Estimation and Evaluation of the Optimal Dynamic Treatment Rule: Practical Considerations, Performance Illustrations, and Application to Criminal Justice Interventions

by

Lina Maria Montoya

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Biostatistics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Associate Professor Maya Petersen, Chair
Professor Mark van der Laan
Professor Jennifer Skeem
Professor Elvin Geng

Fall 2020

Estimation and Evaluation of the Optimal Dynamic Treatment Rule: Practical Considerations, Performance Illustrations, and Application to Criminal Justice Interventions

Abstract

Estimation and Evaluation of the Optimal Dynamic Treatment Rule: Practical Considerations, Performance Illustrations, and Application to Criminal Justice Interventions

by

Lina Maria Montoya

Doctor of Philosophy in Biostatistics

University of California, Berkeley

Associate Professor Maya Petersen, Chair

The optimal dynamic treatment rule (ODTR) framework offers an approach for understanding which kinds of patients respond best to specific treatments – in other words, treatment effect heterogeneity. Further, given an (optimal) dynamic treatment rule, it may be of interest to evaluate that rule – that is, to ask the causal question: what is the expected outcome had every subject received treatment according to that rule? Following the "causal roadmap," in this dissertation, we causally and statistically define the ODTR and its value. Building on work by Luedtke and van der Laan, we provide an introduction to and show finite-sample performance for (1) estimating the ODTR using the ODTR SuperLearner (Chapter 1); and (2) estimating the value of an (optimal) dynamic treatment rule using different estimators, such as cross-validated targeted maximum likelihood (CV-TMLE; Chapter 2). We additionally augment the ODTR SuperLearner by considering stochastic treatment rules and risk criteria that consider the variability of the value of the rule (Chapter 3). We apply these estimators of the ODTR and its value to the "Interventions" Study, an ongoing randomized controlled trial, to identify whether assigning cognitive behavioral therapy (CBT) to criminal justice-involved adults with mental illness using an ODTR significantly reduces the probability of recidivism, compared to assigning CBT in a non-individualized way. We hope this work contributes to understanding the toolbox of methods that can be used to advance the fields of precision medicine, public policy, and health.

To Iván and Martha

Contents

List of Figures

List of Tables

Acknowledgments

This dissertation has actually been generations in the making. Both sets of my *paisa* grandparents – none of whom finished high school – prioritized education over everything when raising their children. Because of that, my parents – both first-borns of a set of 5 siblings – went to medical school in Medellín, Colombia, met there, and subsequently came to the US upon my father's acceptance to an MPH program. After moving back and forth during my childhood, my parents and I ultimately immigrated to the US at a time when violence in Medellín was at its peak, for a safer life, more opportunity, and a better education for me. Their unrelenting support is the reason why I received a stellar college education at Johns Hopkins University and graduate education at UC Berkeley, where I started my PhD in the Fall of 2017 – making me the first in my family to enroll in a PhD program.

Three and a half years later, I completed my dissertation for the doctorate... and it took a village.

First, I want to acknowledge the National Institutes of the Health/National Institute of Allergy and Infectious Diseases for funding my doctorate and allowing me protected time for research (F31 AI140962). Thank you to the ADAPT-R team; I have learned so much from each of you, and I am so grateful that you let me into your study family. Thank you to everyone in the Biostatistics community at UC Berkeley, both past and present. It's been such a privilege and honor to a part of a group where everyone is so smart, creative, and, most importantly, supportive and kind. A special thanks to Yoshika, Catherine, Carrie, and Helen, who helped push me across the dissertation finish line.

Every member of my dissertation committee is brilliant and has been instrumental in my growth as an academic and as a person. Dr. Elvin Geng, thank you for all of the extraordinary opportunities you have afforded me – from inviting me into several study teams, to letting me travel to Kenya, to giving me one of my first leadership positions as part of ADAPT-R. Dr. Mark van der Laan, thank you for teaching me how to think rigorously and critically in a field where it's easy to think there are cookbook, black-and-white answers. Dr. Jennifer Skeem, thank you for giving meaning and motivation to what I do, and for being a phenomenal mentor and human being. I can't wait to keep working with you. And finally, Dr. Maya Petersen: I hope someday you know how grateful I feel that you took me on as your student, and how much I look up to you. And I hope someday you understand the impact you've made on me, my family, and my community. A deep thank you, Maya.

To my family – all of the descendants of Eduardo, Magola, Mario and Luz Stella – you all keep me grounded and are one of my greatest sources of joy. A special thank you to the Velez-Newman family: thank you for taking the role of godparents so seriously, and for giving me a warm, safe, and loving home away from home in California.

Though I am an only child, I do have several sisters. Dana, Courtney, and Priscah – I admire you all so much and feel so lucky to be your friend. Thank you to Sarah, my primary pomodoro buddy and dear cousin; your unwavering support and friendship is so valuable to me, and I can't thank you enough for being who you are. To Caitie: thank you for your

countless edits (you are a brilliant writer), for letting me narrate my life to you for nearly 25 years, and for your commitment to a friendship that is so outstandingly precious to me. And to my *hermana*-sister, Alejandra: I am not exaggerating when I say I could not have done this doctorate without you. Your friendship means the world to me.

My parents are the main reason I am where I am, and it feels overwhelming to even begin to describe in words how grateful I am for everything they have done for me; thus, I dedicate this dissertation to them. Papi y Mami, I love you both so much. A special thanks for learning the difference between "casual" and "causal," and for hearing me present this work countless times – to the point where you now know what a blip function is.

Finally, to Juan Manuel, my closest companion throughout the dissertation journey, and the person I've seen every day since the pandemic started. I am so grateful for your consistent wisdom, support, and love, and I hope you feel this dissertation is a product of your efforts, too. Thank you for letting me weave my life with yours – it's been a total blast.

Introduction

Over the past decade, there has been a marked increase in interest in developing methods for uncovering treatment effect heterogeneity [28, 25]. This stems from the insight that units – patients, participants, users – in a given population often have diverse characteristics, motivations, and needs, and thus will respond to treatments differently. This, paired with goals of the data science era (such as more powerful machinery for prediction) and big data era (such as the collection of richer data sets), have contributed to the recent popularity of developing data-driven methods for improving treatment decisions. In particular, the optimal dynamic treatment rule (ODTR) framework offers a way to identify the treatment option that works best for each kind of person [52, 69], and research aimed at estimating and evaluating the ODTR has grown in recent years, especially within the fields of statistics, machine learning, and causal inference [28].

Work by Luedtke and van der Laan [48] responds to the proliferation and diversity of algorithms for estimating the ODTR by employing the ensemble machine learning/SuperLearner philosophy – that a library of algorithms work in tandem to achieve a certain prediction goal [63]. In the ODTR SuperLearner, a library of ODTR algorithms "team up" to predict which treatment works best for which kind of person. van der Laan and Luedtke [33] additionally lay the theoretical groundwork for evaluating (via the targeted maximum likelihood estimation, or TMLE, framework) such rules, which allow practitioners to determine if administering a treatment in this personalized way is more beneficial than simply giving everyone the same treatment, regardless of covariate profiles. In other words, by evaluating an optimal rule, we can infer if there is any meaningful treatment effect heterogeneity in that population.

Luedtke and van der Laan's seminal work is the foundation of this dissertation. Here, we aim to, first, provide a distilled introduction to and description of the ODTR SuperLearner, in addition to a list of practical considerations for implementing the algorithm. Importantly, we show its finite-sample performance under different library, risk, and metalearner configurations. Chapter 2 focuses on evaluating dynamic treatment rules, and in particular the ODTR, using different estimators. We list the conditions necessary for obtaining adequate inference for different target parameters that correspond to the value of the (optimal) dynamic treatment rule, and the importance of targeting and sample-splitting when evaluating these parameters in the presence of algorithm overfitting. Chapter 3 extends the ODTR SuperLearner to include stochastic rules in its library and a new risk criterion, both of which consider the variability of the expected value of candidate rules, pointing to improvements in finite-sample performance of estimators for the true value of the true ODTR under these extensions. Alexander R. Luedtke, PhD and Jeremy R. Coyle, PhD were co-authors on Chapter 1; Maya L. Petersen, MD, PhD, Mark J. van der Laan, PhD, and Jennifer L. Skeem, PhD were co-authors on Chapters 1, 2, and 3.

Several commonalities between the three chapters will become apparent to the reader. First, every chapter follows the causal roadmap, developed by Petersen and van der Laan [57]. The causal roadmap is an extremely helpful tool for answering causal questions, and transparently doing so under the inevitable constraints of certain data generating processes. Second, throughout, the SL.ODTR package (https://github.com/lmmontoya/SL.ODTR), written by this dissertation's author, is used to estimate all aspects of the aforementioned chapters, and any simulations presented in this dissertation can be found on that GitHub page, as well. Finally, data from the "Interventions" study – an ongoing randomized controlled trial officially called the Correctional Intervention for People with Mental Illness – is used for illustration in each of the chapters. In this trial, 441 (and eventually 720) criminal justice-involved adults with mental illness – a heterogeneous group with diverse symptoms, risk factors, and other treatment-relevant characteristics [82, 83] – are either randomized to cognitive behavioral therapy (CBT) or treatment as usual (TAU), and re-arrest is collected one year after randomization occurs, as a measure of recidivism. Tools presented in this dissertation could ultimately shed light on how to tailor mental health interventions to offenders with mental illness, to ultimately reduce recidivism outcomes.

Throughout, we also emphasize our "big picture" hope for this work – that it serves in helping understand the toolbox of methods available for precision health/medicine/public policy, and ultimately contributes to maximally improving people's outcomes.

Chapter 1

The Optimal Dynamic Treatment Rule SuperLearner: Considerations, Performance, and Application

1.1 Introduction

The primary objective of a clinical trial is often to evaluate the overall, average effect of a treatment on an outcome in a given population [16, 25, 28]. To accomplish this objective in the point treatment setting, baseline covariate, treatment, and outcome data are often collected and the average treatment effect (ATE) is estimated, quantifying the average impact of the treatment in a population. Researchers may then interpret the impact of the treatment as beneficial, neutral, or harmful. In this interpretation, the treatment's impact is one-size-fits-all; in other words, the effect of the treatment is interpreted as the same for everyone in the study population. But, it may be the case that an intervention tends to yield better outcomes for certain kinds of people but not for others. For example, because justice-involved people with mental illness are a heterogeneous group with diverse symptoms, risk factors, and other treatment-relevant characteristics [82, 83], assigning Cognitive Behavioral Therapy (CBT) may decrease the probability of recidivism for individuals with high risk of recidivism but not low risk of recidivism [45]. The ATE analysis may lead one to conclude that there is no treatment effect in a given population, when there is, in fact, a differential treatment effect for levels of variables.

Precision health aims to shift the question from "which treatment works best" to "which treatment works best *for whom*?" (sometimes, it further asks: at what time? And/or at what dose? [28]). The point of moving towards this question is to move towards better subject outcomes. While a range of novel study designs can help to address these questions by generating data in which individualized treatment effects are unconfounded [28, 43, 22], data from classic randomized controlled trials also provide a rich data source for discovering treatment effect heterogeneity. Under the assumption of no unmeasured confounding, the same methods can be applied to observational data.

One way of learning which treatment works best for whom is to estimate effects within subgroups. Following our above example within the field of criminal justice, one could split the sample into subjects who are likely versus unlikely to re-offend, and look at the average effect of CBT on recidivism within these two risk categories. Such a classic subgroup analysis helps to move a step closer to understanding the treatment that works best for whom. However, the need to restrict the number of tests performed and to pre-specify analyses limits traditional subgroup analyses to comparing intervention effects in a small set of subgroups in which heterogeneous treatment effects are expected [25, 44]. In practice, the subject characteristics that are most important for determining the best-suited intervention may not be clear based on background knowledge. Further, effectively predicting the type of intervention that a subject will best respond to may require accounting for a wide range of subject characteristics and complex interactions between them. For instance, identifying the subjects most likely to respond to CBT versus, for example, treatment as usual (TAU) may require considering not only risk level, but also age, educational attainment, sex, substance abuse, psychological distress, and internal motivation to adhere to treatment – as well as various interactions between these. In summary, the challenge is to take a wide range of subject characteristics and flexibly learn how to best combine them into a strategy or rule that assigns to each subject the specific intervention that works best for him or her.

Estimating the optimal dynamic treatment rule (ODTR) for a given population offers a formal approach for learning about heterogeneous treatment effects and developing such a strategy. A dynamic treatment rule can be thought of as a rule or algorithm where the input is subject characteristics and the output is an individualized treatment choice for each subject [2, 34, 67, 9]. An optimal dynamic treatment rule (also known as an optimal treatment regime, optimal strategy, individualized treatment rule, optimal policy, etc.) is the dynamic treatment rule that yields the best overall subject outcomes [52, 69]. In our criminal justice example, a dynamic treatment rule takes as input subject characteristics such as age, criminal history, and education level and outputs a treatment decision – either CBT or TAU. The ODTR is the dynamic treatment rule under which the highest proportion of patients are not re-arrested. It is the most effective and, if one incorporates cost or constraints on resources [46], efficient way of allocating the interventions at our disposal based on measured subject characteristics.

There have been major advances in estimating the ODTR within the fields of statistics and computer science, with important extensions to the case where treatment decisions are made at multiple points in time. Regression-based approaches, such as Q-learning, learn the ODTR by modeling the outcome regression (i.e., the expected outcome given treatment and covariates) directly [52, 41, 80, 50, 64]. Robins and Murphy developed methods of estimating the ODTR by modeling blip-to-reference functions (i.e., the strata-specific effect of the observed treatment versus control) and regret functions (i.e., the strata-specific loss incurred when given the optimal treatment versus the observed treatment), respectively [52, 69, 49]. Direct-estimation approaches to learning the ODTR, such as outcome weighted learning (OWL), aim to search among a large class of candidate rules for the one that yields the best expected outcome [92, 88, 91]. These are examples of broad classes of ODTR estimators; within and outside of them there has been a proliferation of methods to estimate the ODTR (see [28, 29, 89] for reviews of the state of the art in estimating ODTRs and precision medicine).

Given the vast number of methods available for estimating the ODTR, the question becomes: which approach to use? In some settings, some algorithms may work better than others. SuperLearning [35] (or, more specific to prediction, stacked regression [7]) was originally proposed as a method for data-adaptively choosing or combining prediction algorithms. The basic idea is to define a library of candidate algorithms and choose the candidate or the combination of candidates that gives the best performance based on V-fold cross-validation. This requires defining: (1) the algorithms to include in the library, (2) a parametric family of weighted combinations of these algorithms, the "metalearning" step [42], and (3) the choice of performance metric (i.e., risk) as the criterion for selecting the optimal combination of algorithms. Given these three requirements, then one can estimate the risk for each combination of algorithms using V-fold cross-validation, and choose the combination with the lowest cross-validated risk. The SuperLearner framework has been implemented extensively for prediction problems [63, 60, 61, 59], and has been extended to the ODTR setting [48, 15]. In particular, Luedtke and van der Laan showed that in the randomized controlled trial (RCT) and sequential multiple assignment randomized trial (SMART) [29, 1, 43] settings, under the assumption that the loss function is bounded, the ODTR SuperLearner estimator will be asymptotically equivalent to the ODTR estimator chosen by the oracle selector (that is, the ODTR estimator, among the candidate ODTR estimators, that yields the lowest risk under the true data distribution [35]). This implies that the ODTR SuperLearner will asymptotically do as well as or better than any single candidate estimator in the library, provided that none of the candidate algorithms are correctly specified parametric models. If there is a well-specified parametric model in the library, the ODTR SuperLearner estimator of the ODTR will achieve near parametric rates of convergence to the true rule.

These theoretical results lay important groundwork for understanding the asymptotic benefits to using the algorithm; however, less has been published on how the ODTR SuperLearner performs in finite samples, the practical implications of key choices when implementing the algorithm, and illustrations of implementing this algorithm on real RCT data. In this paper, we provide an introduction to the implementation of the ODTR SuperLearner in the point treatment setting, and use simulations to investigate the tradeoffs inherent in these user-supplied choices and how they may differ with varying sample sizes. In particular, for sample sizes 1,000 and 300, we examine: (1) how to select the candidate algorithms for estimating the ODTR; specifically, the costs and benefits to expanding the library to include a wider set of diverse ODTR algorithms, including simple parametric models versus more data adaptive algorithms, and blip-based versus direct estimation algorithms; (2) implications of the choice of parametric family for creating weighted combinations for candidate ODTR learners (i.e., choice of metalearner); and, (3) implications of the choice of risk function used to judge performance and thereby select the optimal weighted combination of candidate learners. Finally, we apply the ODTR SuperLearner to real data generated from the Correctional Intervention for People with Mental Illness, or "Interventions," trial, an ongoing RCT in which justice-involved adults with mental illness were either randomized to CBT or TAU. In applying the ODTR SuperLearner to this sample, we aim to identify which people benefit most from CBT versus TAU, in order to reduce recidivism.

The organization of this article is as follows. First, we step through the causal roadmap (as described in [57]) for defining the true ODTR for a given population. We focus on the case in which baseline covariates are measured, a single binary treatment is randomized, and an outcome is measured. We then give a brief introduction to some estimators of the ODTR, and in particular, describe the SuperLearner approach for estimating the optimal rule that builds on Luedtke and van der Laan's work [48]. We investigate the implications of the three sets of implementation choices outlined above in finite samples using simulations (with corresponding R code illustrating implementation of all estimators considered), and the performance under such options. Lastly, we show results for the ODTR SuperLearner algorithm applied to the "Interventions" Study. We close with concluding remarks and future directions.

1.2 Causal Roadmap and ODTR Framework

Data and Causal Model

Consider point-treatment data where $W \in \mathcal{W}$ are baseline covariates, $A \in \{0,1\}$ is the treatment, and $Y \in \mathbb{R}$ is the outcome measured at the end of the study. Our data can be described by the following structural causal model (SCM), $\mathcal{M}^F$ [56]:

$$\begin{aligned} W &= f_W(U_W) \\ A &= f_A(W, U_A) \\ Y &= f_Y(W, A, U_Y)\ , \end{aligned}$$

where the full data $X = (W, A, Y)$ are endogenous nodes, $U = (U_W, U_A, U_Y) \sim P_U$ are unmeasured exogenous variables, and $f = (f_W, f_A, f_Y)$ are structural equations. If it is known that data were generated from an RCT using simple randomization with equal probability to each arm, then the above structural causal model would state that Y may be affected by both W and A, but that W does not affect A (as in the "Interventions" trial); this can be represented in the above model by letting $U_A \sim \mathit{Bernoulli}(p = 0.5)$ and $A = U_A$. In this point treatment setting, a dynamic treatment rule is a function d that takes as input some function V of the measured baseline covariates W and outputs a treatment decision: $V \rightarrow d(V) \in \{0,1\}$. For the remainder of the paper, we consider the case where $V = W$; in other words, we consider treatment rules that potentially respond to all measured baseline covariates. However, consideration of dynamic rules based on a more restrictive set of baseline covariates is also of frequent practical interest, allowing, for example, for consideration of dynamic rules based on measurements that can be more readily attained; all methods described extend directly to this case. We denote the set of all dynamic treatment rules as $\mathcal{D}$.

The "Interventions" data consist of baseline covariates W, which include intervention site, sex, ethnicity, age, Colorado Symptom Index (CSI) score (a measure of psychiatric symptoms), level of substance use, Level of Service Inventory (LSI) score (a risk score to predict future recidivism that summarizes risk factors like criminal history, educational and employment problems, and attitudes supportive of crime), number of prior adult convictions, most serious offense, Treatment Motivation Questionnaire (TMQ) score (a measure of internal motivation for undergoing treatment), and substance use level; the randomized treatment A, either a manualized Cognitive Behavioral Intervention for people criminal justice system (abbreviated CBT; $A = 1$) or treatment as usual (TAU), primarily psychiatric or correctional services ($A = 0$); and a binary outcome Y of recidivism, an indicator that the person was not re-arrested within one year after study enrollment. In Table 3.1 we show the distribution of the data.

Target Causal Parameter

Let $d(W)$ be a deterministic function that takes as input a vector of baseline covariates, and gives as output a treatment assignment (in this case, either 0 or 1). For a given rule d, we intervene on the above SCM to derive counterfactual outcomes:

$$\begin{aligned} W &= f_W(U_W) \\ A &= d(W) \\ Y_{d(W)} &= f_Y(W, d(W), U_Y) \ . \end{aligned}$$

Here, $Y_{d(W)}$ is the counterfactual outcome for a subject if his/her treatment A were assigned using the dynamic treatment rule $d(W)$; to simplify notation we refer to this counterfactual outcome as Y_d. The counterfactual outcomes for a person if he/she were assigned treatment or given control are denoted Y_1 and Y_0, respectively. Together, the distribution of the exogenous variables P_U and structural equations f imply a distribution of the counterfactual outcomes, and the SCM provides a model for the set of possible counterfactual distributions: $P_{U,X} \in \mathcal{M}^F$.

Our target parameter of interest in this paper is the ODTR, defined as the rule that, among all candidate rules $\mathcal{D}$, yields the best expected outcomes. Using the convention that larger values of Y correspond to better outcomes, an ODTR is defined as a maximizer of $E_{P_{U,X}}[Y_d]$ over all candidate rules

$$d^* \in \arg\max_{d \in \mathcal{D}} E_{P_{U,X}}[Y_d] \ . \tag{1.1}$$

Any such ODTR can be defined in terms of the conditional additive treatment effect (CATE), namely $E_{P_{U,X}}[Y_1 - Y_0|W]$, which is the effect of treatment for a given value of covariates W. Any ODTR assigns treatment 1 and 0 to all strata of covariates for which the CATE is positive and negative, respectively. If the CATE is 0 for a particular W (i.e., there is no treatment effect for that strata of W), the ODTR as defined above may have more than one maximizing rule and therefore may be non-unique; this is why the RHS of equation 1.1 above is a set [47]. An ODTR can take an arbitrary value for strata at which the CATE is 0. If we assume that assigning treatment 0 is preferable to assigning treatment 1 in the absence of a treatment effect, then we would prefer the following ODTR as a function of the CATE:

$$d^*(W) \equiv \mathbb{I}\Big[E_{P_{U,X}}[Y_1 - Y_0|W] > 0\Big] \ .$$

In other words, if a subject's expected counterfactual outcome is better under treatment versus no treatment given his or her covariate profile, then assign treatment; otherwise, assign control. A subject's counterfactual outcome under the ODTR is Y_{d^*}, and the expected outcome had everyone received the treatment assigned by the ODTR is $E_{P_{U,X}}[Y_{d^*}]$.

Following our applied example, Y_1, Y_0, and Y_d are the counterfactual outcomes for a person if he/she were given CBT, TAU, and either CBT or TAU based on the rule d, respectively; here, d^* is the rule for assigning CBT versus TAU using subjects' covariates that would yield the highest probability of no re-arrest, $E_{P_{U,X}}[Y_{d^*}]$.

Identification and Statistical Parameter

We assume that our observed data were generated by sampling n independent observations $O_i \equiv (W_i, A_i, Y_i)$, $i = 1, \ldots, n$, from a data generating system described by $\mathcal{M}^F$ above (e.g., the "Interventions" study consists of 441 i.i.d. observations of O). The likelihood of the observed data can be written as:

$$\mathcal{L}_0(O) = p_{W,0}(W) g_0(A|W) p_{Y,0}(Y|A,W) \ ,$$

where $p_{W,0}$ is the true density of W; g_0 is the true conditional probability of A given W, or the treatment mechanism; $p_{Y,0}$ is the true conditional density of Y given A and W. The distribution of the data P_0 is an element of the statistical model $\mathcal{M}$, which in our RCT example is semi-parametric. Further, if the data are generated from an RCT design, as in the "Interventions" study, then the true g_0 is known, and the backdoor criteria (with the implied randomization assumption [54, 71]), $Y_d \perp A|W \quad \forall d \in \mathcal{D}$, and the positivity assumption, $Pr\Big(\min_{a\in\{0,1\}} g_0(A = a|W) > 0\Big) = 1$, hold by design; in an observational data setting the randomization assumption requires measurement of a sufficient set of baseline covariates, and the positivity assumption may also pose greater challenges [58].

Define $Q(a, w) \equiv E[Y|A = a, W = w]$. Under the above assumption, $E_{P_{U,X}}[Y_d]$ (a parameter of the counterfactual distribution) is identified as $E_0[Q_0(A = d, W)]$ (a parameter of the observed distribution) for any candidate rule d. Thus, the ODTR is identified by

$$d_0^* \in \arg\max_{d \in \mathcal{D}} E_0[Q_0(A = d, W)] \ .$$

In addition, the CATE is identified as $Q_0(1, W) - Q_0(0, W)$, where the latter is sometimes referred to as the blip function $B_0(W)$. Then, the true optimal rule can also be defined as a parameter of the observed data distribution using the blip function:

$$d_0^*(W) \equiv \mathbb{I}[B_0(W) > 0] \ .$$

Analogous to the definition of the ODTR as a function of the CATE, in words, the blip function essentially says that if treatment for a type of subject $W = w$ is effective (i.e., greater than 0), treat that type of person. If not, do not treat him/her. If all subjects were assigned treatment in this way, then this would result in the highest expected outcome, which is the goal.

1.3 Estimation of the ODTR

We denote estimators with a subscript n, so that, for example, an estimator of the true ODTR d_0^* is d_n^*. Estimates are functions of P_n, which is the empirical distribution that gives each observation weight $\frac{1}{n}$; $P_n \in \mathcal{M}_{NP}$, and $\mathcal{M}_{NP}$ is a non-parametric model. In what follows, we briefly describe examples of common methods for estimating the ODTR. We first describe methods that estimate the ODTR via an estimate of the blip function. We then describe methods that directly estimate a rule that maximizes the mean outcome.

Blip-based Approaches

Blip-based approaches aim to learn the blip, which implies an ODTR. A benefit of doing this is that one can look at the distribution of the predicted estimates of the blip for a given sample. Having the blip distribution allows one to identify the patients in a sample who benefit most (or least, or little) from treatment. Additionally, estimating the blip function can allow for estimating the ODTR under resource constraints; for example, an ODTR in which only $k\%$ of the population can receive treatment [46]. Below we illustrate two methods of estimating the ODTR by way of the blip function (i.e., blip-based estimators of the ODTR).

Single stage Q-learning A plug-in estimator naturally follows from the above definition of the optimal rule. One can estimate $Q_n(A, W)$ using any regression-based approach for estimating an outcome regression and predict at $Q_n(1, W)$ and $Q_n(0, W)$. This provides an estimate of the blip: $B_n(W) = Q_n(1, W) - Q_n(0, W)$, which implies an estimate of the optimal rule: $d_n^* = \mathbb{I}[B_n(W) > 0]$ [28, 80, 29, 53].

Estimating the blip function Consider the double-robust pseudo-outcome [77]:

$$D(Q,g) = \frac{2A-1}{g(A|W)}[Y - Q(A,W)] + Q(1,W) - Q(0,W)\ .$$

Importantly, $E_0[D(Q,g)|W] = B_0(W)$ if $Q = Q_0$ or $g = g_0$. Using this result, one could estimate the blip by regressing the pseudo-outcome $D_n(Q_n, g_n)$ (which we abbreviate from here on as D_n) on W using any regression-based approach. As in the previous method, this estimate of the blip implies an estimate of the optimal rule $d_n^* = \mathbb{I}[B_n(W) > 0]$ [69, 48, 33].

Direct Estimation Approaches for Maximizing the Expected Outcome

Instead of estimating the blip function, which implies an ODTR, one could estimate the ODTR directly by selecting a rule d that maximizes the estimated $E_{U,X}[Y_d]$. Below we illustrate outcome weighted learning (OWL) – one example of a direct-estimation method for the ODTR.

Single stage outcome weighted learning We briefly describe the general concept of outcome weighted learning here, but refer to [92] and [78] for a more thorough explanation. The optimal rule defined above as a function of P_0 could equivalently be written as an inverse probability of treatment weighted (IPTW) estimand:

$$d_0^* \in \arg\max_{d \in \mathcal{D}} E_0[Q_0(A = d, W)] = \arg\max_{d \in \mathcal{D}} E_0\left[\frac{Y}{g_0(A|W)}\mathbb{I}[A = d]\right]\ .$$

Written this way, estimating d_0^* could be regarded as a classification problem, where the weighted outcome $\frac{Y}{g(A|W)}$ helps us learn what kind of patients should get treatment: if a certain kind of patient $W = w$ has large weighted outcomes and they were treated according to candidate rule d, future patients with that covariate profile should be treated using that rule. Conversely, the smaller the weighted outcome among patients W who were treated according to d, the larger the "misclassification error" and the less likely those kinds of patients should be treated according to d. This maximization problem is equivalent to the following minimization problem:

$$d_0^* \in \arg\min_{d \in \mathcal{D}} E_0\left[\frac{Y}{g_0(A|W)}\mathbb{I}[A \neq d]\right]\ . \tag{1.2}$$

Now, if patients $W = w$ who did not follow the rule d have large weighted outcomes (and thus larger "misclassification error"), those kinds of patients should be given the opposite treatment that d proposes. Note that in the RCT setting, if one uses the known g_0 and if treatments are given with equal probability, then this reduces to finding the rule that minimizes the mean outcome among patients who did not follow the rule. Equation 1.2

could alternatively be written as a minimization problem for, instead of a rule d, a function f:

$$f_0^* \in \arg\min_{f \in \mathcal{F}} E_0 \left[\frac{Y}{g_0(A|W)} I \left[A \neq \frac{sign(f(W)) + 1}{2} \right] \right] , \tag{1.3}$$

where $sign(x) = -1$ if $x \leq 0$ and $sign(x) = 1$ if $x > 0$. Under the true data distribution P_0, f_0 is the blip function, B_0. In order to solve this minimization problem using data, we can use a plug-in estimator of (1.3); however, since it is a 0-1 function (i.e., it is discontinuous and non-convex), one could use a convex surrogate function to approximate it, to instead minimize:

$$f_n^* \in \arg\min_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^{n} \frac{Y_i}{g_n(A_i|W_i)} \Phi(A_i f(W_i)) + \lambda_n \|f\|^2 , \tag{1.4}$$

where $\Phi(t)$ is the surrogate loss function (e.g., hinge loss, exponential loss, logistic loss), $\|f\|$ is the norm of f, and λ_n is the estimated penalization parameter on f to avoid overfitting of the rule. This can also be generalized with the IPTW function replaced by the augmented IPTW [48, 78]. Once f_n^* is found as the solution to equation (1.4), the estimated ODTR is:

$$d_n^* = sign(f_n^*(W)) .$$

SuperLearner to Estimate ODTR

The overarching goal of SuperLearner is to let a user-supplied library of candidate algorithms, such as specific implementations of the general approaches described above, “team up” to improve estimation of the ODTR. In order to implement the ODTR SuperLearner, there are three user-supplied decisions one must make. First, one must consider the library of candidate algorithms to include. These could include algorithms for estimating the blip function (which imply a rule), algorithms that search for the ODTR directly (such as OWL estimators), static rules that determine treatment regardless of covariates, or combinations of the above classes of algorithms. Second, in what is sometimes referred to as the metalearning step, one can either implement a SuperLearner that chooses one algorithm out of the library of candidate algorithms to include (i.e., “discrete” SuperLearner), or a SuperLearner that is a combination the candidate algorithms (i.e., “continuous” SuperLearner). For the latter, one again has a choice of metalearner; we consider weighted convex combinations of candidate estimators of the blip and combinations of estimates of the rules themselves (through a weighted “majority vote”). Finally, one must choose the risk function used to judge the performance of the weighted combinations of algorithms (estimated using V-fold cross validation). Here, we consider two risk functions: the mean-squared error (MSE) and the mean outcome under the candidate rule.

