Demystifying Optimal Dynamic Treatment Regimes

Erica E. M. Moodie,[1,*] **Thomas S. Richardson**[2] **and David A. Stephens**[3]

[1] Department of Biostatistics, University of Washington, Seattle, WA 98195

[2] Department of Statistics, University of Washington, Seattle, WA 98195

[3] Department of Mathematics, Imperial College London, SW7 2AZ, U.K.

November 7, 2005

SUMMARY

A dynamic regime is a function that takes treatment and response history and baseline covariates as inputs and returns a decision to be made. Robins (2004) and Murphy (2003) have proposed models and developed semi-parametric methods for making inference about the optimal regime in a multi-interval trial that provide clear advantages over traditional parametric approaches. We show that Murphy's model is a special case of Robins' and that the methods are closely related but not completely equivalent; in doing this, we show that Murphy's estimates are not efficient. Interesting features of the methods are highlighted using the Multicenter AIDS Cohort Study (MACS) and through simulation.

Corresponding author's email address: eemm@u.washington.edu
Key words: Optimal dynamic regimes; Sequential randomization; Randomized controlled trials; Optimal structural nested mean models; Treatment algorithms.

1

1. Introduction

In a study aimed at estimating the mean effect of a treatment on a time-dependent outcome, it may be argued that dynamic treatment regimes are the most logical and ethical protocols to consider. A dynamic treatment regime is a function which takes in covariates, treatment and response history as arguments and outputs an action to be taken, providing a list of decision rules for how treatment should be allocated over time. One characteristic of a dynamic regime is that a subject's interval-specific treatment is not known at the start of the regime, since treatment depends on subsequent time-varying variables that may be influenced by earlier treatment.

The problem of finding the optimal dynamic regime is one of sequential decision-making, where an action which appears optimal in the short-term may not be a component of the optimal regime [Lavori (2000)]. We define a regime as optimal if it maximizes the mean response at the end of the final time interval.

There are many examples of adaptive intervention strategies in health care, ranging from treatment of AIDS to encouraging participation in mammography screening for breast cancer (see, for example, Robins (1994)). Yet there is a dearth of randomized trials that have implemented dynamic treatment protocols, due perhaps to the historical lack of theory for the design and analysis of such a trial. Recent work in the area has provided better insight into issues of randomization and sample size calculations [Murphy (2004), Lavori and Dawson (2001), Dawson and Lavori (2004)]. Design considerations for multi-center, sequentially randomized trials with adaptive randomization have been addressed in a Bayesian framework [Thall et al. (2000), Thall and Wathen (2005)]. Alongside the theoretical innovations, within-person sequentially-randomized trials

are being performed for treatment of mental illness [Rush et al. (2003), Schneider et al. (2001)] and cancer [Thall et al. (2000)]. However, the protocols for some of these trials call for analyses which do not take advantage of their sequential nature, but rather treat each phase as a separate trial.

Dynamic programming, also called backwards induction, is a traditional method of solving sequential decision problems [Bellman (1957), Bertsekas and Tsitsiklis (1996)], however in the dynamic regimes context, it requires modelling the longitudinal distribution of all status variables and outcome. The knowledge needed to model this is often unavailable and, by mis-specifying the distribution, it is possible to incorrectly recommend treatment when no treatment effect exists. The methods of Murphy (2003) and Robins (2004) do not suffer from this serious limitation.

Lavori et al. (1994) developed a causal approach to assessing optimal treatment discontinuation time for a binary outcome using propensity scores to adjust for time-varying covariates. Thall et al. (2000) produced a likelihood-based approach to analyze sequentially randomized trials for optimal regimes in the context of prostate cancer treatment, with additional complications: randomization probabilities changed as information from patients accrued and the outcome of interest was death or treatment failure in any interval.

The purpose of this article is to provide a clearer understanding of the models and methods proposed in the optimal dynamic regime literature: g-estimation and Murphy's iterative minimization, and to demonstrate the similarities between what may appear to be very different approaches. The following section introduces our motivating example, looking at the effect of AZT initiation on 12-month CD4 counts. In section 3, the procedure for estimating an optimal

3

dynamic regime is described, beginning with models and the optimal treatments which they imply, followed by an explanation of g-estimation and iterative minimization. The section concludes by contrasting the two methods. The methods are demonstrated using the Multicenter AIDS Cohort Study (MACS) in section 4, alongside simulations that highlight interesting features of the two methods.

2.   Context of the problem and motivating example

To examine the work of Murphy (2003) and Robins (2004) in a simple, two-interval example, we consider a subset of the MACS data [Kaslow et al. (1987)], a longitudinal observational study accumulating information from over 5000 HIV-1 infected homosexual and bisexual men in four U.S. cities beginning in 1984. Participants were invited to return for follow-up every six months, completing a questionnaire and a physical examination including blood work. We restrict our attention to the 2179 HIV-positive, AIDS-free men recruited after March 1986, when zidovudine (AZT) became available. Of those men, 38 (1.7%) were lost to follow-up before one year and 10 (0.5%) had initiated AZT before study entry; these were excluded from the analysis. We follow Hernan et al. (2000) in using last-observation carried forward to account for unobserved information due to missed visits.

To minimize notation, we will consider only two intervals - baseline to six months and six to 12 months into study - and use a single status variable, CD4 count, to determine the optimal rule for prescribing AZT at each interval. However, our development extends to the general case.

2.1   Notation

Treatments are given at two fixed times, t_1 and t_2. X_1 and X_2 are the status variables measured prior to treatment at the beginning of the first and

second intervals, respectively, i.e. at t_1 and t_2. In particular, X_1 represents baseline covariates and X_2 includes time-varying covariates which may depend on treatment received in the first interval. A_j, $j = 1, 2$, is the treatment given subsequent to observing X_j. Y is the outcome observed at the end of the second interval, and larger values of Y are deemed preferable. Thus, the order of occurrence is (X_1, A_1, X_2, A_2, Y) and the data can be depicted by a tree when X and A are categorical (Figure 1a). Let $\mathcal{H}_j$ denote the treatment and response history up to time j, so $\mathcal{H}_1 = X_1$ and $\mathcal{H}_2 = (X_1, A_1, X_2)$. Specific values will be denoted with the lower-case, e.g. $\mathfrak{h}_1 = x_1$.

In our example, X_1 is baseline CD4; X_2 is CD4 cell counts at six months; and Y is CD4 counts at 12 months. Treatment is the indicator of AZT commencement so that $A_1 = 1$ if AZT therapy was initiated between baseline and six months and A_2 is the equivalent indicator for initiation of AZT between six and 12 months. Participants who took AZT did not discontinue treatment, so the rules to be estimated are for starting of AZT (Figure 1b).

[Figure 1 about here.]

Throughout this paper, models will rely on *counterfactuals* (or *potential outcomes*), i.e. a person's outcome had he followed a particular treatment regime - possibly different from the regime that was actually followed. Let $X_2(a_1)$ denote a person's counterfactual status at the beginning of the second interval had treatment a_1 been received by that person and $Y(a_1, a_2)$ denote the end-of-study outcome if he had followed regime (a_1, a_2).

