The Annals of Applied Statistics
2019, Vol. 13, No. 3, 1989–2010
https://doi.org/10.1214/19-AOAS1266

A SEMIPARAMETRIC MODELING APPROACH USING BAYESIAN ADDITIVE REGRESSION TREES WITH AN APPLICATION TO EVALUATE HETEROGENEOUS TREATMENT EFFECTS

BY BRET ZELDOW, VINCENT LO RE III AND JASON ROY

Harvard Medical School, Perelman School of Medicine and Rutgers School of Public Health

Bayesian Additive Regression Trees (BART) is a flexible machine learning algorithm capable of capturing nonlinearities between an outcome and covariates and interactions among covariates. We extend BART to a semiparametric regression framework in which the conditional expectation of an outcome is a function of treatment, its effect modifiers, and confounders. The confounders are allowed to have unspecified functional form, while treatment and effect modifiers that are directly related to the research question are given a linear form. The result is a Bayesian semiparametric linear regression model where the posterior distribution of the parameters of the linear part can be interpreted as in parametric Bayesian regression. This is useful in situations where a subset of the variables are of substantive interest and the others are nuisance variables that we would like to control for. An example of this occurs in causal modeling with the structural mean model (SMM). Under certain causal assumptions, our method can be used as a Bayesian SMM. Our methods are demonstrated with simulation studies and an application to dataset involving adults with HIV/Hepatitis C coinfection who newly initiate antiretroviral therapy. The methods are available in an R package called semibart.

1. Introduction. The number of antiretroviral medications available to persons living with HIV has grown enormously in the past thirty years—a far cry from the few that were available in the early 1990s. Despite detailed guidelines on HIV treatment, challenges in prescribing persist, particularly for individuals with co-morbidities [National Institutes of Health (2018)]. In the United States approximately 25% of patients with HIV also have chronic Hepatitis C virus (HCV) [Centers for Disease Control and Prevention (2017)]. As HCV can lead to liver damage, clinicians must be mindful when prescribing antriretrovirals to patients with HIV/HCV coinfection; improving HIV-related outcomes is ineffectual if accompanied by a fatal decline in liver function.

In previous work, we estimated the effect that certain mitochondrial toxic nucleoside reverse transcriptase inhibitors (mtNRTIs)—didanosine, stavudine, zalcitabine, and zidovidine—had on risk of liver decompensation and death when

Received June 2018; revised May 2019.
Key words and phrases. Bayesian Additive Regression Trees, structural mean model, antiretrovirals.

used as part of an antiretroviral regimen (compared to antiretroviral regimens containing other NRTIs) [Lo Re et al. (2017)]. While these four drugs are no longer recommended for initial treatment, they are still used in resource-limited settings and in salvage regimens. Using Cox marginal structural models, we found that increased cumulative exposure to mtNRTIs was associated with higher risk of decompensation and death. Here, we extend those results and investigate a potential modifier of the effect—fibrinogen-4 (FIB-4).

FIB-4 is a marker of liver injury in which higher values indicate worse liver function. In this paper we ask whether the effect of mtNRTIs on the risk of death (within two years of antiretroviral initiation) changes for individuals with varying FIB-4 levels. To answer this question, we develop a model that has an interpretable, parametric form for the mtNRTIs and its interaction with FIB-4 while remaining nonparametric in the functional form of the confounders.

From a Bayesian standpoint, we can treat the unknown function of the confounders as a random parameter, assign it a prior distribution within an appropriate function space that recognizes our prior knowledge (or lack thereof), and then estimate it using our data. One such prior is the Gaussian process, which induces flexibility through its covariance function [Rasmussen (2006)]. Alternatively, we could assume the function of the confounders is approximated by basis functions like splines or wavelets and assign priors to the coefficients of the bases [Eilers and Marx (1996), Müller et al. (2015)]. Splines, in particular, have been used extensively in Bayesian nonparametric and semiparametric regression. For example, Biller (2000) presented a semiparametric GLM where one variable was modeled using splines and the remaining variables were part of a parametric linear model. Holmes and Mallick (2001) developed a flexible Bayesian piecewise regression using linear splines. The approach in Denison, Mallick and Smith (1998a) involved piecewise polynomials which were able to approximate nonlinearities. Biller and Fahrmeir (2001) introduced a varying-coefficient model with B-splines with adaptive knot locations.

Alongside these Bayesian methods reside two common procedures to predict an outcome given an unknown function of covariates: generalized additive models (GAM) [Hastie and Tibshirani (1990)] and multivariate adaptive regression splines (MARS) [Friedman (1991)]. GAM allows each predictor to have its own functional form using splines. However, any interactions between covariates must be specified by the analyst, which can pose difficulties in high-dimensional problems with multi-way interactions. Bayesian versions of GAM based on P-splines exist [Brezger and Lang (2006)] but are not widely available in statistical software like the frequentist version. MARS is a nonparametric procedure which can automatically detect nonlinearities and interactions through basis functions also based on splines. A Bayesian MARS algorithm has been developed [Denison, Mallick and Smith (1998b)], but also lacks accessible off-the-shelf software. A third option for nonparametric estimation is Bayesian Additive Regression Trees (BART), which, like MARS, allows for nonlinear relationships between an outcome and

covariates and interactions between covariates, while taking a Bayesian approach to estimation [Chipman, George and McCulloch (2010)].

In this paper we introduce a novel Bayesian semiparametric model using BART, which we call semi-BART. Semi-BART partitions the covariate space into two distinct subsets: (1) covariates relevant to the research question such as treatment and effect modifiers (in our example, mtNRTI use and FIB-4) and (2) confounders or other covariates that are not directly relevant to the research question. Semi-BART works by modeling treatment and effect modifiers using linear terms and the confounders with BART. The benefits are easy interpretation of the linear terms and flexibility for the rest. This contrasts with other methods like GAM/MARS or linear regression which require either full flexibility or full parametric specification. We choose to modify BART rather than other nonparametric models (GAM, MARS, random forest, etc.) because it has proved successful in practical settings [van der Laan and Rose ((2011), Chapter 3)], and we wanted a Bayesian method that allowed for direct inference and easily interpretable credible intervals.

Our goal is to provide a new semiparametric tool for use in diverse settings. For example, we can use semi-BART to quantify effect modifiers in personalized medicine applications. We also show how semi-BART is equivalent to a structural mean model (SMM), making it the first Bayesian SMM. And perhaps most importantly, we imagine semi-BART to be a practical substitute for commonly used methods such as linear regression. In the rest of this paper, we provide relevant background to semi-BART (Section 2), show its equivalence to SMMs (Section 3), describe computational details (Section 4), perform simulation studies against competitor models (Section 5), and evaluate FIB-4 as an effect modifier of mtNRTIs (Section 6).

2. Review of Bayesian Additive Regression Trees. BART is an algorithm that uses sum-of-trees to predict a binary or continuous outcome given predictors. For continuous outcomes Y, let $Y = \omega(X) + \epsilon$ where $\epsilon \sim N(0, \sigma^2)$, and $\omega(\cdot)$ is the unknown function relating the covariates X to the outcome Y. For binary Y we use a probit link function so that $\Pr(Y = 1|X) = \Phi(\omega(X))$, where $\Phi(\cdot)$ is the distribution function of a standard normal variable. We write the BART sum-of-trees model as $\omega(x) = \sum_{j=1}^{m} \omega_j(x; T_j, M_j)$, where each $\omega_j(x)$ is a single tree and T_j and M_j are the parameters that represent the tree structure and end node parameters, respectively. Each individual tree is a sequence of binary decisions based on X that yield predictions of Y within clusters of observations with similar covariate patterns. Typically, the number of trees m is chosen to be large, and each tree is restricted to be small through regularization priors. This restricts the influence of any single tree and allows for nonlinearities and interactions that would not be possible with any one tree. In Figure 1 we provide an example of a BART fit to a nonlinear mean function $y = \sin(x) + \epsilon$ for a univariate predictor x re-

