Targeted Learning Methods for Challenges in Liver Diseases

Mark van der Laan, Ph.D.
Professor, Division of Biostatistics
Faculty Director, Center for Targeted Learning in Big Data
School of Public Health, University of California, Berkeley

How can Targeted Learning contribute? Sample Use Cases

- Data-adaptive learning of optimal surrogate endpoints and non-invasive diagnostic biomarkers.
- Data-adaptive learning of optimal individualized treatment regimens
- Efficient and robust analysis of RCT.
- Adaptive trial designs to meet user-specified needs (e.g. increase efficiency, minimize adverse events, learn optimal treatment rule) while preserving validity.
- Address complex data structure of a Natural History Cohort; causal inference from observational data.

Why Targeted Learning?

Traditional approaches in epidemiology and clinical research:

- Fit parametric models for the outcome.
- Report coefficients in the model as effect measures.

- ➢ Reliant on a mis-specified parametric model.
- ➢ Effect interpretation problematic
- ➢ Human subjectivity; prone to overestimate precision.
- ➢ Non-data-adaptive modeling misses opportunities for discovery.

What is Targeted Learning?

An emerging subfield of statistics that integrates:

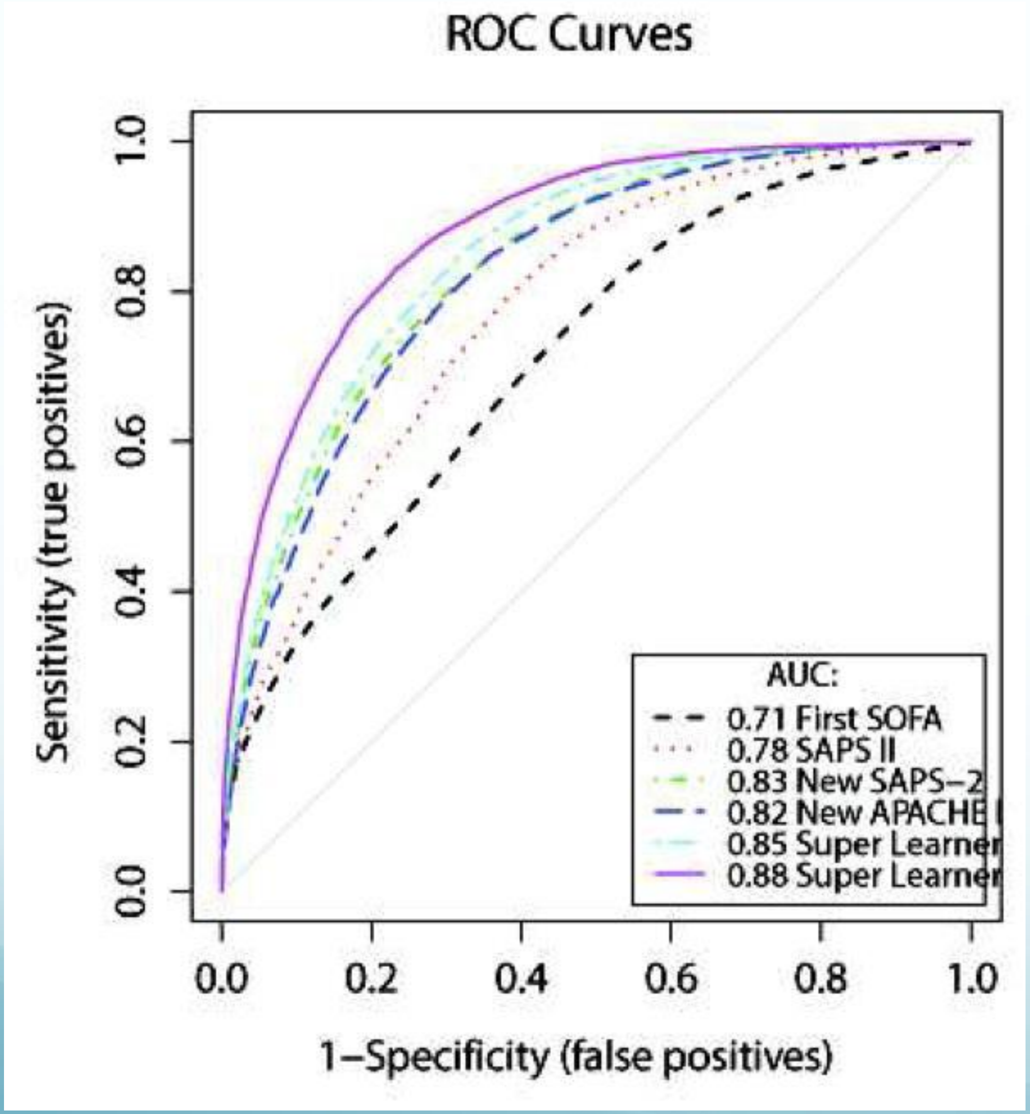
- **Causal Inference** ➔ asking the right questions.
- **Data-Adaptive Estimation**➔ expert knowledge + machine learning prowess. Full potential for prediction and discovery.
- **Robust estimation and Formal Hypothesis Testing** ➔ accurate and actionable information.

