LINK: <https://www.kaggle.com/datasets/fedesoriano/cirrhosis-prediction-dataset>

**Context**

Cirrhosis is a late stage of scarring (fibrosis) of the liver caused by many forms of liver diseases and conditions, such as hepatitis and chronic alcoholism. The following data contains the information collected from the Mayo Clinic trial in primary biliary cirrhosis (PBC) of the liver conducted between 1974 and 1984. A description of the clinical background for the trial and the covariates recorded here is in Chapter 0, especially Section 0.2 of Fleming and Harrington, Counting  
Processes and Survival Analysis, Wiley, 1991. A more extended discussion can be found in Dickson, et al., Hepatology 10:1-7 (1989) and in Markus, et al., N Eng J of Med 320:1709-13 (1989).

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 dataset 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.

**Attribute Information**

1) ID: unique identifier  
2) N\_Days: number of days between registration and the earlier of death, transplantation, or study analysis time in July 1986  
3) Status: status of the patient C (censored), CL (censored due to liver tx), or D (death)  
4) Drug: type of drug D-penicillamine or placebo  
5) Age: age in [days]  
6) Sex: M (male) or F (female)  
7) Ascites: presence of ascites N (No) or Y (Yes)  
8) Hepatomegaly: presence of hepatomegaly N (No) or Y (Yes)  
9) Spiders: presence of spiders N (No) or Y (Yes)  
10) Edema: presence of edema N (no edema and no diuretic therapy for edema), S (edema present without diuretics, or edema resolved by diuretics), or Y (edema despite diuretic therapy)  
11) Bilirubin: serum bilirubin in [mg/dl]  
12) Cholesterol: serum cholesterol in [mg/dl]  
13) Albumin: albumin in [gm/dl]  
14) Copper: urine copper in [ug/day]  
15) Alk\_Phos: alkaline phosphatase in [U/liter]  
16) SGOT: SGOT in [U/ml]  
17) Triglycerides: triglicerides in [mg/dl]  
18) Platelets: platelets per cubic [ml/1000]  
19) Prothrombin: prothrombin time in seconds [s]  
20) Stage: histologic stage of disease (1, 2, 3, or 4)