Discussion weeks 7 - 9

INTRODUCTION: (from Fleming and Harrington, 1991)

Quoting from F&H, Section 0.2:

"Primary biliary cirrhosis of the liver (PBC) is a rare but fatal chronic liver disease of unknown cause, with a prevalence of about 50-cases-per-million population. The primary pathologic event appears to be the destruction of interlobular bile ducts, which may be mediated by immunologic mechanisms. The data are important in two respects. First, controlled clinical trials are difficult to complete in rare diseases, and this case series of patients uniformly diagnosed, treated, and followed is the largest existing for PBC. Second, the data present an opportunity to study the natural history of disease."

"Between January, 1974 and May, 1984, the Mayo Clinic conducted a double-blinded randomized trial for PBC, comparing the drug D-penicillamine (DPCA) with a placebo. There were 424 patients who met the eligibility criteria seen at the Clinic while the trial was open for patient registration. Both the treating physician and the patient agreed to participate in the randomized trial in 312 of the 424 cases. The date of randomization and a large number of clinical, biomedical, serologic, and histologic parameters were recorded for each of the 312 clinical trial patients. The data from the trial were analyzed in 1986 for presentation in the clinical literature. For that analysis, disease and survival status as of July, 1986, were recorded for as many patients as possible. By that date, 125 of the 312 patients had died, with only 11 deaths not attributable to PBC. Eight patients had been lost to follow-up, and 19 had undergone liver transplantation. "

An 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).

OBJECTIVES:

1. Use these data to evaluate the impact of DPCA. We will use Kaplan-Meier survival curves, the log-rank test, and Cox regression to summarize the treatment comparison.
2. Use these data to construct a model that can be used for predicting survival based on a small number of measurements.

The data, mayo.dat, is on the class web page in the Datasets directory.

**Analysis Variables:**

-9 means missing for all variables

===============================================================================

 VARIABLE CODES

===============================================================================

 1. age in years

 2. albumin gm/dl

 3. alkaline phosphatase U/liter

 4. ascites 0 = no

 1 = yes

 5. serum bilirubin mg/dl

 6. serum cholesterol mg/dl

 7. edema 0 = no

 1 = yes

 8. edema treatment 0 = none and no treatment

 0.5 = edema but no treatment,

 or edema resolved

 by treatment

 1 = edema despite treatment

 9. hepatomegaly 0 = no

 1 = yes

 10. time days between

 registration and earlest of

 death, liver transplantation

 and July 1986

 11. platelets count per mm^3 blood/1000

 12. prothrombin time seconds

 13. sex 0 = male

 1 = female

 14. SGOT U/ml

 15. spiders 0 = no

 1 = yes

 16. stage 1,2,3,4

 17. censoring 1 = death

 0 = other

 18. treatment 1 = D-penecillamine

 2 = placebo

 19. trigycerides mg/dl

 20. urine copper micrograms/day

===============================================================================

312 records on the 312 randomized patients

**Week 7: May 13-17, 2013**

1. Discuss the scientific objectives of analysis.
2. Discuss the measurements.
3. Summarize the univariate distribution of each variable (focus on "Mayo Model" predictors)
4. Compare the covariates for the two treatment arms. (focus on "Mayo Model" predictors)
5. For the DPCA treatment analysis formulate the scientific question in terms of a statistical hypothesis and discuss analysis plans (primary analysis and adjusted analysis).

**Week 8: May 20-24, 2013**

1. Use Kaplan-Meier and the log-rank test to summarize the effectiveness of treatment.
2. Use Kaplan-Meier and the log-rank test to summarize the association between survival and each of the other Mayo model covariates.
3. Use Cox regression to obtain an estimate and confidence interval for the relative hazard comparing DCPA to control.
4. Formulate a Cox regression model that that could be used to estimate an adjusted treatment effect and/or consider interactions.

**Week 9: May 27-31, 2012**

1. For the second objective (developing a predictive model) discuss analysis plans.
2. Use Cox regression to build the "Mayo model" as discussed in the article by Dickson et al. (1989).
3. Assess the adequacy of the model.
4. Predict 5 year survival for 52 year old with serum bilirubin = 0.5 mg/dl, serum albumin = 4.5 gm/dl, prothrombin time = 10.1 sec, no edema, not on diuretic therapy (i.e. edemaTx = 0). Assume that S0(5) = 0.9983918 (what subgroup does this S0 value correspond to? Is it meaningful?).
5. Data on additional patients is available to check the prognostic model.
6. Conclusions?