Mark J. van der Laan* and Alexander R. Luedtke

Targeted Learning of the Mean Outcome under an Optimal Dynamic Treatment Rule

Abstract: We consider estimation of and inference for the mean outcome under the optimal dynamic two time-point treatment rule defined as the rule that maximizes the mean outcome under the dynamic treatment, where the candidate rules are restricted to depend only on a user-supplied subset of the baseline and intermediate covariates. This estimation problem is addressed in a statistical model for the data distribution that is nonparametric beyond possible knowledge about the treatment and censoring mechanism. This contrasts from the current literature that relies on parametric assumptions. We establish that the mean of the counterfactual outcome under the optimal dynamic treatment is a pathwise differentiable parameter under conditions, and develop a targeted minimum loss-based estimator (TMLE) of this target parameter. We establish asymptotic linearity and statistical inference for this estimator under specified conditions. In a sequentially randomized trial the statistical inference relies upon a second-order difference between the estimator of the optimal dynamic treatment and the optimal dynamic treatment to be asymptotically negligible, which may be a problematic condition when the rule is based on multivariate time-dependent covariates. To avoid this condition, we also develop TMLEs and statistical inference for data adaptive target parameters that are defined in terms of the mean outcome under the estimate of the optimal dynamic treatment. In particular, we develop a novel cross-validated TMLE approach that provides asymptotic inference under minimal conditions, avoiding the need for any empirical process conditions. We offer simulation results to support our theoretical findings.

Keywords: sequentially randomized controlled trial, cross-validation, dynamic treatment, optimal dynamic treatment, targeted minimum loss-based estimation

DOI 10.1515/jci-2013-0022

1 Introduction

Suppose we observe n in4dependent and identically distributed observations of a time-dependent random variable consisting of baseline covariates, initial treatment and censoring indicator, intermediate covariates, subsequent treatment and censoring indicator, and a final outcome. For example, this could be data generated by a sequentially randomized controlled trial (RCT) in which one follows up a group of subjects, and treatment assignment at two time points is sequentially randomized, where the probability of receiving treatment might be determined by a baseline covariate for the first-line treatment, and time-dependent intermediate covariate (such as a biomarker of interest) for the second-line treatment [1]. Such trials are often called sequential multiple assignment randomized trials (SMART). A dynamic treatment rule deterministically assigns treatment as a function of the available history. If treatment is assigned at two time points, then this dynamic treatment rule consists of two rules, one for each time point [1–4]. The mean outcome under a dynamic treatment is a counterfactual quantity of interest representing what the mean outcome would have been if everybody would have received treatment according to the dynamic treatment rule [5–11]. Dynamic treatments represent prespecified multiple time-point interventions that at each treatment-decision stage are allowed to respond to the currently available treatment and covariate history.

***Corresponding author: Mark J. van der Laan,** University of California – Berkeley, Berkeley, CA, USA, E-mail: laan@berkeley.edu
Alexander R. Luedtke, Department of Biostatistics, University of California – Berkeley, Berkeley, CA, USA,
E-mail: aluedtke@berkeley.edu

Examples of multiple time-point dynamic treatment regimes are given in Lavori and Dawson [12, 13]; Murphy [14]; Rosthøj et al. [15]; Thall et al. [16, 17]; Wagner et al. [18]; Petersen et al. [19]; van der Laan and Petersen [20]; and Robins et al. [21], ranging from rules that change the dose of a drug, change or augment the treatment, to making a decision on when to start a new treatment, in response to the history of the subject.

More recently, SMART designs have been implemented in practice: Lavori and Dawson [12, 22]; Murphy [14]; Thall et al. [16]; Chakraborty et al. [23]; Kasari [24]; Lei et al. [25]; Nahum-Shani et al. [26, 27]; Jones [28]; Lei et al. [25]. For an extensive list of SMARTs, we refer the reader to the website http://methodology.psu.edu/ra/adap-inter/projects. For an excellent and recent overview of the literature on dynamic treatments we refer to Chakraborty and Murphy [29].

We define the optimal dynamic multiple time-point treatment regime as the rule that maximizes the mean outcome under the dynamic treatment, where the candidate rules are restricted to only respond to a user-supplied subset of the baseline and intermediate covariates. The literature on Q-learning shows that we can describe the optimal dynamic treatment among *all* dynamic treatments in a sequential manner [14, 30–33]. The optimal rule can be learned through fitting the likelihood and then calculating the optimal rule under this fit of the likelihood. This approach can be implemented with maximum likelihood estimation based on parametric models. It has been noted (e.g., Robins [32], Chakraborty and Murphy [29]) that the estimator of the parameters of one of the regressions (except the first one) when using parametric regression models is a non-smooth function of the estimator of the parameters of the previous regression, and that this results in non-regularity of the estimators of the parameter vector. This raises challenges for obtaining statistical inference, even when assuming that these parametric regression models are correctly specified. Chakraborty and Murphy [29] discuss various approaches and advances that aim to resolve this delicate issue such as inverting hypothesis testing [32], establishing non-normal limit distributions of the estimators (E. Laber, D. Lizotte, M. Qian, S. Murphy, submitted), or using the m out of n bootstrap.

Murphy [30] and Robins [31, 32] developed structural nested mean models tailored to optimal dynamic treatments. These models assume a parametric model for the "blip function" defined as the additive effect of a blip in current treatment on a counterfactual outcome, conditional on the observed past, in the counterfactual world in which future treatment is assigned optimally. Statistical inference for the parameters of the blip function proceeds accordingly, but Robins [32] points out the irregularity of the estimator, resulting in some serious challenges for statistical inference as referenced above. Structural nested mean models have also been generalized to blip functions that condition on a (counterfactual) subset of the past, thereby allowing the learning of optimal rules that are restricted to only using this subset of the past [32] and Section 6.5 in van der Laan and Robins [34].

An alternative approach, referenced as the direct approach in Chakraborty and Murphy [29], uses marginal structural models (MSMs) for the dynamic regime-specific mean outcome for a user-supplied class of dynamic treatments. If one assumes the marginal structural models are correctly specified, then the parameters of the marginal structural model map into a dynamic treatment that is optimal among the user-supplied class of dynamic regimes. In addition, the MSM also provides the complete dose–response curve, that is, the mean counterfactual outcome for each dynamic treatment in the user-supplied class. This generalization of the original marginal structural models for static interventions to MSMs for dynamic treatments was developed independently by Orellana et al. [35]; van der Laan and Petersen [20]. These articles present inverse probability of treatment and censoring weighted (IPCW) estimators and double robust augmented IPCW estimators based on general longitudinal data structures, allowing for right censoring, time-dependent covariates, and survival outcomes. Double robust estimating equation-based methods that estimate the nuisance parameters with sequential parametric regression models using clever covariates were developed for static intervention MSMs by Bang and Robins [36]. An analogous targeted minimum loss-based estimator (TMLE) [37–39] was developed for marginal structural models for a user-supplied class of dynamic treatments by Petersen et al. [40]. This estimator builds on the TMLE for the mean outcome for a single dynamic treatment developed by van der Laan and Gruber [41]. Additional application papers of interest are [42–44] which involve fitting MSMs for dynamic treatments defined by treatment-tailoring threshold using IPCW methods.

Each of the above referenced approaches for learning an optimal dynamic treatment that also aims to provide statistical inference relies on parametric assumptions: obviously, Q-learning based on parametric models, but also the structural nested mean models and the marginal structural models both rely on parametric models for the blip function and dose–response curve, respectively. As a consequence, even in a SMART, the statistical inference for the optimal dynamic treatment heavily relies on assumptions that are generally believed to be false, and will thus be expected to be biased.

To avoid such biases, we define the statistical model for the data distribution as nonparametric, beyond possible knowledge about the treatment mechanism (e.g., known in an RCT) and censoring mechanism. This forces us to define the optimal dynamic treatment and the corresponding mean outcome as parameters defined on this nonparametric model, and to develop data adaptive estimators of the optimal dynamic treatment. In order to not only consider the most ambitious fully optimal rule, we define the V-optimal rules as the optimal rule that only uses a user-supplied subset V of the available covariates. This allows us to consider suboptimal rules that are easier to estimate and thereby allow for statistical inference for the counterfactual mean outcome under the suboptimal rule. This is analogous to the generalized structural nested mean models whose blip functions only condition on a counterfactual subset of the past. In a companion article we describe how to estimate the V-optimal rule.

In Example 4 of Robins et al. [45], the authors develop an asymptotic confidence set for the optimal treatment regime in an RCT under a large semiparametric model that only assumes that the treatment mechanism is known. This confidence set is certainly of interest and warrants further consideration in the optimal treatment literature. They get this confidence set by deriving the efficient influence curve for the mean squared blip function. They propose selecting a data adaptive estimate of the optimal treatment rule by a particular cross-validation scheme over a set of basis functions, and show that this estimator achieves a data adaptive rate of convergence under smoothness assumptions on the blip function. Our work is distinct from this earlier work in that the earlier work does not directly consider the mean outcome under the optimal rule and only considers data generated by a point treatment RCT.

In this article we describe how to obtain semiparametric inference about the mean outcome under the two time point V-optimal rule. We will show that the mean outcome under the optimal rule is a pathwise differentiable parameter of the data distribution, indicating that it is possible to develop asymptotically linear estimators of this target parameter under conditions. In fact, we obtain the surprising result that the pathwise derivative of this target parameter equals the pathwise derivative of the mean counterfactual outcome under a given dynamic treatment rule set at the optimal rule, treating the latter as known. By a reference to the current literature for double robust and efficient estimation of the mean outcome under a given rule, we then obtain a TMLE for the mean outcome under the optimal rule. Subsequently, we prove asymptotic linearity and efficiency of this TMLE, allowing us to construct confidence intervals for the mean outcome under the optimal dynamic treatment or its contrast with respect to a standard treatment. Thus, contrary to the irregularity of the estimators of the unknown parameters in the semiparametric structural nested mean model, we can construct regular estimators of the mean outcome under the optimal rule in the nonparametric model.

In a SMART the statistical inference would only rely upon a second-order difference between the estimator of the optimal dynamic treatment and the optimal dynamic treatment itself to be asymptotically negligible. This is a reasonable condition if we restrict ourselves to rules only responding to a one-dimensional time-dependent covariate, or if we are willing to make smoothness assumptions. To avoid this condition, we also develop TMLEs and statistical inference for data adaptive target parameters that are defined in terms of the mean outcome under the *estimate* of the optimal dynamic treatment (see van der Laan et al. [46] for a general approach for statistical inference for data adaptive target parameters). In particular, we develop a novel cross-validated TMLE (CV-TMLE) approach that provides asymptotic inference under minimal conditions.

For the sake of presentation, we focus on two time point treatments in this article. In the appendices of our earlier technical reports [47, 48] we generalize these results to general multiple time point treatments, and develop general (sequential) super-learning based on the efficient CV-TMLE of the risk of a candidate estimator. In this appendix we also develop a TMLE of a projection of the blip functions on a parametric working model (with corresponding statistical inference, which presents a result of interest in its own right). We emphasize that this technical report is distinct from our companion paper in this issue, which focuses on the data adaptive estimation of optimal treatment strategies.

1.1 Organization of article

Section 2 defines the mean outcome under the optimal rule as a causal parameter and gives identifiability assumptions under which the causal parameter is identified with a statistical parameter of the observed data distribution.

The remainder of the paper describes strategies to estimate the counterfactual mean outcome under the optimal rule and related quantities. This paper assumes that we have an estimate of the optimal rule in our semiparametric model. In our companion paper we describe how to obtain estimates of the V-optimal rule.

The first part of this article concerns estimation of the mean outcome under the optimal rule. Section 3 establishes the pathwise differentiability of the mean outcome under the V-optimal rule conditions. A closed form expression for the efficient influence curve for this statistical parameter is given, which represents a key ingredient in semiparametric inference for the statistical target parameter. We obtain the surprising result that, under straightforward conditions, estimating the mean outcome under the unknown optimal treatment rule is the same in first order as estimating the mean outcome under the optimal rule when the rule is known from the outset. Section 4 presents the key properties of a TMLE for the mean outcome under the optimal rule, which is presented in detail in “TMLE of the mean outcome under a given rule” in Appendix B due to its similarity to TMLEs presented previously in the literature. Section 5 presents an asymptotic linearity theorem for this TMLE and corresponding statistical inference.

The second part of this article concerns statistical inference for data adaptive target parameters that are defined in terms of the mean outcome under the estimate of the optimal dynamic treatment, thereby avoiding the consistency and rate condition for the fitted V-optimal rule as required for asymptotic linearity of the TMLE of the mean outcome under the actual V-optimal rule. These results are of interest in practice because an estimated, possibly suboptimal, rule will be implemented in the population, not some unknown optimal rule. Section 6 presents an asymptotic linearity theorem for the TMLE presented in Section 4, but now with the target parameter defined as the mean outcome under the estimated rule. In Section 7 we present the CV-TMLE framework. A specific CV-TMLE algorithm is described in “CV-TMLE of the mean outcome under data adaptive V-optimal rule” in Appendix B due to its similarity to CV-TMLEs presented previously in the literature. The CV-TMLE provides asymptotic inference under minimal conditions for the mean outcome under a dynamic treatment fitted on a training sample, averaged across the different splits in training sample and validation sample. Both results allow us to construct confidence intervals that have the correct asymptotic coverage of the random true target parameter, and the fixed mean outcome under the optimal rule under conditions, but statistical inference based on the CV-TMLE does not require an empirical process condition that would put a brake on the allowed data adaptivity of the estimator.

Section 8 presents the simulation methods. The simulations estimate the optimal rule using an ensemble algorithm presented in our companion paper, and then given this estimate apply the estimators of the optimal rule presented in this paper. Section 9 presents the coverage and efficiency of the various estimators in our simulation. Appendix C gives analytic intuition as to why some of the simulation results may have occurred. Section 10 closes with a discussion and directions for future work.

All proofs can be found in Appendix A.

2 Formulation of optimal dynamic treatment estimation problem

Suppose we observe n i.i.d. copies $O_1, \ldots, O_n \in \mathcal{O}$ of

$$O = (L(0), A(0), L(1), A(1), Y) \sim P_0,$$

where $A(j) = (A_1(j), A_2(j))$, $A_1(j)$ is a binary treatment, and $A_2(j)$ is an indicator of not being right censored at "time" j, $j = 0,\ 1$. That is, $A_2(0) = 0$ implies that $(L(1), A_1(1), Y)$ is n ot observed, and $A_2(1) = 0$ implies that Y is not observed. Each time point j has covariates $L(j)$ that precede treatment, $j = 0, 1$, and the outcome of interest is given by Y and occurs after time point 1. For a time-dependent process $X(\cdot)$, we use the notation $\bar{X}(t) = (X(s) : s \leq t)$, where $\bar{X}(-1) = \emptyset$. Let $\mathcal{M}$ be a statistical model that makes no assumptions on the marginal distribution $Q_{0,L(0)}$ of $L(0)$ and the conditional distribution $Q_{0,L(1)}$ of $L(1)$, given $A(0), L(0)$, but might make assumptions on the conditional distributions $g_{0A(j)}$ of $A(j)$, given $\bar{A}(j-1), \bar{L}(j)$, $j = 0, 1$. We will refer to g_0 as the intervention mechanism, which can be factorized in a treatment mechanism g_{01} and censoring mechanism g_{02} as follows:

$$g_0(O) = \prod_{j=1}^{2} g_{01}\big(A_1(j)|\bar{A}(j-1), \bar{L}(j)\big) g_{02}\big(A_2(j)|A_1(j), \bar{A}(j-1), \bar{L}(j)\big).$$

In particular, the data might have been generated by a SMART, in which case g_{01} is known.

Let $V(1)$ be a function of $(L(0), A(0), L(1))$, and let $V(0)$ be a function of $L(0)$. Let $V = (V(0), V(1))$. Consider dynamic treatment rules $V(0) \to d_{A(0)}(V(0)) \in \{0,1\} \times \{1\}$ and $(A(0), V(1)) \to d_{A(1)}(A(0), V(1)) \in \{0,1\} \times \{1\}$ for assigning treatment $A(0)$ and $A(1)$, respectively, where the rule for $A(0)$ is only a function of $V(0)$, and the rule for $A(1)$ is only a function of $(A(0), V(1))$. Note that these rules are restricted to set the censoring indicators $A_2(j) = 1, j = 0, 1$. Let $\mathcal{D}$ be the set of all such rules. We assume that $V(0)$ is a function of $V(1)$ (i.e., observing $V(1)$ includes observing $V(0)$), but in the theorem below we indicate an alternative assumption. For $d \in \mathcal{D}$, we let

$$d(a(0), v) \equiv \big(d_{A(0)}(v(0)), d_{A(1)}(a(0), v(1))\big).$$

If we assume a structural equation model [7] for variables stating that

$$\begin{aligned}
L(0) &= f_{L(0)}\big(U_{L(0)}\big)\\
A(0) &= f_{A(0)}\big(L(0), U_{A(0)}\big)\\
L(1) &= f_{L(1)}\big(L(0), A(0), U_{L(1)}\big)\\
A(1) &= f_{A(1)}\big(\bar{L}(1), A(0), U_{A(1)}\big)\\
Y &= f_Y\big(\bar{L}(1), \bar{A}(1), U_Y\big),
\end{aligned}$$

where the collection of functions $f = (f_{L(0)}, f_{A(0)}, f_{L(1)}, f_{A(1)})$ is unspecified or partially specified, we can define counterfactuals Y_d defined by the modified system in which the equations for $A(0), A(1)$ are replaced by $A(0) = d_{A(0)}(V(0))$ and $A(1) = d_{A(1)}(A(0), V(1))$, respectively. Denote the distribution of these counterfactual quantities as $P_{0,d}$, where we note that $P_{0,d}$ is implied by the collection of functions f and the joint distribution of exogeneous variables $(U_{L(0)}, U_{A(0)}, U_{L(1)}, U_{A(1)}, U_Y)$. We can now define the causally optimal rule under $P_{0,d}$ as $d_0^* = \arg\max_{d \in \mathcal{D}} E_{P_{0,d}} Y_d$. If we assume a sequential randomization assumption stating that $A(0)$ is independent of $U_{L(1)}, U_Y$, given $L(0)$, and $A(1)$ is independent of U_Y, given $\bar{L}(1), A(0)$, then we can identify $P_{0,d}$ with observed data under the distribution P_0 using the G-computation formula:

$$\begin{aligned}
&p_{0,d}(L(0), A(0), L(1), A(1), Y)\\
&\quad \equiv I(A = d(A(0), V)) q_{0,L(0)}(L(0)) q_{0,L(1)}(L(1)|L(0), A(0)) q_{0,Y}\big(Y|\bar{L}(1), \bar{A}(1)\big),
\end{aligned} \tag{1}$$

where $p_{0,d}$ is the density of $P_{0,d}$ and $q_{0,L(0)}$, $q_{0,L(1)}$, and $q_{0,Y}$ are the densities for $Q_{0,L(0)}$, $Q_{0,L(1)}$, and $Q_{0,Y}$, respectively, where $Q_{0,Y}$ represents the distribution of Y given $\bar{L}(1), \bar{A}(1)$. We assume that all densities above are absolutely continuous with respect to some dominating measure μ. We have a similar identifiability result/G-computation formula under the Neyman-Rubin causal model [8]. For the right censoring indicators $A_2(0)$ and $A_2(1)$, we note the parallel between the coarsening at random assumption and the sequential randomization assumption [49]. Thus here we have encoded our missingness assumptions in our causal assumptions.

