

**Statistics 538
Midterm Exam
Winter 2014**

Results for the following exam should be presented in the format of a sequence of reports, one report for each problem. Code used to generate results should be appended to the end of each report. You are not to discuss, in any way, this exam with any other individual. Work on this exam is to be completely your own.

Completed exams are to be submitted by the beginning of class (12:00pm) on Friday, March 7, 2014. A penalty of 20% will be assessed for late exams.

1. Chronic Granulomatous Disease (CGD) is an inherited primary immunodeficiency disease, which increases the body's susceptibility to infections caused by certain bacteria and fungi. Individuals with CGD are unable to fight off common germs and get very sick from infections that would be mild in healthy individuals. You have been provided data from a placebo-controlled trial of gamma interferon in subjects with CGD. Of primary interest is the evaluation of the efficacy of the treatment in comparison to placebo. The data set (cgd.xlsx) consists of the following variables:

- id: subject identification number
- center: enrolling center
- treatment: placebo or gamma interferon
- sex: gender
- age: age in years, at study entry
- height: height in cm at study entry
- weight: weight in kg at study entry
- inherit: pattern of inheritance
- steroids: use of steroids at study entry, 1=yes
- propylac: use of prophylactic antibiotics at study entry
- hos.cat: a categorization of the centers into 4 groups
- time: time (in days) to either infection or end of study
- status: 1=the interval ends with an infection

Provide a complete analysis of this data.

2. In previous studies, the introduction of methotrexate-based chemotherapy has improved median survival for patients with primary central nervous system lymphoma (PCNSL). Using a newly collected set of data (pcnsl.xlsx), you have been asked to verify these findings.

The data consist of the following variables:

DS: treatment (1=methotrexate, 0=cytarabine)
Gender
KFS: Karnofsky performance status
CSFpro: cerebral spinal fluid protein
CSFcyt: cerebral spinal fluid cytology
LDH: serum lactate dehydrogenase
OS: overall survival (in days)
OSc: censoring indicator for overall survival
PFS: progression free survival (in days)
PFSc: censoring indicator for progression free survival

For this data, there are two outcome measures. First, overall survival indicates the time from treatment to death. Second, progression-free survival indicates the time from treatment to the reappearance of cancer.

Provide a complete analysis of this data, for both overall survival and progression-free survival.