*University of California, Berkeley*
U.C. Berkeley Division of Biostatistics Working Paper Series

*Year* 2015 *Paper* 333

# Optimal Dynamic Treatments in Resource-Limited Settings

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

*University of California, Berkeley, Division of Biostatistics, aluedtke@berkeley.edu

†University of California, Berkeley, Division of Biostatistics, laan@berkeley.edu

# Optimal Dynamic Treatments in Resource-Limited Settings

Alexander R. Luedtke and Mark J. van der Laan

**Abstract**

A dynamic treatment rule (DTR) is a treatment rule which assigns treatments to individuals based on (a subset of) their measured covariates. An optimal DTR is the DTR which maximizes the population mean outcome. Previous works in this area have assumed that treatment is an unlimited resource so that the entire population can be treated if this strategy maximizes the population mean outcome. We consider optimal DTRs in settings where the treatment resource is limited so that there is a maximum proportion of the population which can be treated. We give a general closed-form expression for an optimal stochastic DTR in this resource-limited setting, and a closed-form expression for the optimal deterministic DTR under an additional assumption. We also present an estimator of the mean outcome under the optimal stochastic DTR in a large semiparametric model that at most places restrictions on the probability of treatment assignment given covariates. We give conditions under which our estimator is efficient among all regular and asymptotically linear estimators. All of our results are supported by simulations.

# 1 Introduction

Suppose one wishes to maximize the population mean of some outcome using some binary point treatment, where for each individual clinicians have access to (some subset of) measured baseline covariates. Such a treatment strategy is termed a dynamic treatment regime (DTR), and the (counterfactual) population mean outcome under a DTR is referred to as the value of a DTR. The DTR which maximizes the value is referred to as the optimal DTR or the optimal rule. There has been much recent work on this problem in the case where treatment is an unlimited resource (see Murphy, 2003 and Robins, 2004 for early works on the topic, and Chakraborty and Moodie, 2013 for a recent overview). It has been shown that the optimal treatment in this context is given by checking the sign of the average treatment effect conditional on (some subset of) the baseline covariates, also known as the blip function (Robins, 2004).

The optimal DTR assigns treatment to people from a given strata of covariates for which treatment is on average beneficial, and does not assign treatment to this strata otherwise. If treatment is even slightly beneficial to all subsets of the population, then such a treatment strategy would suggest treating the entire population. There are many realistic situations in which such a treatment strategy, or any strategy that treats a large proportion of the population, is not feasible due to limitations on the total amount of the treatment resource. In a discussion of Murphy (2003), Arjas observed that resource constraints may render optimal DTRs of little practical use when the treatment of interest is a social or educational program, though no solution to the constrained problem was given (Arjas et al., 2003).

The mathematical modeling literature has considered the resource allocation problem to a greater extent. Lasry et al. (2011) developed a model to allocate the annual CDC budget for HIV prevention programs to subpopulations which would benefit most from such an intervention. Tao et al. (2012) consider a mathematical model to optimally allocate screening procedures for sexually tranmitted diseases subject to a cost constraint. Though Tao et al. do not frame the problem as a statistical estimation problem, they end up confronting similar optimization challenges to those that we will face. In particular, they confront the (weakly) NP-hard knapsack problem from the combinatorial optimization literature (Karp, 1972; Korte and Vygen, 2012). We will end up avoiding most of the challenges associated with this problem by primarily focusing on stochastic treatment rules, which will reduce to the easier fractional knapsack problem (Dantzig, 1957; Korte and Vygen, 2012). Stochastic DTRs allow the treatment to rely on some external stochastic mech-

anism for individuals in a particular strata of covariates.

We consider a resource constraint under which there is a maximum proportion of the population which can be treated. We primarily focus on evaluating the public health impact of an optimal resource-constrained (R-C) DTR via its value. The value function has been shown to be of interest in several previous works (see, e.g., Zhang et al., 2012; van der Laan and Luedtke, 2014a; Goldberg et al., 2014). Despite the general interest of this quantity, estimating this quantity is challenging for unconstrained deterministic regimes at so-called exceptional laws, i.e. probability distributions at which the blip function is zero in some positive probability strata of covariates (Robins, 2004; a slightly more general assumption is given in Luedtke and van der Laan, 2014b). Chakraborty et al. (2014) showed that confidence intervals for this parameter using $m$-out-of-$n$ bootstrap, though these confidence intervals shrink at a slower than root-$n$ rate. Luedtke and van der Laan (2014b) showed that root-$n$ rate confidence intervals can be developed for this quantity under reasonable conditions in the large semiparametric model which at most places restrictions on the treatment mechanism.

We develop a root-$n$ rate estimator for the optimal R-C value and corresponding confidence intervals in this same large semiparametric model. We show that our estimator is efficient among all regular and asymptotically linear estimators under conditions. When the baseline covariates are continuous and the resource constraint is active, i.e. when the optimal R-C value is less than the optimal unconstrained value, these conditions are far more reasonable than the non-exceptional law assumption needed for regular estimation of the optimal unconstrained value.

We now give a brief outline of the paper. Section 2 defines the statistical estimation problem of interest, gives an expression for the optimal deterministic rule under a condition, and gives a general expression for the optimal stochastic rule. Section 3 presents our estimator of the optimal R-C value. Section 4 presents conditions under which the optimal R-C value is pathwise differentiable, and gives an explicit expression for the canonical gradient under these conditions. Section 5 describes the properties of our estimator, including how to develop confidence intervals for the optimal R-C value. Section 6 presents our simulation methods. Section 7 presents our simulation results. Section 8 closes with a discussion and areas of future research. All proofs are given in the appendix.

## 2 Optimal R-C rule and value

Suppose we observe $n$ independent and identically distributed (i.i.d.) draws from a single time point data structure $(W, A, Y) \sim P_0$, where the covariates $W$ has support $\mathcal{W}$, the treatment $A$ has support $\{0, 1\}$, and the outcome $Y$ has support in the closed unit interval. Our statistical model is nonparametric, beyond possible knowledge of the treatment mechanism, i.e. the probability of treatment given covariates. Little generality is lost with the bound on $Y$, given that any continuous outcome bounded in $[b, c]$ can be rescaled to the unit interval with the linear transformation $(y - b)/(c - b)$. Suppose that treatment are resources are limited so that at most a $\kappa \in (0, 1)$ proportion of the population can receive the treatment $A = 1$. Let $V$ be some function of $W$, and denote the support of $V$ with $\mathcal{V}$. A deterministic treatment rule $\tilde{d}$ takes as input a function of the covariates $v \in \mathcal{V}$ and outputs a binary treatment decision $\tilde{d}(v)$. The stochastic treatment rules considered in this work are maps from $\mathcal{U} \times \mathcal{V}$ to $\{0, 1\}$, where $\mathcal{U}$ is the support of some random variable $U \sim P_U$. If $d$ is a stochastic rule and $u \in \mathcal{U}$ is fixed, then $d(u, \cdot)$ represents a deterministic treatment rule. Throughout this work we will let $U$ be drawn independently of all draws from $P_0$.

For a distribution $P$, let $\bar{Q}_P(a, w) \triangleq E_P[Y|A = a, W = w]$. For notational convenience, we let $\bar{Q}_0 \triangleq \bar{Q}_{P_0}$. Let $\tilde{d}$ be a deterministic treatment regime. For a distribution $P$, let $\tilde{\Psi}_{\tilde{d}} \triangleq E_{P_0}[\bar{Q}_P(\tilde{d}(V), W)]$ represent the value of $\tilde{d}$. Under causal assumptions, this quantity is equal to the counterfactual mean outcome if, possibly contrary to fact, the rule $\tilde{d}$ were implemented in the population (Robins, 1987; Pearl, 2009). The optimal R-C deterministic regime at $P$ is defined as the deterministic regime $\tilde{d}$ which solves the optimization problem

$$\text{Maximize } \tilde{\Psi}_{\tilde{d}}(P) \text{ subject to } E_{P_0}[\tilde{d}(V)] \leq \kappa. \tag{1}$$

For a stochastic regime $d$, let $\Psi_d(P) \triangleq E_{P_U}[\tilde{\Psi}_{d(U,\cdot)}(P)]$ represent the value of $d$. Under causal assumptions, this quantity is equal to the counterfactual mean outcome if, possibly contrary to fact, the stochastic rule $d$ were implemented in the population (see Díaz and van der Laan, 2012 for a similar identification result). The optimal R-C stochastic regime at $P$ is defined as the stochastic treatment regime $d$ which solves the optimization problem

$$\text{Maximize } \Psi_d(P) \text{ subject to } E_{P_U \times P}[d(U, V)] \leq \kappa. \tag{2}$$

We call the optimal value under a R-C stochastic regime $\Psi(P)$. Because any deterministic regime can be written as a stochastic regime which does not rely on the stochastic mechanism $U$, we have that $\Psi(P) \geq \tilde{\Psi}(P)$. The constraint

$E_{P_U \times P}[d(U,V)] \leq \kappa$ above is primarily meant to represent a clinical setting where each patient arrives at the clinic with covariate summary measure $V$, a value of $U$ is drawn from $P_U$ for this patient, and treatment is then assigned according to $d(U,V)$. By Fubini's theorem, this is like rewriting the above constraint as $E_P E_{P_U}[d(U,V)] \leq \kappa$. Nonetheless, this constraint also represents the case where a single value of $U = u$ is drawn for the entire population, and each individual is treated according to the deterministic regime $d(u,\cdot)$, i.e. $E_{P_U} E_P[d(U,V)] \leq \kappa$. This case appears less interesting because, for a fixed $u$, there is no guarantee that $E_P[d(u,V)] \leq \kappa$.

For a distribution $P$, define the blip function as

$$\bar{Q}_{b,P}(v) \triangleq E_P\left[\bar{Q}_P(1,W) - \bar{Q}_P(0,W)\middle| V = v\right].$$

Let $S_P$ represent the survival function of $\bar{Q}_{b,P}$, i.e. $\tau \mapsto Pr_P(\bar{Q}_{b,P} > \tau)$. Let

$$\begin{aligned} \eta_P &\triangleq \inf\{\tau : S_P(\tau) \leq \kappa\} \\ \tau_P &\triangleq \max\{\eta_P, 0\}. \end{aligned} \tag{3}$$

For notational convenience we let $\bar{Q}_{b,0} \triangleq \bar{Q}_{b,P_0}$, $S_0 \triangleq S_{P_0}$, $\eta_0 \triangleq \eta_{P_0}$, and $\tau_0 \triangleq \tau_{P_0}$.

Define the deterministic treatment rule $\tilde{d}_P$ as $v \mapsto I(\bar{Q}_{b,P}(v) > \tau_P)$, and for notational convenience let $\tilde{d}_0 \triangleq \tilde{d}_{P_0}$. We have the following result.

