

Variable selection for causal mediation analysis using LASSO-based methods

Statistical Methods in Medical Research

2021, Vol. 30(6) 1413–1427

© The Author(s) 2021



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/0962280221997505

journals.sagepub.com/home/smmZhaoxin Ye¹, Yeying Zhu¹  and Donna L Coffman²

Abstract

Causal mediation effect estimates can be obtained from marginal structural models using inverse probability weighting with appropriate weights. In order to compute weights, treatment and mediator propensity score models need to be fitted first. If the covariates are high-dimensional, parsimonious propensity score models can be developed by regularization methods including LASSO and its variants. Furthermore, in a mediation setup, more efficient direct or indirect effect estimators can be obtained by using outcome-adaptive LASSO to select variables for propensity score models by incorporating the outcome information. A simulation study is conducted to assess how different regularization methods can affect the performance of estimated natural direct and indirect effect odds ratios. Our simulation results show that regularizing propensity score models by outcome-adaptive LASSO can improve the efficiency of the natural effect estimators and by optimizing balance in the covariates, bias can be reduced in most cases. The regularization methods are then applied to MIMIC-III database, an ICU database developed by MIT.

Keywords

Covariate balancing, electronic health records, marginal structural model, outcome-adaptive LASSO, regularization

I Introduction

I.1 Background

In the field of Biostatistics or medical research, often the focus is to establish causal relations between risk factors and disease outcomes using observational data, for which adjustments must be done to address confounding issues. Marginal structural models using inverse probability weighting is an effective method to handle confounders.^{1,2} This method utilizes propensity score models which are commonly fitted by logistic regression. In a simple binary treatment setting, the propensity score is defined as the probability of being treated given the covariates. Building propensity score models using logistic regression can sometimes be challenging, since data nowadays may have very high dimensionality. Regularization methods including LASSO and its variants, have been developed to solve this problem and to select essential variables for propensity score models.^{3,4}

In this paper, different regularization methods will be applied to a unique database: the MIMIC-III database. In the past 10 years, there has been a trend towards implementation of electronic health record systems in hospitals and the Medical Information Mart for Intensive Care (MIMIC) database developed by the Laboratory for Computational Physiology at MIT is freely accessible under a data agreement.^{5,6,7} The database contains clinical data of patients admitted to the Beth Israel Deaconess Medical Center in Boston, Massachusetts. Medical information including demographics, vital signs, laboratory results and nursing progress notes are available in this database; International Classification of Diseases (ICD-9) codes were documented on patient

¹Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, ON, Canada

²Department of Epidemiology and Biostatistics, College of Public Health, Temple University, Philadelphia, PA, USA

Corresponding author:

Yeying Zhu, Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, ON N2L 3G1, Canada.

Email: yeying.zhu@uwaterloo.ca

discharge. A code repository is available for the MIMIC-III database and researchers are encouraged to make contributions to the code repository, to enhance reproducibility in health research.⁸

Using the MIMIC-III database, we are interested in investigating how the use of transthoracic echocardiography (TTE) affects 28-day mortality for patients with sepsis, a life-threatening condition.⁹ However, diseases can develop from risk factors through a complicated indirect pathway. In other words, the risk factor of interest may affect the disease outcome through intermediate variables. Mediation analysis is a research area that focuses on modeling this indirect pathway to estimate both the direct and indirect effects of the risk factors on the outcome.

One of the earliest approaches to conduct mediation analysis, proposed by Baron and Kenny,¹⁰ models associations between treatment, mediator, and outcome by linear regressions. Direct and indirect effects can then be estimated from the fitted regression models. However, the estimated direct and indirect effects do not readily bear causal meanings. Robins et al.² highlighted that marginal structural models (MSMs) with inverse probability weighting can be used to establish a causal relationship between treatment and outcome in a setting without mediators. Later, VanderWeele,¹¹ Hong et al.,¹² and Lange et al.¹³ extended ideas of Robins et al.² and suggested that causal mediation analysis can be conducted using MSMs with inverse probability weighting. In the following, we will define notation and review the MSM approach for mediation analysis in detail.