More generally, for a distribution $P \in \mathcal{M}$ we can define the G-computation distribution P_d as the distribution with density

$$\begin{aligned}&p_d(L(0), A(0), L(1), A(1), Y)\\ &\quad \equiv I(A = d(A(0), V))q_{L(0)}(L(0))q_{L(1)}(L(1)|L(0), A(0))q_Y(Y|\bar{L}(1), \bar{A}(1)),\end{aligned}$$

where $q_{L(0)}$, $q_{L(1)}$, and q_Y are the counterparts to $q_{0,L(0)}$, $q_{0,L(1)}$, and $q_{0,Y}$, respectively, under P.

For the remainder of this article, if for a static or dynamic intervention d, we use notation L_d (or Y_d, O_d) we mean the random variable with the probability distribution P_d in (1) so that all of our quantities are statistical parameters. For example, the quantity $E_{P_0}(Y_{a(0)a(1)}|V_{a(0)}(1))$ defined in the next theorem denotes the conditional expectation of $Y_{a(0)a(1)}$, given $V_{a(0)}(1)$, under the probability distribution $P_{0,a(0)a(1)}$ (i.e., G-computation formula presented above for the static intervention $(a(0), a(1))$. In addition, if we write down these parameters for some P_d, we will automatically assume the positivity assumption at P required for the G-computation formula to be well defined. For that it will suffice to assume the following positivity assumption at P:

$$\begin{aligned}\Pr_P\left(0< \min_{a_1\in\{0,1\}} g_{0A(0)}(a_1, 1|L(0))\right) &= 1\\ \Pr_P\left(0< \min_{a_1\in\{0,1\}} g_{0A(1)}(a_1, 1|\bar{L}(1), A(0))\right) &= 1. \end{aligned} \tag{2}$$

The strong positivity assumption will be defined as the above assumption, but where the 0 is replaced by a $\delta > 0$.

We now define a statistical parameter representing the mean outcome Y_d under P_d. For any rule $d \in \mathcal{D}$, let

$$\Psi_d(P) \equiv E_{P_d} Y_d.$$

For a distribution P, define the V-optimal rule as

$$d_P = \arg\max_{d\in\mathcal{D}} E_{P_d} Y_d.$$

For simplicity, we will write d_0 instead of d_{P_0} for the V-optimal rule under P_0. Define the parameter mapping $\Psi : \mathcal{M} \to \mathbb{R}$ as $\Psi(P) = E_{P_{d_P}} Y_{d_P}$. The first part of this article is concerned with inference for the parameter

$$\psi_0 \equiv \Psi(P_0) = E_{P_{0,d_0}} Y_{d_0}.$$

Under our identifiability assumptions, d_0 is equal to the causally optimal rule d_0^*. Even if the sequential randomization assumption does not hold, the statistical parameter ψ_0 represents a statistical parameter of interest in its own right. We will not concern ourselves with the sequential randomization assumption for the remainder of this paper.

The next theorem presents an explicit form of the V-optimal individualized treatment rule d_0 as a function of P_0.

Theorem 1. *Suppose $V(0)$ is a function of $V(1)$. The V-optimal rule d_0 can be represented as the following explicit parameter of P_0:*

$$\begin{aligned}
\bar{Q}_{20}(a(0), v(1)) &= \\
&E_{P_0}\left(Y_{a(0),A(1)=(1,1)} | V_{a(0)}(1) = v(1)\right) - E_{P_0}\left(Y_{a(0),A(1)=(0,1)} | V_{a(0)}(1) = v(1)\right) \\
d_{0,A(1)}(A(0), V(1)) &= \left(I(\bar{Q}_{20}(A(0), V(1)) > 0), 1\right) \\
\bar{Q}_{10}(v(0)) &= E_{P_0}\left(Y_{(1,1),d_{0,A(1)}} | V(0)\right) - E_{P_0}\left(Y_{(0,1),d_{0,A(1)}} | V(0)\right) \\
d_{0,A(0)}(V(0)) &= \left(I(\bar{Q}_{10}(V(0)) > 0), 1\right),
\end{aligned}$$

where $a(0) \in \{0,1\} \times \{1\}$. If $V(1)$ does not include $V(0)$, but, for all $(a(0), a(1)) \in \{\{0,1\} \times \{1\}\}^2$,

$$E_{P_0}\left(Y_{a(0),a(1)} | V(0), V_{a(0)}(1)\right) = E_{P_0}\left(Y_{a(0),a(1)} | V_{a(0)}(1)\right), \tag{3}$$

then the above expression for the V-optimal rule d_0 is still true.

3 The efficient influence curve of the mean outcome under *V*-optimal rule

In this section we establish the pathwise differentiability of Ψ and give an explicit expression for the efficient influence curve [34, 50, 51]. Before presenting this result, we give the efficient influence curve for the parameter $\Psi : \mathcal{M} \to \mathbb{R}$ where $\Psi_d(P) \equiv E_P Y_d$ and the rule $d = (d_{A(0)}, d_{A(1)}) \in \mathcal{D}$ is treated as known. This influence curve has previously been presented in the literature [36, 41]. The parameter mapping Ψ_d has efficient influence curve:

$$D^*(d, P) = \sum_{k=0}^{2} D_k^*(d, P)$$

where

$$\begin{aligned}
D_0^*(d, P) &= E_P\left[Y_d | L(0), A(0) = d_{A(0)}(V(0))\right] - E_P Y_d \\
D_1^*(d, P) &= \frac{I\left(A(0) = d_{A(0)}(V(0))\right)}{g_{A(0)}(O)} \\
&\quad \times \left(E_P\left[Y | \bar{A}(1) = d(A(0), V), \bar{L}(1)\right] - E_P\left[Y_d | L(0), A(0) = d_{A(0)}(V(0))\right]\right) \\
D_2^*(d, P) &= \frac{I\left(\bar{A}(1) = d(A(0), V)\right)}{\prod_{j=0}^{1} g_{A(j)}(O)} \left(Y - E_P\left[Y | \bar{A}(1) = d(A(0), V), \bar{L}(1)\right]\right).
\end{aligned} \tag{4}$$

Above $(g_{A(0)}, g_{A(1)})$ is the intervention mechanism under the distribution P. We remind the reader that Y_d has the G-computation distribution from (1) so that:

$$\begin{aligned}
&E_P\left[Y_d | L(0), A(0) = d_{A(0)}(V(0))\right] \\
&\quad = E_P\left[E_P\left[Y | \bar{A}(1) = d(A(0), V), \bar{L}(1)\right] | L(0), A(0) = d_{A(0)}(V(0))\right]
\end{aligned}$$

At times it will be convenient to write $D_k^*(d, Q^d, g)$ instead of $D_k^*(d, P)$, where Q^d represents both of the conditional expectations in the definitions of D_1^* and the marginal distribution of $L(0)$ under P and g represents the intervention mechanism under P. We will denote these conditional expectations under P_0 for a given rule d by Q_0^d. We will similarly at times denote $D^*(d, P)$ by $D^*(d, Q^d, g)$.

Whenever $D^*(P)$ does not contain an argument for a rule d, this $D^*(P)$ refers to the efficient influence curve of the parameter mapping Ψ for which $\Psi(P) = E_P Y_{d_P}$, where the optimal rule d_P under P is not treated as known. Not treating d_P as known means that d_P depends on the input distribution P in the mapping $\Psi(P)$. The following theorem presents the efficient influence curve of Ψ at a distribution P. The main condition on this distribution P is that

$$\max_{a_0(0)\in\{0,1\}} \Pr_P\big(\bar{Q}_2\big((a_0(0),1), V_{a(0)=(a_0(0),1)}\big)=0\big)=0$$

$$\Pr_P\big(\bar{Q}_1(V(0))=0\big)=0, \tag{5}$$

where $\bar{Q}_2$ and $\bar{Q}_1$ are defined analogously to $\bar{Q}_{20}$ and $\bar{Q}_{10}$ in Theorem 1 with the expectations under P_0 replaced by expectations under P. That is, we assume that each of the blip functions under P is nowhere zero with probability 1. Distributions that do not satisfy this assumption have been referred to as "exceptional laws" [32, 52]. These laws are indeed exceptional when one expects that treatment will have a beneficial or harmful effect in all V-strata of individuals. When one only expects that treatment will have an effect on outcome in some but not all strata of individuals then this assumption may be violated. We will make this assumption about P_0 for all subsequent asymptotic linearity results about $E_{P_0}Y_{d_0}$, and we will assume a weaker but still not completely trivial assumption for the data adaptive target parameters in Sections 6 and 7.

Theorem 2. *Suppose $\mathcal{P}\in\mathcal{M}$ such that $Pr_P(|Y|<M)=1$ for some $M<\infty$ and the positivity assumption (2) and (5). Then the parameter $\Psi:\mathcal{M}\to\mathbb{R}$ is pathwise differentiable at P with canonical gradient given by*

$$D^*(P)\equiv D^*(d_P,P)=\sum_{k=0}^{2}D_k^*(d_P,P).$$

That is, $D^*(P)$ equals the efficient influence curve $D^*(d_P,P)$ for the parameter $\Psi_d(P)\equiv E_PY_d$ at the V-optimal rule $d=d_P$, where Ψ_d treats d as given.

The above theorem is proved as Theorem 8 in van der Laan and Luedtke [48] so the proof is omitted here.

We will at times denote $D^*(P)$ by $D^*(Q,g)$, where Q represents Q^{d_P}, along with portions of the likelihood which suffice to compute the V-optimal rule d_P. We denote d_P by d_Q when convenient. We explore which parts of the likelihood suffice to compute the V-optimal rule in our companion paper, though Theorem 1 shows that $\bar{Q}_{20}$ and $\bar{Q}_{10}$ suffice for d_0 (and analogous functions suffice for a more general d_P). We have the following property of the efficient influence curve, which will provide a fundamental ingredient in the analysis of the TMLE presented in the next section.

Theorem 3. *Let d_Q be the V-optimal rule corresponding with Q. For any Q, g, we have*

$$P_0D^*(Q,g)=\Psi(Q_0)-\Psi(Q)+R_{1d_Q}\Big(Q^{d_Q},Q_0^{d_Q},g,g_0\Big)+R_2(Q,Q_0)$$

where for all $d\in\mathcal{D}$

$$R_{1d}(Q^d,Q_0^d,g,g_0)\equiv P_0D^*(d,Q^d,g)-(\Psi_d(Q_0^d)-\Psi_d(Q^d)),$$

$\Psi_d(P)=E_PY_d$ is the statistical target parameter that treats d as known, and $D^*(d,Q_0^d,g_0)$ is the efficient influence curve of Ψ_d at P_0 as given in Theorem 2. In addition,

$$\begin{aligned}R_2(Q,Q_0)&\equiv\Psi_{d_Q}(Q_0^{d_Q})-\Psi_{d_0}(Q_0^{d_0})\\&=E_{P_0}\big(d_{Q,A(0)}-d_{0,A(0)}\big)\big(V(0)\big)\bar{Q}_{10}\big(V(0)\big)\\&\quad+E_{P_0}\big(d_{Q,A(1)}-d_{0,A(1)}\big)\big((0,1),V_{(0,1)}(1)\big)\bar{Q}_{20}\big((0,1),V_{(0,1)}(1)\big)\\&\equiv R_{2A(0)}(Q,Q_0)+R_{2A(1)}(Q,Q_0).\end{aligned}$$

From the study of the statistical target parameter Ψ_d in van der Laan and Gruber [41], we know that $P_0D^*(d,Q^d,g)=\Psi_d(Q_0^d)-\Psi_d(Q^d)+R_{1d}(Q^d,Q_0^d,g,g_0)$, where R_{1d} is a closed form second-order term involving integrals of differences $Q^d-Q_0^d$ times differences $g-g_0$.

The following lemma bounds R_2. We note that this lemma, which concerns how well we can estimate d_0 rather than how well we can make inference about $E_{P_0}Y_{d_0}$, does not require condition (5) to hold. We showed in Theorem 1 that knowing the blip functions $\bar{Q}_{10}$ and $\bar{Q}_{20}$ suffices to define the optimal rule d_0. For general Q, we will let $\bar{Q}_1$ and $\bar{Q}_2$ represent the blip functions under this parameter mapping.

Lemma 1. *Let R_2 be as in Theorem 3. Let $P_{0,(0,1)}$ represent the static intervention-specific G-computation distribution where treatment $(0,1)$ is given at the first time point. Suppose there exist some $\beta_1, \beta_2 > 1$ such that:*

$$E_{P_0}\left[|\bar{Q}_{10}(V(0))|^{-\beta_1} I\left(|\bar{Q}_{10}(V(0))| > 0\right)\right] < \infty$$

$$E_{P_{0,(0,1)}}\left[|\bar{Q}_{20}\left((0,1), V_{(0,1)}(0)\right)|^{-\beta_2} I\left(|\bar{Q}_{20}((0,1), V_{(0,1)}(0))| > 0\right)\right] < \infty, \qquad (6)$$

where the expression in each expectation is taken to be 0 when the indicator is 0. Fix $p \in (1,\infty]$ and define $h : (1,\infty] \times (1,\infty)$ as the function for which $h(p,\beta) = \frac{p(\beta+1)}{p+\beta}$ when $p < \infty$ and $h(p,\beta) = \beta + 1$ otherwise. Then:

$$R_{2A(0)}(Q, Q_0) \leq K_1 \left\|\bar{Q}_1 - \bar{Q}_{10}\right\|_{p,P_0}^{h(p,\beta_1)}$$

$$R_{2A(1)}(Q, Q_0) \leq K_2 \left\|\bar{Q}_2 - \bar{Q}_{20}\right\|_{p,P_{0,(0,1)}}^{h(p,\beta_2)},$$

where $\|\cdot\|_{p,P}$ denotes the $L_{p,P}$ norm for the distribution P and $K_1, K_2 \geq 0$ are finite constants that respectively rely on p, P_0, β_1 and p, $P_{0,(0,1)}$, β_2.

The conditions in (6) are moment bounds which ensure that $\bar{Q}_{10}$ and $\bar{Q}_{20}$ do not put too much mass around zero. To get the tightest bound, we should always choose β_1, β_2 to be as large as possible. We remind the reader that convergence in $L_{p,P}$ implies convergence in $L_{q,P}$ for all distributions P and $1 \leq q \leq p \leq \infty$. Hence there is a trade-off between the chosen bounding norm, $L_{p,P}$, and the rate we need to obtain with respect to that norm so that the term can be expected to be of order $n^{-1/2}$. See Table 1 for some examples of rates of convergence that suffice to give $R_{2A(0)} = o_{P_0}(n^{-1/2})$.

Using the upper bound on $\bar{Q}_{10}$ and applying Cauchy-Schwarz inequality to eq. (15) in the proof of the lemma shows that:

$$R_{2A(0)}(Q, Q_0) \leq \left\|\bar{Q}_1 - \bar{Q}_{10}\right\|_{2,P_0} \sqrt{Pr_{P_0}\left(0 < |\bar{Q}_{10}| < |\bar{Q}_1 - \bar{Q}_{10}|\right)}.$$

Hence $R_{2A(0)} = o_{P_0}(n^{-1/2})$ without any moment condition when $\left\|\bar{Q}_1 - \bar{Q}_{10}\right\|_{2,P_0} = O_{P_0}(n^{-1/2})$, which occurs when one has correctly specified a parametric model for $\bar{Q}_{10}$. In general it is unlikely that one can correctly specify a parametric model for $\bar{Q}_{10}$. In these cases, Lemma 1 shows that the term $R_{2A(0)}$ will still be $o_{P_0}(n^{-1/2})$ if a moment condition holds and $\bar{Q}_{10}$ is estimated at a sufficient rate. The analogue holds for $\bar{Q}_{20}$.

Table 1: Convergence rates of estimators of $\bar{Q}_{10}$ which suffice for $R_{2A(0)}$ to be $o_{P_0}(n^{-1/2})$ according to Lemma 1. The higher the moments of $\bar{Q}_{10}^{-1}$ that are finite, the slower the estimator needs to converge. It is of course preferable to have an estimator which converges according to the P_0 essential supremum than just in L_{2,P_0}, but whether or not there is convergence in L_{∞,P_0} depends on the estimator used and the underlying distribution P_0.

p	β_1	Sufficient L_{p,P_0} convergence rate
2	1	$o_{P_0}(n^{-3/8})$
	2	$o_{P_0}(n^{-1/3})$
	β_1 large	$o_{P_0}(n^{-(1/4+\varepsilon)})$ for small $\varepsilon > 0$
4	1	$o_{P_0}(n^{-5/16})$
	2	$o_{P_0}(n^{-1/4})$
	β_1 large	$o_{P_0}(n^{-(1/8+\varepsilon)})$ for small $\varepsilon > 0$
∞	1	$o_{P_0}(n^{-1/4})$
	2	$o_{P_0}(n^{-1/6})$
	β_1 large	$o_{P_0}(n^{-\frac{1}{2(\beta_1+1)}})$

The bounds given in Lemma 1 are loose. It is not in general necessary to estimate the blip functions $\bar{Q}_{10}$ and $\bar{Q}_{20}$ correctly, only their signs. As an extreme example of the looseness of the bounds, one can have that $\inf_{v(0)} |\bar{Q}_{1n}(v(0)) - \bar{Q}_{10}(v(0))| \to \infty$ as $n \to \infty$ and *still* have that $R_{2A(0)}(Q, Q_n) = 0$ for all n. Nonetheless, these bounds give interpretable sufficient conditions under which the term R_2 converges faster than a root-n rate. We consider methods that do not directly estimate the blip functions in our companion paper.

