## Title

Estimators for the value of the optimal dynamic treatment rule with application to criminal justice interventions

Permalink
https://escholarship.org/uc/item/689925q0

Journal
The International Journal of Biostatistics, 19(1)

ISSN
2194-573X

Authors
Montoya, Lina M
van der Laan, Mark J
Skeem, Jennifer L
et al.

Publication Date
2023-05-01

DOI
10.1515/ijb-2020-0128

Peer reviewed

Lina M. Montoya*, Mark J. van der Laan, Jennifer L. Skeem and Maya L. Petersen

# Estimators for the value of the optimal dynamic treatment rule with application to criminal justice interventions

https://doi.org/10.1515/ijb-2020-0128

Received September 4, 2020; accepted May 6, 2022; published online June 6, 2022

**Abstract:** 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? In this paper, we study the performance of estimators that approximate the true value of: (1) an *a priori* known dynamic treatment rule (2) the true, unknown optimal dynamic treatment rule (ODTR); (3) an estimated ODTR, a so-called "data-adaptive parameter," whose true value depends on the sample. Using simulations of point-treatment data, we specifically investigate: (1) the impact of increasingly data-adaptive estimation of nuisance parameters and/or of the ODTR on performance; (2) the potential for improved efficiency and bias reduction through the use of semiparametric efficient estimators; and, (3) the importance of sample splitting based on the cross-validated targeted maximum likelihood estimator (CV-TMLE) for accurate inference. In the simulations considered, there was very little cost and many benefits to using CV-TMLE to estimate the value of the true and estimated ODTR; importantly, and in contrast to non cross-validated estimators, the performance of CV-TMLE was maintained even when highly data-adaptive algorithms were used to estimate both nuisance parameters and the ODTR. In addition, we apply these estimators for the value of the rule 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.

**Keywords:** causal roadmap; heterogeneous treatment effects; optimal dynamic treatment rule; precision health; value of rule.

## 1 Introduction

There is an interest across disciplines in using both experimental and observational data to uncover treatment effect heterogeneity and quantify the benefits of responding to this heterogeneity when assigning treatments (for example, [1, 2]). Various methods aimed at estimating heterogenous treatment effects (HTEs) aim 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 [3–7]. An optimal dynamic treatment rule (ODTR) is the dynamic treatment rule that yields the highest expected outcome (if higher outcomes are better) [8–10]. In recent years, there has been an increase in literature describing methods to estimate the ODTR, from regression-based techniques to direct-search techniques; see, for example, [11, 12] and [13] for recent overviews of the ODTR literature. One example of a data-adaptive

***Corresponding author: Lina Montoya**, Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7400, USA, E-mail: lmontoya@unc.edu. https://orcid.org/0000-0002-0975-4306
**Mark J. van der Laan**, Division of Biostatistics, University of California Berkeley, Berkeley, USA
**Jennifer L. Skeem**, School of Social Work and Goldman School of Public Policy, University of California Berkeley, Berkeley, USA
**Maya L. Petersen**, Divisions of Biostatistics and Epidemiology, University of California Berkeley, Berkeley, USA

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 [14–16]. In a companion paper, we review this approach, and highlight and investigate the consequences of key choices when implementing this method [17].

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 – a particularly challenging target parameter to estimate; and (3) 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 [6], the inverse probability of treatment weighted (IPTW) estimator [18, 19], the double-robust IPTW estimator (IPTW-DR) [20–22], the targeted maximum likelihood estimator (TMLE) [3, 23–25], and the cross-validated TMLE (CV-TMLE) [25–27].

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. Obtaining inference for the mean outcome under the ODTR has been shown to be difficult due to its lack of smoothness [4, 10, 28]; however, several methods have been proposed for constructing valid confidence intervals for this parameter, such as re-sampling techniques [4, 29, 30]. One approach to inference is to rely on algorithms based 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) [25]. 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 [25].

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 [25, 31]. This relaxes the need for the estimate of the ODTR to converge to the true rule at a fast enough rate. Non-cross-validated estimators of this data-adaptive parameter still require smoothness assumptions on the estimate of the ODTR (and nuisance parameters) for asymptotic linearity. The use of CV-TMLE eliminates these requirements. This means that, at the cost of targeting a distinct, sample-split-specific target parameter, in a randomized experiment, achievement of asymptotic linearity for

CV-TMLE with respect to the sample-split-specific data-adaptive parameter only requires that the estimated ODTR converges to a fixed rule [25].

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. van der Laan and Luedtke [25] 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. Coyle [14] 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 [16]. Both [25] and [14] 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, [30] 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.

We apply these estimators of the value of the rule to the Correctional Intervention for People with Mental Illness, or "Interventions" trial, an ongoing study in which criminal justice-involved adults with mental illness – a heterogeneous group with diverse symptoms, risk factors, and other treatment-relevant characteristics [32, 33] – are either randomized to cognitive behavioral therapy (CBT) or treatment as usual (TAU). Re-arrest, the outcome, is collected one year after randomization occurs, as a measure of recidivism. In our companion paper, we estimated the ODTR using the ODTR SuperLearner algorithm [17] to identify which patients should receive CBT versus TAU. In the current 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 roadmap for answering causal questions [34], and is organized as follows. In the first section, we define the data and causal model, define the causal parameters as functions of counterfactual distributions contained in the causal model, and identify the statistical 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. In the Appendix, we provide a Notation Table with terms frequently used throughout this manuscript and our companion manuscript, in addition to definitions of each of the terms, as a reference to reader.

## 2 Causal roadmap

In this section, we follow the first steps of the roadmap 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 sample), or (b) a sample-split-specific estimate of the ODTR?

## 2.1 Data and models

Structural causal models (SCM, denoted $\mathscr{M}^F$) 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 follow the joint distribution $P_{U,X}$; the SCM describes the set of possible distributions for $P_{U,X}$. The endogenous variables are the covariates $W \in \mathscr{W}$, binary treatment $A \in \mathscr{A} = \{0,1\}$, 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 $U_A \sim \text{Bernoulli}(p = 0.5)$ and $A = U_A$.

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

The density of $O$ can be factorized as $p_0(O) = p_{W,0}(W)g_0(A|W)p_{Y,0}(Y|A, W)$, 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 $\frac{1}{n}$. 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 which we can index with $v = 1, \ldots, V$. For each sample split $v = 1, \ldots, V$, this $v$th data set represents the validation set while 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.

