

Maiying Kong, PhD  
Associate Professor  
School of Public Health and Information Sciences  
University of Louisville

GEE type inference for clustered zero-inflated negative binomial regression with application to dental caries

Use of zero-inflated count data models is common in applications where the number of zero counts exceeds that predicted from a traditional count data model such as Poisson or negative binomial. When count data exhibiting inflated zero counts are correlated among subjects, a natural approach will be to fit a marginal model with the help of generalized estimating equations (GEE) that can incorporate subject-to-subject correlations. A GEE based zero-inflated negative binomial (ZINB) model is proposed to fit clustered counts with excessive zeros. However, the corresponding sandwich variance estimator appears to underestimate the true variance. The theoretical reasons for its failure are explained and a correction under additional modeling assumptions is offered. In addition, a clustered resampling (bootstrap) procedure is proposed to estimate the variance and it is shown that the bootstrap procedure captures the correct variance under no additional model assumptions. Utility of this marginal GEE based ZINB model over two other competing models has been assessed using a thorough simulation study. The resulting inference procedure is applied to study the association between the dental caries and fluoride exposures using a dataset extracted from the Iowa Fluoride Study. A number of risk factors of clinical significance are reliably identified using the proposed model.