

# Association Studies with Known and Unknown Population Structure

Mary Sara McPeck

Assistant Professor,  
Department of Statistics, Department of Human Genetics,  
The University of Chicago

## **Abstract:**

Common diseases such as asthma, diabetes, and hypertension, which currently account for a large portion of the health care burden, are complex in the sense that they are influenced by many factors, both environmental and genetic. One fundamental problem of interest is to understand what the genetic risk factors are that predispose some people to get a particular complex disease.

Technological advances have made it feasible to perform case-control association studies on a genome-wide basis. The observations in these studies can have several sources of dependence, including population structure and relatedness among the sampled individuals, where some of this structure may be known and some unknown. Other characteristics of the data include missing information, and the need to analyze hundreds of thousands or millions of markers in a single study, which puts a premium on computational speed of the methods. We describe a combined approach to these problems which incorporates quasi-likelihood methods for known structure with principal components analysis for unknown structure.