HHS Public Access

Author manuscript

Biometrics. Author manuscript; available in PMC 2017 December 28.

Published in final edited form as:

Biometrics. 2017 December ; 73(4): 1111–1122. doi:10.1111/biom.12679.

Outcome-adaptive lasso: variable selection for causal inference

Susan M Shortreed and

Biostatistics Unit, Group Health Research Institute, Department of Biostatistics, University of Washington, shortreed.s@ghc.org

Ashkan Ertefaie

Department of Biostatistics and Computational Biology, University of Rochester Department of Statistics, The Wharton School, University of Pennsylvania Center for Pharmacoepidemiology Research and Training, University of Pennsylvania, ertefaie@wharton.upenn.edu

Summary

Methodological advancements, including propensity score methods, have resulted in improved unbiased estimation of treatment effects from observational data. Traditionally, a “throw in the kitchen sink” approach has been used to select covariates for inclusion into the propensity score, but recent work shows including unnecessary covariates can impact both the bias and statistical efficiency of propensity score estimators. In particular, the inclusion of covariates that impact exposure but not the outcome, can inflate standard errors without improving bias, while the inclusion of covariates associated with the outcome but unrelated to exposure can improve precision. We propose the outcome-adaptive lasso for selecting appropriate covariates for inclusion in propensity score models to account for confounding bias and maintaining statistical efficiency. This proposed approach can perform variable selection in the presence of a large number of spurious covariates, i.e. covariates unrelated to outcome or exposure. We present theoretical and simulation results indicating that the outcome-adaptive lasso selects the propensity score model that includes all true confounders and predictors of outcome, while excluding other covariates. We illustrate covariate selection using the outcome-adaptive lasso, including comparison to alternative approaches, using simulated data and in a survey of patients using opioid therapy to manage chronic pain.

Keywords

Comparative effectiveness; Model selection; Observational studies; Propensity score

1. Introduction

Methodological advances have improved analytic methods for constructing unbiased treatment effect estimates from observational data (Rubin, 1973, 1974; Robins, 1986; Hernan and Robins, 2006). In particular, propensity score (PS) methods that rely on the PS

Web-Based Supplementary Materials

Web Appendices and Figures referenced in Sections 3, 3.1, 4.1, 4.4, and 4.5 are available with this paper at the Biometrics website on Wiley Online Library, as well as example R code implementing the outcome-adaptive lasso (reference in Section 3.1).

score, defined as the probability of treatment given covariates, are increasingly popular as an effective approach to control for confounding (Rosenbaum and Rubin, 1983). Regardless of the method used for causal inference from observational data, an important assumption is that all confounders of the relationship between treatment and the outcome of interest are measured and included in the PS model. Since excluding important confounders can lead to biased treatment estimates, a "throw in the kitchen sink" mentality has previously been used when including covariates in PS models. While incorporating all confounders is important for unbiased treatment effect estimates, recent work has shown that efficiency losses can accompany the inclusion of extraneous variables (Greenland, 2008; Schisterman et al., 2009; Rotnitzky et al., 2010; Myers et al., 2011; Patrick et al., 2011).

De Luna et al. (2011) and Patrick et al. (2011) highlight the variance inflation caused by including variables associated with exposure (note, 'exposure' and 'treatment' are used interchangeably throughout), but not the outcome and show this may additionally cause bias. Simulation studies in Brookhart et al. (2006) suggest that inclusion of variables related only to the outcome may lead to efficiency gains. This implies that an efficient variable selection method should take in to account both outcome-covariate and treatment-covariate relationships. Wang et al. (2012) proposed Bayesian adjustment for confounding which incorporates the outcome-covariate association via a tuning parameter (Zigler et al., 2013). Wilson and Reich (2014) presented a confounder selection approach that fits a Bayesian regression model and uses the posterior credible region of the regression parameters to form a set of candidate models. The final model is then found by penalizing models that do not include confounders. This method is conservative and often includes treatment predictors that may inflate the variance of the treatment effect (Lin et al., 2015), and tuning the penalty function can be challenging. Zigler and Dominici (2014) propose a Bayesian approach for selecting variables as well as averaging over several possible PS models that may include different sets of covariates. The problem of variable selection in causal inference is also discussed in Robins and Greenland (1986), Judkins et al. (2007), Schneeweiss et al. (2009), Vansteelandt et al. (2010), Van der Laan and Gruber (2010), Rolling and Yang (2013), and Talbot et al. (2015a).

Many recommend using data-driven variable selection methods to build a PS for unbiased and statistically efficient causal inference. There is a vast literature on variable selection methods for prediction, but relatively little work on variable selection for causal inference. We specifically focus on the difference between variable selection for prediction and variable selection for causal inference (i.e. unbiased treatment effect estimation in the presence of confounders). We propose a novel PS variable selection approach, the outcome-adaptive lasso, which is specifically designed for causal inference. Unlike most existing methods, our approach can be used in high-dimensional settings and aims to select covariates to produce an unbiased and statistically efficient PS estimator. We provide evidence based on simulations that the proposed method is efficient in the sense of Brookhart et al. (2006). We describe the problem and approach using the inverse probability of treatment weighted (IPTW) estimator (Hernan and Robins, 2006), although our method is applicable to any PS methodology.

In Section 2, we outline a lasso-based shrinkage method in the context of variable selection for prediction. In Section 2.2 we provide a short overview of causal inference and PS analyses. We then discuss how the variable selection goal for causal inference differs from the variable selection problem for prediction and describe the proposed outcome-adaptive lasso in Section 3. We present results of illustrations assessing the empirical properties of the outcome-adaptive lasso and compare it to alternative approaches using artificial data in Section 4. We also present the results of applying these approaches to assess the effect of long-term opioid use on depressive symptoms. We close with a discussion in Section 5.

2. Preliminaries

The fundamental goal of prediction is to accurately identify a model that predicts the outcome of interest. For many reasons including estimation, statistical efficiency, and interpretation, a parsimonious prediction model is often advantageous. Consider Y a continuous-valued outcome and d predictors, denoted X_j for j=1:d, measured prior to treatment status, A, which is measured prior to the outcome. Lowercase letters refer to possible values of corresponding capital letter random variables. In the prediction setting we would like to identify the $d_0 < d$ covariates that are predictors of the outcome and estimate corresponding coefficients to accurately predict the outcome using this reduced model. Without loss of generality, we assume all covariates have mean 0 with a common standard deviation. The design matrix $\mathbf{X} = (X_1, X_2, .., X_d)$ is defined such that it may include higher order and interaction terms.

2.1 *Adaptive lasso*

The adaptive lasso is an extension of the traditional lasso (Tibshirani, 1996) that uses coefficient specific weights (Zou, 2006). Under certain conditions, Zou (2006) showed the adaptive lasso estimator satisfies the oracle property. Simply put, the oracle property means the adaptive-lasso estimator selects the non-zero coefficients with probability tending to one (i.e., sparsity) and non-zero components are also estimated as if the true (sparse) model were known a priori (i.e., asymptotic normality) (Fan and Li, 2001).