### 2.1.1 Data and models – application to "Interventions" study

The "Interventions" study is a randomized controlled trial (RCT) consisting of 441 i.i.d. observations of the following data generated by a process described 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 1 shows the distribution of the data.

## 2.2 Causal estimands

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

A counterfactual outcome under an arbitrary treatment rule $d$ – an individual's outcome if, possibly contrary to fact, the individual received the treatment that would have been assigned by the treatment rule $d$ (denoted $Y_d$) – is derived under an intervention on the above SCM. Specifically, we consider counterfactual

**Table 1:** Distribution of baseline covariates in the "Interventions" data set, stratified by randomized treatment assignment (TAU denotes Treatment as Usual, CBT denotes Cognitive Behavioral Therapy).

| | **TAU ($A = 0$)** | **CBT ($A = 1$)** |
|---|---|---|
| $N$ | 211 | 230 |
| **No re-arrest** ($Y = 1$) (%) | 128 (60.7) | 143 (62.2) |
| **Site** = San Francisco (%) | 87 (41.2) | 104 (45.2) |
| **Gender** = Female (%) | 38 (18.0) | 37 (16.1) |
| **Ethnicity** = Hispanic (%) | 50 (23.7) | 42 (18.3) |
| **Age** (mean (SD)) | 38.08 (11.05) | 37.01 (11.22) |
| **CSI** (mean (SD)) | 32.35 (11.13) | 33.46 (11.27) |
| **LSI** (mean (SD)) | 5.59 (1.33) | 5.50 (1.48) |
| **SES** (mean (SD)) | 3.81 (1.89) | 3.81 (2.12) |
| **Prior adult convictions** (%) | | |
| 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) |
| **Motivation** (mean (SD)) | 3.22 (1.36) | 3.27 (1.37) |
| **Substance use** (%) | | |
| 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) |

outcomes generated by setting $A$ equal to the following treatment rules: (1) the true ODTR; and, (2) an estimate of the ODTR, either: (a) the sample-specific estimate of the ODTR; or (b) the training sample-specific estimate of the ODTR.

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 $\mathcal{M}^F \to \mathbb{R}$.

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 $d_0^*$:

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

Let $d_n^*\colon \mathcal{M} \to \mathcal{D}$ be an ODTR estimated on the entire sample, and $d_{n,v}^* = d^*(P_{n,-v})\colon \mathcal{M} \to \mathcal{D}$ 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:

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

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

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

where, the true value of this target parameter further depends on the random sample split (noting, again, that the expectation is not over $d^*_{n,v}$).

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 $\mathbb{E}_{P_{U,X}}[Y_0]$.

### 2.2.1 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 prespecified rule $d$ (for example, the simple dynamic treatment rule that gives CBT to those with a high baseline risk score of re-offending and TAU to those with a low baseline risk score of re-offending), where the causal parameter is $\Psi^F_d(P_{U,X})$; (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 $\Psi^F_{d^*_0}(P_{U,X})$; 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 [17]) or (b) a sample-split-specific estimate of the ODTR $d^*_{n,v}$? The causal parameters for (a) and (b) are $\Psi^F_{d^*_n}(P_{U,X})$ and $\Psi^F_{d^*_{n,v}}(P_{U,X})$, respectively.

## 2.3 Identification

Two assumptions are necessary for identification; that is, for determining that the causal estimands (a function of our counterfactual distribution) coincide with 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) positivity assumption: $P(\min_{a\in\{0,1\}} g_0(a|W) > 0) = 1$. 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 [35].

## 2.4 Statistical estimands

We describe statistical estimands corresponding to each of the causal parameters outlined above – each is identified via the G-computation formula.

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 individual (e.g., if certain kinds of individuals neither benefit from nor are harmed by a treatment) [36]. 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", where the true blip function under $P_0$ is defined as $B_0(W) = Q_0(1, W) - Q_0(0, W)$:

$$d^*_0(W) = \mathbb{I}[B_0(W) > 0].$$

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

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

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

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

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

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

# 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 $\psi_{0,d_0^*}$. We use the non-cross-validated estimators (G-computation, IPTW, IPTW-DR, and TMLE) to estimate $\psi_{0,d_n^*}$; we estimate $\psi_{0,d_{n,v}^*}$ with CV-TMLE.

Let $Q_n$ be an estimator of the outcome regression, which could be estimated with, for example, SuperLearner [16]. 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.

## 3.1 Non-cross-validated estimators for estimating $\psi_{0,d}$, $\psi_{0,d_0^*}$, and $\psi_{0,d_n^*}$

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 (see [15, 17] for a description on how to estimate the optimal rule using, for example, the ODTR SuperLearner). Here, $\hat{\Psi}_d(P_n) \equiv \hat{\psi}_d$ is an estimate of the true parameter value $\psi_{0,d}$, based on applying the estimator $\hat{\Psi}_d$ to an empirical distribution based on sampling from $P_0$. We further subscript by each estimator name.

One can use a(n):

- Simple substitution estimator based on the above G-computation formula,

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

- IPTW estimator,

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

- Double-robust IPTW estimator,

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

- or TMLE. We briefly describe one possible TMLE procedure. First, estimate the $i$-specific so-called clever covariate:

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

Then, update the initial fit of $Q_n$ by running a logistic regression of $Y$ (which should be transformed between 0 and 1 if the outcome is continuous [37]) using the logit of $Q_n(d(W), W)$ as offset and weights $H_n$, with maximum likelihood estimation used to estimate the intercept. Denote the predictions from this logistic regression as $Q_n^*(d(W), W)$, from the updated fit. Then, the TMLE estimator is:

