

Retrospective score tests versus prospective score tests for genetic association with case-control data

Yukun Liu

East China Normal University

E-mail: ykliu@sfs.ecnu.edu.cn

Abstract: Since the seminal work by Prentice and Pyke (1979), the prospective logistic likelihood has become the standard method of analysis for retrospectively collected case-control data, in particular for testing the association between a single genetic marker and a disease outcome in genetic case-control studies. When studying multiple genetic markers with relatively small effects, especially those with rare variants, various aggregated approaches based on the same prospective likelihood have been developed to integrate subtle association evidence among all considered markers. In this paper we show that using the score statistic derived from a prospective likelihood is not optimal in the analysis of retrospectively sampled genetic data. We develop the locally most powerful genetic aggregation test derived through the retrospective likelihood under a random effect model assumption. In contrast to the fact that the disease prevalence information cannot be used to improve the efficiency for the estimation of odds ratio parameters in logistic regression models, we show that it can be utilized to enhance the testing power in genetic association studies. Extensive simulations demonstrate the advantages of the proposed method over the existing ones. One real genome-wide association study is analyzed for illustration.