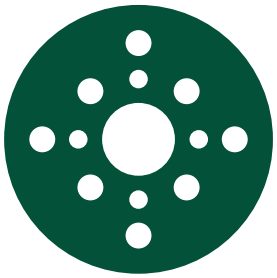
Targeted Machine Learning

Causal inference with TMLE and Super Learner



2026 Training Modules

Targeted ML Solutions

practical information
for general and technical audiences

Targeted Machine Learning Training Modules

Targeted Machine Learning (TML) is a statistical framework for efficient learning from data. Training modules introduce TML, causal inference, and machine learning with a focus on practical understanding of core concepts and how-to guides for learning from randomized controlled trials (RCT) and real-world data (RWD) to create robust evidence.

Trainings are provided live over Zoom. Sessions are recorded and made available to participants. Customized on-site training sessions can be arranged.

Module	Audience	Length
A1. Targeted Machine Learning in Generating and Evaluating Real-World Evidence: The central role of the Targeted Learning Roadmap	General[a]	1/2 day
A2. Causal inference: A comparison of propensity score-based methods and targeted maximum likelihood estimation (TMLE)	General	1/2 day
A3. Super Learning: Data-adaptive machine learning for prediction, modeling, and its role in estimating causal effects	General	1/2 day
A4. Developing a Targeted Learning-based statistical analysis plan	General	1/2 day
B1. Focus on analysis: Point treatment effects using TMLE+SL	Technical[b]	2 hours or 1/2 day
B2. Focus on analysis: Survival analysis using TMLE+SL	Technical	2 hours or 1/2 day
B3. Focus on analysis: Longitudinal treatments, intercurrent events using TMLE+SL	Technical	2 hours or 1/2 day
B4. Focus on analysis: Super Learning for risk prediction	Technical	2 hours or 1/2 day
B5. Focus on analysis: Super Learning for outcome phenotyping	Technical	2 hours or 1/2 day
C1. Hands-on laboratory session special topics using TML Pro software for targeted machine learning *(can be repeated with custom topics such as rare disease, external controls, survival analysis, longitudinal data analysis)*	Technical	2 hours

[a]clinicians, epidemiologists, statisticians, data scientists; [b]statisticians, data analysts

©Targeted ML Solutions, 2026

Module Descriptions

Series A modules provide the necessary background for understanding the challenges of evaluating treatment effects from health care data and how they can be successfully met with the tools of Targeted Machine Learning.

Intended audience: clinicians, epidemiologists, statisticians, data scientists

A1. Targeted Machine Learning in Randomized and Real-World Studies: The central role of the Targeted Learning Estimation Roadmap

This module discusses improving the quality of evidence generated from clinical studies using randomized controlled trial (RCT) or real-world data (RWD), and assessing its level of support for decision-making. The TL estimation roadmap provides a step-by-step guide for causal effect estimation that produces a rich trove of information for assessing whether study findings provide robust scientific evidence for addressing key study questions. Roadmap steps embody elements of a statistical estimand outlined in the ICH E9 Guidelines and the target trial emulation framework.

A2. Causal inference: A comparison of targeted maximum likelihood estimation (TMLE) and propensity score (PS)-based methods

This module introduces fundamental concepts in causal inference including Neyman-Rubin causal model and Pearl's non-parametric structural equation models (NPSEM), the distinction between a causal parameter and a statistical parameter, and statistical approaches for estimating causal effects from data. PS matching, inverse probability of treatment weighting (IPTW), and TMLE + SL will be introduced.

A3. Super Learning: Data-adaptive machine learning for prediction modeling, and its role in causal effect estimation

Traditional parametric models enforce predictor-outcome relationships that are often unrealistic. They are also unsuitable in high dimensional settings. Many machine learning methods are now available, but which one is best depends on underlying characteristics of the data. Super learning is a stacked ensemble approach that can explore them all, relying on cross-validation to select the optimal approach. The course provides a high-level overview of popular machine learning algorithms (tree-based approaches, lasso regression, highly adaptive lasso, neural networks), describes the super learner, and provides practical advice on how to specify a super learner and interpret results.

A4. Developing a Targeted Learning-based statistical analysis plan (SAP)

The SAP for a study that incorporates RWD must explicitly state how issues discussed in the ICH E9(R1) guidelines will be anticipated, ameliorated, and addressed in the study design and statistical analysis. Through a case study we'll discuss the importance of specifying a statistical model that respects the process that gave rise to the data, and defining the causal parameter and the corresponding statistical estimand in observed data. TMLE and Super Learning are recommended methodologies for data-adaptive estimation. Pre-specified sensitivity analyses include a model-free demonstration of how robust the conclusion to is to violations of unavoidable causal assumptions that are not testable from data.

©Targeted ML Solutions, 2026

Series B modules explore technical and practical aspects of using TMLE + SL in point treatment and longitudinal treatment settings, and analyses involving time-to-event data.

