

## Consistent Model Selection and Data-driven Smooth Tests for Longitudinal Data in the Estimating Equations Approach

Lan WANG and Annie QU\*

University of Minnesota and Oregon State University

\*presenter

**Abstract:** Model selection for marginal regression analysis of longitudinal data is challenging due to the presence of correlation and the difficulty of specifying the full likelihood, particularly for correlated categorical data. This paper introduces a novel BIC-type model selection criterion based on the quadratic inference function (Qu, Lindsay and Li, 2000), which does not require the full likelihood or quasiliikelihood. With probability approaching one, the criterion selects the most parsimonious correct model. Although a working correlation matrix is assumed, there is no need to estimate the nuisance parameters in the working correlation matrix; moreover, the model selection procedure is robust against the misspecification of the working correlation matrix. The BIC-type criterion can also be used to construct a data-driven Neyman smooth test for checking the goodness-of-fit of a postulated model. This test is especially useful and often yields much higher power in situations where the classical directional test behaves poorly. The finite sample performance of the model selection and model checking procedures is demonstrated through Monte Carlo studies and analysis of a clinical trial data set.