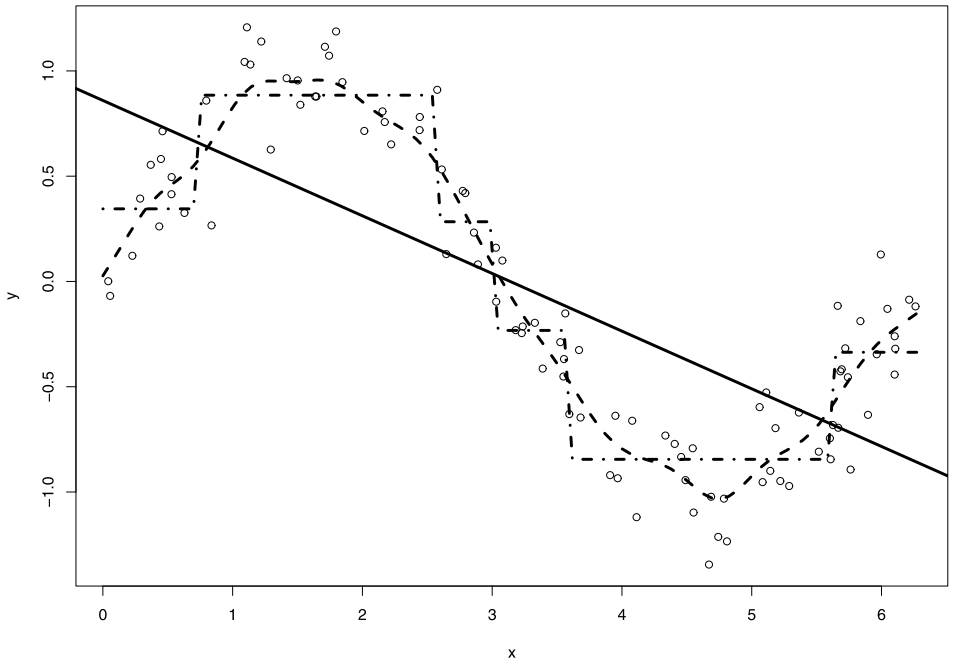

FIG. 1. *Illustration of a BART fit with a univariate predictor space* $x \in [0, 2\pi]$ *and mean response* $y = \sin(x) + \epsilon$. *The solid line is the fit using linear regression, the dashed line is the fit of BART, and the dashed–dotted line is the fit of a single tree.*

stricted to $[0, 2\pi]$, along with the fits of a single regression tree and linear regression.

The Markov Chain Monte Carlo (MCMC) algorithm for BART incorporates Bayesian backfitting [Hastie and Tibshirani (2000)], which we summarize below. Recall that $y_i = \sum_{j=1}^{m} \omega_j(x_i; T_j, M_j) + \epsilon_i$ where ϵ_i is assumed zero-mean normal with unknown variance σ^2. The algorithm alternates between updates to the error variance σ^2 and updates to the trees ω_j. To update σ^2, we find the residuals from the current fit and draw a new value for σ^2. In Chipman, George and McCulloch (2010) and this paper, we use a conjugate inverse χ^2- distribution for the prior of σ^2, so drawing a new value is also a draw from an inverse χ^2- distribution. Second, the trees ω_j are updated one at a time. Starting with ω_1, we compute the residuals of the outcome by subtracting off the fit of the other $m - 1$ trees, $\omega_2, \ldots, \omega_m$. We then propose a modification for the tree ω_1, which is either accepted or rejected by a Metropolis–Hastings step. We update the trees $\omega_2, \ldots, \omega_m$ in the same fashion. More details are available in the original BART paper [Chipman, George and McCulloch (2010)]. In the next section, we propose a semiparametric extension of BART, called semi-BART, where a small subset of covariates have linear functional form and the rest are modeled with BART's sum-of-trees.

3. Semi-BART model.

3.1. *Notation*. Suppose we have n independent observations. Let Y denote the outcome, which can be binary or continuous. Let A denote treatment, which can also be binary or continuous. The remaining covariates we call $\mathbf{X}$. Let $\mathbf{L} = (A, \mathbf{X})$.

3.2. *Semiparametric generalized linear model*. In applied research, often the effects of only a few covariates are of scientific interest, while a larger number of covariates are needed to address confounding. Our model imposes linearity on this small subset of covariates, while retaining flexibility in modeling the rest of the covariates whose exact functional form in relation to the outcome may be considered a nuisance. We partition the predictors into two distinct subsets so that $\mathbf{L} = \mathbf{L}_1 \cup \mathbf{L}_2$ and $\mathbf{L}_1 \cap \mathbf{L}_2 = \varnothing$. Here, $\mathbf{L}_1$ represents nuisance covariates that we must control for but are not of primary interest and $\mathbf{L}_2$ represents covariates that are directly pertinent to the research question, such as treatment A and its effect modifiers. For continuous Y, we write $Y_i = \omega(\mathbf{L}_1) + h(\mathbf{L}_2; \psi) + \epsilon_i$, where $h(\cdot)$ is a parametric function of its covariates in ψ (as in linear regression) but $\omega(\cdot)$ is a function with unspecified form. The errors ϵ_i are assumed independent and identically distributed mean zero and normally distributed with unknown variance σ^2. More generally, we write $g[E(Y|\mathbf{L}_1, \mathbf{L}_2)] = \omega(\mathbf{L}_1) + h(\mathbf{L}_2; \psi)$, for a known link function g. We call this the semi-BART model since we estimate $\omega(\cdot)$ using BART. Note that if $\mathbf{L_1} = \mathbf{L}$ and $\mathbf{L}_2 = \varnothing$, we have a nonparametric BART model. On the other hand if $\mathbf{L_1} = \varnothing$ and $\mathbf{L}_2 = \mathbf{L}$, we have a fully parametric regression model. While there is no restriction on the dimensionality of $\mathbf{L_1}$ and $\mathbf{L_2}$, we assume that $\mathbf{L_1}$ is large enough such that BART is a reasonable choice of an algorithm and that $\mathbf{L_2}$ contains only a few covariates that are of particular interest.

3.3. *Special case*: *Structural mean models*. We also consider a special case of our semiparametric GLM from an observational study with no unmeasured confounders. In doing so we introduce additional notation specific to this section. As before, the exposure of interest is denoted A and can be either binary or continuous. The counterfactual Y^a denotes the outcome that would have been observed under exposure $A = a$. For the special case of binary A, each individual has two counterfactual outcomes—Y^1 and Y^0—but we observe at most one of the two, corresponding to the actual level of exposure received. That is, $Y = AY^1 + (1 - A)Y^0$.

Robins developed structural nested mean models to adjust for time-varying confounding with a longitudinal exposure [Robins (1994, 2000)]. In the case of a point treatment, structural nested mean models are no longer nested and are instead called structural mean models (SMMs). While time-varying confounding with point treatments is not a concern, SMMs still parameterize a useful causal contrast—the average effect of treatment among the treated given the covariates

[Chamberlain (1987), Vansteelandt and Goetghebeur (2003), Vansteelandt and Joffe (2014)]. Write this as:

$$g\{E(Y^a|\mathbf{X}=\mathbf{x}, A=a)\} - g\{E(Y^0|\mathbf{X}=\mathbf{x}, A=a)\} = h^*(x, a; \psi^*), \tag{3.1}$$

where g is a known link function. In this paper, we provide a Bayesian solution to (3.1). To do so, we impose some restrictions on $h^*(\cdot; \psi^*)$, requiring that under no treatment or when there is no treatment effect the function $h^*(\cdot; \psi^*)$ must equal 0. That is, $h^*(x, a; \psi^*)$ satisfies $h^*(x, 0; \psi^*) = h^*(x, a; 0) = 0$. Some examples of $h^*(x, a; \psi^*)$ are $h^*(x, a; \psi^*) = \psi a$ or $h^*(x, a; \psi^*) = (\psi_1 + \psi_2 x_3)a$, when some covariate x_3 modifies the effect of a on y.

While expression (3.1) cannot be evaluated directly due to the unobserved counterfactuals, two assumptions are needed to identify it with observed data [Chamberlain (1987), Vansteelandt and Joffe (2014)].