Let $\ell_n(\beta; Y, X)$ be the negative log-likelihood parametrized by β for a sample of size n. The adaptive lasso estimator is defined as:

$$\hat{\beta}(AL)=\underset{\beta}{\operatorname{argmin}}\left\{\ell_n(\beta;Y,\mathrm{X})+\lambda_n\sum_{j=1}^{d}\hat{\omega}_j|\beta_j|\right\}, \quad (1)$$

where $\hat{\omega}=\left|\widetilde{\beta}_j\right|^{-\gamma}$ such that $\gamma > 0$, λ_n is a regularization parameter, and $\widetilde{\beta}=\operatorname{argmin}_\beta \ell_n(\beta;Y,X)$, i.e., $\hat{\beta}$ is the unpenalized maximum likelihood estimate of β.

2.2 *Causal Inference*

A primary goal of causal inference is to construct unbiased estimators of treatment effects. We focus on the average treatment effect (ATE): $E(Y_{a=1} - Y_{a=0})$ (Rosenbaum and Rubin,

1983). $Y_{a=1}$ denotes the counterfactual outcome under treatment; $Y_{a=0}$ the counterfactual outcome under no treatment (Rosenbaum and Rubin, 1983; Pearl, 2000). Because when certain assumptions hold the PS is a balancing score (Rosenbaum and Rubin, 1983), PS methods are commonly used to estimate treatment effects from observational data. A balancing score is a function of covariates such that conditioned on a particular value of the balancing score individuals who are treated and untreated can be directly compared for unbiased estimation of treatment effects (Rosenbaum and Rubin, 1983).

Four assumptions are required to ensure unbiased treatment effect estimation from observational data. Consistency, requires $Y_{ia} = (Y_i | A_i = a)$, and the stable unit value treatment assumption means an individual's treatment status does not affect another's counterfactual outcome. The positivity assumption states $0 < P(A = 1|\mathbf{X}) < 1$, i.e. no combination of covariates has everyone with those covariate values either treated or untreated. The fourth assumption is known as no unmeasured confounding (or ignorability), which can be written as $A \perp Y_a \mid \mathbf{X}$. That is, conditioned on the set of measured covariates, treatment status is independent of the counterfactual outcome. The last two assumptions are the most important for variable selection in causal inference. Methods must select covariates to ensure the no unmeasured confounding assumption is met, yet exclude covariates that may lead to unnecessary violations (or near violations) of the positivity assumption.

While the PS is defined as the probability of treatment given covariates, its analytic goal is to eliminate bias due to confounding rather than produce a prediction model for the treatment assignment mechanism. Traditionally, the main concern when analyzing observational data was avoiding bias. As larger amounts of data on individuals become available, especially with increased use of electronic health records, variable selection for causal inference becomes more important. Our work was motivated by recent literature discussed in Section 1 on the loss of statistical efficiency that can result from including variables that predict exposure, but are unrelated to the outcome in the PS (Schisterman et al., 2009; Rotnitzky et al., 2010; Myers et al., 2011), and the statistical efficiency gains possible by including outcome predictors not related to treatment (Brookhart et al., 2006; Rotnitzky et al., 2010).

3. Outcome-adaptive lasso

Let $\mathscr{C}$ denote indices of covariates associated with both outcome and exposure, and $\mathscr{P}$ indices of predictors of outcome, but not exposure. We use $\mathscr{I}$ to denote indices of covariates that predict exposure, but are unrelated to outcome, and $\mathscr{S}$ to denote indices of covariates that are unrelated to both exposure and outcome. The ideal PS includes all confounders ($\mathrm{X}_{\mathscr{C}}$) to avoid bias and all predictors ($\mathrm{X}_{\mathscr{P}}$) to increase statistical efficiency, while excluding predictors of exposure not associated with outcome ($\mathrm{X}_{\mathscr{I}}$) and spurious variables ($\mathrm{X}_{\mathscr{S}}$).

Specifically, we use variable selection methods to estimate the following PS:

$$logit\{\pi(\mathrm{X},\hat{\alpha})\} = logit\{P(A=1|\mathrm{X},\hat{\alpha})\} = \sum_{j\in\mathscr{C}} \hat{\alpha}_j X_j + \sum_{j\in\mathscr{P}} \hat{\alpha}_j X_j, \quad (2)$$

where $\hat{\alpha}$ is an estimate of α^*, defined as the population-level parameters in the reduced PS model (2). We refer to (2) as a reduced model because it does not include $X_{\mathscr{I}}$ or $X_{\mathscr{S}}$. Although the coefficients corresponding to $X_{\mathscr{P}}$ are equal to zero in the true treatment generating process, inclusion of these variables in the PS model reduces finite sample bias caused by random confounding and decreases the variance of the estimator. In fact, projecting the treatment onto the space of outcome predictors increases the correlation between the outcome and the PS reducing the variance of the estimator. Standard prediction variable selection techniques applied to the PS model set the coefficients corresponding to $X_{\mathscr{P}}$ to zero. The main challenge is to design a penalty function that imposes heavier penalty on variables that predict only treatment than variables that predict only the outcome.

The outcome-adaptive lasso selects covariates and estimates corresponding coefficients by incorporating information about the outcome-covariate relationships when selecting variables. Assuming a logit model for the PS, the outcome-adaptive lasso estimates are:

$$\hat{\alpha}(OAL)=\underset{\beta}{\operatorname{argmin}}\left[\sum_{i=1}^{n}\left\{-a_i(\mathrm{x_i}^T\alpha)+log(1+e^{x_i^T\alpha})\right\}+\lambda_n\sum_{j=1}^{d}\hat{\omega}_j|\alpha_j|\right]. \quad (3)$$

where $\hat{\omega}_j=\left|\widetilde{\beta}_j\right|^{-\gamma}$ s.t. $\gamma>1$ and $(\widetilde{\beta},\widetilde{\eta})=\operatorname{argmin}_{\beta,\eta}\ell_n(\beta,\eta;Y,\mathrm{X},A)$. We use $\widetilde{\beta}$ to refer to the coefficient estimates of the relationship between the covariates and the outcome conditional on treatment, with $\widetilde{\eta}$ the coefficient estimate corresponding to treatment.

Denote the covariates to include in the estimated PS, $\mathscr{A}=\mathscr{C}\cup\mathscr{P}$, and the covariates to be excluded, $\mathscr{A}^c-\mathscr{I}\cup\mathscr{S}$. Without loss of generality, we assume the indices are ordered such that $\mathscr{A}=\{j:j\in\mathscr{C}\cup\mathscr{P}\}=\{1,2,\ldots,d_0\}$ with $d_0<d=|\mathscr{C}|+|\mathscr{P}|+|\mathscr{I}|+|\mathscr{S}|$. We then write the Fisher Information matrix: $\mathbf{I}(\alpha^*)=\begin{pmatrix}\mathbf{I}_{11}&\mathbf{I}_{12}\\\mathbf{I}_{21}&\mathbf{I}_{22}\end{pmatrix}$, where $\mathbf{I}_{11}$ is a matrix of size d_0 by d_0, and is the Fisher Information matrix for the parsimonious PS. Theorem 1 states the asymptotic behavior of the proposed outcome-adaptive lasso estimates, $\hat{\alpha}(OAL)$.

