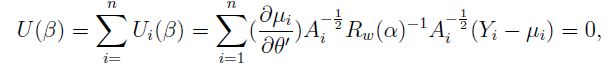
Testing association with longitudinal data is similar to testing association with multiple traits. Zhang et. al. recently published an article about testing association with multiple traits in generalized estimation equations (GEE) framework [1]. Instead of testing a single trait, they were testing the association between multiple neuroimaging and neuropsychological phenotypes as intermediate phenotypes for Alzheimer's disease on a single SNP. Since phenotypes are most likely correlated spatially, authors proposed to use different working correlation matrix in the GEE to adjust for such correlations, for example working independence, exchangeable, autoregression 1-step and unstructured structure. GEE has the nicest property that under misspecification of the correlation structure, i.e. working correlation matrix is not equal to the underlying real correlation matrix (this is usually the case for real data analysis), GEE can still give consistent estimates under mild regulations.

The classical generalized linear model (GLM) is:

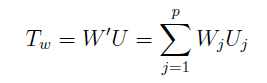
*g*(*µi*) = *ηi* = *Ziϕ* + *Xiβ* = *Hiθ*

with *Hi* = (*Zi, Xi*)*, θ* = (*ϕ', β'*)*'*and *g*(*.*) as a suitable link function. The estimates of *β* and *ϕ* can be obtained by solving the GEE [2]:



The U vector thus contains the information of the association strength between multiple traits and a single SNP in Zhang et. al. They had not taken into consider the SNP-set-based association test, which groups multiple neighboring SNPs into a SNP set (e.g. SNPs within a gene boundary). However, their work within GEE framework can be easily extended to allow for multiple traits - multiple SNPs association test.

In longitudinal data analysis, we have single trait but instead we will have multiple measurement across time points for one single subject. We could use the same candidate working correlation matrix to model the correlations among measurements. By solving the GEE, we can get the U vector which stores the association information between the longitudinal response and SNP(s). Once we get the U vector, we can recourse to a few existing tests including Sum, SSU, Score and UminP tests to calculate the significance level. Let we look at a general form of weighted U test:



*W* here represents a choice of weight vector by that researchers want to penalize/reward each SNP in the SNP set. The Sum test use a vector of 1 as *W*; the SSU test use U' as *W*; Score test is similar to SSU test but adds the denominator part, which is the variance-covariance matrix of U vector; UminP test is similar to Sum test, but only consider the largest absolute value of the whole U vector.

Different tests have their own advantages. For example, when signal effects are in the same direction for a SNP set, Sum test is preferred; when signal effect directions mix, and they are cancelling out each other in a SNP set, SSU test is preferred; when Linkage Disequilibrium (LD) structure matters among SNPs in a SNP set, Score test is preferred; when signals from a SNP set is extremely sparse, UminP is preferred. According to real data situation, one or more tests can be considered together to evaluate the significance level of the association between the longitudinal trait and the SNP set.

# Reference

1. Zhang Y, Xu Z, Shen X, et al. Testing for association with multiple traits in generalized estimation equations, with application to neuroimaging data. Neuroimage 2014;**96**:309-25 doi: 10.1016/j.neuroimage.2014.03.061[published Online First: Epub Date]|.

2. Liang K-Y, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika 1986;**73**(1):13-22