Short Course #1

Sunday, October 23

8 a.m. - Noon

Targeted Learning for data adaptive causal inference in observational and randomized studies

INSTRUCTORS: Mark van der Laan, PhD, and Susan Gruber, PhD, MPH, MS

FORMAT: One 4-hour session, including practicum

TARGET AUDIENCE: Epidemiologists, data scientists, informaticians, data analysts, and statisticians

About the course

Data-driven decision making succeeds when data are transformed into reliable, interpretable information. Targeted learning offers a principled statistical approach utilizing state of the art machine learning to answer questions about health and safety, while still providing statistical inference in terms of confidence intervals and formal testing. The targeted learning framework combines two state of the art methodologies: super learning (SL) for data adaptive machine learning and targeted minimum loss-based estimation (TMLE) for efficient semi-parametric estimation. These provide the building blocks for accurate estimation of the scientific quantity of interest (e.g., treatment effect) and of uncertainty, in order to draw sound statistical conclusions. Real-world applications will be used to illustrate how targeted learning can be used to answer the kinds of complex questions that arise in practice. One TMLE application involves the reanalysis of randomized trials evaluating a treatment for Sepsis, demonstrating a large gain in power. Another TMLE application evaluates different individualized strategies for controlling glucose levels in diabetic patients based on a complex observational longitudinal study. We will also demonstrate SL and TMLE applications in drug safety analysis and using electronic medical record data.

The course will motivate the use of targeted learning for addressing issues commonly found in the analysis of cross-sectional and longitudinal data, and provide a comprehensive overview of the framework. A roadmap will be presented that provides a systematic guide to estimating a relevant statistical quantity that addresses a clear scientific question of interest.

Specific learning objectives

- Understand at a conceptual level the targeted learning approach and its rationale
- Understand the advantages of targeted learning compared with standard methods through case studies and illustrations
- Understand how to implement targeted learning methods using existing software and interpret the results using available open source R packages: *tmle, SuperLearner, ltmle*.

GroupHealth RESEARCH INSTITUTE

DEPARTMENT OF BIOSTATISTICS
SCHOOL OF PUBLIC HEALTH
UNIVERSITY of WASHINGTON

REAGAN - UDALL FOUNDATION FOR THE Food and Drug Administration

About the instructors

Mark J. van der Laan, PhD, is a Hsu/Peace Professor of Biostatistics at the University of California, Berkeley School of Public Health. He is the recipient of the 2005 COPSS Presidents' and Snedecor Awards, as well as the 2004 Spiegelman Award, and is a Founding Editor for the International Journal of Biostatistics and the Journal of Causal Inference. He has authored various books, and his most recent book is Targeted Learning: Causal Inference for Observational and Experimental Data, van der Laan, Rose (2011), Springer: New York.

Susan Gruber, PhD, MPH, MS, is Director of the Biostatistics Center in the Department of Population Medicine at Harvard Medical School and Harvard Pilgrim Health Care Research Institute. She is the author of the *tmle* software package, and has published extensively on targeted learning.