Theorem 1

Suppose $\lambda_n/\sqrt{n}\to 0$ *and* $\lambda_n n^{\gamma/2-1}\to\infty$, for $\gamma>1$; then, under mild regularity conditions, the outcome-adaptive lasso estimates $\hat{\alpha}(OAL)$ *satisfy the following:*

(1) *Consistency in variable selection:* $lim_n P\{\hat{\alpha}_j(OAL)=0|j\in\mathscr{I}\cup\mathscr{S}\}=1$.

(2) *Asymptotic normality:* $\sqrt{n}\left\{\hat{\alpha}(OAL)-\alpha^*_{\mathscr{A}}\right\}\to_d N(0,\mathbf{I}_{11}^{-1})$.

The first part of Theorem 1 states the proposed method forces coefficients corresponding to variables that predict exposure but not outcome and spurious variables to zero. The asymptotic normality result ensures the penalized estimators corresponding to confounders and outcome predictors have behavior similar to the maximum likelihood estimators when the targeted, sparse PS is known a priori. In particular, because $\hat{\omega}_j=O_p(1)$ for $j\in\mathscr{P}$, the

proposed approach does not set coefficients corresponding to outcome predictors to zero. In fact, as shown in the proof (see Web-Based Supplementary Material), these coefficients would be set to zero if $\lambda_n/\sqrt{n} \to \infty$, a stronger condition than required for Theorem 1.

Note, the original adaptive lasso requires $\lambda_n n^{(\gamma-1)/2} \to \infty$ to satisfy the sparsity property (Zou, 2006). However, our proposed method requires a heavier penalty: $\lambda_n n^{\gamma/2-1} \to \infty$. This is because we want to exclude variables that predict exposure, but are unrelated to outcome from the PS model. See Supplementary Materials for the proof of Theorem 1.

3.1 *Selecting λ_n*

Selecting the tuning parameters is important for practical application. The IPTW estimator of the ATE uses the PS to balance the covariate distributions between the exposure groups. We propose selecting λ_n by minimizing a weighted absolute mean difference (wAMD) between the exposure groups:

$$wAMD(\lambda_n) = \sum_{j=1}^{d} |\widetilde{\beta}| \left| \frac{\sum_{i=1}^{n} \hat{\tau}_i^{\lambda_n} X_{ij} A_i}{\sum_{i=1}^{n} \hat{\tau}_i^{\lambda_n} A_i} - \frac{\sum_{i=1}^{n} \hat{\tau}_i^{\lambda_n} X_{ij}(1-A_i)}{\sum_{i=1}^{n} \hat{\tau}_i^{\lambda_n}(1-A_i)} \right|,$$

where $\hat{\tau}^{\lambda_n}$ are the IPTWs constructed using the fitted PS model with variables selected using the outcome-adaptive lasso method with λ_n, denoted $\hat{\pi}^{\lambda_n}(.)$. The $\hat{\tau}^{\lambda_n}$ are defined:

$$\hat{\tau}_i^{\lambda_n} = \frac{A_i}{\hat{\pi}_i^{\lambda_n}\{\mathrm{X}_i, \hat{\alpha}(OAL)\}} + \frac{1-A_i}{1-\hat{\pi}_i^{\lambda_n}\{\mathrm{X}_i, \hat{\alpha}(OAL)\}}, \quad (4)$$

Recall, $\widetilde{\beta}_j$ are the unpenalized estimates of the relationship between the covariates and the outcome conditional on treatment. For large λ_n values, strong penalization of the likelihood occurs, potentially forcing the differences in the covariate means of exposure groups further apart. For those covariates with negligible association with outcome (i.e., small $\widetilde{\beta}_j$) this difference has negligible impact on the wAMD, while differences in means between exposure groups for those covariates with a strong association with outcome will be inflated by the magnitude of their association with the outcome. We include example R code implementing the outcome-adaptive lasso, including selecting λ_n, in Supplementary Materials.

4. Illustrating the outcome-adaptive lasso

We first present simulation results to demonstrate the performance of the outcome-adaptive lasso and compare its performance to alternative approaches. We then describe an analysis to evaluate the impact of long-term opioid therapy on depressive symptoms.

4.1 *Simulation design for generating artificial data*

We modeled simulations assessing the performance of the outcome-adaptive lasso learning approach after illustrations in Zigler and Dominici (2014). For each replicate data set, we simulated n individuals, indexed by i, with d covariates, $\mathbf{X}_i = (X_{i1}, X_{i2}, \ldots, X_{id})$ generated from a multivariate standard Gaussian distribution. We conducted simulations with three correlation structures: the first generated independent covariates; the second predictors that are moderately correlated ($\rho = 0.2$), and the third, predictors that are strongly correlated ($\rho = 0.5$). We generated a binary treatment, A, from a Bernoulli distribution with $logit\{P(A=1)\} = \sum_{j=1}^{d} \nu_j X_j$, and a continuous-valued outcome, Y, such that $Y_i = \eta A + \sum_{j=1}^{d} \beta_j X_j + \epsilon_i$, where $\epsilon_i \sim N(0, 1)$ and η=0 or 2. We varied d and n. To represent situations in which learning is necessary because the ratio of the number of covariates to the sample size is large, we let $n = 200$ with $d = 100$ and $n = 500$ with $d = 200$. To examine performance as the sample size increases, we fix $d = 20$ and let $n = 200$, 500, and 1000.

We used four data generating scenarios, varying the strength of the relationship between covariates and outcome and treatment generating processes. Due to space constraints we describe the parameter values for two of the four data-generating scenarios and describe all simulations (120 different simulation settings) and results in Supplementary Material. For all scenarios, d–6 coefficients were set to 0 in both models (i.e. d–6 covariates were spurious covariates), while the magnitude of coefficient values for the first six covariates were varied. In all scenarios, the first two covariates, X_1 and X_2, are true confounders, while the third and fourth covariates, X_3 and X_4, are in the outcome-generating model but not the exposure-generating model. The fifth and sixth covariates, X_5 and X_6, are in the treatment-generating model but not the outcome-generating model. The first scenario sets: β=(0.6, 0.6, 0.6, 0.6, 0, 0, 0, …, 0) and ν=(1, 1, 0, 0, 1, 1, 0, …, 0). The second scenario generates confounders with a weaker relationship with treatment, setting β=(0.6, 0.6, 0.6, 0.6, 0, 0, 0, …, 0) and ν=(0.4, 0.4, 0, 0, 1, 1, 0, …, 0). While no intercept was included in the data-generating process, we allow for an intercept term in estimation, such that $X_0 = 1$ for all i, and we do not allow ν_0 to shrink.

