MI220 CLASS PROJECT: ANALYSIS OF A PHASE 1 STUDY OF CP-XYZ

Background on CP-XYZ. You are working on the early development of CP-XYZ, for which a Phase 1 study has just been run. The primary objectives of this study were:

- To characterize the maximum concentration of CP-XYZ in plasma following single dose administration ("single dose $C_{\max}$"). (Based on preclinical toxicology there is a belief that the dose limiting toxicity for this drug will be a function of $C_{\max}$.)
- To characterize the dose-response and exposure-response of the drug on a cerebrospinal fluid (CSF) biomarker that may be related to the eventual clinical endpoint.

The key elements of the study protocol that you should be aware of are:

- Subjects were randomized to 10, 25, or 50 mg of CP-XYZ (no placebo).
- The fluid sampling schedule was as follows:
  - Plasma PK samples were collected at 0.5, 1, 2, 3, 4, 6, 8, 10, 12, 24, 48, and 72 hours following administration of the drug.
  - Urine PK samples were collected at baseline and at 8, 24, 48, and 72 hours following administration of the drug.
  - CSF biomarker sample was taken at 24 hours following administration of the drug.

The data from this study comes to you in the form of three comma delimited files.

- `demog.csv` contains the following demographic variables (all assessed at baseline):
  - CRCL: creatinine clearance
  - SEX: male or female
  - WT: weight in kg
  - SubjID: subject identifier
- `dose.csv` indicates the dose levels to which patients were randomized:
  - ID: subject identifier
  - AMT: dose level
- `conc.csv` (variables are self-explanatory)

You only need to worry about the primary objectives of the study, but note that the data sets do include data unrelated to the primary endpoints.

Your Mission. Analyze the data from this study, anticipating that your colleagues are going to ask the following questions. As in real life, these questions are somewhat vague and open-ended. The important thing is that you be prepared to give a well documented response.

- Are there any observations in this data set that appear to have been assayed and/or recorded incorrectly?
- Based on preclinical toxicology, we are concerned that values of $C_{\max} > 5000$ ng/mL may present a long term safety risk. Based only on this consideration, what is the highest dose level you would be comfortable with?
- Do biomarker levels appear to be affected by CP-XYZ?
- Does the 50 mg dose group appear to have a near-maximal biomarker response? Would you be comfortable interpolating the biomarker response at an unstudied dose level such as 35 mg?
- Does the exposure-response (i.e. biomarker response) relationship appear to depend on any of the baseline covariates?

Document your observations and conclusions in a "notebook". Imagine that you are giving this "notebook" to a friend who is going to have to represent you in a team meeting while you are on vacation on Maui. The friend is **not** going to show your notes directly to the team, so **do not** spend time making the notebook pretty, and **do not** worry about eloquence or even grammar. **Do** document your work by including key snippets of your R code with brief comments to explain what your are doing, **do** make liberal use of plots, and **do** provide a clear summary of your key conclusions. As a rough guideline, the total textual content of your notebook should be about 3-6 pages (including code snippets), and in addition there should be about 10 informative plots. This is meant to reflect a real life situation, so we are not going to tell you exactly how to go about this. The only artificial constraint is that you must do everything in R (no editing or merging data sets in Excel!)

Please use a notebook file format that we have a reasonable probability of being able to read, such as:

- a single Word document, with figures embedded, or
- a plain text file with references to external figures (e.g. "see figure pkplot.pdf"). (Please create a zip archive if you have multiple files.)

Good luck!