**Theorem 1.** *If $Pr_P(\bar{Q}_{b,P}(V) = \tau_P) = 0$, then the $\tilde{d}_P$ is an optimal deterministic rule satisfying the resource constraint, i.e. $\tilde{\Psi}_{\tilde{d}_P}(P)$ attains the maximum described in (1).*

One can in fact show that $\tilde{d}_P$ is the $P$ almost surely unique optimal deterministic regime under the stated condition. We do not treat the case where $Pr_P(\bar{Q}_{b,P}(V) = \tau_P) > 0$ for deterministic regimes, since in this case (1) is a more challenging problem: for discrete $V$, (1) is a special case of the $0-1$ knapsack problem, which is NP-hard, though is considered one of the easier problems in this class (Karp, 1972; Korte and Vygen, 2012). Considering the optimization problem over stochastic rather than deterministic regimes yields a fractional knapsack problem, which is known to be solvable in polynomial time (Dantzig, 1957; Korte and Vygen, 2012).

Define the stochastic treatment rule $d_P$ by its distribution with respect to a random variable drawn from $P_U$:

$$Pr_{P_U}\left(d_P(U,v) = 1\right) = \begin{cases} \kappa - S_P(\tau_P), & \text{if } \bar{Q}_{b,P}(v) = \tau_P \text{ and } \tau_P > 0 \\ I(\bar{Q}_{b,P}(v) > \tau_P), & \text{otherwise.} \end{cases}$$

We will let $d_0 \triangleq d_{P_0}$. Note that $\tilde{d}_P(V)$ and $d_P(U,V)$ are $P_U \times P$ almost surely equal if $Pr_P(\bar{Q}_{b,P}(V) = \tau_P) = 0$ or if $\tau_P \leq 0$, and thus have the same value in these settings. It is easy to show that

$$E_{P_U \times P}[d_P(U,V)] = \kappa \text{ if } \tau_P > 0. \tag{4}$$

The following theorem establishes the optimality of the stochastic rule $d_P$ in a resource-limited setting.

**Theorem 2.** *The maximum in (2) is attained at $d = d_P$, i.e. $d_P$ is an optimal stochastic rule.*

Note that the above theorem does not claim that $d_P$ is the unique optimal stochastic regime. For discrete $V$, the above theorem is an immediate consequence of the discussion of the knapsack problem in Dantzig (1957).

In this paper we focus on the value of the optimal stochastic rule. Nonetheless, the techniques that we present in this paper will only yield valid inference in the case where the data is generated according to a distribution $P_0$ for which $Pr_0(\bar{Q}_{b,0}(V) = \tau_0) = 0$. This is analogous to assuming a non-exceptional law in settings where resources are not limited (Robins, 2004; Luedtke and van der Laan, 2014b), though we note that for continuous covariates $V$ this assumption is much more likely if $\tau_0 > 0$. It seems unlikely that the treatment effect in some positive probability strata of covariates will concentrate on some arbitrary (determined by the constraint $\kappa$) value $\tau_0$. Nonetheless, one could deal with situations where $Pr_0(\bar{Q}_{b,0}(V) = \tau_0) > 0$ using similar martingale-based online estimation techniques to those presented in Luedtke and van der Laan (2014b).

# 3 Estimating the optimal optimal R-C value

We now present an estimation strategy for the optimal R-C rule. The upcoming sections justify this strategy and suggest that it will perform well for a wide variety of data generating distributions. The estimation strategy proceeds as follows:

1. Obtain estimates $\bar{Q}_n$, $\bar{Q}_{b,n}$, and $g_n$ of $\bar{Q}_0$, $\bar{Q}_{b,0}$, and $g_0$ using any desired estimation strategy which respects the fact that $Y$ is bounded in the unit interval.

2. Estimate the marginal distributions of $W$ and $V$ with the corresponding empirical distributions.

3. Estimate $S_0$ with the plug-in estimator $S_n$ given by $\tau \mapsto \frac{1}{n}\sum_{i=1}^{n} I\left(\bar{Q}_{b,n}(v_i) > \tau\right)$.

4. Estimate $\eta_0$ with the plug-in estimator $\eta_n \triangleq \inf\{\tau : S_n(\tau) \leq \kappa\}$.

5. Estimate $\tau_0$ with the plug-in estimator given by $\tau_n \triangleq \max\{\eta_n, 0\}$.

6. Estimate $d_0$ with the plug-in estimator $d_n$ with distribution

$$Pr_{P_U}(d_n(U,v)=1) = \begin{cases} \kappa - S_n(\tau_n), & \text{if } \bar{Q}_{b,n}(v) = \tau_n \text{ and } \tau_n > 0 \\ I(\bar{Q}_{b,n}(v) > \tau_n), & \text{otherwise.} \end{cases}$$

7. Run a TMLE for the parameter $\Psi_{d_n}(P_0)$:

   (a) For $\tilde{a} \in \{0,1\}$, define $H(a,w) \triangleq \frac{Pr_{P_U}(d_n(U,v)=a)}{g_n(a|w)}$. Run a univariate logistic regression using:

$$\begin{aligned} \text{Outcome:} &\quad (y_i : i = 1, ..., n) \\ \text{Offset:} &\quad \left(\operatorname{logit} \bar{Q}_n(a_i, w_i) : i = 1, ..., n\right) \\ \text{Covariate:} &\quad (H(a_i, w_i) : i = 1, ..., n)\,. \end{aligned}$$

   Let $\epsilon_n$ represent the estimate of the coefficient for the covariate, i.e.

$$\epsilon_n \triangleq \underset{\epsilon \in \mathbb{R}}{\operatorname{argmax}} \frac{1}{n}\sum_{i=1}^{n} \left[\bar{Q}_n^{\epsilon}(a_i, w_i) \log y_i + \left(1 - \bar{Q}_n^{\epsilon}(a_i, w_i)\right) \log(1 - y_i)\right],$$

   where $\bar{Q}_n^{\epsilon}(a,w) \triangleq \operatorname{logit}^{-1}\left(\operatorname{logit} \bar{Q}_n(a,w) + \epsilon H(a,w)\right)$.

   (b) Define $\bar{Q}_n^* \triangleq \bar{Q}_n^{\epsilon_n}$.

   (c) Estimate $\Psi_{d_n}(P_0)$ using the plug-in estimator given by

$$\Psi_{d_n}(P_n^*) \triangleq \frac{1}{n}\sum_{i=1}^{n}\sum_{a=0}^{1} \bar{Q}_n^*(a, w_i) Pr_{P_U}(d_n(U, v_i) = a).$$

We use $\Psi_{d_n}(P_n^*)$ as our estimate of $\Psi(P_0)$. We will denote this estimator $\hat{\Psi}$, where we have defined $\hat{\Psi}$ so that $\hat{\Psi}(P_n) = \Psi_{d_n}(P_n^*)$. Note that we have used a TMLE for the data dependent parameter $\Psi_{d_n}(P_0)$, which represents the value under a *stochastic* intervention $d_n$. Nonetheless, we assume that $Pr_{P_0}(\bar{Q}_{b,0}(V) = \tau_0) = 0$ for many of the results pertaining to our estimator $\hat{\Psi}$, i.e. we assume that the optimal R-C rule is deterministic. We view estimating the value under a stochastic rather than deterministic intervention as

worthwhile because one can give conditions under which the above estimator is (root-$n$) consistent for $\Psi(P_0)$ at all laws $P_0$, even if non-negligible bias invalidates standard Wald-type confidence intervals for the parameter of interest at laws $P_0$ for which $Pr_{P_0}(\bar{Q}_{b,0}(V) = \tau_0) > 0$.

We will use $P_n^*$ to denote any distribution for which $\bar{Q}_{P_n^*} = \bar{Q}_n^*$, $g_{P_n^*} = g_n$, and $P_n^*$ has the marginal empirical distribution of $W$ for the marginal distribution of $W$. We note that such a distribution $P_n^*$ exists provided that $\bar{Q}_n^*$ and $g_n$ fall in the parameter spaces of $P \mapsto \bar{Q}_P(W)$ and $P \mapsto g_P$, respectively.

In practice we recommend estimating $\bar{Q}_0$ and $\bar{Q}_{b,0}$ using an ensemble method such as super-learning to make an optimal bias-variance trade-off (or, more generally, minimize cross-validated risk) between a mix of parametric models and data adaptive regression algorithms (van der Laan et al., 2007; Luedtke and van der Laan, 2014a). If the treatment mechanism $g_0$ is unknown then we recommend using similar data adpative approaches to obtain the estimate $g_n$. If $g_0$ is known (as in a randomized controlled trial without missingness), then one can either take $g_n = g_0$ or estimate $g_0$ using a correctly specified parametric model, which we expect to increase the efficiency of estimators when the $\bar{Q}_0$ part of the likelihood is misspecified (van der Laan and Robins, 2003; van der Laan and Luedtke, 2014b).

We now outline the main results of this paper, which hold under appropriate consistency and regularity conditions.

- Asymptotic linearity of $\hat{\Psi}$:

$$\hat{\Psi}(P_n) - \Psi(P_0) = \frac{1}{n}\sum_{i=1}^{n} D_0(O_i) + o_{P_0}(n^{-1/2}),$$

  with $D_0$ a known function of $P_0$.

- $\hat{\Psi}$ is an asymptotically efficient estimate of $\Psi(P_0)$.

- One can obtain a consistent estimate $\sigma_n^2$ for the variance of $D_0(O)$. An asymptotically valid 95% confidence intervals for $\Psi(P_0)$ given by $\hat{\Psi}(P_n) \pm 1.96\sigma_n/\sqrt{n}$.

The upcoming sections give the consistency and regularity conditions which imply the above results.

# 4 Canonical gradient of the optimal R-C value

The pathwise derivative of $\Psi$ will provide a key ingredient for analyzing the asymptotic properties of our estimator. We refer the reader to Pfanzagl (1990)

and Bickel et al. (1993) for an overview of the crucial role that the pathwise derivative plays in semiparametric efficiency theory. We remind the reader that an estimator $\hat{\Phi}$ is an asyptotically linear estimator of a parameter $\Phi(P_0)$ with influence curve $IC_{P_0}$ provided that

$$\hat{\Phi}(P_n) - \Phi(P_0) = \frac{1}{n}\sum_{i=1}^{n} IC_{P_0}(O_i) + o_{P_0}(n^{-1/2}).$$

If $\Phi$ is pathwise differentiable with canonical gradient $IC_{P_0}$, then $\hat{\Phi}$ is RAL and asymptotically efficient (minimum variance) among all such RAL estimators of $\Phi(P_0)$ (Pfanzagl, 1990; Bickel et al., 1993).

For $o \in \mathcal{O}$, a deterministic rule $\tilde{d}$, and a real number $\tau$, define

$$D_1(\tilde{d}, P)(o) \triangleq \frac{I(a = \tilde{d}(v))}{g_P(a|w)}\left(y - \bar{Q}_P(a, w)\right)$$
$$D_2(\tilde{d}, P)(o) \triangleq \bar{Q}_P(\tilde{d}(v), w) - E_P\bar{Q}_P(\tilde{d}(V), W),$$

where $g_P(a|W) \triangleq Pr_P(A = a|W)$. We will let $g_0 \triangleq g_{P_0}$. We note that $D_1(\tilde{d}, P) + D_2(\tilde{d}, P)$ is the efficient influence curve of the parameter $\tilde{\Psi}_{\tilde{d}}(P)$.

Let $d$ be some stochastic rule. The canonical gradient of $\Psi_d$ is given by

$$IC_d(P)(o) \triangleq E_{P_U}[D_1(d(U, \cdot), P)(o) + D_2(d(U, \cdot), P)(o)].$$

Define

$$D(d, \tau, P)(o) \triangleq IC_d(P)(o) - \tau\left(E_{P_U}\left[d(U, v)\right] - \kappa\right).$$

For ease of reference, let $D_0 \triangleq D(d_0, \tau_0, P_0)$. The upcoming theorem makes use of the following assumptions.

