

DEPARTMENT OF STATISTICS AND BIOSTATISTICS

**Jason Roy**Center for Casual Inference
Rutgers University*Subject-specific Functional Prediction from
Electronic Health Records using an Enriched
Dirichlet Process Mixture Model***September 19, 2018****3:20 – 4:20pm**

Light refreshments will be served

**110 Frelinghuysen Road
Hill Center, Room 552**

Abstract: In many modern applications, there is interest in predicting subject-specific functions of a variable over time. For example, we might want to know patient-specific trends in a biomarker over time. Modeling is needed if there is measurement error in the variable, or if gaps between data collection times is too wide. We propose a novel semiparametric model for the joint distribution of a continuous longitudinal outcome and the baseline covariates using an enriched Dirichlet process (EDP) prior. This joint model decomposes into subject-specific linear mixed models for the outcome given the covariates and simple marginals for the covariates. The nonparametric EDP prior is placed on the regression and spline coefficients, the error variance, and the parameters governing the predictor space. We predict the outcome at unobserved time points for subjects with data at other time points as well as for completely new subjects with covariates only. We find improved prediction over mixed models with Dirichlet process (DP) priors when there are a large number of covariates. Our method is demonstrated with electronic health records consisting of initiators of second generation antipsychotic medications, which are known to increase the risk of diabetes. We use our model to predict laboratory values indicative of diabetes for each individual and assess incidence of suspected diabetes from the predicted dataset. Our model also serves as a functional clustering algorithm in which subjects are clustered into groups with similar longitudinal trajectories of the outcome over time.

Bio: Jason Roy is Professor of Biostatistics and Chair of the Department of Biostatistics and Epidemiology in Rutgers School of Public Health. He is also director of Rutgers University Biostatistics and Epidemiology Services Center (RUBIES) and co-director of the Center for Causal Inference. Prior to joining Rutgers, Dr. Roy was an Associate Professor in the Department of Biostatistics, Epidemiology, & Informatics and Associate Program Chair of the Biostatistics Graduate Program at the University of Pennsylvania. He received his PhD in 2000 from the University of Michigan. His methodological research interests include causal inference, Bayesian nonparametrics, missing data, and electronic health records. Dr. Roy has collaborated on research in pharmacoepidemiology, health policy, chronic kidney disease, and chronic viral hepatitis infection. He is currently Associate Editor of Biometrics, Journal of the American Statistical Association, and Pharmacoepidemiology and Drug Safety.

