

# Model selection for generalized estimating equations accommodating dropout missingness

Chung-Wei Shen\* (沈仲維) and Yi-Hau Chen (程毅豪)  
Institute of Statistical Science, Academia Sinica

## Abstract

The generalized estimating equation (GEE) has been a popular tool for marginal regression analysis with longitudinal data, and its extension, the weighted GEE approach, can further accommodate data that are missing at random (MAR). Model selection methodologies for GEE, however, have not been systematically developed to allow for missing data. We propose the missing longitudinal information criterion (MLIC) for selection of the mean model, and the MLIC for correlation (MLICC) for selection of the correlation structure in GEE when the outcome data are subject to dropout/monotone missingness and are MAR. Our simulation results reveal that the MLIC and MLICC are effective for variable selection in the mean model and selecting the correlation structure, respectively. We also demonstrate the remarkable drawbacks of naively treating incomplete data as if they were complete and applying the existing GEE model selection method. The utility of proposed method is further illustrated by two real applications involving missing longitudinal outcome data. This is a joint work with Dr. Yi-Hau Chen at Academia Sinica.

Keywords: longitudinal data, missing data, repeated measures