

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

Targeted Learning

Bridging Machine Learning with Causal and Statistical Inference

Mark van der Laan¹ and Nima Hejazi²

¹Division of Biostatistics, University of California, Berkeley

²Department of Biostatistics, Harvard T.H. Chan School of Public Health

15th November 2024

Harvard Catalyst Workshop

Acknowledgements: Susan Gruber, Alan Hubbard, Ivana Malenica, Rachael Phillips, Lars van der Laan



A traditional toolbox for statistics

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

Goal	Type of Data			
	Measurement (from Gaussian Population)	Rank, Score, or Measurement (from Non-Gaussian Population)	Binomial (Two Possible Outcomes)	Survival Time
Describe one group	Mean, SD	Median, interquartile range	Proportion	Kaplan Meier survival curve
Compare one group to a hypothetical value	One-sample t test	Wilcoxon test	Chi-square or Binomial test **	
Compare two unpaired groups	Unpaired t test	Mann-Whitney test	Fisher's test (chi-square for large samples)	Log-rank test or Mantel-Haenszel*
Compare two paired groups	Paired t test	Wilcoxon test	McNemar's test	Conditional proportional hazards regression*
Compare three or more unmatched groups	One-way ANOVA	Kruskal-Wallis test	Chi-square test	Cox proportional hazard regression**
Compare three or more matched groups	Repeated-measures ANOVA	Friedman test	Cochrane Q**	Conditional proportional hazards regression**
Quantify association between two variables	Pearson correlation	Spearman correlation	Contingency coefficients**	
Predict value from another measured variable	Simple linear regression or Nonlinear regression	Nonparametric regression**	Simple logistic regression*	Cox proportional hazard regression*
Predict value from several measured or binomial variables	Multiple linear regression* or Multiple nonlinear regression**		Multiple logistic regression*	Cox proportional hazard regression*



A traditional toolbox for statistics

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

Table 1.1: A few choices of statistical target parameter as implied by some common regression models.

Outcome Type	Model (or Method)	Statistical Parameter
Continuous $Y \in \mathbb{R}$	linear regression	Difference of means
Continuous and positive $Y \in \mathbb{R}^+$	log-linear regression	Ratio of means
Binary $Y \in \{0, 1\}$	logistic regression	Odds ratio
Count $Y \in \mathbb{Z}^+$	Poisson regression	Ratio of means
Survival time $T = \min(T_F, T_C)$	Cox regression	Hazard ratio

The choices given in Table 1.1 are not exhaustive, and several variations on each of the modeling approaches exist, but, in standard practice, even their off-the-shelf varieties are common. Despite their popularity, how does the flow of logic that led us here hold up to scrutiny? In effect, using Table 1.1 as a guide, we have committed to choosing the target of inference based on the type of the outcome variable Y and an inflexible choice of regression model, the latter of which sharply restricts the set of candidate statistical parameters. This *model-based* or *model-forward* approach limits the answers we are able to seek based not on the scientific question of interest, but instead on the type of data and the limitations of common regression models.

- Traditional: Data + Model \rightarrow Question (parameter)
- Agnostic: Question (parameter) + Data \rightarrow Model
- Which of these is better aligned with the scientific process?

Performance of traditional tools

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

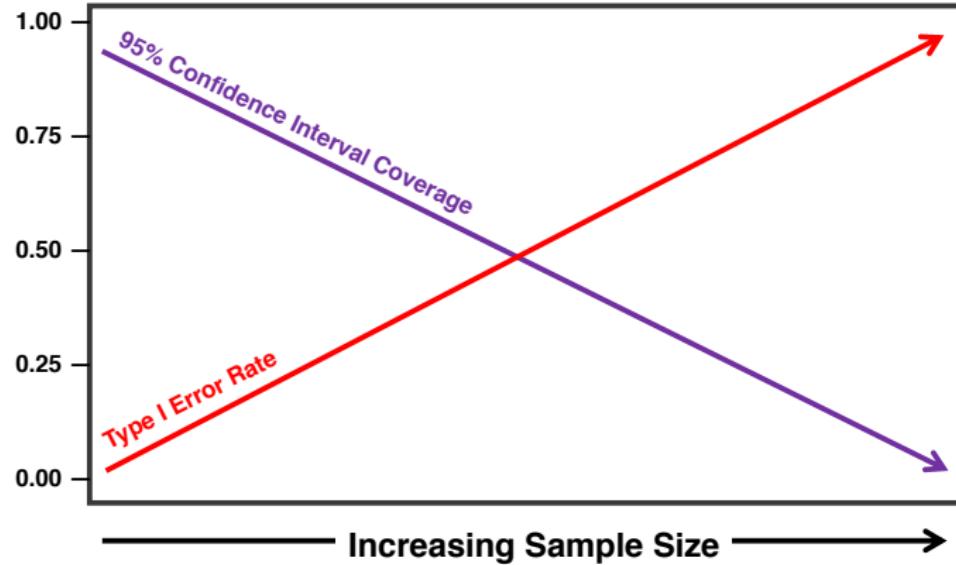
Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion



Performance of traditional tools

Targeted
Learning

Mark van der
Laan and
Nima Hejazi

The Art of
Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic
statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

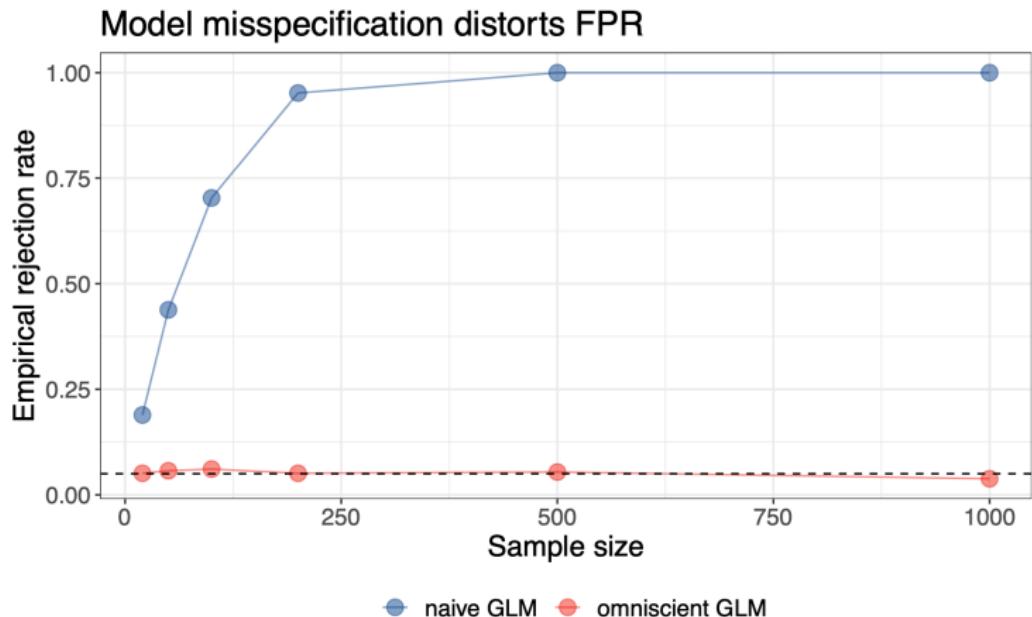
Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion



Truth: No effect of exposure on outcome.

The “art” of statistical inference?

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

Why Most Published Research Findings Are False

John P. A. Ioannidis

False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant

Joseph P. Simmons¹, Leif D. Nelson², and Uri Simonsohn¹

¹The Wharton School, University of Pennsylvania, and ²Haas School of Business, University of California, Berkeley

The Statistical Crisis in Science

Data-dependent analysis—a “garden of forking paths”—explains why many statistically significant comparisons don’t hold up.

Andrew Gelman and Eric Loken



TL answers statistical questions rooted in causality

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

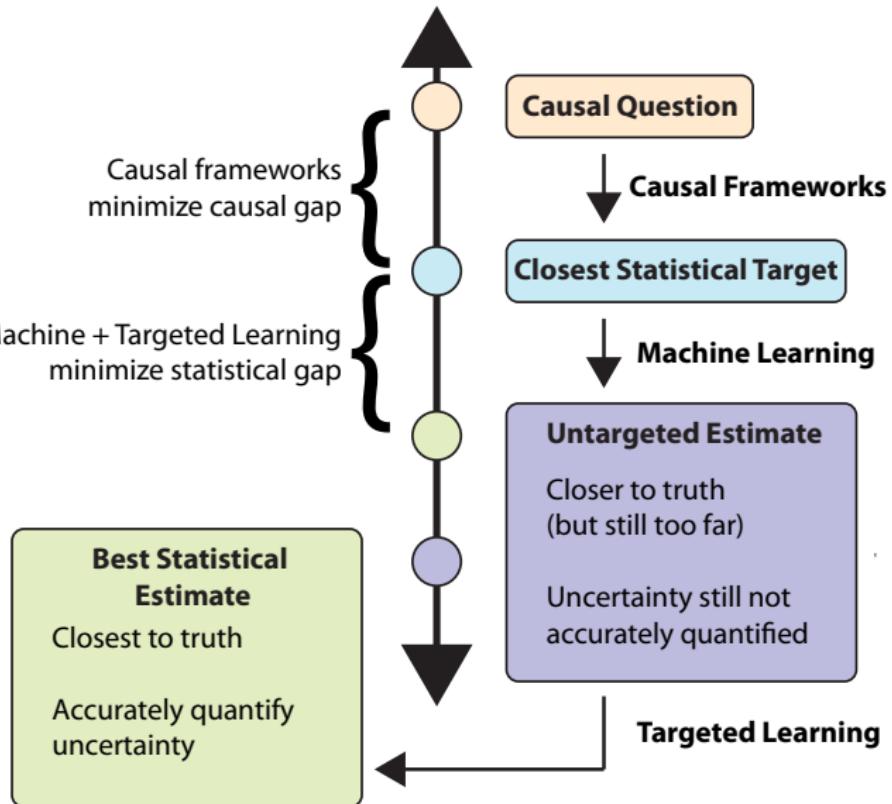
Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion



Public health and medicine use real-world data (RWD) for insight and evidence

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

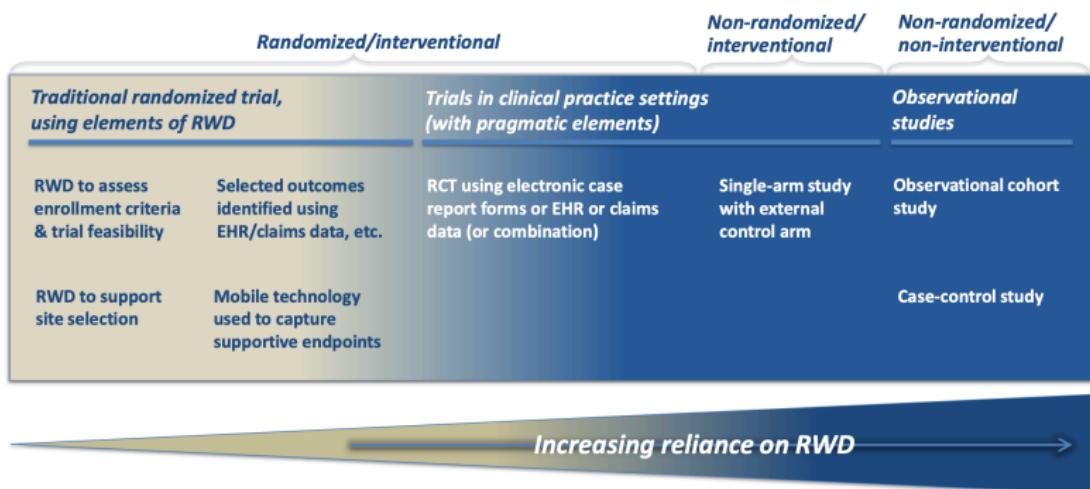
Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion



TL addresses statistical challenges with RWD

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

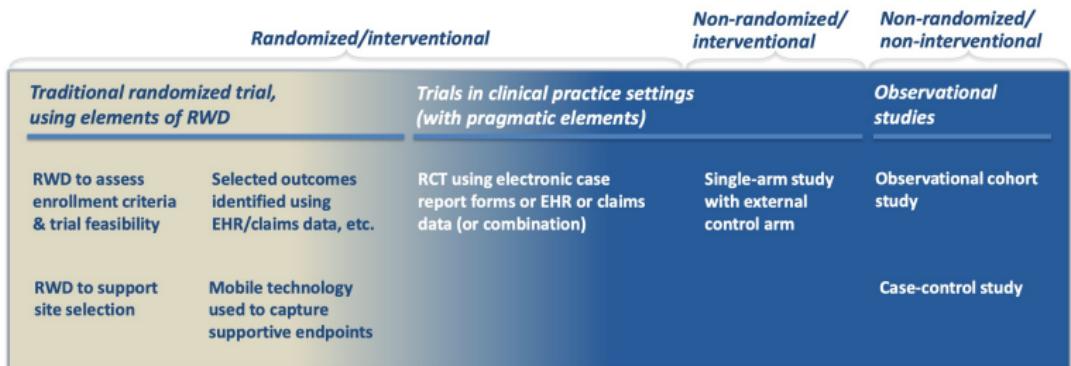
Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion



RWD Challenges

- Selection bias
- Intercurrent events
- Informative missingness
- Treatment by indication
- High dimensional covariates
- Outcome measurement error
- Statistical model misspecification
- Differences between external controls and single trial arm RCT

Targeted Learning path supports regulatory decision making

TL for real-world evidence (RWE) evaluation

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

	<i>Randomized/interventional</i>	<i>Non-randomized/interventional</i>	<i>Non-randomized/non-interventional</i>
<i>Traditional randomized trial, using elements of RWD</i>	<i>Trials in clinical practice settings (with pragmatic elements)</i>		<i>Observational studies</i>
RWD to assess enrollment criteria & trial feasibility	Selected outcomes identified using EHR/claims data, etc.	RCT using electronic case report forms or EHR or claims data (or combination)	Single-arm study with external control arm
RWD to support site selection	Mobile technology used to capture supportive endpoints		Observational cohort study
			Case-control study

RWD Challenges

- ❑ Selection bias
- ❑ Intercurrent events
- ❑ Informative missingness
- ❑ Treatment by indication
- ❑ High dimensional covariates
- ❑ Outcome measurement error
- ❑ Statistical model misspecification
- ❑ Differences between external controls and single trial arm RCT

Targeted Learning path supports regulatory decision making

Targeted Learning

- ✓ Roadmap for causal and statistical inference
- ✓ Realistic statistical model
- ✓ Statistical estimand approximates answer to causal question
- ✓ Flexible estimation and dimension reduction with Super Learner
- ✓ Model-free sensitivity analysis
- ✓ Generate RWE with confidence

TL is a subfield of statistics

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

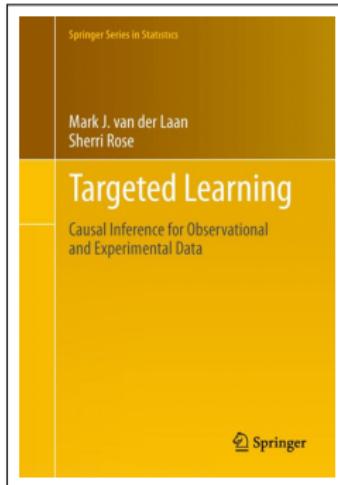
Collaborative TMLE

HAL and A-TMLE

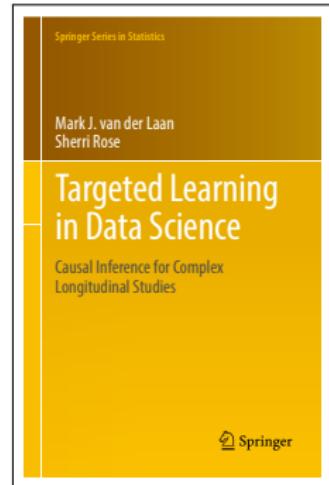
Longitudinal TMLE

Variable Importance

Conclusion



van der Laan & Rose, *Targeted Learning: Causal Inference for Observational and Experimental Data*. New York: Springer, 2011.



van der Laan & Rose, *Targeted Learning in Data Science: Causal Inference for Complex Longitudinal Studies*. New York: Springer, 2018.

Targeted Learning in R with the `tlverse`

Some applications of TL in the real world

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

HIV Testing and Treatment with the Use of a Community Health Approach in Rural Africa

D.V. Havlir, L.B. Balzer, E.D. Charlebois, T.D. Clark, D. Kwarisiima, J. Aiyelo, J. Kabami, N. Sang, T. Liegler, G. Chamie, C.S. Camlin, V. Jain, K. Kadefeo, M. Atukunda, T. Ruel, S.B. Shade, E. Ssemmondo, D.M. Byonanebye, F. Mwangwa, A. Owaramagise, W. Oillo, D. Black, K. Smyman, R. Burger, M. Getahun, J. Achando, B. Awuonda, H. Nakato, J. Kironde, S. Okiror, H. Thirumurthy, C. Koss, L. Brown, C. Marquez, J. Schwab, G. Lavoy, A. Plenty, E. Mugoma Wafula, P. Omanya, Y.-H. Chen, J.F. Rooney, M. Bacon, M. van der Laan, C.R. Cohen, E. Bukusi, M.R. Kamya, and M. Petersen

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Genetic Diversity and Protective Efficacy of the RTS,S/AS01 Malaria Vaccine

D.E. Neafsey, M. Juraska, T. Bedford, D. Benkeser, C. Valim, A. Griggs, M. Lievens, S. Abdulla, S. Adjei, T. Agbenyega, S.T. Agnandji, P. Aide, S. Anderson, D. Ansong, J.J. Aponte, K.P. Asante, P. Bejon, A.J. Birkett, M. Bruls, K.M. Connolly, U. D'Alessandro, C. Dobaño, S. Gesase, B. Greenwood, J. Grimsby, H. Tinto, M.J. Hamel, I. Hoffman, P. Kamthunzi, S. Kariuki, P.G. Kremsner, A. Leach, B. Lell, N.J. Lennon, J. Lusingu, K. Marsh, F. Martinson, J.T. Moiel, E.L. Moss, P. Njuguna, C.F. Ockenhouse, B. Ragama Oguta, W. Otieno, L. Otieno, K. Otieno, S. Owusu-Agyei, D.J. Park, K. Pellié, D. Robbins, C. Russ, E.M. Ryan, J. Sacarlal, B. Sogoloff, H. Sorgho, M. Tanner, T. Theander, I. Valea, S.K. Volkman, Q. Yu, D. Lapierre, B.W. Birren, P.B. Gilbert, and D.F. Wirth

© Health Research and Educational Trust

DOI: 10.1111/1475-6773.12848

METHODS ARTICLE

Robust Machine Learning Variable Importance Analyses of Medical Conditions for Health Care Spending

Sherri Rose 

THE LANCET Respiratory Medicine



Volume 3, Issue 1, January 2015, Pages 42-52

Articles

Mortality prediction in intensive care units with the Super ICU Learner Algorithm (SICULA): a population-based study



Better clinical decisions from observational data

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

Statistics in Medicine

Research Article

Received 24 May 2013,

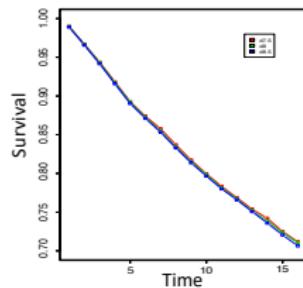
Accepted 5 January 2014

Published online 17 February 2014 in Wiley Online Library

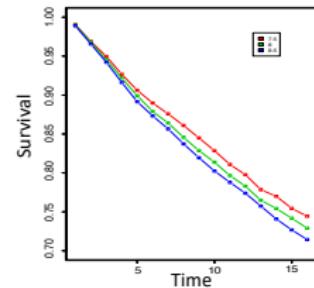
(wileyonlinelibrary.com) DOI: 10.1002/sim.6099

Targeted learning in real-world comparative effectiveness research with time-varying interventions

Romain Neugebauer,^{a,*†} Julie A. Schmittield^a and Mark J. van der Laan^b



Standard methods: No benefit to more aggressive intensification strategy



Targeted Learning: More aggressive intensification protocols result in better outcomes

TL for optimal treatment variable importance in meta-analysis of childhood development studies

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

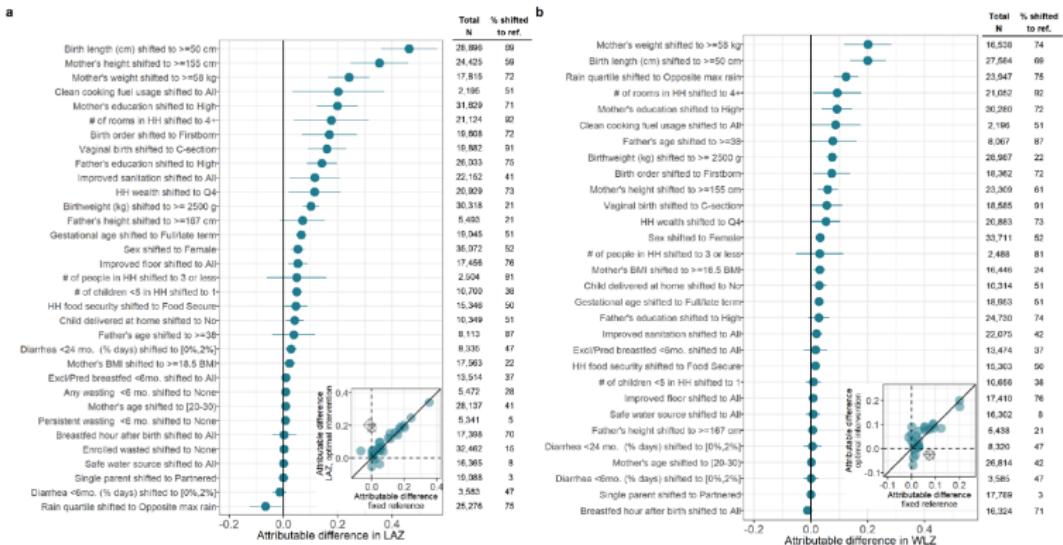


Figure 2 | Rank-ordered attributable differences between child, parental, and household characteristics and population attributable differences in anthropometry Z-scores.