1.2 Notation and review of MSMs

Consider a binary outcome Y and a baseline covariate vector X with dimension p . Define A as a binary treatment variable. Specifically, $A = 1$ indicates the treatment group and $A = 0$ indicates the control group. Define M as a binary mediator, which is on the causal pathway from A to Y . The observed data can be written as $(X_i, A_i, M_i, Y_i), i = 1, \dots, n$. We employ the potential outcomes framework¹⁶ to define causal mediation quantities. $Y_{a,M_{a^*}}$ is defined as the outcome that would have been observed if treatment level is set to a and mediator is set to the value it would have taken if treatment level is set to a^* . Under certain assumptions, the direct effect can be separated from the indirect effect and thus the total effect can be broken down into a natural direct and indirect effect.¹⁵ The natural direct effect (NDE_1) is defined as $E[Y_{1,M_1}] - E[Y_{0,M_1}]$; it can be interpreted as the average outcome difference between the treatment group and the control group while controlling mediator at the level of M_1 , the level the mediator would be under the treatment condition. The natural indirect effect (NIE_1) is defined as $E[Y_{1,M_1}] - E[Y_{1,M_0}]$, which is the average outcome difference between mediator level that would be obtained under treatment (i.e. M_1) and under control (i.e. M_0) conditions, holding treatment at the level of 1. Similarly, we can define NDE_0 as $E[Y_{1,M_0}] - E[Y_{0,M_0}]$ and NIE_0 as $E[Y_{0,M_1}] - E[Y_{0,M_0}]$. Then, the total effect $TE = E[Y_{1,M_1}] - E[Y_{0,M_0}]$ can be written as $NDE_1 + NIE_0$ or $NDE_0 + NIE_1$. On the other hand, since the outcome is binary, it would be more reasonable to define the causal quantities in terms of the odds ratio of the risk or the relative risk. For example, we can define the NDE_1 odds ratio as $\frac{P(Y_{1,M_1}=1)/P(Y_{1,M_1}=0)}{P(Y_{0,M_1}=1)/P(Y_{0,M_1}=0)}$ and the NIE_1 odds ratio as $\frac{P(Y_{1,M_1}=1)/P(Y_{1,M_1}=0)}{P(Y_{1,M_0}=1)/P(Y_{1,M_0}=0)}$. The total effect odds ratio is $\frac{P(Y_{1,M_1}=1)/P(Y_{1,M_1}=0)}{P(Y_{0,M_0}=1)/P(Y_{0,M_0}=0)}$, which can be written, e.g., as NDE_1 odds ratio \times NIE_0 odds ratio.

In this paper, we will follow the MSM approach by Lange et al.¹³ for mediation analysis. Marginal structural models have many advantages in terms of assessing natural direct and indirect effects of a given treatment. Specifically, MSMs can model both the natural direct and indirect effects at the same time without incorporating a mediator into the models, resulting in parsimonious models.¹³ A generalized linear MSM can be written in the following form

$$g(E[Y_{a,M_{a^*}}]) = c_0 + c_1a + c_2a^* + c_3a \cdot a^* \quad (1)$$

where g is a link function and a^* is the level of the treatment at which the mediator is controlled, representing the indirect pathway from the treatment to the outcome through the mediator. First, the original dataset is replicated once and a new variable A^* is created, where $A_i^* = A_i$ for $i = 1, \dots, n$ and $A_i^* = 1 - A_i$ for $i = n + 1, \dots, 2n$. Under the assumption that there are no unmeasured confounders between A and M , A and Y , and M and Y , the parameters in equation (1) can be consistently estimated by inverse probability weighting where the stabilized weights are computed as

$$W_i^S = \frac{P(A = A_i)}{P(A = A_i | X = X_i)} \frac{P(M = M_i | A = A_i^*, X = X_i)}{P(M = M_i | A = A_i, X = X_i)}, i = 1, \dots, 2n \quad (2)$$

If the interaction term (c_3) is zero and weights computed by equation (2) are used in MSMs, then the natural direct and indirect effect odds ratio is estimated as $e^{\hat{c}_1}$ and $e^{\hat{c}_2}$, respectively. Please refer to Lange et al.¹³ for a detailed list of the algorithm.

In order to compute the stabilized weights in equation (2), the treatment and mediator propensity score models need to be fitted. We assume the treatment propensity score model $P(A_i = 1|X_i)$ follows a logistic model: $\text{logit}\{P(A_i = 1|X_i)\} = \sum_{j=1}^p X_{ij}\alpha_j$. Similarly, the mediator propensity score model $P(M_i = 1|X_i, A_i)$ is assumed to follow: $\text{logit}\{P(M_i = 1|X_i, A_i)\} = \sum_{j=1}^p X_{ij}\beta_j + \eta A_i$.

