

Conference on Causal Inference in Longitudinal Studies

September 21-23, 2017

Columbia University

Thursday, September 21, 2017: tutorial

14:15 15:00 Miguel Hernan (Harvard University)

15:00 15:45 Miguel Hernan

15:45 16:00 Coffee break

16:00 16:45 Miguel Hernan

16:45 17:30 Miguel Hernan

Friday, September 22, 2017: research talks

9:00 9:15 David Madigan (Columbia University) - Opening Remarks

Research Talks

9:15 10:00 Jamie Robins (Harvard University)

10:00 10:20 Coffee break

10:20 11:05 Fabrizia Mealli (University of Florence)

11:05 11:50 Stijn Vansteelandt (Ghent University)

11:50 14:00 Lunch break

14:00 14:45 Daniel Almirall (University of Michigan)

14:45 15:30 Mark vander Laan (UC Berkeley)

15:30 15:50 Coffee break

15:50 16:35 Miguel Hernan (Harvard University)

16:35 17:20 Alexander Luedtke (Fred Hutchinson Cancer Research Center)

18:00 Dinner for speakers

Saturday, September 23, 2017: research talks

9:00 9:45 Thomas Richardson (University of Washington)

9:45 10:30 Eric Laber (NC State University)

10:30 10:50 Coffee break

10:50 11:35 Mats Stensrud (University of Oslo)

11:35 12:20 Wenbin Lu (NC State University)

## **Titles and Abstracts**

### **Alexander R. Luedtke**

Title: "Sequential double robustness in right-censored longitudinal models"

Abstract:

Consider estimating the G-formula for the counterfactual mean outcome under a given treatment regime in a longitudinal study. Bang and Robins provided an estimator for this quantity that relies on a sequential regression formulation of this parameter. This approach is doubly robust in that it is consistent if either the outcome regressions or the treatment mechanisms are consistently estimated. We define a stronger notion of double robustness, termed sequential double robustness, for estimators of the longitudinal G-formula. The definition emerges naturally from a more general definition of sequential double robustness for the outcome regression estimators. An outcome regression estimator is sequentially doubly robust (SDR) if, at each subsequent time point, either the outcome regression or the treatment mechanism is consistently estimated. This form of robustness is exactly what one would anticipate is attainable by studying the remainder term of a first-order expansion of the G-formula parameter. We show that a particular implementation of an existing procedure is SDR. We also introduce a novel SDR estimator, whose development involves a novel translation of ideas used in targeted minimum loss-based estimation to the infinite-dimensional setting.

(Joint work with Oleg Sofrygin, Mark van der Laan, and Marco Carone)

### **Mark van der Laan**

Title: Targeted Machine Learning for Causal Inference

Abstract:

We review targeted minimum loss estimation (TMLE), which provides a general template for the construction of asymptotically efficient plug-in estimators of a target estimand for infinite dimensional models. TMLE involves maximizing a parametric likelihood along a so-called least favorable parametric model through an initial estimator (e.g., ensemble super-learner) of the relevant functional of the data distribution. The asymptotic normality and efficiency of the TMLE relies on the asymptotic negligibility of a second-order term. This typically requires the initial estimator to converge at a rate faster than  $n^{-1/4}$ . We propose a new estimator, the Highly Adaptive LASSO (HAL), of the data distribution and its functionals that converges at a sufficient rate regardless of the dimensionality of the data/model, under almost no additional regularity. This allows us to propose a general TMLE that is asymptotically efficient in great generality. We demonstrate the practical performance of HAL and its corresponding TMLE for the average causal effect for dimensions up till 10. We also discuss inference taking into account the higher order contributions of the HAL-TMLE.

## **Fabrizia Mealli**

Title: “Beyond noncompliance: assessing causal effect heterogeneity in longitudinal studies using principal stratification.”