The steps for implementing the ODTR SuperLearner are as follows; they closely follow the implementation of the canonical SuperLearner for regression [37]:

1. Choose J candidate algorithms for estimating the optimal rule $d_{n,j}(W)$ for $j = 1, ..., J$. Candidates can include approaches based on estimating the blip $B_{n,j}(W)$, e.g., candidate regressions of D_n on W, where the candidate regressions might consist of a parametric linear model (corresponding to a classic approach of fitting a parametric outcome regression on A and W) as well as more flexible machine learning type approaches such as neural networks [66], multivariate adaptive regression splines [18], or recursive partitioning and regression trees [5]. Candidate algorithms might also include approaches for estimating the optimal rule directly, such an OWL estimator. Finally, the static treatment rules that treat all patients or treat no patients, regardless of their covariate values, can also be included as candidates. Inclusion of both simple parametric model estimators, as well as static rules such as treating all and treating none, is important to allow for the possibility that the underlying true ODTR may in fact be simple (or well-approximated by a simple rule), and providing less aggressive candidates in the SuperLearner library can help protect against overfitting in finite samples.

2. Split the data into V exhaustive and mutually exclusive folds. Let each fold in turn serve as the validation set and the complement data as the training set.

3. Fit each of the J candidate algorithms on the training set. Importantly, candidate algorithms might depend on nuisance parameters, and those nuisance parameters should be fit on the training set, as well. For example, if a candidate algorithm regresses D_n on W to estimate the blip (which implies an ODTR), then Q and g should be fit and predicted on the training set, and then plugged into D to fit that candidate algorithm on the same training set (this is also called "nested" cross-validation, described in detail by [15]).

4. Predict the estimated blip or the treatment assigned under the estimated ODTR for each observation in the validation set for each algorithm, based on the corresponding training set fit.

5. Choose to either implement the discrete SuperLearner, which selects one algorithm out of the candidate algorithms, or the continuous SuperLearner, which creates a weighted average of the candidate algorithms.

   a) Continuous SuperLearner. Create different convex combinations of the candidate blip or treatment rule estimates that were predicted on the validation set (i.e., convex combinations of the predictions from the previous step). Formally, define an estimator of $B_{n,\alpha}(W)$ or $d_{n,\alpha}(W)$ as a convex combination of the candidate algorithms (indexed by j); each convex combination of algorithms is indexed by a weight vector α. A given convex combination of blip estimates are denoted as:

$$B_{n,\alpha}(W) = \sum_j \alpha_j B_{n,j}(W), \alpha_j \geq 0 \forall j, \sum_j \alpha_j = 1 \ .$$

Alternatively, the predicted treatments under the candidate ODTRs can be combined as a weighted "majority vote" of the convex combination of the candidate rules:

$$d_{n,\alpha}(W) = \mathbb{I}\Big[\sum_j \alpha_j d_{n,j}(W) > \frac{1}{2}\Big], \alpha_j \geq 0 \forall j, \sum_j \alpha_j = 1 \ .$$

b) Discrete SuperLearner. The discrete SuperLearner, which chooses only one candidate algorithm, can be thought of as a special case of the continuous SuperLearner, where algorithms are still combined using a convex combination, but each algorithm weight α_j must be either 0 or 1. Such an approach may be particularly advantageous when sample size is small:

$$B_{n,\alpha}(W) = \sum_j \alpha_j B_{n,j}(W), \alpha_j \in \{0,1\} \forall j, \sum_j \alpha_j = 1$$

$$d_{n,\alpha}(W) = \mathbb{I}\Big[\sum_j \alpha_j d_{n,j}(W) > \frac{1}{2}\Big], \alpha_j \in \{0,1\} \forall j, \sum_j \alpha_j = 1 \ .$$

6. Calculate an estimated risk within each validation set for each combination of algorithms (i.e., for each convex combination indexed by a particular value for α). Here, we discuss two choices of risk functions for step (6) above. First, mean-squared error risk targeting the blip function, which we will refer to as R_{MSE}:

$$R_{MSE} = E[(D(Q,g) - B(W))^2] \ .$$

Because the MSE compares D_n to the predicted blip, the candidate estimators under the MSE risk function must output estimated blip values. Of note, $E_{P_{U,X}}[[Y_1 - Y_0 - B(W)]^2]$ is identified as $R_{MSE_0(B)}$ if either $Q = Q_0$ or $g = g_0$. The second risk function, which we call $R_{E[Y_d]}$, uses the expected rule-specific outcome as criterion:

$$R_{E[Y_d]} = -E[E[Y|A = d, W]] \ .$$

Intuitively, SuperLearner aims to choose the combination of treatment rule algorithms that minimizes a cross-validated empirical risk, so it makes sense to have that risk be the negative of the expected outcome, such that SuperLearner chooses the combination of algorithms that maximizes the expected outcome, since that is the ultimate goal of the ODTR. Candidate estimators for the SuperLearner that use the expected mean outcome under the rule as the risk function can include both blip estimators that imply treatment rules as well as direct estimators of the treatment rules. When the expected rule specific outcome is chosen as the risk function, a further practical choice is how to estimate this quantity; we focus here on a cross-validated targeted maximum likelihood estimator (TMLE) due to its favorable theoretical properties (double robustness, semiparametric efficiency, and greater protection against overfitting through the use of sample splitting [37]); however, one can use any estimator of treatment specific mean outcomes to estimate this quantity [2, 30, 34].

7. Average the risks across the validation sets resulting in one estimated cross validated risk for each candidate convex combination (i.e., each possible choice of α).

8. Choose the estimator (i.e., the convex combination α) that yields the smallest cross-validated empirical risk. Call this "best" weighting of the algorithms α_n.

9. Fit each candidate estimator $B_{n,j}(W)$ of the blip or $d_{n,j}(W)$ of the optimal rule on the entire data set. Generate predictions for each candidate algorithm individually, and then combine them using the weights α_n obtained in the previous step. This is the SuperLearner estimate of the ODTR, where $d^*_{n,B} = \mathbb{I}[B_{n,\alpha_n}(W) > 0]$ or $d^*_{n,d} = d_{n,\alpha_n}(W)$ directly.

We summarize the practical implications of 3 key choices for implementing ODTR SuperLearner here and in Table 2.1. In the first dimension, selection of the candidate algorithms, for illustration we consider having a library with only blip function estimators (called "Blip only" library) or a library with blip estimators, direct-estimation estimators, and static treatment rules that treat everyone or no one (called "Full" library). If one chooses to include direct-search estimators of the ODTR or static rules (i.e., functions that do not output a blip estimate in the process), then one is constrained to using the vote-based metalearner and $R_{E[Y_d]}$ risk function, because the blip-based metalearner and R_{MSE} risk function both rely on estimates of the blip for combining and choosing the algorithms, respectively. For the second dimension, the choice of how to combine algorithms, we consider either the metalearner that (a) selects only one candidate algorithm (called "Discrete"), (b) uses a weighted average to combine predicted blip estimates and directly plugs those into the risk (called "Blip-based"), (c) uses a weighted average to combine predicted treatments under the candidate combinations of rules and creates a weighted majority vote of these candidate rules as input into the risk (called "Vote-based"). The third dimension is the choice of performance measure, risk function – either the MSE (R_{MSE}) or the mean outcome under the candidate rule ($R_{E[Y_d]}$). For the second and third dimensions, if one uses the vote-based metalearner, then the R_{MSE} risk cannot be used because R_{MSE} requires an estimate of the blip to choose the best algorithm, and the vote-based metalearner does not output an estimate of the blip.

Choice 1: Library	**Blip only**						**Full**					
Choice 2: Metalearner	**Discrete**		**Continuous**				**Discrete**		**Continuous**			
			Blip-based		**Vote-based**				**Blip-based**		**Vote-based**	
Choice 3: Risk	R_{MSE}	$R_{E[Y_d]}$	R_{MSE}	$R_{E[Y_d]}$	R_{MSE}	$R_{E[Y_d]}$	R_{MSE}	$R_{E[Y_d]}$	R_{MSE}	$R_{E[Y_d]}$	R_{MSE}	$R_{E[Y_d]}$
Possible?	✓	✓	✓	✓	✗	✓	✗	✓	✗	✗	✗	✓

Table 1.1: Summary of the possible ODTR SuperLearner configurations across the library, metalearner, and risk dimensions. The last row ("Possible?") indicates whether the particular configuration is possible to implement. The checkmarks (✓) in the following table indicate that it is possible to construct that kind of ODTR SuperLearner algorithm; the x-marks (✗) indicate otherwise.

1.4 Simulation: Comparisons and Considerations of SuperLearner ODTR Estimators

We use simulations to: (1) illustrate the potential benefit to estimating the ODTR using a SuperLearner approach, as compared to a more traditional approach to studying treatment effect heterogeneity based on fitting an outcome regression with interaction terms on covariates and treatment, as is often standard practice [16, 86, 27, 87]; and (2) investigate the implications of practical choices when implementing an ODTR SuperLearner in finite samples, including specification of candidate algorithms in the library, choice of metalearner, and choice of risk function.

Data Generating Processes

All simulations were implemented in R [65], and the code, simulated data, and results can be found at https://github.com/lmmontoya/SL.ODTR. We examine these comparisons using two types of data generating processes (DGPs). Each simulation consists of 1,000 iterations of either $n = 1,000$ or $n = 300$, to assess the impacts of the different configurations as a function of sample size. Both DGPs generate the covariates, treatment, and outcome as follows:

$$
\begin{aligned}
W_1, W_2, W_3, W_4 &\sim Normal(\mu = 0, \sigma^2 = 1) \\
A &\sim Bernoulli(p = 0.5) \\
Y &\sim Bernoulli(p) \ .
\end{aligned}
$$

The probability of having a successful outcome differs for the two DGPs, which, in this case, means that the blip functions differ as well. The first DGP is an example of one with a complex blip function, and the second DGP is one with a blip function that is a simpler function of one variable. The first DGP is directly from work by Luedtke and van der Laan

[48, 46, 33], and the second is modified from the first. The probability of a successful outcome for DGP 1 is:

$$\begin{aligned}p = 0.5logit^{-1}(1 - W_1^2 + 3W_2 + \\ 5W_3^2 A - 4.45A) + 0.5logit^{-1}(-0.5 - W_3 + 2W_1W_2 + 3|W_2|A - 1.5A) \ ,\end{aligned}$$

then the true blip function is:

$$\begin{aligned}B_0(W) =& 0.5[logit^{-1}(1 - W_1^2 + 3W_2 + 5W_3^2 - 4.45) \\ &+ logit^{-1}(-0.5 - W_3 + 2W_1W_2 + 3|W_2| - 1.5) \\ &- logit^{-1}(1 - W_1^2 + 3W_2) + logit^{-1}(-0.5 - W_3 + 2W_1W_2)] \ .\end{aligned}$$

For DGP 1, $E_{P_{U,X}}[Y_{d^*}] \approx 0.5626$ and the true optimal proportion treated $E_{P_{U,X}}[d^*] \approx 55.0\%$. $E_{P_{U,X}}[Y_1] \approx 0.4638$ and $E_{P_{U,X}}[Y_0] \approx 0.4643$.

DGP 2's probability of the outcome's success is:

$$p = logit^{-1}(W_1 + 0.1A + W_1A) \ .$$

Thus the true blip function is:

$$B_0(W) = logit^{-1}(W_1 + 0.1 + W_1) - logit^{-1}(W_1) \ .$$

For DGP 2, $E_{P_{U,X}}[Y_{d^*}] \approx 0.5595$ and $E_{P_{U,X}}[d^*] \approx 54.0\%$; $E_{P_{U,X}}[Y_1] \approx 0.5152$ and $E_{P_{U,X}}[Y_0] \approx 0.5000$.

ODTR Estimators

For each data generating process, we consider a number of estimators of the ODTR. First, mirroring epidemiologic practice, we model the outcome as an additive function of the treatment and covariates, and interactions with the treatment and all covariates) [16, 86, 87, 27]. Such an approach translates to using the following parametric model for the outcome regression:

$$h(E[Y|A, W]) = \beta_0 + \sum_{i=1}^{p} \beta_i W_i + \left(\gamma_0 + \sum_{i=1}^{p} \gamma_i W_i\right) A \ ,$$

where $h(.)$ denotes a link function, and p is the number of baseline covariates in W. Using a linear link, the following parametric model for the blip function is implied:

$$B(W) = \gamma_0 + \sum_{i=1}^{p} \gamma_i W_i \ .$$

Next, we examine the finite sample implications of the aforementioned user-supplied choices in implementing a SuperLearner estimator of the ODTR, providing guidance for practical data analysis. First, we examine the choice of library. We consider the library that only combines candidate blip estimators ("Blip only" library; i.e., a library with candidate algorithms suited for regressing D_n on W) versus a library that has blip estimators, direct-estimation algorithms, and static treatment rules ("Full" library). The "Blip only" libraries consist of either:

(a) Simple parametric models only (denoted "Parametric blip models"). This consisted of univariate GLMs with each covariate.

(b) Machine learning algorithms only (denoted "ML blip models"), such as `SL.glm` (generalized linear models), `SL.mean` (the average), `SL.glm.interaction` (generalized linear models with interactions between all pairs of variables), `SL.earth` (multivariate adaptive regression splines [18]), `SL.nnet` (neural networks [66]), `SL.svm` (support vector machines [12]), and `SL.rpart` (recursive partitioning and regression trees [5]) from the SuperLearner package [62]

(c) A combination of (a) and (b) above, denoted "Parametric + ML blip models"

The "Full" library includes other ODTR algorithms like direct-estimation methods, static rules, and other blip-based methods. Specifically, the "Full" library includes either the "ML blip models" or "Parametric + ML blip models" from the "Blip only" libraries above, in addition to Q-learning [29], OWL [92], residual weighted learning (RWL) [94], efficient augmentation and relaxation learning (EARL) [90], optimal classification algorithms [88] (the latter 4 are from the DynTxRegime package [21], with function names `owl, rwl, earl,` and `optimalClass`, respectively), and static rules that treat all patients and no patients, regardless of the patient covariate profiles. For the algorithms from the DynTxRegime package, except for nuisance parameters Q_n and g_n, we use default parameters, and the rule as a function of all covariates. Additionally, for the optimal classification algorithm, the solver method is recursive partitioning for regression trees (`rpart`). Thus, the possible "Full" libraries are:

(d) Algorithms from the "ML blip models" library, plus direct maximizers and static rules, denoted "ML blip models and $E[Y_d]$ maximizers"

(e) All possible algorithms – that is, algorithms from the "Parametric + ML blip models" library, plus direct maximizers and static rules, denoted "All blip models and $E[Y_d]$ maximizers"

Second, we examine the performance of different metalearners for combining the candidate ODTR algorithms. We examine the blip-based metalearner using the "Blip only" libraries, and the discrete and vote-based metalearners using both the "Blip only" libraries and "Full" libraries.

Third, we examine the performance of using either the MSE R_{MSE} or the expected outcome under the candidate rule $R_{E[Y_d]}$ as risk criteria for choosing the optimal linear combination of candidate ODTR algorithms. In particular, CV-TMLE is used for estimating $R_{E[Y_d]}$.

We fully estimate the ODTR SuperLearner by additionally estimating nuisance parameters (as opposed to using the true nuisance parameter functions) in a nested fashion [15] as described above. Specifically, we estimate Q_n and g_n using the canonical SuperLearner [35, 62] and a correctly specified parametric model (i.e., a logistic regression of A on the intercept), respectively. We use 10-fold cross-validation throughout.

Performance Metrics

We measure performance by computing the percent accuracy of the algorithm; that is, in a sample, the proportion of times the treatment assigned by the estimated ODTR matches the true optimal treatment (i.e., the treatment that would have been assigned under the true ODTR) for each observation in the sample, averaged across simulation repetitions. We also evaluate performance metrics of the difference between the true conditional expected outcome under the estimated rule, averaged across the sample, compared to the true mean outcome under the true optimal rule $E_n[Q_0(Y|A = d_n^*, W)] - E_0[Y_{d_0^*}]$ (as an approximation to the regret $E_0[Q_0(Y|A = d_n^*, W)] - E_0[Y_{d_0^*}]$). We compute the mean and variance of this difference across the simulation repetitions. Instead of presenting the raw variance of the regret, we present a variance relative to the regret yielded by estimating the blip, and thus the optimal rule, using as a parametric GLM that models the blip as an additive function of all covariates. Additionally, we compute 2.5^{th} and 97.5^{th} quantiles of the distribution of $E_n[Q_0(Y|A = d_n^*, W)]$ across the simulation repetitions.

Simulation Results

Figure 1.1 displays simulation results (in addition to tables in the appendix). Below we discuss results specific to each DGP, configuration dimension, and sample size. In general, results within each DGP (i.e., across sample sizes) follow generally similar patterns; however, any differences in performance between libraries, metalearners, or risks are more pronounced for the smaller sample size.

DGP 1 Results - "Complex" Blip Function

Above, we showed that DGP 1 yields a blip function that is a complex function of all of the available covariates. Here, for a larger sample size, we would expect a benefit to more aggressive approaches to estimating the ODTR, such as including more flexible machine learning-based approaches in the library of candidates, as well as use of more aggressive metalearners (either vote- or blip-based) over a discrete SuperLearner due to the better ability of these approaches to approximate the true underlying function. That said, for smaller sample sizes, this benefit might be attenuated, or even result in worse performance than simpler alternatives. For this DGP, at sample size of 1,000, indeed we find a benefit to the use of both more aggressive metalearners and larger libraries. Interestingly, however, this benefit is maintained for sample size 300. Specifically, libraries that included data adaptive, machine learning algorithms (as opposed to incorrectly specified parametric models alone) more accurately and precisely approximated the rule, even for sample size of 300. Results also show that for both sample sizes, within the discrete metalearner, the $R_{E[Y_d]}$ risk performed better than R_{MSE} risk, and more saliently, the blip-based and vote-based metalearners performed better than the discrete SuperLearner. Finally, as predicted by theory, all SuperLearner approaches evaluated substantially outperformed a traditional parametric regression approach at both sample sizes. Below we describe results specific to each sample size.

n = **1,000** Libraries that contain machine learning algorithms (i.e., "ML blip models," "Parametric + ML blip models," "ML blip models and EYd maximizers, " and "Blip models and EYd maximizers") overall perform better than libraries with parametric models only (i.e., "Parametric blip models") and the standard GLM (i.e., "GLM"), across all performance metrics. For example, the percent match between the true ODTR and the estimated ODTR spans from 72.0%-77.7% for any libraries with machine learning algorithms, whereas the percent match for libraries with only parametric models is from 56.4% to 58.0%.

There are no stark differences within the libraries that contain machine learning algorithms across the metalearner and risk dimensions, except when using a discrete metalearner and R_{MSE} risk. Specifically, the discrete metalearner that uses R_{MSE} has a higher average regret and relative variance than all other algorithms that use machine learning (e.g., for the "Parametric + ML blip models" "Blip only" library that uses a discrete metalearner, the average regret when using R_{MSE} versus $R_{E[Y_d]}$ is -0.0389 versus -0.0284, respectively, and the relative variance when using R_{MSE} versus $R_{E[Y_d]}$ is 2.137 versus 0.7781, respectively).

n = **300** As expected, given the limited data available to estimate a complex underlying function, both accuracy of treatment assignment and approximated regret (the extent to which the expected outcome under the estimated rule fell short of the best outcomes available) deteriorated with smaller sample sizes. That said, even in this challenging situation of a complex true pattern of treatment effect heterogeneity and limited data with which to discover it, the ODTR SuperLearner would have improved the expected outcome by approximately 4.5% relative to the static rule that treats everyone, an approach that would have been suggested based on estimation of the ATE.

Libraries with only parametric models perform worse than libraries that contain machine learning algorithms in terms of average regret and accuracy. For example, SuperLearner ODTRs that contain libraries with parametric blip models match 54%-55.4% of the time with the true ODTR, while the SuperLearners that contain libraries with machine learning algorithms match 60.9%-66.1% of the time. These results parallel those found with sample size 1,000, except the discrepancy between libraries with machine learning models and parametric models is not as pronounced.

Similar to the $n = 1,000$ case, among the libraries that used machine learning algorithms, using the performance of the rule as risk is better across all performance metrics than using MSE as risk for the discrete metalearner. As long as machine learning methods were included in the library, performance was similar across risk functions and choice of metalearners, with the exception of the MSE risk combined with the discrete metalearner.

DGP 2 Results - "Simple" Blip Function

As shown above, DGP 2 has a true blip function that is a simple function of one covariate (referred to here as a "simple" blip function). Here, the true optimal rule is described by a simple parametric model for the blip; thus, we expect this approach to perform well. However, in practice one is unlikely to be sure that the truth can be well approximated by a simple rule; it is thus of interest to evaluate what price is paid for expanding the library to include more aggressive machine learning algorithms and metalearners. In particular, one might expect that, for smaller sample size, adding machine learning-based candidates and more complex metalearners risks substantial drop-off in performance. However, specifying a library that includes simpler parametric models, in addition to machine learning approaches, may help mitigate this risk. In fact, for this particular DGP, we see, across metalearners and risks, only a small price in performance from adding machine learning algorithms to a library including only simple parametric models. In short, having an ODTR SuperLearner library that also includes parametric models is better than having a library that only consists of data adaptive, machine learning algorithms. Within the libraries that did contain parametric models, particularly for the discrete metalearner, R_{MSE} risk performs slightly better than $R_{E[Y_d]}$ risk; for other metalearners there is little difference in performance in terms of risk. Performance of the metalearners varies slightly by sample size. Below we describe results specific to each sample size.

$n =$ **1,000** In terms of accuracy, the libraries that only contain parametric models perform the best, followed closely by libraries that contain parametric models and machine learning models, followed by the library with only machine learning models. This pattern is evident in the percent match with the true ODTR: for example, within the discrete metalearner with R_{MSE} risk, the percent accuracy is 90.7% for the library with parametric models only ("Parametric blip models" library), 88.8% for the library that contain both parametric models and machine learning models ("Parametric + ML blip models"), and 81.9% for the library that contains machine learning algorithms only ("ML blip models"). This same pattern is apparent in terms of average regret; that is, the libraries that contain parametric blip models or a combination of parametric blip models and machine learning models have the lowest average regret (-0.0041 to -0.0095), while the libraries that only contain machine learning models have the highest average regret (-0.0100 to -0.0138). Modeling the blip with a single, parametric model often used in standard epidemiological analysis (which, in this case, is incorrectly specified) yields an average regret of -0.0100 (higher than the libraries with a combination of parametric models and/or machine learning algorithms, and at the same level as having machine learning algorithms only).

Within the libraries that contain parametric models and use a discrete metalearner, R_{MSE} performs better than $R_{E[Y_d]}$. For example, the mean regret and relative variance for the discrete metalearner that only used parametric models in the library is -0.0041 and 1.0267, respectively, when using R_{MSE} risk, and -.0046 and 1.3174, respectively, when using $R_{E[Y_d]}$ risk. Otherwise, there were no apparent differences in performance by risk.

For libraries that contain parametric models and use R_{MSE}, the discrete ODTR SuperLearner performs better than the blip-based ODTR SuperLearner, with regards to all performance metrics. For example, the average regret and relative variance for the library with only parametric models that uses R_{MSE} was -0.0041 and 1.0267, respectively, when using a discrete metalearner versus -.0059 and 1.1855, respectively, when using the blip-based metalearner. This pattern is also evident for the library that has both parametric models and machine learning algorithms.

$n = \mathbf{300}$ As in the case where $n = 1,000$, the library with only parametric models performs best in terms of accuracy, followed by libraries with parametric models and machine learning models, and finally libraries with only machine learning algorithms; again, however, DGP 1 illustrates the risks of such a strategy. Moreover, even at this small sample size, there is only a small price to pay for adding machine learning-based learners to a library that also includes simple parametric models. For example, for the discrete metalearner that uses R_{MSE}, the percent accuracy for the library that uses only parametric models is 78%, followed by a 75.5% accuracy when there is a combination of parametric models and machine learning, while the library with only machine learning models had a 61.7% match with the true ODTR. While this pattern parallels that of the $n = 1,000$ case, the dropoff in accuracy when the library uses only parametric models versus when the library only uses machine learning algorithms is larger in terms of accuracy in the smaller sample size (16.3% difference) versus the larger sample size (8.8% difference).

Similar to the $n = 1,000$ case, among libraries that contain parametric models and in the discrete metalearner case, R_{MSE} yields slightly better performance results than $R_{E[Y_d]}$. In contrast to the $n = 1,000$ case, for libraries that contain parametric models and used R_{MSE}, the blip-based metalearner performs slightly better than the discrete metalearner, with regards to all performance metrics. For example, the average regret and relative variance for the library with only parametric models that use R_{MSE} is -0.0188 and 1.6102, respectively, when using a blip-based metalearner versus -0.0190 and 1.8109, respectively, when using the discrete metalearner.

1.5 Application of ODTR SuperLearner to "Interventions" Study

In the "Interventions" study, 231 (52.2%) participants received CBT and 210 (47.8%) TAU. Out of the 441 participants, 271 (61.5%) were not re-arrested within the year. The estimated probability of no re-arrest had everyone been assigned CBT is 62.2%, and the estimated probability of no-arrest had everyone been assigned TAU is 60.7%; there was no significant difference between these two probabilities (risk difference: 1.51%, CI: [-8.03%,11.06%]). After adjusting for covariates using TMLE to improve the precision on this ATE estimate [51], the risk difference is, similarly, 1.53% (CI: [-7.31%, 10.37%]).

Figure 1.2 shows subgroup plots for each covariate – that is, the unadjusted difference in probability of no re-arrest between those who received CBT versus TAU, within each covariate group level. One might begin to identify trends towards differential treatment effects; for example, participants may have benefited more from CBT at the San Francisco site, or if they had offended three or more times. Accurate interpretation of any such subgroup analyses, however, requires variance estimates and hypothesis tests with appropriate correction for multiple testing. In addition, as mentioned before, it may be the case that the best way to assign treatment is by using information on more than one variable at a time, and even interactions between those variables.

Thus, we estimated the ODTR on the "Interventions" data to determine which justice-involved adults with mental illness should receive CBT. Specifically, we implemented the ODTR SuperLearner with a blip-only library, a continuous, blip-based metalearner, and $R_{E[Y_d]}$ risk function. We chose a blip-based library in order to generate estimates of the blip, which themselves can be informative. The library for d_n^* consisted of a combination of simple parametric models (univariate GLMs with each covariate) and machine learning algorithms (`SL.glm`, `SL.mean`, `SL.glm.interaction`, `SL.earth`, and `SL.rpart`). As in the simulations, the outcome regression Q_n was estimated using the canonical SuperLearner, g_n was estimated as an intercept-only logistic regression, and we used 10-fold cross validation.

Interestingly, despite implementing a continuous metalearner, the ODTR SuperLearner algorithm assigned all weight on a GLM that modeled the blip as a linear function of only substance use. As shown in Figure 1.3, a plot depicting the distribution of the predicted blip estimates for all participants, the algorithm can be interpreted as such: if a justice-involved person with mental illness has a low substance use score, give him/her CBT; otherwise, give him/her TAU. Under this ODTR estimate, for this sample, 52.38% of the participants would receive CBT.

1.6 Conclusions

We described the ODTR SuperLearner and illustrated its performance for sample DGPs under different configurations of the algorithm and finite sample sizes. These results build on existing work [48, 15] by fully estimating the ODTR and including an expanded Super-

Learner library with not only blip-based regression estimators, but also direct-estimation methods and static interventions. We highlighted the practical choices one must consider when implementing the ODTR SuperLearner across three dimensions: (1) the ODTR SuperLearner library of candidate algorithms, namely, whether to include parametric models, machine learning algorithms, or both; whether to include only estimators that output a predicted blip or include a combination of blip estimators, direct estimators, and static treatment rules, (2) the metalearner that either chooses a single algorithm or combines the algorithms and (3) the risk function that chooses the "best" estimator or combination of estimators of the candidate ODTR algorithms.