We search over several possible λ_n values: $\{n^{-10}, n^{-5}, n^{-1}, n^{-0.75}, n^{-0.5}, n^{-0.25}, n^{0.25}, n^{0.49}\}$, for each data set. We set γ such that $\lambda_n n^{\gamma/2-1} = n^2$ for each λ_n value. This ensures the necessary conditions of Theorem 1 hold and the convergence properties for each of the the λ_n values are equivalent. The λ_n is selected using wAMD (Section 3.1); we denote the outcome-adaptive lasso by `OAL`. To evaluate the impact of the γ convergence factor, we perform a smaller subset of simulations (scenario 1, with no correlation between covariates) varying γ such that $\lambda_n n^{\gamma/2-1} = n^{\{-1,-1/3,1/3,2,5\}}$.

We focus on estimating the ATE, defined as $\theta^* = E(Y_1) - E(Y_0)$. The true ATE is either zero or two in our illustrations, and we use the IPTW estimator (Lunceford and Davidian, 2004):

$$\hat{\theta}=\frac{\sum_{i=1}^{n}\hat{\tau}_i^{\lambda_n}Y_iA_i}{\sum_{i=1}^{n}\hat{\tau}_i^{\lambda_n}A_i}-\frac{\sum_{i=1}^{n}\hat{\tau}_i^{\lambda_n}Y_i(1-A_i)}{\sum_{i=1}^{n}\hat{\tau}_i^{\lambda_n}(1-A_i)}, \quad (5)$$

where $\hat{\tau}^{\lambda_n}$ is defined in (4). We plot the distribution of the ATE estimates using boxplots for simulations with both a large and modest number of covariates and plot the percentage of times each covariate is selected to be included in the PS model (tolerance = 10^{-8}) for illustrations with modest d. We used logistic regression for the PS model. We evaluate the performance of the outcome-adaptive approach and several comparison approaches.

4.2 Comparison variable selection approaches for illustrations using artificial data

We compare the outcome-adaptive lasso approach to variable selection methods designed for causal effect estimation and the traditional adaptive lasso variable selection technique on the PS (`AdL`) as described in Section 2.1 with the coefficient specific weights determined by the maximum likelihood estimates describing the relationship between the covariates and exposure. The λ_n parameter for `AdL` is selected by minimizing the wAMD with weights determined by the absolute values of the coefficients in the PS (as opposed to the outcome model as described in Eq. 4). We select from the same possible λ_n in Section 4.1 and set γ such that $\lambda_n n^{(\gamma-1)/2} = n^2$.

De Luna et al. (2011) propose variable selection for non-parametric estimation of the ATE. We implemented this approach using the `R` package, `CovSel` (Häggström and Persson, 2015; Häggström et al., 2015), using default options. The `CovSel` package offers two algorithms for determining the minimal confounding set, which we label `deLunaA1` (algorithm 1) and `deLunaA2` (algorithm 2) (Häggström et al., 2015). Talbot et al. (2015a) build on Bayesian variable selection approaches proposed by Wang et al. (2012) by formally incorporating the outcome model into their variable selection method. We used the function `ABCEE`, from the `R` package, `BCEE`, developed by the authors (Talbot et al., 2015b) and use all author-recommended defaults (Talbot et al., 2015b). Simulations performed in Talbot et al. (2015b) involved fewer covariates than the illustrations considered here.

The Wilson and Reich (2014) approach is described in Section 1; we used the `R` package `BayesPen` to implement this approach. We used forward selection, including in the PS the variable corresponding to the largest coefficient first, adding covariates one at a time, and stopping when a coefficient had a p-value greater than 0.25 as recommended by Wilson and Reich (2014). We refer to this method as `WR`. Lastly, we implemented a Bayesian model averaging approach (`BAC`) proposed by Wang et al. (2015) from the `R` library `BACR` obtained from `github`. The number of iterations was set to 5000, with a burnin of 500, and a thinning parameter of 10.

We present results for three PS models that include the same covariates across simulations. These fixed PS models were selected to illustrate the impact of covariate inclusion on the IPTW estimator. `Conf` includes only confounders, $X_{\mathscr{C}}$, while `PotConf` includes all *potential*

confounders (i.e., $X_{\mathscr{C}}$, $X_{\mathscr{P}}$, and $X_{\mathscr{I}}$). `Targ` includes only confounders and variables that predict the outcome but not treatment (i.e., $X_{\mathscr{C}}$ and $X_{\mathscr{P}}$). `Targ` should have the smallest standard error and is the model the outcome-adaptive variable selection approach is designed to discover.

4.3 *Results of illustrations using artificial data*

We present the results generating data with no correlation between covariates and with the true ATE equal to 0; remaining results are in Supplementary Material. `OAL` and `WR` methods were applied in all scenarios, while computational and convergence challenges prevented comparison of `BCEE`, `deLuna`, and `BAC` in scenarios with a large number of predictors. `deLuna` and `BCEE` did not reach convergence for large d; and the computational time required to implement `BAC` was prohibitive. The methods varied in computation time, for example a run on one data set generated under scenario 1 (no correlation) with $d = 20$ and $n = 1000$ `OAL` (using the `lqa` function in `R`) took 0.42 seconds, including selecting the tuning parameter. `WR` took 0.05 seconds, `deLuna` 13.92 seconds, `BCEE` 7.35 seconds, and `BAC` 67.61 seconds.

We graphically present IPTW ATE estimates for the 1000 simulations conducted under scenarios 1 and 2 with large d in Figure 1. Box plots sizes show the estimator with the smallest variability is the IPTW estimator that used the `Targ` PS model. As discussed in Section 2.2 this model provides a statistically efficient estimator, while the inclusion of variables related to exposure but not outcome in `PotConf` increases the variability. When $n = 200$ ($d = 100$) performance of the `OAL` estimator is between `Targ` and `PotCont`, but when the sample size increases to $n = 500$ ($d = 200$) `OAL` performs much like `Targ`. `WR` performs similarly to `OAL` for bias, but is slightly more variable.

For illustrations with $d = 20$, boxplots of IPTW estimates and the proportion of times each covariate was included in the PS model (tolerance = 10^{-8}) are in Figures 2 and 3, respectively. Between 15 and 20% of variables related to exposure but not outcome and spurious variables are included in the `OAL` PS model when $n = 200$. This is is halved when $n = 500$, and further decreased when $n = 1000$, as Theorem 1 suggests. Additionally, while these coefficients are greater than the tolerance (10^{-8}), the actual estimates are quite small with little impact on the IPTW estimator. `ABCEE` includes confounders and predictors regardless of sample size and excludes variables unrelated to outcome that predict exposure about 60% of the time and spurious variables under 5%. Both `deLunaA1` and `deLunaA2` (`deLunaA2` not shown) had similar patterns of covariate inclusion and exclusion, including all confounders and excluding spurious variables at similar rates. Both algorithms often excluded variables associated only with outcome (but not exposure) and included variables related to exposure but not outcome about 60% of the time. `BAC` selected all potential confounders, including variables that predict exposure but not outcome, and excluded spurious variables about 20% of the time when n=200, decreasing to about 10% when n=1000. `WR` performed similarly but excluded variables that predict exposure but not outcome about half the time.

