Biometrics 68, 1010–1018
December 2012
DOI: 10.1111/j.1541-0420.2012.01763.x

A Robust Method for Estimating Optimal Treatment Regimes

Baqun Zhang,* Anastasios A. Tsiatis, Eric B. Laber and Marie Davidian

Department of Statistics, North Carolina State University, Raleigh, North Carolina 27695-8203, U.S.A.
**email:* bzhang4@ncsu.edu

Summary. A treatment regime is a rule that assigns a treatment, among a set of possible treatments, to a patient as a function of his/her observed characteristics, hence "personalizing" treatment to the patient. The goal is to identify the optimal treatment regime that, if followed by the entire population of patients, would lead to the best outcome on average. Given data from a clinical trial or observational study, for a single treatment decision, the optimal regime can be found by assuming a regression model for the expected outcome conditional on treatment and covariates, where, for a given set of covariates, the optimal treatment is the one that yields the most favorable expected outcome. However, treatment assignment via such a regime is suspect if the regression model is incorrectly specified. Recognizing that, even if misspecified, such a regression model defines a class of regimes, we instead consider finding the optimal regime within such a class by finding the regime that optimizes an estimator of overall population mean outcome. To take into account possible confounding in an observational study and to increase precision, we use a doubly robust augmented inverse probability weighted estimator for this purpose. Simulations and application to data from a breast cancer clinical trial demonstrate the performance of the method.

Key words: Doubly robust estimator; Inverse probability weighting; Outcome regression; Personalized medicine; Potential outcomes; Propensity score.

1. Introduction

The area of personalized medicine, which is focused on making treatment decisions for an individual patient based on his/her clinical, genomic, and other information, is of considerable current interest. In the simplest case of a single treatment decision, there may be several treatment options, and formalizing this objective involves defining a decision rule, or regime, that takes as input an individual's characteristics and dictates the treatment he/she should receive from among the options available. The optimal regime is that leading to the greatest benefit overall in the patient population; i.e., if followed by the entire population, would result in the most favorable clinical outcome on average.

Deducing optimal treatment regimes using data from a clinical trial or observational study can be informed by identifying patient covariates that exhibit a qualitative interaction with treatment assignment in a statistical model for the outcome of interest; i.e., an interaction in which the treatment effect changes direction depending on covariates. For example, Gail and Simon (1985) considered data from a trial conducted by the National Surgical Adjuvant Breast and Bowel Project (NSABP) comparing L-phenylalanine mustard and 5-fluorouracil (PF) to PF plus tamoxifen (PFT) in patients with primary operable breast cancer and positive nodes (Fisher et al., 1983). The study investigators found "evidence for a heterogeneity in response to PFT therapy that is both age and progesterone receptor dependent." Gail and Simon (1985) proposed a test for qualitative interaction based on partitioning the data into subsets using covariate values and concluded on its basis that young patients (age < 50 years) with progesterone receptor levels <10 femtomole/mg cytosol protein (fmol) achieve better outcomes on PF whereas other patients do better on PFT. However, this approach does not target formally the goal of identifying the optimal regime.

More recently, there has been vigorous research on methods for estimating optimal treatment regimes based on data from clinical trials or observational studies, where a single decision or a series of sequential decisions may be involved (Murphy, 2003; Robins, 2004; Moodie, Richardson, and Stephens, 2007; Robins, Orellana, and Rotnitzky, 2008; Brinkley, Tsiatis, and Anstrom, 2009; Zhao, Kosorok, and Zeng, 2009; Henderson, Ansell, and Alshibani, 2010; Orellana, Rotnitzky, and Robins, 2010; Gunter, Zhu, and Murphy, 2011). In the setting of a single-treatment decision, much of this work involves postulating a model for the regression of outcome on treatment assignment and covariates and then assigning treatment for a patient according to which treatment yields the best estimated mean outcome based on the model and given the patient's particular covariate values. However, this approach is clearly predicated on whether or not the assumed model is correctly specified.

In this article, we focus on the case of a single decision and take an alternative view, considering such a posited regression model as a mechanism for defining a class of treatment regimes but recognizing that the model may in fact be misspecified. Assuming without loss of generality that larger outcomes are preferred, we then base estimation of the optimal regime on maximizing directly an estimator for the overall population mean outcome under regimes in the class. Specifically, we maximize across all regimes in the class a suitable doubly robust augmented inverse probability weighted estimator (AIPWE; e.g., Bang and Robins, 2005). This estimator

takes account of possible confounding in the case of data from an observational study via estimated propensity scores and exploits the postulated outcome regression relationship to gain precision. As we demonstrate, this approach leads to estimated optimal regimes that can achieve comparable performance to those based on correctly specified outcome regression models; outperform those based on simpler mean estimators; and, because of the double robustness property, are protected from misspecification of either the propensity score model or the outcome regression model.

In Section 2, we define a framework in which we may formalize the problem. We introduce the proposed methods in Section 3, and we demonstrate their performance in simulation studies in Section 4 and by application to the NSABP data in Section 5.

2. Framework

Consider a clinical trial or observational study with n subjects sampled from the patient population of interest. Suppose, there are two treatment options, e.g., control and experimental treatment in a clinical trial, and let A, taking values 0 or 1 in accordance with the two options, denote observed treatment received. Let X be a vector of subject characteristics ascertained before treatment, and let Y be the observed outcome of interest, where, as in Section 1, we assume larger values of Y are preferred. The observed data are then (Y_i, A_i, X_i), $i = 1, \ldots, n$, independent and identically distributed (iid) across i. The goal is to use these data to estimate the optimal treatment regime, defined as follows.

In this context, a treatment regime is a function g that maps values of X to {0,1}, so that a patient with covariate value $X = x$ would receive treatment 1 if $g(x) = 1$ and treatment 0 if $g(x) = 0$. A simple example for scalar X is $g(X) = I(X < 50)$. To identify formally the optimal treatment regime, we define potential outcomes $Y^*(0)$ and $Y^*(1)$, representing the outcomes that would be observed were a subject to receive treatment 0 or 1, respectively. As is customary (e.g., Rubin, 1978), we assume that $Y = Y^*(1)A + Y^*(0)(1 - A)$, so that the observed outcome is the potential outcome that would be seen under the treatment actually received. We also assume $\{Y^*(0), Y^*(1)\}$ independent of A conditional on X; i.e., that there are no unmeasured confounders. This is trivially true in a randomized clinical trial but is an unverifiable assumption in an observational study (e.g., Robins, Hernán, and Brumback, 2000). Thus, for $a = 0, 1$, $E\{Y^*(a)\}$ represents the overall population mean were all patients in the population to receive treatment a, and, under these assumptions, it is straightforward to deduce that $E\{Y^*(a)\} = E_X[E\{Y|A = a, X\}]$, where the outer expectation $E_X(\cdot)$ is taken with respect to the marginal distribution of X.

Note that, for arbitrary treatment regime g, we can thus define the potential outcome $Y^*(g) = Y^*(1)g(X) + Y^*(0)\{1 - g(X)\}$ that would be observed if a randomly chosen subject from the population were to be assigned treatment according to g, where we suppress the dependence of $Y^*(g)$ on X. If $\mathcal{G}$ is the class of all such treatment regimes, then we may define the optimal regime, g^{opt}, as the one leading to the largest value of $E\{Y^*(g)\}$ among $g \in \mathcal{G}$; i.e., $g^{\text{opt}}(X) = \arg\max_{g\in\mathcal{G}} E\{Y^*(g)\}$. Under the above assumptions, writing $\mu(a, X) = E(Y|A = a, X)$, $a = 0, 1$, it is straightforward to show that

$$E\{Y^*(g)\} = E_X[\mu(1, X)g(X) + \mu(0, X)\{1 - g(X)\}],$$

and hence the optimal treatment regime is given by

$$g^{\text{opt}}(X) = I\{\mu(1, X) > \mu(0, X)\};$$

i.e., the optimal regime assigns the treatment that yields the larger mean outcome conditional on the value of X. Here, the strict inequality follows from the convention that, in the event $\mu(1, X) = \mu(0, X)$ and viewing treatments 0 and 1 as control and experimental, respectively, a conservative strategy would be to prefer the control.