1. Consistency: If $A = a$, then $Y^a = Y$;
2. Ignorability: $A \perp Y^0|X$.

The consistency assumption asserts that we actually get to see an individual's counterfactual corresponding to the exposure received. Ignorability ensures the exposure A and the counterfactual under no treatment Y^0 are independent given X. Under these two assumptions together with the parametric assumption of $h^*(\cdot)$, the contrast on the left-hand side of (3.1) is identified, and the SMM from (3.1) can be rewritten using observed variables as

$$g\{E(Y|X, A)\} = \omega(\mathbf{L}_1) + h^*(\mathbf{L}_2; \psi^*), \tag{3.2}$$

where $\omega(\mathbf{L}_1)$ is unspecified and $h^*(\mathbf{L}_2; \psi^*)$ is a linear function of $\mathbf{L}_2$ [Chamberlain (1987), Vansteelandt and Joffe (2014)]. Note that the left-hand side of (3.1) is nonparametrically identified with a third assumption, dropping the parametric assumption on $h^*(\cdot)$. That is,

3. Positivity: $\Pr(A = a|\mathbf{X} = \mathbf{x}) > 0 \ \forall x$ such that $\Pr(\mathbf{X} = \mathbf{x}) > 0$.

The positivity assumption states that whenever $\mathbf{X} = \mathbf{x}$ has a positive probability of occurring, there is positive probability that an individual is treated. This assumption is violated in situations where treatment is deterministic at certain levels of $\mathbf{X} = \mathbf{x}$. For example, we restrict our data analysis to the years 2002–2009 to ensure positivity. Prior to 2002, our treatment (mtNRTI use) was near ubiquitous because mtNRTIs were commonly prescribed as first line medications. As a result, if we were to expand our dataset to include earlier years, we could encounter positivity violations.

Let us return to the special case where there is effect modification by a baseline covariate, such as in our data analysis (Section 5) where baseline FIB-4 modifies the effect of treatment (mtNRTI) on death. Let $h^*(x, a; \psi^*) = (\psi_1 + \psi_2 x_3)a$ where a represents a binary indicator of mtNRTI use and x_3 represents FIB-4.

Note that Model (3.2) has no "main effect" for FIB-4. We address this by fitting the model $g\{E(Y|X, A)\} = \{\omega(\mathbf{L}_1) + \psi_3 x_3\} + h^*(x, a; \psi^*)$. This setup still lies within the bounds of semi-BART, described in Section 3.2. An advantage for having a parametric form for x_3 is that researchers interested in quantifying x_3 as an effect modifier can also interpret its main effect. In principle we could have included the effect of x_3 in $\omega(\cdot)$, but we would lose interpretability. We explore the impact of using a linear term for x_3 (as opposed to including x_3 in $\omega(\cdot)$) using simulations, which can be found at https://www.github.com/zeldow/semibart-extras or in the supplementary files [Zeldow, Lo Re III and Roy (2019)].

3.4. *Causal effects with BART: Literature review.* Hill (2011) previously estimated causal effects on the treated using BART. The methods in that paper correspond to our setting in the case of a continuous outcome, in which interest lies in the treatment effect averaged over (possibly) heterogeneous individual-level effects. Semi-BART extends this setup to binary outcomes, continuous-valued treatment, or where low-dimensional summaries of effect modification are of interest, particularly with continuous effect modifiers. In settings with continuous outcomes, binary treatment, and no effect modification (or with a binary effect modifier), the methods presented in Hill (2011) are preferred, whose methods we include in our simulations.

Green and Kern (2012) also modeled treatment heterogeneity using BART with survey experiments. Their methodology is similar to that in Hill (2011) in that effects are calculated using modified datasets. They first fit BART to their dataset, then create two updated versions of the same dataset. The first version has the effect modifier set to a reference level and the second version set to another level. Using the BART fit from the original dataset, they obtain predictions on each of the two modified datasets and then use those predictions to calculate conditional treatment effects. A major difference between our method and the above methods are that we can summarize treatment heterogeneity with a low-dimensional parameter. While this is not desired or appropriate for every application, in some settings it can be useful and efficient.

Hahn, Murray and Carvalho (2018) also developed a method to estimate conditional treatment effects using BART, incorporating propensity score estimates to reduce confounding [Hahn, Murray and Carvalho (2018)]. Like ours, they use a two-pronged regression strategy consisting of a function representing the impact of covariates (like our $\omega(L_1)$) and a part that repesents treatment effects (like our $h(L_2; \psi)$). The ideas in Hahn, Murray and Carvalho (2018) diverge from ours in their intended applications. Hahn, Murray and Carvalho (2018) use a clever prior and the propensity score to improve causal estimates that are biased due to regularization; our method, on the other hand, is intended to appeal to users who otherwise would use linear models.

3.5. *Computations*. The algorithm for semi-BART follows the BART algorithm—briefly reviewed in Section 2—with an additional step. We solve equation (3.2), where $\omega(\mathbf{L}_1)$ can be written as the sum-of-trees $\sum_{j=1}^{m} \omega_j(\mathbf{L}_1; T_j, M_j)$. The parameter T_j contains the structure of the jth tree; for instance, the covariates and rules on which the tree splits. The parameters M_j contain the endnode parameters for the jth tree. For example, the mean of the kth endnode of the jth tree is assumed to be normally distributed with mean μ_{jk} and variance σ_{jk}^2.

For continuous outcomes, we assume independent errors distributed as $N(0, \sigma^2)$ with σ^2 unknown and proceed as follows. Initialize all values including the error variance σ^2, the parameters ψ^*, and the tree structures T_j and M_j for $j = 1, \ldots, m$ and iterate through the following steps. First, update the m trees one at a time. Starting with the first tree $\omega_1(\cdot; T_1, M_1)$, calculate the residuals by subtracting the fit of the remaining $m - 1$ trees at their current parameter values as well as the fit of the linear part $h^*(\mathbf{L}_2; \psi^*)$. That is, for the ith individual, we calculate $y_i^* = y_i - \omega_{-1}(\mathbf{L}_{1i}) - h^*(\mathbf{L}_{2i}; \psi^*)$, where $\omega_{-1}(\mathbf{L}_{1i})$ indicates the fit of the trees except the first tree. As in Chipman, George and McCulloch (2010), a modification of the tree is now proposed. We can grow the tree (breaking one endnode into two endnodes), prune the tree (collapse two endnodes into one), change a splitting rule (for nonterminal nodes), or swap the rules between two nodes. We accept or reject this modification with a Metropolis–Hastings step given the residuals $\mathbf{y}^* = \{y_1^*, \ldots, y_n^*\}$ [Chipman, George and McCulloch (1998)]. Once we have updated $\omega_1(\cdot; T_1, M_1)$, we update $\omega_2(\cdot; T_2, M_2)$ in the same fashion and continue until all m trees are completed.

Next we update ψ^* (in our data analysis, this will include a parameter for treatment, a parameter for FIB-4, and a parameter for their interaction), given a multivariate normal prior for ψ^* so that $p(\psi^*) \sim \text{MVN}(0, \sigma_\psi^2 \mathbf{I})$, where $\mathbf{I}$ is the identity matrix of appropriate dimension and σ_ψ^2 is large enough so that the prior is diffuse. We calculate the residuals after subtracting off the fit of all m trees so that $y_i^* = y_i - \omega(\mathbf{L}_{1i})$. The posterior for ψ is multivariate normal with covariance $\Sigma_\psi = [\frac{\mathbf{L}_2^T \mathbf{L}_2}{\sigma^2} + \frac{\mathbf{I}}{\sigma_\psi^2}]^{-1}$ and mean $\Sigma_\psi [\frac{\mathbf{L}_2 \mathbf{y}^*}{\sigma^2} + \frac{\boldsymbol{\psi}_0}{\sigma_\psi^2}]$, where $\mathbf{y}^*$ is the n-vector of residuals [Gelman et al. (2013)].

Lastly, we update the error variance σ^2. We calculate the residuals given the trees $\omega(\cdot)$ and ψ^* so that $y_i^* = y_i - \omega(\mathbf{L}_{1i}) - h(\mathbf{L}_{2i}; \psi^*)$. With a conjugate scaled inverse χ^2 distribution for σ^2 (parameters ν_0 and λ_0), the posterior is an updated scaled inverse χ^2 distribution with parameters $\nu_n = \nu_0 + n$ and $\lambda_n = \nu_0 \lambda_0 + \langle y^*, y^* \rangle$ where $\langle \rangle$ indicates the dot product. These three steps are repeated until the posterior distributions are well approximated.

Our algorithm for binary outcomes with a probit link uses the underlying latent variable formulation of Albert and Chib (1993), replacing the step in the algorithm that updates the error variance σ^2. Full details of the BART portion of the algorithm are available in Chipman, George and McCulloch (2010), whereas our code for semi-BART is available at https://www.github.com/zeldow/semibart.

