Biometrics 63, 447–455
June 2007

DOI: 10.1111/j.1541-0420.2006.00686.x

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, Washington 98195, U.S.A.
Current address: Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, Québec H3A 1A2, Canada
[2]Department of Statistics, University of Washington, Seattle, Washington 98195, U.S.A.
[3]Department of Mathematics, Imperial College London, London SW7 2AZ, U.K.
[*]*email:* erica.moodie@mcgill.ca

Summary. A dynamic regime is a function that takes treatment and covariate history and baseline covariates as inputs and returns a decision to be made. Murphy (2003, *Journal of the Royal Statistical Society, Series B* **65,** 331–366) and Robins (2004, *Proceedings of the Second Seattle Symposium on Biostatistics*, 189–326) have proposed models and developed semiparametric 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's and that the methods are closely related but not equivalent. Interesting features of the methods are highlighted using the Multicenter AIDS Cohort Study and through simulation.

Key words: Optimal dynamic regimes; Optimal structural nested mean models; Randomized controlled trials; Sequential randomization; Treatment algorithms.

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 treatment and covariate history as arguments and outputs an action to be taken, providing a list of decision rules for how treatment should be allocated over time. A subject's interval-specific treatment is not known at the start of a dynamic 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, e.g., 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 (Lavori and Dawson, 2001; Dawson and Lavori, 2004; Murphy, 2004). Design considerations for multicenter, sequentially randomized trials with adaptive randomization have been addressed in a Bayesian framework (Thall, Millikan, and Sung, 2000; Thall and Wathen, 2005). Alongside the theoretical innovations, within-person sequentially randomized trials are being performed for treatment of mental illness (Schneider et al., 2001; Rush et al., 2003) 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). In the dynamic regimes context, it requires modeling the longitudinal distribution of all covariates and outcome. The knowledge needed to model this is often unavailable and, by misspecifying the distribution, treatment may be incorrectly recommended when no treatment effect exists. Murphy's (2003) and Robins's (2004) methods do not suffer from this serious limitation.

Lavori et al. (1994) assessed the optimal treatment discontinuation time via a causal approach 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 for prostate cancer treatment, where randomization probabilities changed as information from patients accrued.

The purpose of this article is to provide a clearer understanding of the models and methods proposed in the optimal dynamic regime literature—Murphy's iterative minimization and Robins's g-estimation—and to demonstrate the similarities between what may appear to be very different

approaches. The following section introduces our motivating example: the effect of AZT initiation on 12-month CD4 counts. In Section 3, the procedure for estimating an optimal 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 approaches. The methods are demonstrated using the Multicenter AIDS Cohort Study (MACS) in Section 4, alongside simulations that highlight interesting features of the 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 from 1984. Participants were invited to return for a follow-up questionnaire and physical examination every 6 months. 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 1 year and 10 (0.5%) had initiated AZT before study entry; these were excluded from the analysis. We follow Hernán, Brumback, and Robins (2000) in using last observation carried forward to account for missed follow-up visits.

To minimize notation, we consider only two intervals—baseline to 6 months and 6 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; 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 covariate history up to the beginning of the jth interval not including treatment in interval j, so $\mathcal{H}_1 = X_1$ and $\mathcal{H}_2 = (X_1, A_1, X_2)$. Specific values are denoted in lowercase, e.g., $\mathfrak{h}_1 = x_1$. $D_j(\mathcal{H}_j)$ denotes treatment at t_j that depends on history.

In our example, X_1, X_2, and Y are CD4 cell counts at, respectively, baseline, 6 months, and 12 months. Treatment is the indicator of AZT commencement so that $A_1 = 1$ if AZT therapy was initiated between baseline and 6 months and $A_2 = 1$ the equivalent for AZT between 6 and 12 months. Rules to be estimated are for starting of AZT (Figure 1b).

Throughout this article, models use *potential outcomes*, the value of a status or final response variable that would result if a person were assigned to different treatments. Let $X_2(a_1)$ denote a person's potential covariate status at the beginning of the second interval if treatment a_1 is received by that person and $Y(a_1, a_2)$ denote the potential end-of-study outcome if he follows regime (a_1, a_2).