Simulation-based results illustrate the shortcomings of an approach to treatment effect heterogeneity based on approximating the blip as an additive function of the available covariates, or equivalently, modeling the outcome as an additive function of the treatment and covariates, and interactions between the treatment and all covariates, which is common practice in epidemiologic analyses for heterogeneous treatment effects [16, 86, 87, 27]. With respect to choice of library, we recommend specifying a library with both simple parametric models and more aggressive data adaptive algorithms, as well as static rules such as treat all or treat none, allowing for flexible estimation of both simple and complex underlying rules. Inclusion of a full range of algorithms from simple to aggressive was particularly important for small sample sizes. In terms of the choice of metalearner, both vote- and blip-based ensemble learners performed well; a vote-based metalearner has the advantage, however, of allowing for the integration of a larger library of candidate algorithms (including direct estimation approaches) and ease of integration of static rules. Of note, in these simulations, vote- and blip-based metalearners outperformed the discrete ODTR SuperLearner approach, even for sample size 300. However, we caution that this may not always be the case and when sample size is small, a discrete SuperLearner approach may provide benefits – in fact, one could include a convex metalearner as a candidate algorithm. Finally, with respect to choice of risk function, both MSE and the expected outcome under the rule performed well; in practice one might prefer $R_{E[Y_d]}$ because it allows for the use of a larger library of candidate algorithms.

Additionally, as an illustration of how to apply the ODTR SuperLearner to real data, we estimated the ODTR using the "Interventions" study to determine which types of justice-involved adults with mental illness should be assigned CBT versus TAU, to yield the highest probability of no re-arrest. Preliminary results show that the ODTR SuperLearner placed all weight on a simple parametric model with only substance use; thus, the algorithm suggests that, in this sample, participants with lower levels of substance use may benefit more from CBT. We note that this is an example of a case in which the ODTR SuperLearner generated a ODTR estimate that was fully interpretable – although we used a continuous metalearner and thus the SuperLearner could have allowed for combinations of algorithms, the SuperLearner happened to only choose one algorithm: a GLM with substance use as a regressor. To guarantee interpretability in the SuperLearner (for example, if working with practitioners who may want a treatment decision rule that could be easily written down [28, 13]), one could implement the ODTR SuperLearner with a discrete metalearner and a simple parametric library only.

Importantly, in a companion paper, we *evaluate* this ODTR – that is, we ask the causal question: what would have been the probability of no re-arrest had participants been assigned CBT according the ODTR SuperLearner (i.e., using only substance use)? Further, is assigning CBT according to the ODTR SuperLearner significantly better than assigning CBT to everyone or no one? In this way, we can determine if it is of clinical significance to assign CBT according to this rule – namely, if assigning CBT using only substance use scores results in a statistically significant reduction of recidivism, and if so, how much better one does with this ODTR compared to a non-individualized rule (such as giving CBT to all). Under the appropriate causal assumptions, one could use any of the methods for estimating treatment specific means to interpret this estimate as the expected outcome under the true optimal rule or the estimated optimal rule.

Future work could extend these simulations to the multiple treatment (i.e., more than 2 treatment levels) [15] and multiple timepoint setting (i.e., estimating a sequential ODTR from, for example, a SMART design [43, 29, 1] instead of an RCT design). We also wish to apply the ODTR SuperLearner on the full "Interventions" dataset (n = 720), once data collection is finished.

This work contributes to understanding the toolbox of methods for analyzing the heterogeneity in how patients respond to treatment. By learning which patients respond best to what treatment in a flexible manner, we can improve patient outcomes – moving us closer to the goals of precision health.

	TAU ($A = 0$)	CBT ($A = 1$)	p
n	211	230	
No re-arrest ($Y = 1$) (%)	128 (60.7)	143 (62.2)	0.820
Site = San Francisco (%)	87 (41.2)	104 (45.2)	0.455
Gender = Female (%)	38 (18.0)	37 (16.1)	0.682
Ethnicity = Hispanic (%)	50 (23.7)	42 (18.3)	0.198
Age (mean (SD))	38.08 (11.05)	37.01 (11.22)	0.317
CSI (mean (SD))	32.35 (11.13)	33.46 (11.27)	0.300
LSI (mean (SD))	5.59 (1.33)	5.50 (1.48)	0.472
SES (mean (SD))	3.81 (1.89)	3.81 (2.12)	0.995
Prior adult convictions (%)			0.156
Zero to two times	74 (35.1)	93 (40.4)	
Three or more times	134 (63.5)	129 (56.1)	
Missing	3 (1.4)	8 (3.5)	
Most serious offense (mean (SD))	5.29 (2.54)	5.09 (2.52)	0.415
Motivation (mean (SD))	3.22 (1.36)	3.27 (1.37)	0.720
Substance use (%)			0.184
0	53 (25.1)	76 (33.0)	
1	47 (22.3)	55 (23.9)	
2	109 (51.7)	98 (42.6)	
Missing	2 (0.9)	1 (0.4)	

Table 1.2: Distribution of “Interventions” data by treatment assignment.

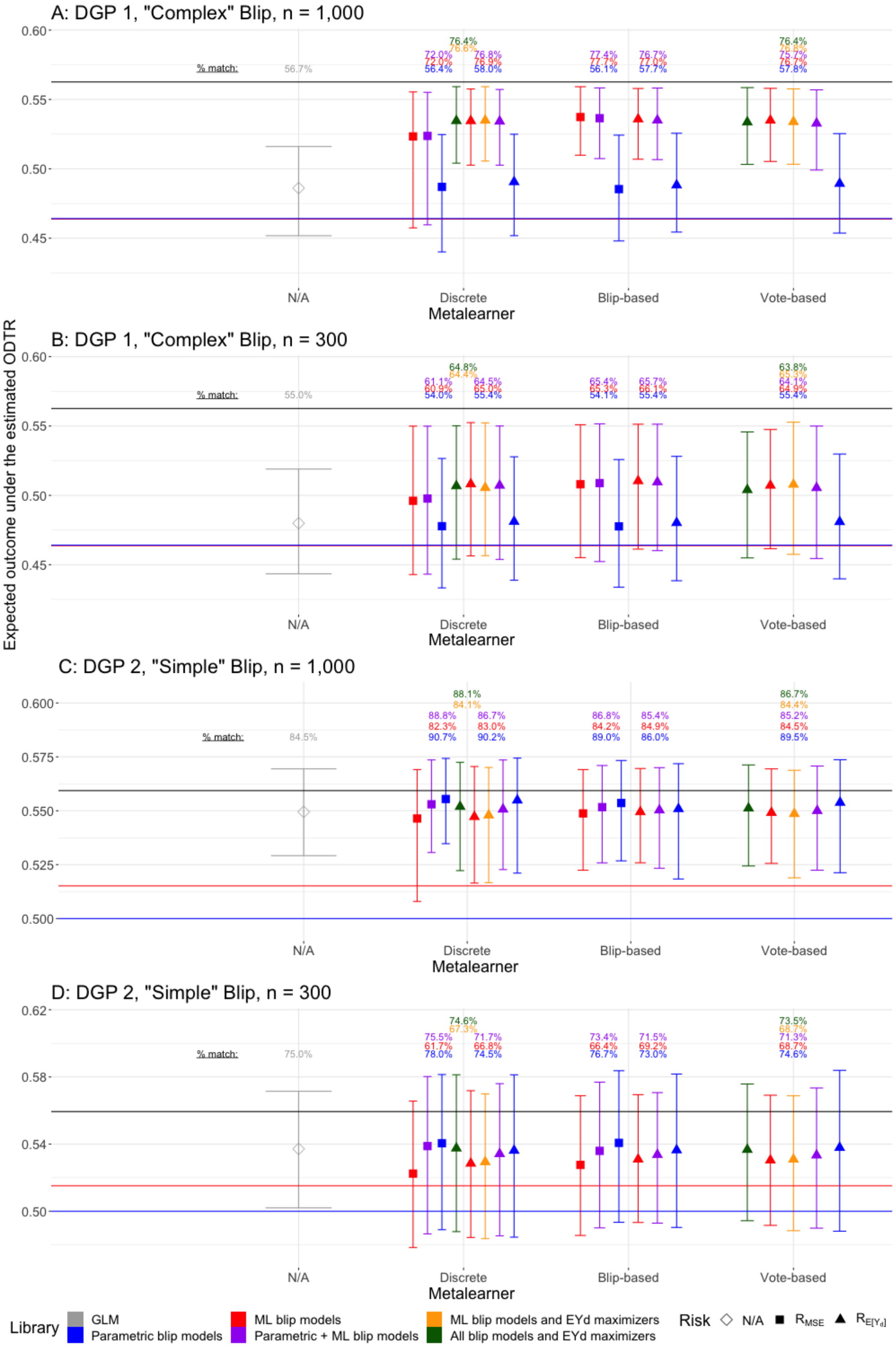

Figure 1.1: (Description on the following page.)

Figure 1.1: Performance of $E_n[Q_0(Y|A = d_n^*, W)]$ (an approximation of the true mean outcome under the estimated ODTR) for DGP 1 (top two) and DGP 2 (bottom two). The horizontal black line depicts $E_{P_{U,X}}[Y_{d_0^*}]$; the horizontal red line depicts $E_{P_{U,X}}[Y_1]$; the horizontal blue line depicts $E_{P_{U,X}}[Y_0]$. We compare the SuperLearner algorithm to an incorrectly specified GLM (in gray, with N/A as the metalearner and a diamond with no fill). We also compare (1) having a SuperLearner library with (a) only algorithms that estimate the blip (i.e., "Blip only" libraries) that only have parametric blip models (blue) or only have machine-learning blip models (red) or both (purple) versus (b) an expanded or "Full" library with blip function regressions estimated via machine learning only (orange-yellow) or machine learning and parametric models (green), with methods that directly estimate the ODTR and static rules, (2) having a metalearner (depicted on the x-axis) either that chooses one algorithm (i.e., the "discrete" SuperLearner) or combines blip predictions/treatment predictions (i.e., the "continuous" SuperLearner), and (3) using the MSE risk function (R_{MSE} as a square) versus the mean outcome under the candidate rule risk function ($R_{E[Y_d]}$ as a triangle).

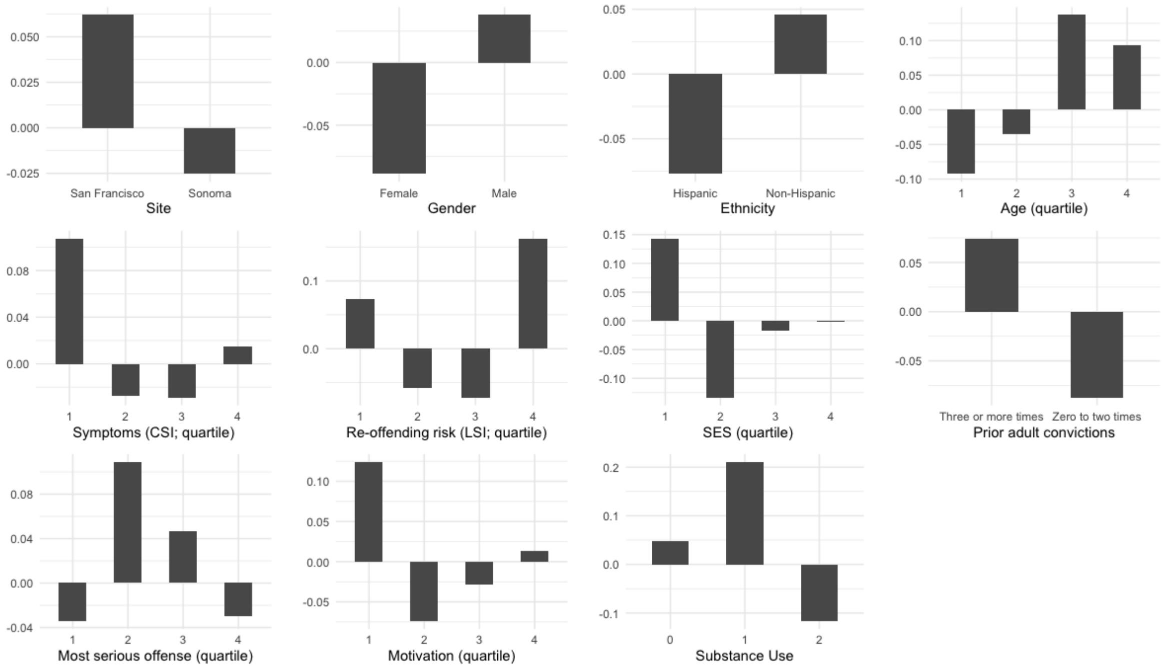

Figure 1.2: Subgroup plots for each of the covariates for the "Interventions" data. The x-axis for each of the plots is the different levels of the covariates; the y-axis is the difference in proportion of people who were not re-arrested between those who received CBT versus TAU, in that covariate subgroup.

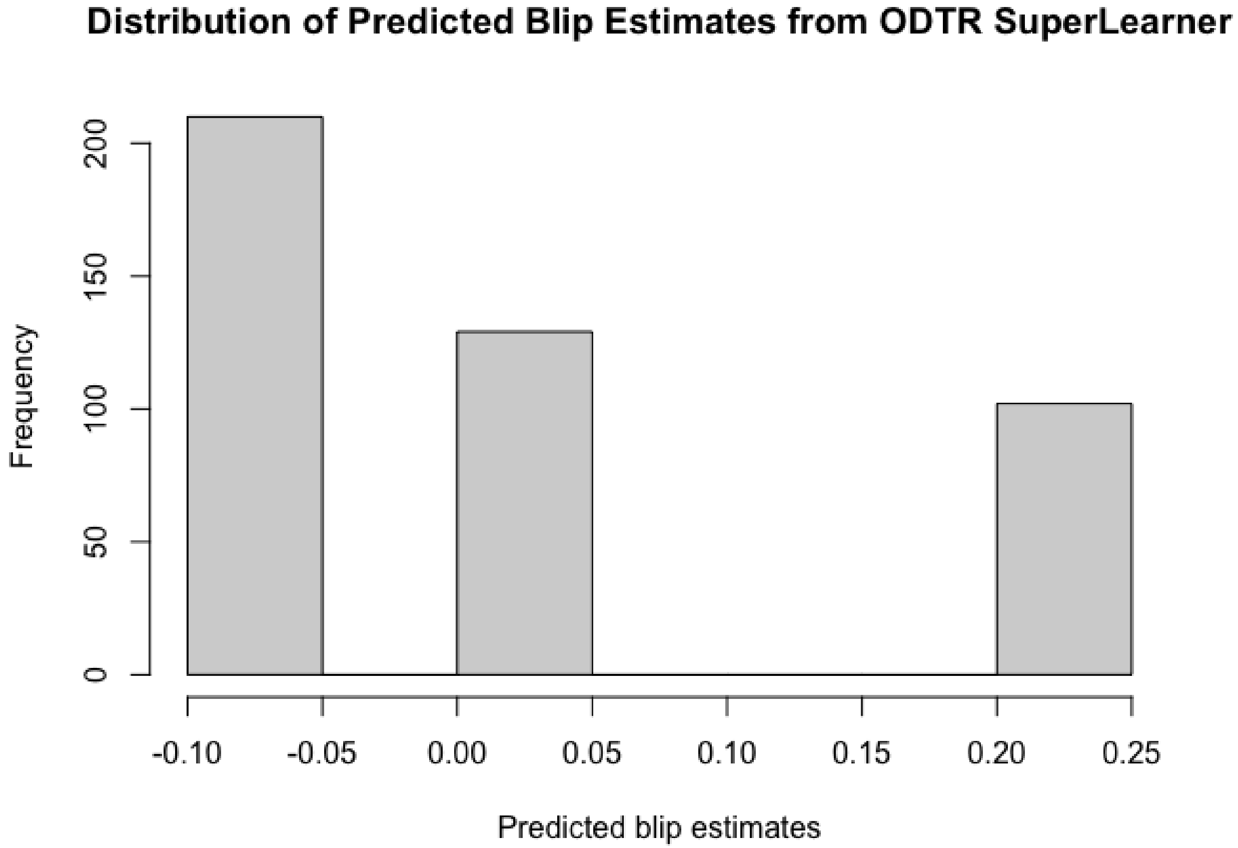

Figure 1.3: Distribution of predicted blip estimates from the ODTR SuperLearner. The frequencies are divided into three groups because the ODTR SuperLearner allocated all coefficient weights to a GLM using substance use, a variable with only 3 treatment levels. One can interpret the ODTR SuperLearner for this sample as follows: CBT may reduce the probability of re-arrest among justice-involved adults with low levels of substance use. Estimation and inference of the value of the ODTR SuperLearner compared to, for example, treating everyone or no one, informs us if there is, in fact, a differential effect by substance use, and thus a benefit to assigning CBT in this individualized way.

Chapter 2

Performance and Application of Estimators for the Value of an Optimal Dynamic Treatment Rule

2.1 Introduction

There is an interest across disciplines in using both experiments and observational data to uncover treatment effect heterogeneity and understand better ways of responding to it [26, 40]. Various methods aimed at estimating heterogenous treatment effects (HTEs) wish to answer the question, "who benefits from which treatment?" One way to uncover HTEs is by using the dynamic treatment rule framework. A dynamic treatment rule is any rule that assigns treatment based on covariates [2, 38, 67, 8, 11]. An optimal dynamic treatment rule (ODTR) is the dynamic treatment rule that yields the highest expected outcome (if higher outcomes are better) [52, 69, 49]. Using data generated from an experiment in which treatment is randomized makes identification of the ODTR more straightforward due to elimination of confounding. In recent years, there has been a increase in literature describing methods to estimate the ODTR, from regression-based techniques to direct-search techniques; see, for example, [28], [29], and [84] for recent overviews of the ODTR literature. One example of a data-adaptive method for estimating the ODTR is the SuperLearner algorithm, an ensemble machine learning approach that aims to best combine a library of candidate treatment rule estimators to work in tandem to yield the ODTR [35, 48, 15].

Once one knows or estimates a rule, it may be of interest to *evaluate* it, which translates to asking the causal question: what is the expected outcome had every person received the treatment assigned to him or her by the (optimal) rule? The causal parameter that answers this question is sometimes referred to as the *value* of the rule. It may be of relevance to learn this quantity in order to determine the benefit of assigning treatment in a more complex way compared to, for example, simply giving everyone treatment (an intervention that is straightforward to implement without the cost or complexity of measuring covariates and personalizing treatment assignment).

In this paper, we examine the following causal parameters, which we identify as statistical estimands, corresponding to the value of an (optimal) rule: 1) the true expected outcome of a given *a priori* known dynamic treatment rule; 2) the true expected outcome under the true, unknown ODTR; and 2) the true expected outcome under the *estimated* ODTR, a so-called "data-adaptive parameter." The latter parameter can be further split into the true expected outcome under a) an ODTR estimated on the entire data at hand, or b) a sample-split specific ODTR, in which, under a cross-validation scheme, the ODTR is estimated on each training set and evaluated, under the true data-generating distribution, on the complementary validation set, with the data-adaptive parameter defined as an average across sample splits.

We discuss several estimators for these estimands. Specifically, we consider the following estimators suited for estimating a treatment-specific mean: the simple substitution estimator of the G-computation formula [67], the inverse probability of treatment weighted (IPTW) estimator [20, 73], the double-robust IPTW estimator (IPTW-DR) [72, 79, 70], the targeted maximum likelihood estimator (TMLE) [2, 75, 37], and the cross-validated TMLE (CV-TMLE) [93, 33, 36].

First, we review the conditions under which asymptotic linearity is achieved for these estimators in the scenario where one wants to evaluate an *a priori* known rule. This provides insight into the common scenario in which one wishes to evaluate the value of a dynamic treatment rule that is pre-specified (based on investigator knowledge or external data sources), rather than learned from the data at hand. Estimators for this parameter require fast enough convergence rates and smoothness assumptions on nuisance parameters, though smoothness assumptions can be relaxed when employing CV-TMLE.

Second, we examine the more ambitious goal of estimating the expected outcome under the true, unknown ODTR, which additionally requires fast enough convergence of the estimate of the ODTR to the true ODTR, and for non cross-validated estimators, smoothness assumptions on ODTR estimators. We refer the reader to the previous chapter and [48] for a discussion of considerations and best practices when implementing the ODTR SuperLearner. Obtaining inference for the mean outcome under the ODTR has been shown to be difficult due to its lack of smoothness [8, 69, 39]; however, several methods have been proposed for constructing valid confidence intervals for this parameter, such as re-sampling techniques [10, 81, 8]. One approach to inference is to rely on parametric models; however, misspecification of these models can bias results. CV-TMLE relaxes the smoothness assumptions needed for inference, allowing one to use a single data set to safely estimate relevant parts of the data distribution (e.g., estimate nuisance parameters and/or the ODTR) and retain valid inference for the target parameter itself (e.g., the mean outcome under the ODTR) [23, 32]. Such internal sample splitting is particularly important if the nuisance parameters or ODTR depend on a high dimensional covariate set or make use of data-adaptive methods [33].

Finally, it may instead be of interest to estimate the true outcome under an estimated ODTR (a data-adaptive parameter) because, in practice, it is the estimated rule that will likely be employed in the population, not the true rule, which is likely unknown [33]. In this case, the only rate condition needed on the estimate of the ODTR is that it converges to a fixed rule. Non-cross-validated estimators of this data-adaptive parameter additionally require smoothness assumptions on the estimate of the ODTR for asymptotic linearity; the CV-TMLE eliminates these requirements, which means that, in a randomized experiment, achievement of asymptotic linearity for CV-TMLE with respect to this data-adaptive parameter only requires that the estimated ODTR converges to a fixed rule [33].

Previous simulation experiments have studied the performance of different estimators for the aforementioned statistical estimands in the setting in which a binary treatment is randomly assigned at a single time point. [33] demonstrated the importance of using an estimator of the value of the rule that uses a targeted bias reduction, such as TMLE and CV-TMLE, in order to improve performance. Of note, when evaluating the estimated rule, the authors used the true treatment mechanism and, as an initial estimate of the outcome regression, either the true outcome regression or a constant value (i.e., an incorrectly specified outcome regression) when employing the (CV-)TMLE. [15] extended these results by "fully" estimating the value of the optimal rule, meaning the nuisance parameters were additionally estimated for both the optimal rule and the value of the rule, using the ensemble machine learning approach SuperLearner [35]. Both [33] and [15] found that, indeed, there exists a positive finite sample bias when using TMLE versus CV-TMLE when estimating the value of the ODTR; in other words, with the rule learned and evaluated on the same data, estimates of the value of the rule may be optimistic, and CV-TMLE corrects this bias. Additionally, recently, [81] showed that cross-validation techniques for estimating the value of the rule, and in particular CV-TMLE, yielded a smaller difference between the true expected value under the true rule and its estimate, versus, for example, bootstrap techniques for evaluating a rule.

The current paper builds on previous work by illustrating, through a simulation study, how the degree of overfitting when estimating the optimal rule and/or nuisance parameters affects the performance of the estimators used for evaluating a rule. We also explore the potential for efficiency improvement and bias reduction through the use of semiparametric efficient estimators, with and without targeting. Finally, we show the importance of sample splitting using CV-TMLE when estimating the aforementioned statistical parameters.

Additionally, we apply these estimators of the value of the rule to the Correctional Intervention for People with Mental Illness, or "Interventions," trial, a ongoing study in which criminal justice-involved adults with mental illness – a heterogeneous group with diverse symptoms, risk factors, and other treatment-relevant characteristics [82, 83] – are either randomized to cognitive behavioral therapy (CBT) or treatment as usual (TAU), and re-arrest is collected one year after randomization occurs, as a measure of recidivism. In a companion paper (the previous chapter), we estimated the ODTR using the ODTR SuperLearner algorithm to identify which patients should receive CBT versus TAU. In this paper, we use CV-TMLE to determine whether administering CBT using the estimated ODTR is more effective in reducing recidivism than assigning CBT in a non-individualized way (for example, giving CBT to all offenders).

This article steps through the causal roadmap for answering causal questions [57], and is thus organized as follows. In the first section, we define the data and causal model, define the causal parameters as functions of the counterfactual distribution, and identify the causal estimands as functions of the observed data distribution. In section 2 we discuss estimation, and in section 3 we discuss inference procedures and conditions for asymptotic linearity. In section 4 we present a simulation study illustrating the performance of these estimators. In section 5 we evaluate the ODTR SuperLearner algorithm that was applied to the "Interventions" Study. Finally, we close with a discussion and future directions.

2.2 Causal Roadmap

In this section, we follow the first steps of the roadmap [57] for answering the causal questions: what would have been the expected outcome had everyone been given treatment according to: 1) any given rule; 2) the true ODTR; and 3) an estimate of the ODTR, which could either be a) a sample-specific estimate of the ODTR (i.e., an ODTR estimated on the entire data), or b) a sample-split-specific estimate of the ODTR?

Data and Models

Structural causal models (SCM) will be used to describe the process that gives rise to variables that are observed (endogenous) and not observed (exogenous). The random variables in the SCM (denoted $\mathcal{M}^F$) follow the joint distribution $P_{U,X}$. The endogenous variables are the covariates $W \in \mathcal{W}$, binary treatment $A \in \mathcal{A} = \{0, 1\}$, and outcome $Y \in \mathbb{R}$. Exogenous variables are denoted $U = (U_W, U_A, U_Y)$. The following structural equations illustrate dependency between the variables:

$$\begin{aligned} W &= f_W(U_W) \\ A &= f_A(U_A, A) \\ Y &= f_Y(U_Y, A, W). \end{aligned}$$

Because we will be focusing on data where treatment is randomly assigned (as in the "Interventions" trial), the above model can be modified by letting $U_A \sim Bernoulli(p = 0.5)$ and $A = U_A$.

We assume the observed data $O_i \equiv (W_i, A_i, Y_i) \sim P_0 \in \mathcal{M}$, $i = 1, \ldots, n$, where P_0 is the observed data distribution and $\mathcal{M}$ is the statistical model, were generated by sampling n i.i.d. times from a data-generating system contained in the SCM $\mathcal{M}^F$ above.

The likelihood of O can be factorized as $\mathcal{L}_0(O) = p_{W,0}(W)g_0(A|W)p_{Y,0}(Y|A, W)$, where, $p_{W,0}$ is the true density of W, $g_0(A|W)$ is the true conditional probability of the treatment A, and $p_{Y,0}$ is the true conditional density of Y.

The empirical distribution P_n gives each observation weight $\frac{1}{n}$; $P_n \in \mathcal{M}_{NP}$, where $\mathcal{M}_{NP}$ is a non-parametric statistical model. Estimates from this empirical distribution are denoted with a subscript n. If V-fold cross-validation is employed, the empirical data are uniformly and at random split into V mutually exclusive sets. For sets $v \in \{1, ..., V\}$, each set of data serves as a validation set; the complement is its training set. Let $P_{n,v}$ be the empirical distribution of the validation sample v, and $P_{n,-v}$ be the empirical distribution of the complementary training set.

Data and Models - Application to "Interventions" Study

The "Interventions" Study is an RCT consisting of 441 i.i.d. observations of the following data generated by the causal model described above: covariates W, which includes intervention site, sex, ethnicity, age, Colorado Symptom Index (CSI) score (a measure of psychiatric symptoms), level of substance use, Level of Service Inventory (LSI) score (a measure of risk for future re-offending), number of prior adult convictions, most serious offense, Treatment Motivation Questionnaire (TMQ) score (a measure of internal motivation for undergoing treatment), and substance use level; the randomized treatment A, which is either a manualized Cognitive Behavioral Intervention for people criminal justice system (abbreviated CBT; $A = 1$) or treatment as usual (TAU), which is mostly psychiatric or correctional services ($A = 0$); and a binary outcome Y of recidivism, an indicator that the person was not re-arrested over a minimum period of one year. Table 3.1 shows the distribution of the data.

Causal Estimands

In this point treatment setting, a dynamic treatment rule in the set of rules $\mathcal{D}$ is a function d that takes as input some function V of the measured baseline covariates W and outputs a treatment decision: $V \rightarrow d(V) \in \{0, 1\}$. It could be the case that $V = W$, in other words, dynamic treatment rules that potentially respond to all measured baseline covariates.

Counterfactual outcomes under a treatment rule d – or a subject's outcome if, possibly contrary to fact, the subject received the treatment that would have been assigned by the treatment rule d – are derived by intervening on the above SCM. Specifically, in parallel with our causal questions above, counterfactual outcomes are generated by setting A equal to the following treatment rules, all in the set $\mathcal{D}$: 1) the true ODTR d_0^*; and, 2) an estimate of the ODTR, either: a) the sample-specific estimate of the ODTR d_n^*; or b) the training sample-specific estimate of the ODTR $d_{n,v}^*$.

The expectation of each of these counterfactual outcomes under the distribution $P_{U,X}$ are the causal parameters of interest in this paper. Each causal estimand is a mapping $\mathcal{M}^F \rightarrow \mathbb{R}$.

The target causal parameter corresponding to the value of a given treatment rule d (from the set of rules $\mathcal{D}$) is:

$$\Psi_d^F(P_{U,X}) \equiv \mathbb{E}_{P_{U,X}}[Y_d].$$

The true ODTR d_0^* is defined as the rule that maximizes the expected counterfactual outcome:

$$d_0^* \in \arg\max_{d \in \mathcal{D}} \Psi_d^F(P_{U,X}).$$

Here, the target causal parameter of interest is the expected outcome under the true ODTR d_0^*:

$$\Psi_{d_0^*}^F(P_{U,X}) \equiv \mathbb{E}_{P_{U,X}}[Y_{d_0^*}].$$

Let $d_n^* : \mathcal{M}_{NP} \to \mathcal{D}$ be an ODTR estimated on the entire sample, and $d_{n,v}^* = d^*(P_{n,-v}) : \mathcal{M}_{NP} \to \mathcal{D}$ be an ODTR estimated on the v^{th} training set. The data-adaptive causal parameters are: a) the expected outcome under a sample-specific estimate of the ODTR:

$$\Psi_{d_n^*}^F(P_{U,X}) \equiv \mathbb{E}_{P_{U,X}}[Y_{d_n^*}],$$

noting that the expectation here is not over d_n^*, i.e., this is $\mathbb{E}_{P_{U,X}}[Y_d]$, with $d = d_n^*$, and b) the average of the expected validation set outcomes under training-set specific estimates of the ODTR:

$$\Psi_{d_{n,CV}^*}^F(P_{U,X}) \equiv \frac{1}{V}\sum_{v=1}^{V} \mathbb{E}_{P_{U,X}}[Y_{d_{n,v}^*}].$$

One might also be interested in comparing the above causal quantities to, for example, the expected outcome had everyone been assigned the treatment $\mathbb{E}_{P_{U,X}}[Y_1]$ or had no one been assigned the treatment $E_{P_{U,X}}[Y_0]$.

