A COMPARISON OF FAMILY-BASED ASSOCIATION TESTS IN THE PRESENCE OF LINKAGE

Jérémie Nsengimana and Jennifer H Barrett

Section of Epidemiology and Biostatistics, Leeds Institute of Molecular Medicine, University of Leeds, UK

Family-based association tests (FBATs) have been proposed to circumvent the problem of population stratification in genetic association analyses. While some FBATs are specific to a type of trait or family structure, others are more versatile. For the analysis of late onset diseases, one appropriate design is to study discordant siblings. For sibships including more than one affected individual, accounting for residual familial correlation is required to ensure validity of the test for association in the presence of linkage. However, it has been suggested that adjusting familial correlation may be superfluous unless the genetic effect under investigation is extremely high.

We present a simulation study to test the validity, power, bias and robustness to model misspecification of five methods: FBAT, Empirical Variance FBAT (EV-FBAT), Sibship Disequilibrium Test (SDT), Conditional Logistic Regression (CLR) and Robust-CLR (R-CLR). FBAT is based on comparing the observed genotype distribution in affected offspring within a general family with that expected under the null hypothesis, conditional on parental genotypes (or sufficient statistics for these if missing). EV-FBAT and R-CLR are modifications of FBAT and CLR respectively designed to account for familial correlation. SDT is a model-free approach based on a sign test examining the proportion of sibships where the putative risk allele is more common in affected than unaffected siblings. Type 1 error was estimated in a range of scenarios based on a log-additive model for disease risk, and power is compared for all valid tests. The effect on type 1 error of distance between marker and disease locus is investigated. Power is estimated at the disease locus and at a nearby marker in incomplete linkage disequilibrium (LD). Since all tests except SDT assume a genetic model, their robustness to model misspecification was tested by analysing the data under dominant and recessive models. The program FBAT is used to compute FBAT, EV-FBAT and SDT statistics, whereas CLR and R-CLR analyses (Wald test) are conducted in STATA v.9.

SDT, EV-FBAT and R-CLR showed correct type 1 error in all designs, while FBAT and CLR were valid only when the odds ratio (OR) per copy of the risk allele was <2. The inflation of type 1 error in FBAT and CLR was stronger the closer the marker and disease locus. EV-FBAT, R-CLR, FBAT and CLR had similar power, while SDT was consistently less powerful. The power of EV-FBAT, R-CLR, FBAT and CLR drops when an incorrect dominant model is assumed, but it remains higher than or comparable to the power of SDT. Under the incorrect recessive model, the power of all these tests drops below that of SDT. While the strength of genetic effect is not estimated using FBAT, EV-FBAT and SDT, estimates of OR with CLR and R-CLR are unbiased when the disease locus is analysed, but they are biased when a nearby marker is analysed. This study shows that ignoring familial correlation invalidates FBATs even for modest effects, and that robust variance estimation produces a valid test at no cost to power, and that, for sibships, R-CLR is a simple but effective method of analysis.