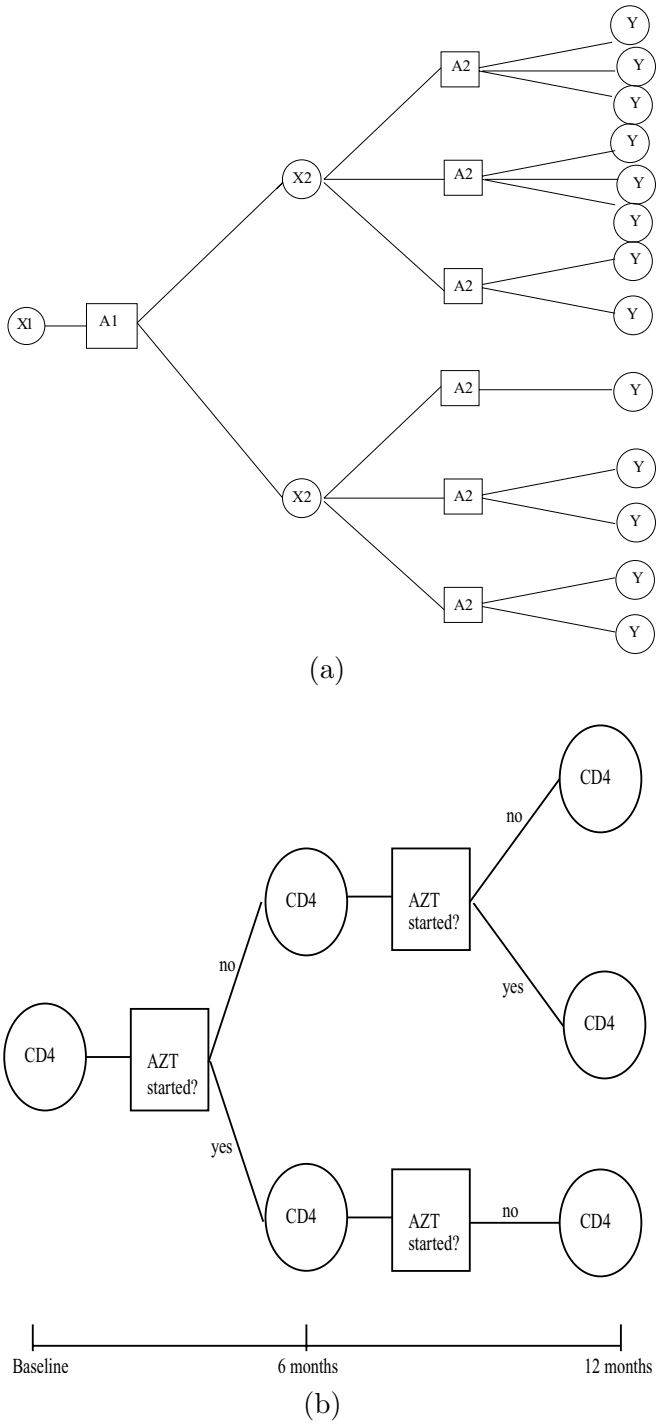

Figure 1. Illustration of data for two intervals: (a) generic and (b) MACS.

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

2.2 *Assumptions*

To estimate the effect of a dynamic regime (optimal or not), we require the following.

(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 fixed 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 $p_1(d_1(\mathfrak{h}_1) \,|\, \mathfrak{h}_1) \times p_2(d_2(\mathfrak{h}_2) \,|\, \mathfrak{h}_2) > 0$. That is, 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 nonparametrically. In terms of a decision tree, no (nonparametric) 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. A rich model was used for the probability of initiating AZT, so it is plausible that there are few or no other variables that confound the association between CD4 and AZT initiation.

3. Steps to Finding the Optimal Regime

We define optimal rules recursively: $d_2^{\text{opt}}(\mathfrak{h}_2) = \max_{d_2} E[Y(a_1, d_2(\mathfrak{h}_2)) \,|\, \mathcal{H}_2 = \mathfrak{h}_2]$, and $d_1^{\text{opt}}(\mathfrak{h}_1) = \max_{d_1} E[Y(d_1(\mathfrak{h}_1), d_2^{\text{opt}}(\mathfrak{h}_1, d_1(\mathfrak{h}_1), X_2(\mathfrak{h}_1, d_1(\mathfrak{h}_1)))) \,|\, \mathcal{H}_1 = \mathfrak{h}_1]$. 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 suboptimal regime had been followed up to that point.

Robins (1986, 1994, 1997) pioneered the field of dynamic treatment regimes. However, Murphy (2003) gave the first method to estimate optimal regimes semiparametrically. Following this, Robins (2004, p. 209–214) 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 (i) definition of the model, (ii) finding the optimal rule implied by the model, and (iii) estimation of model parameters. For (i), Robins uses "blip" models; Murphy uses "regrets." Both are variants of the SNMMs. For (iii), Robins uses g-estimation, while Murphy uses iterative minimization.

3.1 *Step* 1: *Model Definition*

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

Define an *optimal blip-to-reference function* to be the expected difference in outcome when using a reference regime $d_j^{\text{ref}} = d_j^{\text{ref}}(\mathfrak{h}_j)$ instead of a_j at t_j, in persons with treatment and covariate history $\mathfrak{h}_j$ who subsequently receive the optimal regime. At the first interval, $\gamma_1^{d_j^{\text{ref}}}(\mathfrak{h}_1, a_1) = E[Y(a_1, d_2^{\text{opt}}(\mathfrak{h}_1, a_1, X_2(a_1))) - Y(d_1^{\text{ref}}, d_2^{\text{opt}}(\mathfrak{h}_1, d_1^{\text{ref}}, X_2(d_1^{\text{ref}}))) \,|\, \mathcal{H}_1 = x_1]$, and at the second, $\gamma_2^{d_j^{\text{ref}}}(\mathfrak{h}_2, a_2) = E[Y(a_1, a_2) - Y(a_1, d_2^{\text{ref}}(\mathfrak{h}_2)) \,|\, \mathcal{H}_2 = (x_1, a_1, x_2)]$. The term "optimal" refers to treatment subsequent to t_j; what is optimal subsequent to t_j may depend on the treatment received at 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 d_2^{ref} among people with 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, p. 217), takes the reference regime to be the "zero" regime at time j, a substantively meaningful regime such as placebo or standard care. Denote this by $\gamma_j(\mathfrak{h}_j, a_j)$.