Causal Estimands - Application to "Interventions" Study

Analagous to the above causal questions, for the "Interventions" Study, we are interested in asking: what would have been the probability of no re-arrest had everyone been given CBT according to: 1) some pre-specified rule d (for example, the simple dynamic treatment rule that gives CBT to those with high levels of prior education and TAU to those with low levels of prior education), where the causal parameter is $\Psi_d^F(P_{U,X})$; 2) the true ODTR d_0^* (the unknown dynamic treatment rule for assigning CBT that yields the highest probability of no re-arrest), where the causal parameter is $\Psi_{d_0^*}^F(P_{U,X})$; and 3) an estimate of the ODTR specific to the 441 participants in the trial, which could either be a) a sample-specific estimate d_n^* (e.g., the ODTR estimated in the previous chapter) or b) a sample-split-specific estimate of the ODTR $d_{n,CV}^*$? The causal parameters for a) and b) are $\Psi_{d_n^*}^F(P_{U,X})$ and $\Psi_{d_{n,CV}^*}^F(P_{U,X})$, respectively.

Identification

Two assumptions are necessary for identification; that is, for determining that the causal estimands (a function of our counterfactual distribution) are equal to the statistical estimands (a function of our observed data distribution): the 1) randomization assumption, $Y_a \perp A|W \; a \in \{0,1\}$; and 2) the positivity assumption: $Pr(\min_{a\in\{0,1\}} g_0(A = a|W) > 0) = 1$. Both hold if, for example, data are generated from an experiment in which treatment is randomized (as in the "Interventions" trial); for data generated in an observational setting, the randomization assumption requires measurement of all unmeasured confounders, and the positivity assumption should be examined [58].

Statistical Estimands

We describe statistical estimands corresponding to each of the causal parameters outlined above – each is identified via the G-computation formula which corresponds to a mapping $\mathcal{M} \to \mathbb{R}$.

The statistical estimand of the mean outcome under any rule $d \in \mathcal{D}$ is

$$\psi_{0,d} \equiv \Psi_d(P_0) = \mathbb{E}_0[Q_0(d(W), W)],$$

where the function $Q(A, W) = \mathbb{E}[Y|A, W]$ is the outcome regression.

The true optimal rule, as a function of the observed data distribution, is then:

$$d_0^* \in \arg\max_{d\in\mathcal{D}} \Psi_d(P_0).$$

Note that the RHS of this equation is a set because there may be more than one optimal rule for a certain kind of subject (e.g., if certain kinds of subjects neither benefit from nor are harmed by a treatment) [47, 69, 68]. Here, we will assume that when there is no treatment effect, assigning treatment 0 is better than no treatment. Then, the optimal rule can be written as a function of the so-called "blip function", $B(W) = Q(1, W) - Q(0, W)$:

$$d_0^*(W) = \mathbb{I}[B_0(W) > 0].$$

The true mean outcome under the true optimal rule d_0^* is then identified by

$$\psi_{0,d_0^*} \equiv \Psi_{d_0^*}(P_0) = \mathbb{E}_0[Q_0(d_0^*(W), W)].$$

The first data-adaptive parameter we consider, as a function of the observed data, is the true expected outcome under the ODTR estimated on the entire sample d_n^*:

$$\psi_{0,d_n^*} \equiv \Psi_{d_n^*}(P_0) = \mathbb{E}_0[Q_0(d_n^*(W), W)].$$

The second data-adaptive parameter is the average of the validation-set true mean outcomes under the training-set estimated ODTRs $d_{n,v}^*$:

$$\psi_{0,d_{n,CV}^*} \equiv \Psi_{d_{n,CV}^*}(P_0) = \frac{1}{V}\sum_{v=1}^{V} \mathbb{E}_0[Q_0(d_{n,v}^*(W), W)].$$

2.3 Estimation

We describe estimators for each of the statistical parameters above: a simple substitution estimator based on the G-computation formula, an IPTW estimator, a double-robust IPTW estimator (IPTW-DR), a TMLE, and a CV-TMLE. Each of these estimators can be used for estimating $\psi_{0,d}$ and ψ_{0,d_0^*}. We use the non-cross-validated estimators (G-computation, IPTW, IPTW-DR, and TMLE) to estimate ψ_{0,d_n^*}; we estimate $\psi_{0,d^*_{n,CV}}$ with CV-TMLE.

Estimators of these parameters are mappings $\hat{\Psi} : \mathcal{M}_{NP} \rightarrow \mathbb{R}$. For all estimators, let Q_n be an estimator of the outcome regression, which could be estimated with, for example, SuperLearner [35]. In a randomized experiment, the treatment mechanism g_0 is known; thus, one could use this known g_0, or g_n could be a maximum likelihood estimator (MLE) based on a correctly specified model.

We first illustrate each of the non-cross-validated estimators suited for estimating a treatment-specific mean at an arbitrary $d \in \mathcal{D}$, which, for example, could be an *a priori* known rule or an optimal rule estimated on the entire sample d_n^* (see papers from [48] and Chapter 1 for a description on how to estimate the optimal rule using, for example, the ODTR SuperLearner). Here, $\hat{\Psi}_d$ is an estimator of $\psi_{0,d}$; the estimate is denoted $\hat{\Psi}_d(P_n) \equiv \hat{\psi}$. We further subscript by each estimator name.

One can use a simple substitution estimator based on the above G-computation formula [67]:

$$\hat{\psi}_{gcomp,d} = \frac{1}{n}\sum_{i=1}^{n} Q_n(d(W_i), W_i),$$

an IPTW estimator [20, 73]:

$$\hat{\psi}_{IPTW,d} = \frac{1}{n}\sum_{i=1}^{n} \frac{\mathbb{I}[A_i = d(W_i)]}{g_n(A_i|W_i)} Y_i,$$

a double-robust IPTW estimator [72, 79, 70]:

$$\hat{\psi}_{IPTW-DR,d} = \frac{1}{n}\sum_{i=1}^{n} \frac{\mathbb{I}[A_i = d(W_i)]}{g_n(A_i|W_i)}(Y_i - Q_n(A_i, W_i)) + Q_n(d(W_i), W_i),$$

or a TMLE [75, 37, 2, 33]. We briefly describe one possible TMLE procedure. First, estimate the clever covariate for each person:

$$H_{n,i} = \frac{\mathbb{I}[A_i = d(W_i)]}{g_n(A_i|W_i)}.$$

Then, update the initial fit of $Q_n(d(W), W)$ by running a logistic regression of Y (which should be transformed between 0 and 1 if the outcome is continuous [19]) on offset denoted as $Q_n(d(W), W)$ with weights H_n, predicting at $A = d(W)$. Denote the updated fit as $Q_n^*(d(W), W)$. Then, the TMLE estimator is:

$$\hat{\psi}_{TMLE,d} = \frac{1}{n}\sum_{i=1}^{n} Q_n^*(d(W_i), W_i).$$

As previously mentioned, the CV-TMLE can estimate $\psi_{0,d}$, ψ_{0,d_0^*}, and $\psi_{0,d_{n,CV}^*}$ [23, 36, 33, 93, 32]. Instead of illustrating this estimator at d as in the above estimators, we illustrate one type of CV-TMLE procedure for evaluating the mean outcome under sample-split-specific estimates of the ODTR $d_{n,v}^*$ to show on which parts of the data one needs to estimate or predict the ODTR, if estimating ψ_{0,d_0^*} or $\psi_{0,d_{n,CV}^*}$. The same procedure holds for a d that is known, except that the rule need not be estimated on each of the training samples and is simply applied to the validation sets:

1. Split the data into V folds. Let each fold be the validation set and the complement data be the training set.

2. Generate initial estimators of g_0, Q_0, and d_0^* based on the training sample $P_{n,-v}$.

3. Predict the training-set specific fits from the previous step on the validation sample $P_{n,v}$.

4. Using the predictions from the previous step, in each corresponding validation set, update the initial estimator $\hat{\Psi}_{d_{n,v}^*}(P_{n,-v})$ using the TMLE procedure described above to generate $\hat{\Psi}_{d_{n,v}^*}(P_{n,-v}^*)$, a TMLE of $\mathbb{E}_0[Q_0(d_{n,v}^*(W), W)]$.

5. Average over all validation folds to obtain the CV-TMLE, i.e., the estimated mean outcome under the training-sample-split specific estimates of the rules:

$$\hat{\psi}_{CV-TMLE,d_{n,v}^*} = \frac{1}{V}\sum_{v=1}^{V} \hat{\Psi}_{d_{n,v}^*}(P_{n,-v}^*).$$

2.4 Inference

We first discuss the conditions necessary for each the above estimators to be asymptotically linear for $\psi_{0,d}$, ψ_{0,d_n^*}, and $\psi_{0,d_{n,CV}^*}$ in a randomized experiment. Under these conditions, using influence-curve based inference, we describe how to construct 95% confidence intervals with nominal to conservative coverage for the aforementioned statistical estimands of interest.

We do not discuss inference on the G-computation estimator, because in order for it to be asymptotically linear, Q_n must either be equal to Q_0 or be an estimator that converges fast enough to Q_0, neither of which we assume here.

Asymptotic Linearity Conditions for Estimators

We give a brief overview of the conditions needed for asymptotic linearity for each of the estimators with respect to each statistical estimand in the randomized trial setting, and provide a summary of these conditions in Table 2.1. For more details and proofs, we refer the reader to [37] and [33].

An estimator $\hat{\Psi}$ is asymptotically linear for its true value ψ_0 if can be written in the following form:

$$\hat{\psi} - \psi_0 = \frac{1}{n}\sum_{i=1}^{n} IC(O_i) + R_n,$$

where IC is the estimator's influence curve and R_n is a remainder term that is $o_P(1/\sqrt{n})$. An asymptotically linear estimator $\hat{\Psi}$ thus has the following properties: 1) its bias converges to 0 in sample size at a rate faster than $\frac{1}{\sqrt{n}}$; 2) for large n, its distribution is normal, $n^{1/2}(\hat{\psi} - \psi_0) \xrightarrow{d} N(0, \sigma_0^2)$, allowing an estimate of σ_0^2 to be used to construct a Wald-type confidence intervals; and, 3) the asymptotic variance of $n^{1/2}(\hat{\psi} - \psi_0)$ (i.e., σ_0^2) can be well-approximated by the sample variance of its estimated influence curve (or equivalently, $\sigma_n^2 = \frac{1}{n}\sum_i^n IC_n^2(O_i)$, since the mean of an influence curve is 0).

The current randomized experiment scenario guarantees that g_0 is known; here, we consider the case where g_n is an estimate of g_0 based on a correctly specified parametric model. Given this, for an estimand defined as the value of an *a priori* specified rule d, the IPTW estimator is guaranteed to be asymptotically linear for $\psi_{0,d}$; however, it will not be asymptotically efficient. Under a Donsker class assumption on the estimator Q_n, IPTW-DR and TMLE are guaranteed to be asymptotically linear for ψ_0 (due to R_n involving a second-order term that is the product of the difference between Q_n and g_n for Q_0 and g_0, respectively); if Q_n is a consistent estimator of Q_0 with a rate of convergence faster than $1/\sqrt{n}$, IPTW-DR and TMLE are asymptotically efficient. This is also true for CV-TMLE, except Donsker class conditions can be relaxed (in effect allowing for an overfit in the initial estimate of Q_0) [33].

Construction of nominal to conservative confidence intervals around each estimator with respect to the true expected outcome under the true, unknown d_0^* requires additional assumptions. For all estimators, statistical inference for ψ_{0,d_0^*} relies on a second-order difference in R_n between ψ_{0,d_n^*} and ψ_{0,d_0^*} going to 0 at a rate faster $1/\sqrt{n}$. In practice, how hard it is to make this condition hold depends on the extent to which the blip function has density at zero. If the value of the blip is always larger than $\delta > 0$ for some $\delta > 0$, then consistency of Q_n is sufficient; however, if the treatment effect is zero for some covariate levels, then stronger assumptions are required. The non-cross-validated estimators additionally require Donsker conditions on d_n^*. In practice, these conditions on the data-adaptivity of d_n^* hold if, for example, the optimal rule is a function of one covariate, or, if a higher-dimensional covariate set is used, one is willing to make strong smoothness assumptions on, for example, the blip function [33, 48, 31]. CV-TMLE relaxes these Donsker conditions on d_n^*. Thus, in a randomized trial, if employing CV-TMLE for this estimand, the only condition needed is that d_n^* converges fast enough to d_0^*.

For the data-adaptive parameters, the estimators no longer require the strong assumption that d_n^* converges to d_0^* at a certain rate; rather, they only require that d_n^* converges to some fixed rule $d \in \mathcal{D}$ at any rate [33]. This means that, for randomized trial data, the CV-TMLE estimator for $\psi_{0,d^*_{n,CV}}$ is asymptotically linear under essentially no conditions [33].

		Conditions for Asymptotic Linearity:				
Estimands	**Estimators**	$g_n = g_0$ **or** $g_n \xrightarrow{p} g_0$	$Q_n = Q_0$ **or** $Q_n \xrightarrow{p} Q_0$	$\psi_{0,d_n^*} - \psi_{0,d_0^*} = \mathbf{o}_P(\frac{1}{\sqrt{n}})$	Q_n **not overfit**	d_n^* **not overfit**
Value of known rule $\psi_{0,d}$	$\hat{\Psi}_{IPTW,d}$	Satisfied by randomized experiment	Not required	Not required, d known	Not required	Not required, d known
	$\hat{\Psi}_{IPTW-DR,d}$				Required	
	$\hat{\Psi}_{TMLE,d}$				Required	
	$\hat{\Psi}_{CV-TMLE,d}$				Not required	
Value of true ODTR ψ_{0,d_0^*}	$\hat{\Psi}_{IPTW,d_n^*}$			Required	Not required	Required
	$\hat{\Psi}_{IPTW-DR,d_n^*}$				Required	
	$\hat{\Psi}_{TMLE,d_n^*}$				Required	
	$\hat{\Psi}_{CV-TMLE,d_{n,v}^*}$				Not required	Not required
Value of sample-specific ODTR estimate ψ_{0,d_n^*}	$\hat{\Psi}_{IPTW,d_n^*}$			Not required; require $d_n^* \xrightarrow{p} d \in \mathcal{D}$	Not required	Required
	$\hat{\Psi}_{IPTW-DR,d_n^*}$				Required	
	$\hat{\Psi}_{TMLE,d_n^*}$				Required	
Value of sample-split-specific ODTR estimate $\psi_{0,d^*_{n,CV}}$	$\hat{\Psi}_{CV-TMLE,d_{n,v}^*}$			Not required; require $d_n^* \xrightarrow{p} d \in \mathcal{D}$	Not required	Not required

Table 2.1: Summary of the conditions needed for asymptotic linearity in the randomized treatment setting for each of the estimators corresponding to each of the estimands.

Construction of Confidence Intervals

Below, we list conservative working influence curves for each estimator at P_n and $d \in \mathcal{D}$. The actual estimators' influence curves when an MLE of g_n based on a correctly specified parametric model is used (as can be guaranteed when treatment is randomized) are the working influence curves presented below minus a tangent space projection term [37, 31]. Thus, under the conditions stated above, the sample variance of the following working influence curves at a correctly specified g_n yield conservative estimates of the asymptotic variance of the estimators, which yields conservative confidence interval coverage.

The IPTW estimator's working influence curve estimate is:

$$\widehat{IC}_{IPTW,d} = \frac{\mathbb{I}[A = d]}{g_n(A|W)} Y - \hat{\psi}_{IPTW,d}.$$

The influence curve of the TMLE and double-robust IPTW estimator is the *efficient* influence curve for the treatment-specific mean [31, 85, 4]; the corresponding working influence curve estimates are:

$$\widehat{IC}_{IPTW-DR,d} = \frac{\mathbb{I}[A=d]}{g_n(A|W)}(Y - Q_n(A,W)) + Q_n(d(W),W) - \hat{\psi}_{IPTW-DR,d},$$

$$\widehat{IC}_{TMLE,d} = \frac{\mathbb{I}[A=d]}{g_n(A|W)}(Y - Q_n^*(A,W)) + Q_n^*(d(W),W) - \hat{\psi}_{TMLE,d}.$$

As stated above, for these non-cross-validated estimators, the asymptotic variance can be conservatively estimated with the sample variance of the estimated influence curve: $\sigma_n^2 = \frac{1}{n}\sum_i^n \widehat{IC}^2(O_i)$.

For the IPTW-DR and TMLE estimators, one can underestimate the estimator's variance if Q_0 is estimated data-adaptively on the same data on which the sample variance of the estimated influence curve is evaluated. Through sample splitting, CV-TMLE confidence intervals protect against overfitting incurred by using the data twice – for both estimation and evaluation [37]. Then the fold-specific estimate of the working influence curve for CV-TMLE at the training-set-specific estimated ODTR is:

$$\widehat{IC}_{v,d^*_{n,v}} = \frac{\mathbb{I}[A_{-v} = d^*_{n,v}(W_{-v})]}{g_n(A_{-v}|W_{-v})}(Y_{-v} - Q^*_{n,v}(A_{-v},W_{-v})) + Q^*_{n,v}(d^*_{n,v}(W_{-v}),W_{-v}) - \hat{\Psi}(P^*_{n,v}),$$

and the fold-specific estimate of the variance of the fold-specific estimator is:

$$\sigma^2_{n,v} = \frac{1}{n_v - 1}\sum_{i=1}^{n_v} \widehat{IC}^2_{v,d^*_{n,v}}(O_i);$$

thus, the asymptotic variance of the CV-TMLE $\hat{\psi}_{CV-TMLE,d^*_{n,v}}$ can be conservatively estimated with:

$$\sigma^2_{n,CV-TMLE} = \frac{1}{V}\sum_{v=1}^{V}\sigma^2_{n,v}.$$

In sum, for each estimator $\hat{\Psi}$ and its corresponding working influence curve estimate IC_n, we obtain conservative inference on the value of the rule by constructing confidence intervals in the following way:

$$\hat{\psi} \pm \Phi^{-1}(0.975)\frac{\sigma_n}{\sqrt{n}}.$$

2.5 Simulation Study

Using simulations, we evaluate the performance of various estimators of the value of the rule in finite samples. In particular, we investigate: 1) the impact of increasingly data-adaptive estimation of nuisance parameters and (if applicable) the ODTR; 2) the potential for efficiency and bias improvement through the use of semiparametric efficient estimators; and, 3) the importance of sample splitting, in particular via a cross-validated-targeted maximum likelihood estimator (CV-TMLE).

Data Generating Process

All simulations were implemented in R [65], and the code, simulated data, and results can be found at https://github.com/lmmontoya/SL.ODTR. We examine these comparisons using the following data generating process (DGPs) (also used in the previous chapter and [33, 48]). Each simulation consists of 1,000 iterations of n=1,000. Mimicking a randomized experiment, the covariates, treatment and outcome are generated as follows:

$$\begin{aligned}
W_1, W_2, W_3, W_4 \sim & Normal(\mu = 0, \sigma^2 = 1)\\
A \sim & Bernoulli(p = 0.5)\\
Y \sim & Bernoulli(p)\ .\\
p = & 0.5 logit^{-1}(1 - W_1^2 + 3W_2 + 5W_3^2 A - 4.45A) +\\
& 0.5 logit^{-1}(-0.5 - W_3 + 2W_1 W_2 + 3|W_2|A - 1.5A)\ ,
\end{aligned}$$

then the true blip function is:

$$\begin{aligned}
B_0(W) = & 0.5[logit^{-1}(1 - W_1^2 + 3W_2 + 5W_3^2 - 4.45) +\\
& logit^{-1}(-0.5 - W_3 + 2W_1W_2 + 3|W_2| - 1.5)\\
& - logit^{-1}(1 - W_1^2 + 3W_2) + logit^{-1}(-0.5 - W_3 + 2W_1W_2)]\ .
\end{aligned}$$

Here, the true expected outcome under the true ODTR $\Psi^F_{d_0^*}(P_{U,X}) \approx 0.5626$ and the true optimal proportion treated $\mathbb{E}_{P_{U,X}}[d_0^*] \approx 55.0\%$. The mean outcome had everyone and no one been treated is, respectively, $\mathbb{E}_{P_{U,X}}[Y_1] \approx 0.4638$ and $\mathbb{E}_{P_{U,X}}[Y_0] \approx 0.4643$.

Estimator Configurations

We estimate each of the statistical estimands using the following estimators with inference based on the conservative working influence curves describe above: IPTW, IPTW-DR, TMLE, and CV-TMLE. The G-computation estimator is also employed, but confidence intervals are not generated.

A correctly specified logistic regression is used to estimate the nuisance parameter g_0. SuperLearner is used to estimate Q_0 and the ODTR [35, 48]. The ODTR is estimated using a "blip-only" library, using a blip-based metalearner (i.e., an approach to creating an ensemble of candidate ODTR algorithms), and selecting the mean outcome under the candidate rule as the risk function (see Chapter 1). Three libraries are considered that correspond to varying levels of data-adaptiveness, or potential for overfitting.

1. "GLMs - least data adaptive"
   - Q_n library: four logistic regressions, each with a main terms W_j and A, and with an interaction W_j times A, for $j \in \{1, .., 4\}$
   - d_n^* library: univariate linear regressions with each covariate
2. "ML + GLMs - moderately data adaptive"
   - Q_n and d_n^* library: all algorithms in the "GLMs - least data adaptive" Q_n and d_n^* libraries, respectively, in addition to the algorithms `SL.glm` (generalized linear models), `SL.mean` (the average), `SL.glm.interaction` (generalized linear models with interactions between all pairs of variables), `SL.earth` (multivariate adaptive regression splines [18]), `SL.nnet` (neural networks [66]), `SL.svm` (support vector machines [12]), and `SL.rpart` (recursive partitioning and regression trees [5]) from the SuperLearner package [62]
3. "ML + GLMs - most data adaptive"
   - Q_n and d_n^* library: all algorithms in the "ML + GLMs - moderately data adaptive" Q_n and d_n^* libraries, respectively, in addition to `SL.randomForest` [6]

Performance Metrics

Using measures of bias, variance, mean squared error (MSE) and 95% confidence interval coverage, we evaluate the ability of each of the estimators to approximate: 1) the true expected outcome under an *a priori* known rule d, i.e., $\psi_{0,d}$; 2) the true expected outcome under the true, unknown ODTR ψ_{0,d_0^*}; 3) the true expected outcome under an ODTR estimated on: a) the entire sample and evaluated on the entire sample ψ_{0,d_n^*}; or b) estimated on each of the training sets, evaluated and averaged over each of the validation sets $\psi_{0,d_{n,CV}^*}$.

First, we estimate the target parameter $\psi_{0,d}$. This illustrates the performance of these estimators of the value of a rule when the rule is known *a priori*, either because the rule is known to be of interest or it was estimated on other data not included in the current sample. In this case, we choose d to be the true ODTR, that is, $d = d_0^*$. We note that it is highly unlikely that in practice d_0^* is known *a priori*, and stress that the only reason we examine the performance of estimators $\hat{\psi}_{d=d_0^*}$ with respect to ψ_{0,d_0^*} is to illustrate how well these estimators evaluate a given pre-specified rule. However, illustrating this using the true rule d_0^* in a simulation facilitates comparison of estimator performance across estimands, showing, for example, the price in performance one pays for targeting the more ambitious parameter that seeks to estimate both the rule itself and its true value. Said another way, if we see that estimator performance for $\hat{\psi}_{d=d_0^*}$ with respect to ψ_{0,d_0^*} is good, then the only issue left with estimating ψ_{0,d_0^*} is estimating d_0^* well.

Next, we estimate the same target parameter ψ_{0,d_0^*} in the more realistic scenario where the true ODTR d_0^* is unknown. We therefore first estimate the ODTR and then apply each of the estimators of the value of the rule under the estimated ODTR (where the rule is either estimated on the entire sample $\hat{\psi}_{d_n^*}$ or, for CV-TMLE, estimated on each sample split $\hat{\psi}_{d_{n,v}^*}$). Performance of the estimators with respect to ψ_{0,d_0^*} reflects how well both the rule and its value are estimated.

Finally, we treat as target parameter the true expected outcome under the estimated optimal rule, i.e., the data-adaptive parameters ψ_{0,d_n^*} or, for CV-TMLE, $\psi_{0,d_{n,CV}^*}$. This illustrates estimator performance for data-adaptive parameters whose true values depend on the sample, and for which it is of interest to estimate their value using the same sample on which the rule was learned. Note that the target parameter value in this case is specific to the sample at hand (the "truth" will vary from sample to sample); thus, performance calculations are calculated with respect to the true sample-specific or sample-split specific mean outcome. For example, for confidence interval coverage, across the 1,000 simulations, we calculated the proportion of times the confidence interval around the estimated value of the estimated rule covered the true value of the estimated rule – where both the confidence interval around the estimate and the true value of the estimated rule are *specific to each sample.* Furthermore, the data-adaptive parameter will vary between the non-cross-validated estimators (whose data-adaptive parameter is the sample-specific parameter ψ_{0,d_n^*}) and CV-TMLE (whose data-adaptive parameter is the sample-split specific parameter $\psi_{0,d_{n,CV}^*}$), and as such, is not only a function of the sample, but also of the split.

Simulation Results

Results - Value of a Known Dynamic Treatment Regime

Bias, variance, MSE, and confidence interval coverage metrics for estimating $\psi_{0,d}$ in the scenario where d is known *a priori* illustrate the performance of each of the estimators for estimating the value of a given pre-specified rule; for illustration, we use the true optimal rule d_0^*. Thus, only estimation of nuisance parameters g and/or Q were needed for this parameter.

The untargeted G-computation formula exhibited considerable bias if either misspecified parametric models or a SuperLearning approach was used to estimate the outcome regression – regardless of the degree of data-adaptiveness in estimating this nuisance parameter Q. For example, when the Q_n library consisted of only parametric regressions, the mean difference between the G-computation estimate and the truth was -9.09% (i.e., 104.44-940.00 times that of the bias of alternative estimators). We note that this result is in contrast to that of estimating the treatment specific mean for any static regime, in which treatment assignment is not a function of covariates (e.g., $\mathbb{E}_0[Q_0(A = 1, W)]$) from data generated from a randomized experiment; in this case, the G-computation estimator under certain misspecified parametric models is a TMLE, and is therefore unbiased [76].

As expected, the IPTW estimator, although unbiased, was less efficient than the double robust estimators – specifically, throughout, the IPTW estimator's variance was 1.33-1.80 times that of the variance of alternative estimators. Additionally, the IPTW-DR and TMLE were unbiased (as expected, given the double-robustness of these estimators) if the outcome regression was estimated using either a misspecified parametric model or a SuperLearner with a less data-adaptive library. However, both estimators were biased (i.e., -0.84% and -0.75% bias for IPTW-DR and TMLE, respectively) with less than nominal confidence interval coverage (i.e., 90.1% and 90.6% coverage for IPTW-DR and TMLE, respectively) when a more data-adaptive library was used to estimate the outcome regression – a result likely due to overfitting Q_n.

Sample-splitting via CV-TMLE removed the non-cross-validated estimators' bias (-0.01%, or 0.001-0.17 times the bias relative to alternative estimators) and generated better confidence interval coverage (93.6%) under the presence of overfitting for Q_n, at no cost to variance.

Results - Value of the True, Unknown ODTR

No estimator performed well when both the ODTR itself and its value were estimated using the same sample (i.e., estimators $\hat{\psi}_{d_n^*}$ or $\hat{\psi}_{d_{n,v}^*}$ for ψ_{0,d_0^*}). This was evident particularly in terms of increased bias when a less data-adaptive library was used to estimate Q_0 and d_0^*, and in terms of both increased bias and variance when a more aggressive library was used to estimate Q_0 and d_0^*. Notably, however, CV-TMLE performed the best with respect to all performance metrics under the most data-adaptive approaches. A large component of the bias in this case was due to the rate of convergence from d_n^* to d_0^* for any SuperLearner library, and therefore confidence interval coverage of the true value under the true ODTR around any estimated value of the estimated rule did not approach 95% (confidence interval coverage under the least, moderately, and most data adaptive libraries ranged from 14.70%-45.0%, 66.50%-76.10%, and 31.00%-68.60%, respectively).

Although the focus of these simulations was not optimizing estimation of the ODTR, we note that, consistent with results from the previous chapter, the least biased estimators of the true value of the true ODTR are ones that use a combination of parametric models and machine learning algorithms in the estimation of Q_0 and d_0.

Results - Value of an Estimated ODTR

We evaluated the performance of the non-cross-validated estimators (IPTW, IPTW-DR, and TMLE, i.e., $\hat{\psi}_{d_n^*}$) of the data-adaptive parameter (i.e., ψ_{0,d_n^*}) – a parameter that depends on the optimal rule specific to the sample at hand. All non-cross-validated estimators overestimated the value of the rule (i.e., positive bias), regardless of the SuperLearner library. In addition, the bias increased as the library for estimating the ODTR became more data-adaptive. For example, for the most data-adaptive SuperLearner library configuration, TMLE exhibited a bias of 13.46%, variance of 0.0108, MSE of 0.0307, and 15.7% confidence interval coverage.