Counterfactuals adhere to the *axiom of consistency*: $X_2(a_1) = X_2$ whenever treatment a_1 is actually received and $Y(a_1, a_2) = Y$ whenever a_1 and a_2 are received. That is, the actual and counterfactual status are equal when the

5

regime in question is the regime actually received and similarly for outcome.

2.2 *Assumptions*

To estimate the effect of any dynamic regime (optimal or otherwise), we require:

1. *Stable Unit Treatment Value Assumption (SUTVA)*: A subject's outcome is not influenced by other subjects' treatment allocation [Rubin (1978)].

2. *No unmeasured confounders* : For any regime (a_1, a_2),
   $A_1 \perp (X_2(a_1), Y(a_1, a_2)) | \mathcal{H}_1$ and $A_2 \perp Y(a_1, a_2) | \mathcal{H}_2$ [Robins (1997)].

Assumption 2 (also called *sequential ignorability*) always holds under sequential randomization, that is, when treatment is randomly assigned at each interval with known probabilities (which may be a function of history).

Without further assumptions the optimal regime may only be estimated from among the set of *feasible* regimes [Robins (1994)]: Let $p_j(a_j | \mathfrak{h}_j)$ denote the conditional probability of receiving treatment a_j given history $\mathfrak{h}_j$ and let $f(\cdot)$ denote the density function of $\mathfrak{h}_2 = (x_1, a_1, x_2)$. Then for all $\mathfrak{h}_2$ with $f(\mathfrak{h}_2) > 0$, a *feasible regime* $(d_1(\mathfrak{h}_1), d_2(\mathfrak{h}_2))$ satisfies

$$p_1(d_1(\mathfrak{h}_1)|\mathfrak{h}_1) \times p_2(d_2(\mathfrak{h}_2)|\mathfrak{h}_2) > 0.$$

I.e., feasibility requires some subjects to have followed regime $(d_1(\mathfrak{h}_1), d_2(\mathfrak{h}_2))$ for the analyst to be able to estimate its performance non-parametrically. In terms of a decision tree, no (non-parametric) inference can be made of the effect of following a particular branch if no one followed that path. In particular, we cannot make inference about AZT discontinuation in the MACS data-set since no discontinuations were observed in the first year of study (Figure 1b).

It is unlikely that SUTVA is violated in the MACS example, as participants were drawn from four large cities. Further, we model the probability of initiating AZT as a function of CD4 count, presence of symptoms (such as thrush, herpes zoster, etc), pneumonia prophylaxis at the last visit, and year of cohort entry. It is plausible that there are few or no other variables that confound the association between CD4 and AZT initiation; we proceed, assuming that this is so.

3. Steps to finding the optimal regime

We define optimal rules recursively as follows:

$$d_2^{opt}(\mathfrak{h}_2) = \max_{d_2} E\big[Y\big(a_1, d_2(\mathfrak{h}_2)\big)\big|\mathcal{H}_2 = \mathfrak{h}_2\big],$$

$$d_1^{opt}(\mathfrak{h}_1) = \max_{d_1} E\left[Y\Big(d_1(\mathfrak{h}_1), d_2^{opt}\big(\mathfrak{h}_1, d_1(\mathfrak{h}_1), X_2(\mathfrak{h}_1, d_1(\mathfrak{h}_1))\big)\Big)\big|\mathcal{H}_1 = \mathfrak{h}_1\right].$$

Optimal regimes are defined for any sequence of treatment and covariate history, even a sequence $\mathfrak{h}_2$ that might not be possible to observe had the optimal regime been followed by all participants from the first interval. Thus, an optimal regime provides information not only on the best treatment choices from the beginning but also on treatment choices that maximize outcomes from a later time, even if a sub-optimal regime had been followed up to that point.

Robins [(1994), (1997)] pioneered the field of dynamic treatment regimes. However Murphy (2003) gave the first method to estimate regimes semi-parametrically. Following this, Robins (2004) produced a number of estimating equations for finding optimal regimes using structural nested mean models (SNMM). The three key steps to identifying the optimal dynamic treatment regime are (1) definition of the model, (2) finding the optimal rule implied by the model, and (3) estimation of model parameters. For (1), Robins uses "blip" models; Murphy uses "regrets". Both are variants of the SNMMs. For (3), Robins uses

g-estimation, while Murphy uses iterative minimization.

3.1 *Step 1: Model definition*

A SNMM defines an expected difference between a person's counterfactual responses on a specific treatment regime from time $j+1$ onwards and on another specific regime from time j conditional on history. We consider a particular class of SNMM's, those with optimal blip functions.

Define an *optimal blip-to-reference function* to be the expected difference in outcome when using a reference regime $d_j^{ref} = d_j^{ref}(\mathfrak{h}_j)$ instead of a_j at time t_j, in persons with treatment and covariate history $\mathfrak{h}_j$ who subsequently receive the optimal regime. At the first time-point, we have:

$$\gamma_1(\mathfrak{h}_1, a_1) = E\left[Y\left(a_1, d_2^{opt}(\mathfrak{h}_1, a_1, X_2(a_1))\right) - Y\left(d_1^{ref}, d_2^{opt}(\mathfrak{h}_1, d_1^{ref}, X_2(d_1^{ref}))\right)\Big|\mathcal{H}_1 = x_1\right],$$

and at the second,

$$\gamma_2(\mathfrak{h}_2, a_2) = E\left[Y(a_1, a_2) - Y(a_1, d_2^{ref}(\mathfrak{h}_2))\big|\mathcal{H}_2 = (x_1, a_1, x_2)\right].$$

The term "optimal" refers to treatment subsequent to t_j. At the second interval there are no subsequent treatments, so the blip is simply the expected difference in outcomes for having taken treatment a_2 as compared to the reference regime, d_2^{ref}, among people with treatment and response history $\mathfrak{h}_2$.

Two special cases of optimal blip-to-reference functions have been used in the dynamic regimes literature and applications:

The *optimal blip-to-zero* function, suggested by Robins (2004), takes the reference regime to be the "zero" regime at time j, a substantively meaningful regime such as placebo or standard care.

Murphy (2003) modelled the regret function. This function is the *negative* of the optimal blip that uses the optimal treatment at time j as the reference

regime. Denote this by

$$\mu_1(\mathfrak{h}_1, a_1) = E\Big[Y\big(a_1, d_2^{opt}(\mathfrak{h}_1, a_1, X_2(a_1))\big)-$$

$$Y\big(d_1^{opt}(\mathfrak{h}_1), d_2^{opt}(\mathfrak{h}_1, d_1^{opt}, X_2(d_1^{opt}))\big)\big|\mathcal{H}_1 = x_1\Big],$$

$$\mu_2(\mathfrak{h}_2, a_2) = E\big[Y\big(a_1, a_2\big) - Y\big(a_1, d_2^{opt}(\mathfrak{h}_2)\big)\big|\mathcal{H}_2 = (x_1, a_1, x_2)\big].$$

The regret at t_j is the expected difference in the outcome had the optimal treatment been taken at t_j, instead of treatment a_j, in participants who followed regime a up to t_j and the optimal regime from t_{j+1} onwards.

