Stony Brook University
The Graduate School

Doctoral Defense Announcement

Abstract

The impact of genotype misclassification errors on the power to detect a genetic association and gene-environment interaction with Cox proportional hazards modeling

By Lin Tung

Genetic model parameters determine the power and sample size required to detect a genetic association and gene-environment interaction with Cox proportional hazards modeling in cohort genetic studies. Detecting a genetic indirect association is largely dependent on the difference in the allele frequencies between the underlying functional variant and the marker locus. Similar to case-control genetic association studies, the increase in sample size to detect a genetic indirect association is approximately $1/r^2$, where the linkage disequilibrium parameter r^2 is the square of the correlation between alleles in coupling at the disease and marker locus. Detecting a gene-environment interaction for dominant and recessive modes of inheritance for a direct association study is feasible for common disease allele frequencies ($p_d > 0.10$) and moderate effect sizes.

The impact of each genotyping misclassification error upon the increase in sample size required to maintain constant Type I and II error rates can be calculated mathematically through a first-order linear Taylor series expansion. We find that, consistent with previous genotyping errors research in case-control genetic association studies, any misclassification of the more common homozygote is the most deleterious in terms of increase in sample size (or equivalently loss of power) to detect a genetic indirect association. The total required sample size to detect a genetic indirect association in the presence of genotyping errors may be partitioned into the sample size required for a direct association study, the increase in sample size due to an indirect association study, and the increase in sample size due to genotyping errors. For high r^2 (typically $r^2 > 0.95$) and approximately equal allele frequencies, the increase in sample size due to genotyping error rates as low as 2% is larger than the increase in sample size due to an indirect association study. In the detection of a gene-environment interaction, any misclassification of a subject without the at-risk genotype as having the at-risk genotype is the most deleterious. Such errors require an indefinitely large increase in required sample size as the SNP minor allele frequency approaches 0.

Date: July 6, 2007
Time: 11:00 AM
Program: Statistics
Dissertation Advisor: Dr. Stephen J. Finch
Place: Applied Mathematics and Statistics Seminar Room (Room 1-122, Math Tower)