The CV-TMLE (i.e., $\hat{\psi}_{CV-TMLE,d^*_{n,v}}$) with respect to the data-adaptive parameter $\psi_{0,d^*_{n,CV}}$ removed the bias incurred by estimating and evaluating the ODTR on the same sample, at little cost to no cost to variance. For example, for the most data-adaptive SuperLearner library configuration, CV-TMLE had a bias of 0.04% (0.001-0.0006 times that of alternative estimators), variance of 0.0007 (0.07-1.00 times that of alternative estimators), MSE of 0.0005 (0.01-0.06 times that of alternative estimators), and 94.8% confidence interval coverage.

2.6 Evaluating the Estimated ODTR for the "Interventions" Study

In our companion paper (the previous chapter), we estimated the ODTR on the "Interventions" data (n = 441) using the ODTR SuperLearner. The library for d^*_n consisted of a combination of simple parametric models and machine learning algorithms (`SL.glm`, `SL.mean`, `SL.glm.interaction`, `SL.earth`, and `SL.rpart`), and we used the same library for Q_n. The ODTR algorithm allocated all coefficient weight on a simple GLM with only substance use; this means that the estimated ODTR can be interpreted as: give CBT to those with low substance use scores and TAU to those with high substance use scores.

In this paper, we *evaluate* this estimated ODTR using CV-TMLE. Specifically, we aim to determine if administering CBT under this individualized rule is better than administering CBT in a non-individualized way – i.e., simply giving all participants CBT or no participants CBT.

The CV-TMLE estimate of the probability of no re-arrest under the ODTR SuperLearner is 61.37% (CI: [54.82%, 67.93%]). However, this probability is not significantly different than the CV-TMLE estimate of the static rule in which everyone receives CBT (difference: -0.35%, CI: [-6.40%, 5.71%]) and no one receives CBT (difference: -0.18%, CI: [-7.06%, 6.68%]). Estimates and confidence intervals of these CV-TMLE estimates are illustrated in Figure 2.2. Thus, there is insufficient evidence to conclude that assigning CBT using the ODTR SuperLearner is better than assigning CBT in a non-individualized way.

2.7 Conclusions

The aim of this paper was to illustrate the performance of different estimators that can be used to evaluate dynamic treatment rules, and in particular, the ODTR. At sample size 1,000, we saw a small price and many benefits to using CV-TMLE in order to estimate the following parameters: 1) the true value of a given *a priori* known rule; 2) the true value of the true, unknown ODTR; and, 3) the true value of an estimated ODTR (a data-adaptive parameter). In addition, we illustrated how to implement the CV-TMLE estimator to evaluate the ODTR using the "Interventions" data as an applied example.

When evaluating estimators' performance for the value of a known rule, CV-TMLE performed well, irrespective of how data-adaptive the algorithms used for estimating nuisance parameters were. Although no estimator under an estimated ODTR yielded satisfactory performance for a target parameter corresponding to the true value of the true ODTR, CV-TMLE performed the best when nuisance parameters and ODTRs were estimated using the most data-adaptive algorithms, while non-cross-validated estimators yielded overly optimistic and highly variable results. Finally, no other estimator except CV-TMLE performed well when estimating a data-adaptive parameter – a parameter that may be of interest if: 1) one believes one's estimate of the ODTR will not converge appropriately to its truth (as was the case for these estimators of the ODTR under the current DGP); and 2) one cares more about the performance of the estimated ODTR that is generated by the sample at hand (as opposed to the true, but unknown, ODTR).

Future directions for simulations should evaluate results under varying sample sizes. In particular, for small sample sizes and thus less support in the data, it may be that case that we pay a price in performance by sample splitting. Additionally, future work could extend these simulations to the multiple time-point setting to evaluate the *sequential* ODTR that could be generated from, for example, a SMART design [43, 29, 1] instead of an single time-point experiment.

As an illustration of how to apply the ODTR SuperLearner to real data, we estimated the ODTR using the "Interventions" Study to determine which types of criminal justice-involved adults with mental illness should be assigned CBT versus TAU, to yield the highest probability of no re-arrest. In our applied example using the "Interventions" data, preliminary results suggest the probability of recidivism if treatment were assigned using the ODTR algorithm (i.e., in an individualized way) is not significantly different from probability of recidivism if all had been assigned treatment or no treatment (i.e., in a non-individualized way). This may indicate an absence of strong heterogeneous treatment effects by the measured variables, or it may reflect limitations in power to detect such effects due to preliminary sample sizes. In future work, we will apply the ODTR SuperLearner and evaluate it on the full sample size (n = 720).

This work contributes to statistical methods for understanding treatment effect heterogeneity, and in particular, how much improvement we might make in outcomes if interventions are assigned according to an ODTR. It is of great practical relevance to study estimators of these parameters, which allow us to determine the benefit of assigning treatment in a more individualized way compared to, for example, simply giving all subjects treatment.

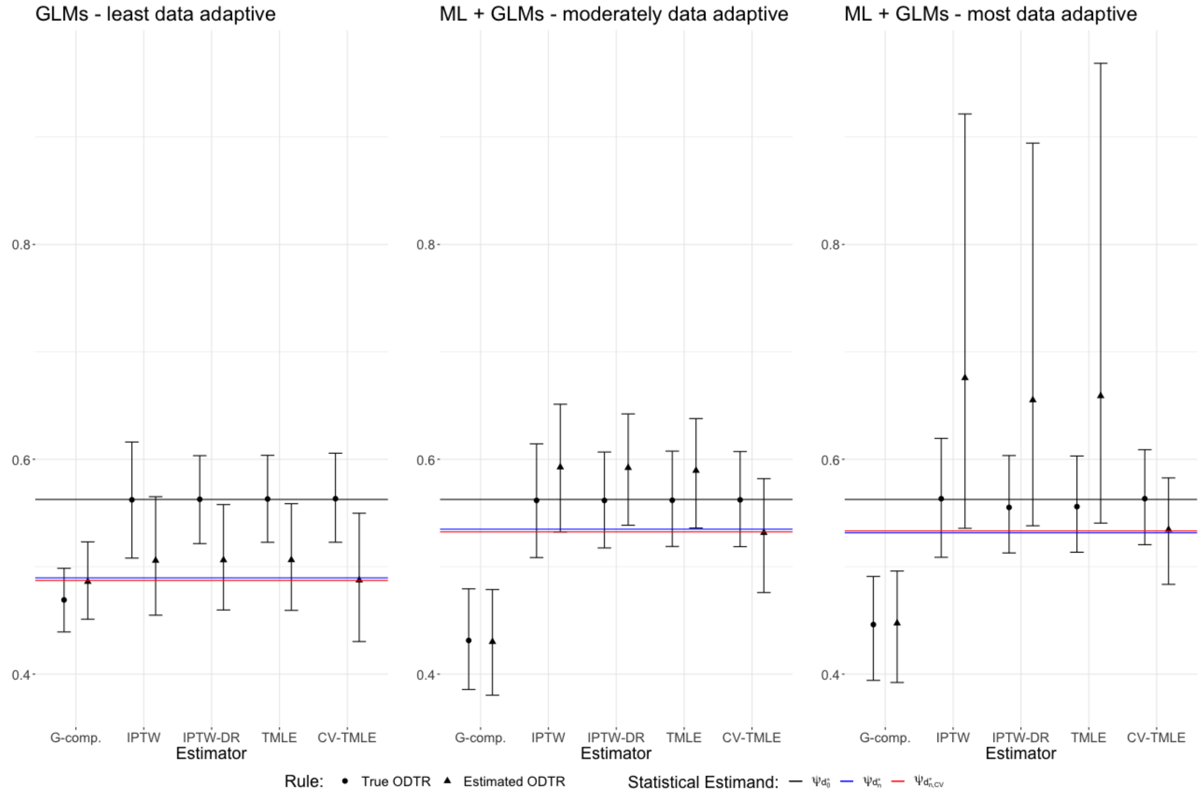

Figure 2.1: Performance of the value of the rule for 3 SuperLearner library configurations with increasing (left to right) levels of data-adaptivity used for estimating Q_0 and/or d_0^* ("GLM - least data adaptive", "ML + GLMs - moderately data adaptive", "ML + GLMs - most data adaptive"). The horizontal black line depicts the true mean outcome under the true ODTR ψ_{0,d_0^*}; the blue and red lines are the data-adaptive parameters ψ_{0,d_n^*} and $\psi_{0,d_{n,CV}^*}$, respectively, averaged over each of the 1,000 simulated samples. Points with error bars show the distribution of the estimators across the 1,000 simulated samples (G-computation estimator, IPTW estimator, TMLE, and CV-TMLE); the points (circles and triangles) show the estimates averaged over the samples, and error bars show the 2.5^{th} and 97.5^{th} quantiles of the distribution of each estimator across the simulation repetitions. The circles depict the estimators under a known rule $\hat{\psi}_{d=d_0^*}$ and the triangles illustrate the estimators under an estimated rule, either $\hat{\psi}_{d_n^*}$ or $\hat{\psi}_{d_{n,v}^*}$ (for CV-TMLE).

Library	Estimator	Bias	Variance	MSE	Coverage
GLMs	G-comp.	-0.0940	0.0003	0.0091	-
	IPTW	0.0009	0.0008	0.0008	95.3%
	IPTW-DR	0.0001	0.0005	0.0005	93.7%
	TMLE	0.0002	0.0005	0.0005	93.7%
	CV-TMLE	0.0004	0.0005	0.0005	93.7%
ML + GLMs not aggressive	G-comp.	-0.1298	0.0006	0.0175	-
	IPTW	0.0002	0.0008	0.0008	94.7%
	IPTW-DR	-0.0009	0.0006	0.0006	94.0%
	TMLE	-0.0011	0.0005	0.0005	93.6%
	CV-TMLE	-0.0009	0.0005	0.0005	93.2%
ML + GLMs aggressive	G-comp.	-0.1180	0.0006	0.0146	-
	IPTW	-0.0006	0.0009	0.0009	94.0%
	IPTW-DR	-0.0084	0.0005	0.0006	90.1%
	TMLE	-0.0075	0.0005	0.0006	90.6%
	CV-TMLE	-0.0001	0.0005	0.0005	93.6%

Table 2.2: Performance metrics (bias, variance, MSE, confidence interval coverage) of each estimator $\hat{\psi}_{d=d_0^*}$ for ψ_{0,d_0^*}, for each library configuration of Q_n.

Library	Estimator	Bias	Variance	MSE	Coverage
GLMs	G-comp.	-0.0773	0.0004	0.0064	-
	IPTW	-0.0558	0.0008	0.0039	45.0%
	IPTW-DR	-0.0565	0.0006	0.0038	30.1%
	TMLE	-0.0565	0.0006	0.0038	29.8%
	CV-TMLE	-0.0764	0.0009	0.0067	14.7%
ML + GLMs not aggressive	G-comp.	-0.1306	0.0007	0.0178	-
	IPTW	0.0334	0.0010	0.0021	76.1%
	IPTW-DR	0.0327	0.0008	0.0019	66.5%
	TMLE	0.0298	0.0008	0.0016	71.3%
	CV-TMLE	-0.0308	0.0007	0.0017	69.0%
ML + GLMs aggressive	G-comp.	-0.1161	0.0007	0.0142	-
	IPTW	0.1236	0.0109	0.0262	31.0%
	IPTW-DR	0.1010	0.0092	0.0194	33.0%
	TMLE	0.1031	0.0108	0.0214	33.6%
	CV-TMLE	-0.0316	0.0007	0.0017	68.6%

Table 2.3: Performance metrics (bias, variance, MSE, confidence interval coverage) of each estimator $\hat{\psi}_{d_n^*}$ (G-computation, IPTW, IPTW-DR, TMLE) or $\hat{\psi}_{d_{n,v}^*}$ (CV-TMLE) for ψ_{0,d_0^*}, for each library configuration of Q_n and d_n^*.

Library	Estimator	Bias	Variance	MSE	Coverage
GLMs	G-comp.	-0.0033	0.0004	0.0006	-
	IPTW	0.0183	0.0008	0.0009	94.3%
	IPTW-DR	0.0175	0.0006	0.0007	90.6%
	TMLE	0.0175	0.0006	0.0007	90.7%
	CV-TMLE	-0.0002	0.0009	0.0005	94.3%
ML + GLMs not aggressive	G-comp.	-0.1027	0.0007	0.0114	-
	IPTW	0.0614	0.0010	0.0046	43.8%
	IPTW-DR	0.0607	0.0008	0.0044	28.9%
	TMLE	0.0578	0.0008	0.0040	30.4%
	CV-TMLE	0.0002	0.0007	0.0005	94.0%
ML + GLMs aggressive	G-comp.	-0.0846	0.0007	0.0081	-
	IPTW	0.1551	0.0109	0.0366	16.3%
	IPTW-DR	0.1325	0.0092	0.0283	15.8%
	TMLE	0.1346	0.0108	0.0307	15.7%
	CV-TMLE	0.0001	0.0007	0.0005	94.8%

Table 2.4: Performance metrics (bias, variance, MSE, confidence interval coverage) of each estimator $\hat{\psi}_{d_n^*}$ (G-computation, IPTW, IPTW-DR, TMLE) for ψ_{0,d_n^*} or $\hat{\psi}_{d_{n,v}^*}$ (CV-TMLE) for $\psi_{0,d_{n,CV}^*}$, for each library configuration of Q_n and d_n^*.

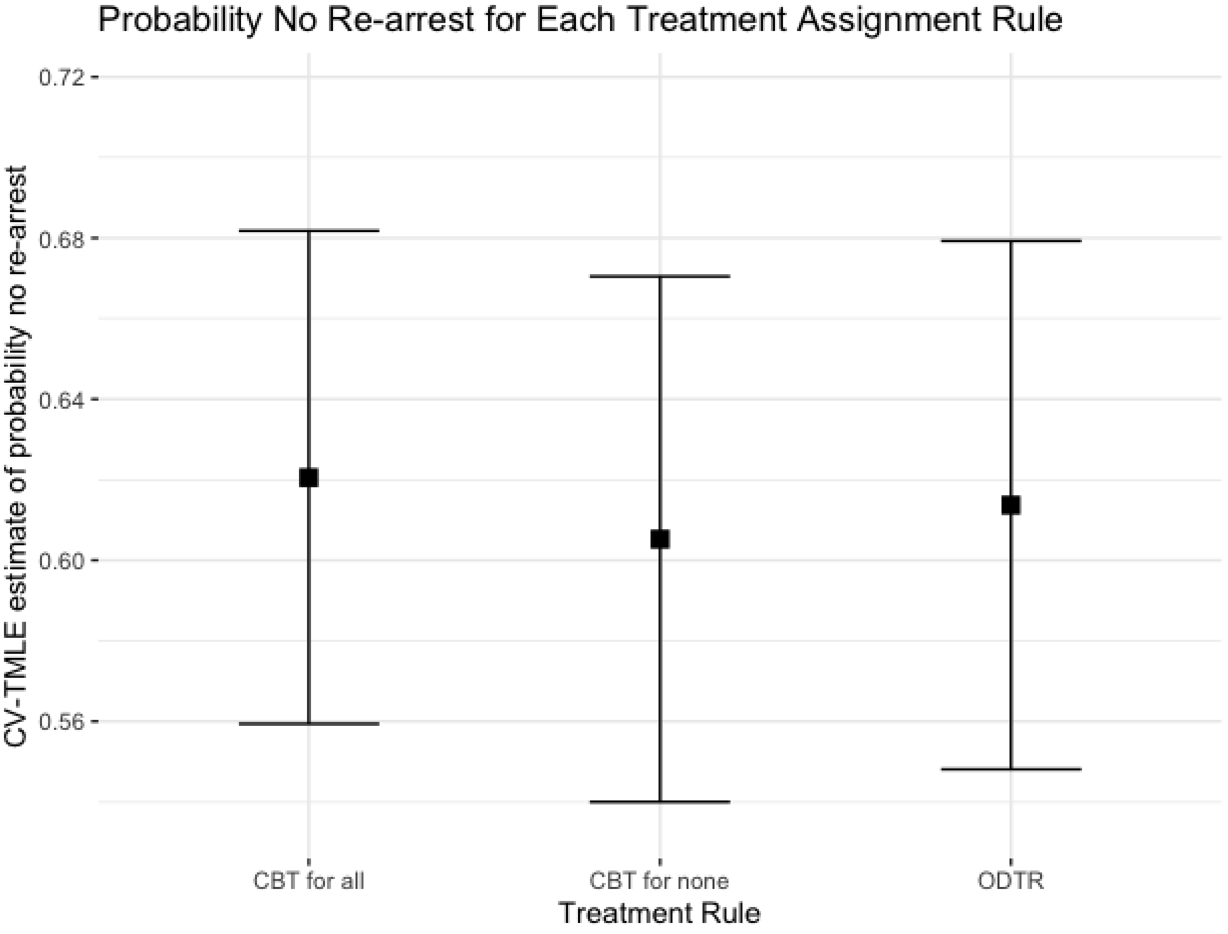

Figure 2.2: CV-TMLE estimates of the probability of no re-arrest under the following treatment rules: give CBT to all, give CBT to none, give CBT according to the ODTR Super-Learner algorithm. The squares are the point estimates and the error bars are 95% confidence intervals on these point estimates. There is no significant difference in the estimated probability of no re-arrest under a treatment regime in which all are given CBT, none are given CBT, and CBT is given using this ODTR.

Chapter 3

Augmenting the Optimal Dynamic Treatment Rule SuperLearner

3.1 Introduction

The optimal dynamic treatment rule (ODTR) framework provides a way of determining which treatment works best for which kinds of patients [52, 69]. Further, by evaluating the ODTR, one can determine whether significant treatment effect heterogeneity exists in a population. Recently, many methods have been developed for estimating the ODTR – algorithms that input patient characteristics (covariates) and output a treatment decision [28, 29]. One such way of estimating the ODTR is the ODTR SuperLearner algorithm, first described in [46], and later in Chapter 1 and [15]. The ODTR SuperLearner considers a library of candidate algorithms for estimating the ODTR, combines those algorithms using a choice of metalearner, and chooses the "best" combination of the candidate ODTRs based on minimizing a choice of risk function.

The above algorithms output a deterministic, discrete (binary, if treatment is 0 or 1) treatment recommendation. However, considering stochastic treatment rule estimators could be helpful in understanding the true optimal rule. Thus, in this paper, we propose to augment the ODTR SuperLearner that only includes a deterministic output to also include candidate algorithms that output a probability that a type of individual should be treated. We do this in two ways: first, by using the strength of the treatment response for a given covariate profile, and second, by regularizing the original deterministic rule with the variance of the value of the rule. Further, we show how to evaluate the mean outcome under a stochastic rule, using cross validated targeted maximum likelihood (CV-TMLE), with inference [37, 36, 24]. We hypothesize that including stochastic candidates in the ODTR SuperLearner library will not harm performance – in terms of bias, variance, mean squared error (MSE), confidence interval coverage, and confidence interval width – of the CV-TMLE of the true value of the true optimal rule, compared to having a library with only deterministic treatment rules. Importantly, we hypothesize that this result will be of benefit in scenarios in which it there is little signal to detect the optimal rule – such as in finite samples and when there is a small amount of treatment effect heterogeneity.

In previous research, we have shown how to estimate the optimal rule using an ODTR SuperLearner that uses as selection criterion (or risk function) the estimate of the value of the candidate rule via CV-TMLE. However, in finite samples, this estimate of the true value of the candidate rule does not take into account the variability of the estimate of the criterion and how it varies from one candidate estimator of the optimal rule to another. Thus, we additionally introduce a new criterion for selecting the ODTR SuperLearner that incorporates the variability of the value of that rule: the upper bound of a 95% confidence interval for the CV-TMLE of the true value of the candidate rule. One can imagine an example scenario in which two candidate rules yield the same point estimate of the value of the rule; under this criterion, either candidate rule could be chosen as the optimal one. However, if one candidate rule has less variability in the CV-TMLE of the value of the rule, then its confidence interval will be smaller, and that one will be chosen over the less precise candidate, despite having the same point estimate of the value of the candidate rule. Thus, we hypothesize more penalization of candidate estimators of the optimal rule under the upper bound of the confidence interval risk function versus the point estimate, especially in scenarios where the candidate rule is weakly supported by the data (such as in observational studies) and thus variability of the CV-TMLE will increase. Further, we hypothesize that this penalization will increase precision in the CV-TMLE estimator.

In this paper, we use simulations to show the performance of TMLE and CV-TMLE as estimators for the true value of the true optimal dynamic treatment rule, when the ODTR SuperLearner is used to estimate the ODTR under two novel settings: (1) when the ODTR SuperLearner library is augmented with stochastic rules, and (2) when the upper bound of the confidence interval on the CV-TMLE of the candidate rule is used as a selection criterion (i.e., risk function) for the ODTR SuperLearner. We also apply the augmented ODTR SuperLearner and the CV-TMLE of that rule to the Correctional Intervention for People with Mental Illness, or "Interventions," trial. This is an ongoing randomized controlled trial (RCT) in which criminal justice-involved adults with mental illness – a heterogeneous group with diverse symptoms, risk factors, and other treatment-relevant characteristics [82, 83] – are either randomized to cognitive behavioral therapy (CBT) or treatment as usual (TAU), and re-arrest (the outcome of interest) is collected one year after randomization occurs, as a measure of recidivism.

This article steps through the causal roadmap for answering causal questions [57], and is organized as follows. In the following section, we define the data and causal model, define the target causal parameters, list the assumptions to identify causal parameters as statistical parameters, and provide a statistical formulation in which causal estimands are identified as functions of the observed data distribution (i.e., statistical estimands). In section 3 we discuss estimation (via the ODTR SuperLearner) and evaluation (via CV-TMLE) of the true value of the true optimal rule, including an introduction of a stochastic rule and variance-based risk function augmentation for the ODTR SuperLearner. In section 4 we present a simulation study illustrating the performance of these estimators with respect to the true expected outcome under the true optimal rule. In section 5 we use the augmented ODTR SuperLearner and CV-TMLE to estimate the ODTR and evaluate it, respectively, using the "Interventions" Study. Finally, we close with conclusions and future directions.

3.2 Causal Roadmap

Data and Causal Model

We consider point-treatment data where $W \in \mathcal{W}$ are baseline covariates, $A \in \{0,1\}$ is the treatment, and $Y \in \mathbb{R}$ is the outcome measured at the end of the study. Our data can be described by the following structural causal model (SCM), $\mathcal{M}^F$ [55]:

$$\begin{aligned} W &= f_W(U_W) \\ A &= f_A(W, U_A) \\ Y &= f_Y(W, A, U_Y)\ , \end{aligned}$$

where the full data $X = (W, A, Y)$ are endogenous nodes, $U = (U_W, U_A, U_Y) \sim P_U$ are unmeasured exogenous variables, and $f = (f_W, f_A, f_Y)$ are structural equations. The SCM provides a model for the set of possible counterfactual distributions: $P_{U,X} \in \mathcal{M}^F$.

Here, $f_A(W, U_A) = \mathbb{I}[U_A < g_0(1|W)]$, where $U_A \sim Uniform(0,1)$ and $g_0(A|W) = Pr(A|W)$; in other words, $A \sim Bernoulli(p = g_0(1|W))$. Data could be generated from an RCT using simple randomization with equal probability to each arm, in which case the above structural causal model would state that Y may be affected by both W and A, but that W does not affect A (as in the "Interventions" trial); this can be represented in the above model by letting $g_0(1|W) = 0.5$.

Target Causal Parameters

In this point treatment setting, a given stochastic treatment rule is a function g^* that takes as input measured baseline covariates W and outputs a probability of receiving treatment: $W \rightarrow g^*(1|W)$. We denote the set of all dynamic treatment rules as $\mathcal{G}^*$.

For a given stochastic intervention g^*, we intervene on the above SCM to derive counterfactual outcomes:

$$\begin{aligned} W &= f_W(U_W) \\ A &\sim Bernoulli(p = g^*(1|W)) \\ Y_{g^*} &= f_Y(W, A, U_Y)\ . \end{aligned}$$

Here, Y_{g^*} is the counterfactual outcome for a subject if his/her treatment A were assigned using the stochastic treatment rule g^*. Each causal estimand below is a mapping $\mathcal{M}^F \rightarrow \mathbb{R}$.

The value of an arbitrary stochastic rule g^* is the expected outcome under g^*:

$$\Psi^F_{g^*}(P_{U,X}) \equiv \mathbb{E}_{P_{U,X}}[Y_{g^*}].$$

Given this, our first causal parameter of interest is the stochastic rule that, among all candidate stochastic rules, yields the best (lowest) expected outcomes. The optimal stochastic rule g_0^* is equivalent to the deterministic rule d_0, which yields a 0 or 1 treatment decision:

$$g_0^* \equiv \underset{g^* \in \mathcal{G}^*}{\arg\min}\, \Psi^F_{g^*}(P_{U,X}) = d_0 \in \underset{d \in \mathcal{D}}{\arg\min}\, E_{P_{U,X}}[Y_d],$$

where $\mathcal{D} = \{0, 1\}$. We ultimately aim to quantify the value of the optimal rule, which is the following causal parameter:

$$\Psi^F_{g_0^*}(P_{U,X}) \equiv \mathbb{E}_{P_{U,X}}[Y_{g_0^*}].$$

Assumptions to Identify Causal Parameters as Statistical Parameters

We assume that our observed data were generated by sampling n independent observations $O_i \equiv (W_i, A_i, Y_i)$, $i = 1, \ldots, n$, from a data generating system described by $\mathcal{M}^F$ above (e.g., the "Interventions" study consists of 441 i.i.d. observations of O).

We additionally assume that $U_A \perp U_Y$ holds and either $U_W \perp U_Y$ or $U_W \perp U_A$ holds. Under these independence assumptions, the backdoor criteria (with the implied randomization assumption) holds; that is, $Y_{g^*} \perp A|W \;\forall g^* \in \mathcal{G}^*$ [55].

If the data are generated from an RCT design, as in the "Interventions" study, then the true g_0 is known, and both the backdoor criteria and positivity assumption,

$$Pr\Big(\min_{a \in \{0,1\}} g_0(A = a|W) > 0\Big) = 1,$$

hold by design; in an observational data setting the randomization assumption requires measurement of a sufficient set of baseline covariates, and the positivity assumption may also pose greater challenges [58].

Statistical Formulation

The likelihood of the observed data can be written as:

$$\prod_{i=1}^{n} p_{W,0}(W_i) g_0(A_i|W_i) p_{Y,0}(Y_i|A_i, W_i),$$

where $p_{W,0}$ is the true density of W; g_0 is the true conditional probability of A given W; $p_{Y,0}$ is the true conditional density of Y given A and W. The distribution of the data P_0 is an element of the statistical model $\mathcal{M}$.

Define $\bar{Q}(a, w) \equiv E[Y|A = a, W = w]$. Under the above assumptions, $E_{P_{U,X}}[Y_{g^*}]$ (a parameter of the counterfactual distribution) is identified as the following statistical parameter:

$$\Psi_{g^*}(P_0) = \mathbb{E}_0\left[\sum_{a \in \{0,1\}} \bar{Q}_0(a, W) g^*(a|W)\right].$$

The parameter $\Psi_{g^*} : \mathcal{M} \rightarrow \mathbb{R}$ is pathwise differentiable at P with canonical gradient at P given by

$$D_{g^*}(P) = \frac{g^*(A|W)}{g(A|W)}(Y - \bar{Q}(A,W)) + \sum_a \bar{Q}(a,W)g^*(a|W) - \Psi_{g^*}(P).$$

The exact second order remainder $R_{2,g^*}(P, P_0) \equiv \Psi_{g^*}(P) - \Psi_{g^*}(P_0) + P_0 D_{g^*}(P)$ is

$$\mathbb{E}_0 \left[\sum_a \frac{g^*(a|W)(g - g_0)}{g}(\bar{Q} - \bar{Q}_0)(a,W) \right].$$

Note that $D_{g^*}(P) = D_{g^*,y}(P) + D_{g^*,W}(P)$, where $D_{g^*,y}(P)$ and $D_{g^*,W}(P)$ are components that are scores of the conditional density of Y, given A, W and the marginal density of W, respectively. Let $\sigma^2_{g^*}(P) = \mathbb{E}_P\left[D_{g^*}(P)^2\right]$ be the variance of $D_{g^*}(P)$, which represents the asymptotic variance of the efficient influence curve for $\Psi_{g^*}(P_0)$. We also define $\sigma^2_{g^*,y}(P) = \mathbb{E}_P\left[D_{g^*,y}(P)^2\right]$, which represents the asymptotic variance of the efficient influence curve for the data-adaptive parameter $\frac{1}{n}\sum_{i=1}^n \sum_a \bar{Q}_0(a, W_i)g^*(a|W_i)$; this data-adaptive target parameter can be viewed as another way to measure the value of a given rule.

The true optimal stochastic intervention is identified by $g_0^* = \arg\min_{g^* \in \mathcal{G}^*} \Psi_{g^*}(P_0)$, and for a given a, $g_0^*(a|W) = \mathbb{I}[a = d_0(w)]$, where $d_0(W) = \arg\min_{d \in \mathcal{D}} \bar{Q}_0(A = d, W)$, or equivalently, $d_0(W) = \mathbb{I}[B_0(W) \leq 0]$, where $B_0(W) = \bar{Q}_0(1, W) - \bar{Q}_0(0, W)$ is sometimes referred to as the blip function (noting that in that definition we assume that assigning treatment 1 is preferable to assigning treatment 0 in the absence of a treatment effect).

