Methods to Detect SNP Associations with Family Data: A Comparative Analysis

Laura R. Budde
Mentor: Dr. Michael A. Province

Abstract
Association studies may be used to provide valuable information about the interplay of both genes and environmental factors in the expression of complex traits when conducted within the framework of family sampling designs, but there are also complications that arise from the use of clustered family data. The performance of four methods constructed to deal with non-independent, non-identically distributed errors in family data were compared on simulated data with known generating parameters: two were extensions of the transmission disequilibrium test (TDT) (QTDT and FBAT) and two were extensions of ordinary least squares (OLS) (bootstrap sampling and the "sandwich" estimator). All methods performed well under the null, producing a uniform distribution of p-values when applied to sample data from a homogeneous population where no SNP effect was simulated. When SNP effects were present in the simulated data, both of the OLS-based methods had greater power to detect a subtle association due to a small SNP effect or low sample size. When these methods were applied to simulated data from an admixed population with no true SNP effect, the performance of the four methods were distinctly different. Both TDT-based tests, as expected protected against stratification with FBAT producing an approximately uniform distribution of p-values while QTDT is even more conservative. The bootstrap procedure offers no protection of false positives in this case, while the "sandwich" estimator produced false positive associations in 33% of simulations. The "sandwich" estimator offers the best of both worlds, allowing for greater power of detection of true associations in datasets with small sample size and/or small SNP effects, yet still offers some protection against population stratification.