4. Simulation. We use simulation to compare the performance of semi-BART to competitor models when estimating the regression coefficient for simulated treatment along with the coefficients for its effect modifiers (main effects and interaction terms with treatment). Our competitors were BART (taken from Hill (2011)), GAM, and linear regression for continuous outcomes and probit regression for binary outcomes. For all simulations, we generated 500 datasets at sample sizes of $n = 500$ and $n = 5000$, and we estimated mean bias, 95% credible/confidence interval coverage, and empirical standard deviation (ESD; defined as the standard deviation of point estimates from each simulated dataset). For GAM we used the mcgv package in R along with splines (using the *s* function [the function used to define smooth terms within GAM formulae] with default settings) for continuous covariates [Wood and Wood (2015)]. For BART, we used the bart function from the BayesTree package in R with default settings [Chipman and McCulloch (2010)]. Other R packages implementing BART are available, including the bartMachine, dbarts, and bartCause packages. The linear regression/probit regression models were fit with the lm and glm functions in R. For semi-BART we used 10,000 MCMC iterations including 2500 burn-in iterations and $m = 50$ trees. Data-generating code is available in the Appendix [Zeldow, Lo Re III and Roy (2019)] and also available on https://www.github.com/zeldow/semibart-extras.

4.1. *Scenario* 1: *Continuous outcome with binary treatment and no effect modification.* In the first scenario, we generated data with a continuous outcome, binary treatment, twenty continuous covariates with a block diagonal covariance structure, and four independent binary covariates. The data generating code is available in the Appendix. The outcome was generated from an independent normal variable with variance one and mean $\mu(a, x) = h(a, x; \psi) + \omega(x)$ where $h(a, x; \psi) = \psi_1 a$ and $\omega(x) = 1 + 2x_1 + \sin(\pi x_2 x_{21}) - 2\exp(x_{22}x_{24}) + \log|\cos(\frac{\pi}{2}x_3)| - 1.8\cos(x_4) + 3x_{22}|x_2|^{1.5}$. The parameter ψ_1, which encodes the treatment effect, was set to 2.

The results in Table 1 show that, at the smaller sample size of $n = 500$, all point estimates are slightly biased in the same direction, and the 95% coverage probabilities hovered around 95%. Notably, the ESD for the BART-based methods was over half as small than for GAM or regression. At $n = 5000$ the bias disappeared, and the discrepancy in ESD between the BART-based methods and non-BART methods remained.

4.2. *Scenario* 2: *Continuous outcome with binary treatment and continuous effect modifier.* We randomly generated 30 continuous covariates with mean zero from a multivariate normal distribution with an autoregressive(1) covariance Σ

TABLE 1

Results from simulation study (scenario 1) with no effect modifiers. Bias: mean absolute bias across 500 datasets. Cov: Confidence/credible interval covarage (percent of simulations where the true value falls within the 95% interval). ESD: Empirical standard deviation defined as the standard deviation of the 500 estimates

Method	Parameter	Bias	Cov.	ESD
			$n = 500$	
Semi-BART	ψ_1	−0.02	0.96	0.153
GAM	ψ_1	−0.02	0.94	0.371
BART	ψ_1	−0.02	0.94	0.153
Regression	ψ_1	−0.02	0.95	0.390
			$n = 5000$	
Semi-BART	ψ_1	0.00	0.95	0.036
GAM	ψ_1	0.00	0.94	0.111
BART	ψ_1	0.00	0.92	0.037
Regression	ψ_1	0.01	0.94	0.119

with $\rho = 0.5$ with the diagonal containing ones. That is,

$$\Sigma = \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & \cdots \\ \rho & 1 & \rho & \rho^2 & \cdots \\ \rho^2 & \rho & 1 & \rho & \cdots \\ \rho^3 & \rho^2 & \rho & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

4.2.1. *Part a: Simple treatment mechanism and nonlinear mean function.* Given the covariates $x_1 - x_{30}$, the treatment was generated from a Bernoulli distribution with probability $p_a = \text{logit}^{-1}(0.1 + 0.2x_1 - \sin(x_3)/3 - 0.1x_{22})$. The outcome was generated as independent random normal variables with variance one and mean $\mu(a, x) = h(a, x; \psi) + \omega(x)$ where $h(a, x) = \psi_1 a + \psi_2 a * x_1 + \psi_3 x_1$ and $\omega(x) = 1 + \sin(\pi x_6 x_{21}) - \exp(x_4 x_5/5) + \log|\cos(\frac{\pi}{2}x_7)| - 1.8\cos(x_8) + 0.2x_{10}|x_6|^{1.5}$. The true values for the parameters are $\psi_1 = 2$, $\psi_2 = -1$, and $\psi_3 = 2$.

Results for these simulations are shown in Table 2. The estimated parameters are unbiased and have coverage near 95% for both sample sizes and all estimators. The ESD for all parameters is smaller with semi-BART than it is with GAM or linear regression. This improvement of semi-BART over GAM comes from the fact that covariate interactions are detected in the semi-BART procedure, whereas they must be prespecified in this implementation of GAM. Note that Hill's BART method—part of the simulations in Table 1—was not included since it does not summarize continuous effect modifiers with a single parameter.

TABLE 2

Results from simulation study (scenario 2a) for continuous outcomes with a simple treatment assignment mechanism, a complex outcome process, and a continuous effect modifier. Bias: mean absolute bias across 500 datasets. Cov: Confidence/credible interval covarage (percent of simulations where the true value falls within the 95% interval). ESD: Empirical standard deviation defined as the standard deviation of the 500 estimates. The true parameters are $\psi_1 = 2$, $\psi_2 = -1$, and $\psi_3 = 2$

Method	Parameter	Bias	Cov.	ESD
			$n = 500$	
Semi-BART	ψ_1	−0.01	0.94	0.123
	ψ_2	0.01	0.94	0.121
	ψ_3	0.00	0.96	0.095
GAM	ψ_1	−0.01	0.93	0.135
	ψ_2	0.01	0.94	0.127
	ψ_3	0.00	0.93	0.102
Regression	ψ_1	−0.01	0.94	0.166
	ψ_2	0.01	0.94	0.167
	ψ_3	0.00	0.94	0.127
			$n = 5000$	
Semi-BART	ψ_1	0.00	0.95	0.034
	ψ_2	0.00	0.94	0.033
	ψ_3	0.00	0.96	0.023
GAM	ψ_1	0.00	0.94	0.038
	ψ_2	0.00	0.94	0.039
	ψ_3	0.00	0.96	0.031
Regression	ψ_1	0.00	0.95	0.049
	ψ_2	0.00	0.95	0.049
	ψ_3	0.00	0.95	0.038

4.2.2. *Part b*: *Complex treatment mechanism and complex mean function*. We also performed these simulations with different treatment and outcome data generating functions. Here, given the covariates $x_1 - x_{30}$, the treatment was generated from a Bernoulli distribution with probability $p_a = \text{logit}^{-1}(0.1 + 0.2x_1 - 0.5x_2 - 0.1x_1x_2 + 0.3x_4 + 0.1x_5 + 0.7x_4x_5 - 0.4x_{11}x22 - 0.4x_{10}^2x_{15})$. The outcome was generated from an independent normal distribution with variance one and mean $\mu(a, x) = h(a, x; \psi) + \omega(x)$ where $h(a, x) = \psi_1 a + \psi_2 a * x_1 + \psi_3 x_1$ and $\omega(x) = 1 - x_2 + 2x_3 - 1.5x_4 - 0.5x_5 - 2x_6 + x_3^2 - x_6^2 + 2x_3x_4 - x_2x_6 + 0.5x_5x_6 - 0.2x_2x_3x_4 + x_6x_8x_9 - x_7x_{21}x_{24}x_{25} + x_{10}x_{13}x_{14}x_{26} - x_{24}x_{25}^2x_{10} + 3x_3x_{16}^2 - 3x_4x_{17}^2 + x_3x_4x_9x_{14} - x_3x_4x_9x_{14}^2 + 1.5x_{10}x_{21}$. The true values for the parameters are $\psi_1 = 2$, $\psi_2 = -1$, and $\psi_3 = 2$.

