University of California, Berkeley

U.C. Berkeley Division of Biostatistics Working Paper Series

Year 2007 *Paper* 216

Analyzing Sequentially Randomized Trials Based on Causal Effect Models for Realistic Individualized Treatment Rules

Oliver Bembom* Mark J. van der Laan†

*Division of Biostatistics, School of Public Health, University of California, Berkeley, bembom@gmail.com

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

Analyzing Sequentially Randomized Trials Based on Causal Effect Models for Realistic Individualized Treatment Rules

Oliver Bembom and Mark J. van der Laan

Abstract

In this paper, we argue that causal effect models for realistic individualized treatment rules represent an attractive tool for analyzing sequentially randomized trials. Unlike a number of methods proposed previously, this approach does not rely on the assumption that intermediate outcomes are discrete or that models for the distributions of these intermediate outcomes given the observed past are correctly specified. In addition, it generalizes the methodology for performing pairwise comparisons between individualized treatment rules by allowing the user to posit a marginal structural model for all candidate treatment rules simultaneously. If only a small number of candidate treatment rules are under consideration, a nonparametric marginal structural can be used to conveniently carry out all of the pairwise comparisons of interest in a single step. An appropriately chosen marginal structural model becomes particularly useful, however, as the number of candidate treatment rules increases, in which case an approach based on individual pairwise comparisons would be likely to suffer from too much sampling variability to provide an informative answer. In addition, such causal effect models represent an interesting alternative to methods previously proposed for selecting an optimal individualized treatment rule in that they give the user a sense of how the optimal outcome is estimated to change in the neighborhood of the identified optimum. We discuss an inverse-probability-of-treatment-weighted (IPTW) estimator for these causal effect models that is straightforward to implement using standard statistical software and develop an approach for constructing valid asymptotic confidence intervals based on the influence curve of this estimator. The methodology is illustrated in two simulation studies that are intended to mimic an HIV/AIDS trial.

1 Introduction

The treatment and management of most chronic or relapsing diseases requires an adaptive strategy that repeatedly adjusts a patient's treatment in response to the observed course of illness. In cancer chemotherapy, for instance, physicians often have available a number of possible treatment options that vary widely, however, in their effectiveness for different patients. After observing a patient's response to the first course of chemotherapy, a physician has to decide whether the selected treatment is effective for this patient or whether the patient should be switched to a new treatment. In the management of HIV/AIDS, physicians likewise have a growing number of treatment options to choose from. In this case, however, the effectiveness of a selected treatment is also likely to change over time for a given patient: While the virus infecting the patient may be initially susceptible to the treatment, it may acquire drug resistance mutations over time that render the treatment ineffective.

In such cases, candidate adaptive strategies, also referred to as dynamic or indivualized treatment rules, are based on a number of different decisions: Which treatment should be used to initially treat a patient? Which treatment should the patient be switched to if the first-line treatment fails? Given an observed intermediate outcome such as change in tumor size or CD4 count, what threshold should be used to decide that the current treatment is failing?

In recent years, sequentially randomized trials have been proposed as a means for investigating such questions (Thall et al., 2000; Lavori and Dawson, 2000, 2004; Murphy, 2005). Conventional randomized trials that randomize patients once to a candidate treatment that is then intended to be followed for the duration of the trial are aimed at comparing static treatment regimens, i.e. regimens that do not change in response to a patient's observed course of illness. Sequentially randomized trials, in contrast, repeatedly re-randomize patients to a set of eligible treatment options over the course of the trial and are specifically aimed at comparing individualized treatment rules. Since the randomization probabilities are allowed to depend on a subject's history of response to treatment, subjects that have responded poorly to their current treatment can be assigned a low probability of being randomized to the same treatment again. Such a design can thus also be hoped to reduce the high rates of drop-out and patient non-compliance that plague conventional trials (Lavori and Dawson, 2000).

A number of sequentially randomized trials are currently being conducted or have already been completed. These include the CATIE trial for atypical antipsychotic medications in patients with Alzheimer's disease (Schneider et al., 2006), the CATIE trial for antipsychotic medications in patients with schizophrenia (Swartz et al., 2007), the STAR*D trial for the treatment of depression (Rush and Fava, 2003; Lavori et al., 2001), and phase II trials for prostate cancer chemotherapy drugs at the MD Anderson Cancer Center (Thall et al., 2000). Most of these trials are exploratory in the sense that they are aimed at developing promising candidate individualized rules that can then be studied in more detail in confirmatory trials.

Statisticians have begun to develop methods for comparing a small number of candidate individualized treatment rules (Thall et al., 2000; Lavori and Dawson, 2000, 2004; Murphy et al., 2001; Hernan et al., 2006). In many cases, however, the collection of candidate treatment rules is sizeable, as, for instance, in the case of threshold switching rules that switch a patient to a new treatment if an observed intermediate outcome such as the change in CD4 count falls below a certain threshold θ. In such cases, a comparison of all candidate treatment rules based on these methods becomes either infeasible or yields highly variable results that are unlikely to provide an informative answer.

van der Laan and Petersen (2007) recently proposed a methodology for comparing individualized treatment rules that addresses this problem by relying on a marginal structural model (MSM). While the authors present their methodology in the context of analyzing observational data, we will argue here that this approach provides a particularly attractive tool for analyzing sequentially randomized trials. Apart from the ability to parsimoniously contrast a large number of candidate individualized treatment rules, we will highlight a number of other advantages over previously proposed methods. These include the reliance on a minimum number of modelling assumptions, the ease of implementation using standard statistical software, and the ability to gain a sense of how candidate treatment rules in the neighborhood of an identified optimal rule compare to that optimal rule. Thus an analysis based on this approach might, for instance, allow an investigator to conclude that, while an optimal rule would switch patients to a salvage regimen if their CD4

count drops by more than 50 cells/μl, there is little evidence that rules based on thresholds between 10 and 90 cells/μl would yield significantly worse results.

van der Laan and Petersen (2007) propose to use the bootstrap for obtaining confidence intervals and p-values for their parameter estimates. We here show that, in the context of sequentially randomized trials, valid asymptotic confidence intervals and p-values can in fact be obtained in closed form based on the influence curve of the proposed estimator. Avoidance of the often computationally intensive bootstrap should further add to the appeal of this methodology. In addition, we illustrate in the context of the threshold switching problem that it is often possible to obtain confidence intervals not only for the expected outcome under the optimal switching threshold, but also for the optimal threshold itself. Such confidence intervals would give the investigator a sense of how precisely the optimal threshold can be estimated based on the available data, a feature that is not provided by previously proposed methods.

The remainder of this article is organized as follows. After reviewing the counterfactual framework for causal inference in the next section, we provide a more detailed comparison of previously proposed methods for analyzing sequentially randomized to that developed in van der Laan and Petersen (2007). Section 4 then describes in detail how to perform estimation based on this latter approach. We next illustrate the methodology in the context of two simulation studies that are meant to mimic a sequentially randomized HIV/AIDS trial. In section 6, we close with a discussion of a number of potential extensions to the methodology presented here.

2 Counterfactual framework for causal inference

In this section, we review the counterfactual framework for causal inference that forms the basis for most of the approaches proposed for comparing individualized treatment rules. We use the notation introduced by Murphy et al. (2001). Suppose that for each subject a total of k treatment assignments $A_1, A_2, \ldots, A_k$ are made. For the sake of simplicity, we assume that only a finite number of possible treatment options are available. Let S_1 denote information available on the subject at baseline that could be used to assign treatment A_1; for $1 < j \leq k$, S_j denotes an intermediate outcome measured at time point j. The observed data structure on a given subject thus consists of

$$O = (S_1, A_1, S_2, A_2, \ldots, S_k, A_k, S_{k+1}). \tag{1}$$

We use the notation $\bar{S}(j) = (S_1, \ldots, S_j)$ to signify information available through time point j; in particular, we use $\bar{S} = \bar{S}_{k+1}$ to denote the entire history of S. The outcome of interest is given by $Y = u(\bar{S}, \bar{A})$ for some known real-value summary function u. Y might for example be the difference between the CD4 count measured at the end of the trial and that measured at baseline. The observed data now consist of n i.i.d. copies O.

The counterfactual framework for causal inference was first introduced by Neyman (1923), extended to causal effects of time-dependent treatments by Rubin (1978), and then further extended to a formal theory of causal inference for direct and indirect effects of time-dependent treatments from experimental and observational longitudinal studies by Robins (1986, 1987). Within this framework, the observed data structure O is viewed as a censored version of a hypothetical full data structure $X = (\bar{S}(\bar{a}) : \bar{a} \in \mathcal{A})$ that contains the outcome process $\bar{S}(\bar{a})$ we would have observed on this subject had she been assigned the static treatment regimen $\bar{a}$ for all $\bar{a}$ in the collection $\mathcal{A}$ of possible static treatment regimens. The consistency assumption $O = (\bar{A}, \bar{S}(\bar{A}))$ states that the observed data consist of the observed treatment process $\bar{A}$ along with the counterfactual outcome process $\bar{S}(\bar{A})$ corresponding to that treatment. The distribution P_0 of the observed data O can thus be indexed by the distribution F_X of the full data structure X and the conditional distribution $g(\cdot \mid X) = P(\bar{A} = \cdot \mid X)$ of the treatment assignment given the full data; g is commonly referred to as the treatment mechanism.

As defined in Robins (1986), an individualized treatment rule d is a sequence of decision rules $d = (d_1, \ldots, d_k)$ that for each time point j map the available information $(\bar{S}_j, \bar{A}_{j-1})$ into a treatment option a_j. As with static treatment regimens, we use the notation $\bar{S}(d)$ to denote the outcome process we would have observed on the subject had she followed the individualized rule d, i.e. $\bar{S}(d) = \bar{S}(\bar{a})$ where $\bar{a} = (a_1 =$

$d(S_1), a_2 = d(\bar{S}_2(a_1), a_1), \ldots)$. We can now compare two candidate individualized treatment rules d_1 and d_2 on the basis of the mean outcomes $E[Y(d_1)]$ and $E[Y(d_2)]$ for the two hypothetical scenarios that all subjects in the target population were to follow rules d_1 and d_2, respectively.

Note that parameters of this type are defined as parameters of the full data distribution F_X. Apart from the consistency assumption, two additional assumptions are generally required in the counterfactual framework for causal inference in order to identify such parameters on the basis of the observed data distribution P_0. The Sequential Randomization assumption states that treatment assignment at time j is independent of the full data X given the information $(\bar{A}_{j-1}, \bar{S}_j)$ available just prior to this treatment assignment. The Experimental Treatment Assignment (ETA) assumption states that all candidate individualized treatment rules under consideration are realistic in the sense that all subjects have positive probability of following any one of them. Since the treatment assignment probabilities are under the control of the investigator, these two assumptions are in general trivially satisfied, as long as some care is taken to define the collection of candidate treatment rules in agreement with the design of the trial. A protocol, for example, that stipulates that any subject with a drop in CD4 count of 50 cells/μl or greater cannot be assigned to their current treatment again does not allow the investigator to study an individualized rule that requires some subjects with a drop in CD4 count of 60 cells/μl to continue on their current treatment. Such a rule would not be realistic given the design of the trial.

