GENETIC ANALYSES USING FAMILY-BASED SURVEY DATA

Yan Li
University of Maryland at College Park, MD, USA, yli6@umd.edu

In population-based household health surveys and case-control studies that collect disease risk factors and DNA samples, for example, the National Health and Nutrition Examination Survey (NHANES) and The US Kidney Cancer Study (KCS), a case-control study, offer the opportunity to unbiasedly estimate genetic frequencies, test for Hardy-Weinberg equilibrium, or study genetic associations with prevalent health related conditions (phenotypes) in well-defined target populations. These types of genetic analyses are typically not populationbased in most genetic studies, as these studies consist of nonrepresentative samples of the target population. Household surveys, such as NHANES, however, collect population representative sample using multistage geographical cluster sampling, where at the last stage blood-related individuals are often selected from the same sampled households. These types of data presents challenges to conducting genetic analysis because observations are correlated due to two types of clustering with one induced by the geographical cluster sampling, and the other induced by biological inheritance among multiple participants within the same sampled household. Populationbased case-control studies, such as KCS, sample individuals at differential rates, which requires sample weights incorporated in analyses in order to obtain unbiased inferences. We have developed efficient statistical methods to address the cluster sampling and sample weighting effects on genetic inferences. The proposed methods are evaluated analytically and via Monte Carlo simulation studies, and illustrated using data from the Hispanic Health and Nutrition Survey, NHANES and KCS.