3. Robust Method

To exploit the developments in the previous section, an obvious approach is to posit a regression model for $\mu(A, X) = E(Y|A, X)$, for example, a parametric model $\mu(A, X; \beta)$ for finite-dimensional parameter β, and to estimate β by $\widehat{\beta}$ obtained via some appropriate method; e.g., least squares or generalized least squares. Assuming the model is correctly specified, so that $\mu(A, X) = \mu(A, X; \beta_0)$ for some β_0, the optimal regime is then $g(X, \beta_0)$, where $g(X, \beta) = I\{\mu(1, X, \beta) > \mu(0, X, \beta)\}$, and it is natural to estimate the optimal treatment regime by $\widehat{g}^{\text{opt}}_{\text{reg}}(X) = g(X, \widehat{\beta})$, which we denote as the regression estimator. An obvious estimator for the overall mean outcome under the optimal regime, $E\{Y^*(g^{\text{opt}})\}$, is then

$$n^{-1}\sum_{i=1}^{n}[\,\mu(1, X_i, \widehat{\beta})\widehat{g}^{\text{opt}}_{\text{reg}}(X_i) + \mu(0, X_i, \widehat{\beta})\big\{1 - \widehat{g}^{\text{opt}}_{\text{reg}}(X_i)\big\}\,]. \quad (1)$$

Of course, whether or not $\widehat{g}^{\text{opt}}_{\text{reg}}$ is a credible estimator for the true optimal regime g^{opt} depends critically on whether or not the model $\mu(A, X; \beta)$ is correct. If it is not, then treatment assignment on its basis will not lead to $E\{Y^*(g^{\text{opt}})\}$, nor will (1) be an appropriate estimator for this maximum achievable mean outcome.

A posited model $\mu(A, X; \beta)$, whether correct or not, may be viewed as defining the class of treatment regimes indexed by β, $\mathcal{G}_\beta$, say, with elements of the form $g(X, \beta)$. In fact, in many instances, only a subset of elements of X and β may define the regime, and the class may be simplified. For example, if $\mu(A, X; \beta) = \exp\{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + A(\beta_3 + \beta_4 X_1 + \beta_5 X_2)\}$, it is straightforward to show that elements in the class are of the form $I(\beta_3 + \beta_4 X_1 + \beta_5 X_2 > 0)$, which may be rewritten in terms of $\eta_0 = -\beta_3/\beta_5$ and $\eta_1 = -\beta_4/\beta_5$ as either $I(X_2 > \eta_0 + \eta_1 X_1)$ or $I(X_2 < \eta_0 + \eta_1 X_1)$ depending on the sign of β_5. This suggests considering directly regimes of the form $g_\eta(X) = g(X, \eta)$ in a class $\mathcal{G}_\eta$, say, indexed by a parameter η. In the event, the regimes in $\mathcal{G}_\eta$ are derived from a regression model $\mu(A, X; \beta)$, as above, $\eta = \eta(\beta)$ is a many-to-one function of β, and $\mathcal{G}_\eta$ will contain g^{opt} if $\mu(A, X; \beta)$ is correct. Thus, estimating the value $\eta^{\text{opt}} = \arg\max_\eta E\{Y^*(g_\eta)\}$ defining the regime $g^{\text{opt}}_\eta(X) = g(X, \eta^{\text{opt}})$ will yield an estimator for g^{opt}. For complex regression models involving high-dimensional X, the resulting regimes may be difficult to interpret or implement broadly; e.g., if some elements of X are not routinely collected in practice. Then, alternatively, it may be desirable to specify directly a class of regimes indexed by

a parameter η and depending on a key subset of elements of X based on clinical practice, cost, and interpretability, without reference to a regression model. In this case, g^{opt} may or may not be in $\mathcal{G}_\eta$. However, although the regime g_η^{opt} defined by η^{opt} may not be the same as g^{opt}, when attention focuses on the feasible class $\mathcal{G}_\eta$, estimation of g_η^{opt} is still of considerable interest. When the regimes in $\mathcal{G}_\eta$ are derived from a misspecified regression model, $\eta(\widehat{\beta})$ may or may not converge in probability to η^{opt}, and the resulting estimator for the optimal regime based on $\widehat{\beta}$ can exhibit very poor performance, as we demonstrate in Section 4 (see also Qian and Murphy, 2011, section 3), suggesting the need for an alternative approach to estimation of η^{opt}.

Based on these considerations, our approach is to identify an estimator for $E\{Y^*(g_\eta)\}$ and to maximize it directly in η to obtain an estimator $\widehat{\eta}^{\text{opt}}$ for η^{opt} and thus an estimator $\widehat{g}_\eta^{\text{opt}}(X) = g(X, \widehat{\eta}^{\text{opt}})$ for g_η^{opt}. To this end, for fixed η, let $C_\eta = Ag(X,\eta) + (1-A)\{1-g(X,\eta)\}$, so that, when $C_\eta = 1$, $Y = Y^*(g_\eta)$, so that $Y^*(g_\eta)$ is observed; otherwise, if $C_\eta = 0$, then $Y^*(g_\eta)$ is "missing." By analogy to a standard missing data problem as in Cao, Tsiatis, and Davidian (2009), we can conceive of "full data" $\{Y^*(g_\eta), X\}$ and "observed data" $\{C_\eta, C_\eta Y^*(g_\eta), X\} = \{C_\eta, C_\eta Y, X\}$. Note that, $Y^*(g_\eta)$ is a function of $\{Y^*(1), Y^*(0), X\}$, and C_η is a function of $\{A, X\}$. Under the assumptions in Section 2, as $\{Y^*(1), Y^*(0)\}$ is independent of A conditional on X, it follows that $Y^*(g_\eta)$ is independent of C_η conditional on X, which corresponds to the assumption of "missing at random," so that $\text{pr}\{C_\eta = 1 | Y^*(g_\eta), X\} = \text{pr}(C_\eta = 1|X)$. Let $\pi(X) = \text{pr}(A = 1|X)$ denote the propensity score for treatment. It is then straightforward to obtain $\text{pr}(C_\eta = 1|X) = \pi_c(X;\eta) = \pi(X)g(X,\eta) + \{1-\pi(X)\}\{1-g(X,\eta)\}$.

In a randomized trial, $\pi(X)$ is known and is ordinarily a constant; in an observational study, $\pi(X)$ is unknown. In the latter case, as is customary, we may posit a parametric model $\pi(X;\gamma)$, such as the logistic regression model $\pi(X;\gamma) = \exp(\gamma^T \tilde{X})/\{1+\exp(\gamma^T \tilde{X})\}$, $\tilde{X} = (1, X^T)^T$; and estimate γ via the maximum likelihood (ML) estimator $\widehat{\gamma}$ based on the iid (A_i, X_i), $i = 1, \ldots, n$. We may thus estimate $\pi_c(X;\eta)$ by $\pi_c(X;\eta,\widehat{\gamma}) = \pi(X;\widehat{\gamma})g(X,\eta) + \{1-\pi(X;\widehat{\gamma})\}\{1-g(X,\eta)\}$. Note that, although the restricted class $\mathcal{G}_\eta$ may depend on X only through a specific subset of its elements, in an observational study the propensity score model $\pi(X,\gamma)$ should be developed based on all of X to ensure that confounding is addressed.