Given this, the causal parameter $\Psi^F_{g_0^*}(P_{U,X})$ can be identified by

$$\Psi_{g_0^*}(P_0) = \mathbb{E}_0 \left[\sum_{a \in \{0,1\}} \bar{Q}_0(a,W)g_0^*(a|W) \right].$$

Further, we define the value of an estimated optimal rule – a data-adaptive parameter. Let P_n be the empirical distribution which gives each observation weight $\frac{1}{n}$; $P_n \in \mathcal{M}_{NP}$, where $\mathcal{M}_{NP}$ is a non-parametric statistical model. Estimators are viewed as mappings applied to P_n. Further, consider a V-fold cross-validation scheme so that the empirical data are uniformly and at random split into V mutually exclusive sets. For sets $v \in \{1, ..., V\}$, each set of data serves as a validation set; the complement is its training set. Let $P_{n,-v}$ be the empirical distribution of the validation sample v, and $P_{n,v}$ be the empirical distribution of the complementary training set. Let $g_n^* = \hat{g}^*(P_n)$ and $d_n = \hat{d}(P_n)$ be a candidate stochastic and deterministic estimator, respectively, of the optimal rule g_0^* and d_0. Similarly, let $g^*_{n,v} = \hat{g}^*(P_{n,v})$ and $d_{n,v} = \hat{d}(P_{n,v})$ resulting estimators of the optimal rule when applying $\hat{g}^*$ and $\hat{d}$, respectively, to the training sample. Then, two data-adaptive parameters of interest could be either (a) $\Psi_{g_n^*}(P_0) = \mathbb{E}_0\left[\sum_{a \in \{0,1\}} \bar{Q}_0(a,W)g_n^*(a|W)\right]$, the true value of the estimated, sample-specific stochastic rule; or (b) the training-sample-specific estimate of the stochastic rule, averaged across sample splits, i.e., $\Psi_{g^*_{n,v,CV}}(P_0) = \frac{1}{V}\sum_{v=1}^V \Psi_{g^*_{n,v}}(P_0)$.

3.3 Estimation

The estimation goal of this paper is to approximate the optimal rule d_0, and its true value $\Psi_{g_0^*}(P_0)$, and provide inference for the latter, as well. Thus, we first describe the ODTR SuperLearner, with an augmented library that includes stochastic rules, as a way to estimate the true optimal rule in finite samples. Then, we describe how to estimate the value of a given stochastic treatment assignment. The discussion of how to evaluate a given stochastic treatment rule leads us to present two risk functions – either the CV-TMLE estimate for the value of the candidate rule or the upper bound of the confidence interval for that CV-TMLE estimate – as selection criteria for the ODTR SuperLearner. Finally, we show how to estimate the true value (estimated using CV-TMLE) of the true optimal rule (estimated using the ODTR SuperLearner).

SuperLearner Estimation of the ODTR

In past research, we have described how to estimate the optimal rule using the data adaptive algorithm, SuperLearner. We refer the reader to Chapter 1 and [46] for an explanation of SuperLearning for estimation of the ODTR (and [37, 63] for an introduction to SuperLearning in general). Importantly, we note that in this past work, the library of candidate rules included as estimators of the optimal rule were deterministic – the predicted output given a covariate profile was always a 0 or 1 treatment decision. In this paper, we aim to extend this SuperLearner by adding candidate estimators of the optimal rule that could output a probability, or a stochastic treatment rule. Additionally, in previous work, we have used as criterion for the selection of the SuperLearner (i.e., the risk function) the expected outcome under the candidate rule. Here, we extend on previous work by introducing a new risk function that incorporates the variability of the value of the rule.

Augmenting the ODTR SuperLearner Library with Stochastic Rules

We build on work described in Chapter 1 by augmenting the SuperLearner library with stochastic rules; here, we discuss two forms of stochastic augmentation. The first leverages the distribution of the blip function, and the second combines the deterministic estimate of the rule with knowledge about the variance of the value of the rule.

Stochastic Intervention Through Blip Transformation The first candidate stochastic intervention estimator transforms a given candidate estimator of the blip $B_n(W)$ by taking its inverse logit:

$$g^*_{n,c}(1|W) = 1 - logit^{-1}(B_n(W)/c).$$

In this way, people who have a lower effect of treatment will have a higher probability of treatment, and vice versa. In the SuperLearner, this estimator is indexed by a non-negative constant $c \in \{0, ..., C\} = \mathcal{C} \in \mathbb{R}_{\geq 0}$. The reason for including c is to introduce a constant that considers and mitigates the spread of the blip. Importantly, $\mathcal{C}$ should include 0, so that

$1 - logit^{-1}(B_n(W)/0) = \mathbb{I}[B_n(W) \leq 0] = d_n$, ensuring that the deterministic rules from the previous section are included in the library of candidate algorithms.

Stochastic Intervention Through Variance Regularization The second candidate stochastic intervention estimator combines candidate estimators of the optimal rule d_n with a minimizer of $\sigma^2_{g^*,y}(P_0)$, i.e., the variance of the data-adaptive measure of the value of a given stochastic rule $\frac{1}{n}\sum_{i=1}^n \sum_a \bar{Q}_0(a, W_i)g^*(a|W_i)$.

Lemma 1 *Let* $\sigma^2_{g^*,y}(P_0) = \mathbb{E}_0\left[\left\{\frac{g^*(A|W)}{g(A|W)}(Y - \bar{Q}(A,W))\right\}^2\right]$ *and* $g^*_{r,P} \equiv \arg\min_{g^*} \sigma^2_{g^*,y}(P)$ *be the minimizer of this variance term. Let* $\sigma^2(A,W) = \mathbb{E}_P\left[(Y - \bar{Q}(A,W))^2|A,W\right]$ *be the conditional variance of* Y*, given* A *and* W*. Then for* $a \in \{0,1\}$ *we have*

$$g^*_{r,P}(a|W) = \frac{g(a|W)/\sigma^2(a,W)}{\sum_{k\in\{0,1\}} g(k|W)/\sigma^2(k,W)}.$$

In particular, if $\sigma^2(a,W)$ *is constant in* a*, then* $g^*_{r,P}(a|W) = g(a|W)$*.*

Proof 1 *First, notice that, for each* w*,* $g^*_{r,P}(\cdot|W=w)$ *is the minimizer of* $\sum_a \frac{g^{*2}(a|w)}{g(a|w)}\sigma^2(a,w)$*. Suppose that this minimizer* g^*_r *is parameterized by* $g^*_r(0)$ *while* $g^*_r(1) = 1 - g^*_r(0)$*. Suppose that* $g^*_r(\cdot|w)$ *is an interior minimum. In that case, we can simply take the derivative with respect to* $g^*_r(0|w)$ *to obtain the equation:*

$$\frac{g^*_r(0|w)}{g(0|w)}\sigma^2(0,w) - (1 - g^*_r(0|w))\frac{\sigma^2(1,w)}{g(1|w)} = 0$$

Suppressing the dependence on w*, we obtain:*

$$\begin{aligned} g^*_r(0) &= \frac{\sigma^2(1)}{\sigma^2(0)}\frac{g(0)}{g(1)} - \frac{\sigma^2(1)}{\sigma^2(0)}\frac{g(0)}{g(1)}\sum_{k\in\{0,1\}} g^*_r(k) \\ &= \frac{\frac{g(0)\sigma^2(1)}{g(1)\sigma^2(0)}}{1 + \frac{g(0)\sigma^2(1)}{g(1)\sigma^2(0)}} \end{aligned}$$

Plugging this in the expression for $g^*_r(0)$ *in the previous displayed equation yields the result.*

From this perspective, the standard error of $\Psi_{g^*}(P_0)$ shrinks the deterministic rule to a mechanism $g^*_{r,P}$, which would equal g for each W for which $\sigma^2(a,W)$ is constant in a. Thus, in the SuperLearner, as candidate stochastic treatment rule estimators of the ODTR, we consider the following convex combination of this regularized rule with a candidate deterministic rule d_n:

$$g^*_{n,\lambda} = (1-\lambda)d_n + \lambda g^*_{n,r}, \lambda \in \Lambda = [0,1]$$

Again, including $\lambda = 0$ in Λ here ensures that the SuperLearner estimator considers the deterministic rules d_n. Thus, in the SuperLearner, we select candidate algorithms based on λ.

SuperLearner Description

Given the aforementioned stochastic rule augmentations to the library, we briefly describe the SuperLearner steps here:

1. Choose J candidate algorithms for estimating the ODTR $d_{n,j}(W)$ for $j = 1, ..., J$. Candidates can include approaches based on estimating the blip $B_{n,j}(W)$, which imply a candidate estimator for a deterministic optimal rule, i.e., $d_{n,j}(W) = \mathbb{I}[B_{n,j}(W) \leq 0]$, or approaches for estimating the optimal rule directly, or static rules.

2. Augment the J candidate algorithms by either taking the inverse logit of $B_{n,j}(W)$ divided by a constant $c \in \mathcal{C}$, i.e., $g^*_{n,c,j}$, or, create a convex combination using $\lambda \in \Lambda$ between the regularized stochastic rule and the candidate deterministic rule $d_{n,j}(W)$, i.e., $g^*_{n,\lambda,j}$.

3. Under a cross-validation scheme, fit each of the augmented candidate algorithms, and any dependent nuisance parameters, on the training set [15].

4. Predict the estimated stochastic rule for each observation in the validation set for each augmented algorithm based on the corresponding training set fit.

5. As measure of performance of a particular candidate estimator $g^*_{n,\lambda,j}(1|W)$ or $g^*_{n,c,j}(1|W)$ of the optimal rule, we use the true risk function $R_0(g^*_n, P_n)$ under the training sample specific estimate of the optimal rule, averaged across sample splits, which can be estimated with a cross-validated estimator $R_{n,CV}(g^*_n, P_n)$. In past research, we have used as measure of performance $R_0(d_n, P_n) = \frac{1}{V}\sum_{v=1}^{V} \mathbb{E}_0\left[\bar{Q}_0(d^*_{n,v}(W), W)\right]$, the true value of the sample-split-specific estimate of the optimal rule, which can be estimated using a cross-validated targeted maximum likelihood estimator (CV-TMLE) [37]. We expand on this more in the next section.

6. Choose the estimator, indexed by j the λ pair or j and c pair, that yields the smallest cross-validated empirical risk, i.e., $(j_n, c_n) = \arg\min_\alpha R_{n,CV}(g^*_{n,c,j}, P_n)$ or $(j_n, \lambda_n) = \arg\min_\alpha R_{n,CV}(g^*_{n,\lambda,j}, P_n)$.

7. Fit each candidate estimator $B_{n,j_n}(W)$ of the blip or $d_{n,j_n}(W)$ of the rule on the entire data set, and then augment them using either c_n or λ_n, respectively. This is the SuperLearner estimate of the optimal rule, where $\hat{g}^*_{SL1}(P_n) = g^*_{n,j_n,c_n}$ or $\hat{g}^*_{SL2}(P_n) = g^*_{n,j_n,\lambda_n}$.

Selection Criteria (Risks) for the ODTR SuperLearner

In this section, we describe two possible risk functions for selecting the SuperLearner: (1) the value of the candidate stochastic rule, and (2) the upper bound of the confidence interval on the value of the candidate stochastic rule. We use CV-TMLE to estimate these quantities, and thus first give an overview of CV-TMLE for estimating the value of a given stochastic rule, and confidence intervals around that estimate.

Brief Overview: TMLE and CV-TMLE for the Value of a Given Stochastic Rule We review how to obtain point estimates of $\Psi_{g^*}(P_0)$ – the true value of an arbitrary stochastic rule g^* – using TMLE or CV-TMLE. In addition, we obtain confidence intervals around CV-TMLE estimates. Though this is not our target parameter of interest, showing this will be helpful in the discussion of possible risk criteria for choosing the ODTR SuperLearner.

TMLE of $\Psi_{g^*}(P_0)$ Let $L(\bar{Q})$ be a loss function so that $\bar{Q}_0 = \arg\min_{\bar{Q}} P_0 L(\bar{Q})$. In addition, let $\{\bar{Q}_\epsilon : \epsilon\}$ be a path through $\bar{Q}$ at $\epsilon = 0$ so that $\frac{d}{d\epsilon}L(\bar{Q}_\epsilon) = D_{g^*}(\bar{Q}_\epsilon, g)$. For example, for $Y \in \{0,1\}$ (noting that if Y is continuous outside of those bounds it should be transformed between 0 and 1 [19]), $L(\bar{Q}) = -Y \log \bar{Q} - (1-Y)\log(1-\bar{Q})$ is the binary log-likelihood loss. Then we can select $\text{Logit}\,\bar{Q}_\epsilon = \text{Logit}\,\bar{Q} + \epsilon g^*/g$. One can also put g^*/g into the weight of the loss, and use instead the intercept model $\text{Logit}\,\bar{Q}_\epsilon = \text{Logit}\,\bar{Q} + \epsilon$. In our implementations, we use the latter. Let $\bar{Q}_n$ be an initial estimator of $\bar{Q}_0$ and g_n be an estimator of g_0. Let $\epsilon_n = \arg\min_\epsilon P_n L(\bar{Q}_\epsilon)$. Then, $\bar{Q}_n^* = \bar{Q}_{n,\epsilon_n}$ is the TMLE of $\bar{Q}_0$ targeted towards $\Psi_{g^*}(P_0)$. Let $Q_{W,n}$ be the empirical probability measure of $W_1, ..., W_n$, and $Q_n^* = (Q_{W,n}, \bar{Q}_n^*)$. Then the TMLE of $\Psi_{g^*}(P_0)$ is given by

$$\hat{\psi}_{g^*} = \frac{1}{n}\sum_{i=1}^{n}\sum_{a\in\{0,1\}} \bar{Q}_n^*(a, W_i) g^*(a|W_i).$$

CV-TMLE of $\Psi_{g^*}(P_0)$ Let $\bar{Q}_{n,v}$ and $g_{n,v}$ be estimators of $\bar{Q}_0$ and g_0, respectively, based on the training sample $P_{n,v}, v = 1, ..., V$. Define $\epsilon_n = \arg\min_\epsilon \frac{1}{V}\sum_v P_{n,v}^1 L(\bar{Q}_{n,v,\epsilon})$. This defines $\bar{Q}_{n,v}^* = \bar{Q}_{n,v,\epsilon_n}$ for each v. Let $Q_{W,n,v}^1$ be the empirical probability distribution of W_i in the validation sample. Then, the CV-TMLE of $\Psi_{g^*}(P_0)$ is given by

$$\hat{\psi}_{g^*,CV} = \frac{1}{V}\sum_{v=1}^{V} \Psi_{g^*}(Q_{W,n,v}^1, \bar{Q}_{n,v}^*).$$

Confidence Interval based on CV-TMLE of $\Psi_{g^*}(P_0)$ The asymptotic variance of $n^{1/2}\left(\hat{\psi}_{g^*,CV} - \Psi_{g^*}(P_0)\right)$ can be estimated with the cross-validated variance of the efficient influence curve:

$$\hat{\sigma}_{g^*}^2 = \frac{1}{V}\sum_{v=1}^{V} P_{n,v}^1\{D_{g^*}(Q_{n,v}^*, g^*)\}^2.$$

A corresponding 95% confidence interval for $\Psi_{g^*}(P_0)$ is given by

$$\hat{\psi}_{g^*,CV} \pm \Phi^{-1}(0.975)\frac{\hat{\sigma}_{g^*}^2}{\sqrt{n}}.$$

Risk Function Through Variance Regularization As described in Section 3.3, in previous research we have estimated the ODTR using a SuperLearner $d_n(W)$ that aims to minimize, over a class of candidates d, $R_0(d, P_n) = \frac{1}{V}\sum_{v=1}^{V} \mathbb{E}_0\left[\bar{Q}_0(d(W), W)\right]$; this risk function is estimated using CV-TMLE, as in Chapter 1. One might also estimate g_0^* with an estimator minimizing, over a class of candidates g^*, the CV-TMLE of $g^* \rightarrow \frac{1}{V}\sum_v \Psi_{g^*}(Q_{n,v}^*)$. However, such an estimator of d_0 does not take into account the uncertainty in the estimator $\bar{Q}_n$ or more directly the chosen estimator of the criterion $\Psi_{g^*}(P_0)$. We argue that one might prefer an estimator of the optimal rule that also takes into account the uncertainty in the estimator of its performance measure, $\Psi_{g^*}(P_0)$.

For that purpose, we define the following risk function for a candidate g^*:

$$R_{g^*}(P) \equiv \Psi_{g^*}(P) + \Phi^{-1}(0.975)\frac{\sigma_{g^*}}{\sqrt{n}},$$

noting that 0.975 is user-supplied.

An estimator $R_{g^*,n} = \hat{\psi}_{g^*,CV} + \Phi^{-1}(0.975)\frac{\hat{\sigma}^2_{g^*}}{\sqrt{n}}$, where $\hat{\psi}_{g^*,CV}$ is the CV-TMLE of $\Psi_{g^*}(P_0)$ and $\hat{\sigma}^2_{g^*}$ is an estimator of $\sigma^2(P_0)$ would correspond with the upper bound of an asymptotic 95% confidence interval for $\Psi_{g^*}(P_0)$ based on the CV-TMLE $\hat{\psi}_{g^*,CV}$.

Risk Options for ODTR SuperLearner Now, let $g_n^* = \hat{g}^*(P_n)$ be a candidate estimator of the optimal rule g_0^*. Let $\bar{Q}_{n,v}$ and $g_{n,v}$ be estimators of $\bar{Q}_0$ and g_0 based on the training sample $P_{n,v}$. Similarly, let $g_{n,v}^* = \hat{g}^*(P_{n,v})$ be the resulting estimator of the optimal rule when applying $\hat{g}^*$ to the training sample. Consider $\text{Logit}\,\bar{Q}_{n,v,\epsilon} = \text{Logit}\,\bar{Q}_{n,v} + \epsilon g_{n,v}^*/g_{n,v}$, and define $\epsilon_n = \arg\min_\epsilon \frac{1}{V}\sum_v P_{n,v}^1 L(\bar{Q}_{n,v,\epsilon})$. This defines $\bar{Q}_{n,v}^* = \bar{Q}_{n,v,\epsilon_n}$ for each v. Then, the CV-TMLE of $\Psi_{g_{n,v,CV}^*}(P_0)$ is given by

$$R_{\hat{g}^*,CV,1} \equiv \hat{\psi}_{\hat{g}^*,CV} = \frac{1}{V}\sum_{v=1}^{V} \Psi_{g_{n,v}^*}(Q_{W,n,v}^1, \bar{Q}_{n,v}^*),$$

which is the first risk criterion we consider for evaluating a candidate estimator $\hat{g}^*$ of g_0^*.

The asymptotic variance of $n^{1/2}\left(\hat{\psi}_{\hat{g}^*,CV} - \Psi_{g_{n,v,CV}^*}(P_0)\right)$ can be estimated with the cross-validated variance of the efficient influence curve:

$$\hat{\sigma}^2_{\hat{g}^*} = \frac{1}{V}\sum_{v=1}^{V} P_{n,v}^1\{D_{g_{n,v}^*}(Q_{n,v}^*, g_{n,v}^*)\}^2.$$

The upper-bound of a confidence interval for $\Psi_{g_{n,v,CV}^*}(P_0)$ is given by:

$$R_{\hat{g}^*,CV,2} \equiv \hat{\psi}_{\hat{g}^*,CV} + \Phi^{-1}(0.975)\frac{\hat{\sigma}^2_{\hat{g}^*}}{\sqrt{n}},$$

which is our second criterion for evaluating a candidate estimator of g_0^*.

As in the ODTR SuperLearner description above in section 3.3, consider now a: (1) library of candidate estimators $g^*_j, J = 1, ..., J$; and a (2) stochastic rule augmentation indexed by either λ or c, i.e., $\hat{g}^*_{j,\lambda}$ or $\hat{g}^*_{j,c}$. Then, the (j, c) or (j, λ) pair for (j_n, c_n) or (j_n, λ_n) could be chosen using either of the two criteria: (1) the point-estimate of the value of the estimated rule $R_{\hat{g}^*,CV,1}$ or (2) the upper-bound of the confidence interval for that point estimate $R_{\hat{g}^*,CV,2}$.

We note that one could additionally augment the SuperLearner to include a metalearner that creates convex combinations between the J candidate algorithms (i.e., the "continuous" SuperLearner); that is, one could additionally introduce a family $\hat{g}^*_\alpha$ such as $\hat{g}^*_\alpha = \sum_j \alpha_j \hat{g}^*_j, \alpha_j \geq 0 \forall j, \sum_j \alpha_j = 1$. Then, the (α, c) or (α, λ) pair for (α_n, c_n) or (α_n, λ_n) could be chosen using either of the two risk criteria.

Estimation and Inference for the True Value of the True Optimal Rule

Let $g^*_{n,v,SL} = \hat{g}^*_{SL}(P_{n,v})$ be the ODTR SuperLearner for g^*_0 based on the training sample $P_{n,v}, v = 1, ..., V$. First, consider $\text{Logit}\, \bar{Q}_{n,v,\epsilon} = \text{Logit}\, \bar{Q}_{n,v} + \epsilon g^*_{n,v,SL}/g_{n,v}$, and then define $\epsilon_n = \arg\min_\epsilon \frac{1}{V} \sum_v P^1_{n,v} L(\bar{Q}_{n,v,\epsilon})$. This defines $\bar{Q}^*_{n,v} = \bar{Q}_{n,v,\epsilon_n}$ for each v. Then, the CV-TMLE of $\Psi_{g^*_0}(P_0)$ is given by

$$\hat{\psi}_{\hat{g}^*_{SL},CV} = \frac{1}{V} \sum_{v=1}^{V} \Psi_{g^*_{n,v,SL}}(Q^1_{W,n,v}, \bar{Q}^*_{n,v}),$$

A corresponding confidence interval for the CV-TMLE of $\Psi_{g^*_0}(P_0)$ is given by

$$\hat{\psi}_{\hat{g}^*_{SL},CV} \pm \Phi^{-1}(0.975) \frac{\hat{\sigma}^2_{\hat{g}^*_{SL},CV}}{\sqrt{n}},$$

where $\hat{\sigma}^2_{\hat{g}^*_{SL},CV} = \frac{1}{V} \sum_{v=1}^{V} P^1_{n,v} \{D_{g^*_{n,v,SL}}(Q^*_{n,v}, g^*_{n,v,SL})\}^2$.

3.4 Simulation Study

We ran a simulation study to examine the performance of the CV-TMLE for $\Psi_{g^*_0}(P_0)$ when the ODTR SuperLearner is used to estimate the true optimal rule, varying two conditions we introduce in this paper: (1) the ODTR SuperLearner library is augmented with candidate stochastic treatment rule estimators; and (2) the ODTR SuperLearner risk function is the upper bound of the confidence interval on the CV-TMLE estimate of the performance of a candidate rule. We did this for simulated RCT data and observational data, and for 3 varying strengths and complexities of the conditional additive treatment effect – 6 data-generating processes (DGPs) total. All simulations were implemented in R [65], and the code, simulated data, and results can be found at https://github.com/lmmontoya/SL.ODTR.

Data Generating Processes

Each simulation consisted of 1,000 iterations of n=441, the same sample size as the "Interventions" study.

The first DGP generated data as follows, where there was a large, positive, marginal treatment effect, and with a blip away from 0 for most covariate levels:

$$\begin{aligned}
W_i &\sim Normal(\mu = 0, \sigma^2 = 1), i \in \{1, ..., 4\} \\
W_j &\sim Bernoulli(p = 0.5), j \in \{5, 6\} \\
W_k &\sim Normal(\mu = 0, \sigma^2 = 20^2), k \in \{7, ..., 10\} \\
A &\sim Bernoulli(p = g_0(1|W)) \\
Y &\sim Bernoulli(p = \bar{Q}_0(A, W) = logit^{-1}(W_1 + 0.01A + 5W_1A)) \ ,
\end{aligned}$$

The true blip function for the first DGP was:

$$B_0(W) = [logit^{-1}(W_1 + 0.01 + 5W_1) - logit^{-1}(W_1)] \ ,$$

and $\Psi_{g_0^*}(P_0) = 38.23\%$.

Second, we examined a DGP in which there was a small, positive, marginal treatment effect, but with a blip close to 0 for all covariates. The data were generated as follows:

$$\begin{aligned}
W_i &\sim Normal(\mu = 0, \sigma^2 = 1), i \in \{1, ..., 4\} \\
W_j &\sim Bernoulli(p = 0.5), j \in \{5, 6\} \\
W_k &\sim Normal(\mu = 0, \sigma^2 = 20^2), k \in \{7, ..., 10\} \\
A &\sim Bernoulli(p = g_0(1|W)) \\
Y &\sim Bernoulli(p = \bar{Q}_0(A, W) = logit^{-1}(W_1 + W_4 + 0.01A)) \ ,
\end{aligned}$$

The true blip function for the second DGP was:

$$\begin{aligned}
B_0(W) = &[logit^{-1}(W_1 + W_4 + 0.01) \\
&- logit^{-1}(W_1 + W_4)] \ ,
\end{aligned}$$

and $\Psi_{g_0^*}(P_0) = 50.00\%$.

Finally, we examined a third DGP, which we have used in previous research [46, 32], which was generated as follows:

$$\begin{aligned}
W_1, W_2, W_3, W_4 &\sim Normal(\mu = 0, \sigma^2 = 1) \\
A &\sim Bernoulli(p = g_0(1|W)) \\
Y &= Bernoulli(p = \bar{Q}_0(A, W) = 0.5logit^{-1}(1 - W_1^2 + 3W_2 + 5W_3^2A - 4.45A) + \\
&\quad 0.5logit^{-1}(-0.5 - W_3 + 2W_1W_2 + 3|W_2|A - 1.5A)) \ .
\end{aligned}$$

The true blip function for the third DGP was:

$$\begin{aligned}B_0(W) =&0.5[logit^{-1}(1 - W_1^2 + 3W_2 + 5W_3^2 - 4.45)\\ &+ logit^{-1}(-0.5 - W_3 + 2W_1W_2 + 3|W_2| - 1.5)\\ &- logit^{-1}(1 - W_1^2 + 3W_2) + logit^{-1}(-0.5 - W_3 + 2W_1W_2)]\ .\end{aligned}$$

Here, the blip varies as a complex function of three baseline covariates; the causal value of the true value under the true optimal rule is $\Psi_{g_0^*}(P_0) = 36.53\%$.

For all DGPs, in the RCT setting, the true treatment mechanism $g_0(1|W) = 0.5$, and in the observational study setting, $g_0(1|W) = logit^{-1}(W1 + W2)$.

Estimator Configurations and Performance Measures

We estimated $\Psi_{g_0^*}(P_0)$ using CV-TMLE, and we used the ODTR SuperLearner to estimate g_0^*.

We used 3 SuperLearner library configurations for the estimation of g_0^*. The first configuration did not include any stochastic rules in the library; this SuperLearner is identical to the one presented in Chapter 1. The second two configurations were libraries that included stochastic rules. The first was via a blip transformation and varying constant $c \in \{0, 0.1, ..., 10\}$. The second was a variance regularization of the deterministic rule, namely, a combination through $\lambda \in \{0, 0.01, ..., 1\}$ of the predicted deterministic rule and the stochastic rule that minimizes the variance of a measure of the value of the rule.

The algorithms used to estimate the blip function B_n were as follows: univariate logistic regressions with each covariate W_i, for $i \in \{1, .., 4\}$, `SL.glm` (generalized linear models), `SL.mean` (the average), `SL.glm.interaction` (generalized linear models with interactions between all pairs of variables), `SL.earth` (multivariate adaptive regression splines [18]), and `SL.rpart` (recursive partitioning and regression trees [5]) from the SuperLearner package [62]. In addition, for DGP 3, we used `SL.nnet` (neural networks [66]), `SL.svm` (support vector machines [12]), `SL.glmnet` (regularized regression [17]), and the highly adaptive LASSO [3, 14].

We used two risk functions to select the ODTR SuperLearner – the value of the candidate rule ($R_{\hat{g}^*,CV,1}$, the point estimate of the CV-TMLE for the candidate rule) and the upper bound of the confidence interval on the value of the candidate rule ($R_{\hat{g}^*,CV,2}$, the upper bound of the confidence interval for the estimate of the CV-TMLE for the candidate rule).

We estimated the outcome regression $\bar{Q}_n$ with the same algorithms as B_n, except instead of including the univariate logistic regressions with each covariate, we included logistic regressions with main terms W_i and A, and with an interaction W_i times A, for $i \in \{1, .., 4\}$. We estimated the treatment mechanism g_n with a correctly specified model: a main terms logistic regression that regressed A on all W.

We used bias, variance, MSE, confidence interval coverage, and average confidence interval width across the 1,000 repetitions to measure the performance of the CV-TMLE as an estimator for the true value of the true ODTR $\Psi_{g_0^*}(P_0)$, under the aforementioned configurations.

Simulation Study Results

Figures 3.1, 3.2, and 3.3 show the distribution of either λ or c, thereby showing how many times stochastic rules were introduced in the estimation of the ODTR. For DGP 1, in which there was strong treatment effect heterogeneity, stochastic rules were infrequently introduced, and they were never utilized in the RCT setting and when the point estimate of the value of the rule was used as risk. DGPs 2 and 3, under both the RCT and observational study settings, utilized stochastic rules at least once in their 1,000 iterations. Stochastic treatment rules were used the most when the variance-regularized stochastic rule was introduced, with an upper bound of the confidence interval as the risk criterion. This was especially evident under an observational study and when there was a small to null blip function (DGP 2), where, out of 1,000 iterations, 514 of those output a stochastic rule (i.e., $\lambda > 0$ in the SuperLearner).

A summary of the performance results of the CV-TMLE for $\Psi_{g_0^*}(P_0)$ can be found in Tables 3.2, 3.3 and 3.4. For all DGP, study, and risk settings, the average confidence interval width was largest when no there was no stochastic rule augmentation in the ODTR SuperLearner, indicating that there is a reduction in standard error when stochastic rules are included.