C1) $g_0$ satisfies the positivity assumption: $Pr_0(0 < g_0(1|W) < 1) = 1$.

C2) $\bar{Q}_{b,0}(W)$ has density $f_0$ at $\eta_0$, and $0 < f_0(\eta_0) < \infty$.

C3) $S_0$ is continuous in a neighborhood of $\eta_0$.

C4) $Pr_0(\bar{Q}_{b,0}(V) = \tau) = 0$ for all $\tau$ in a neighborhood of $\tau_0$.

We now present the canonical gradient of the optimal R-C value.

**Theorem 3.** *Suppose C1) through C4). Then $\Psi$ is pathwise differentiable at $P_0$ with canonical gradient $D_0$.*

Note that C3) implies that $Pr_0(\bar{Q}_{b,0}(V) = \tau_0) = 0$. Thus $d_0$ is (almost surely) deterministic and the expectation over $P_U$ in the definition of $D_0$ is superfluous. Nonetheless, this representation will prove useful when we seek to show that our estimator solves the empirical estimating equation defined by an estimate of $D(d_0, \tau_0, P_0)$.

When the resource constraint is active, i.e. $\tau_0 > 0$, the above theorem shows that $\Psi$ has an additional component over the optimal value parameter when no resource constraints are present (van der Laan and Luedtke, 2014a). The additional component is $\tau_0 \times (E_{P_U}[d_0(U, v)] - \kappa)$, and is the portion of the derivative that relies on the fact that $d_0$ is estimated and falls on the edge of the parameter space. We note that it is possible that the variance of $D_0(O)$ is greater than the variance of $IC_{d_0}(P_0)(O)$. If $\tau_0 = 0$ then these two variances are the same, so suppose $\tau_0 > 0$. Then, provided that $Pr_0(\bar{Q}_{b,0}(V) = \tau_0) = 0$, we have that

$$
\begin{aligned}
&Var_{P_0}\left(D_0(O)\right) - Var_{P_0}\left(IC_{d_0}(P_0)\right) \\
&= \tau_0\kappa(1-\kappa)\left(\tau_0 - 2E_{P_0}\left[\bar{Q}_0(1,W)\middle|\, \tilde{d}_0(V) = 1\right] + 2E_{P_0}\left[\bar{Q}_0(0,W)\middle|\, \tilde{d}_0(V) = 0\right]\right).
\end{aligned}
$$

For any $\kappa \in (0, 1)$, it is possible to exhibit a distribution $P_0$ which satisfies the conditions of Theorem 3 and for which $Var_{P_0}(D_0(O)) > Var_{P_0}(IC_{d_0}(P_0)(O))$. Perhaps more surprisingly, it is also possible to exhibit a distribution $P_0$ which satisfies the conditions of Theorem 3 and for which $Var_{P_0}(D_0(O)) < Var_{P_0}(IC_{d_0}(P_0)(O))$. We will formally show this in a forthcoming work on a data adaptive parameter (see van der Laan et al., 2013; van der Laan and Luedtke, 2014b) which approximately solves (2). We omit the discussion here because the focus of this work is on considering the estimating the value from the optimization problem (2), rather than discussing how this procedure relates to the estimation of other parameters.

# 5 Results about the proposed estimator

We now show that $\hat{\Psi}$ is an asymptotically linear estimator for $\Psi(P_0)$ with influence curve $D_0$ provided our estimates of the needed parts of $P_0$ satisfy consistency and regularity conditions. Our theoretical results are presented in Section 5.1, and the conditions of our main theorem are discussed in Section 5.2.

## 5.1 Inference for $\Psi(P_0)$

For any distributions $P$ and $P_0$ satisfying positivity, stochastic intervention $d$, and real number $\tau$, define the following second-order remainder terms:

$$R_{10}(d,P) \triangleq E_{P_U \times P_0}\left[\left(1 - \frac{g_0(d|W)}{g(d|W)}\right)\left(\bar{Q}_P(d,W) - \bar{Q}_0(d,W)\right)\right]$$
$$R_{20}(d) \triangleq E_{P_U \times P_0}\left[(d-d_0)(\bar{Q}_{b,0}(V) - \tau_0)\right].$$

Above the reliance of $d$ and $d_0$ on $(U,V)$ is omitted in the notation. Let $R_0(d,P) \triangleq R_{10}(d,P) + R_{20}(d)$. The upcoming theorem will make use of the following assumptions.

C5) $g_0$ satisfies the strong positivity assumption: $Pr_0(\delta < g_0(1|W) < 1-\delta) = 1$ for some $\delta > 0$.

C6) $g_n$ satisfies the strong positivity assumption for a fixed $\delta > 0$ with probability approaching 1: there exists some $\delta > 0$ such that, with probability approaching 1, $Pr_0(\delta < g_n(1|W) < 1-\delta) = 1$.

C7) $R_0(d_n, P_n^*) = o_{P_0}(n^{-1/2})$.

C8) $E_{P_0}\left[(D(d_n,\tau_0,P_n^*)(O) - D_0(O))^2\right] = o_{P_0}(1)$.

C9) $D(d_n,\tau_0,P_n^*)$ belongs to a $P_0$-Donsker class $\mathcal{D}$ with probability approaching 1.

C10) $\frac{1}{n}\sum_{i=1}^n D(d_n,\tau_0,P_n^*)(O_i) = o_{P_0}(n^{-1/2})$.

We note that the $\tau_0$ in the final condition above only enters the expression in the sum as a multiplicative constant in front of $-E_{P_U}[d(U,v_i)] - \kappa$.

**Theorem 4** ($\hat{\Psi}$ is asymptotically linear)**.** *Suppose C2) through C10). Then $\hat{\Psi}$ is a RAL estimator of $\Psi(P_0)$ with influence curve $D_0$, i.e.*

$$\hat{\Psi}(P_n) - \Psi(P_0) = \frac{1}{n}\sum_{i=1}^n D_0(O_i) + o_{P_0}(n^{-1/2}).$$

*Further, $\hat{\Psi}$ is efficient among all such RAL estimators of $\Psi(P_0)$.*

Let $\sigma_0^2 \triangleq Var_{P_0}(D_0)$. By the central limit theorem, $\sqrt{n}\left(\hat{\Psi}(P_n) - \Psi(P_0)\right)$ converges in distribution to a $N(0,\sigma_0^2)$ distribution. Let $\sigma_n^2 \triangleq \frac{1}{n}\sum_{i=1}^n D(d_n,\tau_n,P_n^*)(O_i)^2$ be an estimate of $\sigma_0^2$. We now give the following lemma, which gives sufficient conditions for the consistency of $\tau_n$ for $\tau_0$.

**Lemma 5** (Consistency of $\tau_n$)**.** *Suppose C2) and C3). Also suppose $\bar{Q}_{b,n}$ is consistent for $\bar{Q}_{b,0}$ in $L^1(P_0)$ and that the estimate $\bar{Q}_{b,n}$ belongs to a $P_0$ Glivenko Cantelli class with probability approaching* 1*. Then $\tau_n \to \tau_0$ in probability.*

It is easy to verify that conditions similar to those of Theorem 4, combined with the convergence of $\tau_n$ to $\tau_0$ as considered in the above lemma, imply that $\sigma_n \to \sigma_0$ in probability. Under these conditions, an asymptotically valid two-sided $1-\alpha$ confidence interval is given by

$$\hat{\Psi}(P_n) \pm z_{1-\alpha/2}\frac{\sigma_n}{\sqrt{n}},$$

where $z_{1-\alpha/2}$ denotes the $1-\alpha/2$ quantile of a $N(0,1)$ random variable.

## 5.2 Discussion of conditions of Theorem 4

**Conditions C2) and C3).** These are standard conditions used when attempting to estimate the $\kappa$-quantile $\eta_0$, defined in (3). Provided good estimation of $\bar{Q}_{b,0}$, these conditions ensure that gathering a large amount of data will enable one to get a good estimate of the $\kappa$-quantile of the random variable $\bar{Q}_{b,0}$. See Lemma 5 for an indication of what is meant by "good estimation" of $\bar{Q}_{b,0}$. It seems reasonable to expect that these conditions will hold when $V$ contains continuous random variables and $\eta_0 \neq 0$, since we are essentially assuming that $\bar{Q}_{b,0}$ is not degenerate at the arbirtrary (determined by $\kappa$) point $\eta_0$.

**Condition C4).** If $\tau_0 > 0$, then C4) is implied by C3). If $\tau_0 = 0$, then C4) is like assuming a non-exceptional law, i.e. that the probability of a there being no treatment effect in a strata of $V$ is zero. Because $\tau_0$ is not known from the outset, we require something slightly stronger, namely that the probability of *any specific* small treatment effect is zero in a strata of $V$ is zero. Note that this condition does *not* prohibit the treatment effect from being small, e.g. $Pr_0(|\bar{Q}_{b,0}(V)| < \tau) > 0$ for all $\tau > 0$, but rather it prohibits there existing a sequence $\tau_m \downarrow 0$ with the property that $Pr_0(\bar{Q}_{b,0}(V) = \tau_m) > 0$ infinitely often. Thus this condition does not really seem any stronger than assuming a non-exceptional law. If one is concerned about such exceptional laws then we suggest adapting the methods in (Luedtke and van der Laan, 2014b) to the R-C setting.

**Condition C5).** This condition assumes that people from each strata of covariates have a reasonable (at least a $\delta > 0$) probability of treatment.

**Condition C6).** This condition requires that our estimates of $g_0$ respect the fact that each strata of covariates has a reasonable probability of treatment.

**Condition C7).** This condition is satisfied if $R_{10}(d_n, P_n^*) = o_{P_0}(n^{-1/2})$ and $R_{20}(d_n) = o_{P_0}(n^{-1/2})$. The term $R_{10}(d_n, P_n^*)$ takes the form of a typical double robust term that is small if either $g_0$ or $\bar{Q}_0$ is estimated well, and is second-order, i.e. one might hope that $R_{10}(d_n, P_n^*) = o_{P_0}(n^{-1/2})$, if both both $g_0$ and $\bar{Q}_0$ are estimated well. One can upper bound this remainder with a product of the $L^2(P_0)$ rates of convergence of these two quantities using the Cauchy-Schwarz inequality. If $g_0$ is known, then one can take $g_n = g_0$ and this term is zero.

Ensuring that $R_{20}(d_n) = o_{P_0}(n^{-1/2})$ requires a little more work but will still prove to be a reasonable condition. We will use the following margin assumption for some $\alpha > 0$:

$$Pr_0\left(0 < |\bar{Q}_{b,0} - \tau_0| \leq t\right) \lesssim t^\alpha \text{ for all } t > 0, \tag{5}$$

where "$\lesssim$" denotes less than or equal to up to a multiplicative constant. This margin assumption is analogous to that used in Audibert and Tsybakov (2007). The following result relates the rate of convergence of $R_{20}(d_n)$ to the rate at which $\bar{Q}_{b,n} - \tau_n$ converges to $\bar{Q}_{b,0} - \tau_0$.

