Liability threshold modeling increases power in case-control association studies

Genetic case-control association studies often include data on covariates, such as body mass index (BMI) or age, that may modify the underlying genetic risk of case or control samples. For example, in type 2 diabetes, odds ratios estimated from low-BMI cases are larger than those estimated from high-BMI cases. An unanswered question is how to optimally use this information to maximize statistical power. In this study we show via simulation that our approach to fitting liability threshold models and computing association statistics, which accounts for disease prevalence and non-random ascertainment, can use this information to increase power. Our method outperforms standard case-control association tests, case-control tests with covariates, tests of gene x covariate interaction, and tests that restrict to a subset of samples. We investigate empirical case-control studies of type 2 diabetes, prostate cancer, breast cancer, rheumatoid arthritis, age-related macular degeneration, and end-stage kidney disease over a total of 78,256 samples. In these data sets, liability threshold modeling outperforms logistic regression for 104 of the 140 known associated variants investigated (p-value < $10^{-9}$). The improvement varied across diseases with a 17% median increase in test statistics, corresponding to a greater than 25% increase in power. Application of liability threshold modeling to future case-control association studies of these diseases, or other diseases with analogous effects of covariates on genetic risk, will yield a substantial increase in power for disease gene discovery.