Murphy (2003) modeled the *regret* function, which is the *negative* of the optimal blip where the reference regime is the optimal treatment at t_j: $\mu_1(\mathfrak{h}_1, a_1) = E[Y(a_1, d_2^{\text{opt}}(\mathfrak{h}_1, a_1, X_2(a_1))) - Y(d_1^{\text{opt}}(\mathfrak{h}_1), d_2^{\text{opt}}(\mathfrak{h}_1, d_1^{\text{opt}}, X_2(d_1^{\text{opt}}))) \,|\, \mathcal{H}_1 = x_1]$, and $\mu_2(\mathfrak{h}_2, a_2) = E[Y(a_1, a_2) - Y(a_1, d_2^{\text{opt}}(\mathfrak{h}_2)) \,|\, \mathcal{H}_2 = (x_1, a_1, x_2)]$. 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} onward.

Optimal blip functions and regrets are mathematically equivalent. For binary treatment and continuous outcome, the correspondence is $\mu_j(\mathfrak{h}_j, a_j) = \max_a \gamma_j^{d_j^{\text{ref}}}(\mathfrak{h}_j, a) - \gamma_j^{d_j^{\text{ref}}}(\mathfrak{h}_j, a_j)$ and $\gamma_j^{d_j^{\text{ref}}}(\mathfrak{h}_j, a_j) = \mu_j(\mathfrak{h}_j, d_j^{\text{ref}}) - \mu_j(\mathfrak{h}_j, a_j)$. It is apparent from the equations that a regret that is smooth in its arguments (or parameters) implies a smooth optimal blip; the converse does not hold (see Robins, 2004, Section 6.1 for discussion of this correspondence). Optimal blips and regrets compare potential outcomes in which treatment at t_{j+1} and thereafter is optimal; regrets additionally posit that treatment at t_j is optimal. Henceforth, take "optimal blips" to mean optimal blip-to-zero ($d_j^{\text{ref}} = 0$) functions.

While Robins advocates optimal blip-to-zero functions and Murphy regrets, they are equivalent. However, it is important to be aware that simple forms for either model can lead to complex (perhaps unlikely) forms for observables. For example, if $Y(d_1^{\text{opt}}, d_2^{\text{opt}})$ depends linearly on status variables, X_j, and we assume a linear blip or (equivalently) a piecewise linear regret, this implies that the observed outcome, $Y(a_1, a_2)$, is piecewise linear in X_j and not necessarily continuous.

3.2 *Step* 2: *Identification of the Optimal Rules*

Given the *true* optimal blip or regret parameterized by ψ, it is straightforward to identify the optimal regime $d_j^{\text{opt}}(\mathfrak{h}_j, a_j; \psi) = \arg\max_{a_j} \gamma_j(\mathfrak{h}_j, a_j; \psi)$ for all j, or the regime $d_j^{\text{opt}}(\mathfrak{h}_j, a_j; \psi)$ such that $\mu_j(\mathfrak{h}_j, d_j^{\text{opt}}(\mathfrak{h}_j, a_j; \psi)) = 0$.

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: $\mathcal{D}_j(\gamma) = \{d_j(\cdot) \,|\, d_j(\mathfrak{h}_j) = \arg\max_{a_j} \gamma_j(\mathfrak{h}_j, a_j; \psi)$ for some $\psi\}$. $\mathcal{D}_j(\mu)$ is the set of optimal rules that are compatible with regret $\mu_j(\mathfrak{h}_j, a_j; \psi)$: $\mathcal{D}_j(\mu) = \{d_j(\cdot) \,|\, \mu_j(\mathfrak{h}_j, d_j(\mathfrak{h}_j); \psi) = 0$ for some $\psi\}$. $\mathcal{D}_j(\gamma) = \mathcal{D}_j(\mu)$ when the blip and regret are equivalent.

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 $\mathcal{D}_j(\tilde{\mu}) = \{d_j(\cdot) | d_j(\mathfrak{h}_j) = \arg\min_{a_j} \tilde{\mu}_j(\mathfrak{h}_j, a_j; \psi)$ 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 using either method of estimation if parameterization of the true SNMM is poor. For example, suppose the true regret is $\mu_j(\mathfrak{h}_j, a_j) = |\psi_0 + \psi_1 x_j| \times (a_j - I[\psi_0 + \psi_1 x_j > 0])^2$ with treatment a_j binary and $\mathcal{D}_j(\mu) = \{I[\psi_0 + \psi_1 x_j > 0]\}$. Further suppose $\psi_1 > 0$ so that treatment is beneficial if X_j is above the threshold $\beta = -\psi_0/\psi_1$ (note that the threshold given by Robins, 2004, p. 245, is incorrect; numerator and denominator are transposed). We may reparameterize the regret to obtain the threshold, β, by $\mu_j^*(\mathfrak{h}_j, a_j) = |x_j - \beta| \times (a_j - I[x_j - \beta > 0])^2$, which gives $\mathcal{D}_j(\mu^*) = \{I[x_j - \beta > 0]\}$. However, if $\psi_1 < 0$ so that now subjects should be treated when $x_j < \beta, \mu_j^*(\mathfrak{h}_j, a_j) = |x_j - \beta| \times (a_j - I[x_j - \beta < 0])^2$, then $\mathcal{D}_j(\mu^*) = \{I[x_j - \beta < 0]\}$. Thus, using the reparameterized regret (or the corresponding blip) requires knowing in advance whether it is optimal to treat for high- or low-status values. 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, such as the two-parameter regret in this example. (See reply to discussion in Murphy, 2003.)

