1. Primary Biliary Cirrhosis (PBC) Study. Consider the publicly available dataset “PBC.dat”. We **restrict the data to the first 312** PBC randomized patients (i.e., id =1 to 312). The variables in this dataset are:



1. Create a new variable ‘delta’ such that delta=0 if status is 0 or 1. Otherwise delta=1.
2. Create a new variable ’trt’ such that trt=0 if drug=2; trt=1 if drug=1.
3. Create a new variable ‘age\_yr’ such that age\_yr=age/365. This is patient age in years.
4. Create a new variable ‘logbili’ such that logbili=log(bili).
5. With censoring indicator ‘delta’, plot the Kaplan-Meier curves by ‘trt’. Use informative labels and legends.
6. What is the estimated (and 95% confidence interval for) median survival time by ‘trt’?
7. Is there a difference in survival function between placebo and drug groups?
8. Is there a difference in survival function between placebo and drug groups, controlling for stage as a potential confounding factor?
9. Fit a Cox regression model with covariates logbili, albumin, age\_yr, and edema. Interpret the HR associated with age\_yr.
10. Generate deviance residual by age\_yr plot to examine outlier(s).
11. Conduct statistical tests to assess whether the PH assumption is satisfied for logbili, albumin, age\_yr, and edema.