Optimal blip-to-zero functions and regrets correspond directly:

$$\mu_j(\mathfrak{h}_j, a_j) = \underset{a}{max}\, \gamma(\mathfrak{h}_j, a) - \gamma(\mathfrak{h}_j, a_j),$$

$$\gamma_j(\mathfrak{h}_j, a_j) = \mu_j(\mathfrak{h}_j, d_j^{ref}) - \mu_j(\mathfrak{h}_j, a_j).$$

It can be shown that if the regret is smooth in its arguments (or parameters), the optimal blip-to-zero will be also; the converse does not hold. Both optimal blips and regrets compare the counterfactual outcomes in which treatment at time $j+1$ and thereafter is optimal; regrets additionally posit that treatment at time j is optimal. Henceforth, take $d_j^{ref} = 0$ in all optimal blip functions except regrets.

While Robins generally advocates optimal blip-to-zero functions and Murphy, regrets, a simple transformation can be used to go from one to the other. However, it is important to be aware that simple forms for either model can lead to complex - and perhaps unlikely - forms for observables. For example, if outcome under the optimal regime depends linearly on status variables, X_j, and we assume a linear blip function, this implies that observed outcome, Y, is piece-wise linear in X_j, and not necessarily continuous.

3.2 *Step 2: Identification of the optimal rules*

If given the true form of the optimal blip (or alternatively, of the regret function) parameterized by ψ, it is straight-forward to identify the optimal regime. For all j, it is:

$$d_j^{opt}(\mathfrak{h}_j, a_j; \psi) = \arg\max_{a_j} \gamma_j(\mathfrak{h}_j, a_j; \psi);$$

or, using regrets, it is the regime $d_j^{opt}(\mathfrak{h}_j, a_j; \psi)$ such that

$$\mu_j(\mathfrak{h}_j, d_j^{opt}(\mathfrak{h}_j, a_j; \psi)) = 0.$$

Define $\mathcal{D}_j(\gamma)$ to be the set of rules, d_j^{opt}, that are optimal under the optimal blip function model $\gamma_j(\mathfrak{h}_j, a_j; \psi)$ as ψ is varied:

$$\mathcal{D}_j(\gamma) = \{d_j(\cdot) | d_j(\mathfrak{h}_j) = \arg\max_{a_j} \gamma_j(\mathfrak{h}_j, a_j; \psi) \text{ for some } \psi\}.$$

Let $\mathcal{D}_j(\mu)$ be the set of optimal rules that are compatible with regret $\mu_j(\mathfrak{h}_j, a_j; \psi)$:

$$\mathcal{D}_j(\mu) = \{d_j(\cdot) | \mu_j(\mathfrak{h}_j, d_j(\mathfrak{h}_j); \psi) = 0 \text{ for some } \psi\}.$$

Murphy [(2003), p.345] models the regret for a discrete decision by a smooth approximation, $\text{expit}(x) = e^x(e^x + 1)^{-1}$, to facilitate estimation. Using an approximation, $\tilde{\mu}_j(\mathfrak{h}_j, a_j)$, to the true regret model, $\mu_j(\mathfrak{h}_j, a_j)$, let

$$\mathcal{D}_j(\tilde{\mu}) = \{d_j(\cdot) | d_j(\mathfrak{h}_j) = \arg\min_{a_j} \tilde{\mu}_j(\mathfrak{h}_j, a_j; \psi) \text{ for some } \psi\}$$

denote the set of optimal rules that are compatible with $\tilde{\mu}_j(\mathfrak{h}_j, a_j)$. *The approximate regret may not equal zero at the optimal regime.*

Problems may arise if parameterization of the true SNMM is poorly chosen. For example, suppose the optimal blip, $\gamma_j(\mathcal{H}_j; \psi) = a_j f(X_j; \psi)$, is such that $f(X_j; \psi) = \psi_0 + \psi_1 X_j$ with treatment a_j binary. The corresponding regret is

$\mu_j(\mathcal{H}_j, a_j) = |\psi_0 + \psi_1 X_j| \times (a_j - I[\psi_0 + \psi_1 X_j > 0])^2$ and $\mathcal{D}_j(\gamma) = \mathcal{D}_j(\mu) = \{I[\psi_0 + \psi_1 X_j > 0]\}$. Suppose $\psi_1 > 0$ so that treatment is beneficial if X_j is above the threshold $\beta = -\psi_0/\psi_1$. We may re-parameterize the regret to obtain the threshold, β: $\mu_j^*(\mathcal{H}_j, a_j) = |X_j - \beta| \times (a_j - I[X_j - \beta > 0])^2$, which gives $\mathcal{D}_j(\mu^*) = \{I[X_j - \beta > 0]\}$. However, if $\psi_1 < 0$ so that now subjects should be treated when $X_j < \beta$, $\mu_j^*(\mathfrak{h}_j, a_j) = |X_j - \beta| \times (a_j - I[X_j - \beta < 0])^2$ and so $\mathcal{D}_j(\mu^*) = \{I[X_j - \beta < 0]\}$. Thus, one consequence of using the re-parameterized regret in this form is that whether it is optimal to treat for high or low status values must be known in advance. Incorrectly specifying the direction can lead to false conclusions such as failure to detect a treatment effect. This can be overcome by using a richer class of models for the regret, such as the two-parameter model in this example. (See reply to discussion in [Murphy (2003)].)

3.3 *Step 3: Estimation*
3.3.1. g-estimation

Robins (2004) proposes finding the parameters ψ of the optimal blip-to-zero function or regret function via g-estimation. Define

$$H_1(\psi) = Y + \sum_{j=1}^{2} \left[\gamma_j(\mathcal{H}_j, d_j^{opt}; \psi) - \gamma_j(\mathcal{H}_j, a_j; \psi) \right],$$

$$H_2(\psi) = Y + \gamma_2(\mathcal{H}_2, d_2^{opt}; \psi) - \gamma_2(\mathcal{H}_2, a_2; \psi).$$

$H_j(\psi)$ is a patient's actual outcome adjusted by the expected difference between the average outcome for someone with treatment and covariate history $\mathfrak{h}_j$ who is treated optimally from time t_j and someone with history $(\mathfrak{h}_j, a_j)$ who is subsequently treated optimally from time t_{j+1}.

Under additive local rank preservation (LRP),

$$H_2(\psi) = Y\big(a_1, d_2^{opt}(X_1, a_1, X_2(a_1))\big),$$

$$H_1(\psi) = Y\left(d_1^{opt}(X_1), d_2^{opt}\left(X_1, d_1^{opt}(X_1), X_2(d_1^{opt}(X_1))\right)\right)$$

[Robins (2004)]. Loosely, LRP states that the ranking of patients' outcomes under a particular regime is the same as their ranking under another regime, conditional on history. LRP is additive if the difference in person's outcome should he be treated with one regime instead of the other given history equals the expected difference. Rank preservation provides a simplistic situation in which the parameters of a SNMM may be interpreted at the individual level. However SNMM's may be used without making such assumptions via a population-level interpretation in terms of average causal effects.