Evaluating counterfactual HIV infection risk after shifting vaccine-induced immune response markers

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

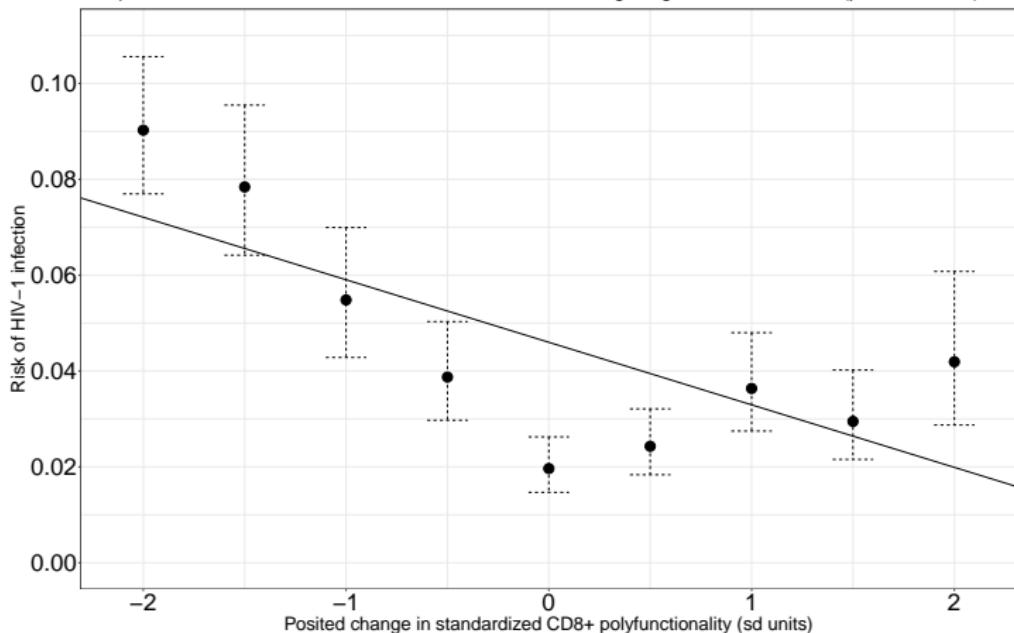
HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

TML estimates of mean counterfactual HIV-1 infection risk under shifted CD8+ polyfunctionality with pointwise confidence intervals and summarization via working marginal structural model ($\hat{\beta}_{TML} = -0.013$)



Hejazi et al. "Efficient nonparametric inference on the effects of stochastic interventions under two-phase sampling, with applications to vaccine efficacy trials" *Biometrics* (2021). 10.1111/biom.13375

Evaluating COVID-19 vaccine efficacy after shifting immune correlates of protection

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

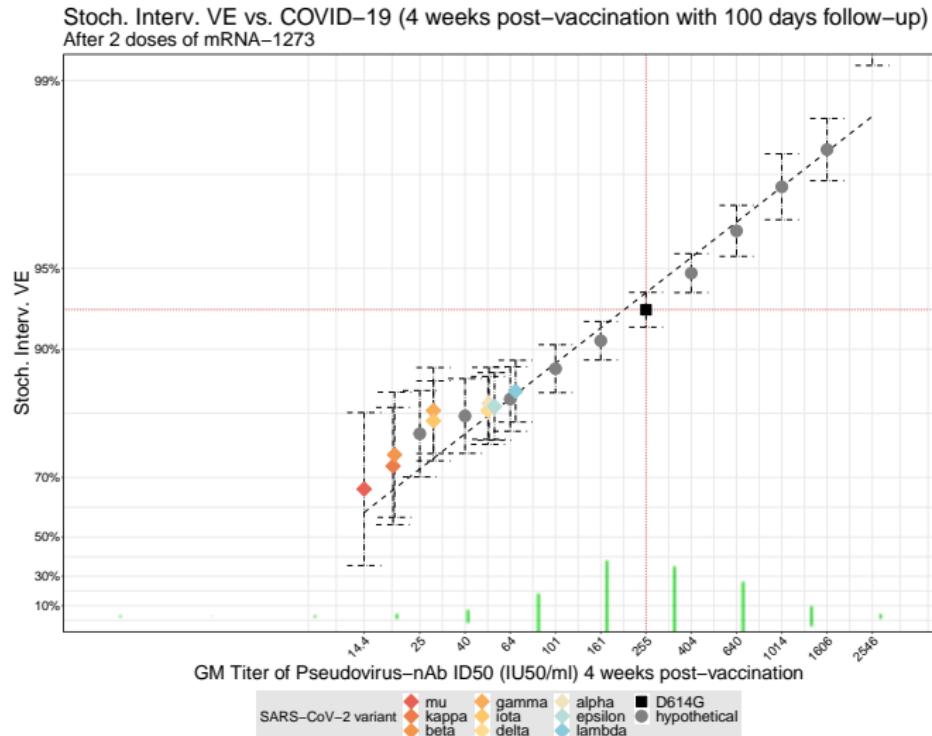
Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion



Hejazi et al. "Stochastic interventional approach to assessing immune correlates of protection: Application to the COVE mRNA-1273 vaccine trial" *IJID* (2023).

Evaluating (in)direct effects of adaptive dosing strategies for OUD

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

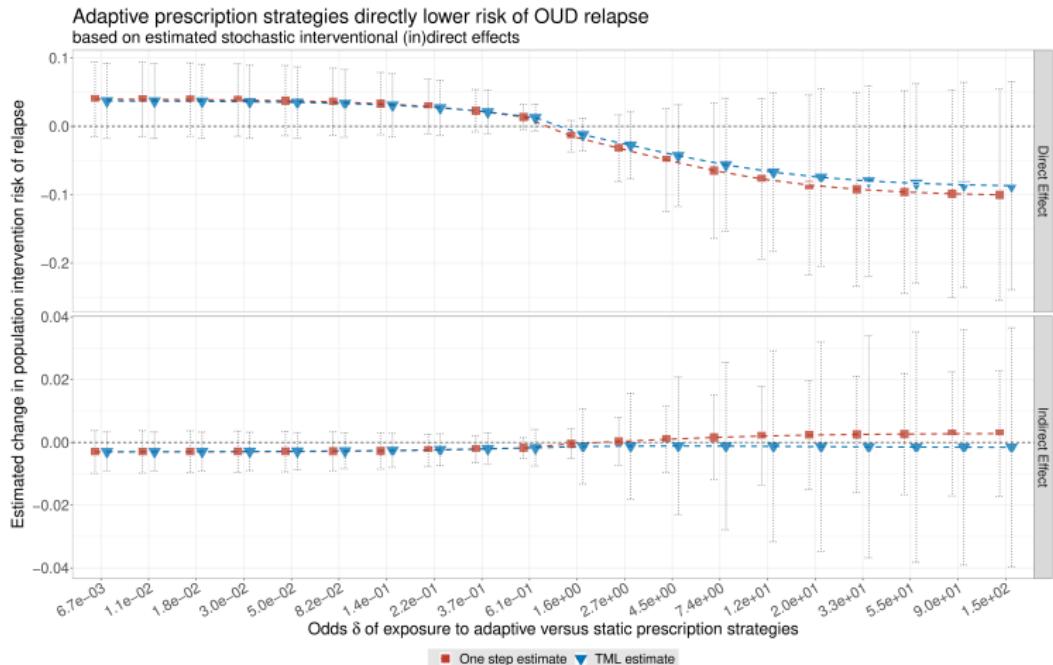
Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion



Hejazi et al., “Nonparametric causal mediation analysis for stochastic interventional (in)direct effects.” *Biostatistics* (2023).



Estimating impacts on AKI of delay-in-intubation policies

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

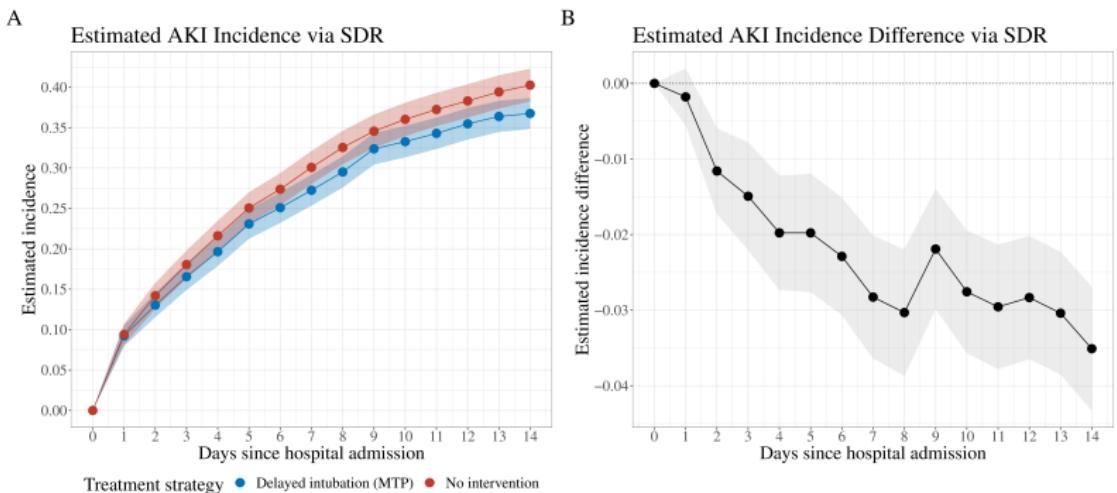
Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion



Diaz, Hoffman, and Hejazi, "Causal survival analysis with longitudinal modified treatment policies." *LiDA* (2024).

Introduction to the tlverse

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

- Describe study
- Specify a realistic statistical model
- Define estimand
- Causal estimand
- Statistical estimand
- Construct estimator
- Obtain inference
- Sensitivity analysis

Advanced TL

- Collaborative TMLE
- HAL and A-TMLE
- Longitudinal TMLE
- Variable Importance

Conclusion

<https://tlverse.org/catalyst2024-workshop/>

A roadmap for learning from data

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

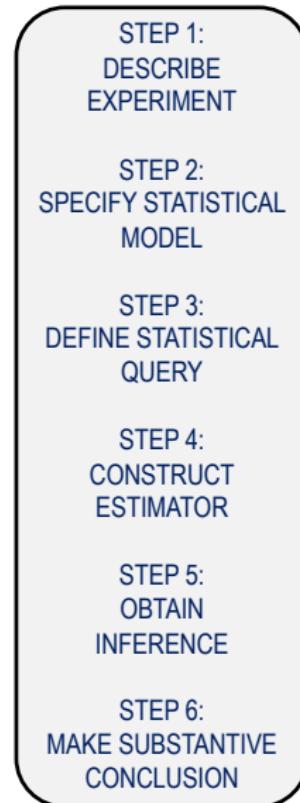
Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion



Step 1: What is the data-generating experiment?

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

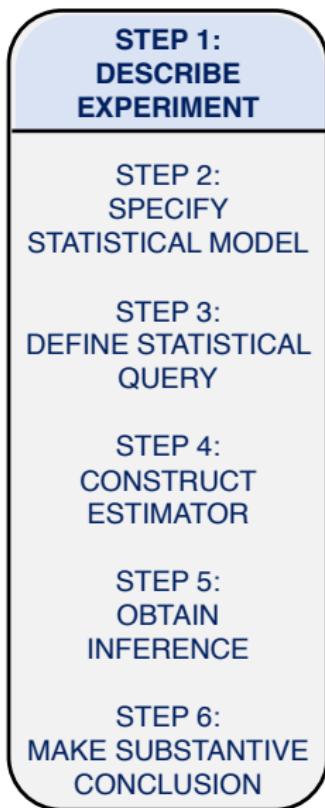
Collaborative TMLE

HAL and A-TMLE

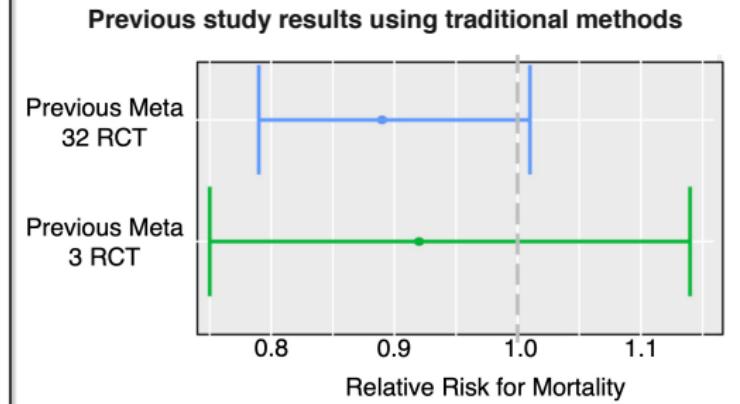
Longitudinal TMLE

Variable Importance

Conclusion



Three multi-national RCTs assessing impact of corticosteroids on mortality among septic shock patients



Target population of interest

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

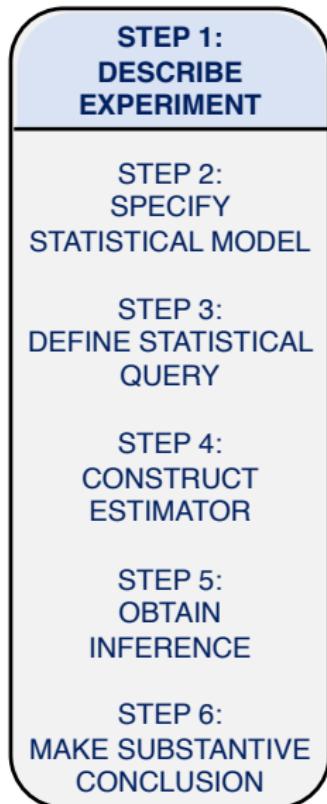
Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion



Three multi-national RCTs assessing impact of corticosteroids on mortality among septic shock patients

Pooled sample of $n = 1,300$ adults in septic shock

Observed data structure

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

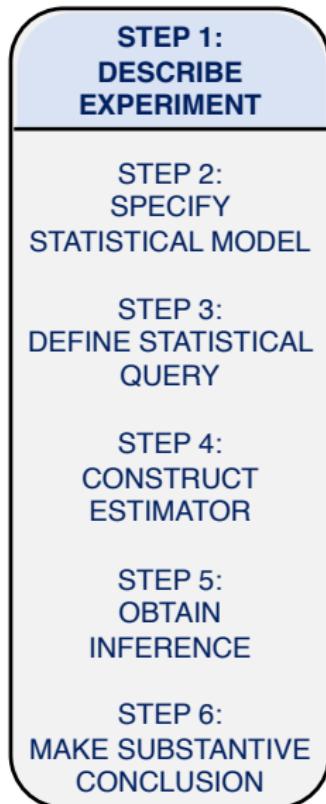
Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion



Three multi-national RCTs assessing impact of corticosteroids on mortality among septic shock patients

Pooled sample of $n = 1,300$ adults in septic shock

BMI	age	sex	steroid treatment	1-month mortality
21	65	F	1	1
22.3	28	F	0	0
19.4	49	F	1	0
24	77	M	1	0

Directed Acyclic Graph (DAG)

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

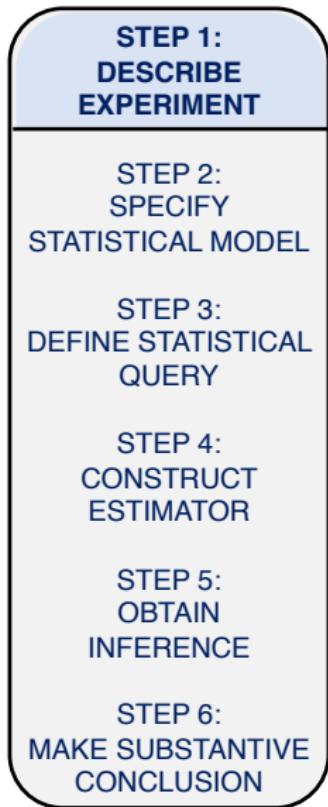
Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

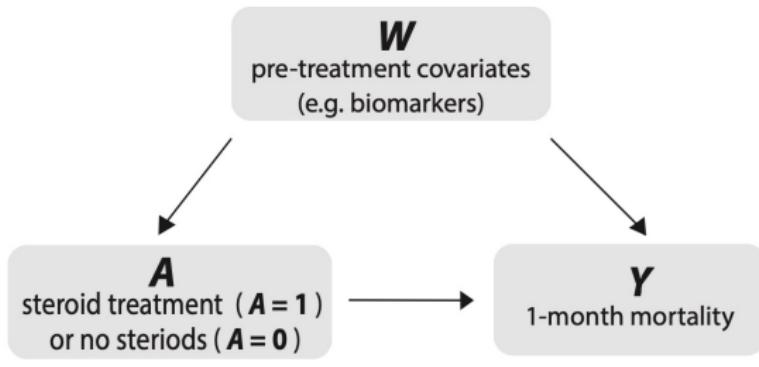
Variable Importance

Conclusion



Three multi-national RCTs assessing impact of corticosteroids on mortality among septic shock patients

Pooled sample of $n = 1,300$ adults in septic shock



Step 2: What is known about stochastic relations of the observed variables?