**Theorem 6.** *If (5) holds for some $\alpha > 0$, then*

*i)* $|R_{20}(d_n)| \lesssim \left\|(\bar{Q}_{b,n} - \tau_n) - (\bar{Q}_{b,0} - \tau_0)\right\|_{2,P_0}^{2(1+\alpha)/(2+\alpha)}$

*ii)* $|R_{20}(d_n)| \lesssim \left\|(\bar{Q}_{b,n} - \tau_n) - (\bar{Q}_{b,0} - \tau_0)\right\|_{\infty,P_0}^{1+\alpha}$.

The above is similar to Lemma 5.2 in Audibert and Tsybakov (2007), and a similar result was proved in the context of optimal dynamic treatment regimes without resource constraints in Luedtke and van der Laan (2014b). If $S_0$ has a finite derivative at $\tau_0$, as is given by C2), then one can take $\alpha = 1$. The above theorem then implies that $R_{20}(d_n) = o_{P_0}(n^{-1/2})$ if either $\left\|(\bar{Q}_{b,n} - \tau_n) - (\bar{Q}_{b,0} - \tau_0)\right\|_{2,P_0}$ is $o_{P_0}(n^{3/8})$ or $\left\|(\bar{Q}_{b,n} - \tau_n) - (\bar{Q}_{b,0} - \tau_0)\right\|_{\infty,P_0}$ is $o_{P_0}(n^{1/4})$.

**Condition C8).** This is a mild consistency condition which is implied by the $L^2(P_0)$ consistency of $d_n$, $g_n$, and $\bar{Q}_n^*$ to $d_0$, $g_0$, and $\bar{Q}_0$. We note that the consistency of the intial (unfluctuated) estimate $\bar{Q}_n$ for $\bar{Q}_0$ will imply the consistency of $\bar{Q}_n^*$ to $\bar{Q}_0$ given C6), since in this case $\epsilon_n \to 0$ in probability,

and thus $\left\|\bar{Q}_n^* - \bar{Q}_n\right\|_{2,P_0} \to 0$ in probability.

**Condition C9).** This condition places restrictions on how data adaptive the estimators of $d_0$, $g_0$, and $\bar{Q}_0$ can be. We refer the reader to Section 2.10 of van der Vaart and Wellner (1996) for conditions under which the estimates of $d_0$, $g_0$, and $\bar{Q}_0$ belonging to Donsker classes implies that $D(d_n, \tau_0, P_n^*)$ belongs to a Donsker class. We note that this condition was avoided for estimating the value function using a cross-validated TMLE in van der Laan and Luedtke (2014b) and using an online estimator of the value function in Luedtke and van der Laan (2014b), and using either technique will allow one to avoid the condition here as well.

**Condition C10).** Using the notation $Pf = \int f(o)dP(o)$ for any distribution $P$ and function $f : \mathcal{O} \to \mathbb{R}$, we have that

$$P_n D(d_n, \tau_0, P_n^*) = P_n D_1(d_n, P_n^*) + P_n D_2(d_n, P_n^*) - \tau_0 \left( \frac{1}{n} \sum_{i=1}^{n} E_{P_U}[d_n(U, v_i)] - \kappa \right).$$

The first term is zero by the fluctuation step of the TMLE algorithm and the second term on the right is zero because $P_n^*$ uses the empirical distribution of $W$ for the marginal distribution of $W$. If $\tau_0 = 0$ then clearly the third term is zero, so suppose $\tau_0 > 0$. Combining (4) and the fact that $d_n$ is a substitution estimator shows that the third term is 0 with probability approaching 1 provided that $\tau_n > 0$ with probability approaching 1. This will of course occur if $\tau_n \to \tau_0 > 0$ in probability, for which Lemma 5 gives sufficient conditions.

# 6 Simulation methods

We simulated i.i.d. draws from two data generating distributions at sample sizes 100, 200, and 1000. For each sample size and distribution we considered resource constraints $\kappa = 0.1$ and $\kappa = 0.9$. We ran 2000 Monte Carlo draws of each simulation setting. All simulations were run in `R` (R Core Team, 2014).

We first present the two data generating distributions considered, and then present the estimation strategies used.

## 6.1 Data generating distributions

### Simulation 1

Our first data generating distribution is identical to the single time point simulation considered in van der Laan and Luedtke (2014b) and Luedtke and van der Laan (2014a). The outcome is binary and the baseline covariate vector $W = (W_1, ..., W_4)$ is four dimensional for this distribution, with

$$\begin{aligned}
&W_1, W_2, W_3, W_4 \overset{i.i.d.}{\sim} N(0,1) \\
&A|W \sim \text{Bernoulli}(1/2) \\
&\text{logit}\left(E_{P_0}\left[Y|A,W,H=0\right]\right) = 1 - W_1^2 + 3W_2 + A\left(5W_3^2 - 4.45\right) \\
&\text{logit}\left(E_{P_0}\left[Y|A,W,H=1\right]\right) = -0.5 - W_3 + 2W_1W_2 + A\left(3|W_2| - 1.5\right),
\end{aligned}$$

where $H$ is an unobserved Bernoulli(1/2) variable independent of $A, W$. For this distribution $E_{P_0}[\bar{Q}_0(0,W)] \approx E_{P_0}[\bar{Q}_0(1,W)] \approx 0.464$.

We consider two choices for $V$, namely $V = W_3$, and $V = W_1, ..., W_4$. We obtained estimates of the approximate optimal R-C optimal value for this data generating distribution using $10^7$ Monte Carlo draws. When $\kappa = 0.1$, $\Psi(P_0) \approx 0.493$ for $V = W_3$ and $\Psi(P_0) \approx 0.511$ for $V = W_1, ..., W_4$. When $\kappa = 0.9$, $\Psi(P_0) \approx 0.536$ for $V = W_3$ and $\Psi(P_0) \approx 0.563$ for $V = W_1, ..., W_4$. We note that the resource constraint is not active ($\tau_0 = 0$) when $\kappa = 0.9$ for either choice of $V$.

### Simulation 2

Our second data generating distribution is a very similar to one of the distributions considered in (Luedtke and van der Laan, 2014b), though has been modified so that the treatment effect is positive for all values of the covariate. The data is generated as follows:

$$\begin{aligned}
&W \sim \text{Uniform}(-1,1) \\
&A|W \sim \text{Bernoulli}(1/2) \\
&Y|A,W \sim \text{Bernoulli}(\bar{Q}_0(A,W)),
\end{aligned}$$

where for $\tilde{W} \triangleq W + 5/6$ we define

$$\bar{Q}_0(A,W) - \frac{6}{10} \triangleq \begin{cases} 0, & \text{if } A = 1 \text{ and } -1/2 \leq W \leq 1/3 \\ -\tilde{W}^3 + \tilde{W}^2 - \frac{1}{3}\tilde{W} + \frac{1}{27}, & \text{if } A = 1 \text{ and } W < -1/2 \\ -W^3 + W^2 - \frac{1}{3}W + \frac{1}{27}, & \text{if } A = 1 \text{ and } W > 1/3 \\ -\frac{3}{10}, & \text{otherwise.} \end{cases}$$

For this distribution $E_{P_0}[\bar{Q}_0(0,W)] = 0.3$ and $E_{P_0}[\bar{Q}_0(1,W)] \approx 0.583$.

We use $V = W$. This simulation is an example of a case where $\bar{Q}_{b,0}(V) > 0$ almost surely, so any constraint on resources will reduce the optimal value from its unconstrained value of 0.583. In particular, we have that $\Psi(P_0) \approx 0.337$ when $\kappa = 0.1$ and $\Psi(P_0) \approx 0.572$ when $\kappa = 0.9$.

## 6.2 Estimating nuisance functions

We treated $g_0$ as known in both simulations and let $g_n = g_0$. We estimated $\bar{Q}_0$ using the super-learner algorithm with the quasi-log-likelihood loss function (`family=binomial`) and a candidate library of data adaptive (`SL.gam` and `SL.nnet`) and parametric algorithms (`SL.bayesglm`, `SL.glm`, `SL.glm.interaction`, `SL.mean`, `SL.step`, `SL.step.interaction`, and `SL.step.forward`). We refer the reader to Table 2 in the technical report Luedtke and van der Laan (2014a) for a brief description of these algorithms. We estimated $\bar{Q}_{b,0}$ by running a super-learner using the squared error loss function and the same candidate algorithms and used $W$ to predict the outcome $\tilde{Y} \triangleq \frac{2A-1}{g_0(A|W)}(Y - \bar{Y}_n) + \bar{Y}_n$, where $\bar{Y}_n$ represents the sample mean of $Y$ from the $n$ observations. See Luedtke and van der Laan (2014a) for a justification of this estimation scheme.

Once we had our estimates $\bar{Q}_n$, $\bar{Q}_{b,n}$, and $g_n$ we proceeded with the estimation strategy described in Section 3.

## 6.3 Evaluating performance

We used three methods to evaluate our proposed approach. First, we looked at the coverage of two-sided 95% confidence intervals for the optimal R-C value. Second, we report the average confidence interval widths. Finally, we looked at the power of the $\alpha = 0.025$ level test $H_0 : \Psi(P_0) = \mu_0$ against $H_1 : \Psi(P_0) > \mu_0$, where $\mu_0 \triangleq E_0[\bar{Q}_0(0,W)]$ is treated as a known quantity. Under causal assumptions, $\mu_0$ can be identified with the counterfactual quantity representing the population mean outcome if, possibly contrary to fact, no one receives treatment. In which treatment is not currently being implemented, one could substitute the population mean outcome (if known) for $\mu_0$. Our test of significance consisted of checking of the lower bound in the two-sided 95% confidence interval is greater than $\mu_0$. If an estimator of $\Psi(P_0)$ is low-powered in testing $H_0$ against $H_1$ then clearly the estimator will have little practical value.

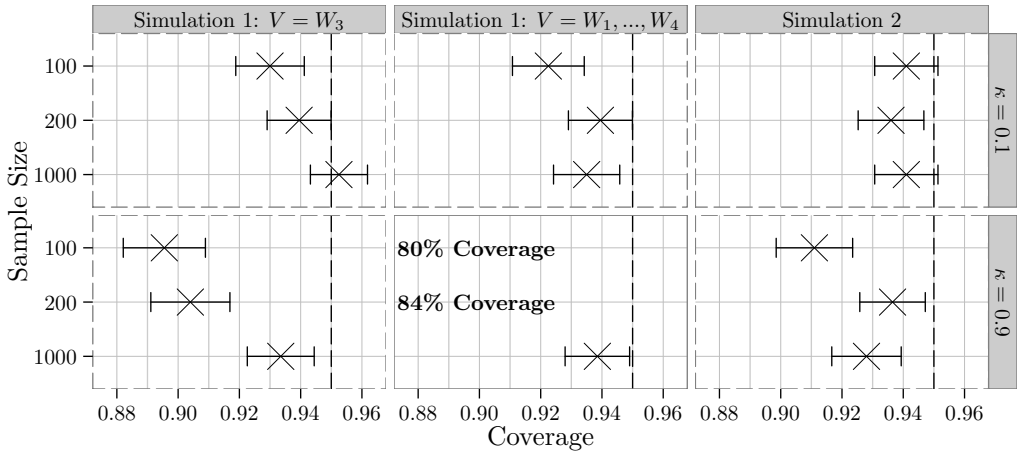


Figure 1: Coverage of two-sided 95% confidence intervals. As expected, coverage increases with sample size. The coverage tends to be better for $\kappa = 0.1$ than for $\kappa = 0.9$, though the estimator performed well at the largest sample size (1000) for all simulations and choices of $\kappa$. Error bars indicate 95% confidence intervals to account for uncertainty from the finite number of Monte Carlo draws.