4 TMLE of the mean outcome under *V*-optimal rule

Throughout this and the next section we assume that condition (5) holds at P_0. Our proposed TMLE is to first estimate the optimal rule d_0, giving us an estimated rule $d_n(A(0), V) = d_{n,A(0)}(V(0)), d_{n,A(1)}(A(0), V(1))$, and subsequently apply the TMLE of EY_d for a fixed rule d at $d = d_n$ as presented in van der Laan and Gruber [41]. This TMLE is an analogue of the double robust estimating equation method presented in Bang and Robins [36]: see also Petersen et al. [40] for a generalization of the TMLE to marginal structural models for dynamic treatments. In a companion paper we describe a data adaptive estimator of d_0. In this paper we take d_n as given. We review the TMLE for $\Psi_d(P_0) = E_{P_0} Y_d$ at a fixed rule d in "TMLE of the mean outcome under a given rule" in Appendix B. Observations which are only partially observed due to right censoring do not cause a problem for the TMLE. In particular, the TMLE only uses individuals who are not right censored at the first or second time point to obtain initial estimates of $E_{P_0}[Y_d | A(0) = d_{A(0)}(V(0)), L(0)]$ and $E_{P_0}[Y | \bar{A}(1) = d(A(0), V), \bar{L}(1)]$ in (4), respectively. See the appendix for details.

Here we note some of the key properties of the TMLE. Let $Q_n^{d_n *}$ consist of the empirical distribution $Q_{L(0),n}$ of $L(0)$, a regression function $l(0) \mapsto E_n^*[Y_d | L(0) = l(0)]$ that estimates $E_{P_0}[Y_d | L(0)]$, and a regression function

$$\left(a(0), \bar{l}(1)\right) \mapsto E_n^*\left[Y | \bar{A}(1) = d(a(0), v), \bar{L}(1) = \bar{l}(1)\right]$$

that estimates $E_{P_0}[Y | \bar{A}(1) = d(A(0), V), \bar{L}(1)]$, where we note that v is a function of $\bar{l}(1)$. In the appendix we describe our proposed algorithm to get the estimates in $Q_n^{d_n *}$. The proposed TMLE for $\psi_0 = E_{P_0} Y_{d_0}$ is given by

$$\psi_{d_n,n}^* = \Psi_{d_n}(Q_n^{d_n *}) = \frac{1}{n} \sum_{i=1}^{n} E_n^*\left[Y_{d_n} | L(0) = l(0)_i\right],$$

where we have applied the TMLE in the appendix to the case where $d = d_n$, treating d_n as known. Note that $\Psi_{d_n}(Q_n^{d_n *})$ is a plug-in estimator in that it is obtained by plugging $Q^{d_n^*}$ into the parameter mapping $Q^d \mapsto \Psi_d(Q^d)$ for $d = d_n$. We expect our plug-in estimator to give reasonable estimates in finite samples because it naturally respects the constraints of our model. In the next section we show that this estimator also enjoys many desirable asymptotic properties.

Recall that $D^*(d, Q^d, g)$ is the efficient influence curve for the target parameter $E_{P_0} Y_d$ which treats d as fixed, and Theorem 2 showed that $D^*(d_0, Q_0^{d_0}, g_0)$ is the efficient influence curve of the target parameter EY_{d_0} where d_0 is the V-optimal rule. The TMLE $(d_n, Q_n^{d_n *})$ described in the appendix solves the efficient influence curve estimating equation:

$$P_n D^*\left(d_n, Q_n^{d_n *}, g_n\right) = 0. \tag{7}$$

Further, one can show using standard M-estimator analysis that the targeted $Q_n^{d_n *}$ proposed in the appendix maintains the same rate of convergence as the initial estimator $Q_n^{d_n}$ under very mild conditions. We do not concern ourselves with these conditions in this paper, and will instead state all conditions directly in terms of $Q_n^{d_n *}$. The above will be a key ingredient in proving the asymptotic linearity of the TMLE for $\psi_0 = E_{P_0} Y_{d_0}$.

5 Asymptotic efficiency of the TMLE of the mean outcome under the *V*-optimal rule

We now wish to analyze the TMLE $\psi_n^* = \Psi_{d_n}(Q_n^{d_n *})$ of $\psi_0 = \Psi_{d_0}(Q_0^{d_0}) = \Psi(Q_0)$. We first give a representation that will allow us to prove the asymptotic linearity of the TMLE under conditions. The result allows $Q_n^{d_n *}$ to be misspecified, even though the intervention mechanism g_0 and the rule d_n are assumed to be consistent for g_0 and d_0, respectively.

Theorem 4. *Assume $Y \in [0,1]$, the strong positivity assumption, condition (5) at P_0, $D_n^* \equiv D^*(d_n, Q_n^{d_n *}, g_n)$ falls in a P_0-Donsker class with probability tending to 1, $P_0\{D_n^* - D^*(d_0, Q^{d_0}, g_0)\}^2$ converges to zero in probability for some Q^{d_0}, and*

$$R_2(Q_n, Q_0) = o_{P_0}(1/\sqrt{n}),$$

where R_2 is defined in Theorem 3 and an upper bound is established in Lemma 1. Then

$$\psi_n^* - \psi_0 = (P_n - P_0)D^*(d_0, Q^{d_0}, g_0) + R_{1d_n}\left(Q_n^{d_n}, Q_0^{d_n}, g_n, g_0\right) + o_{P_0}\left(n^{-\frac{1}{2}}\right), \tag{8}$$

where R_{1d} is defined in Theorem 3.

The proof of the above theorem, which is given in the appendix, makes use of the fact that the TMLE satisfies (7). We now give two sets of conditions which control the remainder term R_{1d_n} in (8) to prove the asymptotic linearity of the TMLE. The first result is an immediate consequence of the fact that $R_{1d_n}(Q_n^{d_n}, Q_0^{d_n}, g_n, g_0) = 0$ whenever $g_n = g_0$.

Corollary 1. *Suppose the conditions of Theorem 4 further suppose that $g_n = g_0$ (i.e., RCT). Then:*

$$\psi_n^* - \psi_0 = (P_n - P_0)D^*(d_0, Q^{d_0}, g_0) + o_{P_0}(n^{-1/2})$$

That is, ψ_n^* is asymptotically linear with influence curve $D^*(d_0, Q^{d_0}, g_0)$.

The next corollary is more general in that it applies to situations where the intervention mechanism g_0 is estimated from the data. The above result emerges as a special case.

Corollary 2. *Suppose all of the conditions of Theorem 4 hold, and that*

$$R_{1d_n}\left(Q_n^{d_n *}, Q_0^{d_n}, g_n, g_0\right) - R_{1d_n}\left(Q^{d_n}, Q_0^{d_n}, g_n, g_0\right) = o_{P_0}(1/\sqrt{n})$$

for some Q^{d_n}. In addition, we assume the following asymptotic linearity condition on a smooth functional of g_n:

$$R_{1d_n}\left(Q^{d_n}, Q_0^{d_n}, g_n, g_0\right) = (P_n - P_0)D_g(P_0) + o_{P_0}(1/\sqrt{n}), \tag{9}$$

for some function $D_g(P_0)(O) \in L_0^2(P_0) \equiv \{h : P_0 h = 0, P_0 h^2 < \infty\}$. Then,

$$\psi_n^* - \psi_0 = (P_n - P_0)\{D^*(d_0, Q^{d_0}, g_0) + D_g(P_0)\} + o_{P_0}(1/\sqrt{n}). \tag{10}$$

If it is also know that g_n is an MLE of g_0 according to a correctly specified model G for g_0 with tangent space $T_g(P_0)$ at P_0, then (9) holds with

$$D_g(P_0) = -\Pi\left(D^*(d_0, Q^{d_0}, g_0) | T_g(P_0)\right), \tag{11}$$

where $\Pi(\cdot | T_g(P_0))$ denotes the projection operator onto $T_g(P_0) \subset L_0^2(P_0)$ in the Hilbert space $L_0^2(P_0)$.

Equation (11) is a corollary of Theorem 2.3 of van der Laan and Robins [34]. The rest of the theorem is the result of a simple rearrangement of terms, so the proof is omitted.

Condition (9) is trivially satisfied in a randomized clinical trial without missingness, where we can take $g_n = g_0$ and thus $D_g(P_0)$ is the constant function 0. Nonetheless, (11) suggests that it would be better to estimate g_0 using a parametric model that contains the true (known) intervention mechanism. For example, at each time point one may use a main terms linear logistic regression with treatment and covariate histories as predictors. If $Q_n^{d_n}$ consistently estimates $Q_0^{d_0}$, then $D^*(d_0, Q^{d_0}, g_0)$ is orthogonal to $T_g(P_0)$ and hence the projection in (11) is the constant function 0. Otherwise the projection will decrease the variance of $\psi_n^* - \psi_0$ without affecting asymptotic bias, thereby increasing the asymptotic efficiency of the estimator. One can then use an empirical estimate of the variance of $D^*(d_0, Q^{d_0}, g_0)$ to get asymptotically conservative confidence intervals for ψ_0.

5.1 Asymptotic linearity of TMLE in a SMART setting

Suppose the data is generated by a sequential RCT and there is no missingness so that g_0 is known. Further suppose that (5) holds at P_0, that is, that treating at each time point has either a positive or negative effect with probability 1, regardless of the choice of the regimen at earlier time points. In addition, assume that $V(0)$ and $V(1)$ are both univariate scores, and assume condition (3) so that the optimal rule $d_{0,A(1)}$ based on $(A(0), V(0), V(1))$ is the same as the optimal rule $d_{0,A(1)}$ based on $A(0), V(1)$: for example, $V(1)$ is the same score as $V(0)$ but measured at the next time point, so that it is reasonable to assume that an effect of $V(0)$ on Y will be fully blocked by $V(1)$. Suppose we want to use the data of the RCT to learn the V-optimal rule d_0 and provide statistical inference for $E_{P_0} Y_{d_0}$. Further suppose that the moment conditions in Lemma 1 hold with $\beta_1 = \beta_2 = 2$. Since both $V(0)$ and $V(1)$ are one-dimensional, using kernel smoothers or sieve-based estimation to generate a library of candidate estimators for the sequential loss-based super-learner of the blip functions $(\bar{Q}_{10}, \bar{Q}_{20})$ described in our companion paper, we can obtain an estimator $\bar{Q}_n = (\bar{Q}_{1n}, \bar{Q}_{2n})$ of $\bar{Q}_0 = (\bar{Q}_{10}, \bar{Q}_{20})$ that converges in L_2 at a rate such as $n^{-2/5}$ under the assumption that $\bar{Q}_{10}, \bar{Q}_{20}$ are continuously differentiable with a uniformly bounded derivative, or at a better rate under additional smoothness assumptions. As a consequence, in this case $R_2(Q_n, Q_0) = O_{P_0}(n^{-3/5}) = o_{P_0}(n^{-1/2})$ by Lemma 1. As a consequence, all conditions of Theorem 4 hold, and it follows that the proposed TMLE is asymptotically linear with influence curve $D^*(d_0, Q^{d_0}, g_0)$, where Q^{d_0} is the possibly misspecified limit of $Q^{d_n *}$ in the TMLE. To conclude, sequential RCTs allow us to learn V-optimal rules at adaptive optimal rates of convergence, and allow valid asymptotic statistical inference for $E_{P_0} Y_{d_0}$. If $V(j)$ is higher dimensional, then one will have to rely on enough smoothness assumptions on the blip functions and/or moment conditions on $1/|\bar{Q}_{10}|$ and $1/|\bar{Q}_{20}|$ from Lemma 1 in order to guarantee that $R_2(Q_n, Q_0) = o_{P_0}(1/\sqrt{n})$.

If there is right censoring, then $g_0 = g_{01} g_{02}$ factors in a treatment mechanism g_{01} and censoring mechanism g_{02}, where g_{01} is known, but g_{02} is typically not known. Having a lot of knowledge about how censoring depends on the observed past might make it possible to obtain a good estimator of g_{02}. In that case, the above conclusions still apply, but one now estimates the nuisance parameters of the loss function (e.g., one uses a double robust loss function in which g_{02} is replaced by an estimator, see our companion paper).

5.2 Statistical inference

Suppose one wishes to estimate the mean outcome under the optimal rule $E_{P_0} Y_{d_0}$ and that (5) holds. Above we developed the TMLE ψ_n^* for $E_{P_0} Y_{d_0}$. By Corollary 1, if $g_n = g_0$ is known, this TMLE of ψ_0 is asymptotically linear with influence curve $IC(P_0) = D^*(d_0, Q^{d_0}, g_0)$. If g_n is an MLE according to a model with tangent space $T_g(P_0)$, then the TMLE is asymptotically linear with influence curve

$$IC(P_0) - \Pi\big(IC(P_0) | T_g(P_0)\big),$$

so that one could use $IC(P_0)$ as a conservative influence curve. Let IC_n be an estimator of this influence curve $IC(P_0)$ obtained by plugging in the available estimates of its unknown components. The asymptotic variance of the TMLE ψ_n^* of ψ_0 can now be (conservatively) estimated with

$$\sigma_n^2 = \frac{1}{n}\sum_{i=1}^{n} IC_n^2(O_i).$$

An asymptotic 95% confidence interval for ψ_0 is given by $\psi_n^* \pm 1.96\sigma_n/\sqrt{n}$.

6 Statistical inference for mean outcome under data adaptively determined dynamic treatment

Let $\hat{d}\colon \mathcal{M} \to \mathcal{D}$ be an estimator that maps an empirical distribution into an individualized treatment rule. See our companion paper for examples of possible estimators $\hat{d}$. Let $d_n = \hat{d}(P_n)$ be the estimated rule. Up until now we have been concerned with statistical inference for $E_{P_0}Y_{d_0}$, where d_0 is the unknown V-optimal rule while d_n is a best estimator of this rule. As a consequence, statistical inference for $E_{P_0}Y_{d_0}$ based on the TMLE relied on consistency of d_n to d_0, but also relied on the rate of convergence at which d_n converges to d_0, that is, $R_2(Q_n, Q_0) = o_{P_0}(1/\sqrt{n})$. In this section we present statistical inference for the data adaptive target parameter

$$\psi_{0n} = \Psi_{d_n}(P_0) = E_{P_0}Y_d|_{d=d_n}.$$

That is, we construct an estimator $\hat{\Psi}_{\hat{d}(P_n)}(P_n)$ of $\Psi_{\hat{d}(P_n)}(P_0)$ and a confidence interval so that

$$\lim_{n\to\infty} \Pr_{P_0}\left(\Psi_{\hat{d}(P_n)}(P_0) \in \hat{\Psi}_{\hat{d}(P_n)}(P_n) \pm 1.96\hat{\sigma}(P_n)/\sqrt{n}\right) = 0.95,$$

where $\hat{\sigma}(P_n)$ is a consistent estimator of the standard error of $\hat{\Psi}_{\hat{d}(P_n)}(P_n)$. Note that in this definition of the confidence interval the target parameter is itself also a random variable through the data P_n.

We do not assume that (5) holds in this section, but we do implicitly make the weaker assumption that $d_n \to d_1$ for some $d_1 \in \mathcal{D}$ in assumption (12) of Theorem 5. Statistical inference will be based on the same TMLE of $\Psi_d(P_0)$ at $d = d_n$, and our variance estimator will also be the same, but since the target is not $\Psi_{d_0}(P_0)$ but $\Psi_{d_n}(P_0)$, there will be no need for d_n to even be consistent for d_0, let alone converge at a particular rate. As a consequence, this approach is particularly appropriate in cases where V is high dimensional so that it is not reasonable to expect that d_n converges to d_0 at the required rate. Another motivation for this data adaptive target parameter is that, even when statistical inference for $E_{P_0}Y_{d_0}$ is feasible, one might be interested in statistical inference for the mean outcome under the concretely available rule d_n instead of under the unknown rule d_0.

As shown in the proof of Theorem 3, $P_0D^*(d_n, Q_n^*, g_n) = \psi_{0n} - \psi_n^* + R_{1d_n}(Q_n^{d_n*}, Q_0^{d_n}, g_n, g_0)$. Further, $P_nD^*(d_n, Q_n^{d_n*}, g_n) = 0$, which yields

$$\psi_n^* - \psi_{0n} = (P_n - P_0)D^*\left(d_n, Q_n^{d_n*}, g_n\right) + R_{1d_n}\left(Q_n^{d_n*}, Q_0^{d_n}, g_n, g_0\right).$$

This relation is key to the proof of the following theorem, which is analogous to Theorem 4. Note crucially that the theorem does not have any conditions on the remainder term R_2, nor does it require that d_n converge to the optimal rule d_0.

Theorem 5. *Assume $Y \in [0,1]$. Let $\hat{d}(P_n) \in \mathcal{D}$ with probability tending to 1, and assume the strong positivity assumption. Let $\psi_{0n} = \Psi_{d_n}(P_0) = E_{P_0}Y_d|_{d=d_n}$ be the data adaptive target parameter of interest. Let R_{1d} be as defined in Theorem 3.*

Assume $D_n^* \equiv D^*(d_n, Q_n^*, g_n)$ falls in a P_0 -Donsker class with probability tending to 1,

$$P_0\{D_n^* - D^*(d_1, Q^{d_1}, g_0)\}^2 = o_{P_0}(1) \tag{12}$$

for some $d_1 \in \mathcal{D}$ and Q^{d_1}. Then,

$$\psi_n^* - \psi_{0n} = (P_n - P_0)D^*(d_1, Q^{d_1}, g_0) + R_{1d_n}\left(Q_n^{d_n*}, Q_0^{d_n}, g_n, g_0\right) + o_{P_0}(n^{-1/2}).$$

If $g_n = g_0$ (i.e., RCT), then $R_{1d_n}(Q_n^{d_n*}, Q_0^{d_n}, g_n, g_0) = 0$, so that ψ_n^* is asymptotically linear with influence curve $D^*(d_1, Q, g_0)$.