Results for these simulations are shown in Table 3. At $n = 500$, semi-BART yielded biased estimates (average of −0.07) for ψ_3, the main effect of the effect

TABLE 3

Results from simulation study (scenario 2b) for continuous outcomes with a complex treatment assignment mechanism, a complex outcome process, and a continuous effect modifier. Bias: mean absolute bias across 500 datasets. Cov: Confidence/credible interval covarage (percent of simulations where the true value falls within the 95% interval). ESD: Empirical standard deviation defined as the standard deviation of the 500 estimates. The true parameters are $\psi_1 = 2$, $\psi_2 = -1$, and $\psi_3 = 2$

Method	Parameter	Bias	Cov.	ESD
			$n = 500$	
Semi-BART	ψ_1	0.00	0.92	0.460
	ψ_2	−0.01	0.92	0.459
	ψ_3	−0.12	0.90	0.361
GAM	ψ_1	−0.05	0.95	0.654
	ψ_2	−0.02	0.92	0.672
	ψ_3	0.00	0.93	0.520
Regression	ψ_1	0.36	0.92	0.731
	ψ_2	−0.03	0.95	0.741
	ψ_3	−0.01	0.95	0.582
			$n = 5000$	
Semi-BART	ψ_1	0.00	0.96	0.081
	ψ_2	0.00	0.93	0.090
	ψ_3	−0.03	0.89	0.071
GAM	ψ_1	−0.07	0.94	0.213
	ψ_2	0.00	0.95	0.200
	ψ_3	0.00	0.97	0.151
Regression	ψ_1	0.39	0.60	0.230
	ψ_2	−0.01	0.95	0.220
	ψ_3	−0.01	0.96	0.166

modifier. On the other hand, GAM and linear regression were unbiased for ψ_3 but had varying degrees of bias for the treatment effect ψ_1 of -0.05 and 0.36, respectively. Semi-BART had slight undercoverage for all parameters—90% to 92%. At $n = 5000$, semi-BART was unbiased for ψ_1 and ψ_2, and the bias of ψ_3 attenuated (-0.12 down to -0.03). For GAM and linear regression, the bias of ψ_1 persisted. Coverage rates were all around 95% save for ψ_3 using semi-BART, which was at 89%. The ESD was notably smaller for semi-BART than the competitors.

4.3. *Scenario* 3: *Binary outcome with binary treatment and continuous effect modifier.* As in scenario 2, we randomly generated 30 continuous covariates with mean zero from a multivariate normal distribution with an autoregressive(1) covariance structure with $\rho = 0.5$ and the diagonal containing ones. The treatment was generated from a Bernoulli distribution with probability $p_a = \text{logit}^{-1}(0.1 + 0.2x_1 - \sin(x_3)/3 - 0.1x_{22})$. The outcome was gener-

TABLE 4

Results from simulation study (scenario 3) for binary outcomes. Bias: mean absolute bias across 500 datasets. Cov: Confidence/credible interval covarage (percent of simulations where the true value falls within the 95% interval). ESD: Empirical standard deviation defined as the standard deviation of the 500 estimates. The true parameter values are $\psi_1 = 0.3$, $\psi_2 = -0.1$, *and* $\psi_3 = 0.1$

Method	Parameter	Bias	Cov.	ESD
			$n = 500$	
Semi-BART	ψ_1	0.03	0.92	0.144
	ψ_2	0.00	0.94	0.140
	ψ_3	0.00	0.93	0.106
Regression	ψ_1	−0.01	0.93	0.131
	ψ_2	0.01	0.94	0.127
	ψ_3	−0.01	0.94	0.101
			$n = 5000$	
Semi-BART	ψ_1	0.00	0.94	0.039
	ψ_2	0.00	0.95	0.039
	ψ_3	0.00	0.94	0.029
Regression	ψ_1	−0.03	0.84	0.038
	ψ_2	0.01	0.93	0.036
	ψ_3	−0.01	0.93	0.029

ated from Bernoulli distribution with probability $p_y(a, x) = \phi[h(a, x; \psi) + \omega(x)]$ with $h(a, x; \psi) = \psi_1 a + \psi_2 a x_1 + \psi_3 x_1$ and $\omega(x) = 0.1 - \sin(\pi x_6 x_{21}/4) + \exp(x_6/5)x_{11}/4 - 0.12x_8 x_9 x_{21} + 0.05x_7 x_9 x_{10}^2$. The true values for the parameters of interest are $\psi_1 = 0.3$, $\psi_2 = -0.1$, and $\psi_3 = 0.1$.

The results for these simulations are shown in Table 4. For semi-BART, there is some bias on ψ_1 at $n = 500$, but this vanishes at $n = 5000$. Overall, bias is small and coverage good for both probit regression and semi-BART. Using probit regression is slightly more efficient than semi-BART at $n = 500$ (based on ESD) but these differences mostly disappear at $n = 5000$.

4.4. *Scenario* 4: *Misspecified models*. Lastly, we examine the properties of semi-BART when the model is misspecified in two ways: (1) we incorrectly incorporate a covariate into the linear part when it should be handled with BART and (2) we simulate data with a heavy tailed distribution (Student's t distribution with 2 degrees of freedom) when all our estimation methods assume a normal error term.

4.5. *Misspecified linear term*. For the first misspecified model with the incorrect linear part, we randomly generated 30 continuous covariates with mean zero from a multivariate normal distribution with an autoregressive(1) covariance structure. The treatment was generated from a Bernoulli distribution with probability $p_a = \text{logit}^{-1}(0.1 + 0.2x_1 - \sin(x_3)/3 - 0.1x_{22})$. The outcome was generated

TABLE 5

Results from simulation study (scenario 4) for misspecified models in which the linear part contains too many covariates. We show results for the treatment effect, which is known by design. Bias: mean absolute bias across 500 datasets. Cov: Confidence/credible interval covarage (percent of simulations where the true value falls within the 95% interval). ESD: Empirical standard deviation defined as the standard deviation of the 500 estimates

Method	Parameter	Bias	Cov.	ESD
			$n = 500$	
Semi-BART	ψ_1	0.01	0.95	0.143
GAM	ψ_1	0.00	0.92	0.153
Regression	ψ_1	0.01	0.96	0.177
			$n = 5000$	
Semi-BART	ψ_1	0.00	0.92	0.041
GAM	ψ_1	0.00	0.93	0.048
Regression	ψ_1	0.00	0.95	0.060

from a normal distribution with variance one and mean $\mu(a, x) = h(a, x; \psi) + \omega(x)$ with $h(a, x; \psi) = \psi_1 a$ and $\omega(x) = 1 + \sin(\pi x_6 x_{21}) - \exp(x_4 x_5/5) + \log|\cos(\pi x_7/2)| - 1.8\cos(x_8) + 0.2x_{10}|x_6|^{1.5} + x_1 x_2 - 0.5x_1^2 - \cos(x_1)$. However, we posited the relationship $h(a, x; \psi) = \psi_1 a + \psi_2 a x_1 + \psi_3 x_1$. Since the effect of x_1 is actually contained in $\omega(x)$, this is a misspecified model. The true value of ψ_1 was 2.

The results for these simulations are shown in Table 5. All methods have no bias and good coverage for ψ_1. There is a slight improvement in terms of ESD for semi-BART compared to its competitors.

4.6. *Misspecified error term.* For the final set of simulations, we used the the same data-generating mechanism as in scenario 2, save for the error term which was generated from a t-distribution with 2 degrees of freedom. Results are displayed in Table 6. The point estimates remain unbiased, including those from semi-BART. As we saw in the other simulations, semi-BART has lower ESD for all point estimates.

5. Data application. To illustrate our method, we analyzed data from the Veterans Aging Cohort Study (VACS) in the years 2002 to 2009, which is a cohort of patients being treated at Veterans Affairs facilities in the United States. Our study sample consisted of patients with HIV/Hepatitis C coinfection who were newly initiating antiretrovirals (including at least one nucleoside reverse transcriptase inhibitor [NRTI]) and had at least six months of observations recorded in VACS prior to initiation. Certain NRTIs are known to cause mitochrondial toxicity. These mitochrondial toxic NRTIs (mtNRTIs) include didanosine, stavudine, zidovudine, and zalcitabine [Soriano et al. (2008)]. While these drugs are no longer part of first

TABLE 6

Results from simulation study (scenario 4) for misspecified models in which the error term is generated from a heavy-tailed t-distribution with 2 degrees of freedom. Bias: mean absolute bias across 500 datasets. Cov: Confidence/credible interval covarage (percent of simulations where the true value falls within the 95% interval). ESD: Empirical standard deviation defined as the standard deviation of the 500 estimates

Method	Parameter	Bias	Cov.	ESD
			$n = 500$	
Semi-BART	ψ_1	−0.01	0.94	0.221
	ψ_2	−0.01	0.95	0.211
	ψ_3	0.02	0.96	0.157
GAM	ψ_1	0.00	0.94	0.352
	ψ_2	−0.02	0.93	0.349
	ψ_3	0.02	0.94	0.277
Regression	ψ_1	0.00	0.95	0.363
	ψ_2	−0.01	0.95	0.357
	ψ_3	0.02	0.96	0.281
			$n = 5000$	
Semi-BART	ψ_1	0.00	0.98	0.062
	ψ_2	−0.02	0.97	0.059
	ψ_3	0.00	0.95	0.050
GAM	ψ_1	0.00	0.95	0.102
	ψ_2	−0.01	0.95	0.103
	ψ_3	0.01	0.94	0.088
Regression	ψ_1	0.00	0.96	0.104
	ψ_2	−0.01	0.96	0.107
	ψ_3	0.01	0.94	0.092

line HIV treatment regimens, they are still used in resource-limited settings or in salvage regimens [Günthard et al. (2016)].