For the purpose of estimation, specify $S_j(a_j) = s_j(a_j, \mathfrak{h}_j) \in \mathbb{R}^{dim(\psi_j)}$ which depends on variables which are thought to interact with treatment to influence outcome. For example, if the optimal blip at the second interval is linear,

$$\gamma_2(\mathfrak{h}_2, a_2) = a_2(\psi_0 + \psi_1 x_2 + \psi_2 a_1 + \psi_3 x_2 a_1),$$

the analyst may choose $S_2(a_2) = \frac{\partial}{\partial \psi}\gamma_2(\mathfrak{h}_2, a_2) = a_2 \cdot (1, x_2, a_1, x_2 a_1)^T$. Let

$$U(\psi, s) = \sum_{j=1}^{2} H_j(\psi)\{S_j(A_j) - E[S_j(A_j)|\mathcal{H}_j]\}, \tag{1}$$

with the probability of being treated modelled (perhaps non-parametrically) by $p_j(a_j|\mathfrak{h}_j; \alpha)$. $E[U(\psi, s)] = 0$ is an unbiased estimating equation from which consistent, asymptotically Normal estimates $\hat{\psi}$ of ψ may be found under standard regularity conditions provided the treatment model is correct and the optimal regime is unique, though see [Robins (2004)]. The intuition behind equation (1) is that counterfactual outcomes under different treatment regimes at time j are independent of any function of actual treatment conditional on prior treatment and covariates (by Assumption 2). These estimators are not efficient.

Robins (2004) refined equation (1) to gain efficiency. Let

$$U^\dagger(\psi, s, \hat{\alpha}) = \sum_j^2 (H_j(\psi) - E[H_j(\psi)|\mathfrak{h}_j])\{S_j(A_j) - E[S_j(A_j)|\mathfrak{h}_j]\}. \quad (2)$$

The inclusion of $E[H_j(\psi)|\mathfrak{h}_j]$ in (2) gives estimates which are more efficient than those found using (1), however these estimates are still not efficient. Robins proves that estimates found by (2) are consistent provided *either* $E[H_j(\psi)|\mathfrak{h}_j]$ *or* $p_j(a_j|\mathfrak{h}_j)$ is correctly modelled, and thus is said to be doubly-robust. Semi-parametric efficient estimates can be found with good choice of $S(A_j)$, although its form is often complex.

Correct specification of $E[H_j(\psi)|\mathfrak{h}_j]$ requires knowing the functional dependence of outcome on history. For simplicity, consider the case of binary treatment. As noted earlier, $\gamma_j(\mathfrak{h}_j; \psi) = a_j f(x_j; \psi)$ and $\mu_j(\mathfrak{h}_j, a_j) = |f(x_j; \psi)| \times (a_j - I[f(x_j; \psi) > 0])^2$ specify the same SNMM so that if $\gamma_j(\mathfrak{h}_j; \psi)$ is linear in $\mathfrak{h}_j$, $\mu_j(\mathfrak{h}_j, a_j)$ is piece-wise linear. Expressing $H_j(\psi)$ as $Y + \sum_{m=j}^2 \mu_m(\mathfrak{h}_m, a_m)$, we see that if the mean of Y depends linearly on $\mathfrak{h}_j$, then $E[H_j(\psi)|\mathfrak{h}_j]$ is piece-wise linear with discontinuities and changes in slope occurring at optimal rule thresholds. Simulations suggest that even when $p_j(a_j|\mathfrak{h}_j)$ is wrongly specified, using an incorrect model for $E[H_j(\psi)|\mathfrak{h}_j]$ in (2) yields estimates that are less biased and less variable than using no model at all, i.e., using (1) (see Table 1).

3.3.2. Recursive, closed-form g-estimation

In general, search algorithms are required to find the values of $\hat{\psi}$ to satisfy the g-estimating equation. Exact solutions can be found when optimal blips are linear in ψ and parameters are not common (shared) between intervals. An example of blip functions that are linear in ψ but *do* have common parameters between intervals is $\gamma_1(x_1, a_1) = a_1(\psi_0 + \psi_1 x_1)$ and $\gamma_2(\mathfrak{h}_2, a_2) = a_2(\psi_0 + \psi_1 x_2 + \psi_2 a_1)$,

13

since ψ_0 and ψ_1 appear in the blip functions of both intervals.

We may also use the modification

$$H_{\mathrm{mod},1}(\psi) = Y - \gamma_1(\mathfrak{h}_1, a_1; \psi) + \left[\gamma_2(\mathfrak{h}_2, d_2^{opt}; \psi) - \gamma_2(\mathfrak{h}_2, a_2; \psi) \right]$$

$$H_{\mathrm{mod},2}(\psi) = Y - \gamma_2(\mathfrak{h}_2, a_2; \psi)$$

in equation (1) or (2) without changing the consistency of the resulting g-estimates. Under LRP, $H_{\mathrm{mod},1}(\psi) = Y\big(0, d_2^{opt}(X_1, 0, X_2(0))\big)$ and $H_{\mathrm{mod},2}(\psi) = Y\big(a_1, d_2^{opt}(X_1, a_1, X_2(a_1))\big)$. This modification allows recursive estimation when parameters are not shared: find first $\hat{\psi}_2$ at the last interval then plug $\hat{\psi}_2$ into $H_{\mathrm{mod},1}(\psi)$ to find $\hat{\psi}_1$. Postulating models for the mean dependence of Y on $\mathfrak{h}_j$ and for $\gamma_j(\mathfrak{h}_j; \psi)$ is sufficient to determine a model for $E[H_{\mathrm{mod},j}(\psi)|\mathfrak{h}_j]$.

3.3.3. Iterative Minimization for Optimal Regimes (IMOR)

Murphy (2003) developed a method that estimates the parameters of the optimal regime, ψ, by searching for $(\hat{\psi}, \hat{c})$ which satisfy

$$\sum_{j=1}^{2} \mathbb{P}_n \Big[Y + \hat{c} + \sum_{l=1}^{2} \mu_l(\mathfrak{h}_l, a_l; \hat{\psi}) - \sum_a \mu_j(\mathfrak{h}_j, a; \hat{\psi}) p_j(a|\mathfrak{h}_j; \hat{\alpha}) \Big]^2$$

$$\leq \sum_{j=1}^{2} \mathbb{P}_n \Big[Y + c + \sum_{l=1, l \neq j}^{K} \mu_l(\mathfrak{h}_l, a_l; \hat{\psi}) + \mu_j(\mathfrak{h}_j, a_j; \psi)$$
$$- \sum_a \mu_j(\mathfrak{h}_j, a; \psi) p_j(a|\mathfrak{h}_j; \hat{\alpha}) \Big]^2 \quad (3)$$

for all c and all ψ, where $\mathbb{P}_n(f) = n^{-1} \sum_{i=1}^{n} f(X_i)$ is the empirical average. $\hat{\psi}$ is consistent for ψ provided the treatment allocation probabilities, $p_j(a_j|\mathfrak{h}_j)$, are correctly estimated. $\hat{\psi}$ is not efficient, as will be seen in the next sub-section.