Abstract:

The talk will focus on settings with intermediate variables and longitudinal responses, where principal stratification can be used to define causal estimands and formally express structural and distributional assumptions. Different modes of inference for principal causal effects will be reviewed, including (Bayesian) model-based approaches. Identification results for graphical models with latent variables will be adapted and extended for the identification of principal strata effects, as principal strata within observed groups take the form of a latent variable. Conditional independence structures among covariates, multivariate, and longitudinal responses, that may vary across observed and latent groups, are shown to be a powerful tool that can flexibly adapt to many empirical settings. An empirical example is provided, using data from a social experiment, where focus is on assessing causal effects of a training program on employment status over time in the presence of noncompliance.

## **Daniel Almirall**

Title: “Assessing Time-Varying Causal Effect Moderation with Intensive Longitudinal Intervention Data: With Application to Mobile Health.”

Abstract:

In mobile health (mHealth) for behavior change and maintenance, interventions are frequent and momentary. Typically, a great deal of information on patient states (e.g., stress), environmental factors, and behavioral responses is generated over time. Such intensive longitudinal intervention data is often collected by self-report or passively with the aid of sensors. One way in which intensive longitudinal intervention data may aid the design of a mobile intervention is the examination of effect moderation; that is, inference about which factors strengthen or weaken the response to just-in-time interventions. In this setting, treatments, outcomes, and candidate moderators are time-varying. This talk (1) introduces a definition for moderated causal effects in terms of potential outcomes which is suitable for intensive longitudinal data (the causal effects are connected to a generalization of the structural nested mean model); (2) it develops an extremely easy-to-use weighted and centered regression approach for investigating these moderated effects (the approach can be used with standard software); and (3) it compares the new method with standard longitudinal modeling methods. The approach is illustrated using BASICS-Mobile, a smartphone-based intervention designed to curb heavy drinking and smoking among college students.

## **James M Robins**

Title: Identification, Inference, and Open Issues in Longitudinal Causality

Abstract:

I will survey progress in longitudinal causality over the past 30 years taking a historical perspective with a focus on important open issues.

## **Mats Julius Stensrud**

Title: Hazard ratios with a causal interpretation: Adjusting for selection bias using frailty models.

It is well known that Cox proportional hazards models may suffer from selection bias. Assessing the magnitude of this bias, however, is not straightforward. I will present an approach to adjust for selection bias, even when unmeasured factors influence survival. The approach is based on frailty theory, and the unobserved risk factors are assumed to follow a parametric distribution in the population. Importantly, we can estimate the parameters of this distribution using published, real-life data on familial risks. By applying this approach, I will assess the magnitude of survival bias in real-life scenarios. In particular, I will discuss situations in which the survival bias is potentially severe.

## **Thomas Richardson**

Title: Structural Nested Mean Models for Binary Treatments and Outcomes

Abstract: Structural Nested Mean Models (SNMMs), introduced by Robins (1994), model contrasts of potential outcomes for a final response conditional on prior history. These models have a simple causal interpretation, lead to direct tests of the global causal null, and can model effect modification by time-dependent covariates. However, their application to binary outcomes has been impeded by the difficulty of specifying compatible nuisance models. In this talk we will introduce a simple parametrization of a multiplicative SNMM using a generalized odds product.

(This is joint work with Linbo Wang and James M. Robins, Harvard School of Public Health.)

## **Eric Laber**

Title: List-based treatment regimes

Abstract: A treatment regime formalizes personalized medicine as a function from individual patient characteristics to a recommended treatment. A high-quality treatment regime can improve patient outcomes while reducing cost, resource consumption, and treatment burden. Thus, there is tremendous interest in estimating treatment regimes from observational and randomized studies. However, the development of treatment regimes for application in clinical practice requires the long-term, joint effort of statisticians and clinical scientists. In this collaborative process, the statistician must integrate clinical science into the statistical models underlying a treatment regime and the clinician must scrutinize the estimated treatment regime for scientific validity. To facilitate meaningful information exchange, it is important that estimated treatment regimes be interpretable in a subject-matter context. We propose a simple, yet flexible class of treatment regimes whose members are representable as a short list of if-then statements. Regimes in this class are immediately interpretable and are therefore an appealing choice for broad application in practice. We derive a robust estimator of the optimal regime within this class and demonstrate its finite sample performance using simulation experiments. The proposed method is illustrated with data from two clinical trials.