ST 790, Data Analysis Project
Spring 2017

N.B. This is a "closed" take-home project. Thus, you ***should not*** collaborate with or discuss this project with any other student in the class or with anyone else. Likewise, you ***should not*** seek help or clarification from anyone except the instructor. You ***are permitted*** to use the course notes, books, the internet, and any other materials you like.

The problem: You have been contacted by a group of mental health researchers who have carried out three studies in an adult population of individuals suffering from *schizophrenia*. The researchers are rather desperate, as although the studies are completed and the data ready for analysis, the statistician with whom they had worked on the design and implementation of the studies has just accepted a position at Google and resigned from her role as team statistician. The researchers have heard that you are an expert in longitudinal data analysis and would like your help with analyses of the data and interpretation of the results. They are happy to pay you a substantial consulting fee and to make you a prominent co-author on the articles they intend to submit to scholarly journals reporting on the studies and results.

Background: Schizophrenia is a chronic and severe mental disorder that affects how an individual thinks, feels, and behaves, and sufferers may seem like they are out of touch with reality, experiencing hallucinations, delusions, emotional issues, inability to focus or understand information, and a host of other problems. No known cure for schizophrenia exists, but there are a number of types of treatments that can help people with this illness lead productive lives. Individuals with schizophrenia are ordinarily treated with both medication and psychosocial interventions such as family and group therapy or individual behavioral therapy and counseling. Two main types of antipsychotic drugs are used: conventional agents, such as haloperidol, which treat mainly symptoms such as hallucinations, delusions, and confusion; and "new generation" agents that treat these and other symptoms, such as risperidone. There are also "miscellaneous" antipsychotic drus that work differently than conventional or new generation treatments.

The researchers have been studying a new "miscellaneous" drug, doxapinerdone, which they have reason to believe might be at least as, if not more, effective than existing agents for treating adults 20 years old or older with schizophrenia. In particular, one of the studies they have carried out is a clinical trial in which doxapinerdone was compared to risperidone. They have also worked with a team of pharmacologists on a study of the pharmacokinetics of doxapinerdone. These studies are described below.

Data: The researchers have data from three separate studies, each involving a different set of patients.

- *Pharmacokinetic study:* This is an intensive study of the pharmacokinetics of doxapinerdone conducted by its manufacturer in 100 adult sufferers of schizophrenia. Each received a single 1000 mg oral dose of doxapinerdone at time 0, and blood samples were drawn and assayed for doxapinerdone concentrations at several time points during the next 24 hours. In addition to concentration-time data, the researchers have recorded several subject characteristics, including gender, age, and weight.

  Each subject has also been classified as a possibly poor, intermediate, or extensive metabolizer of drugs like doxapinerdone on the basis of his/her CYP2D6 genetic polymorphism,

referred to as his/her CYP2D6 *phenotype*. The CYP2D6 gene is expressed primarily in the liver and encodes an enzyme, cytochrome P450 2D6, which is implicated in the way an individual metabolizes a drug. Elimination of a drug encompasses both metabolism of the drug and its excretion from the body. If an individual has the poor CYP2D6 phenotype, s/he may exhibit elimination of the drug at a slower rate than an individual who has the extensive (normal) phenotype. Intermediate metabolizers might show elimination rate somewhere between poor and extensive metabolizers. Because for many antipsychotic drugs it is not the drug itself but another compound that is a *byproduct* of metabolism, a so-called *metabolite*, that actually has the therapeutic effect, it is critical to understand the extent to which elimination characteristics of the drug are associated with CYP2D6 phenotype so that appropriate dosing regimens can be developed. If the effect of the drug is mainly through a metabolite, poor metabolizers may require higher, more frequent doses than extensive metabolizers to achieve the desired therapeutic effect. Likewise, extensive metabolizers may experience exaggerated effects and increased risk of adverse events on a standard dosing regimen.

The data are in the file `pk.dat`, with columns

1 Subject ID
2 Gender indicator (= 0 if female, = 1 if male)
3 CYP2D6 phenotype (= 1 if poor, = 2 if intermediate, = 3 if extensive)
4 Weight (kg)
5 Age (years)
6 Doxapinerdone concentration (mg/dL)
7 Time (hours) since dose at time 0.

- *Longitudinal study of doxapinerdone vs. risperidone:* The researchers have conducted a clinical trial comparing the short term effectiveness of doxapinerdone relative to risperidone on schizophrenia severity. Each of 300 adult patients was randomized to begin a regimen of either oral doxapinerdone or oral risperidone at week 0. Each was to return to the clinic weekly (weeks 1, 2, 3, 4, 5, 6) for assessment of his/her schizophrenia severity. Severity was assessed at each clinic visit using the Brief Psychiatric Rating Scale (BPRS), a sum of scores on 18 items on a questionnaire that reflect a subject's behaviors, mood, and feelings. The BPRS is a standard way of measuring severity of schizophrenia and other mental health disorders. Higher BPRS scores indicate higher severity of schizophrenia.

  Three months after the start of the study (so at week 12), each subject was contacted and asked to rate the current (at week 12) severity of his/her symptoms as mild/moderate or high.

  At baseline, information on each individual's gender, weight, age, and smoking status was also collected.

  The data are in file `bprs.dat`, with columns

  1 Subject ID
  2 Treatment (= 0 if doxapinerdone, = 1 if risperidone)
  3 Gender indicator (= 0 if female, = 1 if male)
  4 Weight (kg)
  5 Age (years)
  6 Smoking status (= 0 if non- or former smoker, = 1 if current smoker)
  7 Time (weeks)
  8 BPRS score
  9 Self-reported severity at week 12 (= 0 if mild/moderate, = 1 if high)

Note that some subjects did not complete all clinic visits, although all were contacted successfully to provide self-reported severity at week 12.

