

Joint analysis of binary and quantitative traits with data sharing and outcome-dependent sampling

Gang Zheng

National Heart, Lung and Blood Institute, Bethesda, Maryland, USA

Colin O. Wu

National Heart, Lung and Blood Institute, Bethesda, Maryland, USA

Minjung Kwak

National Heart, Lung and Blood Institute, Bethesda, Maryland, USA

Wenhua Jiang

Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

Jungnam Joo*

National Cancer Center, Korea jooj@ncc.re.kr

Joao A.C. Lima

Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

We study the analysis of a joint association between a genetic marker with both binary (case-control) and quantitative (continuous) traits where the quantitative trait values are only available for the cases due to data sharing and outcome-dependent sampling. Data sharing becomes common in genetic association studies, and the outcome-dependent sampling is the consequence of data sharing, under which a phenotype of interest is not measured for some subgroup. The trend test (or Pearson's test) and F-test are often, respectively, used to analyze the binary and quantitative traits. Because of the outcome-dependent sampling, the usual F-test can be applied using the subgroup with the observed quantitative traits. We propose a modified F-test by also incorporating the genotype frequencies of the subgroup whose traits are not observed. Further, a combination of this modified F-test and Pearson's test is proposed by Fisher's combination of their P-values as a joint analysis. Because of the correlation of the two analyses, we propose to use a Gamma (scaled chi-squared) distribution to fit the asymptotic null distribution for the joint analysis. The proposed modified F-test and the joint analysis can also be applied to test single trait association (either binary or quantitative trait). Through simulations, we identify the situations under which the proposed tests are more powerful than the existing ones. Application to a real dataset of rheumatoid arthritis is presented.

Key Words: case-control; continuous trait; missing data; polygenic pleiotropy; trend test