QIC program and model selection in GEE analyses

James Cui
Department of Epidemiology and Preventive Medicine
Monash University
Melbourne, Australia
james.cui@med.monash.edu.au

Abstract. The generalized estimating equation (GEE) approach is a widely used statistical method in the analysis of longitudinal data in clinical and epidemiological studies. It is an extension of the generalized linear model (GLM) method to correlated data such that valid standard errors of the parameter estimates can be drawn. Unlike the GLM method, which is based on the maximum likelihood theory for independent observations, the GEE method is based on the quasilikelihood theory and no assumption is made about the distribution of response observations. Therefore, Akaike's information criterion, a widely used method for model selection in GLM, is not applicable to GEE directly. However, Pan (Biometrics 2001; 57: 120–125) proposed a model-selection method for GEE and termed it quasilikelihood under the independence model criterion. This criterion can also be used to select the best-working correlation structure. From Pan's methods, I developed a general Stata program, `qic`, that accommodates all the distribution and link functions and correlation structures available in Stata version 9. In this paper, I introduce this program and demonstrate how to use it to select the best working correlation structure and the best subset of covariates through two examples in longitudinal studies.

Keywords: st0126, qic, Akaike's information criterion, GEE, likelihood, model, quasilikelihood under the independence model criterion