Dissertation Paper Abstracts

Ilana Trumble

May 21, 2019

**Paper 1:** Where does the mixed model and GEE fail?

Title: Recommendations for choosing an analysis method that controls Type I error for longitudinal and multilevel studies with Gaussian outcomes

Current approaches for analyzing multilevel and longitudinal data, such as mixed models and generalized estimating equations (GEE), have notable limitations. Specifically, these models can fail to converge, can yield biased parameter estimates, and can produce inflated Type I error rates that can subsequently impact statistical inference. In this paper, we identify specific situations and criteria that induce limitations in convergence, estimation, and inference for both mixed models and GEE. Sample size and missing data are investigated as factors relating to model failure.

**Paper 2:** Quasilexicographic method

Title: Unbiased, consistent estimation and size inference for longitudinal and multilevel models with missing data

Longitudinal and multilevel studies are frequently subject to missing data. Current approaches, such as mixed models and generalized estimating equations (GEE), are known to have convergence problems and inflated type I error rates when applied to typical longitudinal or multilevel data and unstructured covariance matrices. In this paper, we provide a novel method for analyzing such data without these limitations. We name this the *quasilexicographic* method. The quasilexicographic method can be applied to data with missingness and mistimed predictors without inflation of type I error rates. We present derivations for the quasilexicographic model parameter estimates as well as their distributions for continuous outcomes and non-missing predictors. Simulation studies compare the performance of the quasilexicographic model, mixed model, and GEE.

**Paper 3:** Tutorial paper applied to ECHO consortium – thoughts on cohort design

Title: Tutorial on analyzing longitudinal and multilevel studies

There are various ways to analyze longitudinal and multilevel studies, including mixed models, generalized estimating equations (GEE), and our newly introduced novel quasilexicographic model. These methods each have limitations. In this paper, we provide a tutorial for investigators selecting an analysis method for longitudinal or multilevel studies. Specifically, we aim to provide guidance for investigators from the ECHO (Environmental Influences on Child Health Outcomes) consortium. ECHO is a national, NIH funded, collaborative project to harmonize 84 separate cohort studies of mothers and children in the United States to increase the ability to detect environmental influences on childhood health outcomes. ECHO studies are often longitudinal or multilevel in nature. We provide analysis method selection criteria based on sample size, missingness, whether the outcome is normal, and if predictors are repeated or mistimed. We also provide guidance for cohort design based on the appropriateness of these analysis methods.