The proof of the above theorem is nearly identical to the proof of Theorem 4 so is omitted. For general g_n, $R_{1d_n}(Q_n^{d_n*}, Q_0^{d_n}, g_n, g_0) = o_{P_0}(n^{-1/2})$ under an analogous second-order term condition to the one assumed in Corollary 1. As in Corollary 2, the asymptotic efficiency may improve (and will not worsen) when a known intervention mechanism is fit using a correctly specified parametric model. See Theorem 11 in our online technical report for details [47].

7 Statistical inference for the average of sample-split specific mean counterfactual outcomes under data adaptively determined dynamic treatments

Again let $\hat{d} : \mathcal{M} \to \mathcal{D}$ be an estimator that maps an empirical distribution into an individualized treatment rule. Let $B_n \in \{0,1\}^n$ denote a random vector for a cross-validation split, and for a split B_n, let P_{n,B_n}^0 be the empirical distribution of the training sample $\{i : B_n(i) = 0\}$ and P_{n,B_n}^1 is the empirical distribution of the validation sample $\{i : B_n(i) = 1\}$. Consider a J-fold cross-validation scheme. In J-fold cross-validation, the data is split into J mutually exclusive and exhaustive sets of size approximately n/J uniformly at random. Each set is then used as the validation set once, with the union of all other sets serving as the training set. With probability $1/J$, B_n has value 1 in all indices in validation set $j \in \{1, ..., J\}$ and 0 for all indices not corresponding to training set j.

In this section, we present a method that provides an estimator and statistical inference for the data adaptive target parameter

$$\tilde{\psi}_{0n} = E_{B_n}\Psi_{\hat{d}(P_{n,B_n}^0)}(P_0).$$

Note that $\tilde{\psi}_{0n}$ is different from the data adaptive target parameter ψ_{0n} presented in the previous section. In particular, this target parameter is defined as the average of data adaptive parameters, where the data adaptive parameters are learned from the training samples of size approximately n/J. In the previous section, the data adaptive target parameter was defined as the mean outcome under the rule d_n which was estimated on the entire data set. Again the target parameter is a random quantity that relies on the sample of size n.

One applies the estimator $\hat{d}$ to each of the J training samples, giving a target parameter value $\Psi_{\hat{d}(P_{n,B_n}^0)}(P_0)$, and our target parameter $\tilde{\psi}_{0n}$ is defined as the average across these J target parameters. Below we present a CV-TMLE $\tilde{\psi}_n^*$ of this data adaptive target parameter $\tilde{\psi}_{0n}$. As in the previous section, we will be able to establish statistical inference for our estimate $\tilde{\psi}_n^*$ without requiring that the estimated rules converge to d_0, nor any rate condition on the estimated rules. Unlike the asymptotic linearity results in all previous sections, the results in this section do not rely on an empirical process condition (i.e., Donsker class condition). That means we obtain valid asymptotic statistical inference under essentially no conditions in a sequential RCT, even when d_n is a highly data adaptive estimator of a V-optimal rule for a possibly high dimensional V. Under a consistency and rate condition (but no empirical process condition) on d_n, we also get inference for $E_{P_0}Y_{d_0}$.

The next subsection defines the general CV-TMLE for data adaptive target parameters. We subsequently present an asymptotic linearity theorem allowing us to construct asymptotic 95% confidence intervals.

7.1 General description of CV-TMLE

Here we give a general overview of the CV-TMLE procedure. In "CV-TMLE of the mean outcome under data adaptive V-optimal rule" in Appendix B we present a particular CV-TMLE which satisfies all of the properties described in this section. Denote the realizations of B_n with $j = 1, \ldots, J$, and let $d_{nj} = \hat{d}(P^0_{n,j})$ for some estimator of the optimal rule $\hat{d}$. Let

$$\left(a(0), \bar{l}(1)\right) \mapsto E_{nj}\left[Y|\bar{A}(1) = d_{nj}(a(0), v), \bar{L}(1) = \bar{l}(1)\right]$$

represent an initial estimate of $E_{P_0}[Y|\bar{A}(1) = d_{nj}(A(0), V), \bar{L}(1)]$ based on the training sample j. Similarly, let $l(0) \mapsto E_{nj}[Y_{d_{nj}}|L(0) = l(0)]$ represent an initial estimate of $E_{P_0}[Y_{d_{nj}}|L(0)]$ based on the training sample j. Finally, let $Q_{L(0),nj}$ represent the empirical distribution of $L(0)$ in validation sample j. We then fluctuate these three regression functions using the following submodels:

$$\left\{E^{(\varepsilon_2)}_{nj}[Y|\bar{A}(1) = d_{nj}(a(0), v), \bar{L}(1) = \bar{l}(1)] : \varepsilon_2 \in \mathbb{R}\right\}$$

$$\left\{E^{(\varepsilon_1)}_{nj}[Y_{d_{nj}}|L(0) = l(0)] : \varepsilon_1 \in \mathbb{R}\right\}$$

$$\left\{Q^{(\varepsilon_0)}_{L(0),nj} : \varepsilon_0 \in \mathbb{R}\right\},$$

where these submodels rely on an estimate g_{nj} of g_0 based on training sample j and are such that:

$$E^{(0)}_{nj}\left[Y|\bar{A}(1) = d_{nj}(a(0), v), \bar{L}(1)\right] = E_{nj}\left[Y|\bar{A}(1) = d_{nj}(a(0), v), \bar{L}(1)\right]$$

$$E^{(0)}_{nj}\left[Y_{d_{nj}}|L(0)\right] = E_{nj}\left[Y_{d_{nj}}|L(0)\right]$$

$$Q^{(0)}_{L(0),nj} = Q_{L(0),nj}.$$

Let $Q^{d_{nj}}_{nj}(\varepsilon)$ represent the parameter mapping that gives the three regression functions above fluctuated by $\varepsilon \equiv (\varepsilon_0, \varepsilon_1, \varepsilon_2)$. For a fixed ε, $Q^{d_{nj}}_{nj}(\varepsilon)$ only relies on P^1_{nj} through the empirical distribution of $L(0)$ in validation sample j. Let ϕ be a valid loss function for Q^d_0 so that $Q^d_0 = \arg\min_{Q^d} P_0\phi(Q^d)$, and let ϕ and the submodels above satisfy

$$D^*(d, Q^d, g) \in \left\langle \frac{d}{d\varepsilon}\phi(Q^d(\varepsilon))\Big|_{\varepsilon=0} \right\rangle,$$

where $\langle f \rangle = \{\sum_j \beta_j f_j : \beta\}$ denotes the linear space spanned by the components of f. We choose ε_n to minimize $P^1_n\phi(Q^{d_{nj}}_{nj}(\varepsilon))$ over $\varepsilon \in \mathbb{R}^3$. We then define the targeted estimate $Q^{d_{nj}*}_{nj} \equiv Q^{d_{nj}}_{nj}(\varepsilon_n)$ of $Q^{d_{nj}}_0$. We note that $Q^{d_{nj}*}_{nj}$ maintains the rate of convergence of Q_{nj} under mild conditions that are standard to M-estimator analysis. The key property that we need from the ε_n and the corresponding update $Q^{d_{nj}*}_{nj}$ is that it (approximately) solves the cross-validated empirical mean of the efficient influence curve:

$$E_{B_n}P^1_{n,B_n}D^*\left(d_{nj}, Q^{d_{nj}*}_{nj}, g_{nj}\right) = o_{P_0}(1/\sqrt{n}). \tag{13}$$

The CV-TMLE implementation presented in the appendix satisfies this equation with $o_{P_0}(1/\sqrt{n})$ replaced by 0. The proposed estimator of $\tilde{\psi}_{0n}$ is given by

$$\tilde{\psi}^*_n \equiv E_{B_n}\Psi_{d_{nj}}\left(Q^{d_{nj}*}_{nj}\right).$$

In the current literature we have referred to this estimator as the CV-TMLE [53–56]. We give a concrete CV-TMLE algorithm for $\tilde{\psi}_n^*$ in "CV-TMLE of the mean outcome under data adaptive V-optimal rule" in Appendix B, but note that other CV-TMLE algorithms can be derived using the approach in this section for different choices of loss function ϕ and submodels.

7.2 Statistical inference based on the CV-TMLE

We now proceed with the analysis of this CV-TMLE $\tilde{\psi}_n^*$ of $\tilde{\psi}_{0n}$. We first give a representation theorem for the CV-TMLE that is analogous to Theorem 5.

Theorem 6. *Let g_{nj} and d_{nj} represent estimates of g_0 and d_0 based on training sample j. Let $Q_{nj}^{d_{nj}*}$ represent a targeted estimate of $Q_0^{d_{nj}}$ as presented in Section 7.1 so that $Q_{nj}^{d_{nj}*}$ satisfies (13). Let R_{1d} be as in Theorem 3. Further suppose that the supremum norm of $\max_j D^*(d_{nj}, Q_{nj}^{d_{nj}*}, g_{nj})$ is bounded by some $M < \infty$ with probability tending to 1, and that*

$$\max_{j\in\{1,\ldots,J\}} P_0\left\{D^*\left(d_{nj}, Q_{nj}^{d_{nj}*}, g_{nj}\right) - D^*\left(d_1, Q^{d_1}, g\right)\right\}^2 \to 0 \textit{ in probability}$$

for some $d_1 \in \mathcal{D}$ and possibly misspecified Q^{d_1} and g. *Then:*

$$\begin{aligned}\tilde{\psi}_n^* - \tilde{\psi}_{0n} = {} & (P_n - P_0)D^*\left(d_1, Q^{d_1}, g^{d_1}\right) \\ & + \frac{1}{J}\sum_{j=1}^{J} R_{1d_{nj}}\left(Q_{nj}^{d_{nj}*}, Q_0^{d_{nj}}, g_{nj}, g_0\right) + o_{P_0}(n^{-1/2}).\end{aligned}$$

Note that d_1 in the above theorem need not be the same as the optimal rule d_0, though later we will discuss the desirable special case where $d_1 = d_0$. The above theorem also does not require that g_0 is known, or even that the limit of our intervention mechanisms g is equal to g_0. Nonetheless, we get the following asymptotic linearity result when $g = g_0$ and g_{nj} satisfies an asymptotic linearity condition on a smooth functional of g_{nj}.

Corollary 3. *Suppose the conditions from Theorem 6 hold with $g = g_0$. Further suppose that*:

$$\frac{1}{J}\sum_{j=1}^{J}\left(R_{1d_{nj}}\left(Q_{nj}^{d_{nj}*}, Q_0^{d_{nj}}, g_{nj}, g_0\right) - R_{1d_{nj}}\left(Q^{d_{nj}}, Q_0^{d_{nj}}, g_{nj}, g_0\right)\right) = o_{P_0}(n^{-1/2}),$$

for some $Q^{d_{nj}}$ and that:

$$\frac{1}{J}\sum_{j=1}^{J} R_{1d_{nj}}\left(Q^{d_{nj}*}, Q_0^{d_{nj}}, g_{nj}, g_0\right) = (P_n - P_0)D_g(P_0) + o_{P_0}(n^{-1/2}). \tag{14}$$

We can conclude that:

$$\tilde{\psi}_n^* - \tilde{\psi}_{0n} = (P_n - P_0)\left(D^*(d_1, Q^{d_1}, g_0) + D_g(P_0)\right) + o_{P_0}(n^{-1/2}).$$

The proof of the above result is just a rearrangement of terms so is omitted. Consider our setting. Suppose g_0 is known so we can have that $g_{nj} = g_0$ for all j. Consider the estimator

$$\sigma_n^2 = \frac{1}{J}\sum_{j=1}^{J} P_{n,j}^1\left\{D^*\left(d_{nj}, Q_{nj}^{d_{nj}*}, g_{nj}\right)\right\}^2$$

of the asymptotic variance $\sigma_0^2 = P_0\{D^*(d_1, Q^{d_1}, g_0)\}^2$ of the CV-TMLE $\tilde{\psi}_n^*$. An asymptotic 95% confidence interval for $\tilde{\psi}_{0n}$ is given by $\tilde{\psi}_n^* \pm 1.95\sigma_n/\sqrt{n}$. This same variance estimator and confidence interval can be used for the case that g_0 is not known and each g_{nj} is an MLE of g_0 according to some model. In that case, it is an asymptotically conservative confidence interval (analogous to eq. (11) applied to Corollary 3).

Now consider the case where d_1 from the above theorem is equal to the optimal rule d_0 and condition (5) holds. For simplicity, also assume that g_0 is known and $g_{nj} = g_0$. Then $R_{1d_{nj}}$ is equal to 0 for all j, so Theorem 6 shows that the CV-TMLE for $\tilde{\psi}_{0n}$ is asymptotically linear with influence curve $D^*(d_1, Q^{d_1}, g_0) = D^*(d_0, Q^{d_0}, g_0)$. If

$$\tilde{\psi}_{0n} - \psi_0 = \frac{1}{J}\sum_{j=1}^{J} R_2(Q_{nj}, Q_0)$$

is second order, that is, $o_{P_0}(n^{-1/2})$, where Q_{nj} is analogous to Q_n but only estimated on the training sample j, then the CV-TMLE is consistent and asymptotically normal estimator of the mean outcome under the optimal rule. If $Q^{d_0} = Q_0^{d_0}$, then the CV-TMLE is also asymptotically efficient among all regular asymptotically linear estimators. One can apply bounds like those in Lemma 1 for each of the J terms above to understand the behavior of $\tilde{\psi}_{0n} - \psi_0$. Note crucially that this result does not rely on the restrictive empirical process conditions used in the previous sections, although it relies on a consistency and rate condition for asymptotic linearity with respect to the non-data adaptive parameter $E_{P_0} Y_{d_0}$.

8 Simulation methods

We start by presenting two single time point simulations. In earlier technical reports we directly describe the single time point problem [47, 48]. Here, we instead note that a single time point optimal treatment is a special case of a two time point treatment when only the second treatment is of interest. In particular, we can see this by taking $L(0) = V(0) = \emptyset$, estimating $\bar{Q}_{2,0}$ without any dependence on $a(0)$, and correctly estimating $\bar{Q}_{1,0}$ with the constant function zero. We note that, in this one time point formulation, we do not need (5) to hold for $\bar{Q}_{10}$, so it may be more natural to view the single time point problem directly and use the single time point pathwise differentiability result in Theorem 2 of van der Laan and Luedtke [48]. We can then let $I(A(0) = d_{n,A(0)}(V(0))) = 1$ for all $A(0), V(0)$ wherever the indicator appears in our calculations. Because the first time point is not of interest, we only describe the second time point treatment mechanism for this simulation. We refer the interested reader to the earlier technical report for a thorough discussion of the single time point case. We then present a two time point data generating distribution to show the effectiveness of our proposed method in the longitudinal setting.

8.1 Data

8.1.1 Single time point

We simulate 1,000 data sets of 1,000 observations from an RCT without missingness. We have that:

$$L_1(1), L_2(1), L_3(1), L_4(1) | A(0) \overset{iid}{\sim} N(0,1)$$

$$A_1(1) | A(0) \sim \text{Bern}(1/2)$$

$$A_2(1) | A_1(1), A(0) \sim \text{Bern}(1)$$

$$\begin{aligned} &\text{logit}\, E_{P_0}\left[Y | \bar{A}(1), \bar{L}(1), H = 0\right] \\ &\quad = 1 - L_1(1)^2 + 3L_2(1) + A_1(1)\left(5L_3(1)^2 - 4.45\right) \end{aligned}$$

$$\begin{aligned}&\text{logit } E_{P_0}\left[Y|\bar{A}(1),\bar{L}(1),H=1\right]\\&\quad=-0.5-L_3(1)+2L_1(1)L_2(1)+A_1(1)(3|L_2(1)|-1.5)\end{aligned}$$

where Y is a Bernoulli random variable and H is an unobserved Bern(1/2) variable independent of $\bar{A}(1),\bar{L}(1)$. The above distribution was selected so that the mean outcomes under static treatments (treating everyone or no one at the second time point) have approximately the same mean outcome of 0.464.

We consider two choices for $V(1)$. For the first we consider $V(1)=L_3(1)$, and for the second we consider $V(1)$ to be the entire covariate history $\bar{L}(1)$. We have shown via Monte Carlo simulation that the optimal rule has mean outcome $E_{P_0}Y_{d_0}\approx 0.536$ when $V(1)=L_3(1)$ and the optimal rule has mean outcome $E_{P_0}Y_{d_0}\approx 0.563$ when $V(1)=(L_1(1),L_2(1),L_3(1),L_4(1))$. One can verify that the blip function at the second time point is nonzero with probability 1 for both choices of $V(1)$.

8.1.2 Two time point

We again simulate 1,000 data sets of 1,000 observations from an RCT without missingness. The observed variables have the following distribution:

$$L_1(0),L_2(0)\overset{iid}{\sim}\text{Unif}(-1,1)$$

$$A_1(0)|L(0)\sim\text{Bern}(1/2)$$

$$A_2(0)|A_1(0),L(0)\sim\text{Bern}(1)$$

$$U_1,U_2|A(0),L(0)\overset{iid}{\sim}\text{Unif}(-1,1)$$

$$L_1(1)|A(0),L(0),U_1,U_2\sim U_1(1.25A_1(0)+0.25)$$

$$L_2(1)|A(0),L(0),L_1(1),U_1,U_2\sim U_2(1.25A_1(0)+0.25)$$

$$A_1(1)|A(0),\bar{L}(1)\sim\text{Bern}(1/2)$$

$$A_2(1)|A(0),A_1(1),\bar{L}(1)\sim\text{Bern}(1)$$

$$Y|\bar{A}(1),\bar{L}(1)\sim\text{Bern}\big(0.4+0.069\,b(\bar{A}(1),\bar{L}(1))\big),$$

where

$$\begin{aligned}b(\bar{A}(1),\bar{L}(1))\equiv\ &0.5A_1(0)\Big(-0.8-3(\text{sgn}(L_1(0))+L_1(0))-L_2(0)^2\Big)\\&+A_1(1)\Big(-0.35+(L_1(1)-0.5)^2\Big)+0.08A_1(0)A_1(1).\end{aligned}$$

Note that $E_{P_0}\left[Y|\bar{A}(1),\bar{L}(1)\right]$ is contained in the unit interval by the bounds on $\bar{A}(1)$ and $\bar{L}(1)$ so that Y is indeed a valid Bernoulli random variable. We will let $V(0)=L(0)$ and $V(1)=(A(0),\bar{L}(1))$. One can verify that (5) is satisfied for this choice of V.