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

## 3.2 CV-TMLE for estimating $\psi_{0,d}$, $\psi_{0,d_0^*}$, and $\psi_{0,d_{n,v}^*}$

As previously mentioned, the CV-TMLE can estimate $\psi_{0,d}$, $\psi_{0,d_0^*}$, and $\psi_{0,d_{n,v}^*}$. Instead of illustrating the CV-TMLE 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 $\psi_{0,d_0^*}$ or $\psi_{0,d_{n,v}^*}$. 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, providing us with $v$-specific sample splits in the validation and training set.
2. For $v = 1, \dots, V$, carry out the following steps:
   (a) Estimate the treatment mechanism, ODTR, and outcome regression on the training set.
   (b) Using the fits from the previous step, generate predictions of the observed treatment assignment, optimal treatment assignment, and outcome under the optimal treatment assignment for observations in the validation set. Denote the corresponding estimates as $g_{n,v}(A|W)$, $d_{n,v}^*(W)$, and $Q_{n,v}(d_{n,v}^*(W), W)$.
   (c) Update $Q_{n,v}$ generated in the previous step by fitting an intercept model (as described in the TMLE updating procedure in the previous subsection) on persons in the validation set. Call the updated fit $Q_{n,v}^*$.
   (d) Generate validation set-specific targeted estimates of the mean outcome under the sample-split-specific estimated rule $d_{n,v}^*$ by evaluating $Q_{n,v}^*$ on data in the validation set. Call the updated estimates $Q_{n,v}^*(d_{n,v}^*(W), W)$.
   (e) Define the $v$th validation set-specific estimate of the mean outcome under the estimate rule as:

   $$\hat{\psi}_{d_{n,v}^*} = \frac{1}{n_v}\sum_{i\in Val(v)} Q_{n,v}^*(d_{n,v}^*(W_i), W_i),$$

   where $n_v$ denotes the number of individuals in the validation set $v$ and $Val(v)$ is the indices $i$ for which $O_i$ is in the validation set.
3. Average over all validation folds to obtain the CV-TMLE, i.e., the estimated mean outcome under the sample-split-specific estimates of ODTR:

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

# 4 Inference

We first discuss the conditions necessary for each the above estimators to be asymptotically linear for $\psi_{0,d}$, $\psi_{0,d_n^*}$, and $\psi_{0,d_{n,v}^*}$ 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.

For more details and proofs, we refer the reader to [24, 25, 38].

## 4.1 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 an informal summary of these conditions in Table 2.

An estimator $\hat{\Psi}$ is asymptotically linear for its true value $\psi_0$ if it can be written in the following form:

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

where $\hat{\psi}$ is the estimate of $\psi_0$, IC is the estimator's influence curve (that is centered to have mean 0, by definition) and $R_n$ is a remainder term that is $o_P(1/\sqrt{n})$. An asymptotically linear estimator $\hat{\Psi}$ thus generally 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 approximately normal, $n^{1/2}(\hat{\psi} - \psi_0) \xrightarrow{d} N\left(0, \sigma_0^2\right)$, allowing an estimate of $\sigma_0^2$ 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., $\sigma_0^2$) can be well-approximated by the sample variance of its estimated influence curve $\mathrm{IC}_n$ (or equivalently, $\sigma_n^2 = \frac{1}{n}\sum_{i=1}^{n}\mathrm{IC}_n^2(O_i)$, since the mean of an influence curve is 0).

### 4.1.1 Conditions for value of a known rule

Our randomized experiment scenario guarantees that $g_0$ is known, and thus $g_n$ can be a maximum likelihood estimate (MLE) of $g_0$ based on a correctly specified parametric model. As a result, 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, this estimator will not be asymptotically efficient.

Let $Pt = \mathbb{E}_P[t(O)]$ for a distribution $P$ and function $t$, and let $\mathrm{IC}^*$ be the efficient influence curve. If $P_0\left\{\mathrm{IC}^*(Q_n, g_n) - \mathrm{IC}^*(Q, g_0)\right\}^2$ converges to zero in probability for a limit $Q$, possibly misspecified, and $\left\{\mathrm{IC}^*(Q, g_0)\colon Q\right\}$ is a $P_0$ Donsker class, such as the class of $d$ variate cadlag functions with a universal bound on the sectional variation norm, then the TMLE and IPTW-DR are asymptotically linear with an influence curve equal to $\mathrm{IC}^*(Q, g_0)$ minus its projection onto the tangent space of the parametric model used for the MLE $g_n$. As a consequence, its asymptotic variance is smaller than or equal to the variance of $\mathrm{IC}^*(Q, g_0)$, and, in particular, if $g_n$ is replaced by $g_0$, then $\mathrm{IC}^*(Q, g_0)$ is the actual influence curve. Further, if $Q_n$ is consistent for $Q_0$ in the sense that $P_0\{\mathrm{IC}(Q_n, g_0) - \mathrm{IC}(Q_0, g_0)\}^2$ converges to zero in probability, then the TMLE and IPTW-DR estimators are also asymptotically efficient.

The above is also true for CV-TMLE, except Donsker class conditions can now be removed (in effect allowing for an overfit in the initial estimate of $Q_0$).

### 4.1.2 Conditions for value of true ODTR

Construction of nominal to conservative confidence intervals around each of the non-cross-validated estimators with respect to the true expected outcome under the true, unknown $d_0^*$ requires additional assumptions.

**Table 2:** Informal summary of the conditions needed for asymptotic linearity in the randomized treatment setting for each of the estimators corresponding to each of the estimands.

| | | **Conditions for Asymptotic Linearity:** | | | |
|---|---|---|---|---|---|
| **Estimands** | **Estimators** | $g_n = g_0$ **or** $g_n$ **is MLE of** $g_0$ **according to correctly specified parametric model** | **Fast enough convergence of** $d_n^*$ **to** $d_0^*$ | $Q_n$ **not overfit** | $d_n$ **not overfit** |
| Value of known rule $\psi_{0,d}$ | $\hat{\Psi}_{\mathrm{IPTW,d}}$ | | Not required, $d$ known | Not required | Not required, $d$ known |
| | $\hat{\Psi}_{\mathrm{IPTW-DR,d}}$ | | | Required | |
| | $\hat{\Psi}_{\mathrm{TMLE,d}}$ | | | Required | |
| | $\hat{\Psi}_{\mathrm{CV-TMLE},d}$ | | | Not required | |
| Value of true ODTR $\psi_{0,d_0^*}$ | $\hat{\Psi}_{\mathrm{IPTW},d_n^*}$ | Satisfied by randomized experiment | Required | Not required | Required |
| | $\hat{\Psi}_{\mathrm{IPTW-DR},d_n^*}$ | | Required | | |
| | $\hat{\Psi}_{\mathrm{TMLE},d_n^*}$ | | Required | | |
| | $\hat{\Psi}_{\mathrm{CV-TMLE},d_{n,v}^*}$ | | Required | Not required | Not required |
| Value of sample- specific ODTR estimate $\psi_{0,d_n^*}$ | $\hat{\Psi}_{\mathrm{IPTW},d_n^*}$ | | Not required; require $d_n^* \xrightarrow{p} d \in \mathscr{D}$ | Not required | Required |
| | $\hat{\Psi}_{\mathrm{IPTW-DR},d_n^*}$ | | | Required | |
| | $\hat{\Psi}_{\mathrm{TMLE},d_n^*}$ | | | Required | |
| Value of sample-split-specific ODTR estimate $\psi_{0,d_{n,v}^*}$ | $\hat{\Psi}_{\mathrm{CV-TMLE},d_{n,v}^*}$ | | Not required; require $d_{n,v}^* \xrightarrow{p} d \in \mathscr{D}$ | Not required | Not required |