Murphy (2003) describes an iterative method of finding solutions to (3), which begins by selecting an initial value of $\hat{\psi}$, say $\hat{\psi}^{(1)}$, then minimizing the RHS of the equation over (ψ, c) to obtain a new value of $\hat{\psi}$, $\hat{\psi}^{(2)}$, and repeating

this until convergence. This method may not produce a monotonically decreasing sequence of RHS values of equation (3). Further, this procedure may not converge to a minimum; profile plots of the RHS of (3) for each parameter in an interval about its estimate provide a useful diagnostic tool.

3.4   *Relating the methods for 2 intervals*

Suppose $X_1, A_1, X_2, A_2,$ and Y are observed where A_j is binary and X_j, Y are univariate for $j = 1, 2$. Further suppose that parameters are not shared across intervals. Robins [(2004), Corollary 9.2] proves that for an optimal blip $\gamma_j(\mathfrak{h}_j, a_j; \psi_j)$, the unique function $q(\mathfrak{h}_j, a_j)$ minimizing

$$
E\Bigg[\Big\{Y - q(\mathfrak{h}_j, a_j) + \sum_{m=j+1}^{2} (\gamma_m(\mathfrak{h}_m, d_m^{opt}; \psi_m) - \gamma_m(\mathfrak{h}_m, a_m; \psi_m))
$$
$$
- E\Big[Y - q(\mathfrak{h}_j, a_j) + \sum_{m=j+1}^{2} (\gamma_m(\mathfrak{h}_m, d_m^{opt}; \psi_m) - \gamma_m(\mathfrak{h}_m, a_m; \psi_m)) \,\Big|\, \mathfrak{h}_m \Big]\Big\}^2 \Bigg]
$$
(4)

subject to $q(\mathfrak{h}_j, 0) = 0$ is $\gamma_j(\mathfrak{h}_j, a_j; \psi_j)$. To make use of (4) to estimate ψ_1, $\hat{\psi}_2$ must have already been found - i.e., estimation is recursive, not simultaneous.

At each interval, g-estimation is equivalent to minimizing (4) by setting its derivative to zero. At the minimum, $q(\mathfrak{h}_j, a_j) = \gamma_j(\mathfrak{h}_j, a_j; \psi_j)$ and so

$$
Y - q(\mathfrak{h}_j, a_j) + \sum_{m=j+1}^{K} \big[\gamma(\mathfrak{h}_m, d_m^{opt}; \psi_m) - \gamma(\mathfrak{h}_m, a_m; \psi_m)\big] = H_{\text{mod},j}(\psi_j).
$$

With $S(a_j) = -\frac{\partial}{\partial \psi_j} q(\mathfrak{h}_j, a_j)$, equation (4) leads to g-estimating equation (2) using the modified version of $H_j(\psi)$.

IMOR is another method of recursive minimization. At any interval j, taking $q(\mathfrak{h}_j, a_j) = \mu_j(\mathfrak{h}_j, a_j; \psi)$ and

$$
c = -\mu_j(\mathfrak{h}_j, 0; \psi_j) - \sum_{m=1}^{j-1} \mu_m(\mathfrak{h}_m, a_m; \hat{\psi}_m)
$$
$$
+ E\Big[\mu_j(\mathfrak{h}_j, 0; \psi_j) + \sum_{m=j+1}^{K} \mu_m(\mathfrak{h}_m, a_m; \hat{\psi}_m) - Y \,\Big|\, \mathfrak{h}_j\Big]
$$

in (4) leads to the RHS of (3) on an interval-by-interval basis. The parameter c in (3) is *not* interval-specific, so the two methods are not identical. This is a critical difference between the methods: IMOR does not model $E[H_{\mathrm{mod},j}(\psi)|\mathfrak{h}_j]$ explicitly, but rather captures the quantity through the regrets and $\hat{c}$. If the researcher correctly specifies the model for $E[H_{\mathrm{mod},j}(\psi)|\mathfrak{h}_j]$, he can obtain more precise estimates using g-estimation as compared to IMOR.

4. Examples

4.1 *Simulation results*

Via simulations we compare the performance of the methods discussed in the previous section, as well as illustrate the double-robustness of g-estimating equation (2). Suppose that patients are accrued in a trial whose purpose is to estimate the optimal rule for AZT initiation. Patients will be randomized to either no treatment or AZT at baseline and those who did not receive treatment at baseline will be re-randomized at 6 months to receive either no treatment or AZT. (Clearly, such a trial would be unlikely given the current understanding of the beneficial effects of AZT!)

Variables are as described in §2.1 and were generated as follows: baseline CD4: $X_1 \sim \mathcal{N}(450, 150)$; six-month CD4: $X_2 \sim \mathcal{N}(1.25X_1, 60)$; and one-year CD4: $Y \sim \mathcal{N}(400 + 1.6X_1, 60) - \mu_1(\mathcal{H}_1, A_1) - \mu_2(\mathcal{H}_2, A_2)$. Note that, as stated before, the observed outcome is not linear in X_1 and X_2. Treatments A_1, A_2 were randomly assigned with equal probability and optimal blips are linear: $\gamma_1(\mathfrak{h}_1, a_1) = a_1(\psi_{10} + \psi_{11}x_1)$, and $\gamma_2(\mathfrak{h}_2, a_2) = a_2(\psi_{20} + \psi_{21}x_2)$. We use $S_j(a_j) = \frac{\partial}{\partial \psi}\gamma_j(\mathfrak{h}_j, a_j)$ in g-estimation for greater similarity to IMOR.

Two models were considered for $E[H_{\mathrm{mod},j}(\psi)|\mathfrak{h}_j]$ when implementing g-estimation (2). The first incorrectly assumed that $E[H_{\mathrm{mod},j}(\psi)|\mathfrak{h}_j]$ depends

16

linearly on all of $\mathfrak{h}_j$. The second, correct model allowed the mean function to be piece-wise, discontinuous linear with inflections at the optimal rule thresholds (see Appendix). All results are presented in Table 1; diagnostic plots for the IMOR approach are in Figure 2.

[Table 1 about here.]

[Figure 2 about here.]

The efficiency gained by using the g-estimating equation (2) instead of (1) is considerable; using the incorrect model for $E[H_{\mathrm{mod},j}(\psi)|\mathfrak{h}_j]$ also leads to reduced efficiency. IMOR estimates are slightly biased and efficiency is lower, although much better than equation (1).

Suppose now that physicians broke protocol, so that the probability of initiating AZT is higher in patients with low CD4 counts: $A_j \sim \mathrm{Binom}(p_j)$, where $p_1 = \mathrm{expit}(2 - .006X_1)$ and $p_2 = \mathrm{expit}(0.8 - .004X_2)$. This new randomization scheme used depends only on the *observed* variable CD4. If the analyst incorrectly assumed complete randomization, only equation (2) using the correct model for $E[H_{\mathrm{mod},j}(\psi)|\mathfrak{h}_j]$ yields unbiased estimates (Table 1). In this example, using a linear model for $E[H_{\mathrm{mod},j}(\psi)|\mathfrak{h}_j]$ in equation (2) on average yields optimal decision thresholds (95% CI) not too far from the truth: begin AZT at baseline for patients with CD4 counts below 226 (194, 258), and begin therapy at 6 months if counts are below 325 (271, 378) as compared to the true best regime thresholds of 250 and 360 counts at the first and second interval, respectively. IMOR and g-estimation using equation (1) are seen to be not at all robust to mis-specification of the treatment model, as expected.