Static treatments yield mean outcomes $E_{P_0}Y_{(0,1),(0,1)}=0.400$, $E_{P_0}Y_{(0,1),(1,1)}\approx 0.395$, $E_{P_0}Y_{(1,1),(0,1)}\approx 0.361$, and $E_{P_0}Y_{(1,1),(1,1)}\approx 0.411$. The true optimal treatment has mean outcome $E_{P_0}Y_{d_0}\approx 0.485$.

8.2 Optimal rule estimation methods

For now suppose we have estimators of the optimal rule with reasonable convergence properties, by which we mean that the true mean outcome under the fitted rule is close to the mean outcome under the optimal rule. In our companion paper in this volume we describe these estimators and show precisely how close these estimators come to achieving the optimal mean outcome. Here we note that our estimation algorithms correspond to using the full candidate library of weighted classification and blip function-based estimators proposed in table 2 of our companion paper, with the weighted log loss function used to determine the convex combination of candidates. We provide oracle inequalities for this estimator in our companion paper, and argue that it represents a powerful approach to data adaptively estimate the optimal rule without over- or underfitting the data. For a sample size n, we denote the rule estimated on the whole sample by d_n, and the rule estimated on training sample j by d_{nj}.

8.3 Inference procedures

We use four procedures to estimate the mean outcome under the fitted rule. All inference procedures rely on the intervention mechanism g_0. We always estimate the intervention mechanism with the true mechanism g_0, as one may do in an RCT without missingness. We do not consider efficiency gains resulting from estimating the known treatment mechanism here.

The first method uses the TMLE described in "TMLE of the mean outcome under a given rule" in Appendix B. The second method uses the analogous estimating equation approach that uses the double robust inverse probability of censoring weighted (DR-IPCW) estimating equation implied by $D^*(d_n, Q_n^{d_n}, g_0)$, where $Q_n^{d_n}$ represents the unfluctuated initial estimates of $Q_0^{d_n}$. See van der Laan and Robins [34] for a general outline of such an estimating equation approach. This approach is valid whenever the TMLE is valid. We also use the CV-TMLE described in "CV-TMLE of the mean outcome under data adaptive V-optimal rule" in Appendix B, where we use a 10-fold cross-validation scheme. Finally, we use the CV-DR-IPCW cross-validated estimating equation implied by $\sum_j P_{n,j}^1 D^*(d_{nj}, Q_{nj}^{d_{nj}}, g_0)$, where $Q_{nj}^{d_{nj}}$ represents the unfluctuated initial estimates of $Q_0^{d_{nj}}$. This approach is valid whenever the CV-TMLE is valid.

All inference procedures also rely on an estimate of Q_0^d for some estimated d. For the two time point case, we use the empirical distribution of $L(0)$ to estimate the marginal distribution of $L(0)$. We compare plugging in both of the true values of $E_{P_0}\left[Y|\bar{A}(1) = d(A(0), V), \bar{L}(1)\right]$ and $E_{P_0}\left[Y_d|L(0), A(0) = d_{A(0)}(V(0))\right]$ as initial estimates with plugging in the incorrectly specified constant function 1/2 as initial estimates.

For the single time point case, we compare plugging in the true value of $E_{P_0}\left[Y|\bar{A}(1) = d(A(0), V), \bar{L}(1)\right]$ with the incorrectly specified constant function 1/2. We always estimate $E_{P_0}\left[Y_d|L(0), A(0) = d_{A(0)}(V(0))\right]$ by averaging

$$(A(0), \bar{L}(1)) \mapsto E_{P_0}\left[Y|\bar{A}(1) = d(A(0), V), \bar{L}(1)\right]$$

over the empirical distribution of $L(1)$ from the entire sample for non-cross-validated methods, and from the training sample for cross-validated methods. The empirical distribution of $L(0)$ will not play a role for the single time point case because $L(0) = \emptyset$.

The procedures used to estimate the optimal rule rely on similar means, and we supply these estimation procedures with the incorrect value 1/2 for these conditional means whenever we supply the inference procedures with the incorrect values of the corresponding conditional means, and with the correct values of the conditional means whenever we supply the inference procedures with the corresponding correct values.

The simulation was implemented in R [57]. The code used to run the simulations is available upon request. We are currently looking to implement the methods in this paper and the companion paper in an R package.

8.4 Evaluating performance

We use the coverage of asymptotic 95% confidence intervals to evaluate the performance of the various methods. As we establish in the earlier parts of this paper, each inference approach yields two interesting target parameters with respect to which we can compute coverage. All approaches give asymptotically valid inference for the mean outcome under the optimal rule under conditions, and thus the coverage with respect to this parameter is assessed across all methods.

The TMLE and DR-IPCW estimating equation-based approaches also estimate the data adaptive target parameter ψ_{0n} as presented in Section 6. Given a fitted rule d_n, we approximate the expected value in this parameter definition using 10^6 Monte Carlo simulations for the single time point case and 5×10^5 Monte Carlo simulations for the two time point case. We then assess confidence interval coverage with respect to this approximation.

The CV-TMLE and cross-validated DR-IPCW estimating equation approaches estimate the data adaptive target parameter $\tilde{\psi}_{0n}$ as presented in Section 7. Given the ten rules estimated on each of the training sets, the expectation over the sample split random variable B_n becomes an average over ten target parameters, one for each estimated rule. Again we estimate the expected value of P_0 using 10^6 Monte Carlo simulations for each of the ten target parameters in the single time point case, and 5×10^5 Monte Carlo simulations in the two time point case.

9 Simulation results

Figure 1 shows that the (CV-)TMLE is more efficient than the (CV-)DR-IPCW estimating equation methods in our single time point simulation, except for the cross-validated methods when $V = L_1(1), \ldots, L_4(1)$ and the regressions are misspecified. Note that the MSEs relative to $E_{P_0} Y_{d_0}$ are the typical $E_{P_0}(\psi_n - \psi_0)^2$ for an estimate ψ_n, while the MSEs relative to the data adaptive parameter are the slightly less typical $E_{P_0}(\psi_n - \psi_{0n})^2$ for the TMLE and DR-IPCW, and $E_{P_0}(\psi_n - \tilde{\psi}_{0n})^2$ for the cross-validated methods. That is, the target parameters vary for each of the 1,000 data sets considered. We also confirmed that, as is typical in missing data problems, the methods in which the conditional means were correctly specified were more efficient than the methods in which the conditional means are incorrectly specified. Figure 2 shows that the (CV-)TMLE in general has better coverage than the (CV-)DR-IPCW estimating equation approaches in our single time point simulation, with the only exception being the CV-TMLE for $E_{P_0} Y_{d_0}$ when the regressions are misspecified and $V = L_1(1), \ldots, L_4(1)$.

Figure 3a shows that the (CV-)TMLE is always more efficient than the (CV-)DR-IPCW estimating equation methods for our two time point simulation. Figure 3b shows that this increased efficiency does not come at the expense of coverage: the (CV-)TMLE always has better coverage than the (CV-) DR-IPCW estimators in our two time point simulation. In general, we see that the cross-validated methods always achieve approximately 95% coverage for the data adaptive parameter. This is to be expected because the cross-validated methods only learn the optimal rule on validation sets, and thus avoid finite sample bias when the conditional means of the outcome are averaged over the validation samples.

It may at first be surprising that the TMLE outperforms the DR-IPCW estimating equation method in a randomized clinical trial, especially given that the CV-TMLE and CV-DR-IPCW achieve similar coverage. In Appendix C we give intuition as to why this may be the case in a single time point randomized clinical trial. In short, this difference in coverage appears to occur because our proposed TMLE only fluctuates the conditional means for individuals who received the fitted treatment, thereby reducing finite sample bias

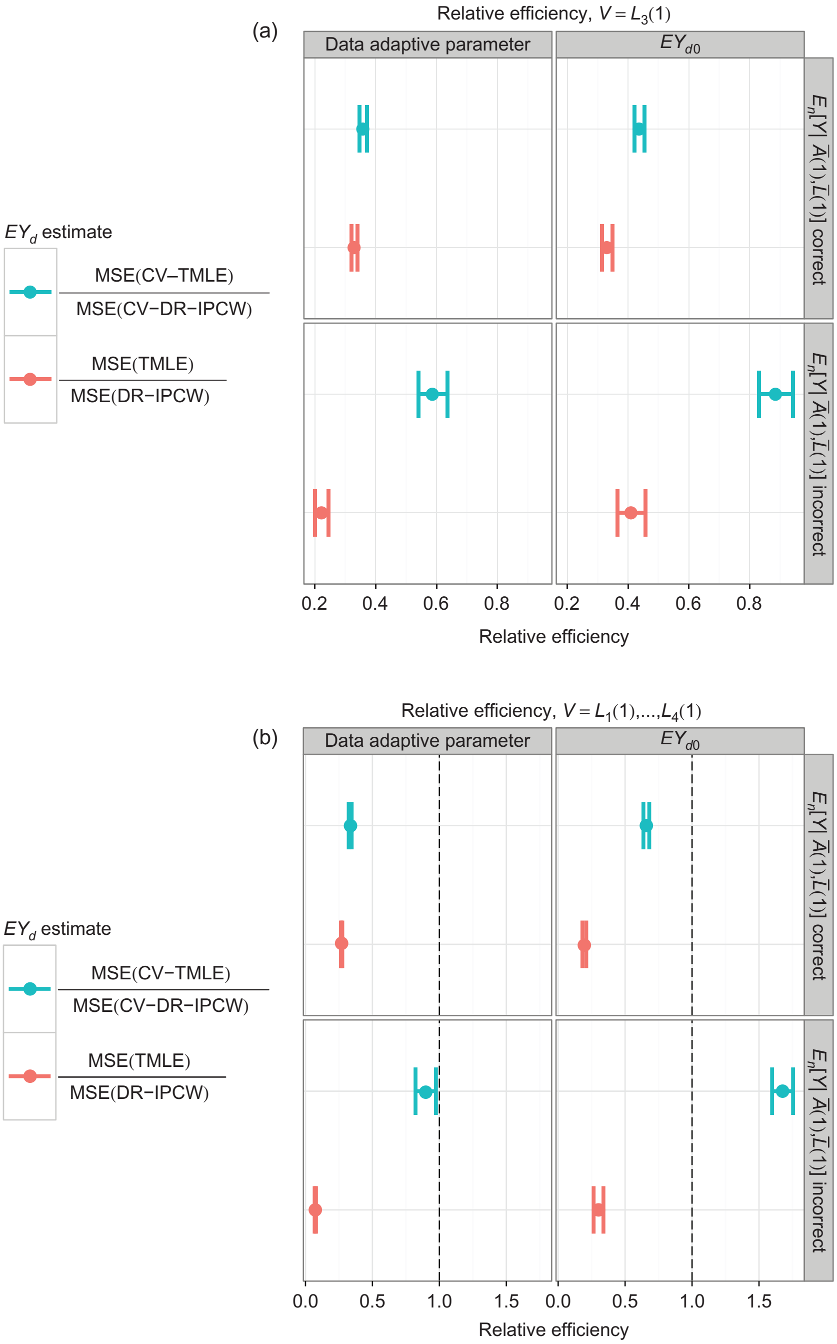

Figure 1: Relative efficiency of TMLE and DR-IPCW methods compared to both $E_{P_0}Y_{d_0}$ and the data adaptive parameter $E_{P_0}(\psi_n - \psi_{0n})^2$ for the TMLE and DR-IPCW, and $E_{P_0}(\psi_n - \tilde{\psi}_{0n})^2$ for the cross-validated methods. Results are provided both for the cases where the estimate $E_n[Y|\bar{A}(1), W]$ of $E_{P_0}[Y|\bar{A}(1), W]$ is correctly specified and the case where this estimate is incorrectly specified with the constant function 1/2. Error bars indicate 95% confidence intervals to account for uncertainty from the finite number of Monte Carlo draws in our simulation. (a) V=L_1(1), (b) V=L_1(1), … , L_4(1).

that may result from estimating the optimal rule on the same sample that is used to estimate the mean outcome under this fitted rule.

We also looked at the average confidence interval width across Monte Carlo simulations for each method and simulation setting. For a given simulation setting, all four estimation methods gave

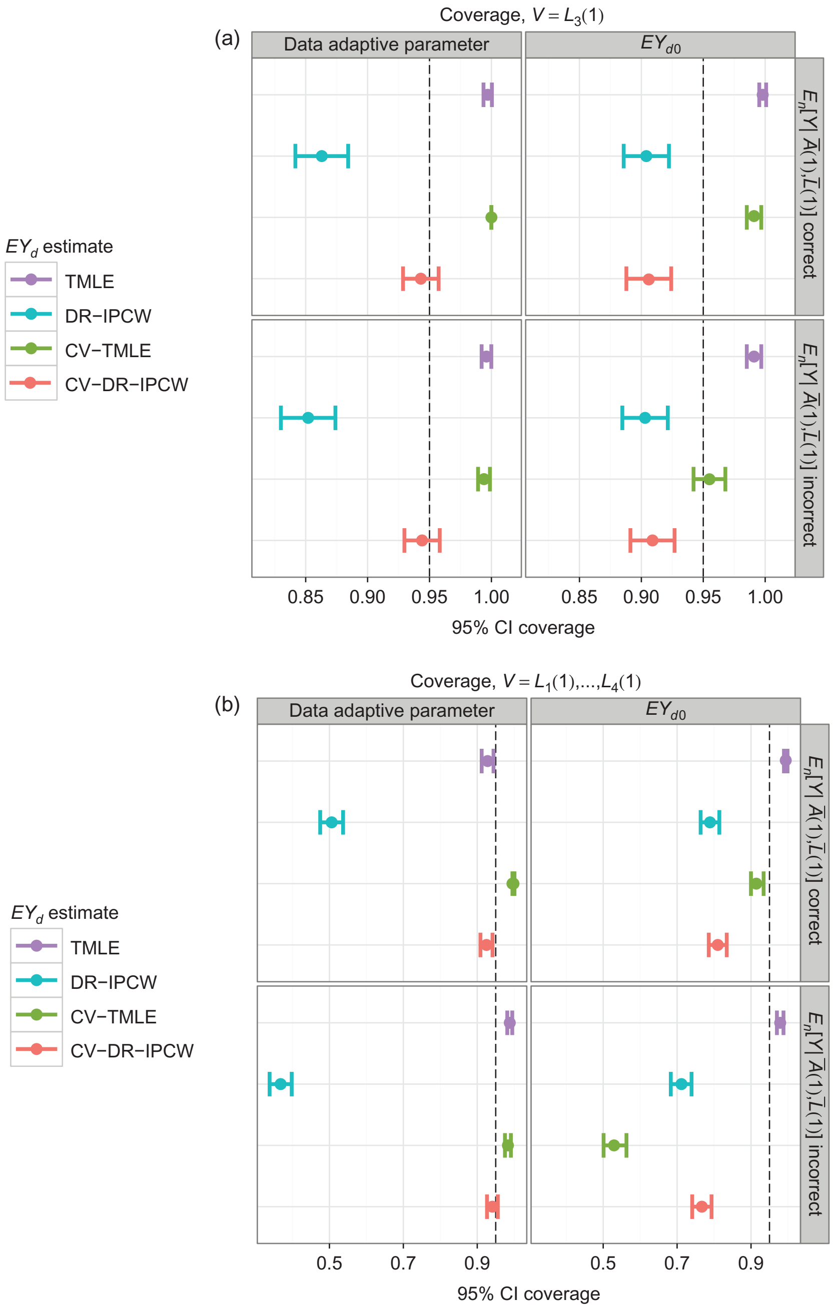

Figure 2: Coverage of 95% confidence intervals from the TMLE and DR-IPCW methods with respect to both $E_{P_0}Y_{d_0}$ and the data adaptive parameter ψ_{0n} for the TMLE and DR-IPCW and $\tilde{\psi_{0n}}$ for the cross-validated methods. Results are provided both for the cases where the estimate $E_n[Y|\bar{A}(1), W]$ of $E_{P_0}[Y|\bar{A}(1), W]$ is correctly specified and the case where this estimate is incorrectly specified with the constant function 1/2. The (CV-)TMLE outperforms the (CV-)DR-IPCW estimating equation approach for almost all settings. Error bars indicate 95% confidence intervals to account for uncertainty from the finite number of Monte Carlo draws in our simulation. (a) V=L_1(1), (b) V=L_1(1), …, L_4(1).

approximately the same (±0.002) average confidence interval width: 0.08 for both single time point simulations, 0.12 for the multiple time point simulation. These average widths show that we can get informatively small confidence intervals from our relatively small sample size of 1,000 individuals. Unlike Figures 1 and 3a, these values should not be used to gauge the efficiency of the proposed estimators since they do not take the true parameter value into account.