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

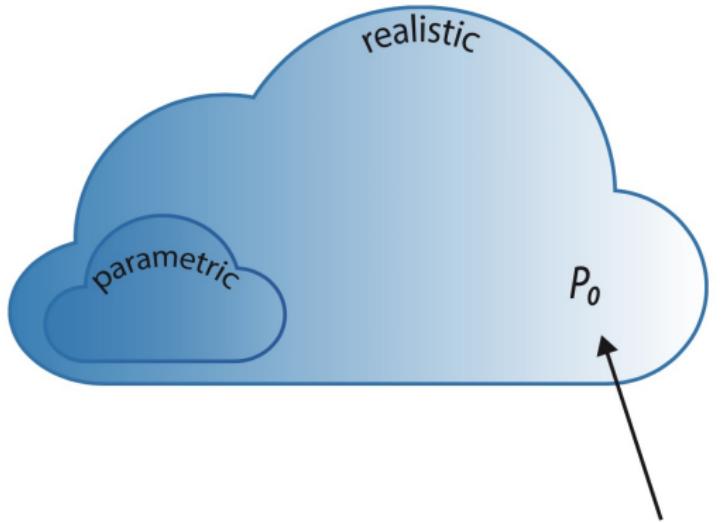
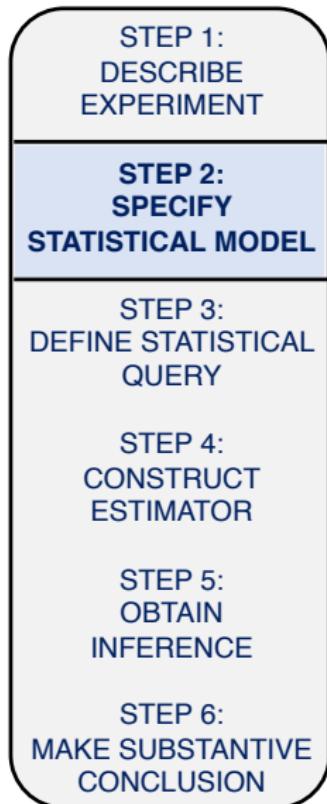
Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion



What happens when the statistical model is misspecified (i.e. does not contain P_0 , the DGP)?

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

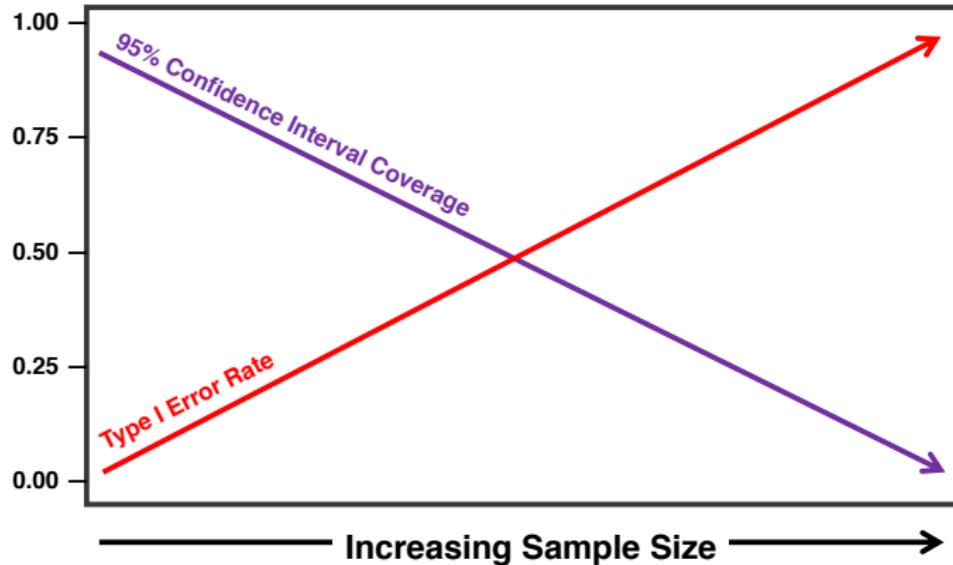
Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion



Step 3a: What is the target causal estimand that we aim to identify from the data?

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

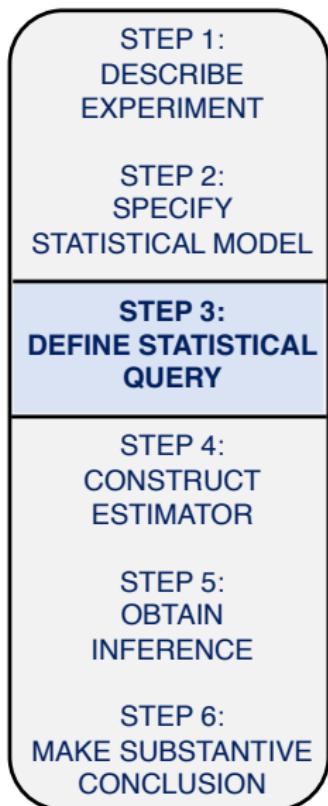
Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion



What is the causal risk ratio in mortality between treatment groups?

$$\psi_{causal} = \frac{P(Y_1 = 1)}{P(Y_0 = 1)}$$

Proportion of subjects in the population of interest that would have died had they all received steroids

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

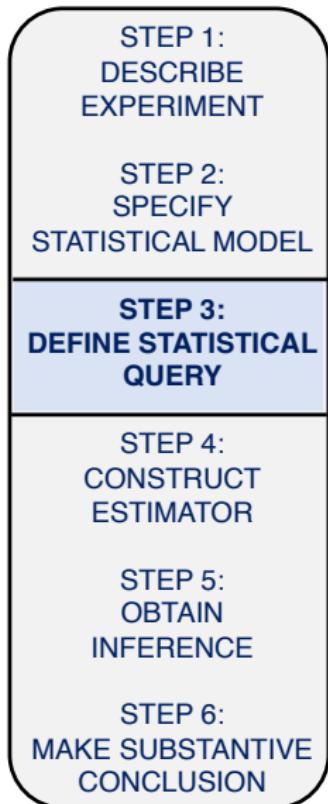
Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

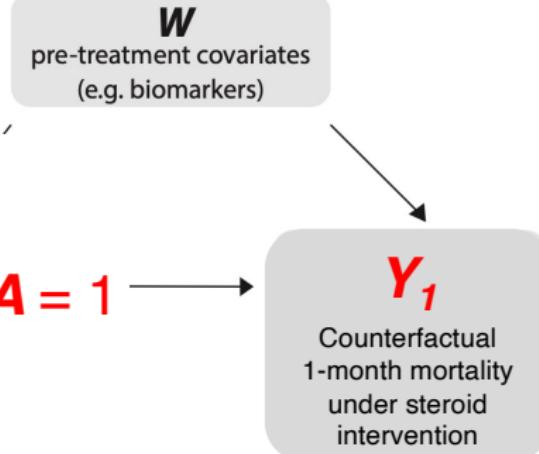
Variable Importance

Conclusion



What is the causal risk ratio in mortality between treatment groups?

$$\psi_{causal} = \frac{P(Y_1 = 1)}{P(Y_0 = 1)}$$



Proportion of subjects in the target population that would have died had they all not received steroids

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

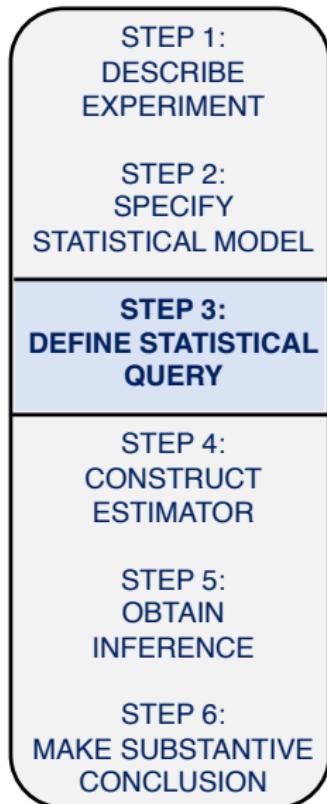
Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

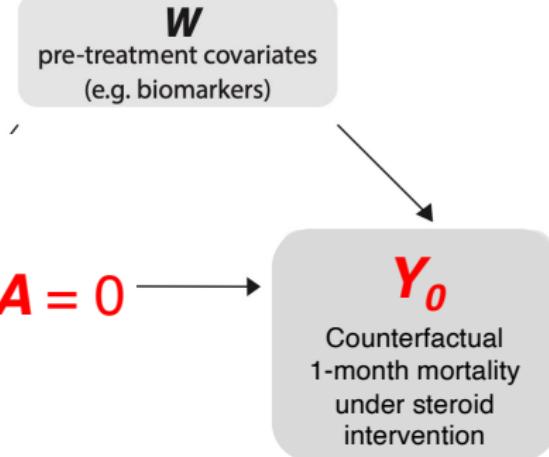
Variable Importance

Conclusion



What is the causal risk ratio in mortality between treatment groups?

$$\psi_{causal} = \frac{P(Y_1 = 1)}{P(Y_0 = 1)}$$



Causal target parameters are functions of the full data under the intervention(s) of interest

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

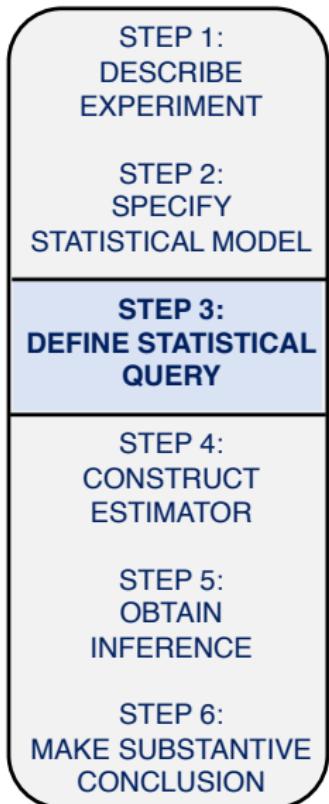
Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion



What is the causal risk ratio in mortality between treatment groups?

$$\psi_{causal} = \frac{P(Y_1 = 1)}{P(Y_0 = 1)}$$

W_1	W_2	W_3	A	Y	Y_1	Y_0
21	65	F	1	1	1	?
22.3	28	F	0	0	?	0
19.4	49	F	1	0	0	?
24	77	M	1	0	0	?

Causal identifiability assumptions must hold in order to interpret the estimate causally

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

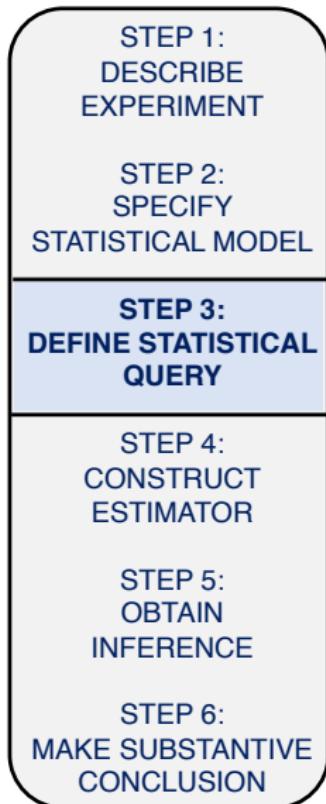
Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion



What is the causal risk ratio in mortality between treatment groups?

$$\psi_{causal} = \frac{P(Y_1 = 1)}{P(Y_0 = 1)}$$

What's needed to ***identify*** ψ_{causal} from the observed data?

1. No unmeasured confounding / randomization / exchangeability

Some causal identifiability assumptions are also necessary for well-defined statistical estimation

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

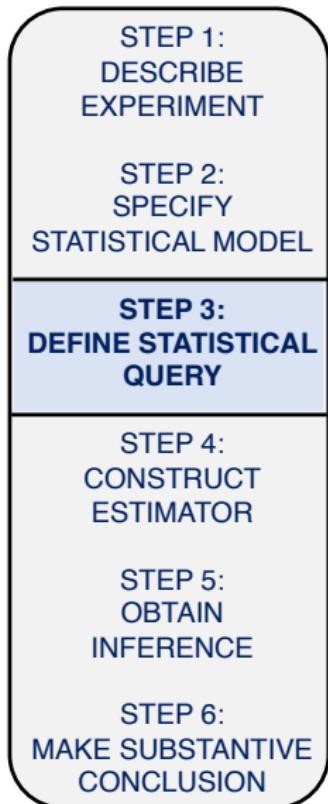
Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion



What is the causal risk ratio in mortality between treatment groups?

$$\psi_{causal} = \frac{P(Y_1 = 1)}{P(Y_0 = 1)}$$

What's needed to ***identify*** ψ_{causal} from the observed data?

1. No unmeasured confounding / randomization / exchangeability
2. Positivity / experimental treatment assignment (ETA) assumption

Identification formula from g-computation

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

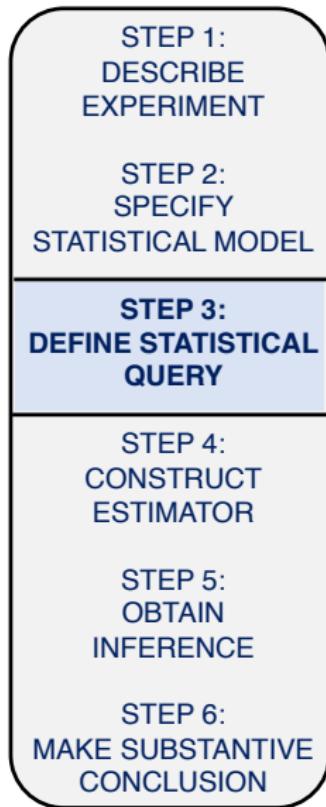
Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion



$$\psi_{causal} = \frac{P(Y_1 = 1)}{P(Y_0 = 1)}$$

$$= \frac{\sum_w P(Y = 1|A = 1, W = w)P(W = w)}{\sum_w P(Y = 1|A = 0, W = w)P(W = w)}$$

$$= \psi_{stat}$$

Step 3b: What is the target statistical estimand that we will learn from the data?

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

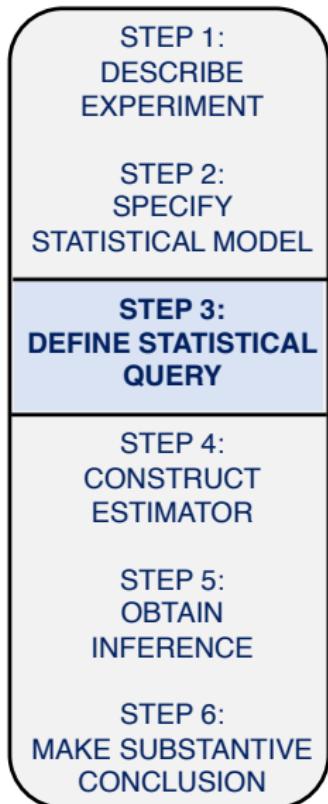
Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion



What is the risk of mortality between treatment groups when adjusting for covariates?

$$\psi_{stat} =$$

$$\frac{\sum_w P(Y = 1|A = 1, W = w)P(W = w)}{\sum_w P(Y = 1|A = 0, W = w)P(W = w)}$$

Step 4: How should we estimate the target estimand?

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

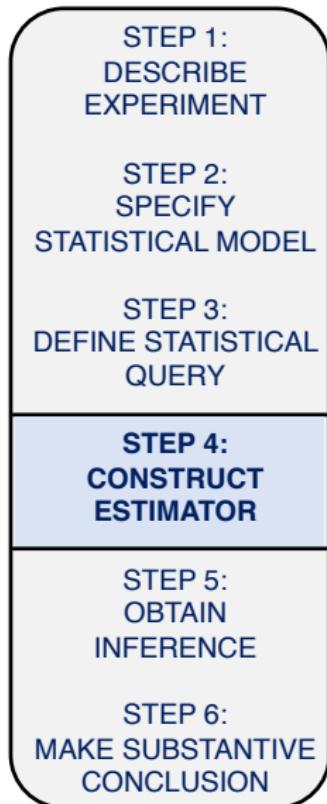
Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion



Statistical properties to consider

- Substitution / plug-in
- Valid inference
- Efficiency
- Ability to optimize finite sample performance

Why plug-in estimation?

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

- Consider a data unit $O = (W, A, Y)$, where $O \sim P_0 \in \mathcal{M}$ for P_0 the data-generating distribution in a model \mathcal{M} .
- A *plug-in* estimator $\Psi(\hat{P}_n)$ is based on the parameter mapping, where the target parameter is $\Psi(P_0)$ and \hat{P}_n is an estimator of P_0 .
- $\Psi(P_0) = \mathbb{E}_0\{\mathbb{E}_0[Y | A = 1, W] - \mathbb{E}_0[Y | A = 0, W]\}$, the ATE functional, can be approximated by recognizing that:
 - Components of P_0 that impact the plug-in estimator are the conditional mean $\bar{Q}_0(A, W) = \mathbb{E}[Y | A, W]$ and the marginal distribution of W , $Q_{0,W}$.
 - Thus, the plug-in estimator depends on estimates of these: $\bar{Q}_n(A, W) = \hat{\mathbb{E}}[Y | A, W]$ and $Q_{n,W} = P_n(W = w)$.
- By construction, plug-in estimators remain within the bounds of the parameter space.

