ST 790, Data Analysis Project
Spring 2018

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: A research team specializing in gerontology has contacted you for assistance with the analysis of data from several studies focused on the treatment of depression in individuals who are 60 or older. All of the studies are longitudinal, and the researchers are not acquainted with any statisticians with experience in longitudinal data analysis. Thus, having heard that you are a noted expert in this area, the researchers have contacted you and are willing to pay you a handsome consulting fee for you to assist them with appropriate analyses and to make you a co-author on any publications that emerge from this work.

Background: Depression can occur at any age and can be an especially difficult condition to treat. Depression in later life can be very different from that in younger individuals. Senior adults often report more physical and cognitive difficulties, including fatigue, headaches, forgetfulness, and slowed thinking, than younger adults, who instead more often contend with the sadness or despair that is typically associated with this disorder. Older adults thus tend to attribute symptoms of depression to "old age," and physicians and family members may miss the signs and view them as reflecting the "slowing down" expected in the elderly. Thus, depression in this population is often underdiagnosed and untreated. At the same time, depression in the elderly often increases the risk of cardiovascular disease and death from a host of illnesses, as well as suicide, especially in elderly white men.

Accordingly, there is considerable recent interest among researchers who study geriatric populations in assessing and treating depression in the elderly. Antidepressant drugs are one option, particularly for those with more severe cases of depression or those for whom other interventions, such as psychotherapy, are not effective, but have been not as well studied in elderly adults as they have been in younger adults. These drugs may take longer to start working in older people, and because the elderly tend to be more sensitive to medications than younger adults and thus experience more side effects, different dosing regimens may be needed.

The majority of patients of all ages are prescribed antidepressants taken orally. In younger adults, there is some evidence that initiating certain types of antidepressant therapy by daily intravenous infusion for several weeks and then switching to the oral version may be more effective than using oral medication from the outset for patients with moderate to severe depression. However, this strategy has not been widely studied in elderly patients.

The researchers are interested in the potential of using a relatively new selective serotonin reuptake inhibitor drug (SSRI; a class of antidepressants), bidalopram, in this way in moderately to severely depressed elderly patients. They have conducted a clinical study in which the efficacy of this strategy has been evaluated by comparing it to placebo. They have also carried out a clinical trial comparing the infusion+oral strategy with bidalopram to oral-only bidalopram. Finally, they have worked with a team of pharmacologists on a study of the pharmacokinetics of bidalopram when administered by intravenous infusion in elderly patients. These studies are described below.

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

- *Pharmacokinetic study:* This is an intensive study of the pharmacokinetics of bidalopram administered by intravenous infusion in 60 subjects aged 60 years or older. Each received a 30 minute (0.5 hour) intravenous infusion of 1000 μg/hour of bidalopram starting at time 0. Blood samples were taken and assayed for bidalopram concentrations at several time points during the infusion (up to 0.5 hours) and then at several subsequent times. In addition to concentration-time data, the researchers recorded several characteristics, including gender, age, and weight, and an indicator of whether or not the subject suffered from renal impairment (kidney dysfunction; individuals whose kidneys are impaired may take longer to eliminate drugs from their systems).

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

  1 Subject ID
  2 Gender indicator (= 0 if female, = 1 if male)
  3 Renal impairment (= 0 if no, = 1 if yes)
  4 Weight (kg)
  5 Age (years)
  6 Bidalopram concentration (ng/ml)
  7 Time (hours) since dose at time 0.

- *Longitudinal study of bidalopram vs. placebo:* The researchers have conducted a clinical study to evaluate the efficacy of a regimen of bidalopram administered by intravenous infusion daily for two weeks followed by daily oral administration in elderly subjects suffering from moderate to severe depression as determined by the Hamilton Rating Scale for Depression (HAM-D). The HAM-D is an observer-rated scale that evaluates core symptoms of depression based on 17 items, where each items scored on either a 0-to-4 scale (0 = none/symptom absent to 4 = most severe) or a 0-to-2 scale (0 = none/absent to 2 = severe). The HAM-D score is the sum of the scores over the 17 items and ranges in value from 0 to 54. The HAM-D score is used to classify an individual as normal (HAM-D < 9), mildly depressed (10 to 13), mildly to moderately depressed (14 to 17), or moderately to severely depressed (> 17).