Exposure to mtNRTIs may increase the risk of hepatic injury which in turn may increase the risk of hepatic decompensation and death [Scourfield et al. (2011)]. The goal of this analysis was to determine if initiating an antiretroviral regimen containing a mtNRTI increased the risk of death within two years of first treatment versus an antiretroviral regimen containing a NRTI that is not a mtNRTI. VACS data contains a number of variables which possibly confound the relationship between mtNRTI use and death including subject demographics, year of antiretroviral initiation, HIV characteristics such as CD4 count and HIV viral load, concomitant medications, and laboratory measures relating to liver function.

In a previous analysis, we analyzed the effect that cumulative exposure to mtNRTI had on risk of death and liver decompensation [Lo Re et al. (2017)]. Data were organized longitudinally in month long increments with an indicator of exposure to mtNRTI within that month and another variable indicating the number of

months cumulatively exposed to mtNRTIs. Using Cox marginal structural models [Robins, Hernan and Brumback (2000)], we calculated hazard ratios for current exposure and cumulative exposure (no exposure [reference], <1 year, 1–3 years, 3–6 years, >6 years) to mtNRTIs. We found evidence of increased risk of death/liver decompensation as cumulative exposure to mtNRTIs increased with estimated hazard ratios of >3 for the highest categories of cumulative exposure compared to no exposure.

One of the covariates included in our analysis is Fibrosis-4 (FIB-4), an index that measures hepatitic fibrosis with higher values indicating larger injury. Specifically, FIB-4 > 3.25 (no units) indicates advanced hepatic fibrosis. FIB-4 is calculated as [Sterling et al. (2006)]:

$$\big[\text{age(years)} \times \text{AST(U/L)}\big]/\big[\text{platelet count}\big(10^9/\text{L}\big) \times \sqrt{\text{ALT(U/L)}}\big].$$

Here, AST stands for aspartate aminotransferase and ALT for alanine aminotransferase. There is some concern in that mtNRTI use in subjects with high FIB-4 will result in higher risk of liver decompensation and death than in subjects who have lower FIB-4. Thus, we consider FIB-4 as a possible effect modifier of the effect of mtNRTIs on death.

The outcome is a binary indicator of death within a two-year period after the subject initiated antiretroviral therapy. We considered only baseline covariates for this analysis. There were some missing values among the predictors that were handled through a single imputation. Our previous work on this data used multiple imputation to handle missing covariates but found that results were very similar across imputations [Lo Re et al. (2017)]. All continuous covariates were centered at interpretable values. For example, age was centered around 50 years and year of study entry was centered at 2005.

In the first analysis we sought to determine the effect of mtNRTI use on death without considering effect modification, and to this extent we fit a Bayesian SMM with a probit link. The estimand can be written as

$$\Phi^{-1}\big\{E\big(Y^a|\mathbf{X}=\mathbf{x}, A=a\big)\big\} - \Phi^{-1}\big\{E\big(Y^0|\mathbf{X}=\mathbf{x}, A=a\big)\big\} = \psi a, \tag{5.1}$$

where Y is the indicator of death, A represents whether mtNRTIs were part of the antiretroviral regimen at baseline ($A = 1$ if an mtNRTI was included in the regimen), and $\mathbf{X}$ all other covariates, including FIB-4. In the second and third analysis, we considered FIB-4 to be an effect modifier, once as a continuous covariate and once as a binary indicator which equaled 1 whenever FIB-4 > 3.25. This estimand can be written as

$$\Phi^{-1}\big\{E\big(Y^a|\mathbf{X}=\mathbf{x}, A=a\big)\big\} - \Phi^{-1}\big\{E\big(Y^0|\mathbf{X}=\mathbf{x}, A=a\big)\big\} = \psi_1 a + \psi_2 a x_1, \tag{5.2}$$

where x_1 corresponds to the appropriate FIB-4 variable.

The analysis was conducted using $m = 50$ trees with 20,000 total iterations (5000 burn-in). The prior distribution on the ψ parameters were independent

Normal$(0, 4^2)$. In the first analysis the mean estimate of the posterior distribution for ψ was 0.15 (95% credible interval [CI]: -0.02, 0.33). Notably the interval includes 0, but the point estimate indicates that subjects initiating antiretroviral therapy with an mtNRTI had greater risk of death within two years than subjects initiating therapy without an mtNRTI. We can interpret this coefficient in terms of $E(Y^0|\mathbf{X} = \mathbf{x}, A = a)$ and $E(Y^a|\mathbf{X} = \mathbf{x}, A = a)$ from equation (5.1). Figure 2a shows the value of $E(Y^1|\mathbf{X} = \mathbf{x}, A = 1)$ as a function of $E(Y^0|\mathbf{X} = \mathbf{x}, A = 1)$ for $\psi = 0.15$. As an example, suppose the unknowable quantity $E(Y^0|\mathbf{X} = \mathbf{x}, A = 1) = 0.20$. This means that subjects treated with a mtNRTI ($A = 1$) with covariates $\mathbf{X} = \mathbf{x}$ would have had a probability of death of 20% within two years had they been untreated ($A = 0$). However, given $\psi = 0.15$ we see that if $E(Y^0|\mathbf{X} = \mathbf{x}, A = 1) = 0.20$ then $E(Y^1|\mathbf{X} = \mathbf{x}, A = 1) = 0.24$, an increase of 4%. One can examine the change in probability for other baseline probabilities $E(Y^0|\mathbf{X} = \mathbf{x}, A = 1)$ by examining the graph in Figure 2a.

We conducted a second analysis with FIB-4 as a continuous effect modifier (centered around 3.25) with the same settings as the previous one. This analysis corresponds to the contrast from equation (5.2). Here, the estimate for the main effect of mtNRTI was $\psi_1 = 0.18$ (0.00, 0.36) and the interaction between mtNRTI use and FIB-4 was $\psi_2 = 0.07$ (0.02, 0.12). The results can be viewed in Figure 2b. Again, for illustration, consider the special case where $E(Y^0|\mathbf{X} = \mathbf{x}, A = 1) = 0.20$. When FIB-4 is 3.25, then $E(Y^1|\mathbf{X} = \mathbf{x}, A = 1) = 0.25$. However, when FIB-4 is 5.25, $E(Y^1|\mathbf{X} = \mathbf{x}, A = 1) = 0.30$.

Finally, we did a third analysis with FIB-4 as a binary effect modifier (>3.25 vs. ≤ 3.25). Here we found that $\psi_1 = 0.07$ (-0.12, 0.26) and $\psi_2 = 0.38$ (0.07, 0.69). These results can be viewed in Figure 2c. Here, we see that if $E(Y^0|\mathbf{X} = \mathbf{x}, A = 1) = 0.20$, then $E(Y^1|\mathbf{X} = \mathbf{x}, A = 1) = 0.22$ for subjects with FIB-4 ≤ 3.25 and $E(Y^1|\mathbf{X} = \mathbf{x}, A = 1) = 0.35$ for subjects with FIB-4 > 3.25.

A summary of all estimates, alongside comparisons those derived from probit regression, is available in Table 7. Runtime for semi-BART in the third analysis was 8.3 minutes on Ubuntu 18.10 with an Intel Core i7 processor @ 2.70 GHz with 8GB RAM. Analysis code is provided on https://www.github.com/zeldow/semibart-extras and in the supplementary files [Zeldow, Lo Re III and Roy (2019)].

6. Discussion. In this paper we presented semi-BART, a new Bayesian semiparametric model, alongside an R package for its implementation that is available on our GitHub repository (https://github.com/zeldow/semibart). To demonstrate our method, we chose a dataset of individuals co-infected with HIV and HCV who newly initiated an antiretroviral regimen. Our aim was to quantify the effect that mtNRTI use and FIB-4 had on two-year death. We found that subjects with higher values of FIB-4, which indicates a greater degree of liver damage, had increased risk of death within two years compared to counterparts with lower values of FIB-4 when their antiretroviral regimen contained an mtNRTI.