Targeted Maximum Likelihood Estimation (TMLE)

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

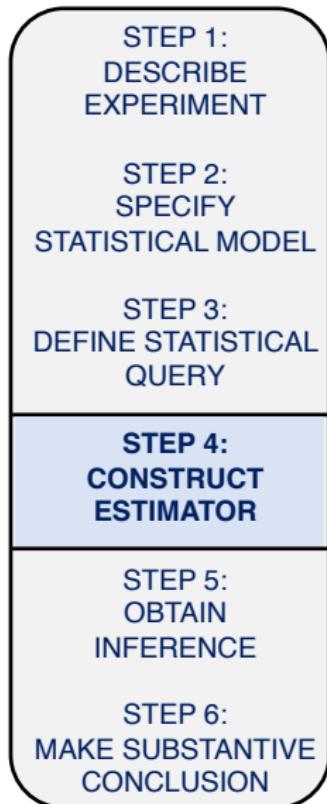
Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion



TMLE

- 1 Initial estimation of $P[Y = 1|A, W]$ with super (machine) learning
- 2 Updating initial estimate to achieve optimal bias-variance trade-off for ψ_{stat}

TMLE estimates are optimal:
plug-in, efficient, unbiased, finite sample robust

Initial estimates via super learner

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

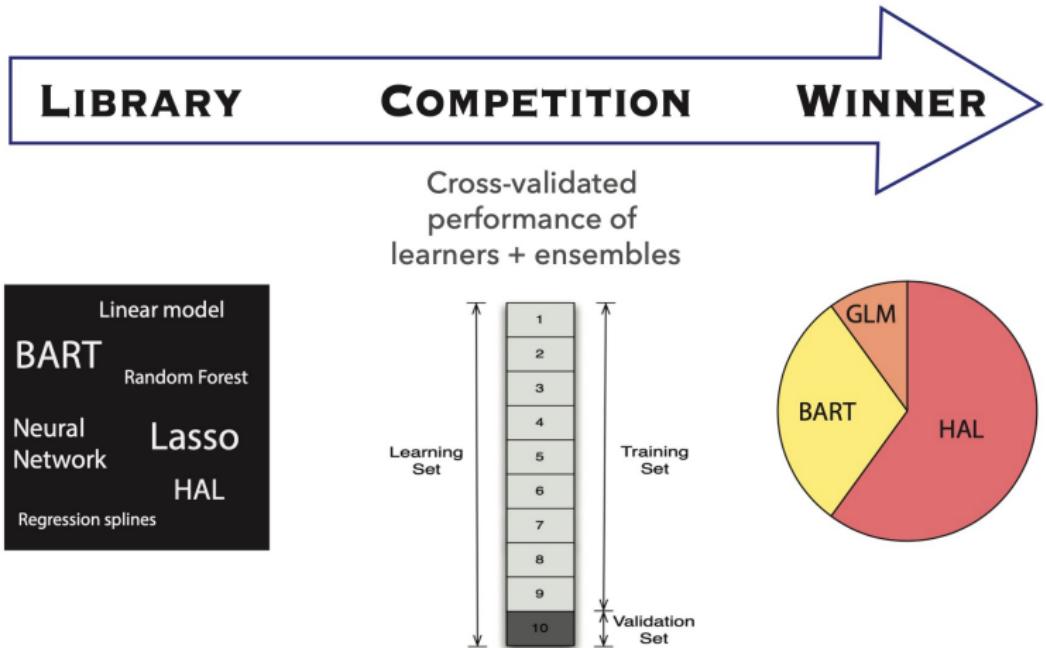
Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion



Hugely advantageous when coupled with NLP-derived covariates with EHR

Cross-validation to choose a winner

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

1	1	1	1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3	3	3	3
4	4	4	4	4	4	4	4	4	4	4
5	5	5	5	5	5	5	5	5	5	5
6	6	6	6	6	6	6	6	6	6	6
7	7	7	7	7	7	7	7	7	7	7
8	8	8	8	8	8	8	8	8	8	8
9	9	9	9	9	9	9	9	9	9	9
10	10	10	10	10	10	10	10	10	10	10

Fold 1 Fold 2 Fold 3 Fold 4 Fold 5 Fold 6 Fold 7 Fold 8 Fold 9 Fold 10

Conclusion

Coding exercise: Super learning with s13

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

<https://tlverse.org/catalyst2024-workshop/s13.html>

Tip: Include the Highly Adaptive Lasso (HAL) as a candidate in the SL library

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

- First estimator to guarantee asymptotically efficient estimation of any pathwise differentiable estimand¹ (e.g., average causal effect or treatment-specific survival), without enforcing local smoothness conditions.
- Mild assumptions, which hold in most practical applications.
- Regularization step can be implemented with standard Lasso software (e.g., `glmnet`).
- Converges to true function at rate $n^{-1/3}(\log n)^{d/2}$.
- Accommodates a variety of function space specifications.

¹An estimand that is a weakly differentiable functional of the density of the data, the case for most causal inference estimands under positivity.



Illustration in Low Dimensions

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

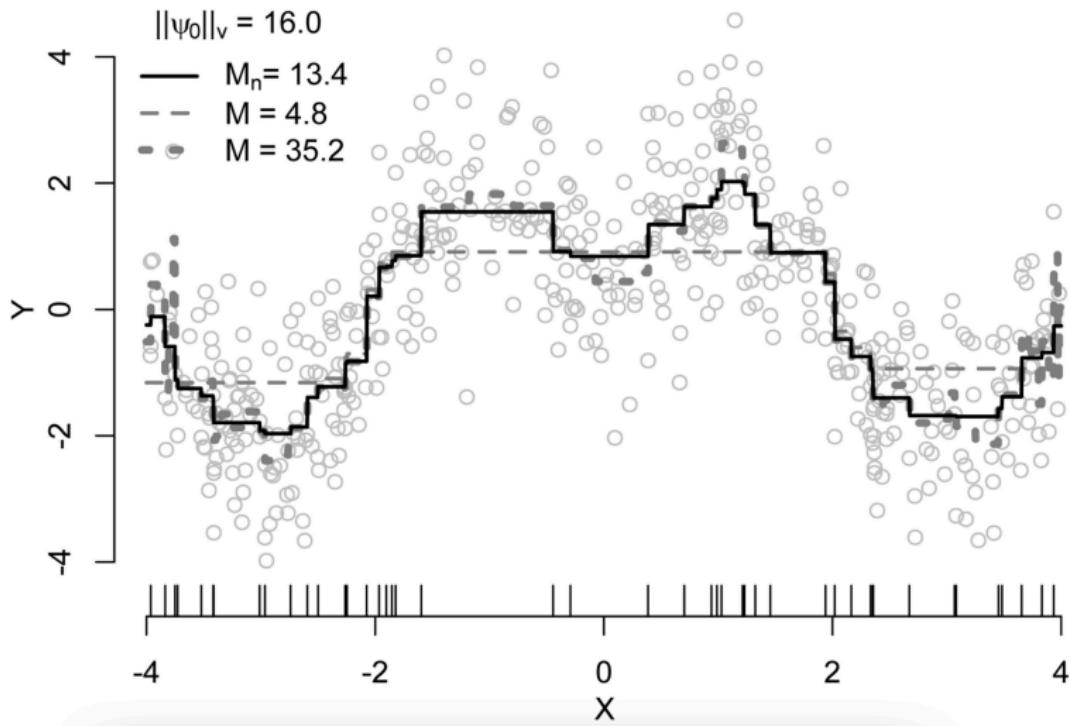
Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion



TMLE: Targeting by following a path of maximal change in target estimand per unit likelihood

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

TMLE Example for the ATE functional

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

- Consider again $O = (W, A, Y)$ where $O \sim P_0 \in \mathcal{M}$.
- $\Psi(P_0) = \mathbb{E}\{\mathbb{E}[Y | A = 1, W] - \mathbb{E}[Y | A = 0, W]\}$, the ATE functional, has a plug-in estimator:

$$\Psi(Q_n) = \frac{1}{n} \sum_{i=1}^n \{\bar{Q}_n(1, W) - \bar{Q}_n(0, W)\},$$

where $\bar{Q}_n(a, W) = \hat{\mathbb{E}}[Y | A = a, W]$.

- The TMLE based on this plug-in estimator is

$$\Psi(Q_n^*) = \frac{1}{n} \sum_{i=1}^n \{\bar{Q}_n^*(1, W) - \bar{Q}_n^*(0, W)\},$$

where \bar{Q}_n^* is an update of the initial estimate $\bar{Q}_n(a, W)$

- The update step uses a parametric fluctuation of the form $\text{logit}(\bar{Q}_n^*) = \text{logit}(\bar{Q}_n) + \epsilon H_n(A, W)$, where the form of $H_n(A, W)$ depends on relevant efficiency theory.

TMLE Example for the ATE functional

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

- The update step uses a parametric fluctuation of the form $\text{logit}(\bar{Q}_n^*) = \text{logit}(\bar{Q}_n) + \epsilon H_n(A, W)$
- $H_n(A, W) = \{I(A = 1)/g_n(W)\} - \{I(A = 0)/(1-g_n(W))\}$, a weight in the efficient influence function (EIF) of regular asymptotically linear (RAL) estimators of $\Psi(P_0)$ wrt \mathcal{M} .
- ϵ_n is a maximum likelihood estimator of ϵ , so that the TMLE is $\Psi(P_n^*)$, where $P_n^* = P_n^0(\epsilon_n)$.

Analysis of TMLE (vdL, Rubin, 2006)

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

- Construct initial estimator \mathbf{P}_n ; determine a least favorable path $\{\mathbf{P}_{n,\epsilon} : \epsilon \in (-\delta, \delta)\} \subset \mathcal{M}$ through \mathbf{P}_n with score $D_{\mathbf{P}_n}^*$ at $\epsilon = 0$.
- Compute MLE $\epsilon_n = \arg \max_\epsilon P_n L(\mathbf{P}_{n,\epsilon})$, where $L(P)$ is a valid loss function so that $\Psi(\arg \min_P P_0 L(P)) = \Psi(P_0)$.
- Let $\mathbf{P}_n^* = \mathbf{P}_{n,\epsilon_n}$. The TMLE is given by $\Psi(\mathbf{P}_n^*)$.
- If the mapping $\Psi(\cdot)$ is pathwise differentiable,
$$\Psi(\mathbf{P}_n^*) - \Psi(P_0) = (P_n - P_0)D_{\mathbf{P}_n^*}^* + R(\mathbf{P}_n^*, P_0),$$
 with
$$R(P, P_0) \equiv \Psi(P) - \Psi(P_0) - (P - P_0)D_P^*$$
 an exact second-order remainder; assume $R(P_n^*, P_0) = o_P(n^{-1/2})$.
- Use super learner to obtain initial estimator \mathbf{P}_n , then the TMLE $\Psi(\mathbf{P}_n^*)$ will be an asymptotically efficient estimator of $\Psi(P_0)$ under regularity conditions.

Asymptotic Linearity of the TMLE

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

- The TMLE $\Psi(P_n^*)$ is a RAL estimator, from which it follows that its difference from $\Psi(P_0)$, the truth, can be approximated by an average of i.i.d. RVs, the EIF, of O :

$$\Psi(P_n^*) - \Psi(P_0) \approx \frac{1}{n} \sum_{i=1}^n \text{IC}(P_0)(O_i) ,$$

where $\text{IC}(P_0)(O)$ is the EIF at $P_0 \in \mathcal{M}$.

- Asymptotic linearity of $\Psi(P_n^*)$ implies a normal sampling distribution of the TMLE (by the CLT):

$$\sqrt{n}(\Psi(P_n^*) - \Psi(P_0)) \xrightarrow{d} \mathcal{N}(0, \sigma^2) ,$$

where $\sigma^2 = \mathbb{V}(\text{IC}(P_0)(O))$ is the asymptotic variance of the EIF at $P_0 \in \mathcal{M}$.

- $\text{IC}(P_0)(O)$ depends on both (g_0, Q_0) , the source of the TMLE update's dependence on g_0 .

How should we approximate the sampling distribution of our estimator?

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

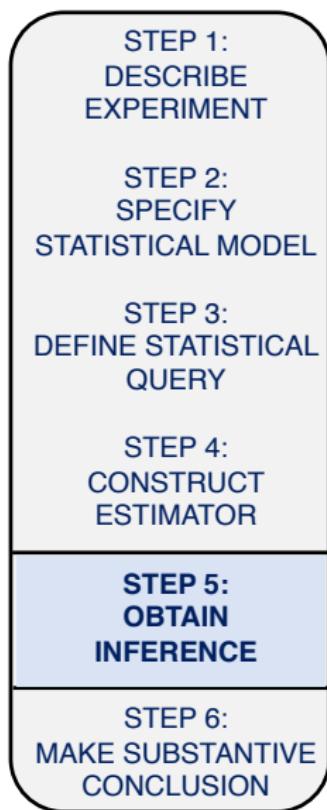
Collaborative TMLE

HAL and A-TMLE

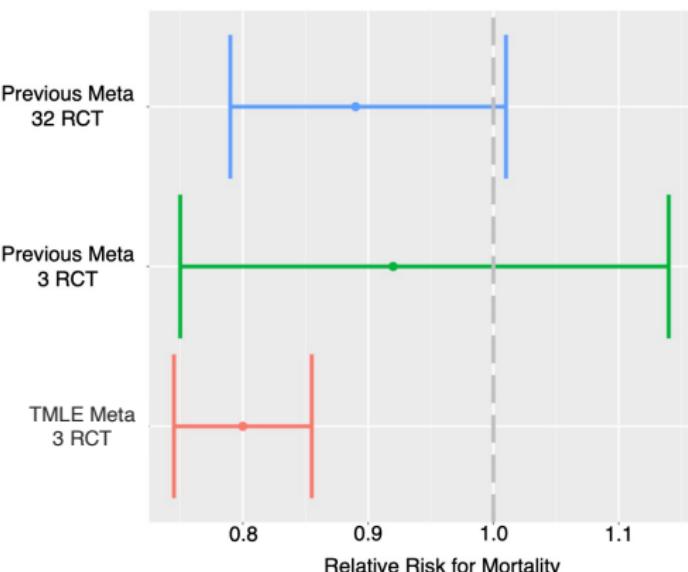
Longitudinal TMLE

Variable Importance

Conclusion



Due to targeting (step ②), the TMLE behaves as the *sample mean* of efficient influence function



Coding exercise: TMLE with tmle3

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

[https:](https://tlverse.org/catalyst2024-workshop/tmle3.html)

//tlverse.org/catalyst2024-workshop/tmle3.html

Arriving at the substantive conclusion

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

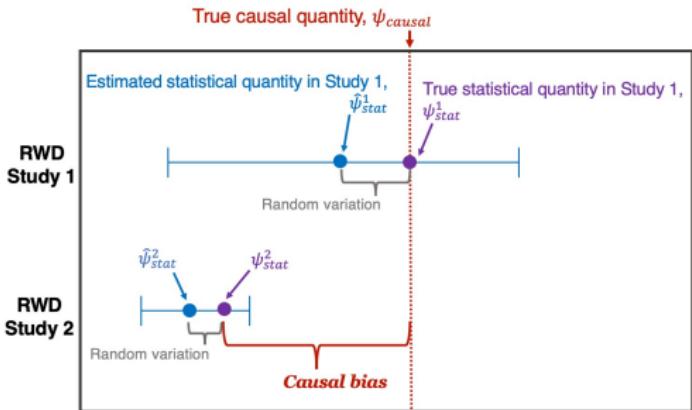
Variable Importance

Conclusion

- STEP 1: DESCRIBE EXPERIMENT
- STEP 2: SPECIFY STATISTICAL MODEL
- STEP 3: DEFINE STATISTICAL QUERY
- STEP 4: CONSTRUCT ESTIMATOR
- STEP 5: OBTAIN INFERENCE
- STEP 6: MAKE SUBSTANTIVE CONCLUSION

Investigate causal bias with sensitivity analysis

Causal bias: Gap between estimate and truth due to violations of any of the causal assumptions (e.g., unmeasured confounding)*



Sensitivity Analysis: Model-free assessment of how reasonable departures from causal assumptions would impact study findings

* Sensitivity analysis can be extended to incorporate statistical bias

TL-based non-parametric sensitivity analysis: RCT with 25% LTFU

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

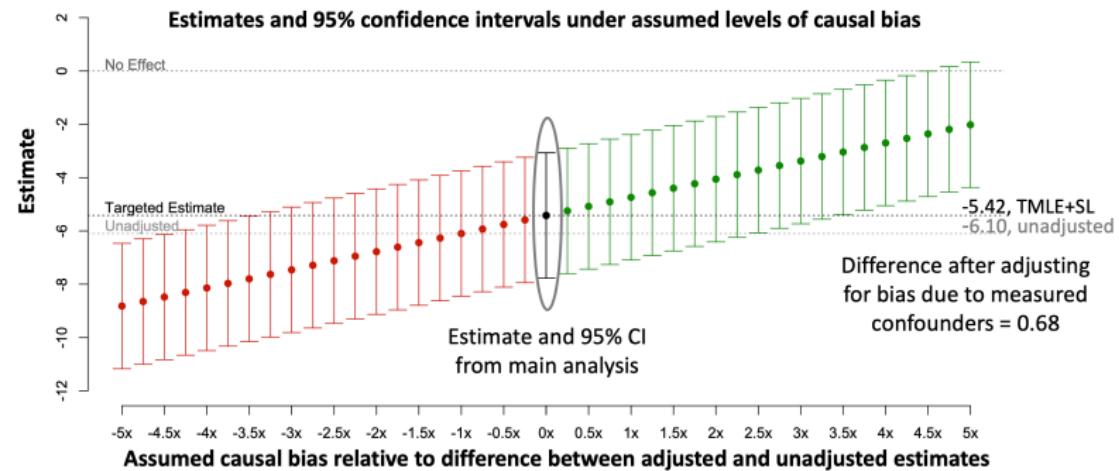
Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion



Courtesy of "Targeted-Learning Based Statistical Analysis Plan" Webinar by Susan Gruber on 28 April 2021

TL-based non-parametric sensitivity analysis: Safety analysis example

Targeted
Learning

Mark van der
Laan and
Nima Hejazi

The Art of
Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic
statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

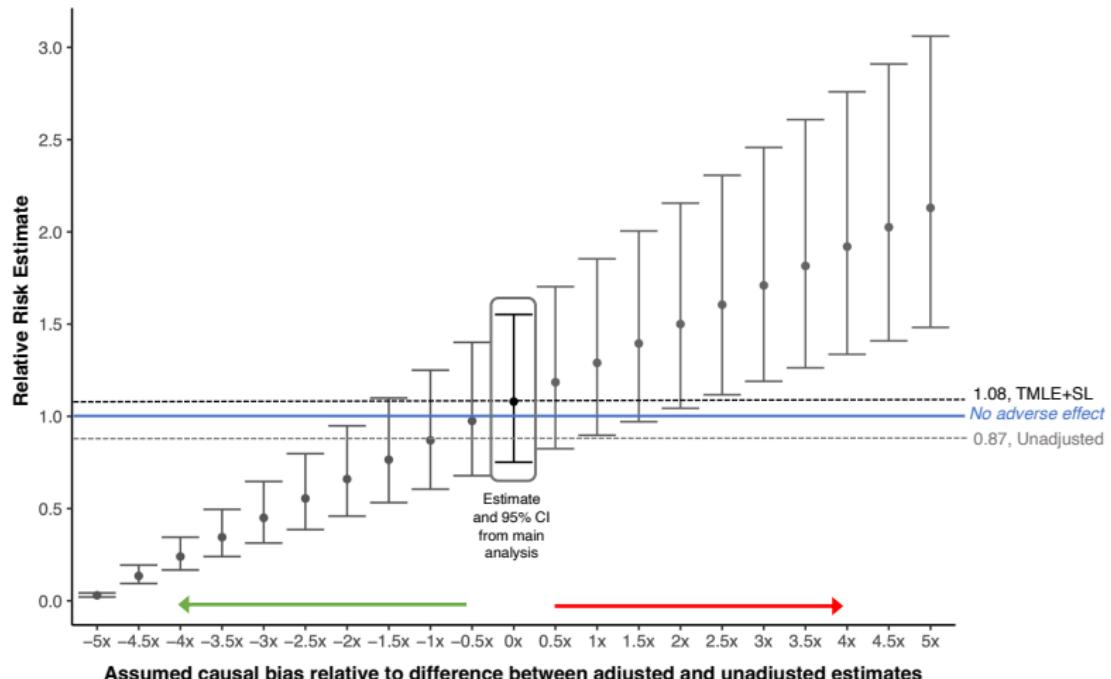
HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

Relative risk estimates and 95% confidence intervals under assumed levels of causal bias



Possibility to refine question of interest and inform future studies

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

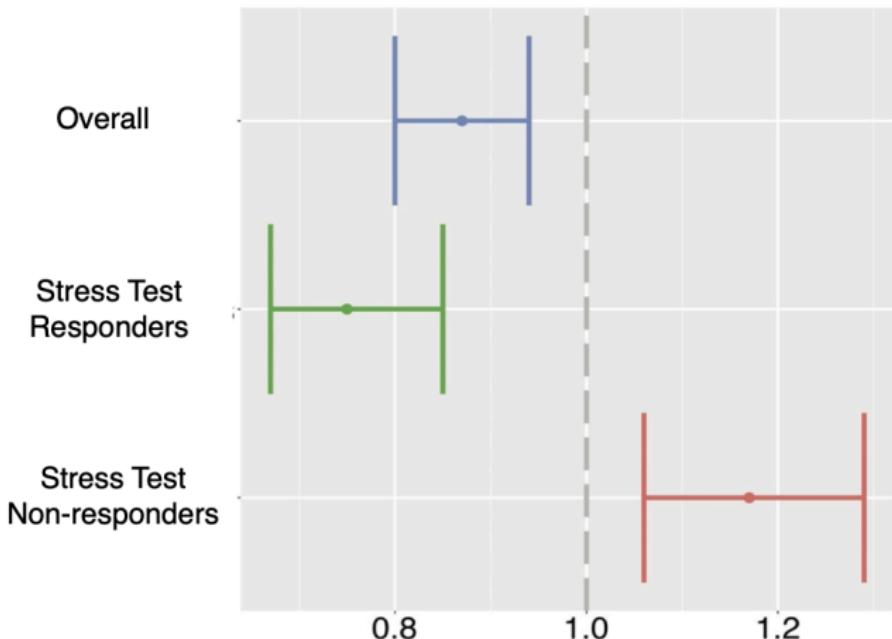
HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

What subgroup of patients in septic shock benefit from corticosteroids?



