**UW SISCER 2020**

**Module 1: Modern Statistical Learning for Observational Data**

**July 14–17, 2020**

While clinical trials provide the highest level of evidence to compare clinical treatments or public health interventions, they are often not feasible due to ethical, logistic or economic constraints. Observational studies provide an opportunity to learn about the effect of interventions for which little or no trial data are available. These studies constitute a potentially rich and relatively cheap source of information. However, in such studies, treatment or intervention allocation may be strongly confounded by other important patient characteristics and much care is needed to disentangle observed relationships and infer causal effects.

In this course, we will provide an overview of modern statistical techniques for analyzing observational data. We will focus primarily on recent advances in the field of targeted learning, which facilitates the use of state-of-the-art machine learning tools to flexibly adjust for confounding while yielding valid statistical inference. In contrast, conventional techniques for confounding adjustment rely on restrictive statistical models and may, therefore, lead to severely biased inference. Use of the Super Learner framework, an implementation of model stacking, will be discussed as a particularly appealing means of performing flexible, pre-specified adjustment for confounding.

We will discuss methods for comparative effectiveness studies for single time-point interventions. We will also introduce the multi time-point extension of these methods and discuss strategies for dealing with missing data. Methods will be illustrated using data from recent observational studies and extracted from electronic medical records. Analyses will be illustrated in R but knowledge of R is not required for this module. In addition to lectures, the course will include in-class, hands-on activities to allow students to familiarize themselves with the methods and tools.

The four-day course is geared towards health science researchers with at least basic experience in data analysis and statistics. A basic understanding of the following concepts would be helpful: confounding, probability (e.g., what is meant by the distribution of random variable, its mean and its variance), statistical inference (confidence intervals, hypothesis tests), and regression (linear and logistic). Advanced knowledge of these topics is useful, but not necessary. Equivalent UW SPH course pre-requisites are BIOS 511/512 (or BIOS 514/515).

**Overview and Schedule**

\* All times are Pacific Daylight Time (PDT)

**Accessing recordings**

Recordings will be put [in this shared Dropbox folder](https://www.dropbox.com/sh/74bx4buewjvorap/AABtpjUocoo4Cm7EhvAx6jFMa?dl=0) and available until September 18, 2020. The password for the Dropbox folder is siscer2020. You are free to download the recordings, but we ask that you do not post the recordings publicly. The recordings corresponding to each session are listed below.

**Tuesday, July 14**

8:30 – 8:45 Introduction to the course Chapter0.pdf 01\_Tuesday1.mp4

8:45 – 10:00 Introduction to causal inference Chapter1.pdf 01\_Tuesday1.mp4

10:00 – 10:15 Break

10:15 – 11:00 Basic identification and estimation of Chapter2.pdf 02\_Tuesday2.mp4

average treatment effects

11:00 – 11:15 Break

11:15 – 12:00 Basic identification and estimation of Chapter2.pdf 03\_Tuesday3.mp4

average treatment effects (cont.)

**Wednesday, July 15**

8:30 – 9:30 Basic identification and estimation of Chapter2.pdf 04\_Wednesday1.mp4

average treatment effects (cont.)

9:30 – 10:00 Super learning Chapter3.pdf 05\_Wednesday2.mp4

10:00 – 10:15 Break

10:15 – 11:00 Super learning Chapter3.pdf 05\_Wednesday2.mp4

11:00 – 11:15 Break

11:15 – 12:00 Super learning Chapter3.pdf 05\_Wednesday2.mp4

**Thursday, July 16**

8:30 – 9:30 Super learning lab Lab1.pdf

9:30 – 10:00 Efficient, doubly robust estimation of Chapter4.pdf

an average treatment effect

10:00 – 10:15 Break

10:15 – 11:00 Efficient, doubly robust estimation of Chapter4.pdf

an average treatment effect

11:00 – 11:15 Break

11:15 – 12:00 Efficient, doubly robust estimation of Chapter4.pdf

an average treatment effect

**Friday, July 17**

8:30 – 9:30 TMLE lab Lab2.pdf

9:30 – 10:00 Identification and inference on the average Chapter5.pdf

 treatment effect of a time-varying

 intervention

10:00 – 10:15 Break

10:15 – 11:00 Identification and inference on the average Chapter5.pdf

 treatment effect of a time-varying

 intervention

11:00 – 11:15 Break

11:15 – 12:00 Identification and inference on the average Chapter5.pdf

 treatment effect of a time-varying

 intervention

**Zoom information**

Call in information is included below. We will use a waiting room to ensure only registered participants are able to access the room. Accordingly, **please enter the name you used to register for the module as your Zoom name when entering the room**.

**Topic: SISCER 2020 Module 1: Modern Statistical Learning for Observational Data
Time: This is a recurring meeting Meet anytime

Join Zoom Meeting**[**https://zoom.us/j/99498759276**](https://zoom.us/j/99498759276)

Meeting ID: 994 9875 9276
One tap mobile
+14702509358,,99498759276# US (Atlanta)
+14703812552,,99498759276# US (Atlanta)

Dial by your location
        +1 470 250 9358 US (Atlanta)
        +1 470 381 2552 US (Atlanta)
        +1 301 715 8592 US (Germantown)
        +1 312 626 6799 US (Chicago)
        +1 929 205 6099 US (New York)
        +1 253 215 8782 US (Tacoma)
        +1 346 248 7799 US (Houston)
        +1 669 900 6833 US (San Jose)
Meeting ID: 994 9875 9276
Find your local number: <https://zoom.us/u/apG7saGdT>

Join by SIP
99498759276@zoomcrc.com

Join by H.323
162.255.37.11 (US West)
162.255.36.11 (US East)
115.114.131.7 (India Mumbai)
115.114.115.7 (India Hyderabad)
213.19.144.110 (EMEA)
103.122.166.55 (Australia)
64.211.144.160 (Brazil)
69.174.57.160 (Canada)
207.226.132.110 (Japan)
Meeting ID: 994 9875 9276

**Other notes**

All course materials are posted on the SISCER website (participant login required).

A remote server will be set up and available for use to run lab code on Thursday and Friday. We will provide details on how to access the server Thursday morning. All that is required to access is a web browser.

Additional material (Chapter6.pdf and Lab3.pdf) are made available and may be covered time-permitting (but it’s doubtful ☺).

If questions come up outside of regular hours, please feel free to ask on the slack channel or via email. We will respond via text or during class the next day.