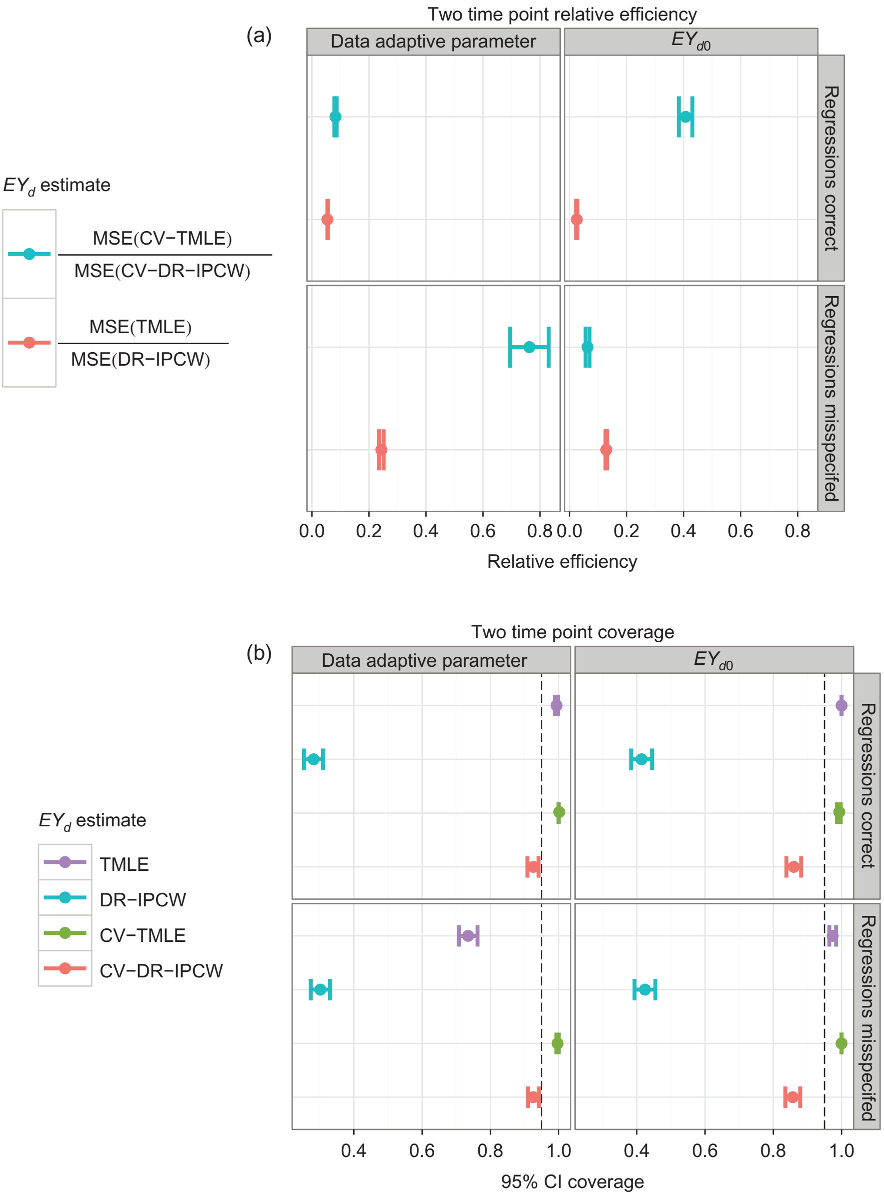

Figure 3: (a) Relative efficiency of TMLE and DR-IPCW methods compared to both $E_{P_0}Y_{d_0}$ and the data adaptive parameter $E_{P_0}(\psi_n - \psi_{0n})^2$ for the TMLE and DR-IPCW, and $E_{P_0}(\psi_n - \tilde{\psi}_{0n})^2$ for the cross-validated methods. (b) Coverage of 95% confidence intervals from the TMLE and DR-IPCW methods with respect to both $E_{P_0}Y_{d_0}$ and the data adaptive parameter ψ_{0n} for the TMLE and DR-IPCW and $\tilde{\psi_{0n}}$ for the cross-validated methods. Both (a) and (b) give results both for the cases where the estimates of $E_{P_0}[Y|\bar{A}(1) = d_n(A(0), V), \bar{L}(1)]$ and $E_{P_0}[Y_{d_n}|L(0)]$ are correctly specified and the case where these estimates are incorrectly specified with the constant function 1/2. Error bars indicate 95% confidence intervals to account for uncertainty from the finite number of Monte Carlo draws in our simulation.

10 Discussion

This article investigated semiparametric statistical inference for the mean outcome under the V-optimal rule and statistical inference for the data adaptive target parameter defined as the mean outcome under a data adaptively determined V-optimal rule (treating the latter as given).

We proved a surprising and useful result stating that the mean outcome under the V-optimal rule is represented by a statistical parameter whose pathwise derivative is identical to what it would have been if the unknown rule had been treated as known, under the condition that the data is generated by a non-exceptional law [52]. As a consequence, the efficient influence curve is immediately known, and any of the efficient estimators for the mean outcome under a given rule can be applied at the estimated rule. In particular, we demonstrate a TMLE, and present asymptotic linearity results. However, the dependence of the statistical target parameter on the unknown rule affects the second-order terms of the TMLE, and, as a consequence, the asymptotic linearity of the TMLE requires that a second-order difference between the estimated rule and the V-optimal rule converges to zero at a rate faster than $1/\sqrt{n}$. We show that this can be expected to hold for rules that are only a function of one continuous score (such as a biomarker), but when V is higher dimensional, only strong smoothness assumptions will guarantee this, so that, even in an RCT, we cannot be guaranteed valid statistical inference for such V-optimal rules.

Therefore, we proceeded to pursue statistical inference for so-called data adaptive target parameters. Specifically, we presented statistical inference for the mean outcome under the dynamic treatment regime we fitted based on the data. We showed that statistical inference for this data adaptive target parameter does not rely on the convergence rate of our estimated rule to the optimal rule, and in fact only requires that the data adaptively fitted rule converges to some (possibly suboptimal) fixed rule. However, even in a sequential RCT, the asymptotic linearity theorem still relies on an empirical process condition that limits the data adaptivity of the estimator of the rule. So, even though the assumptions are much weaker, they can still cause problems in finite samples when V is high dimensional, and possibly even asymptotically.

Therefore, we proceeded with the average of sample split specific target parameters, as in general proposed by van der Laan et al. [46], where we show that statistical inference can now avoid the empirical process condition. Specifically, our data adaptive target parameter is now defined as an average across J sample splits in training and validation sample of the mean outcome under the dynamic treatment fitted on the training sample. We presented CV-TMLE of this data adaptive target parameter, and we established an asymptotic linearity theorem that does not require that the estimated rule is consistent for the optimal rule, let alone at a particular rate. The CV-TMLE also does not require the empirical process condition. As a consequence, in a sequential RCT, this method provides valid asymptotic statistical inference without any conditions, beyond the requirement that the estimated rule converges to some (possibly suboptimal) fixed rule.

We supported our theoretical findings with simulations, both in the single and two time point settings. Our simulations supported our claim that it is easier to have good coverage of the proposed data adaptive target parameters than the mean outcome under the optimal rule, though the results for this harder mean outcome under the optimal rule parameter were also promising. In future work we hope to apply these methods to actual data sets of interest, generated by observational controlled trial as well as RCTs.

It might also be of interest to propose working models for the mean outcome $E_{P_0}[Y_{d_0}|S]$ under the optimal rule, conditional on some baseline covariates $S \subset W$. This is now a function of S, but we would define the target parameter of interest as a projection of this true underlying function on the working model. It would now be of interest to develop TMLE for this finite dimensional pathwise differentiable parameter, and we presume that similar results as we found here might appear. Such parameters provide information about how the mean outcome under the optimal rule are affected by certain baseline characteristics.

Drawing inferences concerning optimal treatment strategies is an important topic that will hopefully help guide future health policy decisions. We believe that working with a large semiparametric model is desirable because it helps to ensure that the projected health benefits from implementing an estimated treatment strategy are not due to bias from a misspecified model. The TMLEs presented in this article have many desirable statistical properties and represent one way to get estimates and make inference in this large model. We look forward to future advances in statistical inference for parameters that involve optimal dynamic treatment regimes.

Acknowledgements: This research was supported by an NIH grant R01 AI074345-06. AL was supported by the Department of Defense (DoD) through the National Defense Science & Engineering Graduate Fellowship (NDSEG) program. The authors would like to thank the anonymous reviewers and Erica Moodie for their invaluable comments and suggestions to improve the quality of the paper. The authors would also like to thank Oleg Sofrygin for valuable discussions.

Funding: National Institute of Allergy and Infectious Diseases, (Grant / Award Number: 'R01 AI074345-06')

Appendix A

Proofs

Proof of Theorem 1. Let $V_d = (V(0), V_d(1))$. For a rule in $\mathcal{D}$, we have

$$\begin{aligned}E_{P_d} Y_d &= E_{P_d} E_{P_d}(Y_d | V_d) \\ &= E_{V_d}\big(E\big(Y_{a(0),a(1)}|V_{a(0)}\big)I\big(a(1) = d_{A(1)}\big(a(0), V_{a(0)}(1)\big)\big)I\big(a(0) = d_{A(0)}(V(0))\big)\big).\end{aligned}$$

For each value of $a(0)$, $V_{a(0)} = (V(0), V_{a(0)}(1))$ and $d_{A(0)}(V(0))$, the inner conditional expectation is maximized over $d_{A(1)}(a(0), V_{a(0)}(1))$ by $d_{0,A(1)}$ as presented in the theorem, where we used that $V(1)$ includes $V(0)$. This proves that $d_{0,A(1)}$ is indeed the optimal rule for assignment of $A(1)$. Suppose now that $V(1)$ does not include $V(0)$, but the stated assumption holds. Then the optimal rule $d_{0,A(1)}$ that is restricted to be a function of $(V(0), V(1), A(0))$ is given by $I(\bar{Q}_{20}(A(0), V(0), V(1)) > 0)$, where

$$\begin{aligned}&\bar{Q}_{20}(a(0), v(0), v(1)) = \\ &E_{P_0}\big(Y_{a(0),A(1)=(1,1)} - Y_{a(0),A(1)=(0,1)}|V_{a(0)}(1) = v(1), V(0) = v(0)\big).\end{aligned}$$

However, by assumption, the latter function only depends on $(a(0), v(0), v(1))$ through $(a(0), v(1))$, and equals $\bar{Q}_{20}(a(0), v(1))$. Thus, we now still have that $d_{0,A(1)}(V) = (I(\bar{Q}_{20}(A(0), V(1)) > 0), 1)$, and, in fact, it is now also an optimal rule among the larger class of rules that are allowed to use $V(0)$ as well.

Given we found $d_{0,A(1)}$, it remains to determine the rule $d_{0,A(0)}$ that maximizes

$$\begin{aligned}&E_{V_d}\Big(E_P\Big(Y_{a(0),d_{0,A(1)}}|V_{a(0)}\Big)I\big(a(0) = d_{A(0)}(V(0))\big)\Big) \\ &= E_{P_0}E\Big(Y_{a(0),d_{0,A(1)}}|V(0)\Big)I\big(a(0) = d_{A(0)}(V(0))\big),\end{aligned}$$

where we used the iterative conditional expectation rule, taking the conditional expectation of $V_{a(0)}$, given $V(0)$. This last expression is maximized over $d_{A(0)}$ by $d_{0,A(0)}$ as presented in the theorem. This completes the proof. □

The following lemma will be useful for proving Theorem 2.

Lemma 1. *Recall the definitions of $\bar{Q}_{20}$ and $\bar{Q}_{10}$ in Theorem 1. We can represent $\Psi(P_0) = E_{P_0} Y_{d_0}$ as follows:*

$$\begin{aligned}\Psi(P_0) = &E_{P_0} Y_{(0,1),(0,1)} + E_{P_0}\big[d_{0,A(1)}\big((0,1), V_{(0,1)}(1)\big)\bar{Q}_{20}\big((0,1), V_{(0,1)}(1)\big)\big] \\ &+ E_{P_0} d_{0,A(0)}(V(0))\bar{Q}_{10}(V(0)).\end{aligned}$$

where $V_{(0,1)}(1)$ is drawn under the G-computation distribution for which treatment $(0,1)$ is given at the first time point.

Proof of Lemma A.1. For a point treatment data structure $O = (L(0), A(0), Y)$ and binary treatment $A(0)$, we have for a rule $V \to d(V)$, $E_{P_0} Y_d = E_{P_0} Y_0 + E_{P_0} d(V)\bar{Q}_0(V)$ with $\bar{Q}_0(V) = E_{P_0}[Y_1 - Y_0 | V]$. This identity is applied twice in the following derivation:

$$\begin{aligned}
\Psi(P_0) &= E_{P_0} Y_{(0,1),d_{0,A(1)}} + E_{P_0} d_{0,A(0)}(V(0))\bar{Q}_{10}(V(0)) \\
&= E_{P_0} E_{P_0}\left[Y_{(0,1),d_{0,A(1)}} | V_{(0,1)}(1)\right] + E_{P_0} d_{0,A(0)}(V(0))\bar{Q}_{10}(V(0)) \\
&= E_{P_0} E_{P_0}\left[Y_{(0,1),(0,1)} | V_{(0,1)}(1)\right] \\
&\quad + E_{P_0} I\left(\bar{Q}_{20}((0,1), V_{(0,1)}(1)\right) > 0)\bar{Q}_{20}(0, V_{(0,1)}(1)) \\
&\quad + E_{P_0} d_{0,A(0)}(V(0))\bar{Q}_{10}(V(0)) \\
&= E_{P_0} E_{P_0}\left[Y_{(0,1),(0,1)} | V_{(0,1)}(1)\right] + E_{P_0} d_{0,A(1)}((0,1), V_{(0,1)}(1))\bar{Q}_{20}(0, V_{(0,1)}(1)) \\
&\quad + E_{P_0} d_{0,A(0)}(V(0))\bar{Q}_{10}(V(0)) \\
&= E_{P_0} Y_{(0,1),(0,1)} + E_{P_0} d_{0,A(1)}((0,1), V_{(0,1)}(1))\bar{Q}_{20}(0, V_{(0,1)}(1)) \\
&\quad + E_{P_0} d_{0,A(0)}(V(0))\bar{Q}_{10}(V(0)).
\end{aligned}$$

□

Proof of Theorem 3. By the definition of R_{1d} we have

$$\begin{aligned}
P_0 D^*(Q,g) &= P_0 D^*(d_Q, Q, g) = \Psi_{d_Q}(Q_0^{d_Q}) - \Psi_{d_Q}(Q^{d_Q}) + R_{1d_Q}(Q^{d_Q}, Q_0^{d_Q}, g, g_0) \\
&= \Psi_{d_0}(Q_0^{d_0}) - \Psi_{d_Q}(Q^{d_Q}) + \{\Psi_{d_Q}(Q_0^{d_Q}) - \Psi_{d_0}(Q_0^{d_0})\} + R_{1d_Q}(Q^{d_Q}, Q_0^{d_Q}, g, g_0) \\
&= \Psi(Q_0) - \Psi(Q) + R_2(Q, Q_0) + R_{1d_Q}(Q^{d_Q}, Q_0^{d_Q}, g, g_0).
\end{aligned}$$

□

Proof of Lemma 1. Below we omit the dependence of $d_{Q,A(0)}$, $d_{0,A(0)}$, $\bar{Q}_1$, and $\bar{Q}_{10}$ on $V(0)$:

$$\begin{aligned}
R_{2A(0)} &= E_{P_0}\left[\left(d_{Q,A(0)} - d_{0,A(0)}\right)\bar{Q}_{10}\right] \\
&\leq E_{P_0}\left|\left(d_{Q,A(0)} - d_{0,A(0)}\right)\bar{Q}_{10}\right| \\
&= E_{P_0}\left|\left(d_{Q,A(0)} - d_{0,A(0)}\right)\bar{Q}_{10} I\left(|\bar{Q}_{10}| \geq |\bar{Q}_1 - \bar{Q}_{10}|\right)\right| \\
&\quad + E_{P_0}\left|\left(d_{Q,A(0)} - d_{0,A(0)}\right)\bar{Q}_{10} I\left(0 < |\bar{Q}_{10}| < |\bar{Q}_1 - \bar{Q}_{10}|\right)\right|.
\end{aligned}$$

The first term in the final equality is always 0 because $d_{Q,A(0)} = d_{0,A(0)}$ whenever the indicator is 1. In the second term, $d_{Q,A(0)} \neq d_{0,A(0)}$ whenever the indicator is 1, so:

$$\begin{aligned}
R_{2A(0)} &\leq E_{P_0}\left[|\bar{Q}_{10}| I\left(0 < |\bar{Q}_{10}| < |\bar{Q}_1 - \bar{Q}_{10}|\right)\right] \\
&\leq E_{P_0}\left[|\bar{Q}_{10}| I\left(0 < |\bar{Q}_{10}|^{\frac{p(\beta_1+1)}{p+\beta_1}} < |\bar{Q}_1 - \bar{Q}_{10}|^{\frac{p(\beta_1+1)}{p+\beta_1}}\right) I\left(|\bar{Q}_{10}| > 0\right)\right] \\
&\leq E_{P_0}\left[|\bar{Q}_1 - \bar{Q}_{10}|^{\frac{p(\beta_1+1)}{p+\beta_1}} |\bar{Q}_{10}|^{-\frac{\beta_1(p-1)}{p+\beta_1}} I\left(|\bar{Q}_{10}| > 0\right)\right] \\
&\leq \left\|\bar{Q}_1 - \bar{Q}_{10}\right\|_{p,P_0}^{\frac{p(\beta_1+1)}{p+\beta_1}} \left\|\bar{Q}_{10}^{-1} I\left(|\bar{Q}_{10}| > 0\right)\right\|_{\beta_1,P_0}^{\frac{\beta_1(p-1)}{p+\beta_1}}
\end{aligned} \tag{15}$$

where the final inequality holds by Hölder's inequality. The above also holds when the limit is taken as $p \to \infty$, yielding the essential supremum result. The result for $R_{2A(1)}$ follows by the same argument. □

Proof of Theorem 4. By Theorem 3, we have

$$P_0 D^*(d_n, Q_n^{d_n *}, g_n) = \psi_0 - \Psi_{d_n}(Q_n^{d_n *}) + R_n,$$

where $R_n = R_{1d_n}(Q_n^{d_n}, Q_0^{d_n}, g_n, g_0) + R_2(Q_n, Q_0)$. Combining this with the fact that $D_n^* \equiv D^*(d_n, Q_n^{d_n *}, g_n)$ has empirical mean 0 yields

$$\psi_n^* - \psi_0 = (P_n - P_0)D_n^* + R_n = (P_n - P_0)D^*(d_0, Q^{d_0}, g_0) + (P_n - P_0)(D_n^* - D^*(d_0, Q^{d_0}, g_0)) + R_n$$

The Donsker condition and the mean square consistency of D_n^* to $D^*(d_0, Q^{d_0}, g_0)$ give

$$(P_n - P_0)\big(D_n^* - D^*(d_0, Q^{d_0}, g_0)\big) = o_{P_0}(n^{-1/2}),$$

see, for example, van der Vaart and Wellner [58]. By assumption, $R_2(Q_n, Q_0) = o_{P_0}(n^{-1/2})$. Thus:

$$\psi_n^* - \psi_0 = (P_n - P_0)D^*\big(d_0, Q^{d_0}, g_0\big) + R_{1d_n}\Big(Q_n^{d_n}, Q_0^{d_n}, g_n, g_0\Big) + o_{P_0}(n^{-1/2})$$

as desired. □

Proof of Theorem 6. For all $j = 1, \ldots, J$, we have that:

$$\begin{aligned}\Psi_{d_{nj}}(Q_{nj}^{d_{nj}*}) - \Psi_{d_{nj}}(Q_0^{d_{nj}*}) = &- P_0 D^*\Big(d_{nj}, Q_{nj}^{d_{nj}*}, g_{nj}\Big)\\ &+ R_{1d_{nj}}\Big(Q_{nj}^{d_{nj}*}, Q_0^{d_{nj}*}, g_{nj}, g_0\Big)\end{aligned}$$

Summing over j and using (13) gives:

$$\tilde\psi_n^* - \tilde\psi_{0n} = \frac{1}{J}\sum_{j=1}^{J}\Big(\Big(P_{n,j}^1 - P_0\Big)D^*\Big(d_{nj}, Q_{nj}^{d_{nj}*}, g_{nj}\Big) + R_{1d_{nj}}\Big(Q_{nj}^{d_{nj}*}, Q_0^{d_{nj}*}, g_{nj}, g_0\Big)\Big).$$

We also have that:

$$\frac{1}{J}\sum_{j=1}^{J}(P_{n,j}^1 - P_0)\Big(D^*(d_{nj}, Q_{nj}^{d_{nj}*}, g_{nj}) - D^*(d_1, Q^{d_1}, g)\Big) = o_{P_0}(n^{-1/2}).$$

The above follows from the first by applying the law of total expectation conditional on the training sample, and then noting that each $\hat{Q}^*(P_{n,B_n}^0, \varepsilon_n)$ only relies on P_{n,B_n}^0 through the finite dimensional parameter ε_n. Because GLM-based parametric classes easily satisfy an entropy integral condition [58], the consistency assumption on $D^*(d_{nj}, Q_{nj}^{d_{nj}*}, g_{nj})$ shows that the above is second order. We refer the reader to Zheng and van der Laan [55] for a detailed proof of the above result for general cross-validation schemes, including J-fold cross-validation.