Following the missing data analogy, we now identify estimators for $E\{Y^*(g_\eta)\}$. For fixed η, the simple inverse probability weighted estimator (IPWE) is given by

$$\begin{aligned} IPWE(\eta) &= n^{-1}\sum_{i=1}^{n} \frac{C_{\eta,i}Y_i}{\pi_c(X_i;\eta,\widehat{\gamma})} \\ &= n^{-1}\sum_{i=1}^{n} \frac{C_{\eta,i}Y_i}{\pi(X_i;\widehat{\gamma})^{A_i}\{1-\pi(X_i;\widehat{\gamma})\}^{1-A_i}}. \end{aligned} \tag{2}$$

As in the missing data context, the estimator (2) is consistent for $E\{Y^*(g_\eta)\}$ if $\pi(X;\gamma)$, and hence $\pi_c(X;\eta,\gamma)$, is correctly specified, but may not be otherwise.

Following Robins, Rotnitzky, and Zhao (1994) and Cao et al. (2009), an alternative estimator that offers protection against such misspecification and improved efficiency is the doubly robust AIPWE

$$\begin{aligned} &AIPWE(\eta) \\ &= n^{-1}\sum_{i=1}^{n}\left\{\frac{C_{\eta,i}Y_i}{\pi_c(X_i;\eta,\widehat{\gamma})} - \frac{C_{\eta,i}-\pi_c(X_i;\eta,\widehat{\gamma})}{\pi_c(X_i;\eta,\widehat{\gamma})}m(X_i;\eta,\widehat{\beta})\right\}. \end{aligned} \tag{3}$$

In (3), $m(X;\eta,\beta) = \mu(1,X,\beta)g(X,\eta) + \mu(0,X,\beta)\{1-g(X,\eta)\}$ is a model for $E\{Y^*(g_\eta)|X\} = \mu(1,X)g(X,\eta) + \mu(0,X)\{1-g(X,\eta)\}$, where $\mu(A,X;\beta)$ is a model for $E(Y|A,X)$, and $\widehat{\beta}$ is an appropriate estimator for β as before. The estimator (3) possesses the double robustness property; i.e., it is consistent for $E\{Y^*(g_\eta)\}$ if either of $\pi(X;\gamma)$, and hence, $\pi_c(X;\eta,\gamma)$, or $\mu(A,X;\beta)$, but not both, is misspecified. Note that, although the regression estimator (1) may be used to estimate $E\{Y^*(g_\eta)\}$ for any arbitrary g_η, regardless of whether or not g_η is derived from a regression model, its consistency hinges critically on correct specification of a model for $E(Y|A,X)$. Likewise, the estimator (2) requires a correct model for $\pi(X;\gamma)$. Thus, relative to these approaches, (3) offers protection against mismodeling of these key quantities. Finally, as shown by Robins et al. (1994), the second term in (3) "augments" the estimator $IPWE(\eta)$ so as to increase asymptotic efficiency; if $\pi(X;\gamma)$ is correctly specified, then the efficient estimator of form (3) is obtained when the regression model is also correct. If $\mu(A,X;\beta)$ is correctly specified, the regression estimator may achieve greater large-sample precision; however, as we demonstrate in Section 4, the gain can be modest.

An estimator for η^{opt} and hence for g_η^{opt} may be obtained by maximizing $AIPWE(\eta)$ in (3) in η to obtain $\widehat{\eta}^{\text{opt}}$ and thus $\widehat{g}_\eta^{\text{opt}}(X) = g(X, \widehat{\eta}^{\text{opt}})$. A corresponding estimator for $E\{Y^*(g_\eta^{\text{opt}})\}$, the population mean outcome using the optimal restricted treatment regime, may be found as $AIPWE(\widehat{\eta}^{\text{opt}})$. Analogous estimators based on $IPWE(\eta)$ may also be obtained; in Section 4, we show that those based on $AIPWE(\eta)$ exhibit superior performance.

Standard errors for these estimators for $E\{Y^*(g_\eta^{\text{opt}})\}$ may be obtained under regularity conditions based on an argument sketched in Web Appendix A. Letting $Q(\eta) = E\{Y^*(g_\eta)\}$ as a function of η, and denoting either estimator by $\widehat{Q}(\eta)$ for arbitrary η, it is shown that

$$n^{1/2}\{\widehat{Q}(\widehat{\eta}^{\text{opt}}) - Q(\eta^{\text{opt}})\} = n^{1/2}\{\widehat{Q}(\eta^{\text{opt}}) - Q(\eta^{\text{opt}})\} + o_p(1), \tag{4}$$

so that the asymptotic variance of the left-hand side of (4) can be approximated by that of the leading term on the right, which can be estimated by the usual sandwich technique (Stefanski and Boos, 2002).

In the situation of a series of sequential decisions, Robins et al. (2008) and Orellana et al. (2010) also consider treatment regimes g_η, say, in a restricted class $\mathcal{G}_\eta$ indexed by a parameter η and propose methods to estimate the optimal regime within the class. These authors motivate their approach in the context of HIV infection, where the goal is to determine the optimal threshold CD4 count η such that, if at any point a subject were to exhibit CD4 count below η, he/she would be administered antiretroviral therapy.

Similar to our approach, the optimal η maximizes $Q(\eta)$ for some outcome of interest. In this more complex time-dependent setting, however, $Q(\eta)$ may be difficult to estimate because the number of subjects in the data treated in accordance with g_η for any fixed η may be quite small. Accordingly, in contrast to our approach, where we maximize an estimator $\widehat{Q}(\eta)$ in η directly, these authors posit a marginal structural mean model for $Q(\eta)$, $M(\eta,\tau)$, say, in terms of a parameter τ; e.g., a quadratic model $M(\eta,\tau)=\tau_0+\tau_1\eta+\tau_2\eta^2$. The estimator $\widehat{\tau}$ is obtained via (augmented) inverse probability weighted estimating equations; see Web Appendix B, and the optimal η is then estimated as $\arg\max_\eta M(\eta,\widehat{\tau})$. In the next section, we compare our approach to this method in the one decision problem.

4. Simulation Studies

We have carried out several simulation studies to evaluate the performance of the proposed methods, each involving 1000 Monte Carlo data sets. We report results here for two scenarios; further results for these and other scenarios are presented in Web Appendices C and D.

For the first scenario, for each data set, we generated $n=500$ observations (Y_i,A_i,X_i), $i=1,\ldots,n$, where $X_i=(X_{i1},X_{i2})^T$ and X_{i1} and X_{i2} were independent and uniformly distributed on $(-1.5,1.5)$; given X_i, A_i were Bernoulli with success probability satisfying $\text{logit}\{\text{pr}(A=1|X)\}=-1.0+0.8X_1^2+0.8X_2^2$, $\text{logit}(u)=\log\{u/(1-u)\}$; and outcomes were generated as $Y_i=\mu(A_i,X_i)+\epsilon_i$ for ϵ_i standard normal and $\mu(A,X)=\exp\{2.0-1.5X_1^2-1.5X_2^2+3.0X_1X_2+A(-0.1-X_1+X_2)\}$. It is straightforward to deduce that the optimal treatment regime $g^{\text{opt}}(X)=I(X_2>X_1+0.1)$, a hyperplane in two-dimensional space. Via Monte Carlo simulation with 10^6 replicates, we obtained $E\{Y^*(g^{\text{opt}})\}=3.71$, $E\{Y^*(0)\}=3.02$, and $E\{Y^*(1)\}=3.14$. Thus, although the strategy of administering treatment 1 to the entire population results in improved mean outcome relative to giving treatment 0 to the entire population, there is added benefit to assigning treatment via g^{opt}, which leads to an 18% increase in mean outcome over treatment 1.