4.4 *Long-term opioid therapy and depressive symptoms*

Some individuals with chronic pain choose to use opioid medication long-term to manage their pain. This therapy includes low-dose medications taken as needed or regular medication use at low or elevated levels. There is limited evidence that continued long-term opioid use, especially high-dose, may lead to increased depression symptoms (Merrill et al., 2012; Scherrer et al., 2014, 2016). We used data from the Middle-Aged/Seniors Chronic Opioid Therapy (MASCOT) longitudinal study (Turner et al., 2016; Von Korff et al., 2016) to address this question. At baseline survey, MASCOT participants filled at least three opioid prescriptions in the last 120 days, with enough medication for 60 days of use (Turner et al., 2016; Von Korff et al., 2016). For this analysis, we restrict our sample to confirmed initiators, i.e., patients with self-reported confirmation of no opioid use for 6 months prior to the first prescription in the last 120 days, or if opioids were used in the past 6 months, that period included at least one month of no use. This one-month period was sufficient to regard subsequent use as a new episode of use. The 8-item patient health questionnaire (PHQ8) was used to measure depressive symptoms at baseline and 4 months later. PHQ8 scores range from 0 to 24, a high score indicates more severe depressive symptoms (Kroenke et al., 2009).

We consider two exposure groups based on opioid use in the time between the baseline survey and the 4-month follow-up interview. Specifically, we compare individuals using opioids at a lower dose and/or intermittently (lower dose) and individuals using opioids regularly and/or at a higher dose (higher dose). We define exposure groups using both electronic pharmacy data and self-reported opioid use. All individuals included in this analysis reported using opioids at least twice a week in the last 28 days at the 4-month follow-up, but the average daily dose, measured in morphine equivalent dose (MED), varied. Individuals in the lower-dose group had an average daily dose in the last 4-months between 5 mg and 15 mg MED; the higher-dose group had average daily doses greater than or equal to 15 mg MED.

The MASCOT study collected a variety of information on patients. A total of 37 covariates were considered for inclusion in the PS model, including demographic, health status and opioid regimen information. The outcome-adaptive lasso was used to learn the PS model predicting opioid exposure using all measured covariates, and λ_n was selected from the same set considered in earlier illustrations using the wAMD. We constructed bootstrap confidence intervals (CIs) for the IPTW ATE estimate as described in Efron (2014). In particular, we used the smoothed non-parametric bootstrap approach to construct a 95% CI that takes into account the model selection procedure (Efron, 2014). We performed 10,000 bootstrap iterations and present the percent of times each covariate was selected for the PS model (tolerance = 10^{-8}). We present the results of applying the Wilson and Reich (2014) method using the same bootstrap procedure to construct a 95% CI. Results of the `BAC` method using 10,000 iterations, with a burnin of 100 and a thinning parameter of 100 are also presented; we report the 95% credible interval for the `BAC` estimate (Wang et al., 2015).

4.5 Evaluating relationship between long-term opioid therapy and depressive symptoms

Our analytic sample consists of 425 individuals with complete baseline data who were using opioids at the 4-month survey. Of 896, who were confirmed initiators at baseline, 73 were excluded for unknown opioid use at 4 months, 364 were excluded for no longer using opioids at the 4-month follow-up, and 34 were excluded for missing baseline information. The results of learning the PS model comparing lower- versus higher-dose long-term opioid use are in Tables 1 and 2. The inclusion probabilities in the `OAL` PS model were relatively high for most covariates, ranging from 29.0% to 100.0%. Covariates known to be strong predictors of depressive symptoms had large selection percentages, including sex, which was selected in 99.2% of the bootstrap replications, and baseline depressive symptoms, which was included 100.0% of the time. `WR` and `BAC` included fewer covariates, with the `WR` approach including only 3 variables more than half of the time. The `OAL` results are not surprising given the setting; based on scientific evidence all covariates collected in the MASCOT study were thought to be associated with opioid use, pain outcomes, or negative psychosocial outcomes.

A few covariates show interesting results in the `OAL` learned PS. For example, while the education and racial makeups of the lower and higher dose exposure categories differed, these covariates were excluded from the PS model about 30% of the time. This means that these covariates were not strong predictors of 4-month depression symptoms in several of the bootstrap replicated data sets. In contrast, while anxiety symptoms are quite balanced between exposure groups, the anxiety covariate is included in the PS 85% of the time, likely because baseline anxiety levels are strongly correlated with 4-month depressive symptoms.

Individuals in the lower-dose opioid groups had an unadjusted 4-month mean PHQ8 score of 5.93, with a standard deviation (sd) of 5.10. Those in the higher-dose group had a 4-month mean PHQ8 score of 6.79 (sd=5.70). We report summary information on the PS and estimated weights in Supplementary Material. The IPTW estimate for the effect of opioid dose on depressive symptoms when the outcome-adaptive lasso was used to select the variables for the PS was 0.13 with a 95% CI of (0.10, 0.17). The IPTW ATE estimate when `WR` was used to learn the PS was 0.41 with a 95% CI of (0.38,0.46) and when `BAC` was used it was 0.25 with a 95 % credible interval of (−0.50, 1.00). While the estimate for the relationship between higher-dose use and depressive symptoms with all methods was positive and CIs for both the `WR` and `OAL` approaches do not include zero, the suggested PHQ8 score increase is small and not clinically meaningful.

5. Discussion

We present a penalization variable selection technique specifically designed for estimating unbiased treatment effects from observational data. The proposed outcome-adaptive penalty function takes into account the association between covariates and the outcome as well as the association between covariates and the exposure. This unique feature allows our proposed approach to exclude variables that predict exposure but are not related to outcome and spurious variables from the PS. Simulation results show the proposed method

outperforms existing methods and provides empirical evidence that the estimated PS converges to the PS that includes only confounders and predictors of outcome.

The appropriateness of considering the outcome model when selecting covariates to account for confounding in observational studies has received much deserved attention. When scientists consider particular coefficient values or p-values, the meaning of the resulting estimates change and scientific integrity can be compromised. We have proposed a data-driven approach to variable selection that uses the outcome model to increase statistical efficiency and relies on objective measures to select the tuning parameter. The metric proposed to select the tuning parameter is different than standard strategies used in the prediction setting. When developing a sparse prediction model, the main goal is to produce high-quality predicted values for the outcome, usually with the purpose of applying the sparse model to new data for which the outcome is unknown. Thus, cross-validation or generalized cross-validation, using misclassification error in the left-out sample, are often used to select the tuning parameter value. In our setting, the goal is a high-quality estimate of the treatment effect using current data. Thus, standard criteria to select the tuning parameter such as the predictive performance of the treatment model do not optimize the desired property of the propensity score. Additionally, procedures such as cross-validation, which focus on out-of-sample performance are designed to optimize criteria in a different sample, may not have ideal characteristics for the causal inference setting (Häggström and de Luna, 2014). Regardless of the criteria for selecting the tuning parameter, in both predictive modeling and causal inference settings, considering variable selection when constructing p-values and CIs is important. We employed the bootstrap to construct CIs accounting for the selection procedure (Efron, 2014), but further work on the exact and asymptotic construction of CIs for the outcome-adaptive lasso could be a topic of future work (Leeb and Potscher, 2005, 2008).