It follows that:

$$\begin{aligned}\tilde\psi_n^* - \tilde\psi_{0n} =& (P_n - P_0)D^*(d_1, Q^{d_1}, g)\\ &+ \frac{1}{J}\sum_{j=1}^{J} R_{1d_{nj}}\Big(Q_{nj}^{d_{nj}*}, Q_0^{d_{nj}*}, g_{nj}, g_0\Big) + o_{P_0}(n^{-1/2}).\end{aligned}$$

□

Appendix B: Estimators of the mean outcome under the optimal rule

TMLE of the mean outcome under a given rule

This TMLE for a fixed dynamic treatment rule has been presented in the literature, but for the sake of being self-contained it will be shortly described here. The TMLE yields a substitution estimator that empirically solves the estimating equations corresponding to the efficient influence curve, analogous to Theorem 2 for general d. By substitution estimator, we mean that the TMLE can be written as the mapping Ψ applied to a particular Q.

Assume without loss of generality that $Y \in [0,1]$. In this section we use lower case letters to emphasize when quantities are the values taken on by random variables rather than the random variables themselves, for example, our sample is given by $(o_1, \ldots, o_n)$, where $o_i = (\bar{l}(1)_i, \bar{a}(1)_i, y_i)$. The indicator for not being right censored at time j for individual i is given by $a_2(j)_i$.

Regress $(y_i : a_2(0)_i = a_2(1)_i = 1)$ on $(\bar{a}(1)_i, \bar{l}(1)_i : a_2(0)_i = a_2(1)_i = 1)$ to get an estimate

$$(a_1(0), a_1(1), \bar{l}(1)) \mapsto E_n\left[Y|\bar{A}(1) = ((a_1(0),1),(a_1(1),1)), \bar{L}(1) = \bar{l}(1)\right]. \tag{16}$$

Note that we have only used individuals who are not right censored at time 1 to obtain this fit. The above regression can be fitted using a data adaptive technique such as super-learning [59]. To estimate $E_{P_0}[Y|\bar{A}(1) = d(a(0),v), \bar{l}(1)]$, use

$$(a(0), \bar{l}(1)) \mapsto E_n\left[Y|\bar{A}(1) = d(a(0),v), \bar{L}(1) = \bar{l}(1)\right],$$

where we remind the reader that we are treating the rule $d = d_n$ as a known function and that v is a function of $\bar{l}(1)$ that sets the indicators for not being censored to 1. Consider the fluctuation submodel

$$\begin{aligned}&\operatorname{logit} E_n^{(\varepsilon_2)}\left[Y|\bar{A}(1) = d(A(0),V), \bar{L}(1)\right]\\ &\quad = \operatorname{logit} E_n\left[Y|\bar{A}(1) = d(A(0),V), \bar{L}(1)\right] + \varepsilon_2 H_2(g_n)(O),\end{aligned}$$

where

$$H_2(g_n)(O) = \frac{I\big(\bar{A}(1) = d(A(0),V)\big)}{\prod_{j=0}^{1} g_{n,A(j)}(O)}.$$

Let ε_{2n} be the estimate for ε_2 obtained by running a univariate logistic regression of $(y_i : i = 1, \ldots, n)$ on $(H_2(g_n)(o_i) : i = 1, \ldots n)$ using

$$\left(\operatorname{logit} E_n\left[Y|\bar{A}(1) = d(a(0)_i, v_i), \bar{L}(1) = \bar{l}(1)_i\right] : i = 1, \ldots, n\right)$$

as offset. This defines a targeted estimate

$$E_n^*\left[Y|\bar{A}(1) = d(A(0),V), \bar{L}(1)\right] \equiv E_n^{(\varepsilon_{2n})}\left[Y|\bar{A}(1) = d(A(0),V), \bar{L}(1)\right] \tag{17}$$

of the regression function, where we remind the reader that the targeted estimate is chosen to ensure that the empirical mean of the component D_2^* is 0 when we plug in the estimate of the intervention mechanism and the targeted estimate of the regression function for the unknown true quantities.

We now develop a targeted estimator of the second regression function in D_1^* to ensure that the substitution estimator of D_1^* will have empirical mean 0. Regress

$$\left(E_n\left[Y|\bar{A}(1) = d(a(0)_i, v_i), \bar{L}(1) = \bar{l}(1)_i\right] : a_2(0)_i = 1\right)$$

on $(l(0)_i, a(0)_i : a_2(0)_i = 1)$ to get the regression function

$$(a_1(0), l(0)) \mapsto E_n\left[E_n\left[Y|\bar{A}(1) = d(A(0),V), \bar{L}(1)\right]\big|A(0) = (a_1(0),1), L(0) = l(0)\right]. \tag{18}$$

One can estimate this quantity using the super-learner algorithm among all individuals who are not right censored at time 0. For honest cross-validation in the super-learner algorithm, the nuisance parameter $E_n\left[Y|\bar{A}(1) = d(A(0),V), \bar{L}(1)\right]$ should be fit on the training samples in the super-learner algorithm. We refer the reader to Appendix B of van der Laan and Gruber [41] for a detailed explanation of this procedure. The same strategy holds for estimating the nuisance parameter g_0 when necessary (e.g., in an observational study).

For an estimate of $E_{P_0}[Y_d|L(0)]$, one can use the regression function above, but with $a(0)$ fixed to $d_{A(0)}(v(0))$, which is itself a function of $l(0)$. We will denote this function by $l(0) \mapsto E_n[Y_d|L(0) = l(0)]$. We now wish to fluctuate this initial estimator so that the plug-in estimator of $D_1^*(P_0)$ has empirical mean 0. In particular, we use the submodel

$$\text{logit}\, E_n^{(\varepsilon_1)}[Y_d|L(0)] = \text{logit}\, E_n[Y_d|L(0)] + \varepsilon_1 H_1(g_n),$$

where

$$H_1(g_n) = \frac{I\big(A(0) = d_{A(0)}(V(0))\big)}{g_{n,A(0)}(O)}.$$

Let ε_{1n} be the estimate for ε_1 obtained by running a univariate logistic regression of

$$\big(E_n^*\big[Y|\bar{A}(1) = d(a(0)_i, v_i), \bar{L}(1) = \bar{l}(1)_i\big] : i = 1, \ldots, n\big)$$

on $(H_1(g_n)(o_i) : i = 1, \ldots, n)$ using $(\text{logit} E_n[Y_d|L(0) = l(0)_i] : i = 1, \ldots, n)$ as offset. A targeted estimate of $E_{P_0}[Y_d|L(0)]$ is given by

$$E_n^*[Y_d|L(0)] \equiv E_n^{(\varepsilon_{1n})}[Y_d|L(0)] \tag{19}$$

Plugging the targeted regressions and g_n into the expression for D_1^* shows that this estimate of D_1^* has empirical mean 0.

Let $Q_{L(0),n}$ be the empirical distribution of $L(0)$, and let Q_n^{d*} be the parameter mapping representing the collection containing $Q_{L(0),n}$ and the targeted regression functions in (17) and (19). This concludes the presentation of the components of the TMLE of $E_{P_0} Y_d$. The discussion of properties of this estimator is continued in the main text.

CV-TMLE of the mean outcome under data adaptive *V*-optimal rule

Let $\hat{d} : \mathcal{M} \to \mathcal{D}$ be an estimator of the V-optimal rule d_0. Firstly, without loss of generality we can assume that $Y \in [0, 1]$. Denote the realizations of B_n with $j = 1, \ldots, J$, and let $d_{nj} \equiv \hat{d}(P_{n,j}^0)$ denote the estimated rule on training sample j. Let

$$(a(0), \bar{l}(1)) \mapsto E_{nj}\big[Y|\bar{A}(1) = d_{nj}(a(0), v), \bar{L}(1) = \bar{l}(1)\big] \tag{20}$$

represent an initial estimate of $E_{P_0}[Y|\bar{A}(1) = d_{nj}(A(0), V), \bar{L}(1)]$ based on the training sample j, obtained analogously to the estimator in (16). Similarly, let g_{nj} represent the estimated intervention mechanism based on this training sample $P_{n,j}^0$, $j = 1, \ldots, J$. Consider the fluctuation submodel

$$\begin{aligned} &\text{logit}\ E_{nj}^{(\varepsilon_2)}\big[Y|\bar{A}(1) = d_{nj}(A(0), V), \bar{L}(1)\big] \\ &\quad = \text{logit}\ E_{nj}\big[Y|\bar{A}(1) = d_{nj}(A(0), V), \bar{L}(1)\big] + \varepsilon_2 H_2(g_{nj})(O) \end{aligned}$$

where

$$H_2(g_{nj})(O) = \frac{I\big(\bar{A}(1) = d_{nj}(A(0), V(1))\big)}{\prod_{l=0}^{1} g_{nj,A(l)}(O)}.$$

Note that the fluctuation ε_2 does not rely on j. Let

$$\varepsilon_{2n} = \arg\min_{\varepsilon_2} \frac{1}{J} \sum_{j=1}^{J} P_{n,j}^1 \tilde{\phi}(E_{nj}^{(\varepsilon_2)}),$$

where $E_{nj}^{(\varepsilon_2)}$ represents the fluctuated function in (20) and

$$-\tilde{\phi}(f)(o) = y\ \log f(o) + (1 - y)\ \log(1 - f(o)). \tag{21}$$

for all $f : \mathcal{O} \to (0,1)$. For each $i = 1, \ldots, n$, let $j(i) \in \{1, \ldots, J\}$ represent the value of B_n for which element i is in the validation set. The fluctuation ε_{2n} can be obtained by fitting a univariate logistic regression of $(y_i : i = 1, \ldots, n)$ on $(H_2(g_{nj(i)})(o_i) : i = 1, \ldots, n)$ using

$$\left(\mathrm{logit} E_{nj(i)}\left[Y|\bar{A}(1) = d_{nj}(a(0)_i, v_i), \bar{L}(1) = \bar{l}(1)_i\right] : i = 1, \ldots, n\right)$$

as offset. Thus each observation i is paired with nuisance parameters that are fit on the training sample which does not contain observation i. This defines a targeted estimate

$$E^*_{nj}\left[Y|\bar{A}(1) = d_{nj}(A(0), V), \bar{L}(1)\right] \equiv E^{(\varepsilon_{2n})}_{nj}\left[Y|\bar{A}(1) = d_{nj}(A(0), V), \bar{L}(1)\right] \tag{22}$$

of $E_{P_0}[Y|\bar{A}(1) = d_{nj}(A(0), V), \bar{L}(1)]$. We note that this targeted estimate only depends on P_n through the training sample $P^0_{n,j}$ and the one-dimensional ε_{2n}.

We now aim to get a targeted estimate of $E_{P_0}[Y_{d_{nj}}|L(0)]$. We can obtain an estimate

$$(a_1(0), l(0)) \mapsto E_{nj}\left[E_{nj}\left[Y|\bar{A}(1) = d_{nj}(A(0), V), \bar{L}(1)\right]|A(0) = (a_1(0), 1), L(0) = l(0)\right] \tag{23}$$

in the same manner as we estimated the quantity in (18), with the caveat that we replace $E_n[Y|\bar{A}(1) = d_{nj}(A(0), V), \bar{L}(1)]$ by $E_{nj}[Y|\bar{A}(1) = d_{nj}(A(0), V), \bar{L}(1)]$ and only fit the regression on samples that are not right censored at time 0 and are in training set j. For an estimate $E_{nj}[Y_{d_{nj}}|L(0)]$ of $E_{P_0}[Y_{d_{nj}}|L(0)]$, we can use the regression function above but with $a(0)$ fixed to $d_{nj,A(0)}(v(0))$.

Consider the fluctuation submodel

$$\mathrm{logit} E^{(\varepsilon_1)}_{nj}\left[Y_{d_{nj}}|L(0)\right] = \mathrm{logit} E_{nj}\left[Y_{d_{nj}}|L(0)\right] + \varepsilon H_1(g_{nj})(O),$$

where

$$H_1(g_{nj})(O) = \frac{I(A(0) = d_{nj,A(0)}(V(0)))}{g_{nj,A(0)}(O)}.$$

Again the fluctuation ε_1 does not rely on j. Let

$$\varepsilon_{1n} = \arg\min_{\varepsilon_1} \frac{1}{J}\sum_{j=1}^{J} P^1_{n,j}\tilde{\phi}(E^{(\varepsilon_1)}_{nj}),$$

where $\tilde{\phi}$ is defined in (21). For each $i = 1, \ldots, n$, again let $j(i) \in \{1, \ldots, J\}$ represent the value of B_n for which element i is in the validation set. The fluctuation ε_{1n} can be obtained by fitting a univariate logistic regression of

$$\left(E^*_{nj(i)}\left[Y|\bar{A}(1) = d_{nj(i)}(a(0)_i, v_i), \bar{l}(1)_i\right] : i = 1, \ldots, n\right)$$

on $(H_1(g_{nj(i)})(o_i) : i = 1, \ldots, n)$ using

$$\left(\mathrm{logit} E_{nj(i)}\left[Y_{d_{nj(i)}}|L(0) = l(0)_i\right] : i = 1, \ldots, n\right)$$

as offset. This defines a targeted estimate

$$E^*_{nj}\left[Y_{d_{nj}}|L(0)\right] \equiv E^{(\varepsilon_{1n})}_{nj}\left[Y_{d_{nj}}|L(0)\right] \tag{24}$$

of $E_{P_0}[Y_{d_{nj}}|L(0)]$. We note that this targeted estimate only depends on P_n through the training sample $P^0_{n,j}$ and the one-dimensional ε_{1n}.

Let $Q_{L(0),nj}$ be the empirical distribution of $L(0)_i$ for the validation sample $P^1_{n,j}$. For all $j = 1, \ldots, J$, let $Q^{d_{nj}*}_{nj}$ be the parameter mapping representing the collection containing $Q_{L(0),nj}$ and the targeted regressions in (22) and (24). This defines an estimator $\psi^*_{nj} = P^1_{n,j}\bar{Q}^*_{1nj}$ of $\psi_{d_{nj}0} = \Psi_{d_{nj}}(P_0)$ for each $j = 1, \ldots, J$. CV-TMLE is now defined as $\psi^*_n = \frac{1}{J}\sum_{j=1}^{J}\psi^*_{nj}$. This CV-TMLE solves the cross-validated efficient influence curve equation:

$$\frac{1}{J}\sum_{j=1}^{J} P^1_{n,j} D^*\left(d_{nj}, Q^{d_{nj}*}_{nj}, g_{nj}\right) = 0.$$

Further, each $Q^{d_{nj}*}_{nj}$ only relies on $P^1_{n,j}$ through the univariate parameters ε_{1n} and ε_{2n}. This will allow us to use the entropy integral arguments presented in Zheng and van der Laan [55] which show that no restrictive empirical process conditions are needed on the initial estimates in (20) and (23).

The only modification relative to the original CV-TMLE presented in Zheng and van der Laan [55] is that in the above description we change our target on each training sample into the training sample-specific target parameter implied by the fit $\hat{d}(P^0_{n,B_n})$ on the training sample, while in the original CV-TMLE formulation, the target would still be $\Psi_{d_0}(P_0)$. With this minor twist, the (same) CV-TMLE is now used to target the average of training sample-specific target parameters averaged across the J training samples. This utilization of CV-TMLE was already used to estimate the average (across training samples) of the true risk of an estimator based on a training sample in van der Laan and Petersen [53] and Díaz and van der Laan [54], so that this just represents a generalization of that application of CV-TMLE to estimate general data adaptive target parameters as proposed in van der Laan et al. [46].

Appendix C: Why the TMLE may have better coverage than the estimating equation approach in a randomized clinical trial

We wrote this section after performing our simulations because we wanted to understand why the TMLE is outperforming the DR-IPCW estimating equation approach by such a wide margin. The two approaches do not typically give such disparate estimates in a randomized clinical trial, so it is natural to ask why this is happening in our simulations. Part of this section is conjecture (which is in line with our simulations), but we offer some justification to support this conjecture.