To estimate the optimal regime, we considered the regression estimator and the estimators based on maximizing $IPWE(\eta)$ in (2) and $AIPWE(\eta)$ in (3) in η, respectively, using both correctly and incorrectly specified models $\mu(A,X;\beta)$ and $\pi(X;\gamma)$. In particular, we considered two posited outcome regression models, which we denote as $\mu_t(A,X;\beta)=\exp\{\beta_0+\beta_1X_1^2+\beta_2X_2^2+\beta_3X_1X_2+A(\beta_4+\beta_5X_1+\beta_6X_2)\}$, corresponding to the correct specification; and $\mu_m(A,X;\beta)=\beta_0+\beta_1X_1+\beta_2X_2+A(\beta_3+\beta_4X_1+\beta_5X_2)$, which is misspecified. We estimated β in each model by least squares. For $\pi(X;\gamma)$, we considered the correctly specified model $\text{logit}\{\pi_t(X;\gamma)\}=\gamma_0+\gamma_1X_1^2+\gamma_2X_2^2$ and an incorrect version $\text{logit}\{\pi_m(X;\gamma)\}=\gamma_0+\gamma_1X_1+\gamma_2X_2$, both of which were fitted via ML. Both outcome regression models define a class of treatment regimes $\mathcal{G}_\eta=\{I(\eta_0+\eta_1X_1+\eta_2X_2>0)\}$, so that clearly $g^{\text{opt}}\in\mathcal{G}_\eta$. Expressed in this form, regimes in $\mathcal{G}_\eta$ do not have a unique representation. Rather than achieving this by taking the coefficient of one of the covariates equal to 1 and redefining η, as in the discussion in Section 3, which yields easily interpretable regimes, for computational convenience in automating the simulations we instead equivalently achieved uniqueness by imposing $\|\eta\|=(\eta^T\eta)^{1/2}=1$. In this case, $g^{\text{opt}}\in\mathcal{G}_\eta$ corresponds to $\eta=(\eta_0,\eta_1,\eta_2)^T=(-0.07,-0.71,0.71)^T$.

Both $IPWE(\eta)$ and $AIPWE(\eta)$ are nonsmooth functions in η; accordingly, the use of traditional optimization methods to maximize these quantities in η may be problematic. Accordingly, we used two approaches to maximization of these quantities: a grid search, as described in Web Appendix C, and the genetic algorithm discussed by Goldberg (1989) and implemented in the `rgenoud` package in R (Mebane and Sekhon, 2011). As noted in the documentation, the latter "combines evolutionary algorithm methods with a derivative-based quasi-Newton approach" to address such difficult optimization problems. In our context, we have found this approach to be computationally efficient in higher dimensions, whereas a direct grid search quickly becomes infeasible for dimensions greater than two. In our implementation using the `genoud` function, we adopted the default settings of all arguments except we took `max=TRUE`; `optim.method = Nelder-Mead`, recommended in the documentation for discontinuous objective functions; and `pop.size = 3000`, which we determined to be sufficiently large to achieve satisfactory results via preliminary testing. We took `starting.values = c(0,0,0)`, and set the `Domains` matrix to be the 3×2 matrix with columns $(-1,-1,-1)^T$ and $(1,1,1)^T$, where, each row corresponds to lower and upper bounds on each element of η, so that the algorithm searched in this region. As above, to identify a unique estimated η^{opt}, we imposed the restriction $\|\eta\|=1$, normalizing the value of $\widehat{\eta}^{\text{opt}}$ obtained from `genoud` for each Monte Carlo data set. We provide further discussion of selection of these tuning parameters in Web Appendix C.

Table 1 shows the results using the genetic algorithm to carry out the maximization for the proposed estimators; results using the grid search were almost identical and are given in Web Appendix C. For the regression estimators, we report $\eta(\widehat{\beta})$. For the proposed estimators based on (2) and (3), results are shown using both correct and incorrect models for the propensity score and outcome regression in different combinations. In accordance with the definitions of $Q(\eta)$ and $\widehat{Q}(\eta)$, $\widehat{Q}(\widehat{\eta}^{\text{opt}})$ for each estimator in the table is the Monte Carlo average (standard deviation) of the estimated values of the true $E\{Y^*(g_\eta^{\text{opt}})\}$ and thus reflects how well each approach does at estimating the true achievable mean outcome under the true optimal regime. In contrast, $Q(\widehat{\eta}^{\text{opt}})$ is a measure of the actual performance of the estimated optimal regime itself. Namely, for each Monte Carlo data set, the true mean outcome that would be achieved if each estimated optimal regime were followed by the entire population was determined by simulation, and the values in this column are the Monte Carlo average (standard deviation) of these simulated quantities. Hence, the values in this column, when compared to the true $E\{Y^*(g_\eta^{\text{opt}})\}=3.71$, reflect the extent to which the estimated optimal regime can achieve the performance of the true optimal regime. We also present Monte Carlo coverage probabilities for 95% Wald confidence intervals for $Q(\eta^{\text{opt}})$ constructed using $\widehat{Q}(\widehat{\eta}^{\text{opt}})$ and standard errors obtained as described in Section 3.

To obtain a graphical depiction of the performance of the estimated optimal regimes, we calculated the ratio

Table 1

Results for the first simulation scenario using 1000 Monte Carlo data sets with $n = 500$. For the true optimal regime $g^{\text{opt}} = g_\eta^{\text{opt}}$ within the class $\mathcal{G}_\eta$, $(\eta_0, \eta_1, \eta_2) = (-0.07, -0.71, 0.71)$ and $E\{Y^*(g_\eta^{\text{opt}})\} = 3.71$. $RG\mu_t$ and $RG\mu_m$ represent the regression method using the correct and incorrect outcome regression models; IPWE is the proposed method using (2); and $AIPWE\mu_t$ and $AIPWE\mu_m$ are the proposed method using (3) and the correct and incorrect outcome regression models, respectively. The columns $\widehat{\eta}_0$, $\widehat{\eta}_1$, and $\widehat{\eta}_2$ show Monte Carlo average estimates, with Monte Carlo standard deviations in parentheses. The column $\widehat{Q}(\widehat{\eta}^{\text{opt}})$ shows Monte Carlo average and standard deviation of the estimated values of the true $E\{Y^*(g_\eta^{\text{opt}})\}$, SE shows the Monte Carlo average of sandwich standard errors for this quantity, Cov. shows the coverage of 95% Wald-type confidence intervals for $Q(\eta^{opt})$, and $Q(\widehat{\eta}^{\text{opt}})$ shows the Monte Carlo average and standard deviation of values $E\{Y^*(\widehat{g}_\eta^{\text{opt}})\}$ obtained using 10^6 Monte Carlo simulations for each data set.