3.3 *Step* 3: *Estimation*

3.3.1 *g-estimation.* Robins (2004, p. 208) proposes finding the parameters ψ of the optimal blip-to-zero function or regret function via *g*-estimation. Define $H_1(\psi) = Y + \sum_{j=1}^{2}[\gamma_j(\mathfrak{h}_j, d_j^{\text{opt}}; \psi) - \gamma_j(\mathfrak{h}_j, a_j; \psi)]$, $H_2(\psi) = Y + \gamma_2(\mathfrak{h}_2, d_2^{\text{opt}}; \psi) - \gamma_2(\mathfrak{h}_2, a_2; \psi)$. $H_j(\psi)$ 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, $H_j(\psi)$ corresponds to a counterfactual outcome: $H_1(\psi) = Y(d_1^{\text{opt}}(x_1), d_2^{\text{opt}}(x_1, d_1^{\text{opt}}(x_1), X_2(d_1^{\text{opt}}(x_1))))$ and $H_2(\psi) = Y(a_1, d_2^{\text{opt}}(x_1, a_1, x_2))$ (Robins, 2004, p. 204). Loosely, local rank preservation states that the ranking of patients' counterfactual outcomes under a particular regime is the same as their ranking under the observed regime, conditional on history. Local rank preservation is additive if the difference in a person's counterfactual outcome should he be treated with one regime instead of another given history equals the expected difference. Note that this is a *counterfactual* difference since it is not directly observable, as it is a difference between two potential outcomes for an individual. Rank preservation provides a simplistic situation in which the parameters of an SNMM may be interpreted at the individual level. SNMMs may be used without this assumption 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 $S_2(a_2) = \frac{\partial}{\partial \psi}\gamma_2(\mathfrak{h}_2, a_2) = a_2(1, x_2, a_1, x_2 a_1)^T$. Let

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

with the probability of being treated modeled (perhaps nonparametrically) by $p_j(a_j \,|\, \mathfrak{h}_j; \alpha)$. $E[U(\psi, s)] = 0$ is an unbiased estimating equation from which consistent estimates $\hat{\psi}$ of ψ may be found. The estimates are asymptotically Normal under standard regularity conditions provided the treatment model is correct and the optimal regime is unique (though see Robins, 2004, p. 219, Appendix 1, for discussion of when this fails to hold). Equation (1) is unbiased since potential outcomes under different treatment regimes at t_j and hence $H_j(\psi)$ are independent of any function of actual treatment conditional on past treatment and covariates (Assumption 2). The estimators are not efficient.

Robins (2004, p. 212) refined equation (1) to gain efficiency. Let

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

The inclusion of $E[H_j(\psi) \,|\, \mathfrak{h}_j]$ in (2) gives estimates which are more efficient than those found using (1) even if its model is misspecified (Robins et al., 1995). 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 modeled, and thus are said to be *doubly robust*. These estimates are still not efficient; semiparametric 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. 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 piecewise 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 piecewise linear with discontinuities and changes in slope occurring at optimal rule thresholds.

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 stationary (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 $\gamma_2(\mathfrak{h}_2, a_2) = a_2(\psi_0 + \psi_1 x_2 + \psi_2 a_1)$, since ψ_0 and ψ_1 appear in the blip functions of both intervals.

Modifications $H_{\text{mod},1}(\psi) = Y - \gamma_1(\mathfrak{h}_1, a_1; \psi) + [\gamma_2(\mathfrak{h}_2, d_2^{\text{opt}}; \psi) - \gamma_2(\mathfrak{h}_2, a_2; \psi)]$ and $H_{\text{mod},2}(\psi) = Y - \gamma_2(\mathfrak{h}_2, a_2; \psi)$ can be used in (1) or (2) without changing the consistency of the resulting *g*-estimates. Under additive local rank preservation, we have the following interpretation: $H_{\text{mod},1}(\psi) = Y(0, d_2^{\text{opt}}(x_1, 0, X_2(0))), H_{\text{mod},2}(\psi) = Y(a_1, 0)$. 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 $H_{\text{mod},1}(\psi)$ to find $\hat{\psi}_1$ (Robins, 2004, p. 219). Postulating models for Y (conditional on $\mathfrak{h}_j$) and $\gamma_j(\mathfrak{h}_j; \psi)$ is sufficient to determine the form of $E[H_{\text{mod},j}(\psi) \mid \mathfrak{h}_j]$.

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 for all $(\psi^*,\ c^*)$

$$\sum_{j=1}^{2} \mathbb{P}_n \left[Y + \hat{c} + \sum_{l=1}^{2} \mu_l(\mathfrak{h}_l, a_l; \hat{\psi}) - \sum_{a} \mu_j(\mathfrak{h}_j, a; \hat{\psi}) p_j(a \mid \mathfrak{h}_j; \hat{\alpha}) \right]^2$$
$$\leq \sum_{j=1}^{2} \mathbb{P}_n \left[Y + c^* + \sum_{l=1, l\neq j}^{2} \mu_l(\mathfrak{h}_l, a_l; \hat{\psi}) + \mu_j(\mathfrak{h}_j, a_j; \psi^*) \right.$$
$$\left. - \sum_{a} \mu_j(\mathfrak{h}_j, a; \psi^*) p_j(a \mid \mathfrak{h}_j; \hat{\alpha}) \right]^2, \tag{3}$$

where $\mathbb{P}_n(f) = n^{-1} \sum_{i=1}^{n} f(X_i)$ is the empirical average. The scalar c improves stability of the minimization but is not required. IMOR estimates are consistent for ψ provided the treatment allocation models, $p_j(a_j \mid \mathfrak{h}_j)$, are correctly specified. See the next section for a discussion of efficiency.