Outcome blind simulations to a priori define TMLE: Wyss et al., 2023

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

Step 4. Estimation and Inference using TMLE + SL

TMLE software default settings are robust across many situations, but not optimized for your analysis

Recommendation

Compare **coverage, type I error, power** for different specifications using outcome blind simulations, or plasmode simulations based on external data (pilot, Phase 1) sharing key features with your data

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Step 4. Estimation and Inference using TMLE + SL

Case study 1: Evaluating safety and efficacy with **high dimensional RWD**

$$\psi_{RD}^{stat} = E[E(Y | A = 1, W)] - E[E(Y | A = 0, W)]$$

Challenges

- Lack of baseline randomization
- How to model propensity score and outcome regression?
 - Thousands of covariates in linked EHR + claims
 - Traditional parametric model too high dimensional, unknown functional form



Outcome-regression weighted LASSO (OAL)

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

Shortreed & Ertefaie (2017) proposed outcome-regression weighted Lasso (OAL) for propensity score (PS) estimation:

- Fit unpenalized linear model for $\mathbb{E}(Y | A, W)$:

$$(\hat{\alpha}, \hat{\eta}) = \arg \min_{\alpha, \eta} I_n(\alpha; Y, A, W).$$

where η is the coefficient for A , and α is the coefficient for W .

- Denote the coefficient for variable W_j in the outcome regression with $\hat{\alpha}_j$.
- Fit PS with Lasso using regularization term

$$\lambda \sum_j ||\alpha_j||^{-\gamma} ||\beta_j|| \text{ instead of usual } \lambda \sum_j |\beta_j| .$$

HAL-based OAL for PS Estimation

Targeted
Learning

Mark van der
Laan and
Nima Hejazi

The Art of
Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic
statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

The theoretical property of OAL relies on the ***correct parametric formula***, which is often unknown in practice. We extend OAL to outcome-regression weighted HAL (OHAL):

- ① Compute the outcome regression using Lasso of form $\sum_j \alpha_n(j) \phi_j(W, A)$.
- ② Use as basis functions for the PS $\{\phi_j(1, W), \phi_j(0, W) : j\}$.
- ③ Both of these two basis functions will be associated with same weight $\alpha_n(j)$.
- ④ Compute the propensity score using a Lasso logistic regression using the above basis functions. The L_1 -constraint for β_j 's, the coefficient for ϕ_j , is defined as the weighted L_1 -norm above.
- ⑤ Or simply define $\|\beta\|_1 = \sum_{j, \alpha_n(j) \neq 0} |\beta(j)|$.

C-TMLE to select L_1 -norm

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

- Tune the L_1 -norm with C-TMLE: i.e. optimize in λ increase in likelihood of TMLE-step using $g_{n,\lambda}$.
- Can be combined with standard lasso to yield a collaborative-controlled (CC) regularization for TMLE.
- Similarly can be used in a collaborative-controlled outcome-adaptive lasso (CC-OAL) for TMLE.

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

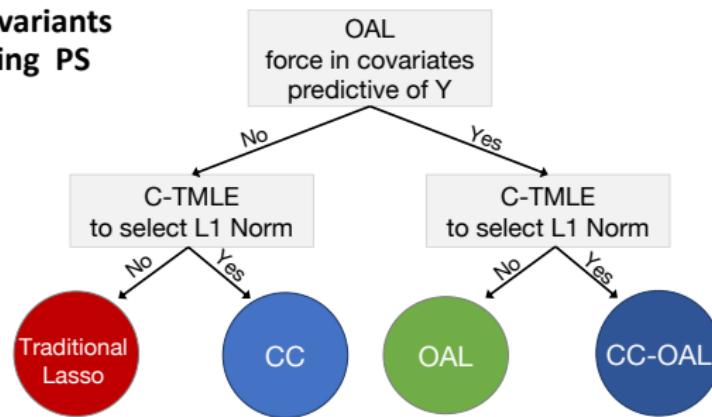
HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Step 4. Estimation and Inference using TMLE + SL

Four Lasso variants for estimating PS



Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

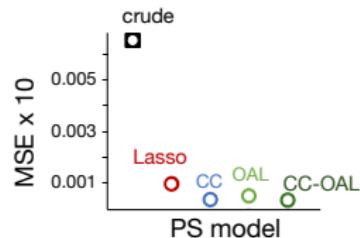
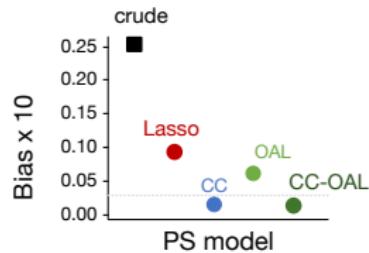
Longitudinal TMLE

Variable Importance

Conclusion

Step 4. Estimation and Inference using TMLE + SL

Plasmode study results



Collaborative control greatly reduced bias and improved MSE

- Less regularization captured more relevant confounder information in PS

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Step 4. Estimation and Inference using TMLE + SL

Case study 1: Data analysis

- Analyze study data
 - Outcome regression: Lasso with expert-selected covariates forced in
 - PS: Collaborative control Outcome-adaptive lasso (CC-OAL)
 - force in covariates selected in the outcome, others as candidates
 - C-TMLE to choose L1 norm
- Result: $RD = 0.005$ (95% CI: -0.027, 0.038)
crude $RD = 0.024$ (95% CI: 0.018, 0.030)

Crude estimate indicates NSAIDS increase risk for AKI

Adjustment moves point estimate towards the null

8

Conclusion



Zero-order Highly Adaptive Lasso (HAL): A nonparametric MLE

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

Key Idea

- Any d -dimensional cadlag function (i.e. right-continuous, left limits) can be represented as an infinite linear combination of spline basis functions.
- The variation norm / complexity of a function is the L_1 -norm of the vector of coefficients.

Zero-order HAL converges to true function at rate $n^{-1/3}(\log n)^{d/2}$ (only assuming finite variation norm)

Formal representation of cadlag function as linear combination of zero-order splines

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

- A cadlag function $f \in D^{(0)}([0, 1]^d)$ can be represented as

$$f(x) = f(0) + \sum_{s \subset \{1, \dots, d\}} \int_{(0_s, x_s]} df_s(u),$$

where $f_s(x_s) = f(x(j))I(j \in s) : j = 1, \dots, d$ is the section implied by setting coordinates outside s equal to zero, and $x_s = (x(j) : j \in s)$.

- Moreover, we define the sectional variation norm of f as the sum over s of the variation norms of f_s :

$$\|f\|_v^* = \sum_s \|f_s\|_v = |f(0)| + \sum_s \int_{(0_s, 1_s]} |df_s(u)|.$$

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

- Now, notice that this writes f as an infinite linear combination of $x \rightarrow I(x_s \geq u)$ of zero order splines with knot-point $u \in (0_s, 1_s]$ and coefficient $df_s(u)$, and that the sectional variation norm is the L_1 -norm in this representation.

Zero-order HAL performance for d=3

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

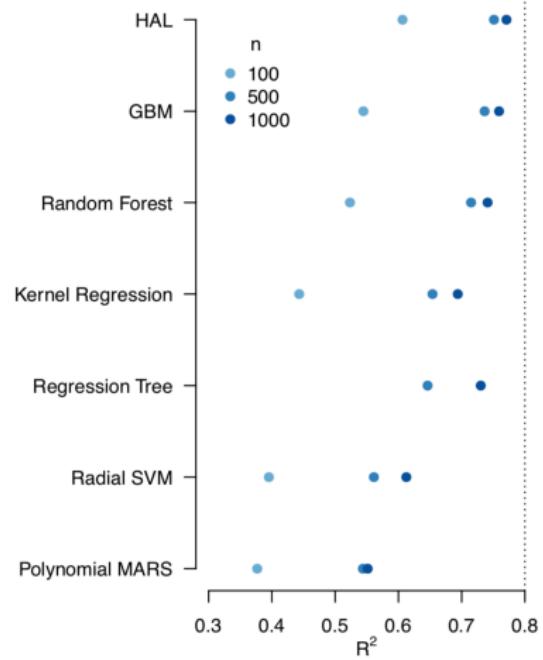
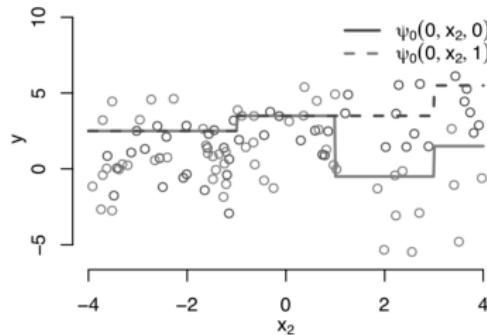
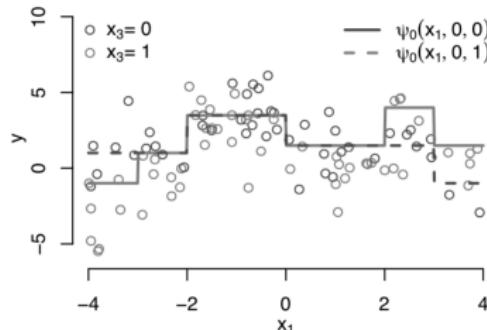
HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

$$\begin{aligned}\psi_0(x) = & -2x_3 \mathbb{I}(x_1 < -3) + 2.5 \mathbb{I}(x_1 > -2) - 2 \mathbb{I}(x_1 > 0) + 2.5x_3 \mathbb{I}(x_1 > 2) \\ & - 2.5 \mathbb{I}(x_1 > 3) + \mathbb{I}(x_2 > -1) - 4x_3 \mathbb{I}(x_2 > 1) + 2 \mathbb{I}(x_2 > 3)\end{aligned}$$



HAL Provides Estimators of Large Variety of Nuisance Parameters Needed in Causal Inference

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

- Causal Inference requires statistical estimation of nuisance functions.
- In particular, it requires estimation of conditional means; conditional densities or cumulative distribution functions; intensities, conditional hazards. Standard loss functions can be employed for these.
- Moreover, it is often possible to define risk functions that define target functions of interest itself. For example, a variety of risk functions have been proposed for the conditional treatment effect. One can then apply HAL to minimize the empirical risk function.
- In these cases, estimation of the empirical risk function requires itself nuisance parameter estimation, where again HAL could be used.

How to develop a HAL-MLE: parametrize target function in terms of unrestricted function

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

- Suppose one is interested in a functional parameter $Q(P)$ such as a conditional density.
- Select a loss function or risk function so that $Q(P) = \arg \min_Q R_P(Q)$, where, for example, $R_P(Q) = PL(Q)$ for some loss $L(Q)$.
- Q might be constrained in some ways. Therefore, find a parametrization $Q = Q_f$ in terms of an unconstrained function f . For example, parametrize a conditional density in terms of conditional hazard and represent the latter as $\exp(f)$.
- Now, model f as a linear combination of zero order splines and compute the MLE

$$\beta_n = \arg \min_{\beta, \|\beta\|_1 < C_n} R_n \left(Q_{\sum_j \beta(j) \phi_j} \right), \text{ where } R_n(Q) \text{ is an estimate of the risk } R_{P_0}(Q) \text{ such as } R_n(Q) = P_n L(Q).$$

- The HAL-MLE of Q_0 is then given by $Q_n = Q_{f_n}$ with



Additive models within the cadlag function space to define subspace specific HALs

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

- Instead of selecting the richest set of basis function such as the ones implied by knot-points $\{X_i(s) : i = 1, \dots, n, s \subset \{1, \dots, d\}\}$, one can only model a subset \mathcal{S}_1 of all the additive functions $\tilde{f}_s(x_s) = \int I(x_s \geq u) df_s(u)$ by not including all subsets s in \mathcal{S}_1 .
- If \mathcal{S}_1 represents the collection of all subsets we include, then this defines additive models $f(x_s) = f(0) + \sum_{s \in \mathcal{S}_1} \tilde{f}_s(x_s)$. One can then define a corresponding HAL-MLE f_{n, \mathcal{S}_1} .
- More generally, we define $D^{(0)}(\mathcal{R}^{0,*})$ as the linear span of $\{\phi_j : j \in \mathcal{R}^{0,*}\}$, and by choosing the richest set \mathcal{R}^0 , we have $D^{(0)}(\mathcal{R}^0) = D^{(0)}([0, 1]^d)$.
- We use $D_M^{(0)}(\mathcal{R}^{0,*})$ when bounding L_1 -norm by M .

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

- Every choice of subset of basis functions then implies an HAL-MLE.
- One could also first use an initial ML-algorithm, such as MARS, to learn the family of subsets, \mathcal{S}_1 , that appears to be needed, and then compute the resulting HAL-MLE f_{n,\mathcal{S}_1} .
- Of course, the screening algorithm is then part of algorithm, which needs to be respected when using cross-validation to select the L_1 -norm or select tuning parameter of the initial screening.

Rate of convergence of zero order HAL

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

- Let $Q_n = Q_{f_n}$ be the HAL-MLE.
- Let $d_0(Q, Q_0) = P_0 L(Q) - P_0 L(Q_0)$ be the loss based dissimilarity.
- We have

$$\begin{aligned} d_0(Q_n, Q_0) &= P_n \{L(Q_n) - L(Q_0)\} \\ &\quad - (P_n - P_0) \{L(Q_n) - L(Q_0)\} \\ &\leq -(P_n - P_0) \{L(Q_n) - L(Q_0)\}. \end{aligned}$$

- Let $\mathcal{F} = \{L(Q_f) - L(Q_0) : f \in D_M^{(0)}([0, 1]^d)\}$.
- The (known) covering number for $D_M^{(0)}([0, 1]^d)$ implies the same covering number for \mathcal{F} .

Using empirical process theory:

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

- Let $\mathcal{F}(\delta) = \{f \in \mathcal{F} : P_0 f^2 \leq \delta^2\}$.
- $\sup_{f \in \mathcal{F}(\delta)} |n^{1/2}(P_n - P_0)f|$ can be bounded by the entropy integral $J(\delta, \mathcal{F}(\delta)) = \int_{(0,\delta]} \sqrt{\log N(\epsilon, \mathcal{F}, L^2)} d\epsilon$, which behaves as $\delta^{1/2}$ up till $\log \delta$ -factor.
- Using that $P_0(L(Q) - L(Q_0))^2 \leq Cd_0(Q, Q_0)$, we can then apply an iterative proof bounding

$$d_0(Q_n, Q_0) \leq n^{-1/2} \sup_{f \in \mathcal{F}(\delta^k)} |n^{1/2}(P_n - P_0)f| \quad (1)$$

$$\sim n^{-1/2} J(\delta^k, \mathcal{F}), \quad (2)$$

starting with $\delta^0 = 1$, $\delta^1 = n^{-1/4}$, and so on.

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

- The monotone improving (in rate) sequence converges to the fixed point δ^* of equation

$$\delta^2 = n^{-1/2} J(\delta, \mathcal{F}).$$

- This δ^* equals the rate of convergence for $d_0^{1/2}(Q_n, Q_0)$.
- We find $d_0(Q_n, Q_0) = O_P(n^{-2/3}(\log n)^d)$.

Basic R hal9001 Functionality

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

- ① Load package and data:

```
library(hal9001)  
data(mtcars)
```

- ② Create numeric vector for dependent variable:

```
Y <- mtcars[, "mpg"]
```

- ③ Create dataframe or matrix of predictors:

```
X <- mtcars[, c("cyl", "disp", "hp", "wt")]
```

- ④ Fit HAL:

```
hal_fit <- fit_hal(X=X, Y=Y)
```

Note: default max_degree=3 considers no more than 3-way interactions, and default reduce_basis=0 places no restrictions on the minimum proportion of 1's in basis functions.

Summary table of hal9001 HAL fit

Targeted
Learning

Mark van der
Laan and
Nima Hejazi

The Art of
Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic
statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

```
summary(hal_fit)$table
```

coef	term
35.4070	Intercept
-4.0801	$I(disp >= 440)$
-4.0118	$I(disp >= 78.7)$
-2.7170	$I(wt >= 1.513)$
-2.4454	$I(wt >= 3.215)$
-1.8184	$I(disp >= 71.1)$
-1.7208	$I(hp >= 180)$
-1.6830	$I(disp >= 95.1)$
-1.6039	$I(hp >= 66)*I(wt >= 2.2)$
-1.5623	$I(wt >= 2.2)$
1.3785	$I(disp >= 351)$
-1.2444	$I(hp >= 175)$
-1.1888	$I(disp >= 301)$
-0.9026	$I(hp >= 123)$
0.7336	$I(hp >= 52)$
-0.5810	$I(disp >= 120.1)*I(hp >= 97)$
-0.4395	$I(disp >= 108)*I(hp >= 93)$

Specifying hal9001 model formulas

Targeted
Learning

Mark van der
Laan and
Nima Hejazi

The Art of
Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic
statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Example: Observe $O = (W_1, W_2, A, Y) \sim P_0$

R code: `fit_hal(Y, X, formula, data, ...)`

Additive model formula:

$Y \sim .$ or $Y \sim h(W_1) + h(W_2) + h(A)$ **Bi-additive model**

formula:

$Y \sim .^2$ or

$Y \sim h(W_1) + h(W_2) + h(A) + h(W_1, W_2) + h(W_1, A) + h(W_2, A)$ **Only**

interactions with A formula:

$Y \sim h(.) + h(., A)$ or

$Y \sim h(W_1) + h(W_2) + h(A) + h(W_1, A) + h(W_2, A)$ **Monotone** \uparrow **(i)**

\downarrow **(d) formula examples:** $Y \sim i(.)$ or $Y \sim i(.) + i(., .)$ or

$Y \sim i(W_1) + d(W_2) + i(A)$

Conclusion

Possible HAL fits under various smoothness orders

Targeted
Learning

Mark van der
Laan and
Nima Hejazi

The Art of
Statistics

TL in Action

TL Roadmap

Describe study
Specify a realistic
statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

Example: Observe $(W, A, Y) \sim P_0$

R code: `fit_hal(Y, X, data, formula, k, ...)`

Example fits for 0-order smoothness, $k=0$:

Additive model:

$$Y = \mathbb{I}(W > 0.5) + \mathbb{I}(W > 0.3) + \mathbb{I}(A > 0)$$

Bi-additive model:

$$Y = \mathbb{I}(W > 0.5) + \mathbb{I}(A > 0) + \mathbb{I}(W > 0.5, A > 0)$$

Example fits for 1st-order smoothness, $k=1$:

Additive model:

$$Y = \mathbb{I}(W > 0.5)[W - 0.5] + \mathbb{I}(W > 0.3)[W - 0.3] + \mathbb{I}(A > 0)[A - 0]$$

Bi-additive model:

$$Y = \mathbb{I}(W > 0.5)[W - 0.5] + \mathbb{I}(A > 0)[A - 0] + \mathbb{I}(W > 0.5, A > 0)[W - 0.5][A - 0]$$

Derivation of first-order spline representation of cadlag function

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

- In the presentation of $f \in D_M^{(0)}([0, 1]^d)$ there appear integrals $\int_{(0(s), x(s))} df_s(u)$.
- Assume and write $df_s(u) = df_s/dudu$; assume the RN-derivative $f_s^{(1)} = df_s/du \in D_M^{(0)}([0, 1]^d)$, and plug-in the zero order spline representation for

$$f_s^{(1)}(u) = f_s^{(1)}(0) + \sum_{s_1 \subset s} \int_{(0(s_1), u(s_1))} df_{s_1}^{(1)}(u_1).$$

- Apply Fubini's theorem to the double integrals over (u, u_1) to obtain a representation in terms of tensor products of **first** order splines.