Method	$\widehat{\eta}_0$	$\widehat{\eta}_1$	$\widehat{\eta}_2$	$\widehat{Q}(\widehat{\eta}^{\text{opt}})$	SE	Cov.	$Q(\widehat{\eta}^{\text{opt}})$
$RG\mu_t$	−0.07 (0.02)	−0.71 (0.01)	0.71 (0.01)	3.70 (0.14)	–	–	3.71 (0.00)
$RG\mu_m$	−0.51 (0.26)	−0.49 (0.32)	0.46 (0.33)	3.44 (0.18)	–	–	3.27 (0.19)
			PS correct				
IPWE	−0.07 (0.15)	−0.69 (0.11)	0.68 (0.11)	4.01 (0.26)	0.28	86.1	3.63 (0.07)
$AIPWE\mu_t$	−0.07 (0.05)	−0.71 (0.03)	0.70 (0.03)	3.72 (0.15)	0.15	94.7	3.70 (0.01)
$AIPWE\mu_m$	−0.06 (0.12)	−0.69 (0.12)	0.69 (0.13)	3.85 (0.21)	0.23	91.8	3.66 (0.07)
			PS incorrect				
IPWE	−0.38 (0.22)	−0.56 (0.30)	0.55 (0.31)	4.06 (0.22)	0.23	69.4	3.42 (0.20)
$AIPWE\mu_t$	−0.07 (0.05)	−0.70 (0.02)	0.70 (0.02)	3.72 (0.15)	0.15	95.2	3.70 (0.01)
$AIPWE\mu_m$	−0.23 (0.22)	−0.62 (0.25)	0.61 (0.27)	3.81 (0.18)	0.19	94.1	3.57 (0.20)

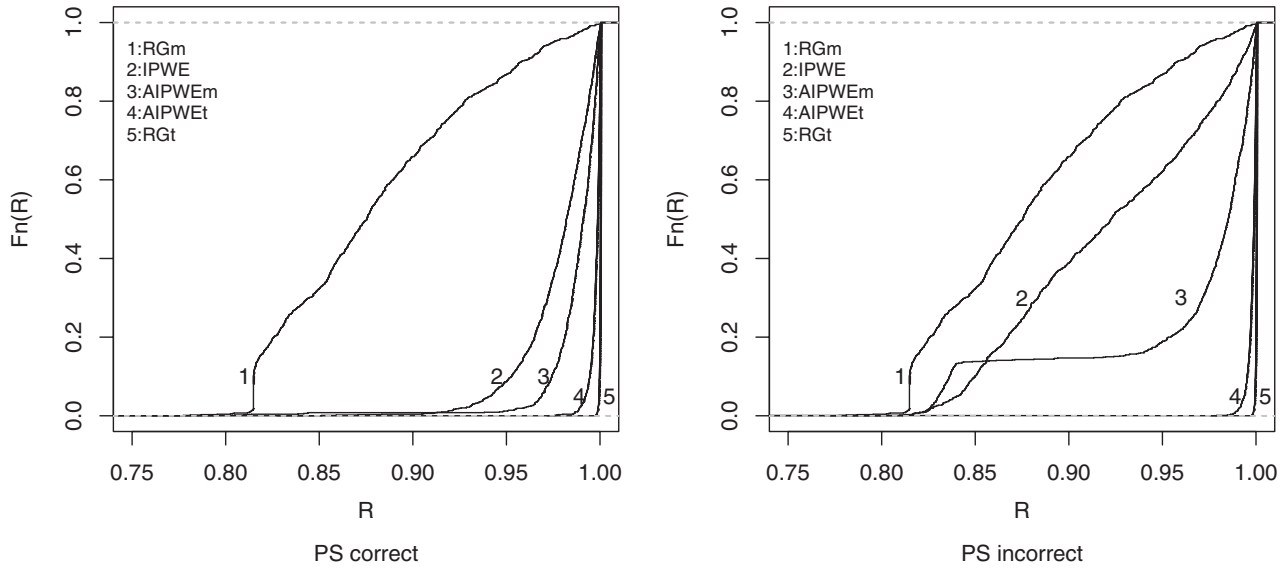

Figure 1. Empirical cdfs across 1000 Monte Carlo data sets using correct and incorrect propensity score (PS) models of the quantities $Q(\widehat{\eta}^{\text{opt}})/E\{Y^*(g_\eta^{\text{opt}})\}$ for each estimator for the first simulation scenario. RGt and RGm denote the regression estimator with correct and misspecified model $\mu(A, X; \beta)$, respectively; AIPWEt and AIPWEm denote the estimator based on (3) with correct and misspecified model $\mu(A, X; \beta)$, respectively; and IPWE denotes the estimator based on (2).

$Q(\widehat{\eta}^{\text{opt}})/E\{Y^*(g^{\text{opt}})\} = Q(\widehat{\eta}^{\text{opt}})/E\{Y^*(g_\eta^{\text{opt}})\}$ for each Monte Carlo data set, which gives the proportion of benefit the estimated regime can achieve if used in the entire population relative to using the true optimal regime. The empirical cumulative distribution function (cdf) of these ratios for each estimator is presented in Figure 1; by definition, "good" estimators should admit empirical cdfs that concentrate at 1.00.

From Table 1, the regression estimator based on a postulated outcome regression that includes the truth yields an estimator for $g^{\text{opt}} = g_\eta^{\text{opt}}$ that virtually achieves the performance of the true optimal regime. However, when the regression model is misspecified, the resulting estimated regime is far from the optimal and leads to relatively poor performance. In contrast, the proposed methods based on $AIPWE(\eta)$ in (3) result in an estimated regime that is almost identical to g^{opt} on average

and performs almost identically to the true optimal regime on the basis of mean outcome, regardless of whether or not the propensity score model is correct. The estimator based on $IPWE(\eta)$ in (2) also yields an estimated regime close to the optimal when the propensity model is correct, but, relative to the regression estimator and AIPWE, is inefficient in estimating the achieved mean outcome under the true optimal regime, and the resulting estimated regime is outperformed by these competing estimators in terms of true mean outcome achieved. When the propensity model is misspecified, this estimator shows a degradation in performance similar to that exhibited by the regression estimator using an incorrect regression model. The estimator based on $AIPWE(\eta)$ in (3) with both propensity and regression model misspecified performs no worse.

The IPWE shows some upward bias in estimation of $E\{Y^*(g_\eta^{\rm opt})\}$, and 95% confidence intervals exhibit undercoverage as a result. Intervals based on the AIPWE show better performance, with some undercoverage when the regression model is misspecified. In Web Appendix C, we present results for $n = 200$ and 1000, which are similar; additional simulations, not shown, with $n = 10,000$, yielded negligible bias and nominal coverage for all estimators, suggesting that this is a sample size issue. Because these estimators involve finding the maximum of a nonsmooth function, this behavior is not unexpected.

Figure 1 shows the performance of all estimators under correct and incorrect propensity score models in panels (a) and (b), respectively, and reiterates graphically the poor performance of the regression estimator under misspecification and the almost identical performance of the regression estimator under a correct outcome model and the AIPWE regardless of whether or not the propensity model is misspecified.

In the second scenario, for each data set, we again generated (Y_i, A_i, X_i), $i = 1, \ldots, n = 500$, where the elements of $X_i = (X_{i1}, X_{i2})^T$ were independent with X_{i1} uniform on $(0, 2)$ and X_{i2} standard normal, and A_i was Bernoulli with $\text{logit}\{\text{pr}(A = 1|X)\} = -1.0 + 0.5X_1^2 + 0.5X_2^2$. Outcomes were generated as $Y_i = \mu(A_i, X_i) + \epsilon_i$ for ϵ_i standard normal and $\mu(A, X) = \exp[2.0 - 0.2X_1 + 0.2X_2 + A\{2.0\,\text{sign}(X_2 - X_1^2 + 1.0)/(2.0 + |X_2 - X_1^2 + 1.0|)\}]$, a rather complicated relationship with corresponding $g^{\rm opt}(X) = I(X_2 > X_1^2 - 1.0)$, for which Monte Carlo simulation using 10^6 replicates yielded $E\{Y^*(g^{\rm opt})\} = 9.50$, whereas $E\{Y^*(0)\} = 6.21$ and $E\{Y^*(1)\} = 8.16$. As it would be unlikely that an analyst would correctly identify the true relationship $\mu(A, X)$, we considered two plausible misspecified working regression models, $\mu_1(A, X; \beta) = \exp\{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + A(\beta_3 + \beta_4 X_1 + \beta_5 X_2)\}$ and $\mu_2(A, X; \beta) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + A(\beta_3 + \beta_4 X_1 + \beta_5 X_2)$, both of which induce the class of treatment regimes $\mathcal{G}_\eta$ with elements of form $I(\eta_0 + \eta_1 X_1 + \eta_2 X_2 > 0)$, where we again take $\|\eta\| = 1$. Correct and incorrect propensity score models were specified as in the first scenario.