**Table 3:** Performance metrics (bias, variance, MSE, confidence interval coverage) of each estimator $\hat{\psi}_{d=d_0^*}$ (G-computation, Inverse Probability of Treatment Weighting [IPTW], Double-robust IPTW [IPTW-DR], Targeted Maximum Likelihood Estimation [TMLE], Cross-Validated TMLE [CV-TMLE]) of the true expected outcome under a given, known dynamic treatment rule $\left(\psi_{0,d_0^*}\right)$, for each library configuration used to estimate the outcome regression ($Q_n$).

| Library | Estimator | Bias | Variance | MSE | Coverage |
|---|---|---:|---:|---:|---:|
| GLMs − least data adaptive | G-comp. | −0.0935 | 0.0002 | 0.0090 | – |
| | IPTW | −0.0004 | 0.0008 | 0.0008 | 95.80% |
| | IPTW-DR | 0.0002 | 0.0004 | 0.0004 | 95.80% |
| | TMLE | 0.0004 | 0.0004 | 0.0004 | 95.80% |
| | CV-TMLE | 0.0007 | 0.0005 | 0.0005 | 95.30% |
| ML + GLMs − moderately data adaptive | G-comp. | −0.1313 | 0.0006 | 0.0179 | – |
| | IPTW | −0.0009 | 0.0007 | 0.0007 | 96.30% |
| | IPTW-DR | −0.0009 | 0.0005 | 0.0005 | 95.00% |
| | TMLE | −0.0008 | 0.0005 | 0.0005 | 94.80% |
| | CV-TMLE | −0.0004 | 0.0005 | 0.0005 | 94.90% |
| ML + GLMs − most data adaptive | G-comp. | −0.1165 | 0.0006 | 0.0142 | – |
| | IPTW | 0.0007 | 0.0008 | 0.0008 | 95.20% |
| | IPTW-DR | −0.0074 | 0.0005 | 0.0006 | 90.50% |
| | TMLE | −0.0067 | 0.0005 | 0.0006 | 91.00% |
| | CV-TMLE | 0.0007 | 0.0005 | 0.0005 | 94.70% |

**Table 4:** Performance metrics (bias, variance, MSE, confidence interval coverage) of each estimator $\hat{\psi}_{d_n^*}$ (G-computation, Inverse Probability of Treatment Weighting [IPTW], Double-robust IPTW [IPTW-DR], Targeted Maximum Likelihood Estimation [TMLE]) or $\hat{\psi}_{d_{n,v}^*}$ (Cross-validated-TMLE [CV-TMLE]) of the true expected outcome under the true optimal dynamic treatment rule (ODTR; $\psi_{0,d_0^*}$), for each library configuration used to estimate the outcome regression ($Q_n$) and the ODTR $\left(d_n^*\right)$.

| Library | Estimator | Bias | Variance | MSE | Coverage |
|---|---|---:|---:|---:|---:|
| GLMs − least data adaptive | G-comp. | −0.0765 | 0.0003 | 0.0062 | – |
| | IPTW | −0.0569 | 0.0008 | 0.0041 | 45.70% |
| | IPTW-DR | −0.0565 | 0.0007 | 0.0038 | 29.80% |
| | TMLE | −0.0563 | 0.0007 | 0.0038 | 29.40% |
| | CV-TMLE | −0.0752 | 0.0009 | 0.0066 | 14.00% |
| ML + GLMs − moderately data adaptive | G-comp. | −0.1325 | 0.0007 | 0.0182 | – |
| | IPTW | 0.0300 | 0.0010 | 0.0019 | 79.50% |
| | IPTW-DR | 0.0295 | 0.0008 | 0.0016 | 69.90% |
| | TMLE | 0.0268 | 0.0007 | 0.0014 | 72.20% |
| | CV-TMLE | −0.0310 | 0.0007 | 0.0017 | 70.30% |
| ML + GLMs − most data adaptive | G-comp. | −0.1152 | 0.0007 | 0.0140 | – |
| | IPTW | 0.1131 | 0.0114 | 0.0242 | 38.30% |
| | IPTW-DR | 0.0925 | 0.0096 | 0.0181 | 38.90% |
| | TMLE | 0.0963 | 0.0118 | 0.0211 | 38.50% |
| | CV-TMLE | −0.0281 | 0.0007 | 0.0015 | 71.90% |

For these estimators, statistical inference for $\psi_{0,d_0^*}$ relies on a second-order difference in $R_n$ between $\psi_{0,d_n^*}$ and $\psi_{0,d_0^*}$ 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 as a random variable (function of $W$) 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 values that have positive probability of occurring, then stronger assumptions are required [15, 25]. The non-cross-validated estimators additionally require Donsker class conditions on $\mathrm{IC}_{d_n^*}(Q_n, g_n)$, thereby also restricting the adaptivity of $d_n^*$ (informally, that $d_n^*$ not be an overfit of $d_0^*$). In practice, these conditions on the data-adaptivity of $d_n^*$ hold if, for example, the optimal rule is a function

**Table 5:** Performance metrics (bias, variance, MSE, confidence interval coverage) of each estimator $\hat{\psi}_{d_n^*}$ (G-computation, Inverse Probability of Treatment Weighting [IPTW], Double-robust IPTW [IPTW-DR], Targeted Maximum Likelihood Estimation [TMLE]) of the true expected outcome under the sample-specific estimate of the optimal dynamic treatment rule (ODTR; $\psi_{0,d_n^*}$) or $\hat{\psi}_{d_{n,v}^*}$ (Cross-validated-TMLE [CV-TMLE]) of the true expected outcome under the sample-split-specific estimate of the ODTR $\left(\psi_{0,d_{n,v}^*}\right)$, for each library configuration used to estimate the outcome regression ($Q_n$) and the ODTR $\left(d_n^*\right)$.