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

- As an example, let's do it for a univariate function:

$$\begin{aligned}f(x) &= f(0) + \int_{(0,x]} df(u) \\&= f(0) + \int_{(0,x]} f^{(1)}(u) du \\&= f(0) + \int_{(0,x]} \{f^{(1)}(0) + \int_{(0,u]} df^{(1)}(u_1)\} du \\&= f(0) + xf^{(1)}(0) + \int_{u_1} \int_u I(u \leq x) I(u_1 \leq u) du df^{(1)}(u_1) \\&= f(0) + f^{(1)}(0)x + \int_{u_1} I(x \geq u_1)(x - u_1) df^{(1)}(u_1),\end{aligned}$$

which is a linear combination of first order splines

$\phi_{u_1}^1(x) = I(x \geq u_1)(x - u_1)$, including $\phi_0^1(x) = x$ implied by knot-point $u_1 = 0$.

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

- In this manner we obtain the following first order spline representation for a function $f \in D^{(1)}([0, 1]^d)$ (a cadlag function satisfying our first orders smoothness assumption):

$$f(x) = f(0) + \sum_{s \subset \{1, \dots, d\}} \phi_0^1(x(s)) f_s^{(1)}(0) \\ + \sum_{s, s_1 \subset s, |s_1| > 0} \bar{\phi}_{s, s_1}(x) \int \phi_{u(s_1)}^1(x(s_1)) f_{s, s_1}^{(1)}(du).$$

- Here $\bar{\phi}_{s, s_1}(x(s/s_1)) = \prod_{l \in s/s_1} x(l)$ and $f_{s, s_1}^{(1)}$ is the s_1 -section of $f_s^{(1)} = df_s/du$.

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

- Notice that finite linear part in this representation of f is just parametric model in x_1, \dots, x_d and its interactions (e.g., for $d = 2$, x_1, x_2, x_1x_2). The remaining infinite linear part is linear combination in tensor products of first order splines with interior knot-points in $(0(s_1), 1(s_1)]$ while having knots at 0 for components x_j with $j \in s \setminus s_1$.
- The L_1 -norm of all coefficients $f_{s,s_1}^{(1)}(du)$ and $f_s^{(1)}(0)$ defines our first order sectional variation norm $\|f\|_{v,1}^*$.
- $D_M^{(1)}([0, 1]^d)$ is defined as all functions $f \in D^{(0)}([0, 1]^d)$ satisfying this first order smoothness with $\|f\|_{v,1}^* \leq M$.

(fine enough) Finite dimensional first order spline working model

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

- For each $s \subset \{1, \dots, d\}$, we can select knot-points $\mathcal{R}^1(s) \equiv \{(0(s/s_1), X_i(s_1)) : i = 1, \dots, n, s_1 \subset s\} \subset [0(s), 1(s)]$. The corresponding first order splines are:

$$\{\phi_{u(s)}^1 : u(s) \in \mathcal{R}^1(s)\}.$$

For s_1 the empty set this yields $\{0(s) : s \subset \{1, \dots, d\}\}$, giving the interactions $\phi_{0(s)}^1 = \prod_{j \in s} x_j$.

- The total set of N first order splines is thus;

$$\mathcal{R}_N^1 \equiv \{\phi_{u(s)}^1 : u(s) \in \mathcal{R}^1(s), s \subset \{1, \dots, d\}\}.$$

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

- The initial starting model of linear combinations of first order splines is then:

$$D^{(1)}(\mathcal{R}_N^1) = \left\{ \sum_{j \in \mathcal{R}_N^1} \beta(j) \phi_j : \beta \right\}.$$

- This working model $D^{(1)}(\mathcal{R}_N^1) \subset D^{(1)}([0, 1]^d)$ represents a close approximation of $D^{(1)}([0, 1]^d)$ (providing sup-norm approximations going as fast as $n^{-1/2}$).

First-order spline HAL

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

- We then define

$$\beta_n = \arg \min_{\beta, |\beta|_1 \leq C_n} P_n L(Q_{f_{N,\beta}}),$$

and first order HAL $Q_n = Q_{f_{N,\beta_n}}$.

Rate of convergence for first order HAL

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

- One can select J zero order splines $\mathcal{R}^0(J)$ so that $D^{(0)}(\mathcal{R}^0(J))$ yields a $O^+(1/J)$ L^2 -approximation of $D^{(0)}([0, 1]^d)$.
- This explains the dimension free rate of convergence $O^+(n^{-1/3})$ for the zero-order HAL.
- One can select J first order splines, $\mathcal{R}^1(J)$, so that $D^{(1)}(\mathcal{R}^1(J))$ yields a $O^+(1/J^2)$ **supremum norm** approximation of $D^{(1)}([0, 1]^d)$.
- The entropy integral $J(\delta, D_M^{(1)}([0, 1]^d), \|\cdot\|_\infty) = O^+(\delta^{3/4})$ instead of $O^+(\delta^{1/2})$ for $D_M^{(0)}([0, 1]^d)$.
- Our rate of convergence proof yields

$$d_0(Q_n, Q_0) = O_P^+(n^{-4/5}).$$

New Results on higher order HAL-MLE

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

- So first order HAL converges at rate $n^{-2/5}$ pointwise, ignoring $\log n$ -factors.
- At cost of another $\log n$ -factor this rate is uniform in x .
- Pointwise and simultaneous confidence intervals follow.
- Any of these HAL-estimators result in plug-in estimators of smooth target features that converge at $n^{-1/2}$ and are asymptotically normal, and either efficient and super-efficient (on set of measure zero!).
- Inference can be based on nonparametric bootstrap.

Efficient HAL-TMLE (vdL, Rubin, 2006)

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

- Construct initial estimator \mathbf{P}_n ; determine a least favorable path $\{\mathbf{P}_{n,\epsilon} : \epsilon \in (-\delta, \delta)\} \subset \mathcal{M}$ through \mathbf{P}_n with score $D_{\mathbf{P}_n}^*$ at $\epsilon = 0$.
- Compute MLE $\epsilon_n = \arg \max_\epsilon P_n L(\mathbf{P}_{n,\epsilon})$, where $L(P)$ is a valid loss function so that $\Psi(\arg \min_P P_0 L(P)) = \Psi(P_0)$.
- Let $\mathbf{P}_n^* = \mathbf{P}_{n,\epsilon_n}$. The TMLE is given by $\Psi(\mathbf{P}_n^*)$.
- By construction,
$$\Psi(\mathbf{P}_n^*) - \Psi(P_0) = (P_n - P_0) D_{\mathbf{P}_n^*}^* + R(\mathbf{P}_n^*, P_0),$$
 with
$$R(P, P_0) \equiv \Psi(P) - \Psi(P_0) - (P - P_0) D_P^*$$
 an exact second order remainder.
- Using the Highly Adaptive Lasso as initial estimator \mathbf{P}_n , we are guaranteed that $\Psi(\mathbf{P}_n^*)$ is asymptotically efficient estimator of $\Psi(P_0)$ (van der Laan, 2017).

HAL-TMLE of Target Estimands is Asymptotically Efficient

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

- TMLE is two-stage method for constructing plug-in efficient estimators $\Psi(P_n^*)$ of a pathwise differentiable target estimand $\Psi(P_0)$.
- The HAL-TMLE (using HAL as initial estimator) is efficient in great generality (vdL, 15).
- The only necessary model assumptions are:
 - The true nuisance parameters have finite sectional variation norm
 - The loss functions of the true nuisance parameters are uniformly bounded, so that oracle inequality applies
 - The strong positivity assumption holds

Example: Asymptotic efficiency of (zero-order) HAL-TMLE for treatment-specific mean / ATE

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

Consider the HAL-TMLE of $EY_1 = EE(Y | A = 1, W)$ based on $(W, A, Y) \sim P_0$ in a nonparametric statistical model. It is asymptotically efficient if

- ① $\delta < P_0(A = 1 | W)$ for some $\delta > 0$
- ② $W \rightarrow E_0(Y | A = 1, W)$ and $W \rightarrow P_0(A = 1 | W)$ are cadlag
- ③ $W \rightarrow E_0(Y | A = 1, W)$ and $W \rightarrow P_0(A = 1 | W)$ have finite sectional variation norm.

Undersmoothed HAL-MLE is efficient uniformly over large class of target estimands

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

- HAL-MLE is efficient for pathwise differentiable target estimands, if L_1 -norm is chosen large enough.
- Due to being an MLE, it solves a large class of score equations, in particular, efficient scores corresponding with target estimands.
- By undersmoothing enough, it uniformly solves a class of score equations that approximates all scores with finite variation norm. As a consequence, it is a globally efficient MLE across most pathwise differentiable features.
- This results can be applied to different assumed subspaces (i.e., additive models of form $D^{(0)}(\mathcal{R}(d))$) or $D^{(0)}([0, 1]^d)$.

Nonparametric Bootstrap of HAL-TMLE or HAL-MLE

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

- Fix M at the cross-validation selector M_n or another selector.
- Draw 10,000 samples of size n from empirical measure P_n . For each bootstrap sample $P_n^\#$, recompute the HAL-TMLE(M), say $\mathbf{P}_{n,M}^{\#*}$.
- The HAL on bootstrap sample can be restricted to only include indicator basis functions that were selected by HAL-MLE(M) on original data.
- Use sampling distribution of $\psi_{n,M}^{\#*} = \Psi(\mathbf{P}_{n,M}^{\#*})$, conditional on P_n , to construct 0.95-confidence interval.
- Increase M till plateau in confidence interval for optimal coverage.

Bootstrap works for HAL-TMLE or HAL-MLE

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

- Conditional on the data ($P_n : n \geq 1$), the bootstrap sampling distribution of ψ_n^* converges to optimal normal limit distribution $N(0, \sigma_0^2)$.
- The approximation error of bootstrap is driven by performance of nonparametric bootstrap for an empirical process indexed by Donsker class (i.e., cadlag functions with sectional variation norm bounded by M).
- This suggests robust finite sample behavior of the nonparametric bootstrap.

Adaptive TMLE

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

- Data adaptively learn a model $\mathcal{M}_n \subset \mathcal{M}$. Make sure that $d(P_0, \mathcal{M}_n) = o_P(n^{-1/4})$ (e.g., use HAL, or cross-validation selection among a large family of submodels).
- Define the projection parameter $\Psi_{\mathcal{M}_n} : \mathcal{M} \rightarrow \mathbb{R}$ defined by $\Psi_{\mathcal{M}_n}(P_0) = \Psi(\Pi_{\mathcal{M}_n} P_0)$, where $P_{0,n} \equiv \Pi_{\mathcal{M}_n}(P_0) = \arg \min_{P \in \mathcal{M}_n} P_0 L(P)$ is the log-likelihood (or loss based) projection of P_0 onto \mathcal{M}_n .
- Construct a TMLE $\Psi_{\mathcal{M}_n}(P_n^*)$ of $\Psi_{\mathcal{M}_n}(P_0)$ based on canonical gradient $D_{\mathcal{M}_n, P}^*$ of this projection parameter. Generally speaking, for log-likelihood loss functions $\Psi_{\mathcal{M}_n}(P_n^*) = \Psi(P_n^*)$ with $P_n^* \in \mathcal{M}_n$.
- Provide confidence intervals for $\Psi_{\mathcal{M}_n}(P_0)$ as usual based on $D_{\mathcal{M}_n, P_0}$.
- Possibly cross-fit, like CV-TMLE.

Adaptive TMLE of ATE

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

- Let $O = (W, A, Y)$; \mathcal{M} nonparametric;
 $\Psi(P) = E_P \tau_P(W)$;
 $\tau_P = E_P(Y | A = 1, W) - E_P(Y | A = 0, W)$.
- Consider loss function $L_{m_0, g_0}(\tau)$ indexed by nuisance parameters $m_0 = E_0(Y | W)$ and $\bar{g}_0 = P_0(A = 1 | W)$:

$$L_{m,g}(\tau) = (Y - m(W) - (A - g(1 | W))\tau)^2 .$$

- Double robust loss for CATE:** We have
 $\tau_0 = \arg \min_{\tau} P_0 L_{m,g}(\tau)$ if $m = m_0$ or $g = g_0$.

Adaptive TMLE of ATE

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

- Let $\sum_j \beta(j)\phi_j$ be a high dimensional linear combination of splines ϕ_j as in HAL. Define

$$\beta_n = \arg \min_{\beta, \|\beta\|_1 < C_n} P_n L_{m_n, g_n} \left(\sum_j \beta_j \phi_j \right).$$

Then, refit $\sum_{j, \beta_n(j) \neq 0} \beta_j \phi_j$ unpenalized (relax-HAL).

- Let $\tau_n = \sum_j \beta_n(j)\phi_j$. $\psi_n = P_n \tau_n$ is an adaptive TMLE where the learned working model \mathcal{M}_n for P_0 is the semiparametric regression model selected by HAL.
- It is super-efficient corresponding with oracle model \mathcal{M}_0 the limit of HAL-working model \mathcal{M}_n (**or just efficient if $\mathcal{M}_0 = \mathcal{M}$**).

Conclusion

Integrating RCT with Observational Data

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

- Suppose we observe data on two studies indicated by $S \in \{0, 1\}$: (S_i, W_i, A_i, Y_i) , $i = 1, \dots, n$.
- In study $S = 1$, treatment is randomized (RCT), so that the ATE can be robustly estimated.
- Study $S = 0$ is an external RWD study.
- We make no other assumptions on likelihood beyond possible knowledge of $g(A | S, W)$.
- However, the RCT is underpowered due to small control arm or small overall sample size.
- Therefore, a TMLE of the ATE based on $S = 1$ only would have large confidence intervals lacking power at reasonable alternatives.
- Can we utilize the external study $S = 0$ from the real world to obtain a more efficient estimator of the ATE? Without adding assumptions!

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

- Dang et al (2023) developed an ES-CV-TMLE. Rejects RWD if biased relative to RCT.
- Here, we develop an Adaptive TMLE that learns bias in RWD and corrects pooled estimator.

Motivating example: oral semaglutide on cardiovascular outcomes (Dang et al. 2023)

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

- *Causal question:* **risk of MACE within one year** if all patients in the target population were prescribed **oral semaglutide** plus standard-of-care compared to if all patients were prescribed **standard-of-care** alone?
- *Set up:* a single **RCT** (PIONEER 6) vs. a **hybrid** randomized-external data study.
- *Results:*
 - Under the single RCT:
-1.30% (95% CI -2.60%, 0.00%).
 - Under the hybrid design using ES-CVTMLE:
-1.53% (95% CI -2.75%, -0.30%).
- *Conclusion:* importance of the **Causal Roadmap**, using **RWD** to inform regulatory approval process, **power gain** of test for superiority, **reduced participation-time** without initiation of a GLP1-RA.

Causal and statistical estimand

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

- RCT-ATE $\Psi_0^F = E_P E_P(Y_1 - Y_0 \mid S = 1, W)$.
- Due to $S = 1$ being an RCT, we have Ψ_0^F equals
$$\psi_0 = E_0(E_0(Y \mid S = 1, W, A = 1) - E_0(Y \mid S = 1, W, A = 0)).$$
- Note in outer expectation, we take the average over the **pooled covariate distribution**.
- *Oral semaglutide case study:* What would the risk difference be **if all patients in the target population were enrolled** in the trial? The target population should **not** contain individuals with 0 probability of being enrolled in the trial. Factors to consider: inclusion/exclusion criteria of PIONEER 6, time-frame, patient characteristics, healthcare engagement, etc. (Dang et al. 2023)

Decomposition of the target estimand as difference pooled-ATE and bias term (Dang et al. 2023)

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

- 1st component, **pooled-ATE** estimand:

$$\tilde{\Psi}(P) = E_P \tau_P(W).$$

- 2nd component, **bias** estimand: $\Psi^{\#}(P) = \tilde{\Psi}(P) - \Psi(P)$:

$$\Psi^{\#}(P) = E_P[\Pi(0 | W, 0)\tau_P^S(W, 0) - \Pi(0 | W, 1)\tau_P^S(W, 1)],$$

where

$$\Pi(s | W, A) = P(S = s | W, A),$$

and

$$\tau_P^S(W, A) = E_P(Y | S = 1, W, A) - E_P(Y | S = 0, W, A).$$

- Thus

$$\Psi(P_0) = \tilde{\Psi}(P_0) - \Psi^{\#}(P_0) \text{ bias correction.}$$

Adaptive-TMLE for bias estimand $\Psi^\#$

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

- For pooled ATE $\tilde{\Psi}(P_0)$, we could either use a regular TMLE of pooled ATE or the above A-TMLE.
- For the bias estimand $\Psi^\#(P_0)$, analogue to A-TMLE for ATE, we will use relax-HAL to learn a **working submodel for the conditional effect τ_0^S of the study indicator S** on the outcome Y (implying a submodel \mathcal{M}_n).
- The corresponding A-TMLE of bias estimand also includes targeting an initial estimator Π_n of $\Pi_0 = P_0(S | W, A)$, and one estimates the expectation over W with the empirical mean.

Benefits of using A-TMLE include

- Nominal **type I error control** without extra assumptions.
- Not only estimates the magnitude of the bias, also takes advantage of the **learnable structure of the bias**. Allows **full utilization** of both RCT and RWD data, more gain in efficiency.
- The bias working model **mitigates large variances** due to inverse-weighting, finite sample robust.

