



The University of Chicago  
Department of Statistics

Ph.D. Seminar

---

**ZUOHENG WANG**

Department of Statistics  
The University of Chicago

**Statistical Methods for Genetic Association Mapping  
of Complex Traits with Related Individuals**

**WEDNESDAY, July 8, 2009, at 3:00 PM**  
**110 Eckhart Hall, 5734 S. University Avenue**

**ABSTRACT**

We develop statistical methods to address both dependent and partially-observed data and apply these methods to problems in haplotype-based association analysis of complex traits in related individuals. We consider a general setting in which the complete data are dependent with marginal distributions following a generalized linear model. We form a vector  $\mathbf{Z}$  whose elements are conditional expectations of the elements of the complete-data vector, given selected functions of the incomplete data. Assuming that the covariance matrix of  $\mathbf{Z}$  is available, we form an optimal linear estimating function based on  $\mathbf{Z}$ , which we solve by an iterative method. This approach allows us to address key difficulties in the haplotype frequency estimation and testing problems in related individuals: (1) dependence that is known but can be complicated; (2) data that are incomplete for structural reasons, as well as possibly missing, with different amounts of information for different observations; (3) the need for computational speed in order to analyze large numbers of markers; (4) a well-established null model, but an alternative model that is unknown and is problematic to fully specify in related individuals. We apply the method to test for association of haplotypes with alcoholism in the GAW 14 COGA data set.

We then address the problem of testing for association with untyped variants. Current genome-wide association studies only genotype a subset of all genomic variants. However, many of the untyped variants can be well-predicted from typed variants, with information on the joint distribution of typed and untyped variants available from an external reference panel such as HapMap. Incorporation of such external information can allow one to perform tests of association between untyped variants and phenotype, thereby making more efficient use of the available genotype data. We introduce a new method, ATRIUM, to test for association between untyped SNPs and phenotype, based on genome screen data in case-control samples in which some individuals are related. ATRIUM uses information from an external reference panel to specify a direction in which to perform a 1-df test of association with an untyped SNP. It properly accounts for uncertainty in haplotype information based on unphased genotype data and for dependence of genotypes across related individuals. We demonstrate that ATRIUM is robust in that it maintains the nominal type I error rate even when the external reference sample is not well-matched to the case-control sample. We apply the method to detect association between type 2 diabetes and variants on chromosome 10 in the Framingham SHARe data.