| Library | Estimator | Bias | Variance | MSE | Coverage |
|---|---|---|---|---|---|
| GLMs − least data adaptive | G-comp. | −0.0035 | 0.0003 | 0.0004 | – |
| | IPTW | 0.0162 | 0.0008 | 0.0011 | 94.90% |
| | IPTW-DR | 0.0166 | 0.0007 | 0.0009 | 90.60% |
| | TMLE | 0.0167 | 0.0007 | 0.0009 | 90.50% |
| | CV-TMLE | 0.0002 | 0.0009 | 0.0009 | 93.90% |
| ML + GLMs − moderately data adaptive | G-comp. | −0.1048 | 0.0007 | 0.0117 | – |
| | IPTW | 0.0577 | 0.001 | 0.0043 | 48.00% |
| | IPTW-DR | 0.0572 | 0.0008 | 0.0041 | 33.00% |
| | TMLE | 0.0545 | 0.0007 | 0.0037 | 33.90% |
| | CV-TMLE | −0.0008 | 0.0007 | 0.0007 | 93.90% |
| ML + GLMs − most data adaptive | G-comp. | −0.0842 | 0.0007 | 0.0078 | – |
| | IPTW | 0.1442 | 0.0114 | 0.0322 | 25.00% |
| | IPTW-DR | 0.1236 | 0.0096 | 0.0248 | 22.60% |
| | TMLE | 0.1274 | 0.0118 | 0.0280 | 22.50% |
| | CV-TMLE | 0.0013 | 0.0007 | 0.0007 | 93.90% |

of one covariate, or, if a higher-dimensional covariate set is used, one is willing to make strong smoothness assumptions, for example, on the blip function.

CV-TMLE also relaxes these additional Donsker conditions on $d_n^*$. Thus, in a randomized trial, if employing CV-TMLE for $\psi_{0,d_0^*}$, the only condition needed is that $\psi_{0,d_{n,v}^*} - \psi_{0,d_0^*} = o_P(n^{-1/2})$. This condition is carefully addressed in [15] and can be expressed in terms of a condition on $P(|B_0(W)| < x)$ to converge to zero at a fast enough rate in $x$ as $x$ approaches 0 (a so-called margin assumption).

### 4.1.3 Conditions for value of sample-(split)-specific ODTR estimate

For the data-adaptive parameters, the asymptotic study of the non-cross-validated estimators no longer requires the strong assumption that $d_n^*$ converges to $d_0^*$ at a fast enough rate; rather, they only require that $d_n^*$ converges to some fixed rule $d \in \mathcal{D}$ at any rate.

Similarly, CV-TMLE only requires the weak consistency condition that $d_{n,v}^*$ converges to some fixed rule $d \in \mathcal{D}$ at any rate. This means that, for randomized trial data, and under the above analogue $L^2(P_0)$ consistency conditions, the CV-TMLE estimator for the data adaptive parameter $\psi_{0,d_{n,v}^*}$ is asymptotically linear under essentially no conditions.

## 4.2 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. 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:

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

The influence curve of the TMLE and double-robust IPTW estimator is the *efficient* influence curve for the treatment-specific mean (39, 40); the corresponding working influence curve estimates are:

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

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

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: $\sigma_n^2 = \frac{1}{n}\sum_i^n \widehat{\mathrm{IC}}^2(O_i)$.

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. Then the fold-specific estimate of the working influence curve for CV-TMLE is based on estimating $d_0^*$, $Q_0$, and $g_0$ on the $v$th training sample, evaluated on the complementary validation sample:

$$\widehat{\mathrm{IC}}_{v,d^*_{n,v}}(O) = \frac{\mathbb{I}\left[A = d^*_{n,v}(W)\right]}{g_{n,v}(A|W)}(Y - Q^*_{n,v}(A, W)) + Q^*_{n,v}(d^*_{n,v}(W), W) - \hat{\psi}_{d^*_{n,v}},$$

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

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

where, as before, $n_v$ denotes the number of individuals in validation set $v$; thus, the asymptotic variance of the CV-TMLE $\hat{\psi}_{\mathrm{CV-TMLE},d^*_{n,v}}$ can be conservatively estimated with:

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

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

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

## 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 (where 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).

## 5.1 Data generating process

All simulations were implemented in R [41], and the code, simulated data, and results can be found at https://github.com/lmmontoya/SL.ODTR. In the future, we plan to integrate the SL.ODTR software to the Targeted Learning software ecosystem, tlverse [42]. We examine these comparisons using the following data generating process (DGPs) (also used in [15, 17, 25]). Each simulation consists of 1000 iterations of $n = 1000$ observations. Mimicking a randomized experiment, the (independent) covariates, treatment and outcome are generated as follows:

$$\begin{aligned}
W_1, W_2, W_3, W_4 \sim & \text{Normal}(\mu = 0, \sigma^2 = 1),\\
A \sim & \text{Bernoulli}(p = 0.5),\\
Y \sim & \text{Bernoulli}(p = Q_0(A, W)),\\
\text{where } Q_0(A, W) = & 0.5 \,\text{expit}\left(1 - W_1^2 + 3W_2 + 5W_3^2 A - 4.45A\right)\\
& + 0.5 \,\text{expit}(-0.5 - W_3 + 2W_1W_2 + 3|W_2|A - 1.5A),
\end{aligned}$$

then the true blip function is:

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

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 $\mathbb{E}_{P_{U,X}}\left[d_0^*\right] \approx 55.0\%$. The mean outcome had everyone and no one been treated are, respectively, $\mathbb{E}_{P_{U,X}}[Y_1] \approx 0.4638$ and $\mathbb{E}_{P_{U,X}}[Y_0] \approx 0.4643$.

In the Appendix, we illustrate results on the same simulation procedure, but with dependent covariates.

## 5.2 Estimator configurations

We estimate each of the statistical estimands using the IPTW, IPTW-DR, TMLE, and CV-TMLE estimators, with inference based on the conservative working influence curves describe above. 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 $g_0$, reflecting the RCT setting. SuperLearner is used to estimate $Q_0$ and $d_0^*$. 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 using the value of the candidate rule as the risk function [17]. Three libraries are considered that correspond to varying levels of data-adaptiveness, or potential for overfitting.