Murphy (2003) described an iterative method for finding solutions to (3): select an initial value of $\hat{\psi}$, say $\hat{\psi}^{(1)}$, and then minimize the right-hand side (RHS) of the equation over $(\psi^*,\ c^*)$ to obtain a new value of $\hat{\psi}, \hat{\psi}^{(2)}$, and repeat this until convergence. This may not produce a monotonically decreasing sequence of RHS values of equation (3) and may not converge to a minimum; profile plots of the RHS of (3) for each parameter about its estimate provide a useful diagnostic tool.

3.4 *Relating the Methods for Two 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

$$E\left[\left\{Y - q\big(\mathfrak{h}_j, a_j\big) + \sum_{m=j+1}^{2} \big(\gamma_m\big(\mathfrak{h}_m, d_m^{\text{opt}}; \psi_m\big) - \gamma_m\big(\mathfrak{h}_m, a_m; \psi_m\big)\big) \right.\right.$$
$$- E\left[Y - q\big(\mathfrak{h}_j, a_j\big) + \sum_{m=j+1}^{2} \right.$$
$$\left.\left.\left. \times \big(\gamma_m\big(\mathfrak{h}_m, d_m^{\text{opt}}; \psi_m\big) - \gamma_m\big(\mathfrak{h}_m, a_m; \psi_m\big)\big) \;\middle|\; \mathfrak{h}_m \right] \right\}^2 \right] \tag{4}$$

subject to $q(\mathfrak{h}_j, 0) = 0$ is $\gamma_j(\mathfrak{h}_j, a_j; \psi_j)$. To use (4) to estimate $\psi_1, \hat{\psi}_2$ 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

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

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 $q(\mathfrak{h}_j, a_j) = \gamma_j(\mathfrak{h}_j, a_j; \psi_j) = \mu_j(\mathfrak{h}_j, 0; \psi_j) - \mu_j(\mathfrak{h}_j, a_j; \psi_j)$ in (4) leads to the RHS of (3) with $-c = E[H_{\text{mod},1}(\psi) \mid \mathfrak{h}_1] + \mu_1(\mathfrak{h}_1, 0; \psi_1) - E[\mu_1(\mathfrak{h}_1, a_1; \psi_1) \mid \mathfrak{h}_1]$ in the first interval and $-c = E[H_{\text{mod},2}(\psi) \mid \mathfrak{h}_2] + \mu_2(\mathfrak{h}_2, 0; \psi_2) + \mu_1(\mathfrak{h}_1, a_1; \hat{\psi}_1) - E[\mu_2(\mathfrak{h}_2, a_2; \psi_2) \mid \mathfrak{h}_2]$ in the second. The parameter c in (3) is *not* interval specific, so the methods are not identical. This is a critical difference: IMOR does not model $E[H_{\text{mod},j}(\psi) \mid \mathfrak{h}_j]$ explicitly, but rather captures the quantity through regrets and c, which does not vary with covariates in $\mathfrak{h}_j$. It is clear that, *under the null hypothesis* of no treatment effect, $c = E[H_{\text{mod},j}(\psi)] = E[Y]$, thus IMOR and *g*-estimation (2) modeling $E[H_{\text{mod},j}(\psi) \mid \mathfrak{h}_j]$ with a constant are equivalent.

Regarding relative efficiency of the methods, we make the following points:

(i) IMOR is a special case of *g*-estimating equation (2) under the null hypothesis of no treatment effect, modeling $E[H_{\text{mod},j}(\psi) \mid \mathfrak{h}_j]$ by a constant.

(ii) Under regularity conditions, estimates from equation (2) are the most efficient among the class of *g*-estimates that use a given function $S(a_j)$ (here, $-\frac{\partial}{\partial \psi_j} q(\mathfrak{h}_j, a_j)$) when the treatment and expected counterfactual models are correctly specified (Robins, 2004, Theorems 3.3(ii) and 3.4).

(iii) In general, equation (2) does not satisfy regularity conditions due to nondifferentiability of the estimating equation in a neighborhood of $\psi = 0$. However, the conditions hold for constant blip functions, $\gamma_j(a_j) = a_j \psi_j$, which posit no treatment interactions. (See E. E. M. Moodie and T. S. Richardson, unpublished manuscript and Robins, 2004, p. 225 for further discussion.)

Table 1
AZT initiation and CD4 *cell counts: g-estimation and IMOR for* 1000 *data sets of sample sizes* 500 *and* 1000

		Correct model for $p_j(a \mid \mathfrak{h}_j; \hat{\alpha})$				Incorrect model for $p_j(a \mid \mathfrak{h}_j; \hat{\alpha})$			
Estimate	ψ	$\hat{\psi}$	SE	rMSE	Cov.*	$\hat{\psi}$	SE	rMSE	Cov.*
				$n = 500$					
g-estimation	$\psi_{10} = 250$	225.76	304.96	407.10	96.5	2782.53	478.10	2577.95	0.0
equation (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-estimation†	$\psi_{10} = 250$	247.03	24.29	32.78	94.9	197.49	21.05	58.23	34.8
equation (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-estimation‡	$\psi_{10} = 250$	250.01	17.17	23.18	95.1	250.81	17.02	25.17	89.4
equation (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-estimation	$\psi_{10} = 250$	237.09	211.23	288.04	94.3	2770.40	335.93	2542.86	0.0
equation (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-estimation†	$\psi_{10} = 250$	247.74	16.85	23.10	93.6	198.08	14.96	54.46	12.9
equation (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-estimation‡	$\psi_{10} = 250$	249.45	12.16	16.68	94.9	250.00	12.04	17.71	89.3
equation (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.
†$E[H_{\text{mod},j}(\psi) \mid \mathfrak{h}_j]$ linear in $\mathfrak{h}_j$ (incorrect model).
‡$E[H_{\text{mod},j}(\psi) \mid \mathfrak{h}_j]$ piecewise linear (correct model).