3 Comparison with other approaches

The notion of realistic individualized treatment rules was first introduced by Robins (1986) in the context of a study aiming to estimate the effect of workplace exposure to chemicals. Since subjects that are unable to come work on a given day cannot be exposed to these chemicals, static treatment rules that assign a fixed exposure level to all subjects cannot be identified from such a study, making it necessary to focus instead on realistic individualized treatment rules.

In recent years, a number of methods have been proposed for estimating the mean counterfactual outcome $E[Y(d)]$ corresponding to a given individualized treatment rule d. Lavori and Dawson (2000) consider a sequentially randomized trial aimed at finding an optimal criterion for switching patients who fail on a first anti-depressive drug to a second drug. The intermediate outcome S_j records the number of time periods through time point j during which symptoms of depression were present. The candidate treatment rules under consideration switch a patient to the new drug once S_j reaches a threshold θ. The authors propose to estimate $E[Y(d)]$ by imputing the full data structure X for each subject. For a subject that reaches an intermediate score of $S_j = s$ and then remains on her current treatment, for instance, the future outcome process for the rule $\theta = s$ is unobserved, but can be imputed by sampling from the outcome processes of subjects who did switch at time j and who had the same intermediate score of $S_j = s$. This approach yields a valid estimate of $E[Y(d)]$ without any additional assumptions only as long as the number of values that the intermediate outcome S_j can take on is small compared to the number of subjects. If S_j records a subject's CD4 count, for instance, it will be very unlikely that the data set contains subjects who switched at time point j and who had the exact same CD4 count S_j as the subject who did not switch at j.

In essence, this imputation approach relies on estimating the distributions $P(S_j \mid \bar{S}_{j-1}, \bar{A}_j)$, $j \leq k$. If S_j can take on only a fairly small number of values, such estimates can be obtained non-parametrically, i.e. in the absence of any modeling assumptions. In general, however, modeling assumptions cannot be avoided in order to obtain useful estimates. One might, for example, have to assume that the CD4 count at time point l follows a normal distribution with a given mean that is a particular function of past CD4 counts and treatment assignments. The estimates of $E[Y(d)]$ provided by this imputation approach, however, are now only valid if these models are correctly specified.

Lavori and Dawson (2004) propose to estimate $E[Y(d)]$ by G-computation (Robins, 1989b), an approach that uses an estimate of the distributions $P(S_j \mid \bar{S}_{j-1}, \bar{A}_j)$, $j \leq k$, in order to generate a large number of realizations of $Y(d)$ for each subject that can then be averaged to obtain the expected outcome for this subject under rule d. An estimate of $E[Y(d)]$ is then obtained by averaging these expected outcomes over all subjects in the study. Thall et al. (2000) propose a similar approach for studying switching rules in

the context of cancer chemotherapy rules. G-computation is very similar to the imputation methodology proposed by Lavori and Dawson (2000) in that both approaches yield a valid estimate of $E[Y(d)]$ only if S_j is discrete or if our modeling assumptions on $P(S_j \mid \bar{S}_{j-1}, \bar{A}_j)$ are correct.

Murphy et al. (2001) propose inverse-probability-of-treatment-weighted (IPTW) as well as double robust estimating functions for estimating the parameter β in the model $E[Y(d) \mid V] = m(V \mid \beta)$. Here V is a subset of the baseline covariates S_1 so that the mean counterfactual outcome under the rule d is allowed to be different for different values of V. An investigator might, for example, be interested in understanding how the mean counterfactual outcome under rule d changes as a function of age at baseline. IPTW estimators are based on the idea of weighting each subject by the inverse of the probability of following her observed treatment in order to generate a new sample in which treatment assignment is independent of any characteristics that are predictive of the outcome (Robins and Rotnizky, 1992; Robins, 1993). In an observational setting, the treatment mechanism $P(A_j \mid \bar{S}_j, \bar{A}_{j-1})$, $j \leq k$ has to be estimated in order to obtain the desired weights, with valid estimates of the parameter of interest depending on correct specification of the models used for this purpose. In a randomized trial, however, the treatment assignment probabilities are known to the investigator so that valid estimates of $E[Y(d)]$ or $E[Y(d) \mid V]$, for instance, can always be obtained, without requiring that S_j consists of a single discrete covariate. This increased flexibility and robustness should in general make IPTW estimators a more appealing choice in the context of randomized trials than methods based on estimating $P(S_j \mid \bar{S}_{j-1}, \bar{A}_j)$ such as those proposed by Lavori and Dawson (2000, 2004); Thall et al. (2000). The double robust estimating functions proposed by Murphy et al. (2001) can provide an increase in efficiency, but offer the same robustness properties as the IPTW estimating functions in the context of a randomized trial.

Two candidate individualized treatment rules d_1 and d_2 can be compared on the basis of their estimated mean counterfactual outcomes $E[Y(d_1)]$ and $E[Y(d_2)]$. Murphy (2005) develops a test statistic based on the IPTW estimator for testing the hypothesis that these two counterfactual means are equal. Hernan et al. (2006) give a review of this approach in the survival setting. If more than two candidate individualized treatment rules are under consideration, we can select an optimal rule by performing a series of pairwise comparisons to an appropriately chosen reference rule. This approach becomes problematic, however, if the number of candidate treatment rules to compare is sizeable. Consider, for instance, the threshold switching problem introduced above in which we consider rules that switch a patient to a new treatment option once her CD4 count has dropped by at least some value θ. In principle, the parameter θ indexing this set of candidate rules is continuous so that we will be comparing an infinite number of individualized treatment rules. But even if we restrict ourselves by, for instance, only considering rules indexed by $\theta \in \{-200, -195, \ldots, -10, -5, 0\}$, the sampling variability of the resulting estimates is likely to be too large to arrive at an informative answer. In such cases, it would generally be necessary to rely on modeling assumptions about how the the expected mean counterfactual outcome $E[Y(d)]$ changes as function of θ.

van der Laan and Petersen (2007) recently introduced a class of causal effect models that allow the investigator to do just that. These models are of the form

$$E[Y(d) \mid V] = m(d, V \mid \beta), \tag{2}$$

for some parameterization $(d, V) \rightarrow m(d, V \mid \beta)$ indexed by a Euclidean parameter β. They thus model the conditional mean counterfactual outcome $E[Y(d) \mid V]$ for all candidate rules d simultaneously. We note that the corresponding estimators were also developed independently by Orellana et al. (2007). In working with these models, we may follow one of two different approaches. First, we may be willing to assume that the true data-generating distribution actually satisfies this model. In that case, the parameter of interest is given by β in (2). If one believes, however, that such a model is unrealistic, it is more honest to view $m(d, V \mid \beta)$ as a working model that is only used to define a smooth version of $E[Y(d) \mid V]$. This can be done by defining the parameter of interest as the projection of the true conditional mean function $E[Y(d) \mid V]$ onto the working model:

$$\beta_h(F_X) \equiv \arg\min_{\beta} \sum_{d,V} h(d, V) \Big[E_{F_X}(Y_d \mid V) - m(d, V \mid \beta) \Big]^2, \tag{3}$$

where $h(d, V)$ is a user-supplied weight function (Neugebauer and van der Laan, 2007).

We note that the pair-wise comparison approach proposed by Murphy (2005) and Hernan et al. (2006) for the case of a relatively small number of candidate rules $\{d_0, d_1, \ldots, d_p\}$ represents a special case of this methodology that can be accommodated by employing a non-parametric model of the form

$$m(d, V \mid \beta) = \beta_0 + \beta_1 I(d = d_1) + \ldots + \beta_p I(d = d_p), \tag{4}$$

where d_0 is taken as the reference rule. The parameter β_j then gives the difference $E[Y(d_j)] - E[Y(d_0)]$.

We also note that the causal effect models introduced in van der Laan and Petersen (2007) represent an interesting alternative to methods currently used for selecting an optimal individualized treatment rule. Murphy (2003) and Robins (2003) propose a dynamic programming approach to this problem that is based on structural nested models (Robins, 1989a, 1994, 1997, 1999). This approach selects an optimal treatment rule and gives an estimate of the expected counterfactual outcome corresponding to that rule, but it does not give the user a sense for how this expected counterfactual outcome changes in the neighborhood of the selected rule. In the context of selecting a switching threshold, for instance, we would not know if increasing the selected threshold by say 10 cells/μl would be estimated to result in a 1% or a 50% reduction in the expected outcome. The causal effect models discussed here, however, allow us to examine the expected outcome corresponding to any candidate treatment rule, not just the one selected as the optimal rule. Identifying the optimal individualized treatment rule can be a very ambitious task. Often it seems more realistic to identify a range of plausible treatment rules that are estimated to yield favorable outcomes along with a sense for the outcome varies over that range.

van der Laan and Petersen (2007) develop IPTW and double robust estimating functions for estimating parameters of interest defined based on such models. As pointed out before, IPTW estimators represent an appealing option in the context of randomized trials, with double robust estimators affording no extra gain in robustness in this case. Compared to double robust estimators, they offer the advantage of being fairly straightforward to implement using standard statistical software. In the following section, we review the IPTW estimating functions and describe how they can be implemented in practice. In addition, we develop a method for obtaining inference for the IPTW estimator in closed form based on its influence curve.

4 Estimation

Since the parameter of interest is a parameter of the data-generating distribution F_X of the full data structure, we can apply the general estimating function methodology of Robins and Rotnizky (1992) and van der Laan and Robins (2003) to obtain the class of IPTW and double robust estimating functions. In order to obtain the class of IPTW estimating functions which we will focus on here, two steps are required. First, we need to find the class of full-data estimating functions, i.e. the class of estimating functions we would use if we in fact observed the full data structure X on each subject. Second, we need to map these estimating functions into estimating functions of the observed data structure O that have the property that their conditional expectation given X gives back the class of full-data estimating functions.

Let $\mathcal{A}^*$ denote the set of candidate individualized treatment rules. If we assume model (2), the class of full-data estimating functions is then given by the usual class of estimating functions for repeated measures regression:

$$\left\{ \sum_{d \in \mathcal{A}^*} h(d, V) \frac{\partial}{\partial \beta} m(d, V \mid \beta) \Big[Y_d - m(d, V \mid \beta) \Big] : h \right\}. \tag{5}$$

If we instead define the parameter of interest as in (3), the single full-data estimating function is given by

$$\sum_{d \in \mathcal{A}^*} h(d, V) \frac{\partial}{\partial \beta} m(d, V \mid \beta) \Big[Y_d - m(d, V \mid \beta) \Big]. \tag{6}$$

A given full-data estimating function can now be mapped into the following IPTW observed-data estimating

function whose conditional expectation given X equals the original full-data estimating function:

$$D_{h,IPTW}(O \mid g_0, \beta_0) = \sum_{d \in \mathcal{A}^*} \frac{I(\bar{A} = d(\bar{L}))}{g_0(\bar{A} \mid X)} h(d, V) \frac{\partial}{\partial \beta} m(d, V \mid \beta) \Big[Y - m(d, V \mid \beta) \Big], \tag{7}$$

where $g_0(\bar{a} \mid X) \equiv P(\bar{A} = \bar{a} \mid X)$ denotes the known treatment mechanism. Commonly an estimate of $P(\bar{A} = \bar{a} \mid V)$ is used for the factor $h(d, V)$ in IPTW estimating functions in order to obtain so-called stabilized weights (e.g., Hernan et al., 2000), but this can generally be avoided in randomized trials if the investigator ensures that treatment assignment probabilities that are bounded away from zero. In that case, one can simply use $h(d, V) = 1$.

