ADNI Biostatistics Conference Call
4 August 2009

Present on call: Laurel Beckett, Danielle Harvey, Bongin Yoo, Bill Billings, John Kornak.

Danielle reported on the data submission. Most data came in by July 1, as requested. We have near-complete data on MRI for some variables and some labs, and also for PET. There are some new variables developed since last major analysis in Spring, for example, 3 from Norman Foster's lab, including one that is cross-validated.

We are currently running summary reports and checks on all data submitted for the analytic plan. One set of problems that has shown up is inconsistent dates, especially in the clinical dataset; these can be uncovered by looking for inconsistent sequence within person or inconsistencies between different forms from the same or near-same visit. Danielle is forwarding errors to the UCSF group to get them corrected in primary data.

We expect the standard reports to go out next week. (Most of them have been run; we are just formatting and polishing up.)

We would like to know what level of detail people want. We are running separate reports of each individual variable. Then we prepare a summary of all the general findings and comparative summary.

Bill Billings would like to get a copy of the sample size calculations document. Danielle has also written a paper on methods for comparing potential markers. She will send these to Bill.

Laurel reports that we are planning many regression models, to try to address the scientific hypotheses the group has laid out for ADNI and ADNI II. We have organized the variables into groups, based on the stage of progression at which we expect to start seeing changes, and what that should predict down the line. The earliest markers should be CSF markers (especially Abeta and ratios) and PIB. Next, we expect to see changes in the FDG PET and MRI measures. Finally, the changes should show up in cognitive and functional measures. Changes may be modified by genetics and/ or education. The models reflect this series of hypotheses (as laid out in our analytic plan).

Statistical challenges show up in several places. The cognitive outcomes we will be using (CDR sum of boxes, MMSE, ADAS-COG, Verbal Learning) have plenty of repeated measures, out to 3 years in some cases. But the CDR and MMSE have floor/ ceiling effects. We think we can adapt to this by taking 30-MMSE to analyze "errors" rather than correct scores, and then using generalized linear models with a variance-proportional-to-mean error structure.

Most of the imaging measures also have repeated measures and are amenable to standard linear models approaches. Some markers, however, either have only two time points (CSF, PIB) or are reported as differences alone (boundary shift integral). Analyses of simple differences will likely use standard regression modeling. Analysis of two-time-point data will be more complicated. Common approaches that use difference as the outcome and baseline level as a covariate are suspect in the case where baseline level is itself a reflection of the speed of progression. Using percent change is not a good idea in cases where the denominator may be small and the measure unstable. Our

current thinking is that we will need to review each such marker individually, look at the bivariate distribution of the two observations, and choose an analytic strategy that captures the salient features of change in that marker. John Kornak concurs.

We have a smart site and are using it internally to share work in progress. For the next call, we will send out summaries of progress, and ask for feedback from the group.

The next call is 1 September, 10 AM Pacific Time.