Note that, here, in contrast to the first scenario, $\mathcal{G}_\eta$ does not contain $g^{\rm opt}$. Thus, $g_\eta^{\rm opt}$ represents the optimal regime within the class $\mathcal{G}_\eta$ but may not achieve the same performance as the overall $g^{\rm opt}$. Via Monte Carlo simulation using 10^6 replicates, we found that $g_\eta^{\rm opt} = I(0.66 - 0.67X_1 + 0.33X_2 > 0)$ and $E\{Y^*(g_\eta^{\rm opt})\} = 9.33$ (<9.50), so that $g_\eta^{\rm opt}$ results in less than a 2% reduction in mean outcome relative to the overall optimal regime.

Table 2 shows results for this scenario, where again the genetic algorithm described previously was used to implement the estimators based on (2) and (3); results using the grid search were similar and are shown in Web Appendix C. The regression estimators based on both incorrect working outcome regression models yield estimated optimal regimes in the class that are far from achieving the performance of the true optimal regime. In contrast, the proposed estimators based on both (2) and (3) exhibit better performance, with a considerable gain in efficiency for those based on $AIPWE(\eta)$ over that using $IPWE(\eta)$. Evidently, augmentation using an incorrect outcome regression model leads to considerable gains over the IPWE regardless of whether or not the propensity score model is correct. Confidence intervals for $E\{Y^*(g_\eta^{\rm opt})\}$ when the propensity model was correctly specified achieve the nominal level; not unexpectedly, those based on the AIPWE with misspecified propensity yield poor performance.

In Figure 2, because the true optimal regime $g^{\rm opt}$ is unachievable if we restrict to the feasible class $\mathcal{G}_\eta$, we plot the empirical cdfs of the ratios $Q(\widehat{\eta}^{\rm opt})/E\{Y^*(g_\eta^{\rm opt})\}$, which are now different from the ratios $Q(\widehat{\eta}^{\rm opt})/E\{Y^*(g^{\rm opt})\}$. Because $g_\eta^{\rm opt}$ and $g^{\rm opt}$ lead to overall mean outcomes that differ by less than 2%, these ratios are also informative of the performance of the estimated regimes relative to the true optimal regime. The figure provides graphical corroboration of the results in Table 2, namely, that the AIPWE may lead to more reliable inference on the optimal regime than the regression estimator or IPWE, exhibiting the desired robustness to misspecification of one or both models.

Overall, these simulations, along with many others we have conducted, suggest that, although the regression estimator (1) leads to valid inference on the optimal treatment regime when the outcome regression model on which it is based is correctly specified, it can suffer serious degradation of performance if it is not. Likewise, the IPWE based on (2) can perform poorly if the propensity score model is incorrect. The proposed methods based on the AIPWE using (3) exhibit robustness to misspecification of both models and lead to reliable and precise inference on the true optimal regime, either overall or in a class of interest.

The computational burden associated with our methods is minimal. With η three-dimensional and using `pop.size = 3000` in the `genoud` function, processing of one data set took 10–20 seconds; with `pop.size = 10,000`, this took 70–90 seconds. For higher dimensional η, e.g., eight-dimensional, say using `pop.size = 10,000`, results were achieved in 2–3 minutes. A grid search is much less computationally efficient, taking several minutes with two-dimensional η and is infeasible for larger dimensions. These results were obtained using R on a PC with an Intel(R) Core(TM)2 Duo CPU T8300@2.4GHz and 2GB RAM.

In Web Appendix E, we present the results of simulations comparing our methods to the marginal structural mean model approach of Robins et al. (2008) and Orellana et al. (2010) discussed at the end of Section 3. In a scenario where the model $M(\eta, \tau)$ is correctly specified, this approach and ours exhibit comparable performance. When $M(\eta, \tau)$ was incorrect, our approaches, and in particular AIPWE, showed

Table 2

Results for the second simulation scenario using 1000 Monte Carlo data sets with $n = 500$. For the true optimal regime g_η^{opt} within the class $\mathcal{G}_\eta$, $(\eta_0, \eta_1, \eta_2) = (0.66, -0.67, 0.33)$ and $E\{Y^*(g_\eta^{\text{opt}})\} = 9.33$. All other quantities are analogous to those in Table 1, with μ_1 and μ_2 denoting the given estimator using the misspecified models $\mu_1(A, X; \beta)$ and $\mu_2(A, X; \beta)$, respectively.

Method	$\widehat{\eta}_0$	$\widehat{\eta}_1$	$\widehat{\eta}_2$	$\widehat{Q}(\widehat{\eta}^{\text{opt}})$	SE	Cov.	$Q(\widehat{\eta}^{\text{opt}})$
$RG\mu_1$	0.80 (0.03)	−0.56 (0.04)	0.20 (0.06)	8.10 (0.27)	–	–	8.72 (0.17))
$RG\mu_2$	0.80 (0.01)	−0.55 (0.02)	0.23 (0.04)	8.47 (0.27)	–	–	8.68 (0.07)
			PS correct				
IPWE	0.64 (0.05)	−0.66 (0.08)	0.36 (0.11)	9.53 (0.52)	0.44	93.8	9.17 (0.18)
$AIPWE\mu_1$	0.65 (0.02)	−0.67 (0.01)	0.34 (0.05)	9.38 (0.27)	0.28	95.7	9.28 (0.05)
$AIPWE\mu_2$	0.66 (0.02)	−0.67 (0.02)	0.34 (0.05)	9.39 (0.26)	0.27	95.6	9.29 (0.05)
			PS incorrect				
IPWE	0.58 (0.13)	−0.67 (0.06)	0.42 (0.12)	9.48 (0.32)	0.34	95.0	9.03 (0.35)
$AIPWE\mu_1$	0.65 (0.02)	−0.67 (0.02)	0.36 (0.05)	8.99 (0.27)	0.27	75.5	9.27 (0.06)
$AIPWE\mu_2$	0.65 (0.02)	−0.67 (0.02)	0.36 (0.05)	9.10 (0.25)	0.26	84.1	9.27 (0.05)

