stage-specific care.
- **Bilirubin, Copper, SGOT, Prothrombin, Albumin**: Monitor liver function and metabolic health closely.
  - The protective impact of copper: Copper deficiency is proved to be a risk factor for mortality in patients with advanced liver disease.
- **Interaction**
  - **Negative Albumin-Copper Interaction:** a synergistic negative impact on survival.
    - Monitor High-Risk Patients
    - Conduct Biological Investigation
  - Investigate other interactions further for personalized treatments.

Limitation

- **Missing data**: A total of 147 observations have missing values, which could be addressed through imputation techniques.
- **Imbalanced Data**:
  - Unbalanced distribution of sex (80-90% female vs male)
  - Right-skewed distribution of Bilirubin
- **High censoring rate**: More than 50% censored data

Thank You!