1. "GLMs – least data adaptive"
   - $Q$ library: four logistic regressions, each with a main terms $W_j$ and $A$, and with an interaction $W_j$ times $A$, for $j \in \{1, \dots, 4\}$
   - $d^*$ library: univariate linear regressions with each covariate
2. "ML + GLMs – moderately data adaptive"
   - $Q$ and $d^*$ library: all algorithms in the "GLMs – least data adaptive" $Q$ and $d^*$ 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 [43]), `SL.nnet` (neural networks [44]), `SL.svm` (support vector machines [45]), and `SL.rpart` (recursive partitioning and regression trees [46]) from the SuperLearner package [47]
3. "ML + GLMs – most data adaptive"
   - $Q$ and $d^*$ library: all algorithms in the "ML + GLMs – moderately data adaptive" $Q$ and $d^*$ libraries, respectively, in addition to `SL.randomForest` [48]

## 5.3 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 $d$: $\psi_{0,d}$; (2) the true expected outcome under the true, unknown ODTR: $\psi_{0,d_0^*}$; (3) the true expected outcome under an ODTR: (a) estimated on the entire sample and evaluated on the entire sample: $\psi_{0,d_n^*}$; or (b) estimated on each of the training sets, evaluated on the corresponding validation sets, and averaged: $\psi_{0,d_{n,v}^*}$.

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, $d = d_0^*$, as if the true optimal rule were known. 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 $\psi_{0,d_0^*}$ 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 the presented estimands, showing, for example, the price in performance one pays for targeting the more ambitious parameter that seeks to estimate both the optimal rule itself and its true value. Said another way, if we see that estimator performance for $\hat{\psi}_{d=d_0^*}$ with respect to $\psi_{0,d_0^*}$ is good, then the only issue left with estimating $\psi_{0,d_0^*}$ is estimating $d_0^*$ well.

Next, we estimate the same target parameter $\psi_{0,d_0^*}$ 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 $\psi_{0,d_0^*}$ 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 $\psi_{0,d_n^*}$, or, for CV-TMLE, $\psi_{0,d_{n,v}^*}$. 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 with respect to the true sample-specific or sample-split-specific mean outcome. For example, for confidence interval coverage, across the 1000 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 $\psi_{0,d_n^*}$) and CV-TMLE (whose data-adaptive parameter is the sample-split-specific parameter $\psi_{0,d_{n,v}^*}$), and as such, is not only a function of the sample, but also of the split.

## 5.4 Simulation results

Simulation results are displayed in Figure 1 and in Table 3, 4 and 5. In the next section, we describe the performance of each estimator under each library configuration for the estimands of interest.

### 5.4.1 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. We used $V = 10$ folds when generating cross-validated estimates. See Table 3 and Figure 1 for detailed results.

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 −9.35% (i.e., 133.57−467.50 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, for example, $\mathbb{E}_0[Q_0(1, W)]$) from data generated from a randomized experiment. In the latter, the G-computation estimator under certain misspecified parametric models is a TMLE, and is therefore unbiased [49].

As expected, the IPTW estimator, although unbiased, was less efficient than alternative estimators – specifically, the IPTW estimator's variance was 1.40−2.00 times that of the variance of double-robust 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 regression based on a misspecified parametric model or a SuperLearner with a less data-adaptive library. However, both estimators were biased (i.e., −0.74% and −0.67% bias for IPTW-DR and TMLE, respectively) with less than nominal confidence interval coverage (i.e., 90.5% and 91.0% 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 (−0.07%, or 0.006−0.100 times the bias relative to alternative double-robust estimators) and generated better confidence interval coverage (94.7%) under the presence of overfitting for $Q_n$, at no cost to variance.

### 5.4.2 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 $\psi_{0,d_0^*}$); see Table 4 and Figure 1 for more results. 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. As a result, 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 14.00%–45.70%, 69.90%–79.50%, and 38.30%–71.90%, respectively).

Although the focus of these simulations was not optimizing estimation of the ODTR, we note that, consistent with results from [17], the least biased estimators of the true value of the true ODTR are ones that use a combination of regressions based on parametric models and machine learning algorithms in the estimation of $Q_0$ and $d_0$.

### 5.4.3 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., $\psi_{0,d_n^*}$) – a parameter that depends on the optimal rule specific to the sample at hand (see Table 5 and Figure 1 for detailed results). 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 12.74%, variance of 0.0118, MSE of 0.0280, and 22.50% confidence interval coverage.

The CV-TMLE (i.e., $\hat{\psi}_{\mathrm{CV-TMLE},d^*_{n,v}}$) with respect to the data-adaptive parameter $\psi_{0,d^*_{n,v}}$ 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 0.13% (0.0090–0.0154 times that of alternative estimators), variance of 0.0007 (0.06–1.00 times that of alternative estimators), MSE of 0.0007 (0.02–0.09 times that of alternative estimators), and 93.9% confidence interval coverage.

## 6 Evaluating the estimated ODTR for the "Interventions" study

In our companion paper, we estimated the ODTR on the "Interventions" data ($n = 441$) using the ODTR SuperLearner. The library for $d^*_n$ consisted of a combination of algorithms based on simple parametric models and machine learning (`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. Thus, there is insufficient evidence to conclude that assigning CBT using the ODTR SuperLearner is better than assigning CBT in a non-individualized way.