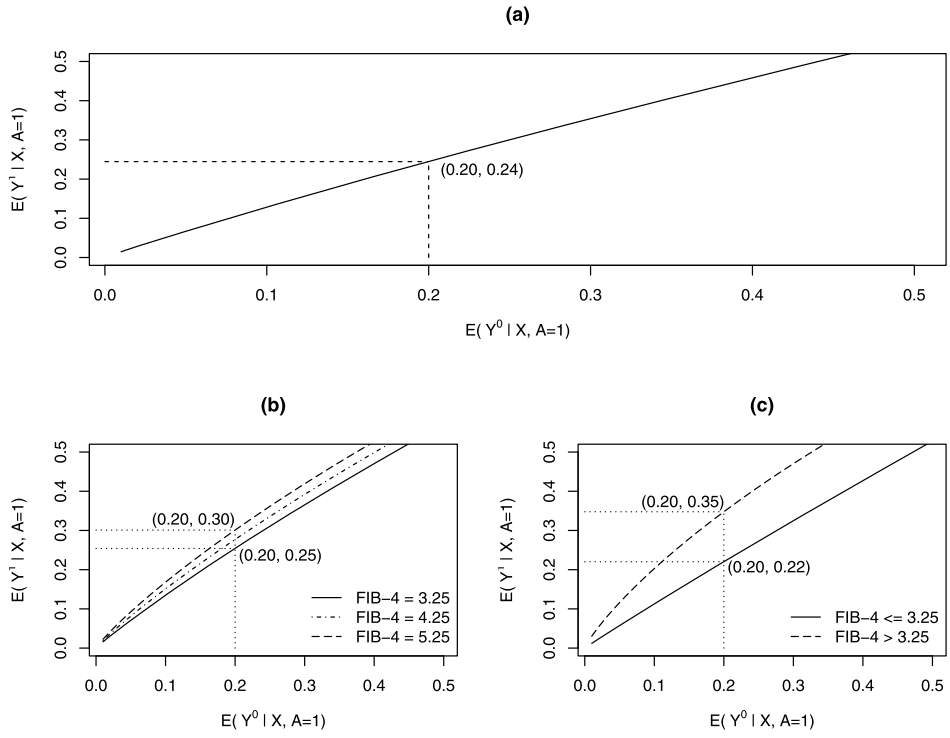

FIG. 2. *Results of data application showing the effect of having an mtNRTI in an antiretroviral regimen on two-year death.* $A = 1$ *indicates receipt of an mtNRTI and* $A = 0$ *indicates no receipt of an mtNRTI. The* x*-axis represents* $E(Y^0|\mathbf{X}, A = 1)$ *which is the mean probability of death if the treated* $A = 1$ *had in fact been untreated* $A = 0$ *given* $\mathbf{X}$*. This quantity is unknown so we consider a spectrum of reasonable values. The* y*-axis represents* $E(Y^1|\mathbf{X}, A = 1)$ *and gives the effect of treatment* A *on death relative to the* x*-axis.* (*a*) *We show effect of mtNRTI on death with no effect modifiers. If* $E(Y^0|\mathbf{X}, A = 1) = 0.20$ *then* $E(Y^1|\mathbf{X}, A = 1) = 0.24$*. For other values of* $E(Y^0|\mathbf{X}, A = 1)$*, identify the value on the* x*-axis, draw a vertical line until it hits the causal curve, then draw a horizontal line from that point to the* y*-axis.* (*b*) *We consider the effect modification of mtNRTI on death by continuous FIB*-4*. Assuming* $E(Y^0|\mathbf{X} = \mathbf{x}, A = 1) = 0.20$*, treatment increases the causal risk of death to* 25% *for subjects with FIB*-4 $= 3.25$ (*solid line*)*. For subjects with FIB*-4 $= 4.25$ (*dotted–dashed line*)*, the causal risk of death increases to* 27%*. The causal risk of death for individuals with FIB*-4 $= 5.25$ (*dashed line*) *is* 30%*.* (*c*) *We consider the effect modification of mtNRTI on death by a dichotomized FIB*-4*. The solid line indicates the causal effect curve when FIB*-4 ≤ 3.25*. Assuming* $E(Y^0|\mathbf{X} = \mathbf{x}, A = 1) = 0.20$*, we find that treatment increases the mean risk to* 22%*. The mean risk of death for individuals with high FIB*-4 > 3.25 (*dashed line*) *is even higher at* 35%*.*

Semi-BART allows for flexible estimation of the nuisance component while being parametric for covariates that are relevant to the research question (treatment, effect modifiers, etc.), providing a viable and intuitive alternative to parametric regression. Because of this, we are able to obtain low-dimensional summaries of effect modification with a BART-based method. Under some causal as-

TABLE 7

Comparison of point estimates 95% *confidence/credible intervals from our data analysis using semi-BART and probit regression. The outcome is a binary indicator of death.* ψ_1 *is the parameter for the treatment* (*mtNRTI use*) *effect and* ψ_2 *is the parameter for the interaction between mtNRTI use and FIB-*4 (*binary or continuous*)

Analysis	Parameter	Semi-BART	Probit regression
No effect modifier	$\hat{\psi}_1$	0.15 (−0.02, 0.33)	0.18 (0.01, 0.35)
Continuous effect modifier	$\hat{\psi}_1$	0.18 (0.00, 0.36)	0.20 (0.03, 0.37)
	$\hat{\psi}_2$	0.07 (0.02, 0.12)	0.06 (0.01, 0.11)
Binary effect modifier	$\hat{\psi}_1$	0.07 (−0.12, 0.26)	0.10 (−0.08, 0.29)
	$\hat{\psi}_2$	0.38 (0.07, 0.69)	0.34 (0.04, 0.64)

sumptions, this model can be interpreted as a SMM, which also provides the first Bayesian SMM. This is particularly useful in the case of binary outcomes where g-estimation is not possible. Vansteelandt and Goetghebeur (2003) provided approaches for estimating SMMs with binary outcomes with frequentist procedures; our method is consistent with their suggestions but incorporates the flexibility of BART. In the simulations we performed, we saw that semi-BART does particularly well with continuous outcomes where the true generating model has many nonlinearities.

Our method has some limitations we want to point out for any potential users. While we provide a framework that is agnostic to the functional form of nuisance covariates, we impose a more rigid model on the key scientific covariates. It may well be the case that linear parameterizations are not appropriate for all data applications so we must take care in using semi-BART, just as we would in using a standard linear model. On the other hand, the linear predictor in semi-BART is entirely user-specified and can easily accomodate squared terms, basis expansions, or any number of flexible reparameterizations.

In our simulations with binary outcomes, we found little difference in our estimates using semi-BART versus probit regression. Although it is reassuring that semi-BART works as well as parametric regression, we want to better understand the reasons why we are seeing equivalent—rather than superior—performance of semi-BART for binary outcomes when outcome processes are complex. Futhermore, our R package only supports a multivariate normal prior for the regression parameters and an inverse-chi square distribution for the error variance. Users wanting other types of priors would need to modify the code directly. In the future, we also want to extend the semi-BART R package to handle other common link functions such as logit or log.

From a causal inference perspective, another limitation of our method in the causal setting is that semi-BART does not currently accommodate instrumental

variables or longitudinal treatment measures, which are frequently used components of structural nested models. Furthermore, semi-BART is not doubly robust. Double robustness is an attractive feature because it gives the analyst two chances to get unbiased causal estimates. Typically, we have to either correctly specify either the outcome process or the treatment process. In semi-BART there is no modeling of the treatment process so it is the outcome model that must be correctly specified. There has been research on Bayesian double robust estimation [Saarela, Belzile and Stephens (2016)] and groundwork has been laid for double robust estimation of structural nested models (Vansteelandt and Joffe ((2014), see Section 6.1)), but more work is needed before semi-BART has double robustness properties.

We feel that semi-BART can benefit researchers across many discplines. In particular, we hope that semi-BART can be a viable alternative to the researcher who uses linear regression as the default statistical method in their research. Secondly, we also hope that researchers who prefer flexible machine learning algorithms, such as BART, but need interpretable coefficients such as a treatment effect and its modifiers that semi-BART is a dependable option.

Acknowledgments. We would like to thank Edward Kennedy, Alisa Stephens-Shields and Laura Hatfield for reviewing and giving comments on this manuscript. We are also indebted to the authors of the BayesTree package (Robert McCulloch and Hugh Chipman) for writing, maintaining and allowing us to modify their code to suit our purposes. Lastly, we'd like to thank the anonymous reviewers for insightful and helpful comments which ultimately made this paper better.