  The study involved 200 adults age 60 years or older with HAM-D score equal to 17 or greater at baseline. Subjects were randomly assigned to receive a daily intravenous infusion for two weeks starting at baseline followed by oral administration of either a placebo or bidalopram. At weeks 2, 4, 6, 8, 10, and 12, the HAM-D score was to be ascertained on each subject.

  At the end of the study (week 12), each subject was also asked to provide his/her own self-assessment of whether or not he/she was feeling moderately to severely depressed.

  At baseline, information on each individual's gender, age, marital status, whether or not he/she had suffered previous depression as a younger adult, and whether or not he/she was currently suffering from chronic pain was also recorded. Depression is known to occur at higher rates among elderly individuals who are single/widowed, have suffered previous bouts of depression, or contend with significant pain due to other conditions.

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

1 Subject ID
2 Treatment (= 0 if placebo, = 1 if bidalopram)
3 Gender indicator (= 0 if female, = 1 if male)
4 Age (years)
5 Marital status (= 0 if married, = 1 if single/widowed)
6 Previous depression (= 0 if no, = 1 if yes)
7 Chronic pain (= 0 if no, =1 if yes)
8 Time (weeks)
9 HAM-D score
10 Self-assessment at week 12 (= 1 if moderately/severely depressed, = 0 otherwise).

Note that some subjects apparently did not complete all HAM-D assessments, but the self-assessment was obtained from all at week 12.

- *Longitudinal study of infusion strategy vs. oral bidalopram:* The researchers also conducted a trial to compare the effectiveness of the bidalopram regimen involving initiation by intravenous infusion followed by oral administration to the usual oral-only regimen. 300 subjects living in several continuing care retirement communities who had been diagnosed previously as suffering from depression were recruited to participate and were randomized to the two regimens. At baseline (week 0), prior to start of the regimens, each was evaluated for depression using the Geriatric Depression Scale (GDS), which is based on 15 "yes/no" questions; a score of ≥ 5 "yes" responses indicates depression.

  At weeks 2, 4, and 8 thereafter, each subject was again evaluated for depression using the GDS.

  At baseline, information on each individual's gender, age, whether or not he/she suffered from mild dementia, and whether or not he/she was currently suffering from chronic pain was also recorded.

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

  1 Subject ID
  2 Gender indicator (= 0 if female, = 1 if male)
  3 Age (years)
  4 Dementia (= 0 if none, = 1 if mild)
  5 Chronic pain (= 0 if no, = 1 if yes)
  6 Treatment (= 0 if oral-only regimen, = 1 if infusion regimen)
  7 Time (weeks)
  8 GDS score (= 0 if GDS < 5, = 1 if GDS ≥ 5)

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 bidalopram are systematically associated with subject characteristics? Which characteristics?
- Please provide estimates of the typical values of bidalopram clearance rate for individuals in this population who do not suffer from renal impairment and for individuals who do suffer from renal impairment. Are these typical values different depending on the gender, age, or weight of an individual? Please also provide an estimate of the extent to which bidalopram clearance rate varies in this population.

- What subject characteristics are associated with average HAM-D score prior to treatment with bidalopram or placebo? What is the nature of the pattern of change of HAM-D score after the start of each regimen? In particular, is there evidence that HAM-D score decreases over the study period for either regimen? Is the average rate of change of HAM-D score at any time different for elderly individuals following the bidalopram regimen and those receiving placebo? Is the average rate of change associated with any of the subject characteristics recorded in the study? What is the average rate of change of HAM-D score for subjects receiving bidalopram?

- Are the odds that an individual from this population taking either bidalopram or placebo assesses him/herself to be moderately or severely depressed at the end of the study (week 12) associated with his/her individual rate of change in HAM-D score during the study? Can you describe this association? Namely, are the odds lower the more dramatic the decrease in HAM-D score during the study?

- In the population of older adults previously diagnosed with depression and living in continuing care retirement communities, what is the proportion who are depressed (as reflected by GDS) prior to treatment with bidalopram? Is this proportion different for those suffering mild dementia and those who are not? For those suffering chronic pain and those who are not? For males and females? Can you provide estimates of these proportions?

- Do the odds of being depressed (as reflected by GDS) decrease over the study period under treatment with either of the infusion or oral-only regimens of bidalopram in this population? Do they decrease more rapidly under the infusion regimen than under the oral regimen?

Your job: You will receive the handsome consulting fee offered by the researchers by conducting appropriate analyses of the data from each study and providing them with 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*.
- As discussed in Appendix F of the course notes, 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* 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.)