In conclusion, we may say that if the null hypothesis holds and we estimate a constant blip model (which trivially is correctly specified), then g-estimation is more efficient than IMOR when $E[H_{\text{mod},j}(\psi) \mid \mathfrak{h}_j] = E[Y \mid \mathfrak{h}_j]$ depends on $\mathfrak{h}_j$ and is correctly specified in (2). If $E[H_{\text{mod},j}(\psi) \mid \mathfrak{h}_j]$ is constant, IMOR and g-estimation (2) are efficient within the class of g-estimates that use $S(a_j) = -\frac{\partial}{\partial \psi_j} q(\mathfrak{h}_j, a_j)$ when the treatment and expected counterfactual models are correctly specified and $E[H_{\text{mod},j}(\psi) \mid \mathfrak{h}_j]$ is constant. In other cases, the situation is not clear; the violation of regularity conditions is an issue in both approaches.

4. Examples

4.1 *Simulation Results*

Via simulations, we compare the performance of the methods discussed, and illustrate the double-robustness of g-estimating equation (2). Suppose that patients are accrued in a trial so as 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 Section 2.1 and were generated as follows: baseline CD4: $X_1 \sim \mathcal{N}(450, 150)$; 6-month CD4: $X_2 \sim \mathcal{N}(1.25X_1, 60)$; and 1-year CD4: $Y \sim \mathcal{N}(400 + 1.6X_1, 60) - \mu_1(\mathcal{H}_1, A_1) - \mu_2(\mathcal{H}_2, A_2)$. Treatments A_1, A_2 were randomly assigned with equal probability and optimal blips are linear: $\gamma_1(\mathfrak{h}_1, a_1) = a_1(\psi_{10} + \psi_{11}x_1), \gamma_2(\mathfrak{h}_2, a_2) = a_2(\psi_{20} + \psi_{21}x_2)$ with corresponding regrets, $\mu_j(\mathfrak{h}_j, a_j)$. As noted before, the outcome is not linear in X_1 and 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.