SUPPLEMENTARY MATERIAL

Supplement A: R code for semi-BART manuscript (DOI: 10.1214/19-AOAS1266SUPP; .zip). The supplement contains R code for the simulations, analysis code for our data application, and R code for some additional simulations performed.

REFERENCES

ALBERT, J. H. and CHIB, S. (1993). Bayesian analysis of binary and polychotomous response data. *J. Amer. Statist. Assoc.* **88** 669–679. MR1224394

BILLER, C. (2000). Adaptive Bayesian regression splines in semiparametric generalized linear models. *J. Comput. Graph. Statist.* **9** 122–140. MR1819868

BILLER, C. and FAHRMEIR, L. (2001). Bayesian varying-coefficient models using adaptive regression splines. *Stat. Model.* **1** 195–211.

BREZGER, A. and LANG, S. (2006). Generalized structured additive regression based on Bayesian P-splines. *Comput. Statist. Data Anal.* **50** 967–991. MR2210741

CENTERS FOR DISEASE CONTROL AND PREVENTION (2017). HIV and viral hepatitis. South Carolina State Documents Depository.

CHAMBERLAIN, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions. *J. Econometrics* **34** 305–334. MR0888070

CHIPMAN, H. A., GEORGE, E. I. and MCCULLOCH, R. E. (1998). Bayesian CART model search. *J. Amer. Statist. Assoc.* **93** 935–948.

CHIPMAN, H. A., GEORGE, E. I. and MCCULLOCH, R. E. (2010). BART: Bayesian additive regression trees. *Ann. Appl. Stat.* **4** 266–298. MR2758172

CHIPMAN, H. and MCCULLOCH, R. (2010). BayesTree: Bayesian methods for tree based models. R package version 0.3-1.1. Available at http://CRAN.R-project.org/package=BayesTree.

DENISON, D. G. T., MALLICK, B. K. and SMITH, A. F. M. (1998a). Automatic Bayesian curve fitting. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **60** 333–350. MR1616029

DENISON, D. G., MALLICK, B. K. and SMITH, A. F. (1998b). Bayesian mars. *Stat. Comput.* **8** 337–346.

EILERS, P. H. C. and MARX, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statist. Sci.* **11** 89–121. MR1435485

FRIEDMAN, J. H. (1991). Multivariate adaptive regression splines. *Ann. Statist.* **19** 1–67. MR1091842

GELMAN, A., CARLIN, J. B., STERN, H. S., DUNSON, D. B., VEHTARI, A. and RUBIN, D. B. (2013). *Bayesian Data Analysis*, 3rd ed. *Texts in Statistical Science Series*. CRC Press/CRC, Boca Raton, FL. MR2027492

GREEN, D. P. and KERN, H. L. (2012). Modeling heterogeneous treatment effects in survey experiments with Bayesian Additive Regression Trees. *Public Opin. Q.* **76** 491–511.

GÜNTHARD, H. F., SAAG, M. S., BENSON, C. A., DEL RIO, C., ERON, J. J., GALLANT, J. E., HOY, J. F., MUGAVERO, M. J., SAX, P. E. et al. (2016). Antiretroviral drugs for treatment and prevention of HIV infection in adults: 2016 recommendations of the International Antiviral Society—USA panel. *J. Amer. Medical Assoc.* **316** 191–210.

HAHN, P. R., MURRAY, J. S. and CARVALHO, C. M. (2018). Bayesian regression tree models for causal inference: Regularization, confounding, and heterogeneous effects. Available at arXiv:1706.09523.

HASTIE, T. J. and TIBSHIRANI, R. J. (1990). *Generalized Additive Models*. CRC Press, London. MR1082147

HASTIE, T. and TIBSHIRANI, R. (2000). Bayesian backfitting. *Statist. Sci.* **15** 196–223. MR1820768

HILL, J. L. (2011). Bayesian nonparametric modeling for causal inference. *J. Comput. Graph. Statist.* **20** 217–240. MR2816546

HOLMES, C. C. and MALLICK, B. K. (2001). Bayesian regression with multivariate linear splines. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **63** 3–17. MR1811987

LO RE, V., ZELDOW, B., KALLAN, M. J., TATE, J. P., CARBONARI, D. M., HENNESSY, S., KOSTMAN, J. R., LIM, J. K., GOETZ, M. B. et al. (2017). Risk of liver decompensation with cumulative use of mitochondrial toxic nucleoside analogues in HIV/hepatitis C virus coinfection. *Pharmacoepidemiol. Drug Saf.* **26** 1172–1181.

MÜLLER, P., QUINTANA, F. A., JARA, A. and HANSON, T. (2015). *Bayesian Nonparametric Data Analysis*. *Springer Series in Statistics*. Springer, Cham. MR3309338

NATIONAL INSTITUTES OF HEALTH (2018). Panel on antiretroviral guidelines for adults and adolescents. Guidelines for the Use of Antiretroviral Agents in Adults and Adolescents Living with HIV, Dept. Health and Human Services. Available at http://aidsinfo.nih.gov/contentfiles/lvguidelines/AdultandAdolescentGL.pdf. Accessed: 2019-03-01.

RASMUSSEN, C. E. (2006). *Gaussian Processes for Machine Learning*. MIT Press, Cambridge, MA.

ROBINS, J. M. (1994). Correcting for non-compliance in randomized trials using structural nested mean models. *Comm. Statist. Theory Methods* **23** 2379–2412. MR1293185

ROBINS, J. M. (2000). Marginal structural models versus structural nested models as tools for causal inference. In *Statistical Models in Epidemiology, the Environment, and Clinical Trials* (*Minneapolis, MN*, 1997). *IMA Vol. Math. Appl.* **116** 95–133. Springer, New York. MR1731682

ROBINS, J. M., HERNAN, M. A. and BRUMBACK, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology* 550–560.

SAARELA, O., BELZILE, L. R. and STEPHENS, D. A. (2016). A Bayesian view of doubly robust causal inference. *Biometrika* **103** 667–681. MR3551791

SCOURFIELD, A., JACKSON, A., WATERS, L., GAZZARD, B. and NELSON, M. (2011). The value of screening HIV-infected individuals for didanosine-related liver disease? *Antivir. Ther.* **16** 941–942.

SORIANO, V., PUOTI, M., GARCIA-GASCÓ, P., ROCKSTROH, J. K., BENHAMOU, Y., BARREIRO, P. and MCGOVERN, B. (2008). Antiretroviral drugs and liver injury. *AIDS* **22** 1–13.

STERLING, R. K., LISSEN, E., CLUMECK, N., SOLA, R., CORREA, M. C., MONTANER, J., SULKOWSKI, M. S., TORRIANI, F. J., DIETERICH, D. T. et al. (2006). Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV coinfection. *Hepatology* **43** 1317–1325.

VANSTEELANDT, S. and GOETGHEBEUR, E. (2003). Causal inference with generalized structural mean models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **65** 817–835. MR2017872

VANSTEELANDT, S. and JOFFE, M. (2014). Structural nested models and G-estimation: The partially realized promise. *Statist. Sci.* **29** 707–731. MR3300367

VAN DER LAAN, M. J. and ROSE, S. (2011). *Targeted Learning: Causal Inference for Observational and Experimental Data. Springer Series in Statistics*. Springer, New York. MR2867111

WOOD, S. and WOOD, M. S. (2015). Package 'mgcv.' R package Version 1-7. Available at http://CRAN.R-project.org/package=mgcv.

ZELDOW, B., LO RE III, V. and ROY, J. (2019). Supplement to "A semiparametric modeling approach using Bayesian Additive Regression Trees with an application to evaluate heterogeneous treatment effects." DOI:10.1214/19-AOAS1266SUPP.

B. ZELDOW
DEPARTMENT OF HEALTH CARE POLICY
HARVARD MEDICAL SCHOOL
180 LONGWOOD AVE
BOSTON, MASSACHUSETTS 02115
USA
E-MAIL: zeldow@pm.me

V. LO RE III
DEPARTMENT OF MEDICINE
PERELMAN SCHOOL OF MEDICINE
UNIVERSITY OF PENNSYLVANIA
PHILADELPHIA, PENNSYLVANIA 19104
USA

J. ROY
DEPARTMENT OF BIOSTATISTICS AND EPIDEMIOLOGY
RUTGERS SCHOOL OF PUBLIC HEALTH
PISCATAWAY, NEW JERSEY 08854
USA