Two part methodology: Super Learning + Targeted Minimum Loss Estimation

Part 1: Super Learning

- Combines a library of learning algorithms (including machine learning) to build a better predictor.
- Uses Cross-validation to find best combination.
- ➔ Does at least as well as the best member in library.
- ➔ Incorporates expert knowledge in the library; let data pick the best combination.
- ➔ Allows data-adaptive modeling of data-generating process. No unrealistic or rigid modeling assumptions.

Super Learning in Action



Mortality prediction in ICU using the Super ICU Learning Algorithm: a population-based study.

Pirracchio, et.al.
Lancet Resp. Med. 2015 Jan; 3(1): 42-52

- Improved performance over all existing methods for predicting mortality of patients in ICU.

Part 2: Targeted Minimum Loss Estimation

- Updates initial estimators from Super Learner.
- Tailor bias-variance trade-off towards the target parameter of interest.
- Incorporates outcome model and exposure model.
- Robust estimates against model-misspecification ➔ **increased accuracy**.
- Efficient estimator ➔ **increased precision, learns more with less**.
- Framework for Formal Hypothesis Testing ➔ **actionable information for decision making**

Sample Uses Cases of Targeted Learning

- Data-adaptive learning of optimal surrogate endpoints and non-invasive diagnostic biomarkers.
- Data-adaptive learning of optimal individualized treatment regimens
- Efficient and robust analysis of RCT.
- Innovative adaptive trial designs to meet user-specified needs (e.g. increase efficiency, minimize adverse events) while preserving validity.
- Address complex data structure of a Natural History Cohort; causal inference from observational data.

Data-adaptive learning of surrogate markers

- From RCT or observational data
- Learn most predictive non-invasive diagnostic biomarkers of underlying disease status.
- Learn surrogate endpoint that best capture the intervention effect on endpoint of interest.
- Incorporates subject-matter expertise with machine learning prowess.
- Avoid unrealistic assumptions on unknown outcome-surrogate interactions ➔ maximize potentials for discovery
- Optimality-measure independent of model-specification ➔ robust interpretation

Data-adaptive learning of optimal individualized treatment strategies

- From RCT or from observational data
- Learn the best treatment/management strategy to prescribe regimen based on individual profile. → precision medicine
- Formal hypothesis testing for effect of optimal strategy.
- Parallel application to learning optimal subgroup to receive treatment
- Incorporates subject-matter expertise with machine learning prowess.
- Avoids unrealistic parametric assumptions on the exposure-outcome relation.

Design and Analysis of RCT

- Increased efficiency and addresses heterogeneity by adjusting for covariates.
- Robust effect estimates even under mis-specified outcome models.
- Adaptive designs to meet user-specified goals (eg. Maximize efficiency, minimize adverse events), while preserving validity.
- Adaptive designs tailored to learn optimal treatment rule, or optimal surrogate markers.

Analysis of the Natural History Cohort

- Complex longitudinal data structure.
- Have rigorous methods to deal with ‘messy’ data: informative missing data, time-varying confounding, etc.
- Have formal framework for drawing causal inference from observational data.
- Increased accuracy and precision through doubly robust and efficient estimators.
- Potentials for regulatory considerations of previously ‘untestable’ clinical questions.

Thank you

- More on the methodology: Targeted Learning Book, van der Laan and Rose, Springer 2011.

- The Center for Targeted Learning in Big Data:

Faculty Director: Mark van der Laan, Ph.D. laan@berkeley.edu

Executive Director: Wenjing Zheng, Ph.D. wenjing.zheng@berkeley.edu