The corresponding IPTW estimator β_n^{IPTW} is defined as the solution of the estimating equation

$$0 = \frac{1}{n} \sum_{i=1}^{n} \sum_{d \in \mathcal{A}^*} \frac{I(\bar{A}_i = d(\bar{L}_i))}{g_0(\bar{A}_i \mid X_i)} h(d, V_i) \frac{\partial}{\partial \beta} m(d, V \mid \beta) \Big[Y_i - m(d, V_i \mid \beta) \Big]. \tag{8}$$

Since this solution is equivalent to

$$\beta_n^{IPTW} = \arg\min_{\beta} \sum_{i=1}^{n} \sum_{d \in \mathcal{A}^*} \frac{I(\bar{A}_i = d(\bar{L}_i))}{g_0(\bar{A}_i \mid X_i)} h(d, V_i) \Big[Y_i - m(d, V_i \mid \beta) \Big]^2, \tag{9}$$

β_n^{IPTW} can be obtained by creating a new data set that for each subject contains one line for each candidate treatment regimen $d \in \mathcal{A}^*$ and then regressing these derived observations on the model $m(d, V \mid \beta)$ using weights $h(d, V_i)/g_0(\bar{A}_i \mid X_i)$. Apart from the first step, this estimator therefore requires the same amount of programming as the IPTW estimator introduced by Murphy (2003).

van der Laan and Petersen (2007) suggest using the bootstrap to obtain confidence intervals and p-values for the estimates provided by this IPTW estimator. Since the treatment mechanism is known in randomized trials, however, we can also base inference in a straightforward manner on an asymptotic normal distribution that can be derived based on the influence curve of the estimator. Recall that the estimator β_n^{IPTW} is asymptotically linear with influence curve $IC^{IPTW}(O \mid g_0, \beta_)$ if we can write

$$\sqrt{n}(\beta_n^{IPTW} - \beta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} IC^{IPTW}(O_i \mid g_0, \beta_0) + o_p(1). \tag{10}$$

This implies in particular that

$$\sqrt{n}(\beta_n^{IPTW} - \beta_0) \Rightarrow N(0, \Sigma^2) \tag{11}$$

with $\Sigma^2 = Var(IC^{IPTW}(O \mid g_0, \beta_0))$. In order to base inference on this limiting distribution, we need to obtain an estimate of the variance Σ^2 of the influence curve of the estimator.

Since β_n^{IPTW} is defined as the solution of an estimating equation derived from an estimating function that lies in the orthogonal complement of the nuisance tangent space, its influence curve is given by an appropriately standardized version of the estimating function itself (Bickel et al., 1993):

$$IC^{IPTW}(O \mid g_0, \beta_0) = -c^{-1} D^{IPTW}(O \mid g_0, \beta_0), \tag{12}$$

where, as elaborated in section A of the appendix, the standardizing constant $-c^{-1}$ is defined by

$$\begin{aligned} c &= \frac{\partial}{\partial \beta^T} E D^{IPTW}(O \mid g_0, \beta) \Big|_{\beta = \beta_0} \\ &= -E_V \left\{ \sum_{d \in \mathcal{A}^*} h(d, V) \frac{\partial}{\partial \beta} m(d, V \mid \beta) \Big|_{\beta = \beta_0} \frac{\partial}{\partial \beta^T} m(d, V \mid \beta) \Big|_{\beta = \beta_0} \right\}. \end{aligned} \tag{13}$$

An estimate c_n of c can be obtained by replacing the expectation over V by the empirical mean over the sample:

$$c_n = -\frac{1}{n}\sum_{i=1}^{n}\left\{\sum_{d\in\mathcal{A}^*} h(d,V_i)\left.\frac{\partial}{\partial\beta}m(d,V_i\mid\beta)\right|_{\beta=\beta_0}\left.\frac{\partial}{\partial\beta^T}m(d,V_i\mid\beta)\right|_{\beta=\beta_0}\right\}. \tag{14}$$

In the case of a linear marginal structural model, for instance, with $m(d,V\mid\beta)=Z_d\beta$ for a p-dimensional column vector β and a p-dimensional row vector Z_d, we have that

$$c_n = -\frac{1}{n}\sum_{i=1}^{n}\left\{\sum_{d\in\mathcal{A}^*} h(d,V_i)Z_{d,i}^T Z_{d,i}\right\}. \tag{15}$$

If we let $\mathbf{Z}$ denote the design matrix corresponding to the model $m(d,V\mid\beta)$ obtained by stacking the row vectors $Z_{d,i}$ and we use $h(d,V)=1$, c_n can thus be obtained as

$$c_n = -\frac{1}{n}\mathbf{Z}^T\mathbf{Z}. \tag{16}$$

An estimate Σ_n^2 of Σ^2 can then be obtained as the sample variance of $IC_n^{IPTW}(O\mid g_0,\beta_n) = -c_n^{-1}D^{IPTW}(O\mid g_0,\beta_n^{IPTW})$. An asymptotic 95% confidence interval for β_j, for example, can now be constructed as

$$\beta_{j,n}^{IPTW} \pm 1.96\sqrt{\frac{\Sigma_n^2(j,j)}{n}}, \tag{17}$$

where $\Sigma_n^2(j,j)$ is the entry in cell (j,j) of Σ_n^2.

Even though the true treatment mechanism g_0 is known to the investigator, one may want to consider obtaining an estimate g_n of g_0 from the sample and substituting g_n for g_0 in the estimator (8). Since the true treatment mechanism is known, it is straightforward to obtain an estimate g_n that is consistent, guaranteeing that the resulting estimator still provides a valid estimate of β. At the same time, however, it is known that an estimator based on g_n can be more efficient than an estimator based on g_0 (van der Laan and Robins, 2003). While this seems somewhat counter-intuitive at first, one can think of the estimator based on g_n as adjusting for empirical confounding that is caused by any chance imbalance between treatment groups in a covariate that has an independent effect on the outcome.

The influence curve for the estimator based on g_n is given by (12) minus its projection on the tangent space corresponding to the model for the treatment mechanism. If that model is non-parametric, the influence curve of the resulting IPTW estimator is equal to the efficient influence curve, making the estimator in fact efficient (van der Laan and Robins, 2003). In that case, the efficient influence curve, derived in van der Laan and Petersen (2007) can be used for inference. If the model for the treatment mechanism is not non-parametric, however, the projection of the influence curve (12) onto the scores of that model must be worked out by hand. In practice, it may be more reasonable to be somewhat conservative and base inference for an inefficient estimator based on g_n on the influence curve (12). Alternatively, one may use the bootstrap for inference.

5 Simulation studies

We now present two simulation studies that illustrate the methodology described in this article. These simulation studies are intended to mimic a sequentially randomized trial that studies the effect of four candidate drugs on the change in CD4 count relative to baseline among HIV-positive patients. The observed data on each patient consist of $O=(S_1,A_1,S_2,A_2,S_3)$, where S_j denotes the CD4 count measured at time point j and A_j denotes the treatment assigned at time point j. The outcome of interest is given by the change in CD4 count relative to baseline observed at the end of the study, $Y\equiv S_3-S_1$.

In both simulation studies, we are interested in comparing indivualized treatment rules that start a patient on some drug a_1 and then switch her to drug a_2 if we decide that she is not responding well enough

to a_1. Specifically, we consider rules that switch a patient to a new drug if the change in CD4 count $S_2 - S_1$ observed between the first two time points falls below some threshold θ. In the first simulation study, θ is assumed to be known ahead of time so that we are tasked with comparing the twelve switching rules $\{d(a_1, a_2) : a_1 \neq a_2\}$ obtained by selecting a initial treatment option and a drug to switch to if that first option fails. The second simulation study mimics a trial that is aimed at finding not only the best choices for a_1 and a_2, but also for the switching threshold θ. The candidate treatment rules $d(a_1, a_2, \theta)$ in this more ambitious trial are thus indexed by three parameters. Apart from the treatment mechanism, the two simulation studies agree in the following setup.

The expected response of a given patient to a candidate drug l depends on whether the virus infecting the patient is susceptible or resistant to that drug. We use $U_l = 1$ and $U_l = 0$ to denote susceptibility and resistance to drug l, respectively. For the sake of simplicity, U_l is assumed to remain constant over the course of the trial. A patient is susceptible to drug one with probability $P(U_1 = 1) = 0.7$. Drugs one and two are assumed to work according to a similar mechanism so that resistance to drug two is highly correlated to resistance to drug one. In particular, we assume that a patient who is susceptible to drug one is also susceptible to drug two with probability $P(U_2 = 1 \mid U_1 = 1) = 0.9$ and that a patient who is resistant to drug one is also resistant to drug two with probability $P(U_2 = 0 \mid U_1 = 0) = 0.95$. Drugs three and four are assumed to represent a second mechanistic class so that U_3 and U_4 are generated in the same manner as U_1 and U_2, but independent of these first two susceptibility scores. Note that $U = (U_1, U_2, U_3, U_4)$ is not observed by the investigator.

Baseline CD4 counts S_1 are generated from a uniform distribution over the interval from 200 to 800. The CD4 count between two time points j and $j + 1$ is assumed to drop by an average of $\beta_0 = -40$ cells/μl if the patient is untreated or treated with a drug to which the virus is resistant to. If the patient is treated with drug one and the virus is susceptible to drug one, the CD4 count is assumed to increase by an average of 10 cells/μl between two time points, corresponding to an effect of $\gamma_1 = 50$ for drug one. The effects of the remaining three drugs are assumed to be $\gamma_2 = 60$, $\gamma_3 = 50$, and $\gamma_4 = 40$, respectively. Specifically, a patient's CD4 count at time $j + 1$ is generated from a $N(\mu_j, \sigma^2)$ distribution, where $\sigma^2 = 100$ and

$$\begin{aligned} \mu_j &= S_j + \gamma_0 + \gamma_1 U_1 I(A_j = 1) + \gamma_2 U_2 I(A_j = 2) + \\ &\quad \gamma_3 U_3 I(A_j = 3) + \gamma_4 U_4 I(A_j = 4). \end{aligned} \tag{18}$$

5.1 Example 1: Switching rules

In this simulation, we assume that the switching threshold θ is fixed ahead of time at some value θ_0. Based on expert opinion, one may, for example, believe that patients with changes in CD4 counts below $\theta_0 = -40$ should always be switched to one of the remaining treatment options. At baseline, patients are randomized to the four candidate drugs with equal probability, i.e. $P(A_1 = a_1) = 0.25$ for $a_1 = 1, 2, 3, 4$. At $j = 2$, patients whose change in CD4 is greater than or equal to θ_0 are kept on their current drug; patients with a change in CD4 below θ_0 are re-randomized to the remaining three drugs with equal probability:

$$P(A_2 = a_2 \mid S_2, A_1, S_1) = \begin{cases} 1/3 & \text{if } S_2 - S_1 < \theta_0 \text{ and } a_2 \neq A_1 \\ 0 & \text{if } S_2 - S_1 < \theta_0 \text{ and } a_2 = A_1 \\ 1 & \text{if } S_2 - S_1 \geq \theta_0 \text{ and } a_2 = A_1 \\ 0 & \text{if } S_2 - S_1 \geq \theta_0 \text{ and } a_2 \neq A_1 \end{cases} \tag{19}$$

Such "play-the-winner-and-drop-the-loser" treatment strategies reflect typical medical practice and represent an appealing option for randomizing treatments in sequentially randomized trials (Thall et al., 2000).