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

Simulation 1: simple bias

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

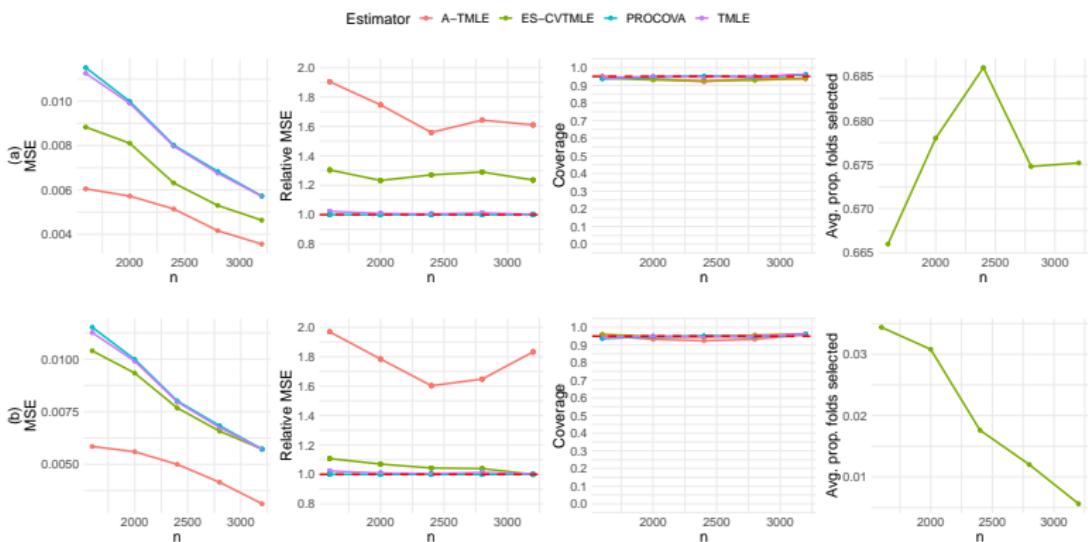
Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion



Simulation 2: complex bias

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

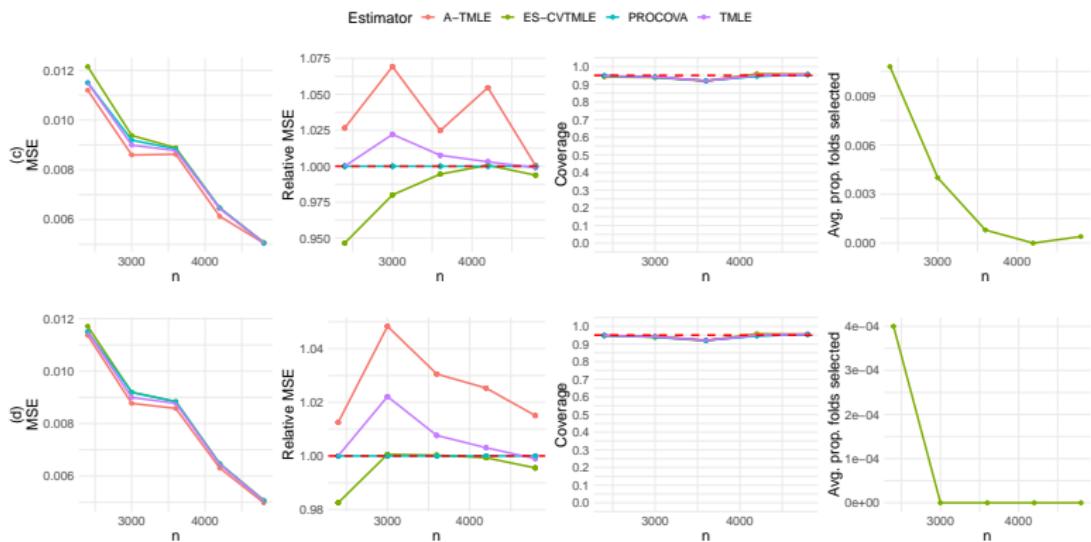
Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion



95% CI width comparisons

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

Table: A-TMLE's 95% CI width as a percentage of other methods'.

Method	Scenario (a)	Scenario (b)	Scenario (c)	Scenario (d)
ES-CVTMLE	66.1%	60.4%	98.3%	98.2%
Regular TMLE	59.1%	61.2%	99.1%	99.1%
PROCVA	59.0%	61.1%	99.0%	99.0%
RCT-only	59.0%	61.1%	99.0%	99.0%

General Longitudinal Data Structure

Targeted
Learning

Mark van der
Laan and
Nima Hejazi

The Art of
Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic
statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

We observe n i.i.d. copies of a longitudinal data structure

$$O = (L(0), A(0), \dots, L(K), A(K), Y = L(K + 1)),$$

where $A(t)$ denotes a discrete valued intervention node, $L(t)$ is an intermediate covariate realized after $A(t - 1)$ and before $A(t)$, $t = 0, \dots, K$, and Y is a final outcome of interest.

Survival example: For example,

$$A(t) = (A_1(t), A_2(t))$$

$$A_1(t) = I(\text{Treated at time } t)$$

$$A_2(t) = I(\min(T, C) \leq t, \Delta = 0) \text{ right-censoring indicator process}$$

$$\Delta = I(T \leq C) \text{ failure indicator}$$

$$Y(t) = I(\min(T, C) \leq t, \Delta = 1) \text{ survival indicator process}$$

$$Y(t) \subset L(t) \quad Y = Y(K + 1).$$

Likelihood and Statistical Model

Targeted
Learning

Mark van der
Laan and
Nima Hejazi

The Art of
Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic
statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

The probability distribution P_0 of O can be factorized according to the time-ordering as

$$\begin{aligned} p_0(O) &= \prod_{t=0}^{K+1} p_0(L(t) \mid Pa(L(t))) \prod_{t=0}^K p_0(A(t) \mid Pa(A(t))) \\ &\equiv \prod_{t=0}^{K+1} q_{0,L(t)}(O) \prod_{t=0}^K g_{0,A(t)}(O) \\ &\equiv q_0 g_0, \end{aligned}$$

where $Pa(L(t)) \equiv (\bar{L}(t-1), \bar{A}(t-1))$ and $Pa(A(t)) \equiv (\bar{L}(t), \bar{A}(t-1))$ denote the parents of $L(t)$ and $A(t)$ in the time-ordered sequence, respectively. The g_0 -factor represents the intervention mechanism.

Statistical Model: We make no assumptions on q_0 , but could make assumptions on g_0 .

Statistical Target Parameter: G -computation Formula for Post-dynamic-Intervention Distribution

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

- $p_0^{g^*} = q_0(o)g^*(o)$ is the G -computation formula for the post-intervention distribution of O under the stochastic intervention $g^* = \prod_{t=0}^K g_{A(t)}^*(O)$.
- In particular, for a dynamic intervention $d = (d_t : t = 0, \dots, K)$ with $d_t(\bar{L}(t), \bar{A}(t-1))$ being the treatment at time t , the G -computation formula is given by

$$p_0^d(I) = \prod_{t=0}^{K+1} q_{0,L(t)}^d(\bar{I}(t)), \quad (3)$$

where

$$q_{L(t)}^d(\bar{I}(t)) = q_{L(t)}(I(t) \mid \bar{I}(t-1), \bar{A}(t-1) = \bar{d}_{t-1}(\bar{I}(t-1))).$$

- Let $L^d = (L(0), L^d(1), \dots, Y^d = L^d(K+1))$ denote the random variable with probability distribution P^d .

A Sequential Regression G-computation Formula (Bang, Robins, 2005)

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

- By the iterative conditional expectation rule (tower rule), we have

$$E_{P^d} Y^d = E \dots E(E(Y^d | \bar{L}^d(K)) | L^d(K-1)) \dots | L(0)).$$

- In addition, the conditional expectation, given $\bar{L}^d(K)$ is equivalent with conditioning on $\bar{L}(K), \bar{A}(K-1) = \bar{d}_{K-1}(\bar{L}(K-1))$.

In this manner, one can represent $E_{P^d} Y^d$ as an iterative conditional expectation, first take conditional expectation, given $\bar{L}^d(K)$ (equivalent with $\bar{L}(K), \bar{A}(K-1)$), then take the conditional expectation, given $\bar{L}^d(K-1)$ (equivalent with $\bar{L}(K-1), \bar{A}(K-2)$), and so on, until the conditional expectation given $L(0)$, and finally take the mean over $L(0)$.

Conclusion

TMLE

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

- A likelihood based TMLE was developed (van der Laan, Stitelman, 2010).
- A sequential regression TMLE $\Psi(Q_n^*)$ was developed for EY_d in van der Laan, Gruber (2012).
- The latter builds on Bang and Robins (2005) by putting their innovative double robust efficient estimating equation method, which uses sequential clever covariate regressions to estimate the nuisance parameters of estimating equation, into a TMLE framework.
- A TMLE for Euclidean summary measures of $(EY_d : d \in \mathcal{D})$ defined by marginal structural working models is developed in Petersen et al. (2013);
- A new (analogue to sequential regression) TMLE allowing for *continuous valued monitoring and time till event* is coming (Rijtgaard, van der Laan, 2019).

A real-world CER study comparing different rules for treatment intensification for diabetes

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

- Data extracted from diabetes registries of 7 HMO research network sites:
 - Kaiser Permanente
 - Group Health Cooperative
 - HealthPartners
- **Enrollment period:** Jan 1st 2001 to Jun 30th 2009

Enrollment criteria:

- past A1c < 7% (glucose level) while on 2+ oral agents or basal insulin
- $7\% \leq \text{latest A1c} \leq 8.5\%$ (study entry when glycemia was no longer reined in)

Longitudinal data

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

- **Follow-up** til the earliest of Jun 30th 2010, death, health plan disenrollment, or the failure date
- **Failure** defined as onset/progression of albuminuria (a microvascular complication)
- **Treatment** is the indicator being on "treatment intensification" (TI)
- $n \approx 51,000$ with a median follow-up of 2.5 years.
- **Target estimand:** What would survival look like if treatment is intensified when $A1c < x\%$ for various levels $x = 7, 7.5, 8, 8.5?$

Better clinical decisions from observational data

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

Statistics
in Medicine

Research Article

Received 24 May 2013,

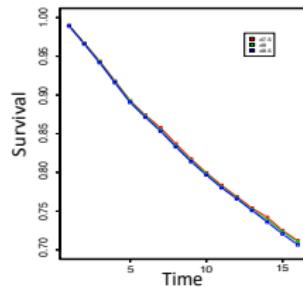
Accepted 5 January 2014

Published online 17 February 2014 in Wiley Online Library

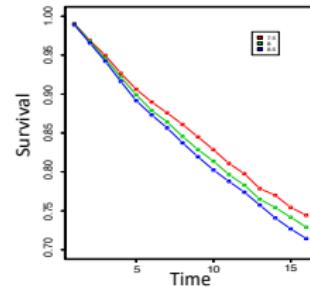
(wileyonlinelibrary.com) DOI: 10.1002/sim.6099

Targeted learning in real-world comparative effectiveness research with time-varying interventions

Romain Neugebauer,^{a*}[†] Julie A. Schmittield^a and Mark J. van der Laan^b



Standard methods: No benefit to more aggressive intensification strategy



Targeted Learning: More aggressive intensification protocols result in better outcomes

Deep LTMLE: Toru Shirakawa, Yi Li, Sky Qiu, vdL

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

- LTMLE (Bang, Robins, 2005, Petersen et al., Gruber, vdL, etc, relies on **sequential regression**).
- Targeted **maximum likelihood** requires estimation of all conditional densities (vdL 2010a,b; Stitelman et al (2011)).
- Rytgaard, Gerdts, vdL (2023) **combines** TM-likelihood based estimation of **intensities** with estimation of a **conditional mean function integrating over time dependent covariates**.
- Toru et al. Deep LTMLE utilize **transformer architecture for temporal difference learning**, simultaneously modeling and estimating the sequential regressions:
 - 1) *Computationally superior*; 2) continuous time monitoring; 3) large histories; 4) TMLE update step uses transformer gradient descent algorithm.

Predicting risk for COVID-19 infection, hospitalization or death

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

- For any subject in study, define $t = 0$ as January 2022: one might consider other $t = 0$ times individualized by an enrollment criterion).
- $K + 1$ is **number of months of follow up** after $t = 0$ at which we want the outcome Y of interest.
- $L(0)$ represents **baseline history** for this subject, including any past events, medical history etc.
- Define as **intervention nodes** $A(t)$ indicator of being censored by time t (e.g. death by other causes, end of study).
- Define **outcome** Y as the indicator of observing a COVID-19 infection by time $t = K + 1$.
- Define **time-dependent covariates** $L(t)$ as anything observed in month t , before $A(t)$: including COVID-19 Infection/Hospitalization/Death, vaccination, medical treatments etc.

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

- Our target is $E(Y_{\bar{a}=0} \mid L(0))$: **conditional risk of COVID-19 infection by time $K + 1$ in counterfactual world without right-censoring.**
- So our goal is to predict risk of COVID-19 event by $K + 1$ -months in world without right-censoring.
- One may look at various **other outcomes** of interest such as hospitalization, death, composite outcomes etc.

Super learner to predict COVID-19 risk from baseline information

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

- One can set up a collection of logistic regression machine learning algorithms, including a variety of HAL-model fits.
- One has to select a **measure of performance**, such as AUC, log-likelihood, or MSE.
- To deal with drop-out (right-censoring), we can use **inverse probability of censoring weights** for each candidate algorithm and for evaluating the cross-validated performance of the algorithm.
- The super-learner will output the best performing prediction function, its measure of performance (e.g., AUC) with a confidence interval.

Variable importance analysis using TMLE

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

- Let $W = L(0)$.
- One has to **define a measure of importance** for each variable W_j .
- For example, for **binary** W_j , one can use the **ATE** estimand, which measures change in prediction due to $W_j = 0$ versus $W_j = 1$ keeping $W(-j)$ constant.
- For **continuous** W_j one can use the **shift-intervention** estimand, which measures change in prediction due to shifting W_j to $W_j + \delta$, keeping $W(-j)$ constant.
- One can compute TMLE of these W_j -specific variable importance measures, with confidence intervals and p -values.
- One can use **multiple testing adjustments** to control family wise error or false discovery rate.

Targeted Learning

Mark van der Laan and Nima Hejazi



The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

The Roadmap in Action: Case Studies & Ongoing Projects

Berkeley
UNIVERSITY OF CALIFORNIA

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

1

FDA Sentinel Innovation Center



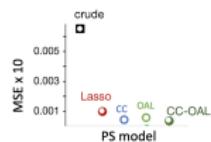
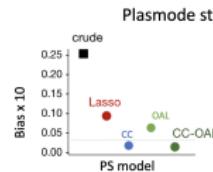
- Safety evaluation with high dimensional data:**

Wyss et al. (2024), AJE, Targeted Learning with an Undersmoothed Lasso Propensity Score Model for Large Scale Covariate Adjustment in Healthcare Database Studies.

- Subset calibration/two-stage designs:**

-Ongoing project evaluating methods such as the two-stage design TMLE for study designs that involve a subset of subjects with carefully curated confounders and or outcomes, and a remaining set of subjects.

-This is a common type of design to obtain desired causal identification from RWD while still gaining efficiency from the less curated data set.



Collaborative control greatly reduced bias and improved MSE

- Less regularization captured more relevant confounder information in PS

These projects involve multi-author working groups with FDA/Pharma/Academics/Kaiser Permanente.

The Sentinel Innovation Center is funded by the FDA through the Department of Health and Human Services (HHS) Task order 75F40119D10037.

Berkeley
UNIVERSITY OF CALIFORNIA

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Comparative Effectiveness: Targeted Learning FDA Demonstration Project

Resulted in various
collaborative relations with
FDA statisticians

Berkeley
UNIVERSITY OF CALIFORNIA

TL for RWE multi-year FDA-funded project

TL on YouTube



Publications with FDA co-authors

Targeted learning: Towards a future informed by real-world evidence

Susan Gruber, Rachael V. Phillips, Hana Lee, Martin Ho, John Concato, Mark J. van der Laan

Published in the *Journal of Clinical Epidemiology* in 2017

Abstract: The goal of improving real-world evidence with Targeted Learning

Susan Gruber, Rachael V. Phillips, Hana Lee, John Concato, Mark van der Laan

Published in the *Journal of Clinical Epidemiology* in 2017

Developing a Targeted Learning-Based Statistical Analysis Plan

Susan Gruber, Rachael V. Phillips, Hana Lee, John Concato, Mark van der Laan

Published in the *Journal of Clinical Epidemiology* in 2017

Over 250 FDA short course attendees

SHORT COURSE ANNOUNCEMENT



A Targeted Learning Framework for Causal Effect Estimation using Real-World Data

Funded by FDABAA-19-00123-A3

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

Comparative Effectiveness: GLP1-RA and Dementia in the Danish National Registry

Population: Insulin-naïve T2-DM initiating 2nd-line therapy (N=104,928)

Causal Estimand: Difference in 5 year incident dementia under hypothetical protocols:

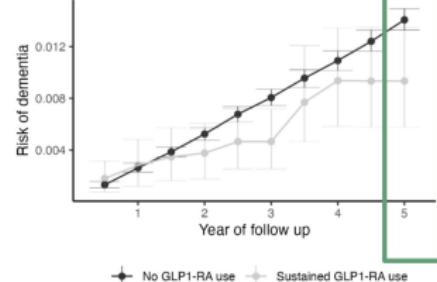
- continuous GLP1-RA vs. no- GLP1-RA use
- no censoring

Identification: Baseline & post-baseline confounder adjustment

Estimation & Inference: Longitudinal TMLE w/ machine learning; simulation-based specification

Nance et al, 2023, [arXiv:2310.03235](https://arxiv.org/abs/2310.03235)

5-yr RD (95% CI): -0.005(-0.009,-0.001)



Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

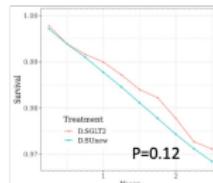
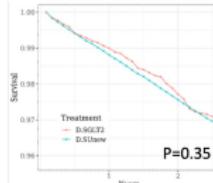
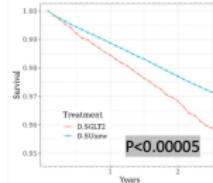
Variable Importance

Conclusion

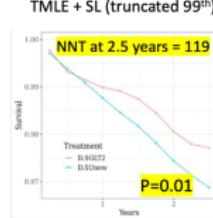
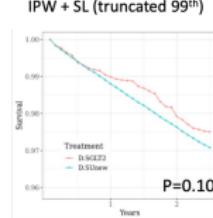
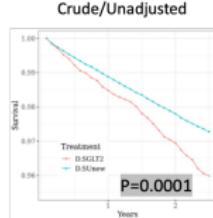
Causal Effect of SLGT2/SU on MACE, controlling for >100 baseline confounders, and time-dependent confounders of drop-out



ITT



PP



Berkeley
UNIVERSITY OF CALIFORNIA

Preliminary results from the ON TARGET DM Study (PCORI #DB-2020C2-20318, Neugebauer & O'Connor)

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

RCT analysis: Improved Precision



THE NEW ENGLAND JOURNAL OF MEDICINE

ORIGINAL ARTICLE

SEARCH Study (NCT:0186460)

- Pragmatic Cluster RCT
 - N=320,000, 16 communities
- TMLE & "Adaptive pre-specification"
 - Flexible pre-specified adjustment
- Primary pre-specified analysis in trial
- Efficiency gains
 - Substantial for some outcomes
 - Never harmful



HIV Testing and Treatment with the Use of a Community Health Approach in Rural Africa

D.V. Havlir, L.B. Balzer, E.D. Charlebois, T.D. Clark, D. Kwarisima, J. Ayieko,

Focusing Stage 2 estimation:	Effect	95%CI	Efficiency*
Three-year cumulative HIV Incidence			
Unadjusted	0.98	(0.66, 1.45)	-
TMLE with Adaptive Prespec.	0.96	(0.80, 1.17)	4.6
Incidence of HIV-associated TB or death			
Unadjusted	0.79	(0.64, 0.98)	-
TMLE with Adaptive Prespec.	0.80	(0.69, 0.91)	2.6
Population-level Viral Suppression			
Unadjusted	1.15	(1.11, 1.20)	-
TMLE with Adaptive Prespec.	1.15	(1.11, 1.20)	1.0

*relative to unadjusted

1



e.g., Balzer, et al Stat Med
2016; Biometrics 2024



Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

RCT analysis: Intercurrent Events/Drop-in

1. LEADER Cardiovascular outcomes trial: RCT comparing Liraglutide vs. SoC for treating diabetes patients at high cardiovascular risk
2. The study showed a significant Cox-PH RH with CI [0.78-0.97] on MACE time to event outcome
3. TMLE reproduced similar effect, with more precision; robust across many subgroups (Chen et al., 2023)
4. There was significant additional drop-in insulin use in the control arm.
5. An L-TMLE stochastic intervention direct effect analysis controlling for the differential post-treatment use of insulin was carried out

Results:

- Significant stochastically controlled additive direct treatment effect
- Non-significant statically controlled additive direct treatment effect

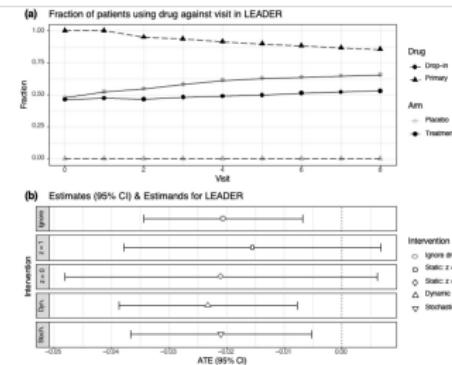


FIGURE 4 (a) Fraction of those at risk using main/drop-in treatment in each arm for LEADER and (b) LTMLE estimates with 95% confidence intervals.