# 7 Simulation results

The proposed estimation strategy performed well overall. Figure 1 demonstrates the coverage of 95% confidence intervals for the optimal R-C value. All methods performed well at all sample sizes for the highly constrained setting where $\kappa = 0.1$. The results were more mixed for the resource constraint $\kappa = 0.9$. All methods performed well at the largest sample size considered. This supports our theoretical results, which were all asymptotic in nature. For Simulation 1, in which the resource constraint was not active for either choice of $V$, the coverage dropped off at lower sample sizes. Coverage was approximately 90% in the two smaller sample sample sizes for $V = W_3$, which may be expected for such an asymptotic method. For the more complex problem of estimating the optimal value when $V = W_1, ..., W_4$ the coverage was somewhat lower (80% when $n = 100$ and 84% when $n = 200$). In Simulation 2, the coverage was better (>91%) for the smaller sample sizes. We note that the resource constraint was still active ($\tau_0 > 0$) when $\kappa = 0.9$ for this simulation, and also that the estimation problem is easier because the baseline covariate was univariate.

We report the average confidence interval widths across the 2000 Monte

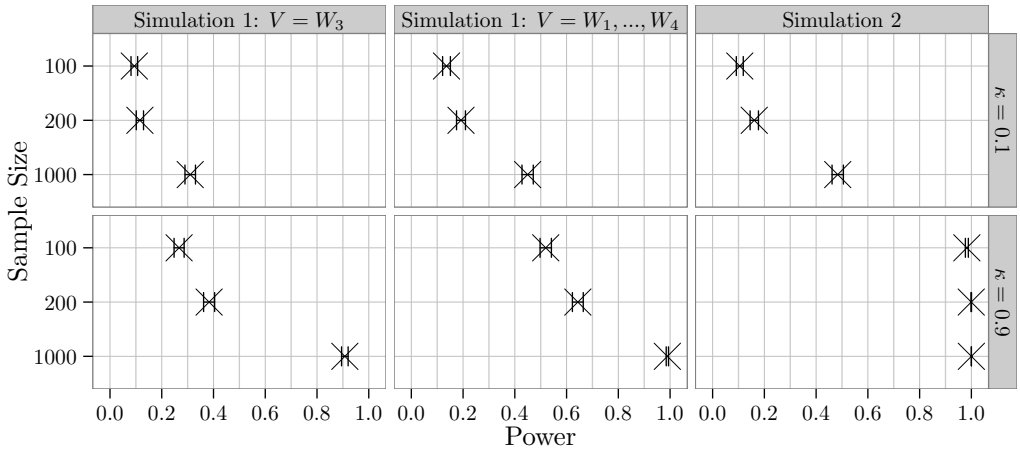


Figure 2: Power of the $\alpha = 0.025$ level test of $H_0 : \Psi(P_0) = \mu_0$ against $H_1 : \Psi(P_0) > \mu_0$, where $\mu_0 = E_{P_0}[\bar{Q}_0(0, W)]$ is treated as known. Power increases with sample size and $\kappa$. Error bars indicate 95% confidence intervals to account for uncertainty from the finite number of Monte Carlo draws.

Carlo draws. For $n = 100$, average confidence interval widths were between 0.25 and 0.26 across all simulations and choices of $\kappa$. For $n = 200$, all average confidence interval widths were between 0.17 and 0.18. For $n = 1000$, all average confidence interval widths were approximately 0.08. We note that the usefulness of such confidence intervals varies across simulations and choices of $\kappa$. When $V = W_3$ and $\kappa = 0.1$ in Simulation 1, the optimal R-C value is approximately 0.493, versus a baseline value $\mu_0 = E_{P_0}[\bar{Q}_0(0, W)]$ of approximately 0.464. Thus here the confidence interval would give the investigator little information, even at a sample size of 1000. In Simulation 2 with $\kappa = 0.9$, on the other hand, the optimal R-C value is approximately 0.572, versus a baseline value of $\mu_0 \approx 0.3$. Thus here all confidence intervals would likely be informative for investigators, even those made for data sets of size 100.

Figure 2 gives the power of the $\alpha = 0.025$ level test $H_0 : \Psi(P_0) = \mu_0$ against the alternative $H_1 : \Psi(P_0) > \mu_0$. Overall our method appears to have reasonable power in this statistical test. We see that power increases with sample size, the key property of consistent statistical tests. We also see that power increases with $\kappa$, which is unsurprising given that $Y$ is binary and $g_0(a|w)$ is 1/2 for all $a, w$. We note that power will not always increase with $\kappa$, for example if $P_0$ is such that $g_0(1|w)$ is very small for individuals with covariate $w$ who are treated at $\kappa = 0.9$ but not at $\kappa = 0.1$. This observation is not meant as a criticism to the estimation scheme that we have presented because

we assume that $\kappa$ will be chosen to reflect real resource constraints, rather than to maximize the power for a test $H_0' : \Psi(P_0) = \mu'$ versus $H_1' : \Psi(P_0) > \mu'$ for some fixed $\mu'$.

We also implemented an estimating equation based estimator for the optimal R-C value and found the two methods performed similarly. We would recommend using the TMLE in practice because it has been shown to be robust to near positivity violations in a wide variety of settings (van der Laan and Rose, 2011). We note that $g_0(1|w) = 1/2$ for all $w$ in both of our simulations, so no near positivity violations occurred. We do not consider the estimation equation approach any further here because the focus of this work is on considering the optimization problem (2), rather than on comparing different estimation frameworks.

# 8 Discussion and future work

We have considered the problem of estimating the optimal resource-constrained value. Under causal assumptions, this parameter can be identified with the maximum attainable population mean outcome under dynamic treatment rules which rely on some summary of measured covariates, subject to the constraint that a maximum proportion $\kappa$ of the population can be treated. We also provided an explicit expression for an optimal stochastic rule under the resource constraint.

We derived the canonical gradient of the optimal R-C value under the key assumption that the treatment effect is not exactly equal to $\tau_0$ in some strata of covariates which occurs with positive probability. The canonical gradient plays a key role in developing asymptotically linear estimators. We found that the canonical gradient of the optimal R-C value has an additional component when compared to the canonical gradient of the optimal unconstrained value when the resource constraint is active, i.e. when $\tau_0 > 0$.

We presented a targeted minimum loss-based estimator for the optimal R-C value. This estimator was designed to solve the empirical mean of an estimate of the canonical gradient. This quickly yielded conditions under which our estimator is RAL, and efficient among all such RAL estimators. All of these results rely on the condition that the treatment effect is not exactly equal to $\tau_0$ for positive probability strata of covariates. This assumption is more plausible than the typical non-exceptional law assumption when the covariates are continuous and the constraint is active because it may be unlikely that the treatment effect concentrates on an arbitrary (determined by $\kappa$) $\tau_0 > 0$. We note that this pseudo-non-exceptional law assumption has implied that the

optimal stochastic rule is almost surely equal to the optimal deterministic rule. Though we have not presented formal theorems here, it is not difficult to derive conditions under which our estimator of the optimal value under a R-C stochastic rule is (root-$n$) consistent even when the treatment effect is equal to $\tau_0$ with positive probability, though the bias will be non-negligible. One could use an analogue of the variance-stabilized online estimator presented in Luedtke and van der Laan (2014b) to get inference for the optimal R-C value in this setting.

Our simulations confirmed our theoretical findings. We found that coverage improved with sample size, with near-nominal coverage at the largest sample size considered. This is not surprising given that most of our analytic results were asymptotic, though we note that the method also performed well at the smaller sample sizes considered. The confidence intervals were informatively tight when one considered the difference between the optimal R-C value and the value under no treatment. Further simulations are needed to fully understand the behavior of this method in practice.

Some resource constraints encountered in practice may not be of the form $E_{P_U \times P_0}[d(U,V)] \leq \kappa$. For example, the cost of distributing the treatment to people may vary based on the values of the covariates. For simplicity assume $V = W$. If $c : \mathcal{W} \to [0, \infty)$ is a cost function, then this constraint may take the form $E_{P_U \times P_0}[c(W)d(U,W)] \leq \kappa$. If $\tau_0 = 0$, then an optimal stochastic rule under such a constraint takes the form $(u,w) \mapsto I(\bar{Q}_{b,0}(w) > 0)$. If $\tau_0 > 0$, then an optimal stochastic rule under such a constraint takes the form $(u,w) \mapsto I(\bar{Q}_{b,0}(w) > \tau_0 c(w))$ for $w$ for which $\bar{Q}_{b,0}(w) \neq \tau_0 c(w)$ or $c(w) = 0$, and randomly distributes the remaining resources uniformly among all remaining $w$. We leave further consideration of this more general resource constraint problem to future work.

In this work our primary focus has been on estimating the optimal value under a resource constraint, rather than the optimal rule under a resource constraint. Nonetheless, our estimation procedure yields an estimate $d_n$ of the optimal R-C rule. It would be interesting to further analyze $d_n$ in future work to better understand how well this estimator will perform, or if there are better estimators which more directly frame the estimation challenge as a (weighted) classification problem (Zhao et al., 2012; Rubin and van der Laan, 2012). Note that we are not guaranteed that $d_n$ satisfies the constraint, i.e. it is quite possible that $E_{P_U \times P_0}[d_n(U,V)] > \kappa$, though concentration inequalities suggest that one can give conditions under which $E_{P_U \times P_0}[d_n(U,V)] - \kappa$ is small with probability approaching 1. One could also seek an optimal rule estimate $d_n'$ which satisfies that, with probability at least $1 - \delta$ for some user-defined $\delta > 0$, $E_{P_U \times P_0}[d_n'(U,V)] \leq \kappa$.

We have not considered the ethical considerations associated with allocating limiting resources to a population. The debate over the appropriate means to distribute limited treatment resources to a population is ongoing (see, e.g., Brock and Wikler, 2009; Macklin and Cowan, 2012; Singh, 2013, for examples in the treatment of HIV/AIDS). Clearly any investigator needs to consider the ethical issues associated with certain resource allocation schemes, though we leave such considerations to experts on the matter. It will be interesting to see if there are settings in which it is possible to transform the outcome so that the statistical problem considered in this paper adheres to the ethical guidelines in those settings.

We have looked to generalize previous works in estimating the value of an optimal dynamic treatment regime to the case where the treatment resource is a limited resource, i.e. where it is not possible to treat the entire population. This work should allow for the application of optimal personalized treatment strategies to many new problems of interest.

# Acknowledgement

This research was supported by NIH grant R01 AI074345-06. AL was supported by the Department of Defense (DoD) through the National Defense Science & Engineering Graduate Fellowship (NDSEG) Program.

# References

E Arjas, C Jennison, A P Dawid, D R Cox, S Senn, R G Cowell, V Didelez, Gill R D, J B Kadane, J M Robins, and S A Murphy. Optimal dynamic treatment regimes - Discussion on the paper by Murphy. *J. R. Stat. Soc. Ser. B*, 65:355—-366, 2003.