We will analyze the data simulated in this fashion based on the marginal structural model

$$E[Y_{d(a_1,a_2)}] = \beta_0 + \beta_{13} I(a_1 = 1, a_2 = 3) + \beta_{14} I(a_1 = 1, a_2 = 4) + \ldots + \beta_{43} I(a_1 = 4, a_2 = 3) \tag{20}$$

where the rule $d(1, 2)$ is taken as a reference and β_{ij} gives the change in mean counterfactual outcome for rule $d(i, j)$ relative to this reference. The hypothesis of equal mean counterfactual outcomes for the two

rules $d(1,2)$ and $d(2,3)$, for instance, thus corresponds to the hypothesis $\beta_{23} = 0$. Note that this marginal structural model makes no additional assumptions on the data-generating distribution. Using a model of the form (20) is thus in fact just a convenient way of carrying out the whole pair-wise comparison approach suggested by Murphy (2005) and Hernan et al. (2006) in a single step.

Table 1: Summary of example 1.

	β_0	β_{13}	β_{14}	β_{21}	β_{23}	β_{24}	β_{31}	β_{32}	β_{34}	β_{41}	β_{42}	β_{43}
Consistency												
1) Truth	-9.1	4.4	3.3	13.4	17.8	16.7	4.4	5.4	-0.3	-10.0	-8.9	-14.4
2) IPTW Limit	-9.1	4.4	3.3	13.4	17.8	16.7	4.4	5.4	-0.3	-10.0	-8.9	-14.4
MSE for g_0												
3) n = 250	62	136	142	217	193	190	180	173	193	163	160	178
4) n = 1000	16	96	97	119	113	116	105	103	110	101	100	106
5) n = 5000	3	87	86	92	91	90	88	88	89	88	89	89
MSE for g_n												
6) n = 250	45	95	96	174	163	164	153	153	162	144	142	153
7) n = 1000	10	85	86	107	105	105	98	98	99	96	96	96
8) n = 5000	2	83	83	88	88	87	86	86	86	87	87	87
Coverage for g_0												
9) n = 250	0.90	0.91	0.91	0.94	0.94	0.94	0.93	0.93	0.94	0.92	0.92	0.93
10) n = 1000	0.94	0.95	0.94	0.94	0.95	0.95	0.94	0.95	0.95	0.95	0.94	0.94
11) n = 5000	0.95	0.95	0.95	0.95	0.94	0.95	0.95	0.95	0.95	0.95	0.95	0.95
Coverage for g_n												
12) n = 250	0.94	0.94	0.95	0.97	0.97	0.97	0.96	0.95	0.97	0.95	0.95	0.96
13) n = 1000	0.98	1.00	1.00	0.99	0.98	0.98	0.98	0.98	0.99	0.98	0.98	0.99
14) n = 5000	0.99	1.00	1.00	0.99	0.98	0.98	0.98	0.99	0.99	0.98	0.98	0.98

Table 1 summarizes the results of this simulation study. The true values of the parameter $(\beta_0, \ldots, \beta_{43})$ can be obtained by Monte-Carlo simulation. The first row of table 1 shows hat the mean counterfactual outcome for the reference rule $d(1,2)$ is given by -9.1. If we choose $a_2 = 3$ instead of $a_2 = 2$ as the salvage drug, the mean counterfactual outcome is increased by $\beta_{13} = 4.4$. Even though drug two is more effective than drug three for a patient whose virus is susceptible ($\gamma_2 = 60$ vs. $\gamma_3 = 50$), cross-resistance between drugs one and two makes drug three a better salvage drug for patients who failed on drug one. Similarly both drugs three and four represent better salvage options than drug one for patients who failed on drug two. A sequentially randomized trial of this kind thus allows us to identify the optimal salvage treatment for each first-line drug. In addition, we are able to identify the rule $d(2,3)$ as the overall optimal rule with a mean counterfactual outcome of 8.7.

We consider both the IPTW estimator based on the true treatment mechanism g_0 and that based on an estimated treatment mechanism g_n. Here g_n is obtained as follows: For $a_1 = 1,2,3,4$, we estimate $P(A_1 = a_1)$ by the empirical proportion of subjects assigned to a_1. Likewise, we estimate $P(A_2 = a_2 \mid S_2 = s_2, A_1 = a_1, S_1 = s_1)$ by the empirical proportion of subjects assigned to a_2 among those subjects with $A_1 = a_1$ and $S_2 - S_1 \geq \theta_0$ or $S_2 - S_1 < \theta_0$ depending on whether $s_2 - s_1 \geq \theta_0$ or $s_2 - s_1 < \theta_0$. To confirm that these two IPTW estimators are consistent, we examine their asymptotic limit by applying them to a data set consisting of 1,000,000 observations. As $n \to \infty$, $g_n \to g_0$ so that the two estimators based on g_0 and g_n are identical in the limit. Row 2 of table 1 shows that they are in fact consistent.

To get a sense of the sampling variability of the two estimators, we computed mean-squared-errors (MSE) for a number of different sample sizes. Rows 3-8 of table 1 show that the parameter β_0 can be estimated with considerably greater precision than the remaining parameters. This makes sense since β_0 represents the mean counterfactual outcome for the rule $d(1,2)$ while the remaining parameters represent contrasts between

the other candidate rules and that reference rule. A comparison of the variance of the two estimators (data not shown) indicates that the IPTW estimator based on g_n achieves a roughly 40 to 60% greater efficiency than the estimator based on g_0. As explained in section 4, this somewhat counter-intuitive results agrees well with general theoretical considerations laid out in van der Laan and Robins (2003), for example.

We also computed estimated coverage probabilities for asymptotic 95% confidence intervals constructed based on the influence curve (12) as described in section 4. Rows 9-11 of table 1 show that the confidence intervals for the estimator based on g_0 have coverage probabilities close to 0.95 even for the fairly small sample size of 250, with further improvements seen as n is increased. Rows 12-14 show that confidence intervals for the IPTW estimator using g_n based on the influence curve (12) are in fact somewhat conservative.

Another criterion for assessing how well the optimal rule can be estimated based on a sample of n observations consists of the mean counterfactual outcome we would expect to see if the selected rule were applied to the entire target population. Specifically, since the true mean counterfactual outcome for each of the twelve rules is known, we can use the IPTW estimator to select a rule for each of a large number of simulated data sets and then simply take the average of the true mean counterfactual outcomes corresponding to the selected rules. Since the IPTW estimator will not always select the true optimal rule, this average counterfactual outcome will tend to be somewhat lower than the mean counterfactual outcome of 8.7 for the true optimal rule. As sample size increases, however, the IPTW estimator should have an increasingly better probability of selecting the true optimal rule so that we would expect the average counterfactual outcome to improve as sample size increases. If the true treatment mechanism g_0 is used, sample sizes of $n = 250$, 1000, and 5000 in fact yield expected counterfactual outcomes of 5.7, 7.9, and 8.4, respectively. Using an estimated treatment mechanism g_n for the same sample sizes yields expected counterfactual outcomes of 6.1, 8.4, and 8.7. Recall that the mean counterfactual outcome for the reference rule $d(1, 2)$ is -9.1. A strategy that selects an optimal rule based on the IPTW estimator can thus yield favorable results for even fairly small sample sizes. As sample size increases, the performance of this strategy begins to approach that of the true optimal rule.

5.2 Example 2: Threshold switching rules

The "play-the-winner-drop-the-loser" design used in example 1 allows us to compare individualized treatment rules $d(a_1, a_2)$ indexed by an initial treatment choice a_1 and a salvage choice a_2. It does not allow us to consider rules $d(a_1, a_2, \theta)$ that switch patients to a_2 if $S_2 - S_1 < \theta$ for a value of θ other than θ_0. In the context of the proposed design, such rules are not realistic since subjects with $S_2 - S_1 < \theta_0$ are always re-randomized to the remaining three drugs and will thus never follow a rule $d(a_1, a_2, \theta)$ for $\theta > \theta_0$. Similarly, subjects with $S_2 - S_1 > \theta_0$ are always kept on their initial treatment a_1 and will thus never follow a rule $d(a_1, a_2, \theta)$ for $\theta < \theta_0$.

In order to compare rules $d(a_1, a_2, \theta)$ for values of θ other than a given θ_0, we can use the following modified version of the design described in example 1. Suppose we can agree on a lower bound $\underline{\theta}$ below which we would consider it unethical to keep patients on their initial treatment. Patients with $S_2 - S_1 < \underline{\theta}$ will thus have to be randomized to the remaining treatment options as in the sequentially randomized trial described earlier. Unlike before, however, patients with acceptable intermediate outcomes $S_2 - S_1 \geq \underline{\theta}$ are not kept on their initial treatment, but are once again randomized to the four candidate drugs:

$$P(A_2 = a_2 \mid S_2, A_1, S_1) = \begin{cases} 0 & \text{if } S_2 - S_1 < \underline{\theta} \text{ and } a_2 = A_1 \\ 1 & \text{if } S_2 - S_1 < \underline{\theta} \text{ and } a_2 \neq A_1 \\ 0.25 & \text{if } S_2 - S_1 \geq \underline{\theta} \text{ for } a_2 = 1, 2, 3, 4. \end{cases} \tag{21}$$

Within this design, all patients have a positive probability of following any individualized treatment rule $d(a_1, a_2, \theta)$ for $\theta \geq \underline{\theta}$. Rules with $\theta < \underline{\theta}$ remain unrealistic and can thus not be identified based on this design. In the simulation study presented in this section, we use $\underline{\theta} = -50$. Apart from the modified treatment mechanism, the data are generated as in example 1.