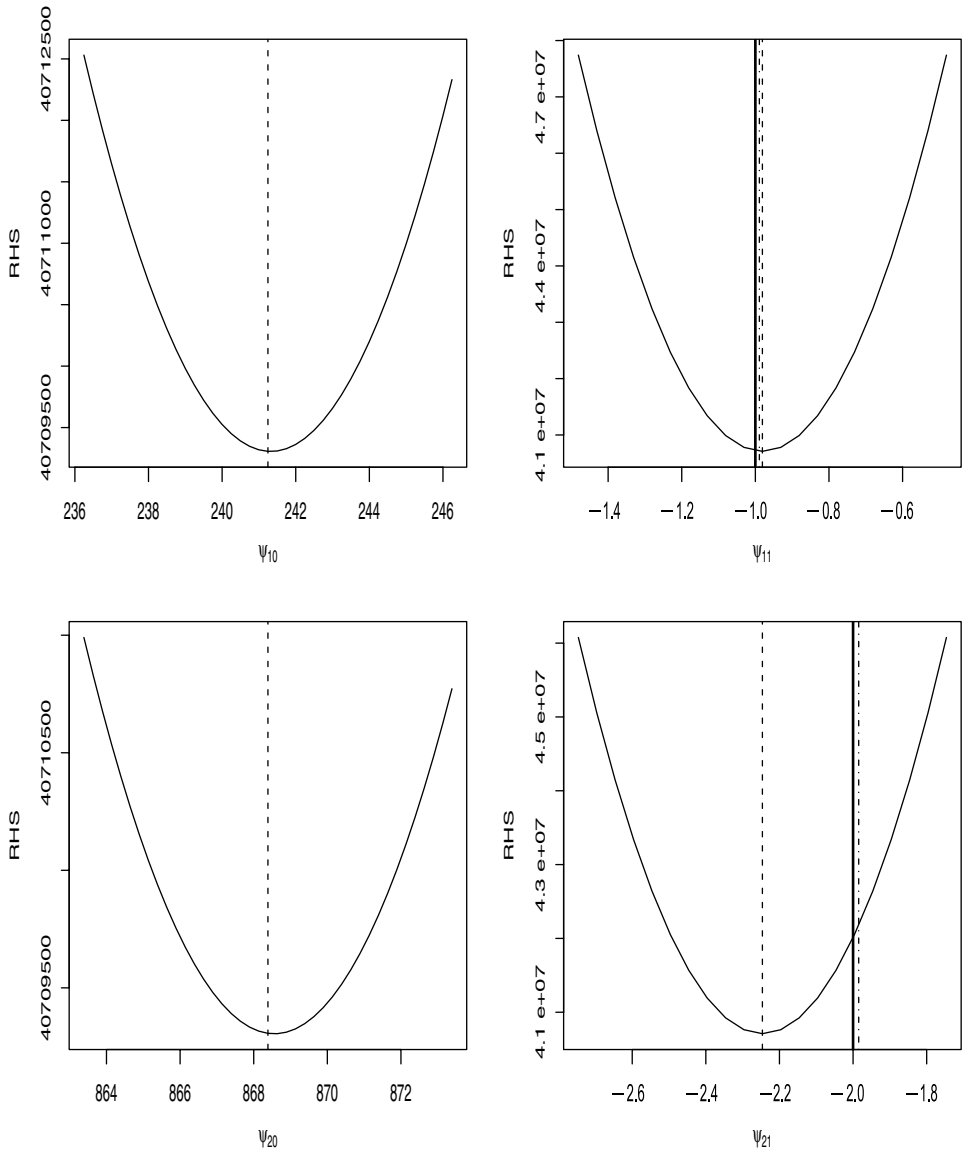

Figure 2. Profiles of the RHS of equation (3) for the IMOR approach from a single simulation. The dashed line is the IMOR estimate, the dashed-and-dotted line from g-estimation (2) using the correct, piecewise model for $E[H_{\text{mod},j}(\psi) \mid \mathfrak{h}_j]$, and the thick black line the truth. For ψ_{10} and ψ_{20}, neither the g-estimate nor the true value occurs within the range plotted; for ψ_{11}, IMOR and the g-estimate nearly coincide.

Two models were considered for $E[H_{\text{mod},j}(\psi) \mid \mathfrak{h}_j]$ for g-estimation with (2). The first incorrectly assumed that $E[H_{\text{mod},j}(\psi) \mid \mathfrak{h}_j]$ depends linearly on all of $\mathfrak{h}_j$. The second, correct model allowed the mean function to be piecewise, discontinuous linear with inflections at the optimal rule thresholds (see Web Appendix). Results are in Table 1; diagnostic plots for IMOR are in Figure 2.

The efficiency gained by using the g-estimating equation (2) instead of (1) is considerable. An incorrect model for $E[H_{\text{mod},j}(\psi) \mid \mathfrak{h}_j]$ reduces efficiency in (2). In these simulations, efficiency of IMOR estimates is less than that of g-estimates from (2), and much better than that from (1).

Suppose now that physicians broke protocol, so that the probability of initiating AZT is higher in patients with low CD4 counts: $A_j \sim \text{Binom}(p_j)$, where $p_1 = \text{expit}(2 - 0.006X_1)$ and $p_2 = \text{expit}(0.8 - 0.004X_2)$. The new randomization scheme depends only on the *observed* variable CD4. If the analyst incorrectly assumed complete randomization, only equation (2) using the correct model for $E[H_{\text{mod},j}(\psi) \mid \mathfrak{h}_j]$ yields unbiased estimates (Table 1). In this example, using a linear model for $E[H_{\text{mod},j}(\psi) \mid \mathfrak{h}_j]$ in (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 thresholds of 250 and 360 counts at the first and second interval, respectively. IMOR and g-estimation (1) are not at all robust to misspecification of the treatment model, as expected.

4.2 *Multicenter AIDS Cohort Study Results*

Turning our attention to the MACS data: 142 (6.7%) participants initiated AZT in the first 6 months of the study; a further 166 (7.8%) began treatment between 6 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).

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 6 months of study. Six-month CD4, use of *Pneumocystis carinii* pneumonia prophylactics in the first 6 months of study, and presence of symptoms at 6 months were predictive of AZT initiation between 6 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 1-year CD4 counts are not sufficient to capture beneficial effects of the therapy or are not a good surrogate for HIV-patient health.

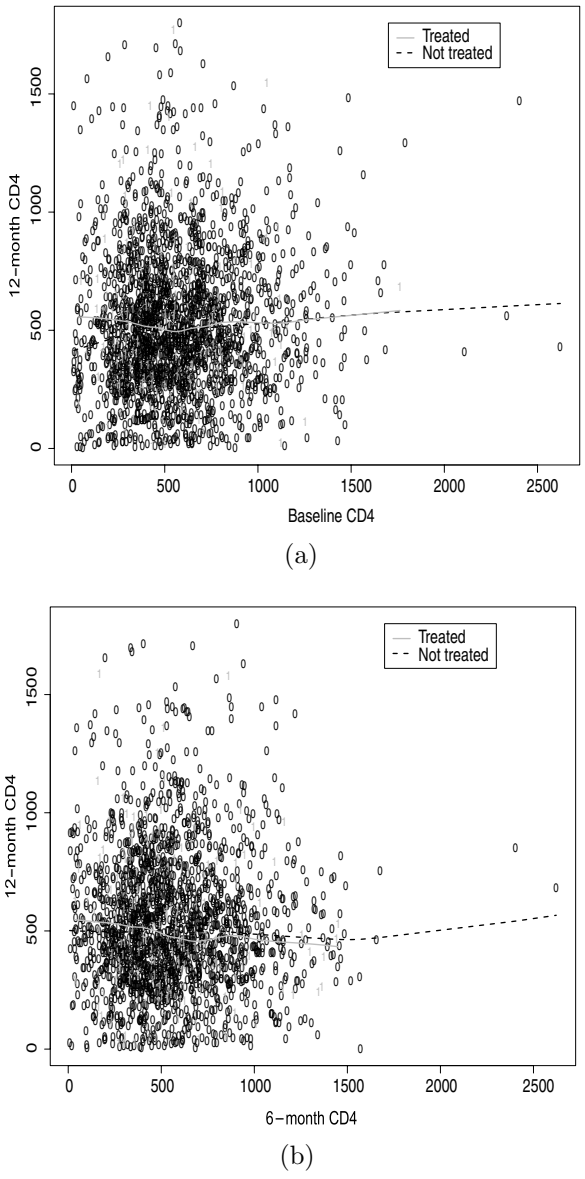

Figure 3. MACS: (a) CD4 at 12 months versus baseline and (b) CD4 at 12 versus 6 months for those who were not treated in the first interval.

