THE GRADUATE COLLEGE OF THE
UNIVERSITY OF OKLAHOMA HEALTH SCIENCES CENTER

ANNOUNCES THE FINAL EXAMINATION OF

SUMMER GALE FRANK

FOR THE DEFENSE OF THE DOCTOR OF PHILOSOPHY DEGREE
GRADUATE COLLEGE
DEPARTMENT OF BIOSTATISTICS AND EPIDEMIOLOGY

Thursday, June 30th, 2016, 9:00 a.m.
Room 144, College of Health Building, OUHSC

The Use of Generalized Method of Moments with Binary Outcomes in the Presence of Time Varying Covariates

COMMITTEE IN CHARGE: Sara K. Vesely, Ph.D, Chair; Michael P. Anderson, Ph.D. Laura A. Beebe, Ph.D; , David M. Thompson, Ph.D., P.T., Eleni L. Tolma, Ph.D.

ABSTRACT:

Introduction - Statistical models for longitudinal data that include covariates that are time-varying can produce biased parameter estimates. The generalized estimating equations (GEE) method employing an independent working covariance structure is considered the "safe" choice to reduce bias, but can produce inefficient estimates of the standard error of a parameter estimate. Alternatively, the generalized method of moments (GMM) can be used to combine estimating equations (EE). Some researchers propose excluding invalid EE from GMM. EE are invalidated by their assumed covariate type or by directly testing standardized residuals (at time t) and weighted covariates (at time s) in a logistic regression model, for each combination of s and t (α=0.05).

Methods - This dissertation proposed identifying invalid EE through the correlation between the studentized residuals at time t and weighted covariate at time s where $s \neq t$ ($\rho \geq 0.2$ and $\alpha = 0.05$); EE on the diagonal were never excluded. Higher correlation thresholds ($\rho = 0.3$ and 0.4) were also explored. We compared the performance of the GEE independence method and two previously developed GMM methods to my proposed GMM method in three real-world analyses and various simulated data scenarios (cluster size= 500 and 1000; time points= 3, 4, and 5; covariate type = I, II, III). Simulations explored various strengths of weights representing the feedback from the previous outcome to the current covariate (feedback weight) and the previous covariate value on the current outcome (time-dependent covariate weight).

Results - The simulation studies demonstrated that higher thresholds for correlation exclude too few EE, producing parameter estimates with larger bias than other methods. When the smallest correlation value ($\rho = 0.2$) was employed, my method had smaller bias and standard error in many simulation scenarios, tending to perform best in extreme combinations (i.e. high feedback weight strength combined with low time-dependent covariate weight strength, or vice-versa). My method had the smallest standard deviation in two real-world analyses and shared the second smallest value in the third with another method.

Discussion - The comparative performance of my method depends upon the strengths of the feedback and time-dependent covariate weights and the correlation threshold used to identify invalid EE to exclude from the GMM calculation.