For very small values of θ, a rule $d(a_1, a_2, \theta)$ will keep all patients on their initial treatment a_1, corresponding the static treatment regimen (a_1, a_1); for very large value of θ, such a rule will switch all patients to

a_2, corresponding to the static treatment regimen (a_1, a_2). Unless one of the drugs a_1 and a_2 is far superior to the other, both of these static rules can be expected to result in a worse mean counterfactual outcome than a dynamic rule that keeps patients responding well to a_1 on a_1 while switching those who respond poorly to a_2. We would therefore expect the function $\theta \rightarrow E[Y(d(a_1, a_2, \theta))]$ to first increase to a maximum $M(a_1, a_2)$ at some value $\theta = m(a_1, a_2)$ before decreasing again. Over a moderate range of candidate values, say $-50 \leq \theta \leq 20$, this function might thus be reasonably well approximated by a quadratic polynomial. Based on this consideration, we will posit the marginal structural model

$$\begin{aligned} E[Y_{d(a_1,a_2,\theta)}] &= I(a_1 = 1, a_2 = 2)\Big[\beta_{12,0} + \beta_{12,1}\theta + \beta_{12,2}\theta^2\Big] + \\ & \quad I(a_1 = 1, a_2 = 3)\Big[\beta_{13,0} + \beta_{13,1}\theta + \beta_{13,2}\theta^2\Big] + \\ & \quad \ldots + I(a_1 = 4, a_2 = 3)\Big[\beta_{43,0} + \beta_{43,1}\theta + \beta_{43,2}\theta^2\Big] \end{aligned} \tag{22}$$

that, for each a_1, a_2, approximates $\theta \rightarrow EY_{d(a_1,a_2,\theta)}$ by a separate quadratic polynomial. As candidate values for θ we consider the set $\Theta = \{-50, -49, \ldots, 19, 20\}$, resulting in a total of 852 candidate individualized treatment rules.

Figure 1 shows the true dependence of $E[Y(d(a_1, a_2, \theta))]$ on θ along with the projection of that function on the quadratic model (22), illustrating that the quadratic model fits the data very well. The true optimal rule, given by $d(2, 3, -16)$, achieves a mean counterfactual outcome of 14.3. Within model (22), the optimal rule is given by $d(2, 3, -13)$, with a corresponding mean counterfactual outcome of 15.5. Recall that the optimal rule $d(2, 3)$ in example 1, based on a switching threshold of $\theta_0 = -40$ achieves a mean counterfactual outcome of 8.7, illustrating that the performance of this rule can be improved somewhat by using a more aggressive threshold. As argued earlier, the causal effect models considered here have the advantage of giving the user a sense of how the mean counterfactual outcome changes in the neighborhood of the selected optimal rule. In this case, we see that a fairly wide range of values for θ in $d(2, 3, \theta)$ achieves mean counterfactual outcomes close to the optimal value. Rules $d(2, 3, \theta)$ for θ between -31 and 8, for instance, all give mean counterfactual outcomes above 12.

We again consider both the IPTW estimator based on the true treatment mechanism g_0 and that based on an estimated treatment mechanism g_n, where g_n is obtained in a similar fashion as in example 1. Figure 1 also shows the asymptotic limit of these two estimators to confirm that they are in fact consistent.

Table 2 shows the MSE for estimating the optimal threshold $m(a_1, a_2) \equiv \arg\max_{\theta \in \Theta} E[Y(d(a_1, a_2, \theta))]$ for a range of different sample sizes. As we might expect, $m(a_1, a_2)$ is hard to estimate precisely if the function $\theta \rightarrow E[Y(d(a_1, a_2, \theta))]$ is relatively flat over the range of candidate values for θ, as illustrated, for instance, by the rules $d(1, 2, \theta)$ and $d(2, 1, \theta)$. If the maximum is more pronounced, however, as seen, for example, in the rules $d(2, 3, \theta)$ or $d(3, 1, \theta)$, $m(a_1, a_2)$ can be estimated with considerably greater precision. Table 3 shows the MSE for estimating the corresponding optimal mean counterfactual outcome $M(a_1, a_2) \equiv \max_{\theta \in \Theta} E[Y(d(a_1, a_2, \theta))]$. Both in estimating $m(a_1, a_2)$ and $M(a_1, a_2)$, the IPTW estimator based on g_n tends to achieve gains in efficiency on the order of 10 to 40% relative to the estimator based on the true g_0 (variance data not shown).

In selecting an optimal value for θ, it would be helpful to have a sense of the sampling variability in our estimate of the function $\theta \rightarrow E[Y(d(a_1, a_2, \theta))]$. For this purpose, we may use the limiting distribution derived in section 4 to construct a simultaneous confidence band for this function (see section B in the appendix). These confidence bands are constructed in such a way that in a large number of repetitions of the trial, the true function $\theta \rightarrow E[Y(d(a_1, a_2, \theta))]$ should be entirely contained in 95% of all confidence bands. Note the difference between these simultaneous confidence bands and point-wise confidence bands that are constructed in such a way that in a large number of repetitions of the trial, $E[Y(d(a_1, a_2, \theta))]$ for any one fixed θ should be contained in 95% of the corresponding confidence intervals. Figure 2 shows an example of the simultaneous confidence bands we constructed for a sample of size $n = 1000$. Note that even at this fairly large sample size, the confidence bands are still quite wide, illustrating that selecting the optimal rule is indeed an ambitious task. As mentioned previously, tackling this problem based on the the causal effect models discussed here rather than the approach introduced by Murphy (2003) has the advantage of

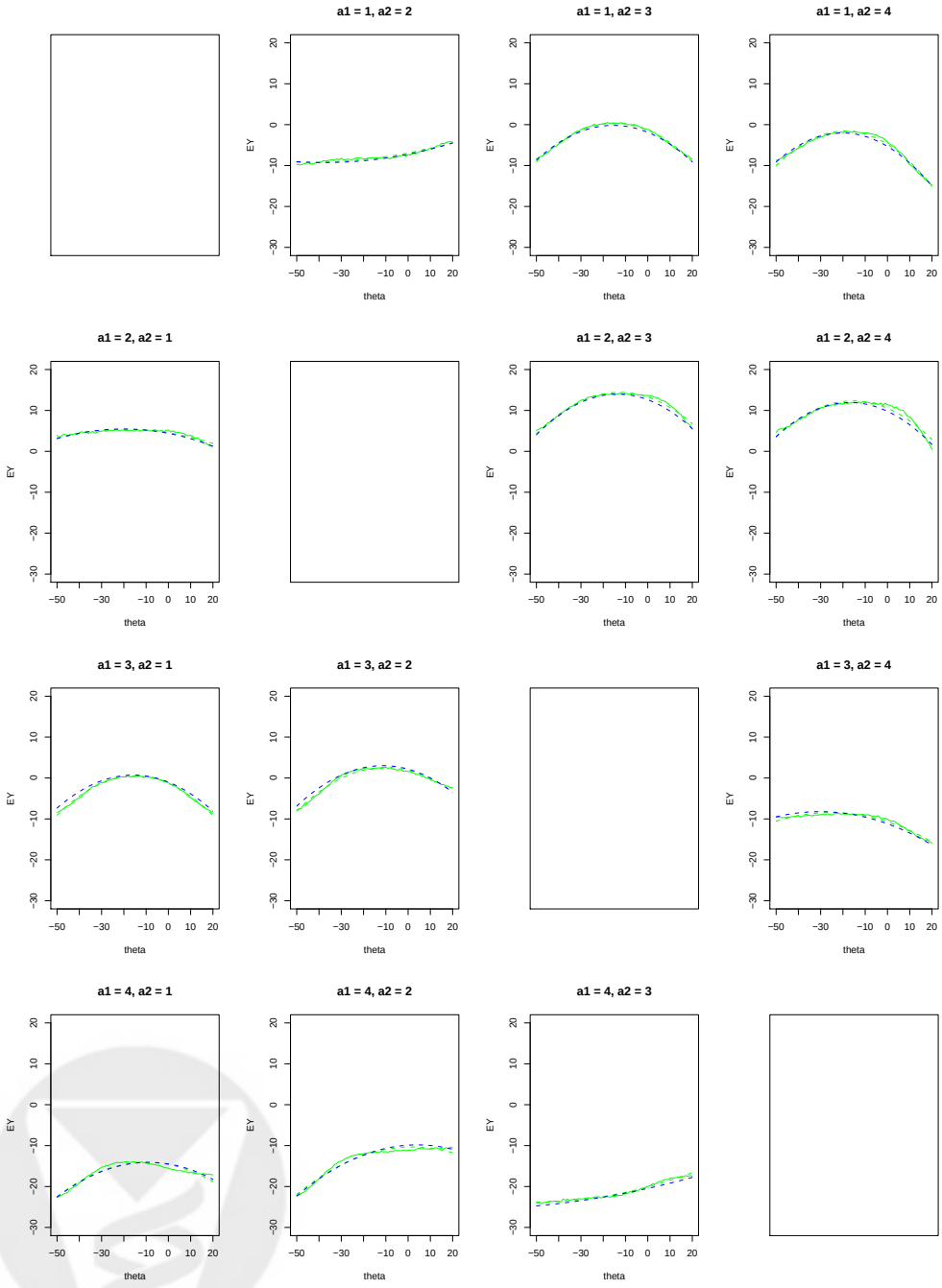

Figure 1: True dependence of $E[Y(d(a_1, a_2, \theta))]$ on θ (solid green line) along with the projection of that function on the quadratic model (22) (dashed green line). The dashed blue line gives the asymptotic limit of the IPTW estimator.

Table 2: MSE for estimating the optimal threshold $m(a_1, a_2)$.

	g_0			g_n		
	n = 250	n = 1000	n = 5000	n = 250	n = 1000	n = 5000
$d(1,2,\theta)$	1562	1027	195	1665	1048	186
$d(1,3,\theta)$	311	70	15	261	50	15
$d(1,4,\theta)$	296	103	31	247	78	29
$d(2,1,\theta)$	651	513	318	642	486	250
$d(2,3,\theta)$	316	85	12	268	62	11
$d(2,4,\theta)$	404	157	48	368	127	44
$d(3,1,\theta)$	327	78	12	287	58	11
$d(3,2,\theta)$	398	101	9	361	71	9
$d(3,4,\theta)$	504	317	163	450	270	118
$d(4,1,\theta)$	549	262	128	484	206	122
$d(4,2,\theta)$	732	343	238	741	339	239
$d(4,3,\theta)$	1108	537	22	1255	582	20

Table 3: MSE for estimating the optimal mean counterfactual outcome $M(a_1, a_2)$.

	g_0			g_n		
	n = 250	n = 1000	n = 5000	n = 250	n = 1000	n = 5000
$d(1,2,\theta)$	123	28	6	123	26	6
$d(1,3,\theta)$	78	20	4	68	14	3
$d(1,4,\theta)$	89	23	5	73	15	3
$d(2,1,\theta)$	188	46	9	170	33	6
$d(2,3,\theta)$	124	32	6	103	20	4
$d(2,4,\theta)$	142	33	7	120	22	4
$d(3,1,\theta)$	82	20	4	71	14	3
$d(3,2,\theta)$	74	18	4	67	14	3
$d(3,4,\theta)$	135	31	6	126	23	4
$d(4,1,\theta)$	52	12	2	50	10	2
$d(4,2,\theta)$	66	15	3	68	14	3
$d(4,3,\theta)$	86	23	5	87	21	5

at least giving the user a better sense of the uncertainty involved in this task. Table 4 summarizes the estimated coverage probabilities of such simultaneous confidence intervals for different sample sizes, showing that reasonable coverage is achieved for sample sizes of $n = 1000$ and greater.