J Y Audibert and A B Tsybakov. Fast learning rates for plug-in classifiers. *Ann. Statist.*, 35(2):608–633, 2007.

P J Bickel, C A J Klaassen, Y Ritov, and J A Wellner. *Efficient and adaptive estimation for semiparametric models*. Johns Hopkins University Press, Baltimore, 1993.

D W Brock and D Wikler. Ethical challenges in long-term funding for HIV/AIDS. *Health Aff.*, 28(6):1666–1676, 2009.

B Chakraborty and E E Moodie. *Statistical Methods for Dynamic Treatment Regimes*. Springer, Berlin Heidelberg New York, 2013.

B Chakraborty, E B Laber, and Y-Q Zhao. Inference about the expected performance of a data-driven dynamic treatment regime. *Clin. Trials*, 11 (4):408–417, 2014.

G B Dantzig. Discrete-variable extremum problems. *Oper. Res.*, 5(2):266–288, 1957.

I Díaz and M J van der Laan. Population intervention causal effects based on stochastic interventions. *Biometrics*, 68(2):541–549, 2012.

Y Goldberg, R Song, D Zeng, and M R Kosorok. Comment on Dynamic treatment regimes: Technical challenges and applications. *Electron. J. Stat.*, 8:1290–1300, 2014.

R M Karp. *Reducibility among combinatorial problems*. Springer, New York Berlin Heidelberg, 1972.

B Korte and J Vygen. *Combinatorial optimization*. Springer, Berlin Heidelberg New York, 5th edition, 2012.

A Lasry, S L Sansom, K A Hicks, and V Uzunangelov. A model for allocating CDCs HIV prevention resources in the United States. *Health Care Manag. Sci.*, 14(1):115–124, 2011.

A R Luedtke and M J van der Laan. Super-learning of an optimal dynamic treatment rule. Technical Report 326, available at `http://www.bepress.com/ucbbiostat/`, Division of Biostatistics, University of California, Berkeley, under review at JCI, 2014a.

A R Luedtke and M J van der Laan. Statistical inference for the mean outcome under a possibly non-unique optimal treatment strategy. Technical Report 332, available at http://biostats.bepress.com/ucbbiostat/paper332/, Division of Biostatistics, University of California, Berkeley, submitted to *Annals of Statistics*, 2014b.

R Macklin and E Cowan. Given financial constraints, it would be unethical to divert antiretroviral drugs from treatment to prevention. *Health Aff.*, 31 (7):1537–1544, 2012.

S A Murphy. Optimal dynamic treatment regimes. *J. R. Stat. Soc. Ser. B*, 65 (2):331–336, 2003.

J Pearl. *Causality: Models, Reasoning and Inference*. Cambridge University Press, New York, 2nd edition, 2009.

J Pfanzagl. *Estimation in semiparametric models*. Springer, Berlin Heidelberg New York, 1990.

R Core Team. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014. URL `http://www.r-project.org/`.

J M Robins. A graphical approach to the identification and estimation of causal parameters in mortality studies with sustained exposure periods. *Comput. Math. Appl.*, 14(9-12):139s—-161s, 1987. ISSN 0097-4943.

J M Robins. Optimal structural nested models for optimal sequential decisions. In D Y Lin and Heagerty P, editors, *Proc. Second Seattle Symp. Biostat.*, volume 179, pages 189–326, 2004.

D B Rubin and M J van der Laan. Statistical issues and limitations in personalized medicine research with clinical trials. *Int. J. Biostat.*, 8:Issue 1, Article 18, 2012.

J A Singh. Antiretroviral resource allocation for HIV prevention. *AIDS*, 27 (6):863–865, 2013.

G Tao, K Zhao, T Gift, F Qiu, and G Chen. Using a resource allocation model to guide better local sexually transmitted disease control and prevention programs. *Oper. Res. Heal. Care*, 1(2):23–29, 2012.

M J van der Laan and A R Luedtke. Targeted learning of an optimal dynamic treatment, and statistical inference for its mean outcome. Technical Report 329, available at `http://www.bepress.com/ucbbiostat/`, Division of Biostatistics, University of California, Berkeley, 2014a.

M J van der Laan and A R Luedtke. Targeted learning of the mean outcome under an optimal dynamic treatment rule. *J. Causal Inference*, (Ahead of print), 2014b. doi: 10.1515/jci-2013-0022.

M J van der Laan and J M Robins. *Unified methods for censored longitudinal data and causality*. Springer, New York Berlin Heidelberg, 2003.

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

M J van der Laan, E Polley, and A Hubbard. Super Learner. *Stat Appl Genet Mol*, 6(1):Article 25, 2007. ISSN 1.

M J van der Laan, A E Hubbard, and S Kherad. Statistical inference for data adaptive target parameters. Technical Report 314, Division of Biostatistics, University of California, Berkeley, 2013.

A W van der Vaart and J A Wellner. *Weak convergence and empirical processes*. Springer, Berlin Heidelberg New York, 1996.

B Zhang, A Tsiatis, M Davidian, M Zhang, and E Laber. A robust method for estimating optimal treatment regimes. *Biometrics*, 68:1010–1018, 2012.

Y Zhao, D Zeng, A Rush, and M Kosorok. Estimating individual treatment rules using outcome weighted learning. *J. Am. Stat. Assoc.*, 107:1106–1118, 2012.

# Appendix: Proofs

## Proofs for Section 2

We first state a simple lemma.

**Lemma A.1.** *For a distribution $P$ and a stochastic rule $d$, we have the following representation for $\Psi_d$:*

$$\Psi_d(P) \triangleq E_{P_U \times P}\left[d(U,V)\bar{Q}_{b,P}(V)\right] + E_P[\bar{Q}_P(0,W)].$$

*Proof of Lemma A.1.* We have that

$$\begin{aligned}\Psi_d(P) &= E_{P_U\times P}[d(U,V)\bar{Q}_P(1,W)] + E_{P_U\times P}[(1-d(U,V))\bar{Q}_P(0,W)] \\ &= E_{P_U\times P}[d(U,V)(\bar{Q}_P(1,W) - \bar{Q}_P(0,W))] + E_P[\bar{Q}_P(0,W)] \\ &= E_{P_U\times P}[d(U,V)\bar{Q}_{b,P}(V)] + E_P[\bar{Q}_P(0,W)],\end{aligned}$$

where the final equality holds by the law of total expectation. □

*Proof of Theorem 1.* This result will be a consequence of Theorem 2. If $Pr_P(\bar{Q}_{b,0}(V) = \tau_P) = 0$, then $d_P(U,V)$ is $P_U \times P$ almost surely equal to $\tilde{d}_P(V)$, and thus $\tilde{\Psi}_{\tilde{d}_P}(P) = \Psi_{d_P}(P)$. Thus $(u,v) \mapsto \tilde{d}_P(v)$ is an optimal stochastic regime. Because the class of deterministic regimes is a subset of the class of stochastic regimes, $\tilde{d}_P$ is an optimal deterministic regime. □

*Proof of Theorem 2.* Let $d$ be some stochastic treatment rule which satisfies the resource constraint. For $(b,c) \in \{0,1\}^2$, define $B_{bc} \triangleq \{(u,v) : d_P(u,v) = b, d(u,v) = c\}$. Note that

$$\begin{aligned}\Psi_{d_P}(P) - \Psi_d(P) &= E_{P_U\times P}\left[(d_P(U,V) - d(U,V))\,\bar{Q}_{b,0}(V)\right] \\ &= E_{P_U\times P}\left[\bar{Q}_{b,0}(V)I((U,V)\in B_{10})\right] - E_{P_U\times P}\left[\bar{Q}_{b,0}(V)I((U,V)\in B_{01})\right]\end{aligned} \tag{A.1}$$

The $\bar{Q}_{b,0}(V)$ in the first term in (A.1) can be upper bounded by $\tau_P$, and in the second term can be lower bounded by $\tau_P$. Thus,

$$\begin{aligned}\Psi_{d_P}(P) - \Psi_d(P) &\geq \tau_P\left[Pr_{P_U\times P}\left((U,V)\in B_{10}\right) - Pr_{P_U\times P}\left((U,V)\in B_{01}\right)\right] \\ &= \tau_P\left[Pr_{P_U\times P}\left((U,V)\in B_{10}\cup B_{11}\right) - Pr_{P_U\times P}\left((U,V)\in B_{01}\cup B_{11}\right)\right] \\ &= \tau_P\left(E_{P_U\times P}\left[d_P(U,V)\right] - E_{P_U\times P}\left[d(U,V)\right]\right).\end{aligned}$$

If $\tau_P = 0$ then the final line is zero. Otherwise, $E_{P_U\times P}[d_P(U,V)] = \kappa$ by (4). Because $d$ satisfies the resource constraint, $E_{P_U\times P}\left[d(U,V)\right] \leq \kappa$ and thus the final line above is at least zero. Thus $\Psi_{d_P}(P) - \Psi_d(P) \geq 0$ for all $\tau_P$. Because $d$ was arbitrary, $d_P$ is an optimal stochastic rule. □

## Proofs for Section 4

*Proof of Theorem 3.* The pathwise derivative of $\Psi(Q)$ is defined as $\frac{d}{d\epsilon}\Psi(Q(\epsilon))\big|_{\epsilon=0}$ along paths $\{P_\epsilon : \epsilon\} \subset \mathcal{M}$. In particular, these paths are chosen so that

$$\begin{aligned} dQ_{W,\epsilon} &= (1 + \epsilon H_W(W))dQ_W, \\ &\quad \text{where } EH_W(W) = 0 \text{ and } C_W \triangleq \sup_w |H_W(w)| < \infty; \\ dQ_{Y,\epsilon}(Y \mid A, W) &= (1 + \epsilon H_Y(Y \mid A, W))dQ_Y(Y \mid A, W), \\ &\quad \text{where } E(H_Y \mid A, W) = 0 \text{ and } \sup_{w,a,y} |H_Y(y \mid a, w)| < \infty. \end{aligned}$$

The parameter $\Psi$ is not sensitive to fluctuations of $g_0(a|w) = Pr_0(a|w)$, and thus we do not need to fluctuate this portion of the likelihood. Let $\bar{Q}_{b,\epsilon} \triangleq \bar{Q}_{b,P_\epsilon}$, $\bar{Q}_\epsilon \triangleq \bar{Q}_{P_\epsilon}$, $d_\epsilon \triangleq d_{P_\epsilon}$, $\eta_\epsilon \triangleq \eta_{P_\epsilon}$, $\tau_\epsilon \triangleq \tau_{P_\epsilon}$, and $S_\epsilon \triangleq S_{P_\epsilon}$. First note that

$$\bar{Q}_{b,\epsilon}(v) = \bar{Q}_{b,0}(v) + \epsilon h_\epsilon(v) \tag{A.2}$$

for an $h_\epsilon$ with

$$\sup_{|\epsilon|<1} \sup_v |h_\epsilon(v)| \triangleq C_1 < \infty. \tag{A.3}$$