We defined the penalty weights in the outcome-adaptive lasso using the MLEs relating covariates and outcome through a linear function $\beta\mathbf{X} + \eta A$; defining the design matrix to include higher order and interaction terms if necessary. When the postulated association form is drastically misspecified, the outcome-adaptive lasso may under-select important covariates. In such cases, defining the weights using different association measures of covariates and outcome would be an interesting topic for future research.

Sometimes treatment does not affect all individuals in the same way. When there is evidence of treatment effect heterogeneity, we may use $\ell_n(\beta, \eta; Y, \mathbf{X}, A) = \ell_\eta(Y, \beta\mathbf{X} + \eta A\mathbf{X})$ to model the covariate-outcome relationship and define the weights in the outcome-adaptive lasso as $\hat{w}_j = (|\widetilde{\beta}_j| + |\widetilde{\eta}_j|)^{-\gamma}$ where $(\widetilde{\beta}, \widetilde{\eta}) = \mathrm{argmin}_{\beta,\eta} \ell_n(Y, \beta\mathbf{X} + \eta A\mathbf{X})$. This imposes heavier penalties on covariates that are neither strong effect modifiers nor have strong main effects on the outcome.

We assumed the number of covariates d was fixed while n goes to infinity. One important expansion of this work is to develop extensions and study the asymptotic behavior of our approach when both the covariate dimension and the sample size can go to infinity. In this situation, the asymptotic properties of the estimators are drastically different, for example,

the hessian matrix of the negative log likelihood is singular (Negahban et al., 2009; Javanmard and Montanari, 2014; Van de Geer et al., 2014).

As access to large amounts of data increases, such as data from health systems' networks with electronic health records on millions of lives, we must develop objective data-driven approaches to variable selection that are specific to the scientific question of interest. Traditional variable selection approaches developed for prediction should be modified to improve bias and statistical efficiency in causal inference settings. The outcome-adaptive lasso takes into account the goal of unbiased treatment effect estimation from observational data to select variables that decrease bias, while increasing statistical efficiency.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

NIH grant 1RO1 AG034181 (PI:Michael Von Korff) from the National Institute on Aging supported MASCOT study data collection and provided support for Dr. Shortreed. Dr. Ertefaie was supported by NSF grant SES 1260782 from the National Science Foundation.

References

Brookhart M, Schneeweiss S, Rothman K, Glynn R, Avorn J, Sturmer T. Variable selection for propensity score models. American Journal of Epidemiology. 2006; 163:1149–1156. [PubMed: 16624967]

De Luna X, Waernbaum I, Richardson T. Covariate selection for the nonparametric estimation of an average treatment effect. Biometrika. 2011; 98:861–875.

Efron B. Estimation and accuracy after model selection. Journal of the American Statistical Association. 2014; 109:991–1007. [PubMed: 25346558]

Fan J, Li R. Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American Statistical Association. 2001; 96:1348–60.

Greenland S. Invited commentary: variable selection versus shrinkage in the control of multiple confounders. American Journal of Epidemiology. 2008; 167:523. [PubMed: 18227100]

Häggström J, de Luna X. Targeted smoothing parameter selection for estimating average causal effects. Computational Statistics. 2014; 29:1727–48.

Häggström, J., Persson, E. Package 'CovSel'. CRAN; 2015.

Häggström J, Persson E, Waernbaum I, de Luna X. CovSel: An R package for covariate selection when estimating average causal effects. Journal of Statistical Software. 2015; 68:1–20.

Hernan M, Robins J. Estimating causal effects in epidemiological data. Journal of Epidemiology and Community Health. 2006; 60:578–586. [PubMed: 16790829]

Javanmard A, Montanari A. Confidence intervals and hypothesis testing for high-dimensional regression. The Journal of Machine Learning Research. 2014; 15:2869–2909.

Judkins D, Morganstein D, Zador P, Piesse A, Barrett B, Mukhopadhyay P. Variable selection and raking in propensity scoring. Statistics in Medicine. 2007; 26:1022–33. [PubMed: 16708347]

Kroenke K, Strine T, Spitzer R, Williams J, Berry J, Mokdad A. The PHQ-8 as a measure of current depression in the general population. Journal of Affective Disorder. 2009; 114:163–73.

Leeb H, Potscher B. Model selection and inference: Facts and fiction. Econometric Theory. 2005; 21:21–59.

Leeb H, Potscher B. Sparse estimators and the oracle property, or the return of hodges estimator. Journal of Econometrics. 2008; 142:201–211.

Lin W, Feng R, Li H. Regularization methods for high-dimensional instrumental variables regression with an application to genetical genomics. Journal of the American Statistical Association. 2015; 110:270–288. [PubMed: 26392642]

Lunceford JK, Davidian M. Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. Statistics in Medicine. 2004; 23:2937–60. [PubMed: 15351954]

Merrill J, Von Korff M, Banta-Green C, Sullivan M, Saunders K, Campbell C, et al. Prescribed opioid difficulties, depression and opioid dose among chronic opioid therapy patients. General Hospital Psychiatry. 2012; 34:581–7. [PubMed: 22959422]

Myers J, Rassen J, Gagne J, Huybrechts K, Schneeweiss S, Rothman K, et al. Effects of adjusting for instrumental variables on bias and precision of effect estimates. American Journal of Epidemiology. 2011; 174:1213–22. [PubMed: 22025356]

Negahban S, Yu B, Wainwright M, Ravikumar P. A unified framework for high-dimensional analysis of *m*-estimators with decomposable regularizers. Advances in Neural Information Processing Systems. 2009:1348–1356.

Patrick A, Schneeweiss S, Brookhart M, Glynn R, Rothman K, Avorn J, et al. The implications of propensity score variable selection strategies in pharmacoepidemiology: An empirical illustration. Pharmacoepidemiology and Drug Safety. 2011; 20:551–9. [PubMed: 21394812]

Pearl, J. Causality. Cambridge University Press; 2000.

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

Robins J, Greenland S. The role of model selection in causal inference from nonexperimental data. American Journal of Epidemiology. 1986; 123:392–402. [PubMed: 3946386]

Rolling C, Yang Y. Model selection for estimating treatment effects. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2013; 76:749–69.

Rosenbaum P, Rubin D. The central role of the propensity score in observational studies for causal effects. Biometrika. 1983; 70:41–55.

Rotnitzky A, Li L, Li X. A note on overadjustment in inverse probability weighted estimation. Biometrika. 2010; 97:1–5.

Rubin D. The use of matched sampling and regression adjustment to remove bias in observational studies. Biometrics. 1973; 29:184–203.

Rubin D. Estimating causal effects of treatment in randomized and nonrandomized studies. Journal of Educational Psychology. 1974; 66:688–701.