Table 4: Estimated coverage probabilities of simultaneous 95% confidence bands.

	g_0			g_n		
	n = 250	n = 1000	n = 5000	n = 250	n = 1000	n = 5000
$d(1,2,\theta)$	0.84	0.93	0.95	0.87	0.96	0.97
$d(1,3,\theta)$	0.83	0.91	0.93	0.86	0.94	0.96
$d(1,4,\theta)$	0.83	0.92	0.93	0.87	0.96	0.96
$d(2,1,\theta)$	0.85	0.92	0.94	0.88	0.96	0.97
$d(2,3,\theta)$	0.84	0.91	0.94	0.86	0.95	0.97
$d(2,4,\theta)$	0.83	0.92	0.93	0.87	0.96	0.97
$d(3,1,\theta)$	0.82	0.93	0.93	0.85	0.95	0.96
$d(3,2,\theta)$	0.82	0.92	0.94	0.84	0.94	0.96
$d(3,4,\theta)$	0.84	0.92	0.94	0.86	0.96	0.97
$d(4,1,\theta)$	0.82	0.92	0.94	0.86	0.95	0.96
$d(4,2,\theta)$	0.82	0.92	0.95	0.84	0.94	0.96
$d(4,3,\theta)$	0.84	0.93	0.94	0.87	0.96	0.97

Another way of getting a sense of the sampling variability of our estimates would be to construct confidence intervals for $m(a_1, a_2)$ and $M(a_1, a_2)$. Given the parametric model (22), this is possible if the polynomial $\beta_{a_1a_2,0} + \beta_{a_1a_2,1}\theta + \beta_{a_1a_2,2}\theta^2$ attains a maximum in the interval $(\min\Theta, \max\Theta)$. In that case, this maximum can be found as an explicit function of the parameters $\beta_{a_1a_2,0}$, $\beta_{a_1a_2,1}$, and $\beta_{a_1a_2,2}$ by setting the first derivative of the polynomial equal to zero. Specifically, this yields the parameters

$$\tilde{m}(a_1, a_2) = \frac{\beta_{a_1a_2,1}}{2\beta_{a_1a_2,2}} \tag{23}$$

and

$$\tilde{M}(a_1, a_2) = \beta_{a_1a_2,0} - \frac{\beta^2_{a_1a_2,1}}{2\beta_{a_1a_2,2}} + \frac{\beta_{a_1a_2,1}}{2}. \tag{24}$$

Confidence intervals for $\tilde{m}(a_1, a_2)$ and $\tilde{M}(a_1, a_2)$ can now be obtained in a straightforward way by applying the δ-method (see section C of the appendix). If the polynomial $\beta_{a_1a_2,0} + \beta_{a_1a_2,1}\theta + \beta_{a_1a_2,2}\theta^2$ does not attain a maximum in the interval $(\min\Theta, \max\Theta)$, the point $(\tilde{m}(a_1, a_2), \tilde{M}(a_1, a_2))$ corresponds either to a global maximum outside of that interval or to a minimum of the polynomial. Even in the former case, a confidence interval for $\tilde{m}(a_1, a_2)$ and $\tilde{M}(a_1, a_2)$ would not be of too much interest since it would be based entirely on extrapolation. In practice, it is therefore necessary to ensure that the estimated optimal threshold $\tilde{m}(a_1, a_2)$ lies between the smallest and the largest candidate value for θ before interpreting confidence intervals obtained in this fashion. Tables 5 and 6 summarize the estimated coverage probabilities for such 95% confidence intervals, where we have suppressed the results for $(a_1 = 1, a_2 = 2)$ and $(a_1 = 4, a_2 = 3)$ since those two polynomials do not attain a maximum in the interval $(-50, 20)$. Unless the function $\theta \to E[Y(d(a_1, a_2, \theta))]$ is quite flat, as, for instance, in the case of the rules $d(2, 1, \theta)$ and $d(3, 4, \theta)$, confidence intervals for $\tilde{m}(a_1, a_2)$ are estimated to achieve coverage probabilities close to the nominal 0.95 level even at a sample size of $n = 250$. The coverage probability of confidence intervals for $\tilde{M}(a_1, a_2)$ is estimated to be somewhat worse at $n = 250$, with reasonable coverage achieved, however, for sample sizes of $n = 1000$ and greater.

As in example 1, we can assess how well the optimal rule can be estimated based on the mean counterfactual outcome we would expect to see if the selected rule were applied to the entire target population. If the true treatment mechanism g_0 is used, sample sizes of $n = 250$, 1000, and 5000 yield expected counterfactual

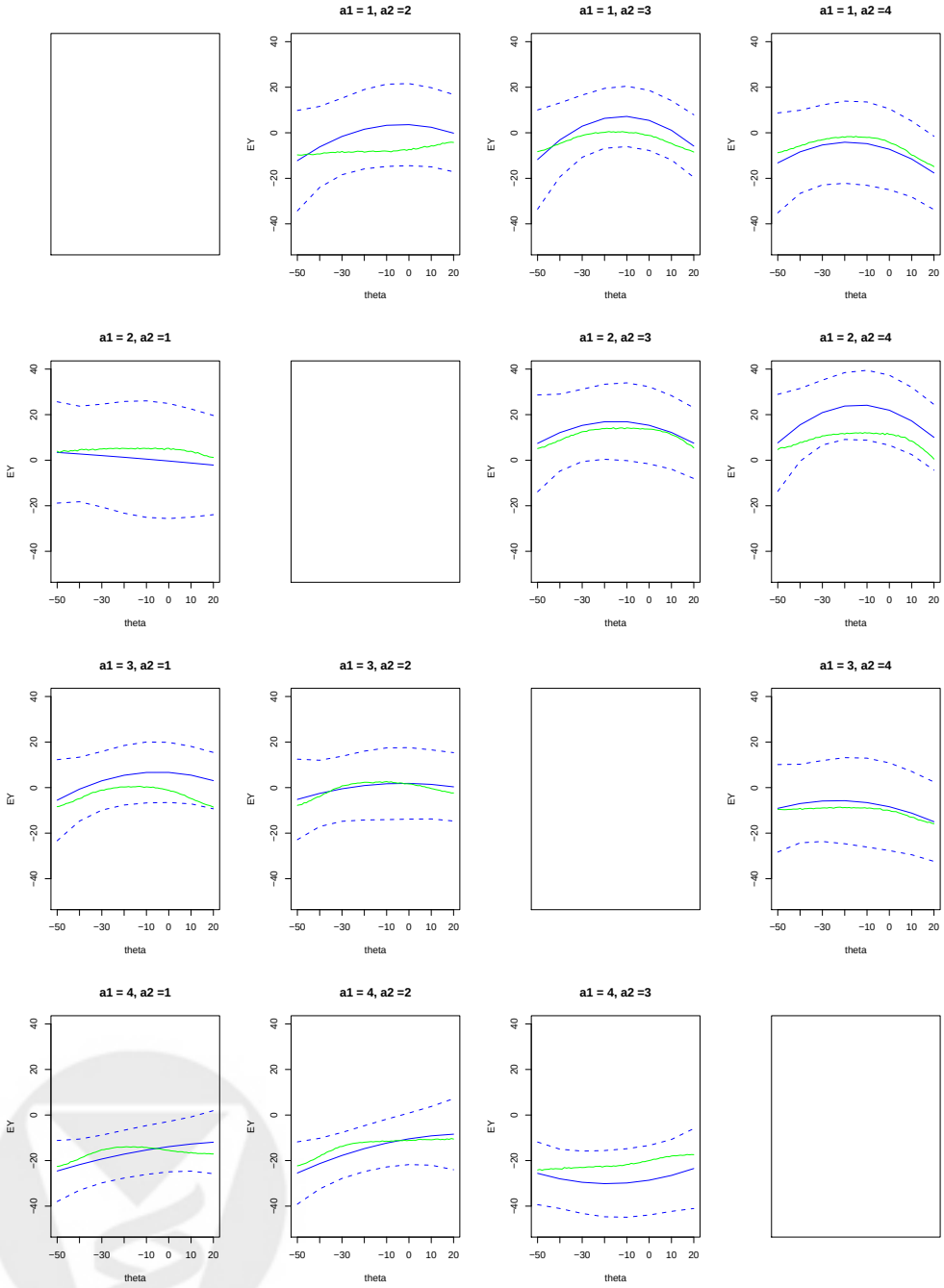

Figure 2: IPTW estimate of the function $\theta \to E[Y(d(a_1, a_2, \theta))]$ derived from a sample of $n = 1000$ subjects (solid blue line) along with simultaneous 95% confidence bands (dashed blue lines). The solid green line shows the true function $\theta \to E[Y(d(a_1, a_2, \theta))]$.

Table 5: Estimated coverage probabilities of 95% confidence intervals for $\tilde{m}(a_1, a_2)$.

	g_0			g_n		
	n = 250	n = 1000	n = 5000	n = 250	n = 1000	n = 5000
$d(1,2,\theta)$	—	—	—	—	—	—
$d(1,3,\theta)$	0.95	0.97	0.96	0.96	0.97	0.96
$d(1,4,\theta)$	0.93	0.95	0.97	0.94	0.96	0.97
$d(2,1,\theta)$	0.85	0.89	0.91	0.87	0.90	0.92
$d(2,3,\theta)$	0.94	0.96	0.98	0.95	0.97	0.98
$d(2,4,\theta)$	0.91	0.95	0.98	0.93	0.96	0.99
$d(3,1,\theta)$	0.96	0.96	0.96	0.96	0.96	0.96
$d(3,2,\theta)$	0.98	0.99	0.95	0.98	0.99	0.96
$d(3,4,\theta)$	0.89	0.90	0.94	0.89	0.92	0.95
$d(4,1,\theta)$	0.99	0.99	0.96	0.99	0.99	0.96
$d(4,2,\theta)$	0.94	0.94	0.97	0.96	0.97	0.98
$d(4,3,\theta)$	—	—	—	—	—	—

Table 6: Estimated coverage probabilities of 95% confidence intervals for $\tilde{M}(a_1, a_2)$.

	g_0			g_n		
	n = 250	n = 1000	n = 5000	n = 250	n = 1000	n = 5000
$d(1,2,\theta)$	—	—	—	—	—	—
$d(1,3,\theta)$	0.90	0.92	0.94	0.91	0.96	0.98
$d(1,4,\theta)$	0.90	0.93	0.94	0.92	0.96	0.97
$d(2,1,\theta)$	0.92	0.94	0.95	0.93	0.98	0.99
$d(2,3,\theta)$	0.89	0.92	0.94	0.90	0.96	0.98
$d(2,4,\theta)$	0.88	0.93	0.93	0.90	0.96	0.97
$d(3,1,\theta)$	0.89	0.93	0.95	0.90	0.97	0.98
$d(3,2,\theta)$	0.89	0.94	0.94	0.90	0.96	0.98
$d(3,4,\theta)$	0.91	0.95	0.94	0.92	0.98	0.98
$d(4,1,\theta)$	0.90	0.94	0.96	0.92	0.97	0.98
$d(4,2,\theta)$	0.92	0.96	0.96	0.93	0.98	0.97
$d(4,3,\theta)$	—	—	—	—	—	—

outcomes of 8.0, 12.0, and 13.5, respectively. Using an estimated treatment mechanism g_n for the same sample sizes yields expected counterfactual outcomes of 8.4, 12.7, and 13.6. Recall that the optimal rules in examples 1 and 2 achieve mean counterfactual outcomes of 8.7 and 14.3, respectively. A strategy based on an estimated optimal rule thus again succeeds quite well, even for moderate sample sizes.