We now offer a heuristic explanation of why the TMLE may have better coverage than the DR-IPCW estimating equation approach when estimating the data adaptive parameter ψ_{0n}. Suppose we have a single time point data structure $O = (W, A, Y)$ drawn according to the distribution P_0 in a randomized clinical trial without missingness. Here we use notation which directly describes the single time point data structure rather than forcing this problem into the longitudinal context as in Section 8.1.1. Let $d_0 = \arg\max_d E_{P_0}E_{P_0}[Y|A = d(V), W]$ for some V that is a function of W. Suppose we observe $o_1, \ldots, o_n$ and let d_n be an estimate of d_0, which is obtained using the methods in our accompanying technical report [47]. For any fixed rule d, the efficient influence curve at some $P \in \mathcal{M}$ is given by

$$\begin{aligned} E_P\left[\frac{I(A = d(V))}{g(A|W)}(Y - E_P[Y|A = d(V), W])\right] \\ + E_P[Y|A = d(V), W] - E_P E_P[Y|A = d(V), W], \end{aligned}$$

where g is the intervention mechanism under P. Again we have that $E_{P_0}Y_{d_0}$ has the same influence curve as above with $d = d_0$ (see our online technical report). Suppose that $g_0 = 1/2$ is known and we have estimated $E_{P_0}[Y|A = d(V), W]$ perfectly, though we continue to work in the model where $E_{P_0}[Y|A = d(V), W]$ is treated as unknown so that simply averaging over this quantity is not appropriate if we want inference or robustness.

For any fixed rule $V \mapsto d(V)$, it is easy to show that

$$E_{P_0}\left[\frac{I(A = d(V))}{g_0(A|W)}(Y - E_{P_0}[Y|A = d(V), W])\right] = 0,$$

where $g_0(a|w)$ represents the probability under P_0 that $A = a$ given $W = w$. Similarly, we expect that

$$\beta_d(P_n) \equiv \frac{1}{n}\sum_{i=1}^{n}\frac{I(a_i = d(v_i))}{g_0(a_i|w_i)}(y_i - E_{P_0}[Y|A = d(v_i), W = w_i]) \approx 0.$$

Further, $E_{P_0}\beta_d(P_n) = 0$ for fixed d, where the expectation is over the observed sample P_n but not the fixed rule d. In the first part of this paper we argued that one can learn an estimated rule d_n on the entire data set, and then treat this rule d_n as known when estimating $E_{P_0}Y_{d_n}$. This is asymptotically valid under the conditions given in this paper, but even if these conditions hold we may expect some finite sample bias. In our simulation this finite sample bias is manifested as

$$E_{P_0}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{I(a_i = d_n(v_i))}{g_0(a_i|w_i)}(y_i - E_{P_0}[Y|A = d_n(v_i), W = w_i])\right] > 0,$$

where the expectation is over the observed sample P_n and the estimated rule d_n. For a single time point simulation with $V = L_3(1)$, this sample average is approximately 0.013 on average across 1,000 simulations. When $V = L_1(1), \ldots, L_1(4)$, this sample average is approximately 0.040 on average across 1,000 simulations. Because this was a follow-up analysis, we ran these simulations on different Monte Carlo draws than those used for our results in the main text. We conjecture that the above phenomenon is not specific to our simulation settings and will occur in more general settings. Our companion paper in this issue explores the estimation of d_0, and a careful look at the mean performance-based loss function presented in that paper will show that indeed one way to make the empirical risk smaller is to choose d_n so that $\beta_{d_n}(P_n) > 0$. Nonetheless, selecting d_n by a cross-validation selector as we propose in our companion paper should help mitigate this issue since β_{d_n} for d_n trained on a training sample should have empirical mean close to 0 in the validation sample.

The DR-IPCW estimating equation gives the estimator:

$$\hat{\Psi}_{EE}^{d_n}(P_n) \equiv \psi_{n,EE} \equiv \beta_{d_n}(P_n) + \frac{1}{n}\sum_{i=1}^{n} E_{P_0}[Y|A = d_n(V_i), W = W_i].$$

This estimator has bias $E_{P_0}\beta_{d_n}(P_n)$, where the expectation is over the random sample P_n and the estimated rule d_n.

Consider the simple linear TMLE which fluctuates $w \mapsto E_{P_0}[Y|A = d_n(v), W = w]$ using the submodel:

$$E_{P_0}^{(\varepsilon)}[Y|A = d_n(V), W] = E_{P_0}[Y|A = d_n(V), W] + \varepsilon\frac{I(A = d_n(V))}{g_0(A|W)}$$

where we recall that $w \mapsto E_{P_0}[Y|A = d_n(v), W = v]$ is being treated as unknown. A valid TMLE is given by choosing ε_n to minimize the mean-squared error between Y and $E_{P_0}^{(\varepsilon)}[Y|A = d_n(V), W]$. When Y is bounded, the logistic fluctuations that we have presented in this paper are preferable to the linear fluctuation because they respect our model constraints. We consider the linear fluctuation here for simplicity. The minimizer ε_n is given by

$$\varepsilon_n = \frac{\frac{1}{n}\sum_i \frac{I(a_i = d_n(v_i))}{g_0(a_i|w_i)}(y_i - E_{P_0}[Y|A = d_n(v_i), W = w_i])}{\frac{1}{n}\sum_i \frac{I(a_i = d_n(v_i))}{g_0(a_i|w_i)^2}}$$

$$= \frac{1}{2}\frac{\beta_{d_n}(P_n)}{\frac{1}{n}\sum_i \frac{I(a_i = d_n(v_i))}{g_0(a_i|w_i)}},$$

if $\frac{1}{n}\sum_i \frac{I(a_i = d_n(v_i))}{g_0(a_i|w_i)} > 0$ and we take $\varepsilon_n = 0$ if $\frac{1}{n}\sum_i \frac{I(a_i = d_n(v_i))}{g_0(a_i|w_i)} = 0$. The denominator above is the same as the denominator in a modified Horvitz-Thompson estimator [60] and, more importantly, appears in one of the terms in the TMLE, which is given by

$$
\begin{aligned}
\psi^*_{n,\mathrm{TMLE}} &\equiv \frac{1}{n}\sum_{i=1}^{n} E^{(\varepsilon_n)}_{P_0}[Y|A=d_n(V),W] \\
&= \frac{1}{n}\sum_{i=1}^{n} E_{P_0}[Y|A=d_n(V),W] + \frac{\varepsilon_n}{n}\sum_{i=1}^{n}\frac{I(A=d_n(V))}{g_0(A|W)} \\
&= \frac{1}{n}\sum_{i=1}^{n} E_{P_0}[Y|A=d_n(V),W] + \frac{\beta_{d_n}(P_n)}{2}.
\end{aligned}
$$

This linear fluctuation TMLE has bias $E_{P_0}\left[\frac{\beta_{d_n}(P_n)}{2}\right]$, which is half the bias of $\hat{\Psi}^{d_n}_{EE}(P_n)$.

The arguments presented in this section are mainly interesting if $E_{P_0}[\beta_{d_n}(P_n)] \neq 0$. We have conjectured that $E_{P_0}[\beta_{d_n}(P_n)] > 0$ for many data generating distributions P_0 and estimators of the optimal rule, though we have not analytically justified this claim. If the conditions of Theorem 5 hold, then this bias will only occur in finite samples. For simplicity we analyzed a different TMLE than the ones presented in this paper. First, we analyzed a TMLE for the single time point problem. We show in our online technical report that the single and multiple time point problems are closely related, so we expect that these results carry over to the two time point case. We have also analyzed a linear rather than logistic fluctuation in this section. We did this simply so we could get a straightforward expression for the bias of the TMLE without having to worry about linearizing the fluctuation submodel in a neighborhood of 0. Similar results should hold for the logistic fluctuations. We also assumed that $E_{P_0}[Y|A=d_n(V),W]$ was estimated perfectly, which of course is not true in practice. Nonetheless, this assumption makes our results clearer because then we do not have to worry about a resulting empirical process term.

The term $\beta_{d_n}(P_n)$ only causes problems because d_n is learned from the same data over which the estimators of $E_{P_0}Y_{d_n}$ are run. The cross-validated approaches that we have presented in this paper do not suffer from this conjectured bias because we can condition on the training sample and then treat d_n as known. For fixed d, $E_{P_0}[\beta_d(P_n)] = 0$ and thus $\beta_d(P_n)$ will not cause problems.

References

1. Robins JM. A new approach to causal inference in mortality studies with sustained exposure periods-application to control of the healthy worker survivor effect. Math Mod 1986;7:1393–512.
2. Robins JM. Information recovery and bias adjustment in proportional hazards regression analysis of randomized trials using surrogate markers. In Proceedings of the Biopharmaceutical Section. American Statistical Association, 1993.
3. Robins JM. Causal inference from complex longitudinal data. In Berkane M, editor. Latent variable modeling and applications to causality. New York: Springer, 1997:69–117.
4. Robins JM. Marginal structural models versus structural nested models as tools for causal inference. In: Halloran ME, Berry D, editors. Statistical models in epidemiology, the environment, and clinical trials (Minneapolis, MN, 1997). New York: Springer, 2000:95–133.
5. Holland PW. Statistics and causal inference. J Am Stat Assoc 1986;810:945–60.
6. Neyman J. Sur les applications de la théorie des probabilites aux experiences agaricales: essay des principle (1923). Excerpts reprinted (1990) in English (D. Dabrowska and T. Speed), trans. Stat Sci 1990;5:463–72.
7. Pearl J. Causality: models, reasoning and inference, 2nd ed. New York: Cambridge University Press, 2009.
8. Robins JM. Addendum to: "A new approach to causal inference in mortality studies with a sustained exposure period–application to control of the healthy worker survivor effect". Comput Math Appl 1987;140:923–45. ISSN 0097-4943
9. Robins JM. A graphical approach to the identification and estimation of causal parameters in mortality studies with sustained exposure periods. J Chron Dis (40, Supplement) 1987;2:139s–161s.
10. Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. J Educ Psychol 1974;66:688–701.
11. Rubin DB. Matched sampling for causal effects. Cambridge, MA: Cambridge University Press, 2006.
12. Lavori P, Dawson R. A design for testing clinical strategies: biased adaptive within-subject randomization. J R Stat Soc Ser A 2000;163:29–38.

13. Lavori P, Dawson R. Adaptive treatment strategies in chronic disease. Annu Rev Med 2008;59:443–53.
14. Murphy S. An experimental design for the development of adaptive treatment strategies. Stat Med 2005;24:1455–81.
15. Rosthøj S, Fullwood C, Henderson R, Stewart S. Estimation of optimal dynamic anticoagulation regimes from observational data: a regret-based approach. Stat Med 2006;88:4197–215.
16. Thall P, Millikan R, Sung H-G. Evaluating multiple treatment courses in clinical trials. Stat Med 2000;19:10111028.
17. Thall P, Sung H, Estey E. Selecting therapeutic strategies based on efficacy and death in multicourse clinical trials. J Am Stat Assoc 2002;39:29–39.
18. Wagner E, Austin B, Davis C, Hindmarsh M, Schaefer J, Bonomi A. Improving chronic illness care: translating evidence into action. Health Aff 2001;20:64–78.
19. Petersen ML, Deeks SG, Martin JN, van der Laan MJ. History-adjusted marginal structural models to estimate time-varying effect modification. Am J Epidemiol 2007;166:985–93.
20. van der Laan MJ, Petersen ML. Causal effect models for realistic individualized treatment and intention to treat rules. Int J Biostat 2007;3:Article 3.
21. Robins J, Orallana L, Rotnitzky A. Estimation and extrapolation of optimal treatment and testing strategies. Stat Med 2008;27:4678–721.
22. Lavori P, Dawson R. Dynamic treatment regimes: practical design considerations. Clin Trials 2004;1:9–20.
23. Chakraborty B, Murphy SA, Strecher V. Inference for non-regular parameters in optimal dynamic treatment regimes. Stat Methods Med Res 2010;19:317–43.
24. Kasari C. Developmental and augmented intervention for facilitating expressive language. ClinicalTrials.gov database, updated Apr. 26, 2012, Natl. Inst:0 accessed July 24, 2013, 2009.
25. Lei H, Nahum-Shani I, Lynch K, Oslin D, Murphy S. A SMART design for building individualized treatment sequences. Annu Rev Clin Psychol 2011;8:21–48.
26. Nahum-Shani I, Qian M, Almirall D, Pelham WE, Gnagy B, Fabiano GA, et al. Experimental design and primary data analysis methods for comparing adaptive interventions. Psychol Methods 2012;17:457–77.
27. Nahum-Shani I, Qian M, Almirall D, Pelham WE, Gnagy B, Fabiano GA, et al. Q-learning: a data analysis method for constructing adaptive interventions. Psychol Methods 2012;17:478–94.
28. Jones H. Reinforcement-based treatment for pregnant drug abusers. ClinicalTrials.gov data base, updated October 19, 2012, Natl. Inst:0 accessed July24, 2013, 2010.
29. Chakraborty B, Murphy SA. Dynamic treatment regimens. Annu Rev Stat Appl 2013;1:1–18.
30. Murphy SA. Optimal dynamic treatment regimes. J R Stat Soc Ser B 2003;65:331–55.
31. Robins JM. Discussion of "optimal dynamic treatment regimes" by Susan A. Murphy. J R Stat Soc Ser B 2003;65: 355–66.
32. Robins JM. Optimal structural nested models for optimal sequential decisions. In Proceedings of the Second Seattle Symposium on Biostatistics 2004:189–326.
33. Sutton R, Sung H. Reinforcement learning: an introduction. Cambridge, MA: MIT Press, 1998.
34. van der Laan MJ, Robins JM. Unified methods for censored longitudinal data and causality. New York: Springer, 2003.
35. Orellana L, Rotnitzky A, Robins JM. Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, part I: main content. Int J Biostat 2010;6:Article 8.
36. Bang H, Robins JM. Doubly robust estimation in missing data and causal inference models. Biometrics 2005;61:962–72.
37. van der Laan MJ. The construction and analysis of adaptive group sequential designs. Technical Report 232, Division of Biostatistics, University of California, Berkeley, CA, 2008.
38. van der Laan MJ, Rose S. Targeted learning: causal inference for observational and experimental data. New York: Springer, 2012.
39. van der Laan MJ, Rubin DB. Targeted maximum likelihood learning. Int J Biostat 2006;2:Article 11.
40. Petersen M, Schwab J, Gruber S, Blaser N, Schomaker M, van der Laan MJ. Targeted minimum loss based estimation of marginal structural working models. J Causal Inference 2013;submitted.
41. van der Laan MJ, Gruber S. Targeted minimum loss based estimation of causal effects of multiple time point interventions. Int J Biostat 2012;8:Article 9.
42. Cotton C, Heagerty P. A data augmentation method for estimating the causal effect of adherence to treatment regimens targeting control of an intermediate measure. Stat Biosci 2011;3:28–44.
43. Hernan MA, Lanoy E, Costagliola D, Robins JM. Comparison of dynamic treatment regimes via inverse probability weighting. Basic Clin Pharmacol 2006;98:237–42.
44. Shortreed S, Moodie E. Estimating the optimal dynamic antipsychotic treatment regime: evidence from the sequential-multiple assignment randomized CATIE Schizophrenia Study. J R Stat Soc C 2012;61:577–99.
45. Robins JM, Li L, Tchetgen E, van der Vaart AW. Higher order influence functions and minimax estimation of non-linear functionals. In Probability and statistics: essays in honor of David A. Freedman. Beachwood, OH: Institute of Mathematical Statistics, 2008:335–421. doi:10.1214/193940307000000527. Available at: http://projecteuclid.org/euclid.imsc/1207580092

46. van der Laan MJ, Hubbard AE, Kherad S. Statistical inference for data adaptive target parameters. Technical Report 314, Division of Biostatistics, University of California, Berkeley, CA, 2013.
47. van der Laan MJ. Targeted learning of an optimal dynamic treatment and statistical inference for its mean outcome. Technical Report 317, UC Berkeley, CA, 2013.
48. van der Laan MJ, Luedtke AR. Targeted learning of an optimal dynamic treatment, and statistical inference for its mean outcome. Technical Report 329, UC Berkeley, CA, 2014.
49. Robins JM, Rotnitzky A, Scharfstein DO. Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In: Halloran ME, Berry D, editors. Statistical models in epidemiology, the environment and clinical trials. IMA Volumes in Mathematics and Its Applications. Springer, 1999.
50. Bickel PJ, Klaassen CA, Ritov Y, Wellner J. Efficient and adaptive estimation for semiparametric models. New York: Springer, 1997.
51. van der Vaart AW. Asymptotic statistics. New York: Cambridge University Press, 1998.
52. Robins J, Rotnitzky A. Discussion of "dynamic treatment regimes: technical challenges and applications. Electron J Stat 2014;8:1273–89. doi:10.1214/14-EJS908. URL http://dx.doi.org/10.1214/14-EJS908
53. Díaz I, van der Laan MJ. Targeted data adaptive estimation of the causal dose response curve. Technical Report 306, Division of Biostatistics, University of California, Berkeley, CA, submitted to JCI, 2013.
54. van der Laan MJ, Petersen ML. Targeted learning. In Zhang C, Ma Y, editors. Ensemble machine learning. New York: Springer, 2012.
55. Zheng W, van der Laan MJ. Asymptotic theory for cross-validated targeted maximum likelihood estimation. Technical Report 273, Division of Biostatistics, University of California, Berkeley, CA, 2010.
56. Zheng W, van der Laan MJ. Cross-validated targeted minimum loss based estimation. In van der Laan MJ, Rose S, editors. Targeted learning: causal inference for observational and experimental studies. New York: Springer, 2012.
57. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2014. Available at: http://www.R-project.org/
58. van der Vaart AW, Wellner JA. Weak convergence and empirical processes. New York: Springer, 1996.
59. van der Laan MJ, Polley E, Hubbard A. Super learner. Stat Appl Genet Mol Biol 2007;6:Article 25.
60. Hernán MA, Robins JM. Estimating causal effects from epidemiological data. J Epidemiol Community Health 2006;60:578–86.