Intended audience: statisticians, data scientists

B1. Focus on analysis: Point treatment effects using TMLE+SL

Understand how to define and evaluate causal effects of point treatments. Observations in the data contain information on baseline covariates (W), treatment (A), a binary or continuous outcome (Y), with potential missingness in Y (indicated by Δ). We'll discuss potential estimands, strategies for handling missing covariate values, and techniques for robust inference.

B2. Focus on analysis: Survival analysis using TMLE+SL

Understand how to define and evaluate causal effects of treatments with time-to-event data. Observations in the data contain information on baseline covariates (W), treatment (A), and time to either a right censoring or outcome event ($\tilde{T}$). Strategies for handling missing covariate values and techniques for robust inference will be discussed.

B3. Focus on analysis: Longitudinal treatments, intercurrent events using TMLE+SL

Understand how to define and evaluate causal effects of time-dependent treatments, or point treatments accounting for intercurrent events. Observations in the data contain information on baseline and time-varying covariates (L_t), time-dependent treatment (A_t), a binary or continuous outcome measured up until time t (Y), with time-dependent right censoring and loss to follow-up (C_t). Strategies for handling missing covariate values and techniques for robust inference will be discussed.

B4. Focus on analysis: Super Learning for risk prediction

This module explores how to use super learning to construct a risk prediction algorithm that can be integrated into health-care settings. The emphasis is on practical issues for defining the super learner based on characteristics of the data and the task at hand and evaluating relevant cross-validated performance metrics, including global goodness of fit and calibration within subgroups.

B5. Focus on analysis: Super Learning for outcome phenotyping

This module explores how to use super learning for outcome phenotyping. The emphasis is on practical issues for defining the super learner based on characteristics of the data and the task at hand and evaluating relevant cross-validated performance metrics (area under the receiver-operating curve, area under the precision-recall curve, positive predictive value, negative predictive value, sensitivity, specificity, F1 measure).

©Targeted ML Solutions, 2026

Series C modules are hands-on laboratory sessions working with the TML Pro software for targeted learning. It can be repeated multiple times for in-depth understanding of how to use the R-based software for TMLE analyses and super learning.

Intended audience: statisticians, data scientists, programmers

C1. Hands-on laboratory session special topics using TMLE + SL in R

These 2-hour hands-on lab sessions focus on TMLE to analyze time-to-event data, longitudinal data, or point treatment data consisting of baseline covariates, treatment, outcome missingness indicator, and outcome. Additional sessions focus on Super learning (SL) for predictive modeling and risk score prediction, or as an integral part of unbiased causal effect estimation. We will utilize a flowchart for specifying a SL for defining the SL library, loss function, and V-fold cross validation scheme. Sample code and data will be provided.

Instructor Bios

Mark van der Laan, PhD
Co-Founder and CEO

Professor of Biostatistics at UC Berkeley and inventor of Targeted Learning and Targeted Maximum Likelihood Estimation (TMLE). A globally recognized leader in causal inference and machine learning, Mark's work underpins much of modern causal AI in healthcare and regulatory science.

Susan Gruber, PhD, MPH, MS
Co-Founder and President

Former Harvard Medical School faculty and Reagan-Udall Foundation leader, Susan is a pioneer in applied TMLE and real-world evidence. Her work bridges advanced methodology with practical decision-making for regulators, industry, and public health.

©Targeted ML Solutions, 2026

2026 Pricing and Registration

Module	Length
A1. Targeted Learning in Generating and Evaluating Real-World Evidence: The central role of the Targeted Learning Roadmap	1/2 day
A2. Causal inference: A comparison of propensity score-based methods and targeted maximum likelihood estimation (TMLE)	1/2 day
A3. Super Learning: Data-adaptive machine learning for prediction, modeling, and its role in estimating causal effects	1/2 day
A4. Developing a Targeted Learning-based statistical analysis plan	1/2 day
B1. Focus on analysis: Point treatment effects using TMLE+SL	2 hours or 1/2 day
B2. Focus on analysis: Survival analysis using TMLE+SL	2 hours or 1/2 day
B3. Focus on analysis: Longitudinal treatments, intercurrent events using TMLE+SL	2 hours or 1/2 day
B4. Focus on analysis: Super Learning for risk prediction	2 hours or 1/2 day
B5. Focus on analysis: Super Learning for outcome phenotyping	2 hours or 1/2 day
C1. Hands-on laboratory session special topics using TMLE + SL in R *(can be repeated with custom topics in targeted learning, causal inference, super learning, TMLE)*	2 hours

Public on-line sessions (open to all registrants)

$450 per participant for 2-hour modules,
$850 per participant for 1/2-day modules

(20% discount for groups of 10 or more registering simultaneously)

Private on-line sessions (restricted to members of your organization)

$500 per participant for 2-hour modules (minimum of 5 participants)
$1000 per participant for 1/2-day modules (minimum of 5 participants)

(20% discount for groups of 10 or more)

Government and academic discounts available

Registration

Individuals registering for public sessions will be sent a registration link.

On-demand half-day sessions are held between the hours of 10 a.m. and 5 p.m. Eastern Time, subject to instructor availability. Flexible scheduling is available.

Contact: *info@targetedML.com*

©Targeted ML Solutions, 2026