For DGP 1, there were no stark differences in performance of the CV-TMLE in terms of bias, variance, MSE, confidence interval coverage, and average confidence interval width with respect to $\Psi_{g_0^*}(P_0)$ between the studies, risk settings and kind of stochastic rule augmentation. The lack of difference between the configurations is indicative of the findings that including stochastic learners and/or using the upper bound of the CV-TMLE confidence interval as a risk function does not hurt performance compared to including only deterministic learners and/or using the point-estimate on the CV-TMLE as a risk function, for both the observational setting and RCT setting in which there is strong treatment effect heterogeneity that is easy to detect.

We see the largest differences in DGP 2, the DGP with a blip value at or near 0 for all values of W, with the most obvious differences in the observational study case. Across the risk dimension, including the upper bound of the confidence interval of the CV-TMLE as the risk, i.e., $R_{\hat{g}^*,CV,2}$, resulted in a lower variance of the CV-TMLE of $\Psi_{g_0^*}(P_0)$ across simulations repetitions compared to using $R_{\hat{g}^*,CV,1}$ as risk, as long as $R_{\hat{g}^*,CV,2}$ was paired with an augmentation of the library by stochastic rules. For example, this was most apparent in the variance across simulation repetitions in the observational study case and when the variance-regularized stochastic rules were included in the library – the variance with $R_{\hat{g}^*,CV,1}$ was 1.33 times that of $R_{\hat{g}^*,CV,2}$. In addition, using the risk $R_{\hat{g}^*,CV,2}$ resulted in lower average confidence interval widths on the CV-TMLE. For example, again, the library that included variance-regularized stochastic rules had an average confidence interval width when using $R_{\hat{g}^*,CV,1}$ that was 1.16 times that of when $R_{\hat{g}^*,CV,2}$ was used.

For DGP 2, inclusion of stochastic rules in the library was beneficial, particularly in terms of variance and MSE across simulation repetitions of the estimator and the standard error. For example, in the observational setting with risk function $R_{\hat{g}^*,CV,2}$, the variance and MSE of the CV-TMLE across simulation repetitions for a library that did not include stochastic rules was 1.33 times that of a library that included variance-regularized stochastic rules and 1.11 times that of a library that included blip-transformed stochastic rules. Further, the average confidence interval width of the CV-TMLE for a library that did not include stochastic rules was 1.16 times that of a library that included variance-regularized stochastic rules and 1.05 times that of a library that included blip-transformed stochastic rules.

ODTR SuperLearner libraries that included the presented candidate stochastic rules decreased the estimator's standard error and thus shrunk confidence intervals; however, the largest improvement came from the variance-regularized stochastic rules. We illustrate this improvement in Figure 3.4. This plot shows – for one $n = 441$ instance of DGP 2, using $\bar{Q}_0$, g_0, and $\sigma_0^2(A, W)$ – the TMLE estimate of $\Psi_{g_0^*}(P_0)$, the upper bound of the TMLE's confidence interval, and the estimated variance of the TMLE's influence curve. In addition, the variance-regulated stochastic rule is a convex combination of the incorrect candidate optimal rule $d =$ treat all (d_0 is actually treat everyone) with g^*_{r,P_0}. The plot shows that as λ moves away from 0, the estimated variance of the influence curve of the TMLE necessarily decreases, which means that the standard error of the TMLE under the variance-regularized rule will be equivalent to or smaller than the standard error of the TMLE when the deterministic rule d is used. This implies that any usage of a stochastic rule will yield a confidence interval that is smaller than when a deterministic rule is used. In this finite sample instance, the $R_{\hat{g}^*,CV,2}$ would choose $\lambda = 0.96$, increasing the stochasticity of the treatment assignment rules, whereas $R_{\hat{g}^*,CV,1}$ would have chosen $\lambda = 0$, the incorrect deterministic rule. Though this is not the case for every instance, as we show in the simulations, it is more common that a stochastic rule is chosen for $R_{\hat{g}^*,CV,2}$ than $R_{\hat{g}^*,CV,1}$.

In DGP 3, we explored the performance of the estimator when the true, underlying blip function was a complex function of baseline covariates. As expected, in both the RCT and observational study setting, because there was a presence of treatment effect heterogeneity, there was no difference in performance between the different risk functions and the addition of stochastic estimators in terms of variance, and a little improvement in the average width of the confidence intervals when stochastic rules were included in the library. Simultaneously, however, there was an increase in bias compared to $\Psi_{g_0^*}(P_0)$, particularly when the stochastic rules were paired with the confidence interval-based risk function, which translated to a drop-off in confidence interval coverage. This was most pronounced in the observational setting when using the variance regularized stochastic rule (e.g., 54.07% coverage including variance regularized stochastic rules and $R_{\hat{g}^*,CV,2}$ versus approximately 63.00% coverage when not including stochastic rules and/or using $R_{\hat{g}^*,CV,1}$). Though not the target parameter of interest, performance for estimators of the data-adaptive parameter $\Psi_{g^*_{n,v,CV}}(P_0)$ did not vary by library or risk configuration; coverage under the RCT study ranged from 95.60% to 96.14%, while coverage under the observational setting ranged from 91.86% to 92.80%.

3.5 Application to "Interventions" Study

The "Interventions" study is an ongoing RCT experiment, in which 441 participants were either randomized to CBT or TAU. Thus far, 231 (52.2%) participants have received CBT and 210 (47.8%) TAU. See Table 3.1 for the distribution of the covariates and outcome (re-arrest by one year after enrollment) by treatment assignment. Out of the 441 participants, 271 (38.5%) were not re-arrested within the year following their treatment assignment. The estimated probability of re-arrest had everyone been assigned CBT is 37.8%, and the estimated probability of re-arrest had everyone been assigned TAU is 39.3%; there was no significant difference between these two probabilities (risk difference: -1.51%, CI: [-11.06%,8.03%]). After adjusting for covariates using TMLE to improve the precision on this ATE estimate [51], the risk difference was, similarly, -1.53% (CI: [-10.37%, 7.49%]).

In Chapter 1, we implemented the original ODTR SuperLearner on this dataset, that output a deterministic rule only, using a blip-only library and a continuous, blip-based metalearner. The SuperLearner consisted of a combination of simple parametric models (univariate GLMs with each covariate) and machine learning algorithms (`SL.glm`, `SL.mean`, `SL.glm.interaction`, `SL.earth`, and `SL.rpart`). The algorithm assigned all weight on a GLM that modeled the blip as a linear function of only substance use.

It may be of interest to include stochastic rules to potentially shrink confidence intervals, as demonstrated in the simulations, though we do not necessarily expect a drastic improvement in this case, since the "Interventions" Study is an RCT. Therefore, we ran the ODTR SuperLearner using the same configurations as in Chapter 1, except we allowed for stochastic rules in the library. Specifically, we included stochastic interventions through the blip transformation and regularization using the variance of the efficient influence curve, as described above. Additionally, we used the mean outcome of the rule and the upper bound of the confidence interval for the mean outcome of the rule as risk functions. The outcome regression $\bar{Q}_n$ (and therefore $\sigma_n^2(A, W)$) was estimated using the canonical outcome prediction SuperLearner [63], g_n was estimated as an intercept-only logistic regression, and we used 10-fold cross validation.

Though stochastic rules were included in the SuperLearner library through both "blip transformations" and "variance regularization," for both risk functions, the SuperLearner ultimately chose the library which did not introduce any stochasticity. In other words, the SuperLearner chose c and λ to be 0. The SuperLearner configuration where the value of the rule is the risk function returned identical results to those in Chapter 1, with all coefficient weight on the variable substance use. In addition, the CV-TMLE estimates of the value of the optimal rule (38.63%, CI = [32.07%, 45.18%]) were not significantly different than estimates of the expected value had everyone been given CBT (difference is 0.35%, CI = [-5.70%, 6.40%]) and had no one been given CBT (difference is 0.19%, CI = [-6.68%, 7.06%]). When the upper bound of the confidence interval on the value of the rule was used as a risk function, the mean outcome under that rule (38.25%, CI = [31.68%, 44.83%]) was, again, not significantly different than estimates of the expected value had everyone been given CBT (difference is -0.98%, CI = [-7.63%, 5.68%]) and had no one been given CBT (difference is

0.18%, CI = [-6.16%, 6.51%]). Under this risk function, all algorithms were given nonzero weight.

3.6 Conclusions

The aim of this paper was to extend the ODTR SuperLearner, as described in Chapter 1 and [46], by (1) augmenting the possible library of candidate algorithms to include stochastic treatment rules, and (2) introducing a new risk criterion for selection of the ODTR SuperLearner – the upper bound of the confidence interval on the value of the rule, in the case when smaller outcomes are better. We hypothesized that these extensions would improve performance of estimation of the true value of the true ODTR in finite samples, specifically by reducing the estimator's variance.

First, in this paper we examined a setting in which conditional additive treatment effect was strong for most covariate profiles, which means that the true ODTR was easy to approximate. In this case, despite including stochastic rules, the algorithm almost always (correctly) chose deterministic rules to estimate the true ODTR (which is always deterministic). Performance under these circumstances was similar across all risk and stochastic rule augmentation conditions, which means that, in the presence of a strong, conditional additive treatment effect, and thus easy detection of the ODTR, the ODTR SuperLearner will not choose stochastic rules, and thus inclusion of them in the library will not hurt performance of estimators for the true value of the true ODTR.

Next, we showed that there are finite-sample benefits to introducing stochastic rules when the conditional additive treatment effect is around 0 for most covariate levels. As compared with the case when all algorithms are deterministic, including stochastic treatment rules in the ODTR SuperLearner reduces confidence interval width (thus increasing precision) around estimates of the mean outcome under the true ODTR and the variance of the estimator across simulation repetitions, while simultaneously preserving confidence interval coverage. Including stochastic rules, paired with the upper bound of the confidence interval on the value of the rule as risk criterion, improves performance the most, especially in the observational study setting, where positivity issues, and therefore increases in the variance of the estimator for the value of the rule, are more likely to arise.

Third, we showed a scenario in which there are costs (in terms of bias) to including both stochastic treatment rules in the ODTR SuperLearner library paired with the upper bound of the confidence interval for the estimate of the value of the rule as risk function. We hypothesize that these costs are due to the complexity of the true blip function, and therefore the difficulty of estimating that underlying model. Thus, in future research we will estimate DGP 3's blip with more flexible and aggressive machine learning algorithms; in particular, a highly adaptive LASSO that allows for more interactions, increasing the L1 norm. We expect that this will decrease bias with little cost to the variance, equalizing performance between all stochastic rule library/risk configurations.

	TAU ($A = 0$)	CBT ($A = 1$)	p
n	211	230	
Re-arrest ($Y = 1$) (%)	83 (39.3)	87 (37.8)	0.820
Site = San Francisco (%)	87 (41.2)	104 (45.2)	0.455
Gender = Female (%)	38 (18.0)	37 (16.1)	0.682
Ethnicity = Hispanic (%)	50 (23.7)	42 (18.3)	0.198
Age (mean (SD))	38.08 (11.05)	37.01 (11.22)	0.317
CSI (mean (SD))	32.35 (11.13)	33.46 (11.27)	0.300
LSI (mean (SD))	5.59 (1.33)	5.50 (1.48)	0.472
SES (mean (SD))	3.81 (1.89)	3.81 (2.12)	0.995
Prior adult convictions (%)			0.156
Zero to two times	74 (35.1)	93 (40.4)	
Three or more times	134 (63.5)	129 (56.1)	
Missing	3 (1.4)	8 (3.5)	
Most serious offense (mean (SD))	5.29 (2.54)	5.09 (2.52)	0.415
Motivation (mean (SD))	3.22 (1.36)	3.27 (1.37)	0.720
Substance use (%)			0.184
0	53 (25.1)	76 (33.0)	
1	47 (22.3)	55 (23.9)	
2	109 (51.7)	98 (42.6)	
Missing	2 (0.9)	1 (0.4)	

Table 3.1: Distribution of “Interventions” data by treatment assignment.

Finally, we applied the augmented ODTR SuperLearner to the “Interventions” study. As shown in the simulations, under RCTs, where there are no positivity issues and thus the variance of the estimate of the value of the rule is less likely to have an inflated variance, there was no advantage (though no disadvantage, either) to including stochastic rules and having a confidence interval-based risk for the ODTR SuperLearner. As expected, since the “Interventions” study is an RCT, for both risk functions, the SuperLearner did not utilize stochastic rules in the estimation of the ODTR. Thus, results were identical to those in Chapter 1.

Extending the SuperLearner to include stochastic rules, as opposed to only outputting a deterministic, discrete (binary, if treatment is 0 or 1) treatment recommendation, could be of interest in an adaptive design; for example, in a covariate-adjusted response adaptive trial design, in which the the ODTR SuperLearner is learned on an initial set of data, and treatment is then assigned based on that learned ODTR on a second set of incoming data [22, 74]. In the second stage, where treatment is assigned to new, incoming subjects, it may be better to assign treatment with some stochasticity based on how certain one is of assigning treatment (as in the work presented in this paper), versus an all-or-nothing rule.

Study	Risk	Stoch. Addition	Bias	Variance	MSE	% Cov.	Avg. CI Width
RCT	Point Est.	λ	0.0122	0.0010	0.0012	91.24	0.0585
		c	0.0117	0.0011	0.0012	91.10	0.0585
		None	0.0122	0.0010	0.0012	91.24	0.0585
	CI	λ	0.0125	0.0010	0.0012	91.14	0.0585
		c	0.0121	0.0011	0.0012	91.10	0.0585
		None	0.0125	0.0010	0.0012	91.14	0.0585
Obs.	Point Est.	λ	0.0190	0.0020	0.0023	85.91	0.0742
		c	0.0192	0.0020	0.0023	85.91	0.0741
		None	0.0190	0.0020	0.0023	85.91	0.0741
	CI	λ	0.0215	0.0021	0.0025	84.76	0.0739
		c	0.0217	0.0021	0.0025	85.07	0.0738
		None	0.0214	0.0021	0.0025	84.86	0.0739

Table 3.2: DGP 1: Performance of CV-TMLE $\hat{\psi}_{\hat{g}^*_{SL},CV}$ with respect to $\Psi_{g^*_0}(P_0)$, the true value of the true optimal rule, for (1) the RCT and observational (Obs.) study setting; (2) the CV-TMLE point estimate of the value of the rule as risk function ($R_{\hat{g}^*,CV,1}$) and the upper bound of the confidence interval of the CV-TMLE as risk function ($R_{\hat{g}^*,CV,2}$); (3) the SuperLearner library that includes stochastic rules through a blip transformation (c), variance regularization (λ), and a library that does not include stochastic rules (None).

Study	Risk	Stoch. Addition	Bias	Variance	MSE	% Cov.	Avg. CI Width
RCT	Point Est.	λ	0.0002	0.0014	0.0014	89.8	0.0611
		c	0.0004	0.0014	0.0014	90.3	0.0609
		None	0.0003	0.0014	0.0014	89.6	0.0616
	CI	λ	0.0004	0.0013	0.0013	90.2	0.0584
		c	0.0005	0.0013	0.0013	90.6	0.0584
		None	0.0004	0.0014	0.0014	89.0	0.0616
Obs.	Point Est.	λ	0.0003	0.0020	0.0020	90.4	0.0767
		c	0.0001	0.0021	0.0021	90.9	0.0773
		None	0.0004	0.0021	0.0021	90.9	0.0782
	CI	λ	0.0006	0.0015	0.0015	92.1	0.0664
		c	0.0005	0.0018	0.0018	92.6	0.0738
		None	0.0004	0.0020	0.0020	92.2	0.0773

Table 3.3: DGP 2: Performance of CV-TMLE $\hat{\psi}_{\hat{g}^*_{SL},CV}$ with respect to $\Psi_{g_0^*}(P_0)$, the true value of the true optimal rule, for (1) the RCT and observational (Obs.) study setting; (2) the CV-TMLE point estimate of the value of the rule as risk function ($R_{\hat{g}^*,CV,1}$) and the upper bound of the confidence interval of the CV-TMLE as risk function ($R_{\hat{g}^*,CV,2}$); (3) the SuperLearner library that includes stochastic rules through a blip transformation (c), variance regularization (λ), and a library that does not include stochastic rules (None).

Study	Risk	Stoch. Addition	Bias	Variance	MSE	% Cov.	Avg. CI Width
RCT	Point Est.	λ	0.0515	0.0017	0.0044	63.19	0.0660
		c	0.0516	0.0017	0.0044	63.20	0.0658
		None	0.0515	0.0017	0.0044	63.30	0.0660
	CI	λ	0.0525	0.0018	0.0046	60.69	0.0654
		c	0.0528	0.0017	0.0045	60.10	0.0646
		None	0.0518	0.0018	0.0045	62.57	0.0660
Obs.	Point Est.	λ	0.0621	0.0025	0.0064	62.73	0.0817
		c	0.0621	0.0025	0.0064	63.57	0.0817
		None	0.0621	0.0025	0.0064	63.88	0.0822
	CI	λ	0.0652	0.0024	0.0066	54.07	0.0747
		c	0.0638	0.0024	0.0065	60.02	0.0785
		None	0.0628	0.0025	0.0065	63.05	0.0807

Table 3.4: DGP 3: Performance of CV-TMLE $\hat{\psi}_{\hat{g}^*_{SL},CV}$ with respect to $\Psi_{g^*_0}(P_0)$, the true value of the true optimal rule, for (1) the RCT and observational (Obs.) study setting; (2) the CV-TMLE point estimate of the value of the rule as risk function ($R_{\hat{g}^*,CV,1}$) and the upper bound of the confidence interval of the CV-TMLE as risk function ($R_{\hat{g}^*,CV,2}$); (3) the SuperLearner library that includes stochastic rules through a blip transformation (c), variance regularization (λ), and a library that does not include stochastic rules (None).

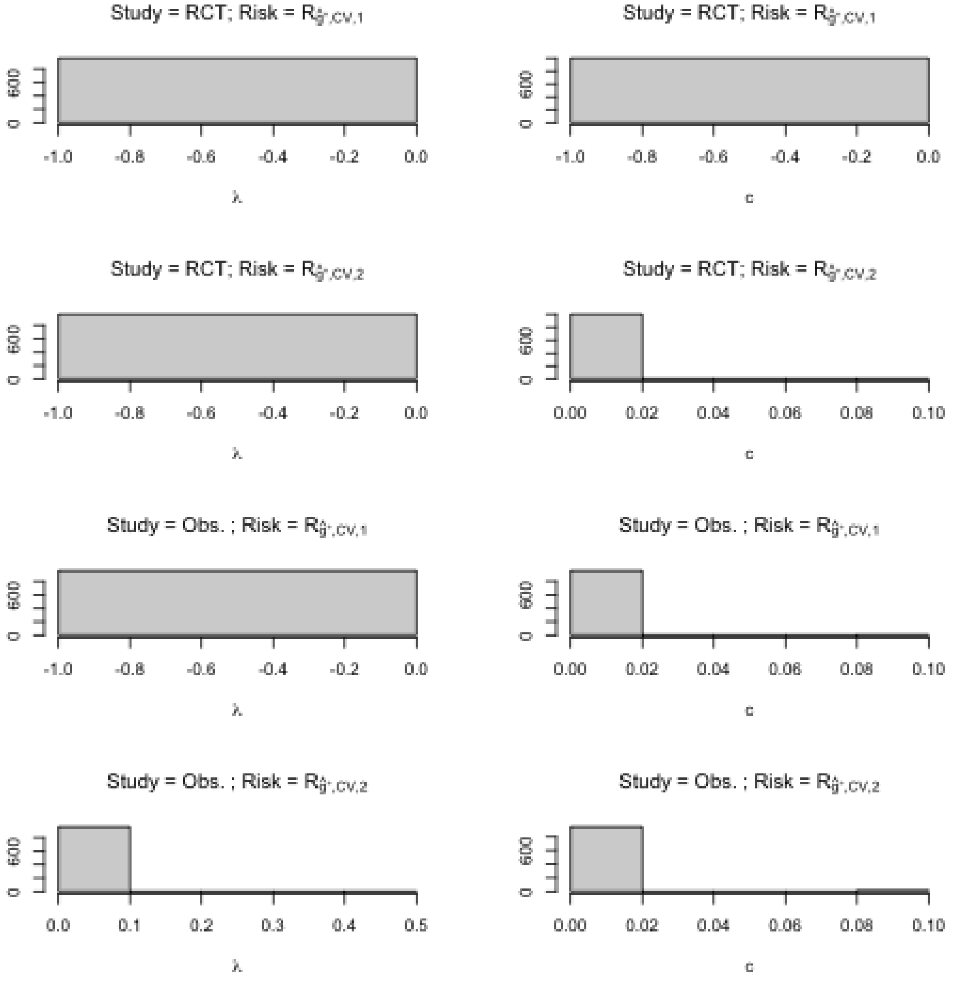

Figure 3.1: DGP 1: Distribution of λ and c for varying study type (RCT and observational) and risk type ($R_{\hat{g}^*,CV,1}$ and $R_{\hat{g}^*,CV,2}$).

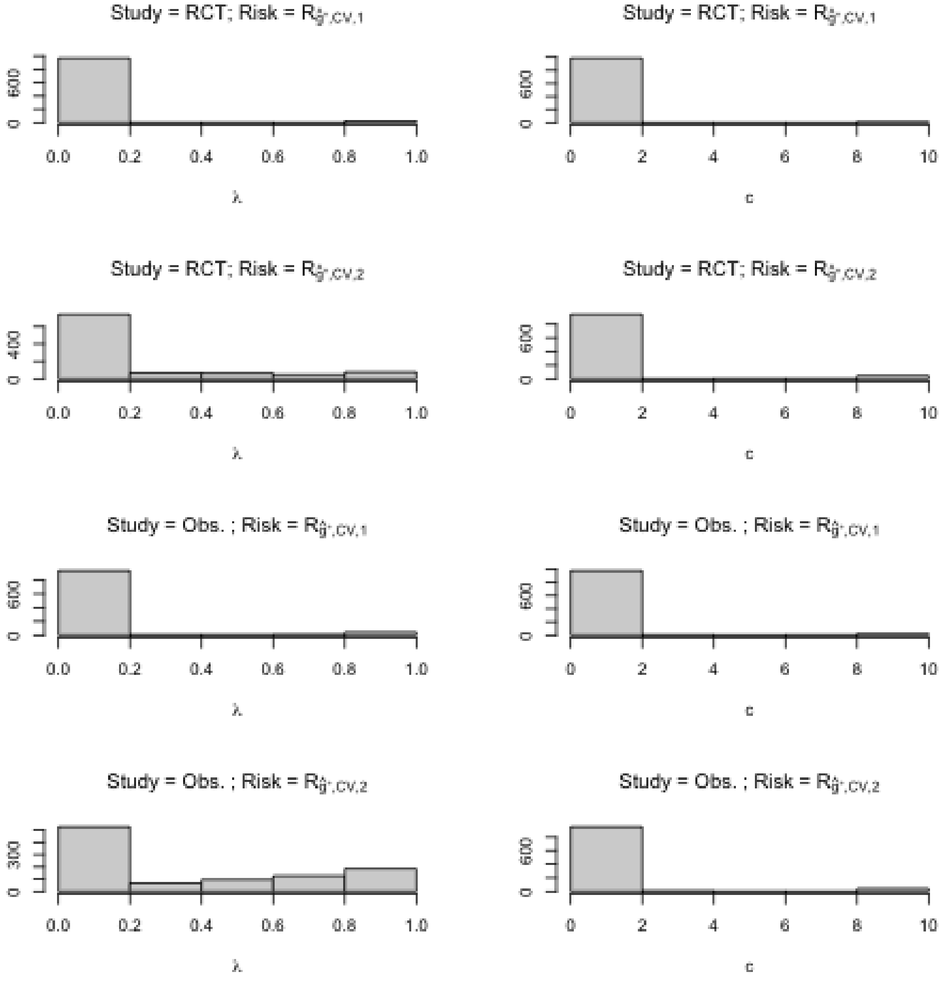

Figure 3.2: DGP 2: Distribution of λ and c for varying study type (RCT and observational) and risk type ($R_{\hat{g}^*,CV,1}$ and $R_{\hat{g}^*,CV,2}$).

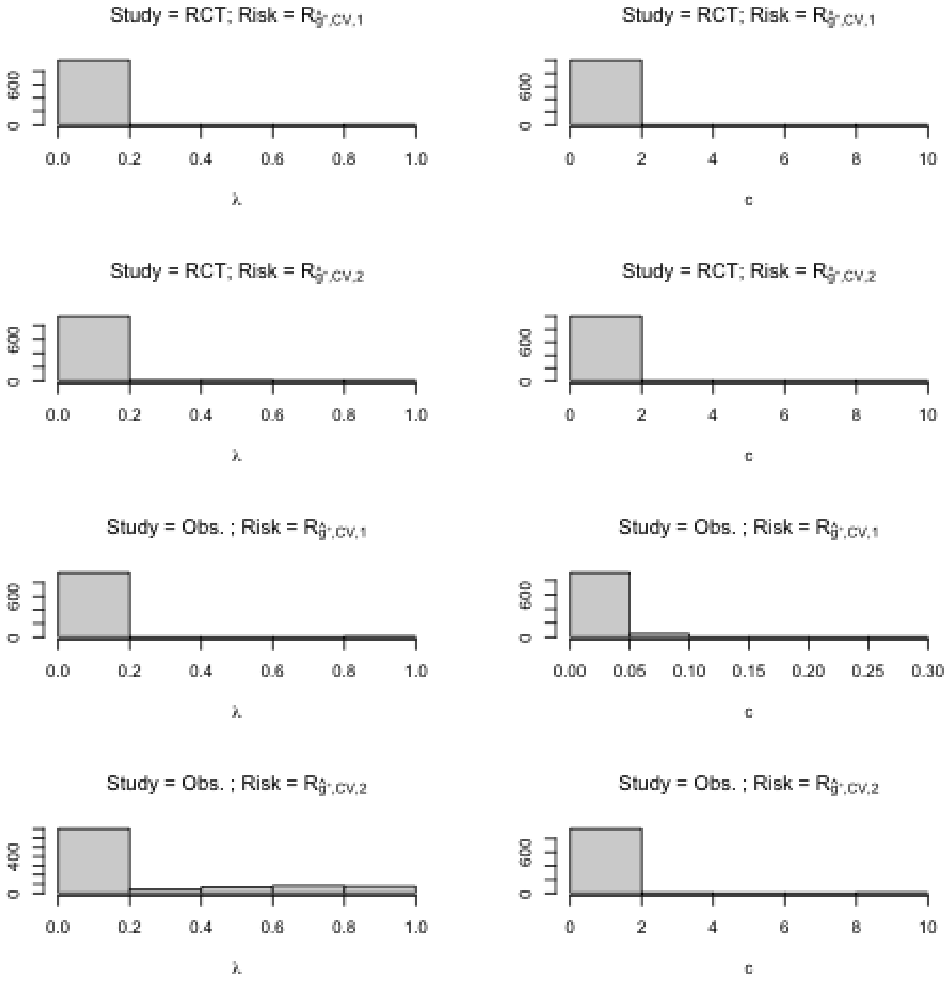

Figure 3.3: DGP 3: Distribution of λ and c for varying study type (RCT and observational) and risk type ($R_{\hat{g}^*,CV,1}$ and $R_{\hat{g}^*,CV,2}$).

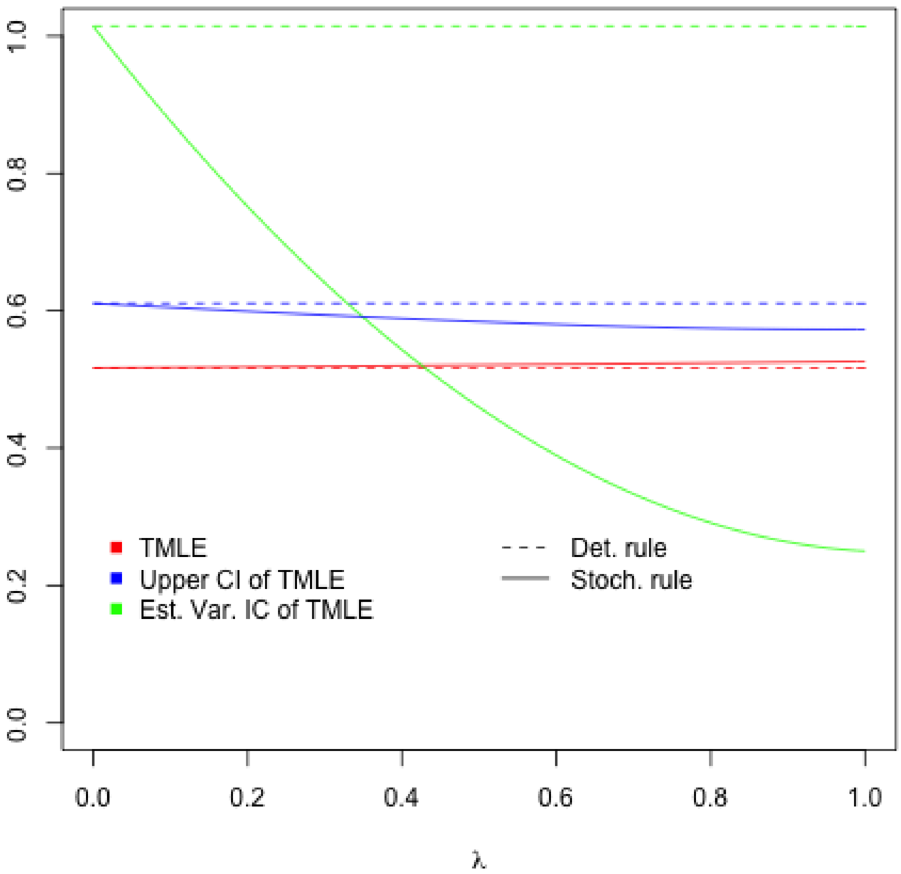

Figure 3.4: Illustration of reduction of the variance of the estimated influence curve as λ increases, as compared to the variance of the estimated influence curve (IC) for TMLE of $\Psi_{g_0^*}(P_0)$ when $\lambda = 0$, i.e., the deterministic rule setting. This is a scenario where $\lambda = 0.96$ would be picked under $R_{\hat{g}^*,CV,2}$, whereas $R_{\hat{g}^*,CV,1}$ would pick $\lambda = 0$.

