Estimation of Gene-Environment Interactions Using Mixture Regression Models with Non-Susceptibility

Chen-Hsin Chen[1,2], Yuh-Chyuan Tsay[1,3] and Wen-Harn Pan[2,4]

[1] Institute of Statistical Science, Academia Sinica, Taipei 11529, Taiwan
[2] Institute of Epidemiology, National Taiwan University, Taipei 10055, Taiwan
[3] Institute of Agronomy, National Taiwan University, Taipei, 10617 Taiwan
[4] Institute of Biomedical Sciences, Academia Sinica, Taipei, 11529 Taiwan

Population-based studies such as biobank projects have been devised to delineate gene-environment interactions on specific complex diseases/disorders in different ethnic groups. Statistical methods for analyzing this kind of data have not been well developed. In most of the cases, outcomes of the disease/disorder under study are treated as either a dichotomous status or a continuous measurement of age at onset. The former approaches do not consider the probability of later onset for observed non-cases, and the latter approaches ignore the disease non-susceptibility in study subjects, who did not inherit the related genes or were never exposed to deleterious environmental factors. These two kinds of approaches may render misleading interpretations.

In this presentation, we advocate a mixture regression model to tackle the disease non-susceptibility and the age-at-onset of susceptible subjects simultaneously. Main effects of candidate genes and potential environmental factors, as well as their interactions, may be fitted in both the logistic model for the probability of susceptibility and the event-time regression model for the distribution of age-at-onset of the susceptible subjects. To illustrate the mixture regression approach, we apply the estimation procedure to evaluate the susceptibility and the incidence of obesity in the Taiwan two-township study. This project recruited non-obese subjects and conducted periodical follow-ups. Adjusting for the two-stage study design, we use the mixture model with left truncated and general interval censored data to analyze single-nucleotide polymorphisms in candidate genes, dietary factors and their interactions. The proposed approach can be applied to general studies and is novel to the interpretation in genetic epidemiology.