Statistical Estimating Functions and Their Biomedical Applications;
The Story of Generalized Estimating Equations

Kung-Yee Liang*
Graduate Institute of Public Health, National Yang-Ming University, Taiwan

Scott Zeger*
Department of Biostatistics, Johns Hopkins University, U.S.A.

The method known as generalized estimating equations (GEE; Liang and Zeger, 1986 Biometrika; Zeger and Liang, 1986, Biometrics) was motivated by demands from public health colleagues for improved methods to analyze longitudinal data with binary and other discrete outcomes. The earlier seminal work on estimating functions by Godambe (1960, Annals of Mathematical Statistics) and on generalized linear models by Nelder and Wedderburn (1972, JRSSA) paved the way for the development of GEE. The goal was to obtain valid inferences about regression parameters while correctly specifying a model for only the first two moments for the multiple outcome variables for each subject. In this sense, GEE can be seen as a multivariate extension of quasi likelihood by Wedderburn (1974, Biometrika). The use of the so-called "sandwich covariance estimator" introduced by Huber (1967, Proceedings of the Fifth Berkeley Symposium on Statistics), White (1982, Econometrica) and Royall (1986, International Statistical Review) enables the investigators to further relax the assumption that the joint second moments must be correctly specified. Numerous simulation studies (e.g., Liang and Hanfelt, 1994, Biometrics) have confirmed that the quality of inference for the regression parameters is insensitive to the correct specification of the second moments in many cases. The work by Liang and Zeger (1995, Statistical Sciences) provides a theoretical justification for the empirical observations involving approximate orthogonality of the estimating functions.

Throughout the years, GEE has been found useful for analyzing correlated data common in public health and clinical research (Liang and Zeger, 1993, Annual Review of Public Health). In the second half of this talk, we demonstrate the use of GEE for analyzing multiple data sets, for example from the Baltimore Eye Survey to assess risk factors associated with visual acuity, and from pre- and post-design for a clinical trial on patients diagnosed with schizophrenia. We close by discussing important modern developments in likelihood, estimating function, and Bayesian inference that extend some of the ideas that motivated GEE and its applications.