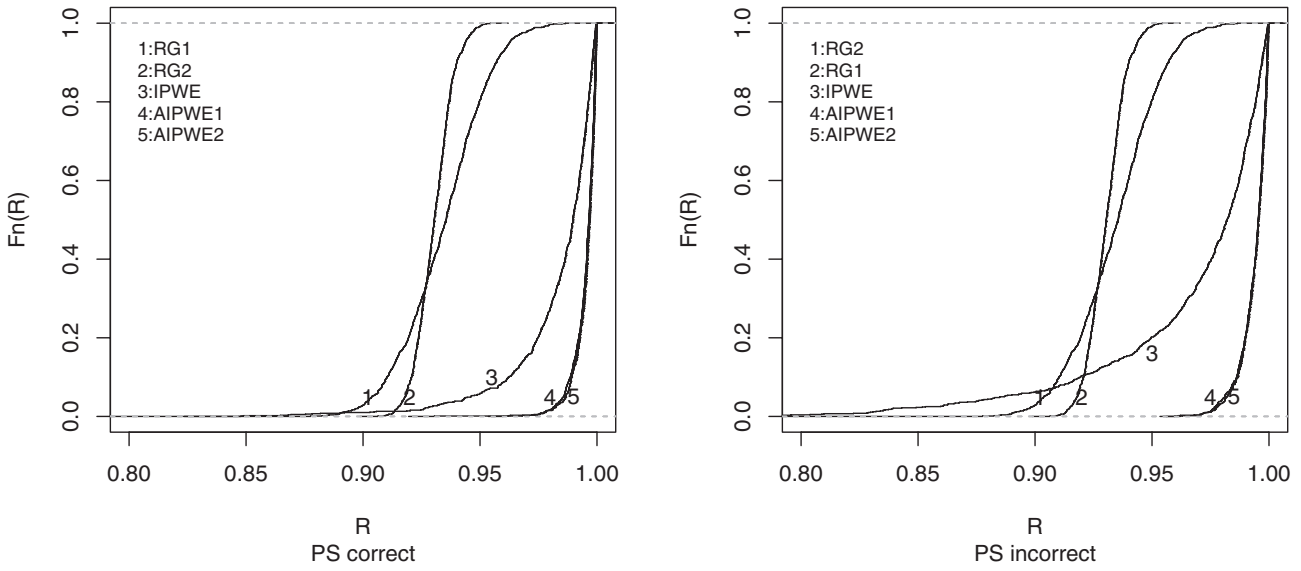

Figure 2. Empirical cdfs across 1000 Monte Carlo data sets using correct and incorrect propensity score (PS) models of the quantities $Q(\widehat{\eta}^{\text{opt}})/E\{Y^*(g_\eta^{\text{opt}})\}$ for each estimator for the second simulation scenario. RG1 and RG2 denote the regression estimator using incorrect models $\mu_1(A, X; \beta)$ and $\mu_2(A, X; \beta)$, respectively; AIPWE1 and AIPWE2 denote the estimator based on (3) using $\mu_1(A, X; \beta)$ and $\mu_2(A, X; \beta)$, respectively; and IPWE denotes the estimator based on (2).

improved relative performance; this is not unexpected, as the posited $M(\eta, \tau)$ in practice will be at best an empirical approximation to $Q(\eta)$.

In the simulation scenarios discussed here, the treatment assignment mechanism depended on X, as would be the case in an observational study. Thus, the analyst must model the propensity score with the risk of misspecifying the true mechanism. In a typical randomized clinical trial, as in the NSABP study discussed next, the true propensity score does not depend on X and may be estimated by the sample proportion assigned to treatment 1. In this setting, then, misspecification of the propensity score is of no concern, and the IPWE and AIPWE estimators will always be consistent, with the latter being more efficient.

5. Application to the NSABP Trial

We apply the proposed methods to data from the NSABP clinical trial introduced in Section 1, with binary outcome $Y = 1$ if a subject survived disease-free to 3 years from baseline, and $Y = 0$ otherwise; and $A = 0$ (1) if a subject was randomized to PF (PFT). We consider estimation of the optimal treatment regime using covariates age (years) and progesterone receptor level (PR, fmol) based on data from the $n = 1276$ patients with complete covariate information. Gail and Simon (1985) applied their testing procedure for qualitative interaction to these data to conclude that there is evidence supporting the regime proposed by Fisher et al. (1983) under which subjects with age < 50 years and PR < 10 fmol receive PF, with all others receiving PFT.

Table 3
Results of fitting the logistic outcome regression model (5) for the NSABP data.

	Estimate	SE	Z value	p-value
Intercept	0.0988	0.4395	0.225	0.82
Age	0.0004	0.0081	0.048	0.96
LPR	0.0992	0.0431	2.304	0.02
Treatment	−1.4580	0.6332	−2.303	0.02
Treatment×age	0.0274	0.0117	2.337	0.02
Treatment×LPR	0.1555	0.0650	2.394	0.02

Because the distribution of PR is very skewed, with PR = 0 for some participants, we consider regimes and models involving LPR = log(PR + 1), and let $X = (X_1, X_2) = (\text{age}, \text{LPR})$. For the regression method, we postulated the logistic regression model

$$\mu(A, X; \beta) = \text{expit}\{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + A(\beta_3 + \beta_4 X_1 + \beta_5 X_2)\} \tag{5}$$

for $E(Y|A, X) = \text{pr}(Y = 1|A, X)$, where $\text{expit}(u) = e^u/(1 + e^u)$. It is straightforward to show that (5) induces the class of regimes of form $I(\beta_3 + \beta_4 X_1 + \beta_5 X_2 > 0)$. Results of fitting this model using the R function `nlm` are presented in Table 3 and show that the interactions of both covariates with assigned treatment are statistically significant at level 0.05. Taking $\beta_4 > 0$ in accordance with the estimate in Table 3, to achieve a unique representation that is straightforward to interpret in practice, we follow the convention in Section 3 and write regimes in the class equivalently as $I(X_1 > \eta_0 + \eta_1 X_2)$, where $\eta_0 = -\beta_3/\beta_4$, $\eta_1 = -\beta_5/\beta_4$. Hence, the estimated optimal regime based on the regression estimator is, in obvious notation, $\widehat{g}^{\text{opt}}_{\eta,\text{reg}}(X) = I(X_1 > 53.2 - 5.68X_2)$; i.e., $I(\text{age} > 53.2 - 5.68 \times \text{LPR})$, which dictates that older patients should receive PFT, where the threshold defining "older" depends on PR level. We may estimate the value of $E\{Y^*(g^{\text{opt}}_\eta)\}$ achieved by the true optimal regime within this class of regimes by substituting $\widehat{g}^{\text{opt}}_\eta$ into any of (1)–(3). As this was a randomized study, for the latter two estimators, which involve estimates of the propensity score $\pi(X)$, $\pi(X)$ may be estimated directly by the sample proportion assigned to PFT; i.e., $\sum_{i=1}^n A_i/n$ for all X. Using the previous notation $\widehat{Q}(\widehat{\eta}^{\text{opt}})$ to denote such estimators for $E\{Y^*(g^{\text{opt}}_\eta)\}$, then, the estimate based on (1) is $\widehat{Q}(\widehat{\eta}^{\text{opt}}) = 0.673$. The estimate based on $AIPWE(\eta)$ in (3) is $\widehat{Q}(\widehat{\eta}^{\text{opt}}) = 0.679$; recall that this estimator is robust to misspecification of the regression model (5).

To implement the proposed estimators based on maximizing $IPWE(\eta)$ and $AIPWE(\eta)$ in (2) and (3), we used $\pi(X)$ estimated as above and the ML fit of $\mu(A, X; \beta)$ in (5) in (3). The estimated regime using $IPWE(\eta)$ is $\widehat{g}^{\text{opt}}_{\eta,IPWE}(X) = I(X_1 > 60.0 - 8.68X_2) = I(\text{age} > 60.0 - 8.68 \times \text{LPR})$, with associated $\widehat{Q}(\widehat{\eta}^{\text{opt}}) = 0.693$ and Wald-type 95% confidence interval using estimated standard errors obtained as in (4) $(0.657, 0.729)$. The corresponding quantities using $AIPWE(\eta)$ are $\widehat{g}^{\text{opt}}_{\eta,AIPWE}(X) = I(X_1 > 60.0 - 7.98X_2) = I(\text{age} > 60.0 - 7.98 \times \text{LPR})$, with $\widehat{Q}(\widehat{\eta}^{\text{opt}}) = 0.695$ and 95% confidence interval $(0.659, 0.730)$. The results are virtually identical, and suggest a close to 70% three-year disease-free survival rate if treatment were to be assigned in accordance with the optimal regime of this form. Figure 3(a) depicts the three regimes graphically, superimposed on the observed values of (age, LPR) for all n subjects.