6 Discussion

In this paper, we argue that IPTW estimators for the causal effect models for realistic individualized treatment rules introduced by van der Laan and Petersen (2007) represent an attractive tool for analyzing sequentially randomized trials. Unlike a number of methods proposed previously, this approach does not rely on the assumption that intermediate outcomes S_j are discrete or that models for the distributions $P(S_j \mid \bar{S}_{j-1}, \bar{A}_j)$, $j \leq k$, are correctly specified (Lavori and Dawson, 2000, 2004; Thall et al., 2000). In addition, it generalizes the methodology introduced by Murphy et al. (2001) by allowing the user to posit a marginal structural model for the mean counterfactual outcome corresponding to all candidate treatment rules simultaneously rather than requiring the user to perform a number of pairwise comparisons. Example 1 illustrates that, if only a small number of candidate treatment rules are under consideration, a non-parametric marginal structural can be used to conveniently carry out all of the pairwise comparisons of interest in a single step. Example 2 shows that an appropriately chosen marginal structural model becomes particularly useful as the number of candidate treatment rules increases. In this case, an approach based on individual pairwise comparisons would be likely to suffer from too much sampling variability to provide an informative answer. Furthermore, the causal effect models discussed here represent an interesting alternative to methods proposed by Murphy (2003) for selecting an optimal individualized treatment rule in that they give the user a sense of how the optimal outcome is estimated to change in the neighborhood of the identified optimum. In our second simulation study, for instance, we are able to construct confidence intervals not only for the optimal mean counterfactual outcome, as would have been possible with the methodology developed by Murphy (2003), but also for the optimal switching threshold θ itself. The IPTW estimators discussed in this article are straightforward to implement using standard statistical software by simply creating a derived data set that imitates the unobserved full data structure and then performing a weighted regression.

In addition, we present an approach for constructing valid asymptotic confidence intervals based on the influence curve of the IPTW estimator. Previously, such confidence intervals had only been developed for the pairwise comparisons of two treatment rules that are mutually exclusive in the sense that no subject can simultaneously follow both rules (Murphy, 2005). Other methods for comparing individualized treatment rules rely on resampling-based approaches for inference. Lavori and Dawson (2000), for instance, rely on multiple imputation (Rubin and Shenker, 1991), while Lavori and Dawson (2004) and van der Laan and Petersen (2007) recommend using the bootstrap. Our simulation studies illustrate that the asymptotic confidence intervals proposed here lead to valid inference even for fairly moderate sample sizes.

A number of possible extensions of the methodology discussed in this article exist. First, van der Laan and Petersen (2007) show how the IPTW estimating function can be made more efficient by subtracting its projection onto the nuisance tangent space corresponding to the treatment mechanism under the randomization assumption $P(A_j = a_j \mid X) = P(A_j = a_j \mid \bar{S}_j, \bar{A}_{j-1})$. van der Laan and Rubin (2006) recently introduced a targeted maximum likelihood methodology that could also be used to obtain a more efficient double robust estimator. While such estimators are not as straightforward to implement as the IPTW estimator, they may offer considerable gains in efficiency, especially if the investigator has measured covariates other than the treatment itself that are highly predictive of the outcome (Moore and van der Laan, 2007).

In some cases, it may be difficult to specify an appropriate marginal structural model solely based on *a priori* considerations as those in example 2. Since a mis-specified model can lead to strongly biased results, one may want to employ the general loss-based estimation methodology developed by van der Laan and Dudoit (2003) to select between different candidate models. Wang et al. (2007) have implemented this approach for selecting an appropriate marginal structural model for static treatment regimens in the point-treatment scenario.

As shown in van der Laan and Petersen (2007), the framework introduced in section 2 can easily be

extended to accommodate the case that individual subjects are followed up at a random number of time points. This would, for instance, be the case in a trial in which the outcome of interest is survival. The same set-up could be applied to the sequentially randomized trial design proposed by Thall et al. (2000) in which the treatment of a patient is terminated once they have responded well or poorly to two successive rounds of treatment with the same drug.

The approach discussed in this paper could also be used in the biased-coin adaptive within-subject (BCAWS) design introduced by Lavori and Dawson (2000) that makes re-randomization probabilities at time j dependent on the observed history $\bar{S}_j$. This allows patients with comparatively poor intermediate outcomes to be preferentially assigned to new treatment options while still keeping a treatment rule identifiable that would require the patient to continue on the current treatment. In this context, we note that the investigator will want to ensure that the resulting randomization probabilities are bounded away from zero as it is known that IPTW estimators become biased in finite samples if treatment assignment probabilities are too close to zero (Neugebauer and van der Laan, 2005). Such a practical violation of the ETA assumption can in general be avoided by making sure that all treatment assignment probabilities are greater than 0.05, for instance.

Other scenarios for sequentially randomized trials in which causal effect models for realistic individualized treatment rules might be useful include trials in which patients are initially assigned to a particular dose of a candidate drug, but are then moved to a higher dose of that same drug if they show no evidence for a response and are moved to a lower dose once they experience side effects. The Systolic Hypertension in the Elderly Program is an example of a randomized trial that examined such treatment rules (Kostis et al., 1997).

Lastly, we note that the methodology proposed by van der Laan and Petersen (2007) also has many applications in the observational setting, in which treatment assignment is not under the control of the investigator. In this context, however, additional assumptions are needed to ensure consistency of the IPTW estimator discussed here yields consistent estimates of causal effects. In particular, it is necessary to assume that all important confounders of the relationship between treatment and the outcome of interest have been measured and that one is able to specify a correct model for the unknown treatment mechanism.

References

Bickel, P., Klaasen, C., Ritov, Y., and Wellner, J. (1993). *Efficient and adaptive estimation for semiparametric models*. The Johns Hopkins University Press.

Hernan, M., Brumback, B., and Robins, J. (2000). Marginal structural models to estimate the causal effect of zidovudine on the survival in hiv-positive men. *Epidemiology*, 11:561–570.

Hernan, M., Lanoy, E., Costagliola, D., and Robins, J. (2006). Comparison of dynamic treatment regimes via inverse probability weighting. *Basic & Clinical Pharmacology & Toxicology*, 98:237–242.

Kostis, J., Davis, B., Cutler, J., Grimm, Jr., R., Berge, K., Cohen, J., Lacy, C., Perry, H., Blaufox, M., Wasertheil-Smoller, S., Black, H., Schron, E., Berkson, D., Curb, J., Smith, W., McDonald, R., and Applegate, W. (1997). Prevention of heart failure by antihypertensive drug treatment in older persons with isolated systolic hypertension: SHEP Cooperative Research Group. *Lancet*, 351:1755–1762.

Lavori, P. and Dawson, R. (2000). A design for testing clinical strategies: biased adaptive within-subject randomization. *Journal of the Royal Statistical Society, Series A*, 163:29–38.

Lavori, P. and Dawson, R. (2004). Dynamic treatment regimes: practical design considerations. *Clinical trials*, 1:9–20.

Lavori, P., Rush, A., Wisniewski, S., Alpert, J., Fava, M., Kupfer, D., Nierenberg, A., Quitkin, F., Sacheim, H., Thase, M., and Trivedi, M. (2001). Strengthening clinical effectiveness trials: Equipoise-stratified randomization. *Biological Psychiatry*, 50(10):792–801.

Moore, K. and van der Laan, M. (2007). Covariate Adjustment in Randomized Trials with Binary Outcomes: Targeted Maximum Likelihood Estimation. Technical Report 215, UC Berkeley Division of Biostatistics Working Paper Series.

Murphy, S. (2003). Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society, Series B*, 65(2):331–355.

Murphy, S. (2005). An experimental design for the development of adaptive treatment strategies. *Statistics in Medicine*, 24:1455–1481.

Murphy, S., van der Laan, M., and Robins, J. (2001). Marginal mean models for dynamic treatment regimens. *Journal of the American Statistical Association*, 96:1410–1424.

Neugebauer, R. and van der Laan, M. (2005). Why prefer double robust estimates in causal inference? *Journal of Statistical Planning and Inference*, 129:405–426.

Neugebauer, R. and van der Laan, M. (2007). Nonparametric causal effects based on marginal structural models. *Journal of Statistical Planning and Inference*, 137(2):419–434.

Neyman, J. (1923). On the application of probability theory to agricultural experiments. *Statistical Science*, 5:465–480 (1990).

Orellana, L., Rotnitzky, A., and Robins, J. (2007). Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes. Technical report, Department of Biostatistics. Harvard School of Public Health.

Robins, J. (1986). A new approach to causal inference in mortality studies with sustained exposure periods - application to control of the healthy survivor effect. *Mathematical Modelling*, 7:1393–1512.

Robins, J. (1987). Addendum to "a new approach to causal inference in mortality studies with sustained exposure periods - application to control of the healthy survivor effect" [math. modelling **7** (1986) 1393-1512]. *Computers and Mathematics with Applications*, 14:923–945.

Robins, J. (1989a). The analysis of randomized and non-randomized AIDS treatment trials using a new approach in causal inference in longitudinal studies. In Sechrest, L., Freeman, H., and Mulley, A., editors, *Health Service Methodology: A Focus on AIDS*. U.S. Public Health Service, National Center for Health Services Research.

Robins, J. (1989b). The control of confounding by intermediate variables. *Statistics in Medicine*, 8:679–701.

Robins, J. (1993). Information recovery and bias adjustment in proportional hazards regression analysis of randomized trials using surrogate markers. In *Proceedings of the Biopharmaceutical section*, pages 24–33. American Statistical Association.

Robins, J. (1994). Correcting for non-compliance in randomized trials using structural nested mean models. *Communications in Statistics*, 23:2379–2412.

Robins, J. (1997). Structural nested failure time models. In Armitage, P., Colton, T., Andersen, P., and Keiding, N., editors, *The Encyclopedia of Biostatistics*. John Wiley and Sons.

Robins, J. (1999). Marginal structural models versus structural nested models as tools for causal inference. In Halloran, M. and Berry, D., editors, *Statistical Models in Epidemiology: The Environment and Clinical Trials*, volume 116, pages 95–134. Springer Verlag.

Robins, J. (2003). Optimal-regime structural nested models. proceedings of the second seattle symposium on biostatistcs. In Lin, D. and Haegerty, P., editors, *Lecture notes in statistics*. Springer.

Robins, J. and Rotnizky, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. In Jewell, N., Dietz, K., and Farewell, V., editors, *AIDS Epidemiology, Methodological issues*, pages 297–331. Bihäuser.

Rubin, D. (1978). Bayesian inference for causal effects: The role of randomization. *Annals of Statistics*, 6:34–58.

Rubin, D. and Shenker, N. (1991). Multiple imputation in health-care data bases: an overview and some applications. *Statistics in Medicine*, 10:585–598.

Rush, A.J. Trivedi, M. and Fava, M. (2003). Depression, IV. *American Journal of Psychiatry*, 160(2):237.