- *Longitudinal study of doxapinerdone in addition to behavioral therapy:* The researchers also conducted a larger study of the effectiveness of doxapinerdone when it prescribed in addition to individual and group behavioral therapy. 500 adult sufferers of schizophrenia experiencing mild or moderate symptoms and participating in individual and group behavioral therapy interventions were randomized to start either a regimen of placebo or doxapinerdone. At baseline (month 0) and at months 1, 3 and 6, each subject was classified by mental health experts as either responding or not responding to his/her treatment (behavioral therapy plus placebo, so behavioral therapy alone, or behavioral therapy plus doxapinerdone).

  At baseline, information on each individual's gender and age was also recorded.

  The data are in the file `response.dat`, with columns

  1 Subject ID
  2 Gender indicator (= 0 if female, = 1 if male)
  3 Age (years)
  4 Treatment (= 0 if placebo + behavioral therapy, = 1 if doxapinerdone + behavioral therapy)
  5 Time (months)
  6 Response status (= 0 if no, = 1 if yes)

Major questions to be addressed: The researchers would like you to address the following questions based on the above studies.

- Is there evidence to suggest that the pharmacokinetic properties of doxapinerdone, such as absorption, distribution, and elimination, are systematically associated with subject characteristics? Which characteristics?
- In particular, is there evidence that the elimination characteristics of doxapinerdone are associated with CYP2D6 phenotype? Can you provide estimates of the typical value of doxapinerdone clearance rate for poor, intermediate, and extensive metabolizers and the extent to which clearance varies in the population of adult schizophrenia sufferers? Are these typical values different depending on the gender, weight, or age of a patient?
- Is there evidence that average BPRS score prior to treatment with doxapinerdone or risperidone is associated with gender, weight, age, or smoking status? What is the nature of the pattern of change of BPRS score after initiation of treatment with each agent? Is the average rate of change in BPRS score among adult schizophrenia patients different for doxapinerdone and risperidone? Is it associated with a patient's gender, weight, age, or smoking status at baseline? What is the average rate of change for each drug for male patients of average weight and age who are smokers? Non-smokers?
- Is it possible to estimate the rate of change in BPRS score for each study participant? Is there any evidence that the odds of experiencing self-reported high severity of schizophrenia after 3 months in this population are associated with an individual's rate of change in BPRS score during the first 6 weeks of treatment with either drug?
- Is there evidence in the data to suggest that the percentage of schizophrenia patients in the population who are classified as responders to behavioral therapy prior to initiation of

placebo or doxapinerdone is different for males and females? Depends on age? What are these percentages?

- Is there evidence to suggest that the odds of being classified as a responder increase over 6 months for either placebo or doxapinerdone? If so, is there evidence that the odds show a more rapid increase among schizophrenia patients taking doxapinerdone in addition to participating in behavioral therapy than in patients participating in behavioral therapy with no medication (so taking only placebo)?

Your job: The researchers will pay you the very large consulting fee to conduct appropriate analyses of the data and to provide a detailed report describing what you did, why you did it, and what inferences can be made on the above questions.

Your report: You should write a formal report *for the researchers* (*not for me*). Although the researchers are familiar with basic statistical models and methods, such as linear and logistic regression, they have *no familiarity* with longitudinal data models or methods. Thus, your report will need to explain what you did, why you did it, what your conclusions are, and any possible pitfalls or caveats at a level that the researchers can understand. They will need to incorporate information from your report into the papers they write, which will be reviewed by others who are also likely to know very little about longitudinal data analysis.

Accordingly, your grade will be based on how well you communicate and justify the statistical modeling and analysis choices you make and how well you explain and interpret the results. A good report will explain in a non-technical way

- Why specialized statistical models and methods are required for longitudinal analysis and why methods familiar to the researchers are not appropriate;
- The basic features of the statistical model you have chosen for each analysis you present and how the researchers' questions can be stated formally in terms of the model;
- Your rationale for the modeling choices you have made and any assumptions that are involved and why these are reasonable;
- The method used to fit your models;
- Each step of your analyses in "layman's" terms (not "technical" terms);
- The results and careful interpretation of the analyses in terms of the science.

The researchers will not be interested in seeing lots of equations, formulæ, and matrices because they will not understand them. Thus, you will need to explain things to them mainly in words, with very few equations and symbols (being sure to define any symbols you do use). The researchers should not have to encounter terminology or concepts that they are not likely to know or that are not explained, nor should they have to search through programs and output, which they are sure not to understand, to find results.

Your report should adhere to the following requirements:

- It should be *typed*.

- A good statistical data analysis report always provides background on the situation, gives a general statement of the problem (even if both you and your collaborators know what it is), summarizes the data (this is often most effective when done graphically), and states clearly the scientific objectives and why they are of interest. It also always summarizes what was done and gives clear a statement of the conclusions regarding questions of interest from a subject-matter perspective. Thus, your report should be organized into a sequence of sections that presents all of this in a logical way.
- Any code that produces results cited in your report and the associated output should be included as an *appendix* to the report. *However*, this is for my information only; your report *should not* refer to the appendix. (I.e., the report should not ask the researchers to go to pages of code or output; any results that the researchers need to see should be cited in the body of your report. Keep in mind that the researchers will know nothing about SAS procedures and R functions you might use nor how to interpret their output.)