Scherrer J, Salas J, Copeland L, Stock E, Abemedani B, Sullivan M, et al. Prescription opioid duration, dose and increased risk of depression in 3 large patient populations. The Annals of Family Medicine. 2016; 14:54–62. [PubMed: 26755784]

Scherrer J, Svrakic D, Freeland K, Chrusciel T, Balasubramanian S, Bucholz K, et al. Prescription opioid analgesics increase risk of depression. Journal of General Internal Medicine. 2014; 29:491–9. [PubMed: 24165926]

Schisterman E, Cole S, Platt R. Overadjustment bias and unnecessary adjustment in epidemiologic studies. Epidemiology. 2009; 20:488. [PubMed: 19525685]

Schneeweiss S, Rassen J, Glynn R, Avorn J, Mogun H, Brookhart M. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. Epidemiology. 2009; 20:512–22. [PubMed: 19487948]

Talbot D, Lefebvre G, Atherton J. The Bayesian causal effect estimation algorithm. The Journal of Causal Inference. 2015a; 3:207–36.

Talbot, D., Lefebvre, G., Atherton, J. Package ‘BCEE’. CRAN; 2015b.

Tibshirani R. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 1996; 58:267–88.

Turner J, Shortreed S, Saunders K, LeResche L, Von Korff M. Association of levels of opioid use with pain and activity interference among patients initiating chronic opioid therapy: A longitudinal study. PAIN. 2016; 154:849–57.

Van de Geer S, Buhlmann P, Ritov Y, Dezeure R. On asymptotically optimal confidence regions and tests for high-dimensional models. The Annals of Statistics. 2014; 42:1166–1202.

Van der Laan M, Gruber S. Collaborative double robust targeted maximum likelihood estimation. The International Journal of Biostatistics. 2010; 6:17.

Vansteelandt S, Bekaert M, Claeskens G. On model selection and model misspecification in causal inference. Statistical Methods in Medical Research. 2010:1477–0334.

Von Korff M, Dublin S, Walker R, Parchman M, Shortreed S, Hansen R, et al. The impact of opioid risk reduction initiatives on high-dose opioid prescribing for chronic opioid therapy patients. The Journal of Pain. 2016; 17:101–10. [PubMed: 26476264]

Wang C, Dominici F, Parmigiani G, Zigler C. Accounting for uncertainty in confounder and effect modifier selection when estimating average causal effects in generalized linear models. Biometrics. 2015; 71:654–65. [PubMed: 25899155]

Wang C, Parmigiani G, Dominici F. Bayesian effect estimation accounting for adjustment uncertainty. Biometrics. 2012; 68:661–671. [PubMed: 22364439]

Wilson A, Reich B. Confounder selection via penalized credible regions. Biometrics. 2014; 70:852–61. [PubMed: 25123966]

Zigler C, Dominici F. Uncertainty in propensity score estimation: Bayesian methods for variable selection and model averaged causal effects. Journal of the American Statistical Association. 2014; 109:95–107. [PubMed: 24696528]

Zigler C, Watts K, Yeh R, Wang Y, Coull B, Dominici F. Model feedback in bayesian propensity score estimation. Biometrics. 2013; 69:263–73. [PubMed: 23379793]

Zou H. The adaptive lasso and its oracle properties. Journal of the American Statistical Association. 2006; 101:1418–1429.

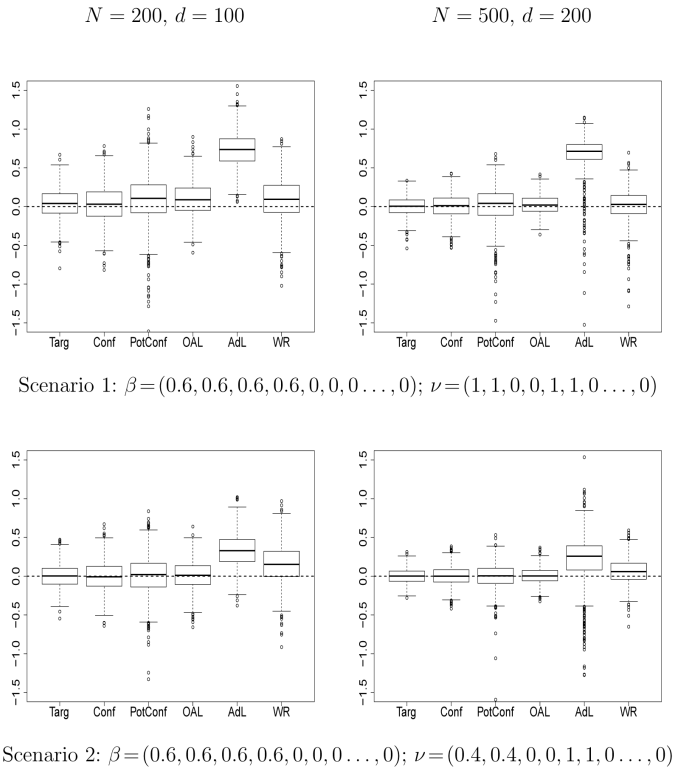

Figure 1.

Illustrations with large *d*. Box plots of 1000 inverse probability weighted estimates for the average treatment effect under scenarios 1 and 2. True treatment effect of 0 is indicated with dotted line.

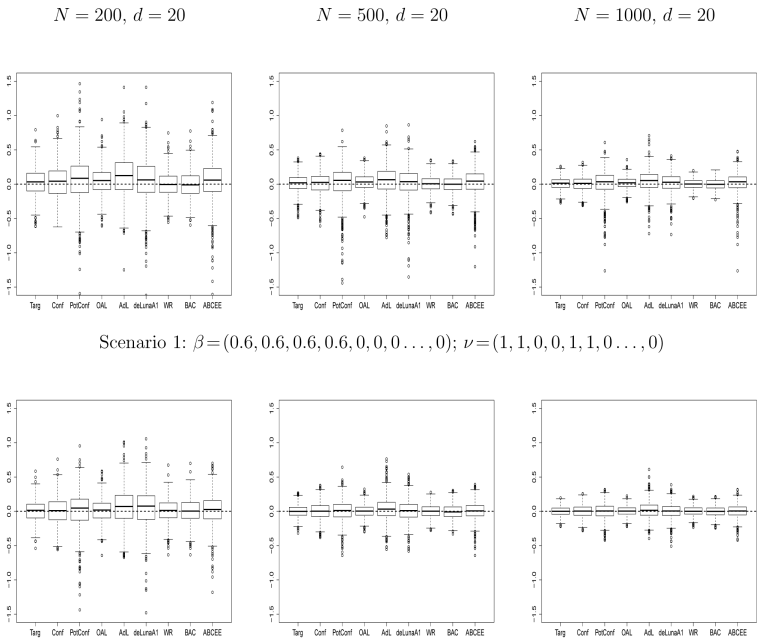

Figure 2.

Illustrations with modest *d*. Box plots of 1000 inverse probability weighted estimates for the average treatment effect under scenarios 1 and 2. True treatment effect of 0 is indicated with dotted line.

