Biostatistics 537: Longitudinal Data Analysis - Fall 2009
Problem Set 5 - Due: Thursday, November 5, 2009

The data for this problem come from a randomized clinical trial to determine the efficacy of the drug auranofin versus placebo in the treatment of rheumatoid arthritis (Bombardier *et al.*, 1986, Auranofin therapy and quality of life in patients with rheumatoid arthritis *American Journal of Medicine, 81,* 565-578).

The outcome of interest is self-assessment of arthritis, coded as 0=poor and 1=good. Individuals were randomized to one of the two treatment groups after baseline self-assessment of arthritis, and then took the treatment for the length of the study. All 294 patients were observed at baseline, and 1 and 3 months after randomization. The self-assessment of arthritis at these three timepoints represent the responses of the dichotomous dependent variable across time. These data and some SAS syntax for this problem are provided at the class website: `http://www.uic.edu/classes/bstt/bstt513`. The variables in the dataset are `ID` (subject ID), `ARTHRIT` (dependent variable coded 0=poor and 1=good), `TXAURA` (treatment group coded 0=placebo and 1=auranofin), and `TIME` (coded 0=baseline, 1=month 1, 2=month 3).

For this problem, following the SAS syntax file on the web, use dummy-codes for the effect of time treating baseline as the reference cell. Let's call the resulting two dummy-codes `TIMEA` and `TIMEB`. Given that randomization was done after the baseline assessment, and therefore subjects are statistically the same at baseline, use the following regressors: `TIMEA, TIMEB, TX*TIMEA, TX*TIMEB`. More discussion on this issue can be found in the excellent book *Applied Longitudinal Analysis* by Fitzmaurice, Laird, and Ware (2004) on pages 126-132.

1. Based on the observed correlations select a reasonable working correlation structure and estimate a GEE logistic regression model with the above 4 regressors. Describe the meaning of the estimated regression coefficients in this model. Is there evidence of a treatment effect? If so, is the effect seen at both follow-ups? How well does the model fit the observed proportions for the two groups at the three timepoints?

2. Estimate a random-intercepts (*i.e.*, mixed) model with the above 4 regressors. Describe the meaning of the estimated regression coefficients in this model. Is there evidence of a treatment effect? If so, is the effect seen at both follow-ups? How well does the model fit the observed proportions for the two groups at the three timepoints?

3. How do the estimates from (1) and (2) relate to each other?

4. What do you conclude about the efficacy of auranofin in the treatment of arthritis? Descriptive statistics, like marginal proportions or odds ratios, may help elucidate your results. Write up your final conclusions in a way that a medical doctor could understand.