17

Turning our attention to the MACS data: 142 (6.7%) participants initiated AZT in the first six months of the study; a further 166 (7.8%) began treatment between six and 12 months. Initial plots show little difference in 12-month CD4 counts between those who were treated and those who were not (Figure 3).

[Figure 3 about here.]

[Table 2 about here.]

Initially, treatment was fit as a function of CD4 at the previous visit only. It is unlikely that this scenario reflects the true decision-making process of physicians, so the analysis was repeated using richer treatment models which were selected using the Bayesian Information Criterion. The richer models found year of study entry and presence of symptoms at baseline as well as baseline CD4 to predict treatment in the first six months of study. Six-month CD4, use of *Pneumocystis Carinii* pneumonia prophylactics in the first six months of study, and presence of symptoms at six months were predictive of AZT initiation between six and 12 months. Neither g-estimation nor IMOR detected any effect of AZT initiation at any time in the first year on 12-month CD4 counts (Table 2). This analysis should not undermine the usefulness of AZT as a treatment for HIV. It may suggest that one-year CD4 counts are not sufficient to capture beneficial effects of the therapy or are not a good surrogate for HIV-patient health.

A naive linear regression of 12-month CD4 counts on baseline CD4, six-month CD4, and treatments A_1 and A_2 picks up a strong association between initiation of AZT in the second interval and outcome ($p < 0.001$): in this model, participants who started AZT between six and 12 months had, on average, 12-

18

month CD4 counts that were 74 (44, 104) *lower* than those who did not initiate therapy. A non-significantly lower average CD4 count was also observed for AZT initiation between zero and six months. Residual plots suggested heteroscedasticity. Log-transforming the outcome did not remove the strong statistical significance of the association, nor did including of the covariates from the richer treatment models used in the dynamic regimes methods or interaction terms.

The negative association between treatment in the second interval and one-year CD4 in linear regression can reasonably be explained by confounding: patients with low CD4 counts were more likely to use AZT. This example nicely demonstrates the utility of dynamic regimes in general, particularly the importance of causal models that are correctly specified under the null hypothesis.

5.   Conclusion

Our paper has clarified the connections between both the models and the methods used to make inference in the context of dynamic treatment regimes. Blip functions are highly flexible and can be used to describe a number of different mean models, including regret functions. While the methods of Robins and Murphy appear to be very different at first glance, we have shown that the methods are based on a similar minimization. In fact, g-estimation and IMOR are conceptually nearly equivalent.

The methods discussed here, along with the advances in theory needed to implement clinical studies of dynamic regimes mentioned in the section 1, have the potential to contribute greatly to the design of treatment protocols for a variety of medical conditions. Robins and Murphy developed their methods in the general, K-interval case and so they are widely applicable although sample size requirements may rapidly become infeasible with a large number of treatment

options, unless it is reasonable to impose stationarity (i.e., parameter-sharing).

We leave the reader with a final word of caution: Model choice should be driven

by practical considerations, however it is important to be aware of the (perhaps

implausible) models for observables which are then implied.

References

Bellman, R. (1957). *Dynamic Programming*. Princeton University Press, Princeton, NJ.

Bertsekas, D. P. and Tsitsiklis, J. N. (1996). *Neuro-dynamic Programming*. Athena Scientific, Belmont.

Brumback, B., Greenland, S., Redman, M., Kiviat, N. and Diehr, P. (2003). The intensity score approach to adjusting for confounding. *Amer. J. Epi.* **59**, 274–285.

Dawson, R. and Lavori, P. W. (2004). Placebo-free designs for evaluating new mental health treatments: the use of adaptive treatment strategies. *Statist. Med.* **23**, 3249–3262.

Hernan, M. A., Brumback, B. and Robins, J. M. (2000). Marginal structural

models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Amer. J. Epi.* **11**, 561–570.

Kaslow, R. A., Ostrow, D. G., Detels, R., Phair, J. P., Polk, B. F. and Rinaldo, C. R. (1987). The Multicenter AIDS Cohort Study: rationale, organization, and selected characteristics of the participants. *Amer. J. Epi.* **126**, 310–318.

Lavori, P. W. (2000). A design for testing clinical strategies: biased adaptive within-subject randomization. *J. Roy. Statist. Soc. A* **163**, 29–38.

Lavori, P. W. and Dawson, R. (2001). Dynamic treatment regimes: practical design considerations. *Statist. Med.* **20**, 1487–1498.

Lavori, P. W., Dawson, R. and Mueller, T. B. (1994). Causal estimation of time-varying treatment effects in observational studies: application to depressive disorder. *Statist. Med.* **13**, 1089–1100.

Murphy, S. A. (2003). Optimal dynamic treatment regimes. *J. Roy. Statist. Soc. B* **65**, 331–366.

Murphy, S. A. (2004). An experimental design for the development of adaptive treatment strategies. *Statist. Med.* **24**, 1455–1481.

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

Robins, J. M. (1997). *Latent Variable Modeling and Applications to Causality*, chapter Causal Inference from Complex Longitudinal Data, pages 69–117. Springer-Verlang, New York.

Robins, J. M. (2004). Optimal structural nested models for optimal sequential decisions. In Lin, D. Y. and Heagerty, P., editors, *Proceedings of the Second Seattle Symposium on Biostatistics*, New York. Springer.

Rubin, D. B. (1978). Bayesian inference for causal effects: The role of random-

ization. *Ann. Statist.* **6**, 34–58.

Rush, A. J., Fava, M., Wisniewski, S. R., Lavori, P. W., Trivedi, M. H. and
Sackeim, H. A. (2003). Sequenced treatment alternatives to relieve depression
(STAR*D): rationale and design. *Con. Clin. Trials* **25**, 119–142.

Schneider, L. S., Tariot, P. N., Lyketsos, C. G., Dagerman, K. S., Davis, K. L.
and Davis, S. (2001). National institute of mental health clinical antipsy-
chotic trials of intervention effectiveness (CATIE): Alzheimer disease trial
methodology. *Am. J. Ger. Psych.* **9**, 346–360.

Thall, P. F., Millikan, R. E. and Sung, H.-G. (2000). Evaluating multiple treat-
ment courses in clinical trials. *Statist. Med.* **19**, 1011–1028.

Thall, P. F. and Wathen, J. K. (2005). Covariate-adjusted adaptive random-
ization in a sarcoma trial with multi-state treatments. *Statist. Med.* **24**,
1947–1964.

Appendix

Assuming a linear model for the optimal blip function with binary treatment
is equivalent to assuming a broken-stick model for the regret. The assumption
of a model for the blip/regret induces a model for the distribution of outcome
(see, for example, Murphy (2003)). Together, the blip or regret model and the
distribution of Y specify the correct form for $E[H_{\mathrm{mod},j}(\psi)|\mathfrak{h}_j]$.