To compare to the regime identified by Fisher et al. (1983) and Gail and Simon (1985), consider the alternative class of regimes defined directly as $1 - I(\text{age} < \eta_0 \text{ and } \text{PR} < \eta_1)$. As noted above, these authors advocated the regime defined by $\eta_0 = 50$ and $\eta_1 = 10$, which, using the estimator for $E\{Y^*(g_\eta)\}$ based substituting this regime in (3) with

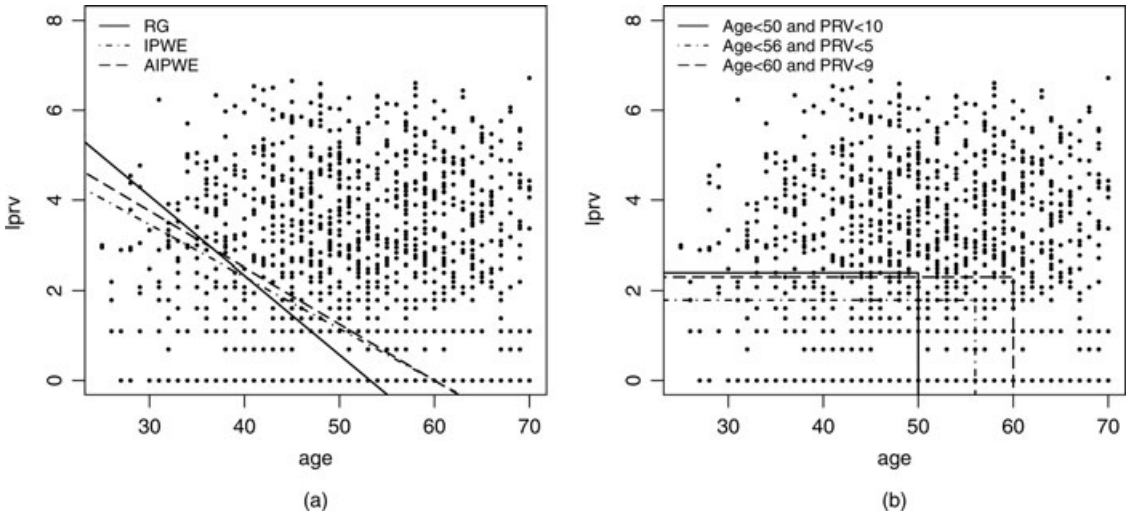

Figure 3. (a) Estimated optimal treatment regimes of the form $I(\text{age} \geq \eta_0 + \eta_1 \text{LPR})$ using the regression estimator (RG), the estimator based on (2) (IPWE), and the estimator based on (3) (AIPWE). (b) The regime identified by Fisher et al., (1983) and Gail and Simon (1985) (solid lines) and optimal regimes of the form $1 - I(\text{age} < \eta_0 \text{ and } \text{PR} < \eta_1)$ estimated based on (2) (dotted-dashed lines) and (3) (long dashed lines).

the estimators for $\pi(X)$ and $\mu(A, X; \beta)$ as above, yields estimated mean outcome 0.679 $(0.643, 0.715)$. Finding $\widehat{\eta}^{\text{opt}}$ maximizing $IPWE(\eta)$ and $AIPWE(\eta)$ in (2) and (3) yields estimated regimes $\widehat{g}^{\text{opt}}_{\eta,IPWE}(X) = 1 - I(\text{age} < 56 \text{ and PR} < 5)$, with $\widehat{Q}(\widehat{\eta}^{\text{opt}}) = 0.681$ $(0.644, 0.717)$; and $\widehat{g}^{\text{opt}}_{\eta,AIPWE}(X) = 1 - I(\text{age} < 60 \text{ and PR} < 9)$, with $\widehat{Q}(\widehat{\eta}^{\text{opt}}) = 0.686(0.651, 0.722)$, respectively. These estimated regimes are shown in Figure 3(b) along with that identified by the investigators. All three regimes suggest that PF should be given to younger patients with low PR levels.

6. Discussion

We have proposed new methods for estimating the optimal treatment regime within a specified class of regimes at a single decision point, where the class may be defined through a postulated model for the regression of outcome on treatment assignment and subject covariates or based on other considerations, such as practice, cost, or simplicity. The methods may be applied to data on outcome, treatment received, and baseline covariates from a clinical trial or observational study. In the latter case, under the assumption of no unmeasured confounders, the methods take account of possible confounding through modeling of the propensity score. Our simulation studies demonstrate that the methods can achieve comparable performance to those based on direct outcome regression modeling with the added benefit of robustness to misspecification of either the outcome regression or propensity model. We have presented the methods for the case of two treatment options, but they may be adapted straightforwardly to more than two treatments.

The proposed methods may be extended to the case of more than one treatment decision; we are currently studying them in this setting and will report the results in a future article.

7. Supplementary Materials

Web Appendices A–E referenced in Sections 3 and 4 are available with this paper at the Biometrics website on Wiley Online Library.

Acknowledgements

This work was supported by NIH grants R37 AI031789, R01 CA051962, R01 CA085848, and P01 CA142538. The authors thank the NSABP for providing the clinical trial data.

References

Bang, H. and Robins, J. M. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics* **61,** 962–972.

Brinkley, J., Tsiatis, A. A., and Anstrom, K. J. (2009). A generalized estimator of the attributable benefit of an optimal treatment regime. *Biometrics* **21,** 512–522.

Cao, W., Tsiatis, A. A., and Davidian, M.(2009). Improving efficiency and robustness of the doubly robust estimator for a population mean with incomplete data. *Biometrika* **96,** 723–734.

Fisher, B., Redmond, C., Brown, A., Wickerham, D. L., Wolmark, N., Allegra, J., Escher, G., Lippman, M., Savlov, E., Wittliff, J., and Fisher, E. R., with the contributions of Plotkin, D., Bowman, D., Wolter, J., Bornstein, R., Desser, R., Frelick, R., and other NSABP Investigators. (1983). Influence of tumor estrogen and progesterone receptor levels on the response to Tamoxifen and chemotherapy in primary breast cancer. *Journal of Clinical Oncology* **1,** 227–241.

Gail, M. and Simon, R. (1985). Testing for qualitative interactions between treatment effects and patient subsets. *Biometrics* **41,** 361–372.

Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning.* Reading, MA: Addison-Wesley.

Gunter, L., Zhu, J., and Murphy, S. A. (2011). Variable selection for qualitative interactions in personalized medicine while controlling the family-wise error rate. *Journal of Biopharmaceutical Statistics* **21**, 1063–1078.

Henderson, R., Ansell, P., and Alshibani, D. (2010). Regret-regression for optimal dynamic treatment regimes. *Biometrics* **66,** 1192–1201.

Mebane, W. R. and Sekhon, J. S. (2011). Genetic optimization using derivatives: The rgenoud package for R. *Journal of Statistical Software* **42,** 1–26.

Moodie, E., Richardson, T. S., and Stephens, D. (2007). Demystifying optimal dynamic treatment regimes. *Biometrics* **63,** 447–455.

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

Orellana, L., Rotnitzky, A., and Robins, J. M. (2010). Dynamic regime marginal structural mean models for estimation of optimal treatment regimes, part I: Main content. *International Journal of Biostatistics* **6,** Issue 2, Article 8, doi: 10.2202/1557-4679.1200

Qian, M. and Murphy, S.A. (2011). Performance guarantees for individualized treatment rules. *Annals of Statistics* **39,** 1180–1210.

Robins, J., Orellana, L., and Rotnitzky, A. (2008). Estimation and extrapolation of optimal treatment and testing strategies. *Statistics in Medicine* **27,** 4678–4721.

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

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association* **89,** 846–866.

Robins, J. M., Hernán, M., and Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology* **11,** 550–560.

Rubin, D. B. (1974). Estimating causal effects of treatments in

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

Stefanski, L. A. and Boos, D. D. (2002). The calculus of M-estimation. *The American Statistician* **56,** 29–38.

Zhao, Y., Kosorok, M. R., and Zeng, D. (2009). Reinforcement learning design for cancer clinical trials. *Statistics in Medicine* **28,** 3294–3315.

Received September 2011. *Revised January* 2012. *Accepted March* 2012.