Hypothetical treatment interventions to handle treatment drop-in in randomized controlled trials

HELENE C. W. RYTGAARD*, EDWIN FONG, JENS M. TARP,

THOMAS A. GERIDS, SOREN RASMUSSEN, MARK J. VAN DER LAAN, HENRIK RAVN

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

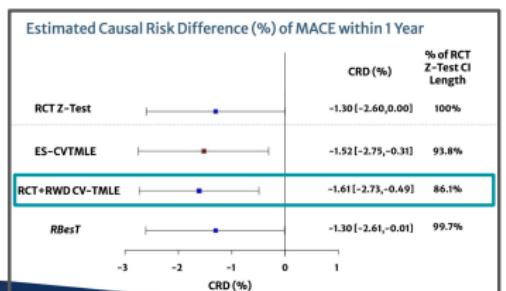
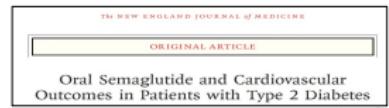
Longitudinal TMLE

Variable Importance

RCTs Augmented with External Controls Ex. Semaglutide and Major Adverse Cardiovascular Event (MACE)



- **Objective:** Potentially augment control arm while controlling Type 1 error
 - Pioneer 6 RCT
 - Optum observational data
- **Experiment-Selector CV-TMLE**
 - Select experiment (RCT only or RCT with RWD) that optimizes bias-variance tradeoff
 - Separate experiment-selection from effect estimation using cross-validation
 - Integrates negative control outcomes



Dang, et al, 2022; arXiv:2210.05802

Dang, et al, 2023, J. Clin & Trans. Science

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

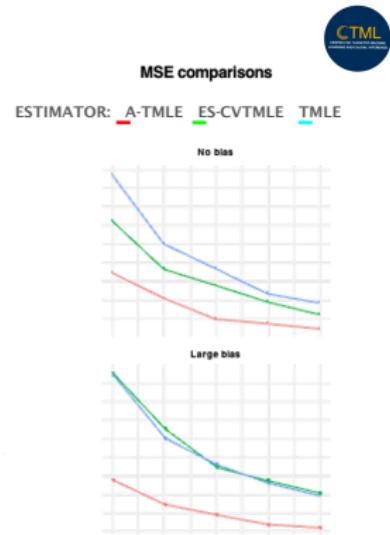
Longitudinal TMLE

Variable Importance

Conclusion

External Controls: Adaptive-TMLE

- Dang et al (2023): ES-CV-TMLE that data adaptively decides to **accept or reject** external data based on evaluating bias relative to gain in variance.
- Adaptive-TMLE is an advance in TMLE.
- It yields a new method for integrating external controls that **estimates the bias** from external data and **corrects** the pooled TMLE of ATE.
- The method is shown to be asymptotically **valid coverage** and type-I error **without any assumptions beyond RCT**.
- Advantage is that it yields a way forward with biased external control data while preserving valid inference.



Adaptive debiased machine learning
using data-driven model selection techniques

Lars van der Laan, Marco Carone, Alex Luedtke, Mark van der Laan

July 25, 2023

arXiv:2307.12544

Concluding Remarks

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

- Roadmap for causal inference and Targeted Learning provides systematic principled approach for generating RWE.
- Integrates all advances in machine learning, statistical theory and causal identification.
- SL and TMLE can be tailored towards particular estimation problem in pre-specified manner using outcome blind simulations.
- HAL provides realistic models; theoretical guarantees, dimension free rates, and bridges TMLE to inference for non-pathwise differentiable target functions such as conditional treatment effects, dose response curves etc.
- Bridges such as Deep LTMLE to deep learning community are crucial.

Active TL collaborations and software

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

- The Center for Targeted Machine Learning (CTML) at UC Berkeley: <https://ctml.berkeley.edu>
- Government: US Food and Drug Administration (FDA) and National Institute of Allergy and Infectious Diseases (NIAID), California Department of Public Health
- Private sector: Genentech, Gilead, Kaiser Permanente, Accenture, Novo Nordisk
- Software:
 - The `tlverse` ecosystem: <https://github.com/tlverse>
 - Additional packages: `tmle`, `ltmle`, `lmtp`, `SuperLearner`

Thank you!

Targeted Learning

Mark van der Laan and Nima Hejazi

The Art of Statistics

TL in Action

TL Roadmap

Describe study

Specify a realistic statistical model

Define estimand

Causal estimand

Statistical estimand

Construct estimator

Obtain inference

Sensitivity analysis

Advanced TL

Collaborative TMLE

HAL and A-TMLE

Longitudinal TMLE

Variable Importance

Conclusion

Feel free to email us to ask questions, request learning resources, or get involved in TL research and/or CTML.

Mark's email: laan@berkeley.edu

Nima's email: nhejazi@hsph.harvard.edu

Targeted Learning

Mark van der
Laan and
Nima Hejazi

Appendix

Higher order spline HAL

- Analogue as above, we can define $D_M^{(k)}([0, 1]^d) \subset D^{(k-1)}([0, 1]^d)$ as class of k -th order smooth functions with k -th order sectional variation norm bounded by M .
- Analogue we obtain a k -th order spline representation for $f \in D^{(k)}([0, 1]^d)$; a finite dimensional linear working model $D^{(k)}(\mathcal{R}_N^k)$ approximating $D^{(k)}([0, 1]^d)$ and a corresponding k -th order spline HAL-MLE:

$$f_n = \arg \min_{f \in D_{M_n}^{(k)}(\mathcal{R}_N^k)} P_n L(Q_f),$$

and $Q_n = Q_{f_n}$.

- With J well chosen k -th order spline basis functions we can obtain an $O^+(1/J^{k+1})$ sup-norm approximation of $D^{(k)}([0, 1]^d)$.

- As a consequence, we now have
 $J(\delta, D_M^{(k)}([0, 1]^d), \|\cdot\|_\infty) = O^+(\delta^{(2k+1)/(2k+2)}).$
- Our rate of convergence proof now yields:

$$d_0(Q_n^k, Q_0) = O_P^+(n^{-2k^*/(2k^*+1)}),$$

with $k^* = k + 1$.

- For example, for $k = 0, 1, 2$, we have the dimension free rates $O_P^+(n^{-1/3})$, $O_P^+(n^{-2/5})$ and $O_P^+(n^{-3/7})$, respectively.

Discrete Super Learner incorporating higher order HAL

Targeted
Learning

Mark van der
Laan and
Nima Hejazi

- By varying smoothness degree k and starting sets $\mathcal{R}^k(j)$ of basis functions, one can define many HAL-MLEs over $D_C^{(k)}(\mathcal{R}^k(j))$.
- The discrete super learner using this library of (k, j) -specific HAL-MLEs will perform asymptotically exact as well as the oracle choice among all these HAL estimators, thereby achieves the rate of convergence of HAL for the unknown smoothness k_0 and smallest subspace $D^{(k_0)}(\mathcal{R}_{j_0}(d))$ containing true Q_0 .
- Thus this cross-validated higher order HAL-MLE is minimax **smoothness adaptive** achieving the minimax smoothness adaptive rate of convergence for univariate function estimation, up till $\log n$ -factors.

Asymptotic normality of higher order HAL

- Let \mathcal{R}_n be a set of J_n k -th order splines providing uniform approximation error $O^+(1/J_n^{k+1})$.
- The non-zero coefficients in the HAL-fit does this for us when we choose a fine enough starting model.
- In addition, HAL will select an adaptive selection that works best for Q_0 .
- Let $D^{(k)}(\mathcal{R}_n)$ be the linear working model. This set yields the $O(1/J_n^{k+1})$ -uniform approximation of Q_0 .
- The HAL-MLE $Q_n = \sum_{j \in \mathcal{R}_n} \beta_n(j) \phi_j$ operates as an MLE of the oracle approximation $Q_{0,n} = \sum_{j \in \mathcal{R}_n} \beta_{0,n}(j) \phi_j$ in this working model.

- In particular, if we do the relax-HAL (refitting the selected working model without L_1 -penalty), then it is an exact MLE of $Q_{0,n}$.
- One can analyze this parametric MLE Q_n^k w.r.t. $Q_{0,n}$ to establish that $(J_n/n)^{1/2}(Q_n - Q_{0,n})(x) \Rightarrow_d N(0, \sigma_0^2(x))$, while, by our uniform approximation result $\|Q_{0,n} - Q_0\|_\infty \sim O^+(1/J_n^{k+1})$.

Asymptotic normality and uniform rates

Targeted
Learning

Mark van der
Laan and
Nima Hejazi

- By selecting $J_n \sim n^{-1/(k+2)}$, the pointwise rate equals $n^{-k^*/(2k^*+1)}$ up till $\log n$ -factors.
- At cost of another $\log n$ -factor this rate is uniform in x .
- Pointwise and uniform confidence intervals follow.
- Beyond inference for Q_0 , these results teach us that we have dimension free **uniform** rates of convergence for $\|Q_n - Q_0\|_\infty = O^+(n^{-k^*/(2k^*+1)})$.

New Results: Asymptotic normality of higher order HAL-MLE itself, van der Laan, 2023

Targeted Learning

Mark van der Laan and Nima Hejazi

- Defined higher order smoothness classes $D_C^{(k)}([0, 1]^d)$ with global complexity measure C , $k = 0, \dots$
- We define MLE over this model (k -th order HAL).
- $D_C^{(k)}([0, 1]^d)$ can be represented as linear span of tensor products of $\leq k$ -th order spline basis functions, where L_1 -norm is bounded by C .
- So implementation can be carried out with `glmnet()`.
- With J basis functions, one obtains uniform approximation error $O(1/J^{k^*})$ up till log J -factor, $k^* = k + 1$.
- HAL-MLE $Q_n = \sum_{j \in \mathcal{R}_n} \beta_n(j) \phi_j$ with J_n non-zero coefficients of its oracle MLE $Q_{0,n} = \sum_{j \in \mathcal{R}_n} \beta_{0,n}(j) \phi_j$ satisfies $(J_n/n)^{1/2}(Q_n - Q_{0,n})(x) \Rightarrow_d N(0, \sigma_0^2(x))$, while $\|Q_{0,n} - Q_0\|_\infty \sim O(1/J_n^{k^*})$.
- By selecting $J_n \sim n^{1/(k^*+1)}$, rate equals $n^{-k^*/(2k^*+1)}$ up till log n -factors. HAL will do this for you.

Super-efficient estimation for smooth features of target function

Targeted
Learning

Mark van der
Laan and
Nima Hejazi

- The smoothness adaptive HAL (using cross-validation to select k and additive submodel $D^{(k)}(\mathcal{R}_{j_0})$) will asymptotically act like an HAL for an oracle smoothness k_0 and oracle subspace $D^{(k_0)}(\mathcal{R}_{j_0}(d))$.
- As a consequence, by undersmoothing within the oracle subspace it will be plug-in efficient for smooth target features w.r.t. the oracle statistical model on the data distribution that assumes $Q_0 \in D^{(k_0)}(\mathcal{R}_{j_0}(d))$.
- That is, $\Psi(Q_n)$ is asymptotically super-efficient, and is an example of the Adaptive MLE of (Lars van der Laan et al., 2023).

Finite sample robust TMLE

- The current literature on TMLE has proposed various modifications of TMLE that regularize the TMLE to be better behaved in finite samples when the support is limited (i.e., practical violation of positivity assumption).
- For example, in censored and causal inference literature: adaptive truncation; regularize TMLE step by not fully solving EIC-equation; collaborative TMLE; outcome-adaptive TMLE; super-efficient TMLE (adjusting in PS for outcome regressions $Q_n(1, W)$, $Q_n(0, W)$, beyond a possible baseline model).
- These proposed regularizations have in common that they all concern targeted estimation of the orthogonal nuisance function $g(P_0)$ that is needed in targeting step, while $\Psi(P_0) = \Psi_1(Q_0)$ only depends on certain factors of likelihood.

- Typically, these variations preserve asymptotic efficiency but adapt the targeting step towards the data to carefully trade-off bias reduction with variance gain.
- Finite sample simulations, theoretical results such as collaborative double robustness etc, have shown that these regularizations are crucial for robust finite sample behavior for poorly supported parameters.
- However, we never developed a truly unifying approach (and super-efficiency theory)!
- Fortunately, Lars did: Lars van der Laan et al. (2023), Adaptive debiased machine learning.

Limitations of Efficient Estimators such as TMLE

Targeted Learning

Mark van der Laan and
Nima Hejazi

- Due to lack of nonparametric support, if \mathcal{M} is close to nonparametric, then $D_{P_0}^*(o)$ can have very large values.
- In that case, confidence intervals for standard TMLE will be wide and no significant finding can be obtained.
- However if we would use plug-in HAL $\Psi(\mathbf{P}_n)$ (possibly super-efficient) we might very well still find a significant true result. Somehow, the targeting step can blow up a good initial estimator.
- This is due to an efficient estimator to be regular along all paths through P_0 , including paths that appear to be contradicting the data.

An efficient estimator cannot adapt to structure in the true data distribution

- For example, suppose $O = (W, A, Y)$ and $\Psi(P) = E_P E_P(Y | A = 1, W)$. A plug-in HAL might end up fitting an additive model for the regression $E(Y | A, W)$ that contains the true regression.
- An efficient estimator of $\Psi(P_0)$ for this additive model would be well supported and have reasonable variance.
- But an efficient estimator for a nonparametric model protects itself against any kind of fluctuation of the data distribution, and as a consequence carries out a bias reduction that (**asymptotically**) holds up uniform in P in a ball around P_0 .
- In other words, it needs to remain asymptotically unbiased under fluctuations adding a 100 way interaction.

Lets reset benchmarks for our estimator

- **Super-efficient estimator:** Suppose that we require that our estimator of $\Psi(P_0)$ is asymptotically linear at any $P_0 \in \mathcal{M}$, and regular along paths through P_0 that stay in an oracle model $\mathcal{M}_0 \subset \mathcal{M}$, approximated by a data adaptive submodel \mathcal{M}_n satisfying $P_0 \in \mathcal{M}_0$.
- Then, we can construct \mathcal{M}_0 -super-efficient estimators that still provide asymptotically valid confidence interval and are still robust under perturbations of P_0 that stay in the model \mathcal{M}_0 .
- **Regularized efficient estimator:** If our data adaptive model \mathcal{M}_n approximates $\mathcal{M}_0 = \mathcal{M}$, let the estimator behave as an efficient estimator under model \mathcal{M}_n that approximates the a-priori specified model \mathcal{M} as sample size converges to infinity but in a way that carefully balances finite sample bias and variance.

Why does A-TMLE work: 1) standard TMLE analysis

- From TMLE analysis we will have that $\Psi_{\mathcal{M}_n}(P_n^*) - \Psi_{\mathcal{M}_n}(P_0)$ behaves as $(P_n - P_0)D_{\mathcal{M}_n, P_0}^*$. Therefore, under an asymptotic stability condition on \mathcal{M}_n so that $D_{\mathcal{M}_n, P_0}^* \rightarrow_p D_{\mathcal{M}_0, P_0}^*$ we have that it behaves as $P_n D_{\mathcal{M}_0, P_0}^*$ and is thus asymptotically normal with mean zero and variance $\sigma_{\mathcal{M}_0}(P_0) = P_0\{D_{\mathcal{M}_0, P_0}^*\}^2$.
- Cross-fitting weakens this need for asymptotic stability. Moreover, one might still have asymptotic normality by standardizing by a variance estimator.
- $D_{\mathcal{M}_0, P_0}^*$ equals the efficient influence curve of $\Psi : \mathcal{M}_0 \rightarrow \mathbb{R}$: i.e. we achieve the efficiency we would achieve with TMLE if we would a priori know that $P_0 \in \mathcal{M}_0$.

Why does A-TMLE work: 2) data adaptive model bias negligible

Targeted Learning

Mark van der Laan and Nima Hejazi

- Let $R_{\mathcal{M}_0}(P, P_0) = \Psi_{\mathcal{M}_0}(P) - \Psi_{\mathcal{M}_0}(P_0) + P_0 D_{\mathcal{M}_0, P}^*$.
- We have

$$\begin{aligned}\Psi_{\mathcal{M}_n}(P_0) - \Psi(P_0) = \\ (P_{0,n} - P_0) \{ D_{\mathcal{M}_0, P_{0,n}}^* - \Pi_n(D_{\mathcal{M}_0, P_{0,n}}^* \mid T_{\mathcal{M}_n}(P_{0,n})) \} \\ + R_{\mathcal{M}_0}(P_{0,n}, P_0),\end{aligned}$$

where the projection Π_n projects $D_{\mathcal{M}_0, P_{0,n}}^*$ onto tangent space of \mathcal{M}_n at $P_{0,n}$.

- This is a very nice second order remainder (i.e., $o_P(n^{-1/2})$).
- Therefore, our adaptive TMLE is asymptotically linear estimator of $\Psi(P_0)$ with (super-efficient) influence curve $D_{\mathcal{M}_0, P_0}^*$.
- Since it operates as an efficient estimator of $\Psi_{\mathcal{M}_n}(P_0)$ it will also be regular along any path through P_0 that stays in the limit oracle model \mathcal{M}_0 .