In the model of §4.1, $E[H_{\mathrm{mod},1}(\psi)|\mathfrak{h}_1] = 400 + 1.6X_1$ if $X_1 \geq 250$ (equiva-
lently, if $\psi_{10} + \psi_{11}X_1 > 0$) and $E[H_{\mathrm{mod},1}(\psi)|\mathfrak{h}_1] = 400 + 1.6X_1 + \psi_{10} + \psi_{11}X_1$
otherwise. To see why this is so, note that if $X_1 \geq 250$, then the patient *should
not* be treated and so the optimal outcome, $400 + 1.6X_1$, will be observed under
$H_{\mathrm{mod},1}(\psi) = Y(0, d_2^{opt})$.

At the second interval, $E[H_{\mathrm{mod},2}(\psi)|\mathfrak{h}_2]$ can take one of four forms, depending on subgroups defined by the value of A_1, $X_1 < 250$, and $X_2 < 360$:

$A_1 = 1$ and $X_1 < 250$, $E[H_{\mathrm{mod},2}(\psi)|\mathfrak{h}_2] = 400 + 1.6X_1$

$A_1 = 1$ and $X_1 \geq 250$, $E[H_{\mathrm{mod},2}(\psi)|\mathfrak{h}_2] = 400 + 1.6X_1 + \psi_{10} + \psi_{11}X_1$

$A_1 = 0$, $X_1 \geq 250$, and $X_2 \geq 360$, $E[H_{\mathrm{mod},2}(\psi)|\mathfrak{h}_2] = 400 + 1.6X_1$

$A_1 = 0$, $X_1 < 250$, and $X_2 \geq 360$,

$\qquad E[H_{\mathrm{mod},2}(\psi)|\mathfrak{h}_2] = 400 + 1.6X_1 + \psi_{10} + \psi_{11}X_1$

$A_1 = 0$, $X_1 \geq 250$, and $X_2 < 360$,

$\qquad E[H_{\mathrm{mod},2}(\psi)|\mathfrak{h}_2] = 400 + 1.6X_1 + \psi_{20} + \psi_{21}X_2$

$A_1 = 0$, $X_1 < 250$, and $X_2 < 360$,

$\qquad E[H_{\mathrm{mod},2}(\psi)|\mathfrak{h}_2] = 400 + 1.6X_1 + \psi_{10} + \psi_{11}X_1 + \psi_{20} + \psi_{21}X_2.$

Note that $A_1 = 1$ implies $A_2 = 0$, i.e. if AZT is started between baseline and six months, it cannot be initiated between six and 12 months. Thus, if a participant initiated AZT in the first interval, only a single treatment possible in the second interval (continue taking AZT) so that this treatment is optimal by virtue of it being the only option.

These derivations have an important implication: the parameters ψ appear *in the design matrix* for $E[H_{\mathrm{mod},j}(\psi)|\mathfrak{h}_j]$ under a linear model. Not only that, parameters from the first interval appear in the design matrix for $E[H_{\mathrm{mod},2}(\psi)|\mathfrak{h}_2]$ which rather takes away from the recursive g-estimation that we would like to use when (ψ_{10}, ψ_{11}) are not assumed to equal (ψ_{20}, ψ_{21}). Thus, using the correct model for $E[H_{\mathrm{mod},j}(\psi)|\mathfrak{h}_j]$ requires iterating between estimating $E[H_{\mathrm{mod},j}(\psi)|\mathfrak{h}_j]$ and solving the g-estimating equations.
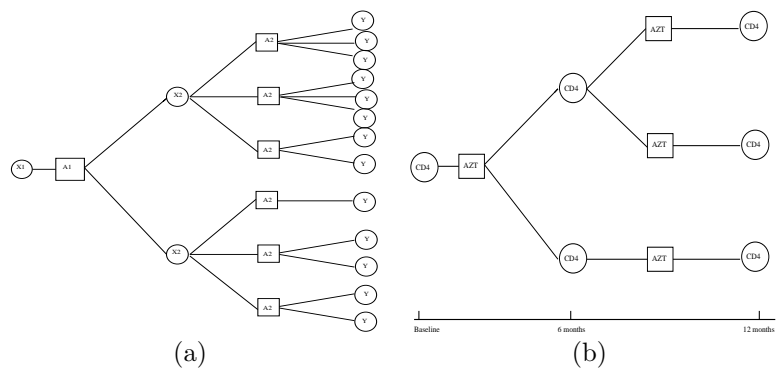
Figure 1. Illustration of data for two intervals: (a) generic and (b) MACS.
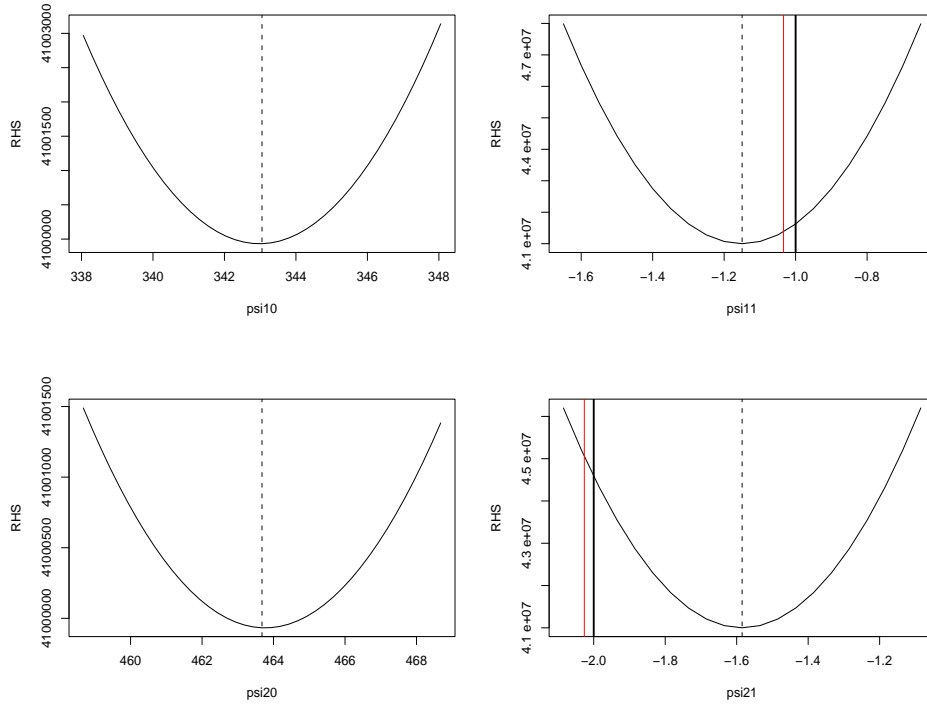
Figure 2. Profiles of the RHS of equation (3) for the IMOR approach from a single simulation. The dashed line is the IMOR estimate, the red line from g-estimation (2) using the correct, piece-wise model for $E[H_{\mathrm{mod},j}(\psi)|\mathfrak{h}_j]$, and the thick black line the truth.
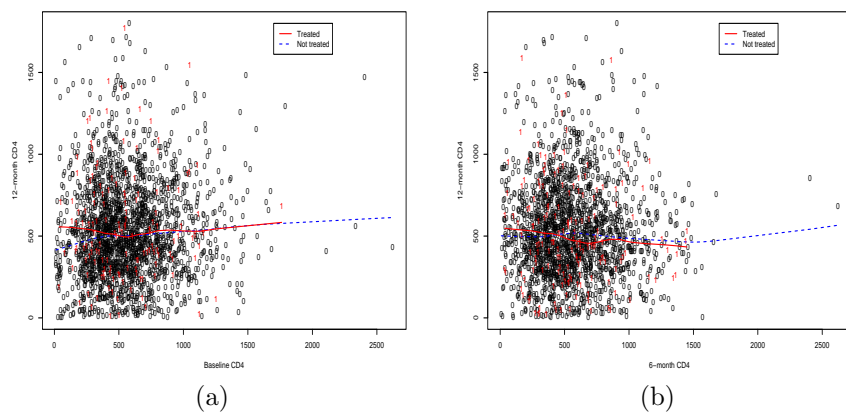
Figure 3. MACS: (a) CD4 at 12 months vs. baseline and (b) CD4 at 12 vs. six months for those who were not treated in the first interval.