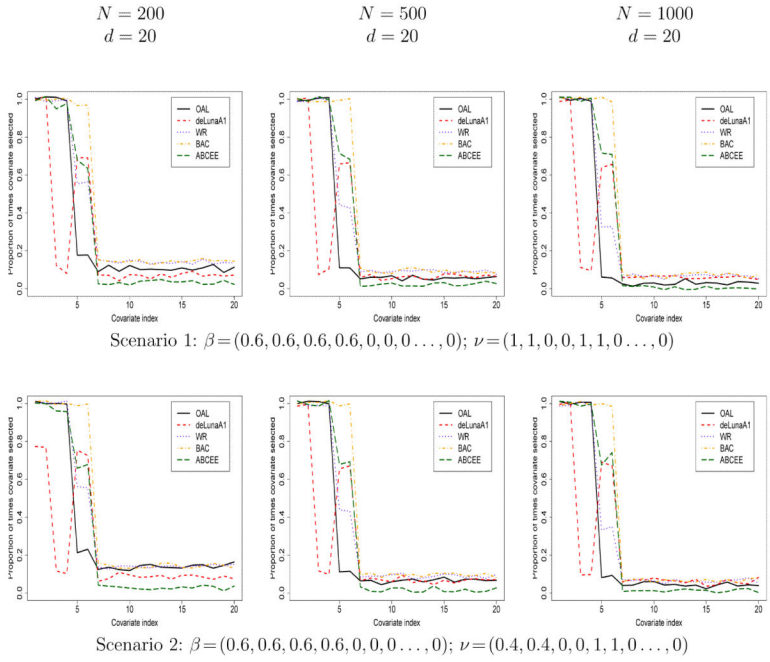

Figure 3.

Illustrations with modest *d*. Probability of covariate selection over 1000 simu lations for PS model under scenarios 1 and 2. `deLunaA1` and `deLunaA2` performed similarly, thus only `deLunaA1` shown.

Table 1

Covariate distribution by opioid use category, means (standard deviations) for continuous-valued and percentages (N) for categorical variable. Last three column reports percent of 10,000 bootstrap iterations covariate was selected for PS model inclusion for `OAL` *and* `WR` and posterior probability of being selecting in model for `BAC`.

	Opioid Use Category		% Selected		
Covariates	**Lower dose N=197**	**Higher dose N=228**	**OAL**	**WR**	**BAC**
Sex: Male	38.6 (64)	61.4 (102)	99.2	0.1	66.1
Age (years)	64.9 (11.7)	62.9 (10.3)	73.4	0.0	31.3
Education: HS graduate	44.5 (134)	55.5 (167)	64.5	0.0	7.4
Race: White, non-Hispanic	44.9 (160)	55.1 (196)	73.4	0.7	6.3
Unemployed due to disability	37.9 (25)	62.1 (41)	95.4	1.3	59.3
BMI (kg/m^2)	32.6 (8.1)	30.8 (7.6)	76.1	0.0	45.4
Charlson comorbidity score					
0	42.3 (80)	43.2 (96)	Ref	Ref	Ref
1	7.9 (15)	6.8 (15)	81.3	0.0	9.4
2	20.1 (38)	17.1 (38)	71.7	1.0	7.4
3+	29.6 (38)	32.9 (59)	97.3	0.0	59.3
Smoking use: current	37 (27)	63 (46)	79.8	0.1	21.3
Anxiety symptom severity scale	1.6 (1.8)	1.7 (1.8)	84.3	0.3	62.7
Depression symptom severity scale	7.1 (5.7)	8.2 (5.9)	100.0	79.1	1.0
Pain interference scale	6 (2.3)	6.4 (2)	34.0	0.0	5.1
Characteristic pain scale	6.5 (1.5)	6.6 (1.4)	29.0	0.0	5.1
Diffuse pain scale	5.9 (2.7)	5.9 (2.6)	32.4	0.0	7.9
Number of pain days in prior 6 months	144.5 (53.7)	143.4 (53.2)	34.0	0.0	7.1
Baseline average daily opioid dose					
Continuous-valued mg MED	9.7 (7.6)	19.2 (13.8)	80.3	0.3	1.0
Less than 20mg MED	89.4 (169)	64.0 (142)	Ref	Ref	Ref
Between 20mg and 40mg MED	9.5 (18)	29.3 (65)	81.8	0.0	49.5
Greater than 40mg	1.1 (2)	6.9 (15)	96.5	2.0	75.4
PODS Concern Sum Score	2.5 (3.6)	3.8 (3.9)	82.7	0.0	11.6

Table 2

Response to the prescribed opioids difficulties scale (PODS) items by opioid use category, percent (N). % Selected is percent of 10,000 bootstrap iterations item was selected for PS model inclusion for OAL *and* WR, and posterior probability of being selecting in model for BAC.

	Opioid Use Category		% Selected		
PODS items	**Lower dose N=197**	**Higher dose N=228**	**OAL**	**WR**	**BAC**
Preoccupied with opioids					
disagree	85.7 (162)	78.8 (175)	92.0	0.9	6.1
neutral	7.4 (14)	8.1 (18)	Ref	Ref	Ref
agree	6.9 (13)	13.1 (29)	87.3	0.1	11.1
Difficulty controlling opioids					
disagree	93.1 (176)	94.1 (209)	88.5	93.5	8.0
neutral	2.6 (5)	2.3 (5)	Ref	Ref	Ref
agree	4.2 (8)	3.6 (8)	98.5	41.1	16.9
Feels need higher dose of opioids					
disagree	79.4 (150)	71.2 (158)	79.0	3.0	7.3
neutral	5.3 (10)	6.8 (15)	Ref	Ref	Ref
agree	15.3 (29)	22.1 (49)	87.0	1.3	4.9
Worried about becoming addicted to opioids					
disagree	84.1 (159)	73.9 (164)	81.6	5.6	7.6
neutral	7.9 (15)	6.8 (15)	Ref	Ref	Ref
agree	7.9 (15)	19.4 (43)	91.4	1.9	5.8
Want to cut down or stop use of opioids					
disagree	57.1 (108)	50.5 (112)	93.3	0.4	14.1
neutral	14.3 (27)	11.7 (26)	Ref	Ref	Ref
agree	28.6 (54)	37.8 (84)	NA	NA	NA
Opioids have caused me to feel depressed, down or anxious					
disagree	85.2 (161)	79.7 (177)	73.7	0.2	4.2
neutral	8.5 (16)	9.9 (22)	Ref	Ref	Ref
agree	6.3 (12)	10.4 (23)	87.5	0.0	8.1
Opioids have caused me to lose interest in usual activities					
disagree	81.5 (154)	79.7 (177)	84.4	5.3	16.8
neutral	9.5 (18)	10.8 (24)	Ref	Ref	Ref
agree	9.0 (17)	9.5 (21)	85.4	0.2	10.2
Opioids have caused me to lose concentration					
disagree	82.0 (155)	77.5 (172)	71.9	0.4	7.5
neutral	8.5 (16)	9.0 (20)	Ref	Ref	Ref
agree	9.5 (18)	13.5 (30)	90.7	0.2	24.5
Opioids have caused me to feel slowed down or sluggishness					
disagree	66.7 (126)	62.2 (138)	68.9	0.1	2.4
neutral	11.6 (22)	9.9 (22)	Ref	Ref	Ref
agree	21.7 (41)	27.9 (62)	72.9	0.1	6.4