Biostat 513: **Analysis Project** DUE: Friday June 14, 2002

DATA: The data for this project can be found on the class web page:
`http://courses.washington.edu/b513/` in the *Discussion Section* directory (ie. click on Discussion Section from the main Biostat 513 page).

Objective: The objective of this project is to have students demonstrate the skills learned in this course by analyzing a biomedical data set. In particular, students should use simple descriptive methods, and both binary response and survival methods.

Description: The data are from a clinical trial conducted to assess a lung cancer treatment. In this trial, males with advanced inoperable lung cancer were randomized to either a standard or a test chemotherapy. The primary end point for therapy comparison was time to death. Only 13 of the 137 survival times were censored (4 were censored at 500 days). As is common in such studies, there was much heterogeneity between patients in, for example, disease extent and pathology, previous treatment of the disease, demographic background and initial health status. The data include information on:

`Treatment`	`1=test, 2=standard`
`Celltype`	`1=squamous, 2=smallcell, 3=adeno, 4=large`
`Survival`	`in days`
`Status`	`1=dead, 0=censored`
`Karnofsky score`	
`Age`	`in years`
`Prior therapy`	`0=no, 10=yes`

Assignment: Working either individually or in groups of 2-3, complete an analysis of these data using two approaches:

1. Outcome #1 = vital status at 90 days. Define an indicator Y that takes the value 1 if the subject has died within ($\leq$) 90 days, and is 0 if they have lived beyond 90 days.

2. Outcome #2 = survival time.

It is typical in treatment studies that the primary aim is to determine if there is evidence for an overall treatment effect. In addition, a common secondary analysis considers adjustment for measured covariates and evaluates whether treatment appears to work better / worse for certain patient subgroups.

Your analysis should provide:

1. Present a summary of covariates by randomization group (standard versus new treatment) – [Table 1].

2. Using 90 day mortality (1=dead, 0=alive), present an unadjusted analysis of the impact of treatment.

3. Using 90 day mortality (1=dead, 0=alive), present an adjusted analysis of the impact of treatment using cell type (categorical), Karnofsky score (as a linear term), age (as a linear term), and prior treatment (categorical) as control variables. Evaluate whether the treatment appears to be more/less effective for patients defined by cell type or Karnofsky score.

4. Using survival time as the outcome, present Kaplan-Meier curves by treatment group – [Figure 1].

5. Using survival time as the outcome, present an adjusted analysis of the impact of treatment using cell type (categorical), Karnofsky score (as a linear term), age (as a linear term), and prior treatment (categorical) as control variables. Evaluate whether the treatment appears to be more/less effective for patients defined by cell type or Karnofsky score.

Written Summary:

1. (50%) Using no more than (2) pages of text, write a summary of the methods and results for the 90 day mortality outcome.

2. (50%) Using no more than (2) pages of text, write a summary of the methods and results for the survival outcome.

3. Include a table listing the summary of covariates by treatment assignment (Table 1), and a table for regression model results if used (logistic, Cox).

4. Include a figure showing survival for the two treatment groups.