Conclusion

In this dissertation, we illustrated how to implement and evaluate the ODTR SuperLearner, with several novel extensions and an ongoing application to the "Interventions" study. In Chapter 1, we described the ODTR SuperLearner, and outlined the possible library, metalearner, and risk configurations for implementing the algorithm. In particular, using simulations of finite-sample data, we argued for the importance of having a SuperLearner library with a diversity of candidate ODTR algorithms, and showed the benefits and drawbacks of having a risk function based on the value of the rule (versus MSE) and a vote-based (versus blip-based) metalearner. In Chapter 2, we showed the advantages of using CV-TMLE to estimate in finite-samples: 1) the true value of an *a priori* known rule; 2) the true value of the true, unknown ODTR; and 3) the true value of an estimated ODTR (a data-adaptive parameter). In particular, the gain of using CV-TMLE was especially evident when estimating the data-adaptive parameter – while other estimators' performance declined dramatically as the library of algorithms used to estimate the ODTR SuperLearner increased in data-adaptiveness, the CV-TMLE retained adequate performance for the true value of the sample-split-specific estimate of the ODTR. In Chapter 3, we augmented the ODTR SuperLearner to: 1) include stochastic treatment rules in the library; and 2) include an additional risk function that takes into account the variability of the estimate of the value of a rule. In particular, we showed the advantages in finite-samples of including these two augmentations, particularly in observational study scenarios and when there is weak treatment effect heterogeneity (i.e., the blip is close to 0 for all covariate values).

In the future, we hope to build on this work in various ways, touched on in each of the chapters. First, we will re-run all analyses with the full "Interventions" dataset, which consists of approximately 720 offenders with mental illness, helping us to gain power in any inferential analyses – especially in the comparison of the value of the ODTR SuperLearner compared to giving everyone CBT or no one CBT.

Next, we aim to extend the SuperLearner ODTR algorithm to the 3-treatment setting and longitudinal setting, e.g., that of a Sequential Multiple Assignment Randomized Trial (SMART). SMART designs provide an ideal opportunity to learn the best treatment sequence [29]. In particular, we wish to implement the 3-treatment/sequential ODTR on a SMART aimed at improving HIV care called "Adaptive Strategies for Preventing and Treating Lapses of Retention in HIV Care." In this SMART, 1,816 HIV-positive patients in Kenya were randomly assigned to a low-intensity "prevention" intervention (SMS messages, transportation vouchers, counseling) when they initiated ART; those who subsequently experienced a lapse in retention were re-randomized to a higher intensity “treatment” intervention (e.g., SMS and voucher, peer navigator, outreach), and those who did not have a lapse in retention were re-randomized to either keep their intervention or discontinue it.

The SL.ODTR software currently allows for estimation of resource-constrainted ODTRs, which answers the question: “who should get treatment, under the constraint that only k% of the population can be treated?” In future research, we hope to formally implement this algorithm on the “Better Information for Health in Zambia” study, which used a sampling approach to randomly select a sample of “lost” patients and intensively seek them out to ascertain their true outcomes across four provinces in Zambia. In this way, we can understand which kinds of patients benefit from sampling for tracing back to care. This is of benefit in this setting because (1) resources for tracing patients in this population are limited, and (2) it is likely that sampling is only neutral or beneficial to the patient, not harmful (thus, the ODTR will indicate everyone should be sampled, not using the resources we have at our disposal in the most efficient way).

Lastly, if one has an ODTR that is well-estimated and significantly improves overall outcomes, in future research, we hope to explore how to most efficiently and effectively implement that learned ODTR on a new cohort of patients. Adaptive trial designs, or trials in which aspects of subsequent experiments are informed by data from earlier experiments, offer a way to study how the ODTR can be used to assign treatment on new patients [74, 22]. In an ODTR-based adaptive design, we learn the ODTR on an initial cohort, and that ODTR informs how to best assign treatment to a new, incoming cohort of patients based on (1) the regime learned on the initial cohort and (2) the new patients’ characteristics. In future research, we hope to examine an ODTR-based adaptive design that assigns treatment to the next cohort using probabilities that are a function of the ODTR learned on the first cohort.

In conclusion, in this dissertation, using causal inference, machine learning, and statistical theory, we described and expanded on ways to learn and evaluate the ODTR. We applied these findings to the “Interventions” Study, to begin to uncover which kinds of justice-involved adults with mental illness benefit more from CBT versus TAU, with the end goal of reducing recidivism. In future research, we aim to apply and expand on findings from this dissertation in various ways, with the ultimate goal of using data to effectively and responsibly determine and administer the best treatment decision for each person.

Bibliography

[1] Daniel Almirall et al. “Introduction to SMART designs for the development of adaptive interventions: with application to weight loss research”. In: *Translational behavioral medicine* 4.3 (2014), pp. 260–274.

[2] Oliver Bembom and Mark J van der Laan. “A practical illustration of the importance of realistic individualized treatment rules in causal inference”. In: *Electronic journal of statistics* 1 (2007), p. 574.

[3] David Benkeser and Mark van der Laan. “The highly adaptive lasso estimator”. In: *2016 IEEE international conference on data science and advanced analytics (DSAA)*. IEEE. 2016, pp. 689–696.

[4] Peter J Bickel et al. *Efficient and adaptive estimation for semiparametric models*. Vol. 4. Johns Hopkins University Press Baltimore, 1993.

[5] Leo Breiman. *Classification and regression trees*. Routledge, 2017.

[6] Leo Breiman. “Random forests”. In: *Machine learning* 45.1 (2001), pp. 5–32.

[7] Leo Breiman. “Stacked regressions”. In: *Machine learning* 24.1 (1996), pp. 49–64.

[8] B. Chakraborty, E. B. Laber, and Y. Zhao. “Inference for optimal dynamic treatment regimes using an adaptive m-out-of-n bootstrap scheme”. In: *Biometrics* 69.3 (2013), pp. 714–23. ISSN: 1541-0420 (Electronic) 0006-341X (Linking). DOI: `10.1111/biom.12052`. URL: `https://www.ncbi.nlm.nih.gov/pubmed/23845276`.

[9] Bibhas Chakraborty and EE Moodie. *Statistical methods for dynamic treatment regimes*. Springer, 2013.

[10] Bibhas Chakraborty, Susan Murphy, and Victor Strecher. “Inference for non-regular parameters in optimal dynamic treatment regimes”. In: *Statistical methods in medical research* 19.3 (2010), pp. 317–343. ISSN: 0962-2802.

[11] Bibhas Chakraborty and Susan A Murphy. “Dynamic treatment regimes”. In: *Annual review of statistics and its application* 1 (2014), pp. 447–464. ISSN: 2326-8298.

[12] Chih-Chung Chang and Chih-Jen Lin. “LIBSVM: A library for support vector machines”. In: *ACM transactions on intelligent systems and technology (TIST)* 2.3 (2011), p. 27.

[13] Zachary D Cohen and Robert J DeRubeis. "Treatment selection in depression". In: *Annual Review of Clinical Psychology* 14 (2018).

[14] Jeremy R Coyle, Nima S Hejazi, and Mark J van der Laan. *hal9001: The scalable highly adaptive lasso*. R package version 0.2.6. 2020. DOI: `10.5281/zenodo.3558313`. URL: `https://github.com/tlverse/hal9001`.

[15] Jeremy Robert Coyle. "Computational Considerations for Targeted Learning". PhD thesis. UC Berkeley, 2017.

[16] Issa J Dahabreh, Rodney Hayward, and David M Kent. "Using group data to treat individuals: understanding heterogeneous treatment effects in the age of precision medicine and patient-centred evidence". In: *International journal of epidemiology* 45.6 (2016), pp. 2184–2193.

[17] Jerome Friedman, Trevor Hastie, and Rob Tibshirani. "Regularization paths for generalized linear models via coordinate descent". In: *Journal of statistical software* 33.1 (2010), p. 1.

[18] Jerome H Friedman et al. "Multivariate adaptive regression splines". In: *The annals of statistics* 19.1 (1991), pp. 1–67.

[19] Susan Gruber and Mark J van der Laan. "A targeted maximum likelihood estimator of a causal effect on a bounded continuous outcome". In: *The International Journal of Biostatistics* 6.1 (2010).

[20] Miguel A. Hernan and James M. Robins. "Estimating causal effects from epidemiological data". In: *J Epidemiol Community Health* 60.7 (2006), pp. 578–86. ISSN: 0143-005X (Print) 0143-005X (Linking). DOI: `10.1136/jech.2004.029496`. URL: `https://www.ncbi.nlm.nih.gov/pubmed/16790829`.

[21] S. T. Holloway et al. *DynTxRegime: Methods for Estimating Optimal Dynamic Treatment Regimes*. R package version 4.1. 2019. URL: `https://CRAN.R-project.org/package=DynTxRegime`.

[22] Feifang Hu and William F Rosenberger. *The theory of response-adaptive randomization in clinical trials*. Vol. 525. John Wiley & Sons, 2006.

[23] A. E. Hubbard, S. Kherad-Pajouh, and M. J. van der Laan. "Statistical Inference for Data Adaptive Target Parameters". In: *Int J Biostat* 12.1 (2016), pp. 3–19. ISSN: 1557-4679 (Electronic) 1557-4679 (Linking). DOI: `10.1515/ijb-2015-0013`. URL: `https://www.ncbi.nlm.nih.gov/pubmed/27227715`.

[24] Alan E Hubbard, Chris J Kennedy, and Mark J van der Laan. "Data-Adaptive Target Parameters". In: *Targeted Learning in Data Science*. Springer, 2018, pp. 125–142.

[25] David M Kent, Ewout Steyerberg, and David van Klaveren. "Personalized evidence based medicine: predictive approaches to heterogeneous treatment effects". In: *Bmj* 363 (2018), k4245.

[26] Muin J Khoury, Michael F Iademarco, and William T Riley. "Precision public health for the era of precision medicine". In: *American journal of preventive medicine* 50.3 (2016), p. 398.

[27] David van Klaveren et al. "Estimates of absolute treatment benefit for individual patients required careful modeling of statistical interactions". In: *Journal of clinical epidemiology* 68.11 (2015), pp. 1366–1374.

[28] Michael R Kosorok and Eric B Laber. "Precision medicine". In: *Annual review of statistics and its application* 6 (2019), pp. 263–286.

[29] Michael R Kosorok and Erica EM Moodie. *Adaptive Treatment Strategies in Practice: Planning Trials and Analyzing Data for Personalized Medicine.* Vol. 21. SIAM, 2015.

[30] Mark J van der Laan and Susan Gruber. "Targeted minimum loss based estimation of an intervention specific mean outcome". In: (2011).

[31] Mark J van der Laan, MJ Laan, and James M Robins. *Unified methods for censored longitudinal data and causality.* Springer Science & Business Media, 2003.

[32] Mark J van der Laan and Alexander R Luedtke. "Targeted learning of an optimal dynamic treatment, and statistical inference for its mean outcome". In: (2014).

[33] Mark J van der Laan and Alexander R Luedtke. "Targeted learning of the mean outcome under an optimal dynamic treatment rule". In: *Journal of causal inference* 3.1 (2015), pp. 61–95.

[34] Mark J van der Laan and Maya L Petersen. "Causal effect models for realistic individualized treatment and intention to treat rules". In: *The international journal of biostatistics* 3.1 (2007).

[35] Mark J van der Laan, Eric C Polley, and Alan E Hubbard. "Super learner". In: *Statistical applications in genetics and molecular biology* 6.1 (2007).

[36] Mark J van der Laan and Sherri Rose. *Targeted Learning in Data Science.* Springer, 2018.

[37] Mark J van der Laan and Sherri Rose. *Targeted learning: causal inference for observational and experimental data.* Springer Science & Business Media, 2011.

[38] Mark J. van der Laan and Maya L. Petersen. "Causal effect models for realistic individualized treatment and intention to treat rules". In: *Int J Biostat* 3.1 (2007), Article 3. ISSN: 1557-4679 (Electronic) 1557-4679 (Linking). URL: `https://www.ncbi.nlm.nih.gov/pubmed/19122793`.

[39] E. Laber and M. Qian. "Evaluating Personalized Treatment Regimes". In: *Methods in Comparative Effectiveness Research.* Ed. by Constantine Gatsonis and Sally C. Morton. Boca Raton, FL: CRC Press LLC : Chapman and Hall/CRC, 2017. Chap. 15, pp. 483–497.

[40] Eric Laber and Marie Davidian. “Dynamic treatment regimes, past, present, and future: A conversation with experts”. In: *Statistical methods in medical research* 26.4 (2017), pp. 1605–1610.

[41] Eric B Laber, Kristin A Linn, and Leonard A Stefanski. “Interactive model building for Q-learning”. In: *Biometrika* 101.4 (2014), pp. 831–847.

[42] Erin LeDell, Mark J van der Laan, and Maya Petersen. “AUC-maximizing ensembles through metalearning”. In: *The international journal of biostatistics* 12.1 (2016), pp. 203–218.

[43] Huitian Lei et al. “A” SMART” design for building individualized treatment sequences”. In: *Annual review of clinical psychology* 8 (2012), pp. 21–48.

[44] Ilya Lipkovich, Alex Dmitrienko, and Ralph B D’Agostino Sr. “Tutorial in biostatistics: data-driven subgroup identification and analysis in clinical trials”. In: *Statistics in medicine* 36.1 (2017), pp. 136–196.

[45] Mark W Lipsey, Nana A Landenberger, and Sandra J Wilson. “Effects of cognitive-behavioral programs for criminal offenders”. In: *Campbell systematic reviews* 3.1 (2007), pp. 1–27.

[46] Alexander R Luedtke and Mark J van der Laan. “Optimal individualized treatments in resource-limited settings”. In: *The international journal of biostatistics* 12.1 (2016), pp. 283–303.

[47] Alexander R Luedtke and Mark J van der Laan. “Statistical inference for the mean outcome under a possibly non-unique optimal treatment strategy”. In: *Annals of statistics* 44.2 (2016), p. 713.

[48] Alexander R Luedtke and Mark J van der Laan. “Super-learning of an optimal dynamic treatment rule”. In: *The international journal of biostatistics* 12.1 (2016), pp. 305–332.

[49] E. E. Moodie, T. S. Richardson, and D. A. Stephens. “Demystifying optimal dynamic treatment regimes”. In: *Biometrics* 63.2 (2007), pp. 447–55. ISSN: 0006-341X (Print) 0006-341X (Linking). DOI: `10.1111/j.1541-0420.2006.00686.x`. URL: `https://www.ncbi.nlm.nih.gov/pubmed/17688497`.

[50] Erica EM Moodie, Bibhas Chakraborty, and Michael S Kramer. “Q-learning for estimating optimal dynamic treatment rules from observational data”. In: *Canadian Journal of Statistics* 40.4 (2012), pp. 629–645.

[51] Kelly L Moore and Mark J van der Laan. “Covariate adjustment in randomized trials with binary outcomes: targeted maximum likelihood estimation”. In: *Statistics in medicine* 28.1 (2009), pp. 39–64.

[52] Susan A Murphy. “Optimal dynamic treatment regimes”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 65.2 (2003), pp. 331–355. ISSN: 1467-9868.

[53] Inbal Nahum-Shani et al. "Q-learning: A data analysis method for constructing adaptive interventions." In: *Psychological methods* 17.4 (2012), p. 478.

[54] Judea Pearl. "An introduction to causal inference". In: *The international journal of biostatistics* 6.2 (2010).

[55] Judea Pearl. *Causality : models, reasoning, and inference*. Cambridge, U.K. ; New York: Cambridge University Press, 2000, xvi, 384 p. ISBN: 0521773628 (hardback).

[56] Judea Pearl. *Causality: models, reasoning and inference*. Vol. 29. Springer, 2000.

[57] Maya L Petersen and Mark J van der Laan. "Causal models and learning from data: integrating causal modeling and statistical estimation". In: *Epidemiology (Cambridge, Mass.)* 25.3 (2014), p. 418.

[58] Maya L Petersen et al. "Diagnosing and responding to violations in the positivity assumption". In: *Statistical methods in medical research* 21.1 (2012), pp. 31–54.

[59] Maya L Petersen et al. "Super learner analysis of electronic adherence data improves viral prediction and may provide strategies for selective HIV RNA monitoring". In: *Journal of acquired immune deficiency syndromes (1999)* 69.1 (2015), p. 109.

[60] Romain Pirracchio, Maya L Petersen, and Mark van der Laan. "Improving propensity score estimators' robustness to model misspecification using super learner". In: *American journal of epidemiology* 181.2 (2014), pp. 108–119.

[61] Romain Pirracchio et al. "Mortality prediction in intensive care units with the Super ICU Learner Algorithm (SICULA): a population-based study". In: *The Lancet Respiratory Medicine* 3.1 (2015), pp. 42–52.

[62] Eric Polley et al. *SuperLearner: Super Learner Prediction*. R package version 2.0-24. 2018. URL: `https://CRAN.R-project.org/package=SuperLearner`.

[63] Eric C Polley and Mark J van der Laan. "Super learner in prediction". In: (2010).

[64] Min Qian and Susan A Murphy. "Performance guarantees for individualized treatment rules". In: *Annals of statistics* 39.2 (2011), p. 1180.

[65] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2018. URL: `https://www.R-project.org/`.

[66] Brian D Ripley and NL Hjort. *Pattern recognition and neural networks*. Cambridge university press, 1996.

[67] James Robins. "A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect". In: *Mathematical modelling* 7.9-12 (1986), pp. 1393–1512.

[68] James Robins, Andrea Rotnitzky, et al. "Discussion of "Dynamic treatment regimes: Technical challenges and applications"". In: *Electronic Journal of Statistics* 8.1 (2014), pp. 1273–1289.

[69] James M Robins. “Optimal structural nested models for optimal sequential decisions”. In: *Proceedings of the second seattle Symposium in Biostatistics*. Springer. 2004, pp. 189–326.

[70] James M Robins. “Robust estimation in sequentially ignorable missing data and causal inference models”. In: *Proceedings of the American Statistical Association*. Vol. 1999. Indianapolis, IN. 2000, pp. 6–10.

[71] James M Robins and Miguel A Hernán. “Estimation of the causal effects of time-varying exposures”. In: *Longitudinal data analysis* 553 (2009), p. 599.

[72] James M Robins, Andrea Rotnitzky, and Lue Ping Zhao. “Estimation of regression coefficients when some regressors are not always observed”. In: *Journal of the American statistical Association* 89.427 (1994), pp. 846–866.

[73] Paul R Rosenbaum and Donald B Rubin. “The central role of the propensity score in observational studies for causal effects”. In: *Biometrika* 70.1 (1983), pp. 41–55.

[74] William F Rosenberger, Oleksandr Sverdlov, and Feifang Hu. “Adaptive randomization for clinical trials”. In: *Journal of biopharmaceutical statistics* 22.4 (2012), pp. 719–736.

[75] Michael Rosenblum and Mark J van der Laan. “Targeted maximum likelihood estimation of the parameter of a marginal structural model”. In: *The international journal of biostatistics* 6.2 (2010).

[76] Michael Rosenblum and Mark J van der Laan. “Using regression models to analyze randomized trials: Asymptotically valid hypothesis tests despite incorrectly specified models”. In: *Biometrics* 65.3 (2009), pp. 937–945.

[77] Daniel Rubin and Mark J van der Laan. “A doubly robust censoring unbiased transformation”. In: *The international journal of biostatistics* 3.1 (2007).

[78] Daniel B Rubin and Mark J van der Laan. “Statistical issues and limitations in personalized medicine research with clinical trials”. In: *The international journal of biostatistics* 8.1 (2012).

[79] Daniel O Scharfstein, Andrea Rotnitzky, and James M Robins. “Theory and Methods-Rejoinder-Adjusting for Nonignorable Drop-Out Using Semiparametric Nonresponse Models”. In: *Journal of the American Statistical Association* 94.448 (1999), pp. 1135–1146.

[80] Phillip J Schulte et al. “Q-and A-learning methods for estimating optimal dynamic treatment regimes”. In: *Statistical science: a review journal of the Institute of Mathematical Statistics* 29.4 (2014), p. 640.

[81] Aniek Sies and Iven Van Mechelen. “Estimating the quality of optimal treatment regimes”. In: *Statistics in medicine* 38.25 (2019), pp. 4925–4938.

[82] Jennifer L Skeem, Sarah Manchak, and Jillian K Peterson. “Correctional policy for offenders with mental illness: Creating a new paradigm for recidivism reduction”. In: *Law and human behavior* 35.2 (2011), pp. 110–126.

[83] Jennifer L Skeem et al. "Offenders with mental illness have criminogenic needs, too: Toward recidivism reduction." In: *Law and human behavior* 38.3 (2014), p. 212.

[84] Anastasios A Tsiatis. *Dynamic Treatment Regimes: Statistical Methods for Precision Medicine*. CRC Press, 2019.

[85] Aad W Van der Vaart. *Asymptotic statistics*. Vol. 3. Cambridge university press, 2000.

[86] Ravi Varadhan et al. "A framework for the analysis of heterogeneity of treatment effect in patient-centered outcomes research". In: *Journal of clinical epidemiology* 66.8 (2013), pp. 818–825.

[87] Salim Yusuf et al. "Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials". In: *Jama* 266.1 (1991), pp. 93–98.

[88] Baqun Zhang et al. "Estimating optimal treatment regimes from a classification perspective". In: *Stat* 1.1 (2012), pp. 103–114.

[89] Ying-Qi Zhao and Eric B Laber. "Estimation of optimal dynamic treatment regimes". In: *Clinical Trials* 11.4 (2014), pp. 400–407.

[90] Ying-Qi Zhao et al. "Efficient augmentation and relaxation learning for individualized treatment rules using observational data." In: *Journal of Machine Learning Research* 20.48 (2019), pp. 1–23.

[91] Ying-Qi Zhao et al. "New statistical learning methods for estimating optimal dynamic treatment regimes". In: *Journal of the American Statistical Association* 110.510 (2015), pp. 583–598.

[92] Yingqi Zhao et al. "Estimating individualized treatment rules using outcome weighted learning". In: *Journal of the American Statistical Association* 107.499 (2012), pp. 1106–1118.

[93] Wenjing Zheng and Mark J van der Laan. "Asymptotic theory for cross-validated targeted maximum likelihood estimation". In: (2010).

[94] Xin Zhou et al. "Residual weighted learning for estimating individualized treatment rules". In: *Journal of the American Statistical Association* 112.517 (2017), pp. 169–187.

Appendix A

Tables of Performance of ODTR SuperLearner Simulations

n	Library		General Metalearner		Risk R	Avg. Regret	Var. Relative to GLM	% Match
1,000	Blip only	GLM	N/A		N/A	-0.0765	1.0000	56.7
		Parametric blip models	Discrete		MSE	-0.0756	2.2119	56.4
					$E[Y_d]$	-0.0721	1.8161	58.0
			Continuous	Blip-based	MSE	-0.0772	1.7008	56.1
					$E[Y_d]$	-0.0744	1.6762	57.7
				Vote-based	$E[Y_d]$	-0.0733	1.6635	57.8
		ML blip models	Discrete		MSE	-0.0393	2.4105	72.0
					$E[Y_d]$	-0.0281	0.7664	76.9
			Continuous	Blip-based	MSE	-0.0253	0.5708	77.7
					$E[Y_d]$	-0.0268	0.6024	77.0
				Vote-based	$E[Y_d]$	-0.0277	0.7006	76.7
		Parametric + ML blip models	Discrete		MSE	-0.0389	2.1327	72.0
					$E[Y_d]$	-0.0284	0.7781	76.8
			Continuous	Blip-based	MSE	-0.0261	0.6493	77.4
					$E[Y_d]$	-0.0276	0.6351	76.7
				Vote-based	$E[Y_d]$	-0.0298	0.8021	75.7
	Full	ML blip models and EYd maximizers	Discrete		$E[Y_d]$	-0.0277	0.7844	76.6
			Continuous	Vote-based	$E[Y_d]$	-0.0287	0.7463	76.8
		All blip models and EYd maximizers	Discrete		$E[Y_d]$	-0.0280	0.8083	76.4
			Continuous	Vote-based	$E[Y_d]$	-0.0290	0.7772	76.4
300	Blip only	GLM	N/A		N/A	-0.0827	1.0000	55.0
		Parametric blip models	Discrete		MSE	-0.0848	1.5846	54.0
					$E[Y_d]$	-0.0815	1.5871	55.4
			Continuous	Blip-based	MSE	-0.0850	1.5509	54.1
					$E[Y_d]$	-0.0824	1.5346	55.4
				Vote-based	$E[Y_d]$	-0.0817	1.6227	55.4
		ML blip models	Discrete		MSE	-0.0665	2.4337	60.9
					$E[Y_d]$	-0.0544	1.5571	65.0
			Continuous	Blip-based	MSE	-0.0546	1.5906	65.3
					$E[Y_d]$	-0.0522	1.3185	66.1
				Vote-based	$E[Y_d]$	-0.0554	1.3437	64.9
		Parametric + ML blip models	Discrete		MSE	-0.0649	2.4343	61.1
					$E[Y_d]$	-0.0555	1.6545	64.5
			Continuous	Blip-based	MSE	-0.0538	1.6002	65.4
					$E[Y_d]$	-0.0530	1.3822	65.7
				Vote-based	$E[Y_d]$	-0.0572	1.5197	64.1
	Full	ML blip models and EYd maximizers	Discrete		$E[Y_d]$	-0.0571	1.6907	64.4
			Continuous	Vote-based	$E[Y_d]$	-0.0547	1.5835	65.3
		All blip models and EYd maximizers	Discrete		$E[Y_d]$	-0.0558	1.6624	64.8
			Continuous	Vote-based	$E[Y_d]$	-0.0586	1.5285	63.8

Table A.1: DGP 1 ("complex blip") results: Performance metrics (average, relative variance) of the approximate regret $E_n[Q_0(Y|A = d_n^*, W)] - E_0[Y_{d_0^*}]$ (the difference between the average true conditional mean outcome under the estimated ODTR versus the true ODTR) for the SuperLearners generated by DGP 1. Percent agreement between the treatment assigned under the true versus estimated ODTR.

n	Library		General Metalearner		Risk R	Avg. Regret	Var. Relative to GLM	% Match
1,000	Blip only	GLM	N/A		N/A	-0.0098	1.0000	84.5
		Parametric blip models	Discrete		MSE	-0.0039	1.0267	90.7
					$E[Y_d]$	-0.0045	1.3174	90.2
			Continuous	Blip-based	MSE	-0.0057	1.1855	89.0
					$E[Y_d]$	-0.0085	1.5630	86.0
				Vote-based	$E[Y_d]$	-0.0055	1.3902	89.5
		ML blip models	Discrete		MSE	-0.0129	1.9508	82.3
					$E[Y_d]$	-0.0121	1.7321	83.0
			Continuous	Blip-based	MSE	-0.0106	1.3294	84.2
					$E[Y_d]$	-0.0099	1.1477	84.9
				Vote-based	$E[Y_d]$	-0.0102	1.1776	84.5
		Parametric + ML blip models	Discrete		MSE	-0.0063	1.1875	88.8
					$E[Y_d]$	-0.0087	1.4343	86.7
			Continuous	Blip-based	MSE	-0.0077	1.2204	86.8
					$E[Y_d]$	-0.0091	1.3069	85.4
				Vote-based	$E[Y_d]$	-0.0094	1.4221	85.2
	Full	ML blip models and EYd maximizers	Discrete		$E[Y_d]$	-0.0114	1.6804	84.1
			Continuous	Vote-based	$E[Y_d]$	-0.0107	1.3875	84.4
		All blip models and EYd maximizers	Discrete		$E[Y_d]$	-0.0074	1.4027	88.1
			Continuous	Vote-based	$E[Y_d]$	-0.0082	1.2160	86.7
300	Blip only	GLM	N/A		N/A	-0.0222	1.0000	75.0
		Parametric blip models	Discrete		MSE	-0.0188	1.8019	78.0
					$E[Y_d]$	-0.0232	2.0950	74.5
			Continuous	Blip-based	MSE	-0.0186	1.6102	76.7
					$E[Y_d]$	-0.0229	1.8938	73.0
				Vote-based	$E[Y_d]$	-0.0215	2.1003	74.6
		ML blip models	Discrete		MSE	-0.0369	1.6533	61.7
					$E[Y_d]$	-0.0309	1.4898	66.8
			Continuous	Blip-based	MSE	-0.0317	1.4242	66.4
					$E[Y_d]$	-0.0284	1.2123	69.2
				Vote-based	$E[Y_d]$	-0.0289	1.2973	68.7
		Parametric + ML blip models	Discrete		MSE	-0.0205	1.8830	75.5
					$E[Y_d]$	-0.0252	1.7460	71.7
			Continuous	Blip-based	MSE	-0.0234	1.5428	73.4
					$E[Y_d]$	-0.0257	1.3637	71.5
				Vote-based	$E[Y_d]$	-0.0260	1.4567	71.3
	Full	ML blip models and EYd maximizers	Discrete		$E[Y_d]$	-0.0300	1.5305	67.3
			Continuous	Vote-based	$E[Y_d]$	-0.0284	1.4512	68.7
		All blip models and EYd maximizers	Discrete		$E[Y_d]$	-0.0218	1.7749	74.6
			Continuous	Vote-based	$E[Y_d]$	-0.0226	1.5302	73.5

Table A.2: DGP 2 ("simple blip") results: Performance metrics (average, relative variance) of the approximate regret $E_n[Q_0(Y|A = d_n^*, W)] - E_0[Y_{d_0^*}]$ (the difference between the average true conditional mean outcome under the estimated ODTR versus the true ODTR) for the SuperLearners generated by DGP 2. Percent agreement between the treatment assigned under the true versus estimated ODTR.