Table 1

*AZT initiation and CD4 cell counts: g-estimation and IMOR for 1000
data-sets of sample sizes 500 and 1000.*

Estimate	ψ	Correct model for $p_j(a\|\mathfrak{h}_j;\hat{\alpha})$				Incorrect model for $p_j(a\|\mathfrak{h}_j;\hat{\alpha})$			
		$\hat{\psi}$	SE	rMSE	Cov.*	$\hat{\psi}$	SE	rMSE	Cov.*
						$n = 500$			
g-est.	$\psi_{10} = 250$	225.76	304.96	407.10	96.5	2782.53	478.10	2577.95	0.0
eqn. (1)	$\psi_{11} = -1.0$	-0.967	0.735	0.984	96.2	-8.648	1.398	7.776	0.0
	$\psi_{20} = 720$	744.87	406.15	549.3	95.1	3172.1	799.21	2584.39	4.9
	$\psi_{21} = -2.0$	-2.060	0.773	1.046	94.9	-7.085	1.671	5.363	4.0
g-est.†	$\psi_{10} = 250$	247.03	24.29	32.78	94.9	197.49	21.05	58.23	34.8
eqn. (2)	$\psi_{11} = -1.0$	-0.995	0.052	0.071	94.5	-0.870	0.055	0.144	35.8
	$\psi_{20} = 720$	721.34	82.35	114.34	92.4	563.92	79.63	183.48	50.6
	$\psi_{21} = -2.0$	-2.003	0.131	0.183	92.4	-1.724	0.141	0.321	51.4
g-est.‡	$\psi_{10} = 250$	250.01	17.17	23.18	95.1	250.81	17.02	25.17	89.4
eqn. (2)	$\psi_{11} = -1.0$	-1.000	0.038	0.051	95.2	-1.002	0.048	0.064	95.4
	$\psi_{20} = 720$	720.30	24.05	33.56	92.6	719.18	28.52	41.23	89.3
	$\psi_{21} = -2.0$	-2.001	0.041	0.056	93.0	-1.999	0.054	0.076	92.2
IMOR	$\psi_{10} = 250$	242.63	98.11	123.7	98.5	-38.79	122.18	316.73	32.2
	$\psi_{11} = -1.0$	-0.986	0.213	0.265	98.8	-0.720	0.261	0.416	86.5
	$\psi_{20} = 720$	716.61	142.34	187.35	96.4	479.47	193.99	345.69	73.7
	$\psi_{21} = -2.0$	-1.995	0.223	0.295	95.9	-1.797	0.328	0.485	88.3
						$n = 1000$			
g-est.	$\psi_{10} = 250$	237.09	211.23	288.04	94.3	2770.40	335.93	2542.86	0.0
eqn. (1)	$\psi_{11} = -1.0$	-0.980	0.511	0.697	94.1	-8.587	0.981	7.651	0.0
	$\psi_{20} = 720$	720.04	284.26	380.78	95.2	3051.94	558.22	2399.22	0.0
	$\psi_{21} = -2.0$	-2.006	0.540	0.724	94.8	-6.824	1.163	4.965	0.0
g-est.†	$\psi_{10} = 250$	247.74	16.85	23.10	93.6	198.08	14.96	54.46	12.9
eqn. (2)	$\psi_{11} = -1.0$	-0.996	0.036	0.050	93.9	-0.871	0.039	0.135	10.1
	$\psi_{20} = 720$	720.82	60.12	82.74	93.0	562.77	59.65	170.63	25.8
	$\psi_{21} = -2.0$	-2.002	0.096	0.132	93.1	-1.723	0.106	0.300	24.0
g-est.‡	$\psi_{10} = 250$	249.45	12.16	16.68	94.9	250.00	12.04	17.71	89.3
eqn. (2)	$\psi_{11} = -1.0$	-0.999	0.027	0.037	94.2	-1.000	0.034	0.046	95.3
	$\psi_{20} = 720$	720.29	17.22	23.73	93.5	720.28	20.30	29.28	90.0
	$\psi_{21} = -2.0$	-2.001	0.029	0.040	94.3	-2.001	0.038	0.054	92.8
IMOR	$\psi_{10} = 250$	245.16	69.46	86.78	98.4	-35.38	87.44	299.31	4.8
	$\psi_{11} = -1.0$	-0.991	0.150	0.186	98.8	-0.727	0.186	0.345	71.3
	$\psi_{20} = 720$	720.57	101.76	134.35	95.7	464.94	139.89	303.99	53.0
	$\psi_{21} = -2.0$	-2.001	0.159	0.211	95.9	-1.770	0.236	0.379	83.2

* Coverage of 95% Wald-type confidence intervals

† $E[H_{\mathrm{mod},j}(\psi)|\mathfrak{h}_j]$ linear in $\mathfrak{h}_j$ (incorrect model)

‡ $E[H_{\mathrm{mod},j}(\psi)|\mathfrak{h}_j]$ piece-wise linear (correct model)

Table 2

AZT initiation and its effects on 12-month CD4 cell counts in the Multicenter AIDS Cohort Study where (a) the treatment model depends only on prior CD4 and (b) a richer treatment model is assumed. For details of the model for $E[H_{mod,j}(\psi)|\mathfrak{h}_j]$, see section 4.2.

	ψ	g-estimate eqn (2) $\hat{\psi}$	95% CI	IMOR $\hat{\psi}$	95% CI
(a)	ψ_{10}	-16.61	(-64.37, 31.16)	-103.79	(-308.43, 100.85)
	ψ_{11}	-0.019	(-0.152, 0.114)	0.177	(-0.457, 0.811)
	ψ_{20}	-39.32	(-85.43, 6.79)	-116.76	(-294.80, 61.27)
	ψ_{21}	-0.063	(-0.192, 0.067)	0.134	(-0.313, 0.581)
(b)	ψ_{10}	1.40	(-53.10, 55.90)	-129.43	(-316.44, 57.57)
	ψ_{11}	-0.046	(-0.183, 0.092)	0.182	(-0.340, 0.705)
	ψ_{20}	-14.44	(-62.30, 33.42)	197.65	(-2635.69, 3031.00)
	ψ_{21}	-0.105	(-0.236, 0.026)	-0.442	(-3.103, 2.219)