Probability of Detecting Disease-Associated SNPs in Genome-Wide Association Studies

Ruth Pfeiffer, PhD

Biostatistics Branch, Div. of Cancer Epidemiology and Genetics, National Cancer Institute , Bethesda, MD, USA

Some investigators use *p*-values in case-control genome-wide association studies (GWASs) to rank and select promising single nucleotide polymorphisms (SNPs) for future study and are not concerned about the frequentist error control properties of the selection procedure. In this talk we first give an introduction to genome-wide association studies and then investigate the probability that a specific disease SNP will be selected for further study in a GWAS and the probability that a selected SNP will be a true disease SNP. We call the probability that the test statistic for a specific disease SNP will have a p-value among the lowest *T* p-values in the sample the detection probability (DP) and study DP for various models of genetic risk. For a genetic odds ratio per minor allele of 1.2 or less, even GWAS with 1000 cases and 1000 controls require T to be impractically large to achieve an acceptable DP. We extend these results to two-stage GWASs where all SNPs are analyzed in stage 1, and the SNPs with the smallest p values are then followed up in a second stage. Large sample sizes for stage 1 are required to achieve acceptable DP, and one-stage designs can be recommended in many settings. We also compare several procedures to combine GWAS data from several different studies both in terms of the power to detect a disease-associated SNP while controlling the genome-wide significance level, and in terms of DP.

This is joint work with Mitchell Gail