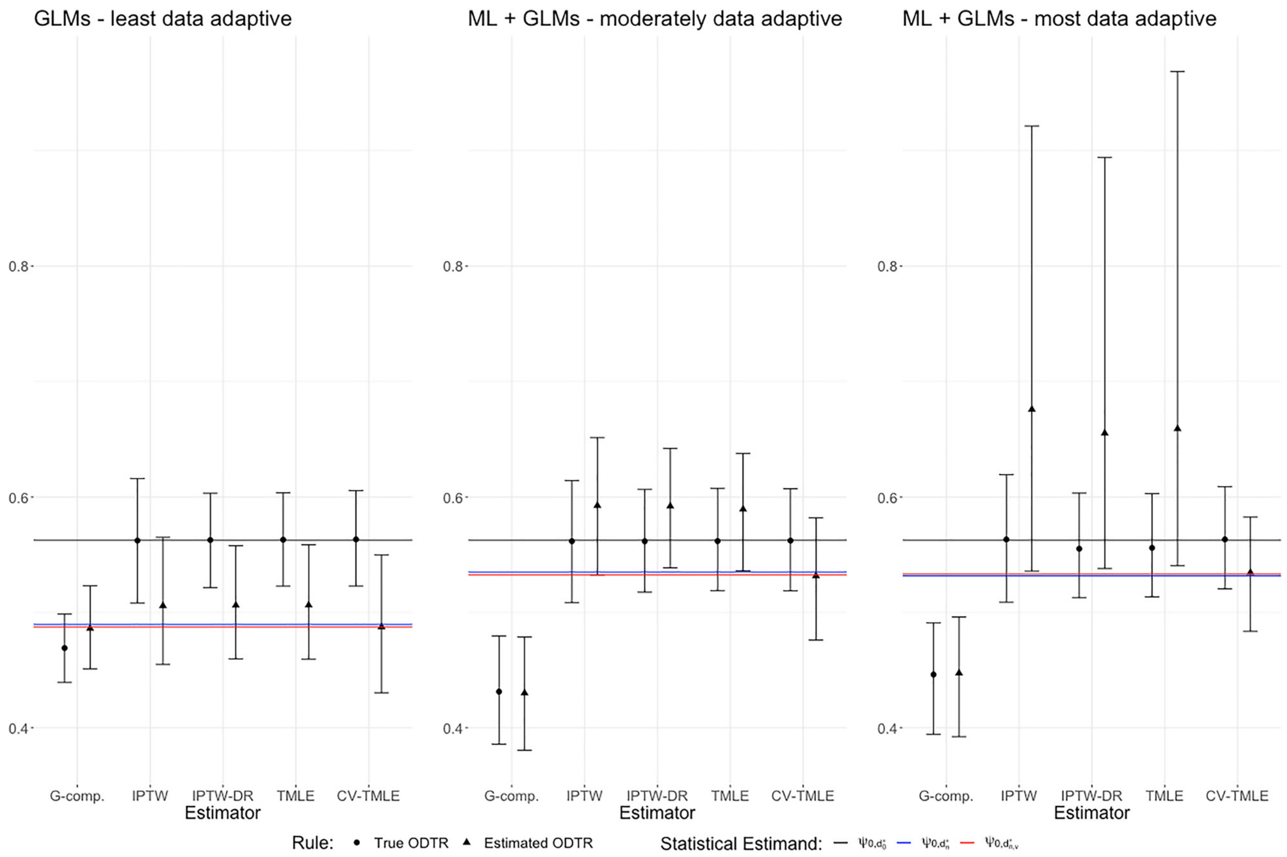

**Figure 1:** Performance of candidate estimators of the value of a given rule, evaluated for 3 SuperLearner library configurations with increasing (left to right) levels of data-adaptivity used for estimating the true outcome regression $Q_0$ and/or the true optimal rule $d_0^*$ ("GLM – least data adaptive", "ML + GLMs – moderately data adaptive", "ML + GLMs – most data adaptive"). The horizontal black line depicts the true mean outcome under the true ODTR $\psi_{0,d_0^*}$; the blue and red lines are true values of the data-adaptive parameters $\psi_{0,d_n^*}$ and $\psi_{0,d_{n,v}^*}$, respectively, averaged over each of the 1000 simulated samples. Points with error bars show the distribution of the estimators across the 1000 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 2.5th and 97.5th 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).

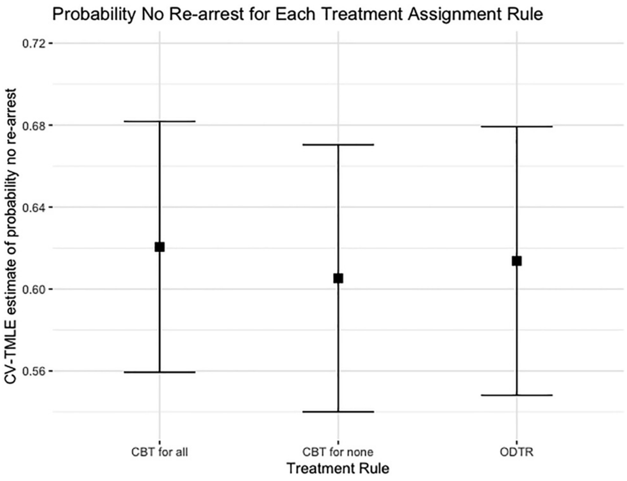

**Figure 2:** Analysis of the "Interventions" study. CV-TMLE estimates of the probability of no re-arrest under the following treatment rules: give cognitive behavioral therapy (CBT) to all, give CBT to none, give CBT according to the ODTR SuperLearner algorithm. The squares are the point estimates and 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.

## 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). Of note, we see similar results when there is dependence between covariates in the DGP, as shown in the Appendix. 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, when nuisance parameters and ODTRs were estimated using the most data-adaptive algorithms, CV-TMLE performed the best among the candidate estimators, while non-cross-validated estimators yielded overly optimistic and highly variable results. Finally, no 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). That said, the superior performance of CV-TMLE does come at the cost of estimating a distinct, data-adaptive parameter that depends not only on the sample at hand, but also the sample split.

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 [12, 50, 51] 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 participants treatment.

**Author contribution:** All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.
**Research funding:** None declared.
**Conflict of interest statement:** The authors declare no conflicts of interest regarding this article.

