**Project 4: analysis plan and sample size calculation**

Wenru Zhou

10 December 2018

**Data analysis plan:**

We will use linear regression to conduct the analysis. The assumption of outcome is normal distribution. If the outcome is skewed, appropriate transformation should be performed. Covariates that will be adjusted in models are age, sex, severity level (CDR of 0 and 0.5), present or absent of E4 allele, BMI, history of hypercholesterolemia, NSAID use; immune-related health conditions. To correct the inflated type I error caused by multiple comparison of six predictors, Bonferroni adjustment should be used and the alpha level is set as 0.008. Power is set as 0.8. We use best subset method to perform model selection (models with highest adjusted R-square and lowest Cp and with main predictors in them).

Aim1:

Hypothesis a: We are going to fit 12 sets of models: there are six kinds of models corresponding to six markers of cytokine and chemokine predictors: IL-6, TNF-alpha, MCP-1, Eotaxin-1, Beta-2 microglobulin, ACT at baseline. For each model, the outcomes are the change in cortical thickness and the change in memory over one year. Additional covariates that should be adjusted are Baseline cortical thickness and baseline memory separately corresponding to change in cortical thickness as the outcome and change in memory as the outcome.

Hypothesis b: All are the same as hypothesis a, except that we will use change in markers as main predictors instead of markers at baseline.

Aim 2:

Hypothesis a: We are going to fit 6 sets of models: there are six kinds of models corresponding to six markers of cytokine and chemokine predictors: IL-6, TNF-alpha, MCP-1, Eotaxin-1, Beta-2 microglobulin, ACT at baseline. For each model, the outcome is amyloid deposition at baseline.

Hypothesis b: We are going to fit 12 sets of models: there are six kinds of models corresponding to six markers of cytokine and chemokine predictors: IL-6, TNF-alpha, MCP-1, Eotaxin-1, Beta-2 microglobulin, ACT at baseline. And for each model we will add amyloid at baseline. Interaction between each marker and amyloid at baseline will be tested. The outcome for each model is the change in memory or change in cortical thickness.

**Sample size and detectable difference calculation:**

Given the fact that the correlation between all outcomes and main predictors are from 0.2 to 0.6. We are going to use correlation=0.2 to estimate the maximum sample size. The initial sample size=298. Considering about 10% loss follow up. The final sample size should be 332 to meet all our need. The markers should be dichotomized into high and low group according to medians. The effect size (Cohen’s q) of slope in interaction between each marker and amyloid, or each marker and cortical thickness is 0.41, corresponding to correlation of 0.39 using inverse Fisher Z-transformation. One limitation is that confounders are not included in sample size calculation, and number of precision predictors and confounders are unknown. Therefore we cannot conclude if these calculations are conservative or not. Results were computed using G\*Power 3.1