Note that C4) implies that $d_0$ is (almost surely) deterministic, i.e. $d_0(U, \cdot)$ is almost surely a fixed function. Let $\tilde{d}$ represent the deterministic rule $v \mapsto I(\bar{Q}_{b,0}(v) > 0)$ to which $d(u, \cdot)$ is (almost surely) equal for all $u$. By Lemma A.1,

$$\begin{aligned} \Psi(P_\epsilon) - \Psi(P_0) &= \int_w \left(E_{P_U}[d_\epsilon(U, V)] - \tilde{d}_0(V)\right) \bar{Q}_{b,\epsilon} dQ_{W,\epsilon} \\ &\quad + \int_w \tilde{d}_0(V) \left(\bar{Q}_{b,\epsilon} dQ_{W,\epsilon} - \bar{Q}_{b,0} dQ_{W,0}\right) \\ &\quad + E_{P_\epsilon}\bar{Q}_\epsilon(0, W) - E_{P_0}\bar{Q}_0(0, W) \\ &= \int_w \left(E_{P_U}[d_\epsilon(U, V)] - \tilde{d}_0(V)\right) \left(\bar{Q}_{b,\epsilon} - \tau_0\right) dQ_{W,\epsilon} \\ &\quad + \tau_0 \int_w \left(E_{P_U}[d_\epsilon(U, V)] dQ_{W,\epsilon} - \tilde{d}_0(V) dQ_{W,0}\right) \\ &\quad - \tau_0 \int_w \tilde{d}_0(V) \left(dQ_{W,\epsilon} - dQ_{W,0}\right) \\ &\quad + \Psi_{d_0}(P_\epsilon) - \Psi_{d_0}(P_0). \end{aligned} \tag{A.4}$$

Dividing the fourth term by $\epsilon$ and taking the limit as $\epsilon \to 0$ gives the pathwise derivative of the mean outcome under the rule that treats $d_0$ as known. The

third term can be written as $-\epsilon\tau_0 \int_w \tilde{d}_0(V) H_W dQ_{W,0}$, and thus the pathwise derivative of this term is $-\int_w \tau_0 \tilde{d}_0(V) H_W dQ_{W,0}$. If $\tau_0 > 0$, then $E_{P_U \times P_0}[\tilde{d}_0(V)] = \kappa$. The pathwise derivative of this term is zero if $\tau_0 = 0$. Thus, for all $\tau_0$,

$$\lim_{\epsilon \to 0} -\frac{1}{\epsilon}\tau_0 \int_w \tilde{d}_0(V) \left(dQ_{W,\epsilon} - dQ_{W,0}\right) = \int_w \left(-\tau_0(\tilde{d}_0(v) - \kappa)\right) H_W(w) dQ_{W,0}(w).$$

Thus the third term in (A.4) generates the $v \mapsto -\tau_0(\tilde{d}_0(v) - \kappa)$ portion of the canonical gradient, or equivalently $v \mapsto -\tau_0(E_{P_U}[d_0(U,v)] - \kappa)$. The remainder of this proof is used to show that the first two terms in (A.4) are $o(\epsilon)$.

**Step 1:** $\eta_\epsilon \to \eta_0$**.**
We refer the reader to (3) for a definition of the quantile $P \mapsto \eta_P$. This is a consequence of the continuity of $S_0$ in a neighborhood of $\eta_0$. For $\gamma > 0$,

$$|\eta_\epsilon - \eta_0| > \gamma \text{ implies that } S_\epsilon(\eta_0 - \gamma) \leq \kappa \text{ or } S_\epsilon(\eta_0 + \gamma) > \kappa. \tag{A.5}$$

For positive constants $C_1$ and $C_W$,

$$S_\epsilon(\eta_0 - \gamma) \geq (1 - C_W|\epsilon|) Pr_0\left(\bar{Q}_{b,\epsilon} > \eta_0 - \gamma\right) \geq (1 - C_W|\epsilon|) S_0(\eta_0 - \gamma + C_1|\epsilon|).$$

Fix $\gamma > 0$ small enough so that $S_0$ is continuous at $\eta_0 - \gamma$. In this case we have that $S_0(\eta_0 - \gamma + C_1|\epsilon|) \to S_0(\eta_0 - \gamma)$ as $\epsilon \to 0$. By the infimum in the definition of $\eta_0$, we know that $S_0(\eta_0 - \gamma) > \kappa$. Thus $S_\epsilon(\eta_0 - \gamma) > \kappa$ for all $|\epsilon|$ small enough.

Similarly, $S_\epsilon(\eta_0 + \gamma) \leq (1 + C_W|\epsilon|) S_0(\eta_0 + \gamma - C_1|\epsilon|)$. Fix $\gamma > 0$ small enough so that $S_0$ is continuous at $\eta_0 + \gamma$. Then $S_0(\eta_0 + \gamma - C_1|\epsilon|) \to S_0(\eta_0 + \gamma)$ as $\epsilon \to 0$. Condition C2) implies the uniqueness of the $\kappa$-quantile of $\bar{Q}_{b,0}$, and thus that $S_0(\eta_0 + \gamma) < \kappa$. It follows that $S_\epsilon(\eta_0 + \gamma) < \kappa$ for all $|\epsilon|$ small enough.

Combining $S_\epsilon(\eta_0 - \gamma) > \kappa$ and $S_\epsilon(\eta_0 + \gamma) < \kappa$ for all $\epsilon$ close to zero with (A.5) shows that $\eta_\epsilon \to \eta_0$ as $\epsilon \to 0$.

**Step 2: Second term of (A.4) is** $0$ **eventually.**
If $\tau_0 = 0$ then the result is immediate, so suppose $\tau_0 > 0$. By the previous step, $\eta_\epsilon \to \eta_0$, which implies that $\tau_\epsilon \to \tau_0 > 0$ by the continuity of the max function. It follows that $\tau_\epsilon > 0$ for $\epsilon$ large enough. By (4), $Pr_{P_U \times P_\epsilon}(d_\epsilon(U,V) = 1) = \kappa$ for all sufficiently small $|\epsilon|$ and $Pr_0(\tilde{d}_0(V) = 1) = \kappa$. Thus the second term of (A.4) is 0 for all $|\epsilon|$ small enough.

**Step 3:** $\tau_\epsilon - \tau_0 = O(\epsilon)$**.**

Note that $\kappa < S_\epsilon(\eta_\epsilon - |\epsilon|) \leq (1+C_W|\epsilon|)S_0(\eta_\epsilon-(1+C_1)|\epsilon|)$. A Taylor expansion of $S_0$ about $\eta_0$ shows that

$$\begin{aligned}\kappa &< (1 + C_W|\epsilon|)\left(S_0(\eta_0) + (\eta_\epsilon - \eta_0 - (1 + C_1)|\epsilon|)(-f_0(\eta_0) + o(1))\right)\\ &= \kappa + (\eta_\epsilon - \eta_0 - (1 + C_1)|\epsilon|)(-f_0(\eta_0) + o(1)) + O(\epsilon)\\ &= \kappa - (\eta_\epsilon - \eta_0)f_0(\eta_0) + o(\eta_\epsilon - \eta_0) + O(\epsilon). \qquad \text{(A.6)}\end{aligned}$$

The fact that $f_0(\eta_0) \in (0, \infty)$ shows that $\eta_\epsilon - \eta_0$ is bounded above by some $O(\epsilon)$ sequence. Similarly, $\kappa \geq S_\epsilon(\eta_\epsilon + |\epsilon|) \geq (1 - C_W|\epsilon|)S_0(\eta_\epsilon + (1 + C_1)|\epsilon|)$. Hence,

$$\begin{aligned}\kappa &\geq (1 - C_W|\epsilon|)\left(S_0(\eta_0) + (\eta_\epsilon - \eta_0 + (1 + C_1)|\epsilon|)(-f_0(\eta_0) + o(1))\right)\\ &= \kappa - (\eta_\epsilon - \eta_0)f_0(\eta_0) + o(\eta_\epsilon - \eta_0) + O(\epsilon).\end{aligned}$$

It follows that $\eta_\epsilon - \eta_0$ is bounded below by some $O(\epsilon)$ sequence. Combining these two bounds shows that $\eta_\epsilon - \eta_0 = O(\epsilon)$, which immediately implies that $\tau_\epsilon - \tau_0 = \max\{O(\epsilon), 0\} = O(\epsilon)$.

**Step 4: First term of (A.4) is $o(\epsilon)$.**
We know that

$$\bar{Q}_{b,0}(V) - \tau_0 + O(\epsilon) \leq \bar{Q}_{b,\epsilon}(V) - \tau_\epsilon \leq \bar{Q}_{b,0}(V) - \tau_0 + O(\epsilon).$$

By C4), it follows that there exists some $\delta > 0$ such that $\sup_{|\epsilon|<\delta} Pr_0(\bar{Q}_{b,\epsilon}(V) = \tau_\epsilon) = 0$. By the absolute continuity of $Q_{W,\epsilon}$ with respect to $Q_{W,0}$, $\sup_{|\epsilon|<\delta} Pr_{P_\epsilon}(\bar{Q}_{b,\epsilon}(V) = \tau_\epsilon) = 0$. It follows that, for all small enough $|\epsilon|$ and almost all $u$, $d_\epsilon(u, v) = I(\bar{Q}_{b,\epsilon}(v) > \tau_\epsilon)$. Hence,

$$\begin{aligned}&\int_w \left(E_{P_U}[d_\epsilon(U, V)] - d_0(V)\right)\left(\bar{Q}_{b,\epsilon} - \tau_0\right) dQ_{W,\epsilon}\\ &= \left|\int_w \left(I(\bar{Q}_{b,\epsilon} > \tau_\epsilon) - I(\bar{Q}_{b,0} > \tau_0)\right)\left(\bar{Q}_{b,\epsilon} - \tau_0\right) dQ_{W,\epsilon}\right|\\ &\leq \int_w \left|I(\bar{Q}_{b,\epsilon} > \tau_\epsilon) - I(\bar{Q}_{b,0} > \tau_0)\right|\left(\left|\bar{Q}_{b,0} - \tau_0\right| + C_1|\epsilon|\right) dQ_{W,\epsilon}\\ &\leq \int_w I(|\bar{Q}_{b,0} - \tau_0| \leq |\bar{Q}_{b,0} - \tau_0 - \bar{Q}_{b,\epsilon} + \tau_\epsilon|)\left(\left|\bar{Q}_{b,0} - \tau_0\right| + C_1|\epsilon|\right) dQ_{W,\epsilon}\\ &= \int_w I(0 < |\bar{Q}_{b,0} - \tau_0| \leq |\bar{Q}_{b,0} - \tau_0 - \bar{Q}_{b,\epsilon} + \tau_\epsilon|)\left(\left|\bar{Q}_{b,0} - \tau_0\right| + C_1|\epsilon|\right) dQ_{W,\epsilon}\\ &\leq O(\epsilon)\int_w I(0 < |\bar{Q}_{b,0} - \tau_0| \leq O(\epsilon)) dQ_{W,\epsilon}\\ &\leq O(\epsilon)(1 + C_W|\epsilon|)Pr_0\left(0 < |\bar{Q}_{b,0} - \tau_0| \leq O(\epsilon)\right),\end{aligned}$$

where the penultimate inequality holds by Step 3 and (A.2). The last line above is $o(\epsilon)$ because $Pr(0 < X \leq \epsilon) \to 0$ as $\epsilon \to 0$ for any random variable $X$. Thus dividing the left-hand side above by $\epsilon$ and taking the limit as $\epsilon \to 0$ yields zero. □

## Proofs for Section 5

We give the following lemma before proving Theorem 4.