# References

1. Khoury MJ, Iademarco MF, Riley WT. Precision public health for the era of precision medicine. Am J Prev Med 2016;50:398–401.
2. Laber EB, Davidian M. Dynamic treatment regimes, past, present, and future: a conversation with experts. Stat Methods Med Res 2017;26:1605–10.
3. Bembom O, van der Laan MJ. A practical illustration of the importance of realistic individualized treatment rules in causal inference. Electronic Journal of Statistics 2007;1:574–96.
4. Chakraborty B, Laber EB, Zhao Y. Inference for optimal dynamic treatment regimes using an adaptive m-out-of-n bootstrap scheme. Biometrics 2013;69:714–23.
5. Chakraborty B, Murphy SA. Dynamic treatment regimes. Annual Review of Statistics and its Application 2014;1:447–64.
6. Robins JM. A new approach to causal inference in mortality studies with a sustained exposure period - application to control of the healthy worker survivor effect. Math Model 1986;7:9–12.
7. van der Laan MJ, Petersen ML. Causal effect models for realistic individualized treatment and intention to treat rules. Int J Biostat 2007;3:1557–4679.
8. Moodie EE, Richardson TS, Stephens DA. Demystifying optimal dynamic treatment regimes. Biometrics 2007;63:447–55.
9. Murphy SA. Optimal dynamic treatment regimes. J Roy Stat Soc B 2003;65:331–55.
10. Robins JM. Optimal structural nested models for optimal sequential decisions. In: Proceedings of the Second Seattle Symposium in Biostatistics. New York: Springer; 2004:189–326 pp.
11. Kosorok MR, Laber EB. Precision medicine. Annual Review of Statistics and its Application 2019;6:263–86.
12. Kosorok MR, Moodie EE. Adaptive Treatment Strategies in Practice: Planning Trials and Analyzing Data for Personalized Medicine. Philadelphia: Society for Industrial and Applied Mathematics; 2015.
13. Tsiatis AA, Davidian M, Holloway ST, Laber EB. Dynamic Treatment Regimes: Statistical Methods for Precision Medicine. Boca Raton: Chapman and Hall/CRC Press; 2019.
14. Coyle JR. Computational considerations for targeted learning [Ph.D. thesis]. Berkeley: University of California; 2017.
15. Luedtke AR, van der Laan MJ. Super-learning of an optimal dynamic treatment rule. Int J Biostat 2016;12:305–32.
16. van der Laan MJ, Polley EC, Hubbard AE. Super learner. Stat Appl Genet Mol Biol 2007;6. https://doi.org/10.2202/1544-6115.1309.
17. Montoya L, van der Laan MJ, Luedtke AR, Skeem JL, Coyle JR, Petersen ML. The optimal dynamic treatment rule superlearner: considerations, performance, and application. Int J Biostat [Forthcoming].
18. Hernan MA, Robins JM. Estimating causal effects from epidemiological data. J Epidemiol Community Health 2006;60:578–86.
19. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika 1983;70:41–55.
20. Robins JM. Robust estimation in sequentially ignorable missing data and causal inference models. In: Proceedings of the American Statistical Association. American Statistical Association; 1999;6–10 pp.
21. Robins JM, Rotnitzky A, Zhao LP. Estimation of regression coefficients when some regressors are not always observed. J Am Stat Assoc 1994;89:846–66.
22. Scharfstein DO, Rotnitzky A, Robins JM. Adjusting for nonignorable drop-out using semiparametric nonresponse models: Rejoinder. J Am Stat Assoc 1999;94:1135–46.
23. Rosenblum M, van der Laan MJ. Targeted maximum likelihood estimation of the parameter of a marginal structural model. Int J Biostat 2010;6:1557–4679.

24. van der Laan MJ, Rose S. Targeted learning: causal inference for observational and experimental data. New York: Springer; 2011.
25. van der Laan MJ, Luedtke AR. Targeted learning of the mean outcome under an optimal dynamic treatment rule. J Causal Inference 2015;3:61–95.
26. van der Laan MJ, Rose S. Targeted learning in data science. New York: Springer; 2018.
27. Zheng W, van der Laan MJ. Asymptotic theory for cross-validated targeted maximum likelihood estimation. In: U.C. Berkeley division of biostatistics working paper series; 2010.
28. Laber EB, Qian M. Evaluating personalized treatment regimes. New York: Chapman and Hall/CRC; 2017: 483–97 pp. chapter 15.
29. Chakraborty B, Murphy S, Strecher V. Inference for non-regular parameters in optimal dynamic treatment regimes. Stat Methods Med Res 2010;19:317–43.
30. Sies A, Van Mechelen I. Estimating the quality of optimal treatment regimes. Stat Med 2019;38:4925–38.
31. Hubbard AE, Kherad-Pajouh S, van der Laan MJ. Statistical inference for data adaptive target parameters. Int J Biostat 2016;12:3–19.
32. Skeem JL, Manchak S, Peterson JK. Correctional policy for offenders with mental illness: creating a new paradigm for recidivism reduction. Law Hum Behav 2011;35:110–26.
33. Skeem JL, Winter E, Kennealy PJ, Louden JE, Tatar JR II. Offenders with mental illness have criminogenic needs, too: toward recidivism reduction. Law Hum Behav 2014;38:212–24.
34. Petersen ML, van der Laan MJ. Causal models and learning from data: integrating causal modeling and statistical estimation. Epidemiology 2014;25:418–26.
35. Petersen ML, Porter KE, Gruber S, Wang Y, van der Laan MJ. Diagnosing and responding to violations in the positivity assumption. Stat Methods Med Res 2012;21:31–54.
36. Luedtke AR, van der Laan MJ. Statistical inference for the mean outcome under a possibly non-unique optimal treatment strategy. Ann Stat 2016;44:713–42.
37. Gruber S, van der Laan MJ. A targeted maximum likelihood estimator of a causal effect on a bounded continuous outcome. Int J Biostat 2010;6:1557–4679.
38. van der Laan MJ, Robins JM. Unified methods for censored longitudinal data and causality. New York: Springer; 2003.
39. van der Vaart AW. Asymptotic statistics. Cambridge: Cambridge University Press; 1998.
40. Bickel PJ, Klaassen CAJ, Ritov Y, Klaassen JA. Efficient and adaptive estimation for semiparametric models. Baltimore: Johns Hopkins University Press; 1993, volume 4.
41. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2020. Available from: https://www.R-project.org/.
42. Coyle J. tlverse: Umbrella package for targeted learning in R; 2021. Available from: https://github.com/tlverse/tlverse, \ignorespacesr\ignorespacespackage\ignorespacesversion\ignorespaces0.0.1.
43. Friedman JH. Multivariate adaptive regression splines. Ann Stat 1991;19:1–67.
44. Ripley BD. Pattern Recognition and Neural Networks. Cambridge: Cambridge University Press; 1996.
45. Chang C-C, Lin C-J. Libsvm: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology 2011;2:1–27.
46. Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and Regression Trees. New York: Routledge; 1984.
47. Polley E, LeDell E, Kennedy C, van der Laan M. SuperLearner: Super Learner Prediction; 2020. Available from: https://CRAN.R-project.org/package=SuperLearner, r package version 2.0-28.
48. Breiman L. Random forests. Mach Learn 2001;45:5–32.
49. Rosenblum M, van der Laan MJ. Using regression models to analyze randomized trials: asymptotically valid hypothesis tests despite incorrectly specified models. Biometrics 2009;65:937–45.
50. Almirall D, Nahum-Shani I, Sherwood NE, Murphy SA. Introduction to SMART designs for the development of adaptive interventions: with application to weight loss research. Translational Behavioral Medicine 2014;4:260–74.
51. Lei H, Nahum-Shani I, Lynch K, Oslin D, Murphy SA. A ”SMART” design for building individualized treatment sequences. Annu Rev Clin Psychol 2012;8:21–48.

**Supplementary Material:** The online version of this article offers supplementary material (https://doi.org/10.1515/ijb-2020-0128).