Schneider, L., Tariot.P.N., Dagerman, K., Davis, S., Hsiao, J., Ismail, S., Lebowitz, B., Lyketsos, C., Ryan, J., Stroup, T., Sultzer, D., Weintraub, D., and Lieberman, J. (2006). Effectiveness of atypical antipsychotic drugs in patients with Alzheimer's disease. *New England Journal of Medicine*, 355(15):1525–1538.

Swartz, M., Perkins, D., Stroup, T., Davis, S., Capuano, G., Rosenheck, R., Reimherr, F., McGee, M., Keefe, R., McEvoy, J., Hsiao, J., and Lieberman, J. (2007). Effects of antipsychotic medications on psychosocial functioning in patients with chronic shizophrenia: Findings from the nimh catie study. *American Journal of Psychiatry*, 164:428–436.

Thall, P., Millikan, R., and Sung, H.-G. (2000). Evaluating multiple treatment courses in clinical trials. *Statistics in Medicine*, 19:1011–1028.

van der Laan, M. and Dudoit, S. (2003). Unified cross-validation methodology for selection among estimators and a general cross-validated adaptive epsilonnet estimator: Finite sample oracle inequalities and examples. Technical report 130, Division of Biostatistics, University of California, Berkeley.

van der Laan, M. and Petersen, M. (2007). Causal effect models for realistic individualized treatment and intention to treat rules. *The International Journal of Biostatistics*, 3(1):Article 3.

van der Laan, M. and Robins, J. (2003). *Unified Methods for Censored Longitudinal Data and Causality*. Springer Series in Statistics. Springer Verlag.

van der Laan, M. and Rubin, D. (2006). Targeted maximum likelihood learning. *The International Journal of Biostatistics*, 2(1):Article 11.

Wang, Y., Bembom, O., and van der Laan, M. (2007). Data-adaptive estimation of the treatment-specific mean. *Journal of Statistical Planning and Inference*, 137(6):1871–1887.

A Normalization constant in the IPTW influence curve

$$
\begin{aligned}
c &= \frac{\partial}{\partial \beta^T} E D^{IPTW}(O \mid g_0, \beta)\Big|_{\beta=\beta_0} \\
&= \frac{\partial}{\partial \beta^T} E\left\{ \sum_{d\in\mathcal{A}^*} \frac{I(\bar{A}=d(\bar{L}))}{g_0(\bar{A}\mid X)} h(d,V) \frac{\partial}{\partial\beta} m(d,V\mid\beta)\Big[Y - m(d,V\mid\beta)\Big]\right\}\Bigg|_{\beta=\beta_0} \\
&= E\left\{ \sum_{d\in\mathcal{A}^*} \frac{I(\bar{A}=d(\bar{L}))}{g_0(\bar{A}\mid X)} h(d,V)\ \frac{\partial^2}{\partial\beta^T\beta} m(d,V\mid\beta)\Big|_{\beta=\beta_0} \Big[Y - m(d,V\mid\beta)\Big]\right\} - \\
&\quad E\left\{ \sum_{d\in\mathcal{A}^*} \frac{I(\bar{A}=d(\bar{L}))}{g_0(\bar{A}\mid X)} h(d,V)\ \frac{\partial}{\partial\beta} m(d,V\mid\beta)\Big|_{\beta=\beta_0} \frac{\partial}{\partial\beta^T} m(d,V\mid\beta)\Big|_{\beta=\beta_0}\right\} \\
&= E_{F_X} E_g\left\{ \sum_{d\in\mathcal{A}^*} \frac{I(\bar{A}=d(\bar{L}))}{g_0(\bar{A}\mid X)} h(d,V)\ \frac{\partial^2}{\partial\beta^T\beta} m(d,V\mid\beta)\Big|_{\beta=\beta_0} \Big[Y_d - m(d,V\mid\beta)\Big| X\right\} - \\
&\quad E_{F_X} E_g\left\{ \sum_{d\in\mathcal{A}^*} \frac{I(\bar{A}=d(\bar{L}))}{g_0(\bar{A}\mid X)} h(d,V)\ \frac{\partial}{\partial\beta} m(d,V\mid\beta)\Big|_{\beta=\beta_0} \frac{\partial}{\partial\beta^T} m(d,V\mid\beta)\Big|_{\beta=\beta_0}\Big| X\right\} \\
&= E_{F_X}\left\{ \sum_{d\in\mathcal{A}^*} h(d,V)\ \frac{\partial^2}{\partial\beta^T\beta} m(d,V\mid\beta)\Big|_{\beta=\beta_0} \Big[Y_d - m(d,V\mid\beta)\Big]\right\} - \\
&\quad E_{F_X}\left\{ \sum_{d\in\mathcal{A}^*} h(d,V)\ \frac{\partial}{\partial\beta} m(d,V\mid\beta)\Big|_{\beta=\beta_0} \frac{\partial}{\partial\beta^T} m(d,V\mid\beta)\Big|_{\beta=\beta_0}\right\} \\
&= E_V\left\{ \sum_{d\in\mathcal{A}^*} h(d,V)\ \frac{\partial^2}{\partial\beta^T\beta} m(d,V\mid\beta)\Big|_{\beta=\beta_0} E\Big[Y_d - m(d,V\mid\beta)\Big| V\Big]\right\} - \\
&\quad E_V\left\{ \sum_{d\in\mathcal{A}^*} h(d,V)\ \frac{\partial}{\partial\beta} m(d,V\mid\beta)\Big|_{\beta=\beta_0} \frac{\partial}{\partial\beta^T} m(d,V\mid\beta)\Big|_{\beta=\beta_0}\right\} \\
&= -E_V\left\{ \sum_{d\in\mathcal{A}^*} h(d,V)\ \frac{\partial}{\partial\beta} m(d,V\mid\beta)\Big|_{\beta=\beta_0} \frac{\partial}{\partial\beta^T} m(d,V\mid\beta)\Big|_{\beta=\beta_0}\right\}. \qquad (25)
\end{aligned}
$$

B Simultaneous confidence bands

For a given a_1, a_2, we have that

$$\sqrt{n}(\beta_n^{IPTW}(a_1,a_2) - \beta_0(a_1,a_2)) \Rightarrow N(0, \Sigma^2(a_1,a_2)), \qquad (26)$$

where $\beta(a_1,a_2) \equiv (\beta_{a_1a_2,0}, \beta_{a_1a_2,1}, \beta_{a_1a_2,2})^T$ and $\Sigma^2(a_1,a_2)$ is the 3×3 matrix containing the corresponding elements of Σ^2. Let $\pi = (\pi_1 = E[Y(d(a_1,a_2,-50))], \pi_2 = E[Y(d(a_1,a_2,-49))], \ldots, \pi_{71} = E[Y(d(a_1,a_2,-20))])^T$. Then, letting

$$B = \begin{pmatrix} 1 & -50 & 2500 \\ 1 & -49 & 2401 \\ \vdots & \vdots & \vdots \\ 1 & 20 & 400 \end{pmatrix}, \qquad (27)$$

we have that

$$\pi = B\beta(a_1,a_2), \qquad (28)$$

so that

$$\sqrt{n}(\pi_n - \pi_0) \Rightarrow N(0, \Sigma^2_\pi(a_1, a_2)), \tag{29}$$

where $\Sigma^2_\pi(a_1, a_2) = B\Sigma^2(a_1, a_2)B^T$. We would like to find a constant a such that

$$P\left(\pi_{j,0} \in \left(\pi_{j,n} \pm a\frac{\sigma_\pi(a_1, a_2)(j, j)}{\sqrt{n}}\right) \quad \forall\, j\right) \to 0.95, \tag{30}$$

where $\sigma_\pi(a_1, a_2)(j, j) = \sqrt{\Sigma^2_\pi(a_1, a_2)(j, j)}$. Since this is equivalent to

$$P\left(\max_j \left|\frac{\sqrt{n}(\pi_{j,n} - \pi_{j,0})}{\sigma_\pi(a_1, a_2)(j, j)}\right| < a\right) \to 0.95, \tag{31}$$

a can be found as follows. Let $\rho(a_1, a_2)$ denote the correlation matrix corresponding to $\Sigma^2_\pi(a_1, a_2)$. Simulate a large number of realizations $Z = (Z_1, \ldots, Z_{71})^T \sim N(0, \rho(a_1, a_2))$. For each Z, let $M_Z = max_j|Z_j|$ denote the largest absolute value of any of its components. Then a can be selected as the 95th quantile of the simulated values M_Z.

C Inference for $\tilde{m}(a_1, a_2)$ and $\tilde{M}(a_1, a_2)$

For a given a_1, a_2, consider the maps

$$\phi_{a_1a_2} : \beta \to \phi_{a_1a_2}(\beta) \equiv \frac{\beta_{a_1a_2,1}}{2\beta_{a_1a_2,2}} \tag{32}$$

and

$$\psi_{a_1a_2} : \beta \to \psi_{a_1a_2}(\beta) \equiv \beta_{a_1a_2,0} - \frac{\beta^2_{a_1a_2,1}}{2\beta_{a_1a_2,2}} + \frac{\beta_{a_1a_2,1}}{2}. \tag{33}$$

If $\beta_{a_1a_2,2} < 0$, then $\phi_{a_1a_2}$ maps β into the value $\tilde{m}(a_1, a_2)$ that maximizes

$$\beta \to E[Y(d(a_1, a_2, \theta)) \mid \beta] = \beta_{a_1a_2,0} + \beta_{a_1a_2,1}\theta + \beta_{a_1a_2,2}\theta^2, \tag{34}$$

and $\psi_{a_1a_2}$ maps β into the corresponding maximum value $\tilde{M}(a_1, a_2) = E[Y(d(a_1, a_2, \theta)) \mid \beta = \tilde{m}(a_1, a_2)]$.

Since

$$\sqrt{n}(\beta^{IPTW}_n - \beta_1) \Rightarrow N(0, \Sigma^2), \tag{35}$$

it follows by the δ-method that

$$\sqrt{n}(\tilde{m}_n(a_1, a_2) - \tilde{m}_0(a_1, a_2)) \Rightarrow N(0, \dot{\phi}^T\Sigma^2\dot{\phi}) \tag{36}$$

and

$$\sqrt{n}(\tilde{M}_n(a_1, a_2) - \tilde{M}_0(a_1, a_2)) \Rightarrow N(0, \dot{\psi}^T\Sigma^2\dot{\psi}), \tag{37}$$

where $\dot{\phi}$ and $\dot{\psi}$ are the gradients of ϕ and ψ, respectively. In particular, we have that

$$\dot{\phi}_{a_1a_2} = \begin{pmatrix} 0 \\ -1/\beta_{a_1a_2,2} \\ \beta_{a_1a_2,1}/2\beta^2_{a_1a_2,2} \end{pmatrix} \tag{38}$$

and

$$\dot{\psi}_{a_1a_2} = \begin{pmatrix} 1 \\ -\beta_{a_1a_2,1}/\beta_{a_1a_2,2} + 0.5 \\ \beta^2_{a_1a_2,1}/2\beta^2_{a_1a_2,2} \end{pmatrix} \tag{39}$$