**Lemma A.2.** *Let $P_0$ and $P$ be distributions which satisfy the positivity assumption and for which $Y$ is bounded in probability. Let $d$ be some stochastic treatment rule and $\tau$ be some real number. We have that $\Psi_d(P) - \Psi(P_0) = -E_{P_0}[D(d, \tau_0, P)(O)] + R_0(d, P)$.*

*Proof of Lemma A.2.* Note that

$$\begin{aligned}
&\Psi_d(P) - \Psi(P_0) + E_{P_0}[D(d, \tau_0, P)(O)] \\
&\quad = \Psi_d(P) - \Psi_d(P_0) + \sum_{j=1}^{2} E_{P_U \times P_0}[D_j(d(U, \cdot), P)(O)] \\
&\qquad + \Psi_d(P_0) - \Psi_{d_0}(P_0) - \tau_0 E_{P_U \times P_0}[d(U, V) - \kappa].
\end{aligned}$$

Standard calculations show that the first term on the right is equal to $R_{10}(d, P)$ (van der Laan and Robins, 2003). If $\tau_0 > 0$, then (4) shows that $\tau_0 E_{P_U \times P_0}[d - \kappa] = \tau_0 E_{P_U \times P_0}[d - d_0]$. If $\tau_0 = 0$, then obviously $\tau_0 E_{P_U \times P_0}[d - \kappa] = \tau_0 E_{P_U \times P_0}[d - d_0]$. Lemma A.1 shows that $\Psi_d(P_0) - \Psi_{d_0}(P_0) = E_{P_U \times P_0}[(d - d_0)\bar{Q}_{b,0}]$. Thus the second line above is equal to $R_{20}(d)$. □

*Proof of Theorem 4.* We make use of empirical process theory notation in this proof so that $Pf = E_P[f(O)]$ for a distribution $P$ and function $f$. We have that

$$\begin{aligned}
&\hat{\Psi}(P_n) - \Psi(P_0) \\
&\quad = -P_0 D(d_n, \tau_0, P_n^*) + R_0(d_n, P_n^*) && \text{(by Lemma A.2)} \\
&\quad = (P_n - P_0) D(d_n, \tau_0, P_n^*) + R_0(d_n, P_n^*) + o_{P_0}(n^{-1/2}) && \text{(by Condition C10))} \\
&\quad = (P_n - P_0) D_0 + (P_n - P_0)(D(d_n, \tau_0, P_n^*) - D_0) + R_0(d_n, P_n^*).
\end{aligned}$$

The middle term on the last line is $o_{P_0}(n^{-1/2})$ by C5), C6), C8), and C9) (van der Vaart and Wellner, 1996), and the third term is $o_{P_0}(n^{-1/2})$ by C7). This yields the asymptotic linearity result. Proposition 1 in Section 3.3 of Bickel et al. (1993) yields the claim about regularity and asymptotic efficiency when conditions C2), C3), C4), and C5) hold (see Theorem 3). □

*Proof of Lemma 5.* We will show that $\eta_n \to \eta_0$ in probability, and then the consistency of $\tau_n$ follows by the continuous mapping theorem. By C3), there exists an open interval $N$ containing $\eta_0$ on which $S_0$ is continuous. Fix $\eta \in N$. Because $\bar{Q}_{b,n}$ belongs to a Glivenko-Cantelli class with probability approaching 1, we have that

$$\begin{aligned} |S_n(\eta) - S_0(\eta)| &= \left|P_n I(\bar{Q}_{b,n} > \eta) - P_0 I(\bar{Q}_{b,0} > \eta)\right| \\ &\leq \left|P_0\left(I(\bar{Q}_{b,n} > \eta) - I(\bar{Q}_{b,0} > \eta)\right)\right| + \left|(P_n - P_0) I(\bar{Q}_{b,n} > \eta)\right| \\ &\leq \underbrace{\left|P_0\left(I(\bar{Q}_{b,n} > \eta) - I(\bar{Q}_{b,0} > \eta)\right)\right|}_{\triangleq T_n(\eta)} + o_{P_0}(1), \end{aligned} \tag{A.7}$$

where we use the notation $Pf = E_P[f(O)]$ for any distribution $P$ and function $f$. Let $Z_n(\eta)(w) \triangleq \left(I(\bar{Q}_{b,n}(w) > \eta) - I(\bar{Q}_{b,0}(w) > \eta)\right)^2$. The following display holds for all $q > 0$:

$$\begin{aligned} T_n(\eta) &\leq P_0 Z_n(\eta) \\ &= P_0 Z_n(\eta) I(|\bar{Q}_{b,0} - \eta| > q) + P_0 Z_n(\eta) I(|\bar{Q}_{b,0} - \eta| \leq q) \\ &= P_0 Z_n(\eta) I(|\bar{Q}_{b,0} - \eta| > q) + P_0 Z_n(\eta) I(0 < |\bar{Q}_{b,0} - \eta| \leq q) \qquad \text{(A.8)} \\ &\leq P_0 Z_n(\eta) I(|\bar{Q}_{b,n} - \bar{Q}_{b,0}| > q) + P_0 Z_n(\eta) I(0 < |\bar{Q}_{b,0} - \eta| \leq q) \qquad \text{(A.9)} \\ &\leq Pr_0\left(|\bar{Q}_{b,n} - \bar{Q}_{b,0}| > q\right) + Pr_0\left(0 < |\bar{Q}_{b,0} - \eta| \leq q\right) \\ &\leq \frac{P_0|\bar{Q}_{b,n} - \bar{Q}_{b,0}|}{q} + Pr_0\left(0 < |\bar{Q}_{b,0} - \eta| \leq q\right). \end{aligned}$$

Above (A.8) holds because C3) implies that $Pr_0(\bar{Q}_{b,0} = \eta) = 0$, (A.9) holds because $Z_n(\eta) = 1$ implies that $|\bar{Q}_{b,n} - \bar{Q}_{b,0}| \geq |\bar{Q}_{b,0} - \eta|$, and the final inequality holds by Markov's inequality. The lemma assumes that $E_{P_0}|\bar{Q}_{b,n} - \bar{Q}_{b,0}| = o_{P_0}(1)$, and thus we can choose a sequence $q_n \downarrow 0$ such that

$$T_n(\eta) \leq Pr_0\left(0 < |\bar{Q}_{b,0} - \eta| \leq q_n\right) + o_{P_0}(1).$$

To see that the first term on the right is $o(1)$, note that $Pr_0(\bar{Q}_{b,0} = \eta) = 0$ combined with the continuity of $S_0$ on $N$ yield that, for $n$ large enough,

$$Pr_0\left(0 < |\bar{Q}_{b,0} - \eta| \leq q_n\right) = S_0(-q_n + \eta) - S_0(q_n + \eta).$$

The right-hand side is $o(1)$, and thus $T_n(\eta) = o_{P_0}(1)$. Plugging this into (A.7) shows that $S_n(\eta) \to S_0(\eta)$ in probability. Recall that $\eta \in N$ was arbitrary.

Fix $\gamma > 0$. For $\gamma$ small enough, $\eta_0 - \gamma$ and $\eta_0 + \gamma$ are contained in $N$. Thus $S_n(\eta_0 - \gamma) \to S_0(\eta_0 - \gamma)$ and $S_n(\eta_0 + \gamma) \to S_0(\eta_0 + \gamma)$ in probability. Further,

$S_0(\eta_0-\gamma) > \kappa$ by the definition of $\eta_0$ and $S_0(\eta_0+\gamma) < \kappa$ by Condition C2). It follows that, with probability approaching 1, $S_n(\eta_0-\gamma) > \kappa$ and $S_n(\eta_0+\gamma) < \kappa$. But $|\eta_n - \eta_0| > \gamma$ implies that $S_n(\eta_0 - \gamma) \leq \kappa$ or $S_n(\eta_0 + \gamma) > \kappa$, and thus $|\eta_n - \eta_0| \leq \gamma$ with probability approaching 1. Thus $\eta_n \to \eta_0$ in probability, and $\tau_n \to \tau_0$ by the continuous mapping theorem. □

*Proof of Theorem 6.* This proof mirrors the proof of Lemma 5.2 in Audibert and Tsybakov (2007). It is also quite similar to the proof of Theorem 7 in Luedtke and van der Laan (2014b), though the proof given in that working technical report is for optimal rules without any resource constraints, and also contains several typographical errors which will be corrected in the final version.

Define $B_n$ to be the function $v \mapsto \bar{Q}_{b,n}(v) - \tau_n$ and $B_0$ to be the function $v \mapsto \bar{Q}_{b,0}(v) - \tau_0$. Below we omit the dependence of $B_n$, $B_0$ on $V$ in the notation and of $d_n$, $d_0$ on $U$ and $V$. For any $t > 0$, we have that

$$\begin{aligned}
|R_{20}(d_n)| &\leq E_{P_U \times P_0}\left[|(d_n - d_0)B_0|\right] \\
&= E_{P_U \times P_0}\left[I\left(d_n \neq d_0\right)|B_0|\right] \\
&= E_{P_U \times P_0}\left[I\left(d_n \neq d_0\right)|B_0| I(0 < |B_0| \leq t)\right] \\
&\quad + E_{P_U \times P_0}\left[I\left(d_n \neq d_0\right)|B_0| I(|B_0| > t)\right] \\
&\leq E_{P_0}\left[|B_n - B_0| I(0 < |B_0| \leq t)\right] \\
&\quad + E_{P_0}\left[|B_n - B_0| I(|B_n - B_0| > t)\right] \\
&\leq \|B_n - B_0\|_{2,P_0} Pr_0(0 < |B_0| \leq t)^{1/2} + \frac{\|B_n - B_0\|_{2,P_0}^2}{t} \\
&\leq \|B_n - B_0\|_{2,P_0} C_0^{1/2} t^{\alpha/2} + \frac{\|B_n - B_0\|_{2,P_0}^2}{t},
\end{aligned}$$

where the second inequality holds because $d_n \neq d_0$ implies that $|B_n - B_0| \geq |B_0|$ when $|B_0| > 0$, the third inequality holds by the Cauchy-Schwarz and Markov inequalities, and the $C_0$ on the final line is the constant implied by (5). The first result follows by plugging $t = \|B_n - B_0\|_{2,P_0}^{2/(2+\alpha)}$ into the upper bound above. We also have that

$$\begin{aligned}
|R_{20}(d_n)| &\leq E_{P_U \times P_0}\left[I(d_n \neq d_0)|B_0|\right] \\
&\leq E_{P_0}\left[I(0 < |B_0| \leq |B_n - B_0|)|B_0|\right] \\
&\leq E_{P_0}\left[I(0 < |B_0| \leq \|B_n - B_0\|_{\infty,P_0})|B_0|\right] \\
&\leq \|B_n - B_0\|_{\infty,P_0} Pr_0\left(0 < |B_0| \leq \|B_n - B_0\|_{\infty,P_0}\right).
\end{aligned}$$

By (5), it follows that $|R_{20}(d_n)| \lesssim \|B_n - B_0\|_{\infty,P_0}^{1+\alpha}$.

□