A naive linear regression of 12-month CD4 on baseline CD4, 6-month CD4, and treatments A_1 and A_2 picks up a strong association between AZT initiation in the second interval and outcome ($p < 0.001$): participants who started AZT between 6 and 12 months had, on average, 12-month CD4 that was 74 (44, 104) cell counts *lower* than those who did not initiate AZT. A nonsignificantly lower mean CD4 count was also observed for AZT initiation between 0 and 6 months. Residual plots suggested heteroscedasticity; log-transforming outcome did not remove the strong statistical significance of the association, nor did including the covariates from the richer treatment model, nor interaction terms.

The negative association between treatment in the second interval and 1-year CD4 in linear regression can reasonably be explained by confounding: patients with low CD4 counts were more likely to use AZT. This example 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 article has clarified the connections between both the models and the methods used to make inference in the context of dynamic treatment regimes. We have provided formulae for transforming between blips and regrets, and elucidated the similarities between IMOR and g-estimation.

The methods discussed here, along with the advances in theory needed to implement clinical studies of dynamic regimes mentioned in Section 1, have the potential to contribute greatly to the design of treatment protocols for a variety of medical conditions. Murphy and Robins developed methods in the general, K-interval case that are widely applicable although sample size requirements may be infeasible with a large number of treatments unless stationarity is assumed. We conclude with a word of caution: model choice should be driven by practical considerations; however, it is important to be aware of the (perhaps implausible) implied models for observables.

6. Supplementary Materials

The Web Appendix referenced in Section 4.1 and sample code for the simulations are available under the Paper Information link at the *Biometrics* website `http://www.tibs.org/biometrics`.

Acknowledgements

We are grateful to Professor James Robins for enlightening discussion and to anonymous referees for detailed reviews and helpful comments. We also wish to thank Professor Miguel Hernán for advice on data processing. Erica Moodie is supported by a BARDS Fellowship from the Merck Company

Table 2
AZT initiation and its effects on 12*-month* CD4 *cell counts in the MACS where* (*a*) *the treatment model depends only on prior* CD4 *and* (*b*) *a richer treatment model is assumed. For details of the model for* $E[H_{\mathrm{mod},j}(\psi) \mid \mathfrak{h}_j]$, *see Section* 4.2.

		g-estimate equation (2)		IMOR	
	ψ	$\hat{\psi}$	95% CI	$\hat{\psi}$	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)

Foundation. Thomas Richardson acknowledges support by NSF grant DMS 0505865 and NIH R01 AI032475. Data were collected by the Multicenter AIDS Cohort Study (MACS) with centers (Principal Investigators) at The Johns Hopkins University Bloomberg School of Public Health (Joseph B. Margolick, Lisa Jacobson), Howard Brown Health Center and Northwestern University Medical School (John Phair), University of California, Los Angeles (Roger Detels), and University of Pittsburgh (Charles Rinaldo). The MACS is funded by the National Institute of Allergy and Infectious Diseases, with additional supplemental funding from the National Cancer Institute and the National Heart, Lung, and Blood Institute, UO1-AI-35042, 5-MO1-RR-00722 (GCRC), UO1-AI-35043, UO1-AI-37984, UO1-AI-35039, UO1-AI-35040, UO1-AI-37613, UO1-AI-35041.

References

Bellman, R. (1957). *Dynamic Programming*. Princeton, New Jersey: Princeton University Press.

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

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

Hernán, 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. *Epidemiology* **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. *American Journal of Epidemiology* **126,** 310–318.

Lavori, P. W. (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. W. and Dawson, R. (2001). Dynamic treatment regimes: Practical design considerations. *Statistics in Medicine* **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. *Statistics in Medicine* **13,** 1089–1100.

Murphy, S. A. (2003). Optimal dynamic treatment regimes (with discussion). *Journal of the Royal Statistical Society, Series B* **65,** 331–366.

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

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

Robins, J. M. (1997). Causal inference from complex longitudinal data. In *Latent Variable Modeling and Applications to Causality*, M. Berkane (ed), 69–117, New York: Springer-Verlag.

Robins, J. M. (2004). Optimal structural nested models for optimal sequential decisions. In *Proceedings of the Second Seattle Symposium on Biostatistics*, D. Y. Lin and P. Heagerty (eds), 189–326. New York: Springer.

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association* **90,** 106–121.

Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *Annals of Statistics* **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. *Controlled Clinical 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 Antipsychotic Trials of Intervention Effectiveness (CATIE): Alzheimer disease trial methodology. *American Journal of Geriatric Psychology* **9,** 346–360.

Thall, P. F. and Wathen, J. K. (2005). Covariate-adjusted adaptive randomization in a sarcoma trial with multistate treatments. *Statistics in Medicine* **24,** 1947–1964.

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

Received March 2005. *Revised July* 2006.
Accepted July 2006.