When the covariate space is large, regularizing the fitted propensity score models is necessary. Parsimonious models are usually preferred, either from an estimation perspective or an inference perspective. In this paper, LASSO and its variants are used to select the important variables for both treatment and mediator propensity score models. For a given model, LASSO conducts variable selection by shrinking the coefficients of some variables to exactly zero, leading to a simpler model.¹⁶ Suppose we have a logistic regression model $P(Y_i = 1) = \text{expit}\{X_i^T \beta\}$. Let $\beta = (\beta_1, \dots, \beta_p)$ be the regression coefficients and λ be a nonnegative regularization parameter, the LASSO estimate $\hat{\beta}$ for logistic regression is defined as

$$\hat{\beta}^{\text{(LASSO)}} = \underset{\beta}{\text{argmin}} \sum_{i=1}^n \{-Y_i(X_i^T \beta) + \log(1 + e^{X_i^T \beta})\} + \lambda \sum_{j=1}^p |\beta_j|$$

Adaptive LASSO (AdaLASSO) is one variant of the traditional LASSO and has better statistical properties than the traditional LASSO. Zou¹⁷ developed an efficient algorithm for computing the AdaLASSO estimator, which is a global optimizer of the objective function

$$\hat{\beta}^{\text{(AdaLASSO)}} = \underset{\beta}{\text{argmin}} \sum_{i=1}^n \{-Y_i(X_i^T \beta) + \log(1 + e^{X_i^T \beta})\} + \lambda_n \sum_{j=1}^p \hat{w}_j |\beta_j|$$

where $\hat{w} = (\hat{w}_1, \dots, \hat{w}_p) = |\hat{\beta}^{(mle)}|^{-\gamma}$, $\gamma > 0$ and $\hat{\beta}^{(mle)}$ is the unpenalized maximum likelihood estimate of β . λ_n is a regularization parameter and γ is a power parameter. With a properly chosen λ_n , AdaLASSO satisfies the so-called oracle property. The first part of the oracle property indicates that AdaLASSO regularizes a given model as if the true underlying model is known.¹⁷ Consequently, when regularizing treatment propensity score models by AdaLASSO, the models should select only a subset of covariates, including all confounders and treatment-related variables. The second part of the oracle property essentially states that the AdaLASSO estimator is asymptotically unbiased, which is a property not shared by traditional LASSO.

In Section 2, outcome-adaptive LASSO which is another variant of LASSO proposed by Shortreed and Ertefaie,⁴ is outlined in detail. We then extend it to the mediation setting to estimate the natural direct and indirect effects. A simulation study is conducted to illustrate the performance of different regularization methods in Section 3. These methods are then applied to analyze a dataset extracted from the MIMIC-III database in Section 4. Section 5 is for discussion and conclusion.

2 Methodology

In the causal inference literature, variables that are only related to outcome are suggested for inclusion in the propensity score models. Rubin¹⁸ stated that including variables that are not related to the outcome in the propensity score models can result in efficiency loss. In a later paper, Brookhart et al.¹⁹ conducted simulation studies and discovered that adding variables that are only related to the outcome in the propensity score models can improve efficiency without increasing bias. Wyss et al.²⁰ also found similar results in a simulation setting with multiple outcomes. Zhu et al.²¹ found that, by achieving balance in the covariates that are related to the outcome, both finite bias and variance of the causal estimates can be reduced. Following similar ideas, outcome-adaptive LASSO (OAL) was developed for estimating the propensity scores in a high-dimensional setting by incorporating outcome information in the penalty function.⁴ In a mediation setting, it is natural to ask if such a principle of variable selection on the covariates still applies. In our simulation section, by examining the performance of several benchmark models, we found the propensity score model including the variables that are marginally related to the outcome variable leads to the most efficient estimates of the natural direct and indirect effects

and yield the smallest mean squared error, compared to other models. This again supports our speculation that the variable selection principle in the non-mediation setting also applies to the mediation framework. Therefore, we would extend the outcome-adaptive LASSO to mediation analysis.

2.1 Outcome-adaptive LASSO for mediation analysis

Outcome-adaptive LASSO is a modified version of adaptive LASSO. When restricting γ to be greater than 1 and $\lambda_n n^{(\gamma/2)-1}$ to go to infinity, outcome-adaptive LASSO puts less penalty weight on covariates that are only related to the outcome and thus tends to include such covariates as well as real confounders into the propensity score models.⁴ We can extend the idea of Shortreed and Ertefaie⁴ to the mediation setting where treatment can affect the outcome through a mediator.

Let \mathcal{C}_A denote indices of treatment-outcome confounders, which are defined as variables that are only related to A and Y. Let \mathcal{C}_M denote indices of mediator-outcome confounders, which are defined as variables that are only related to M and Y. Let \mathcal{C} denote indices of common confounders, which are defined as variables that are related to A, M, and Y. Let \mathcal{P} denote indices of covariates that are only related to Y and \mathcal{Q} denote indices of covariates that are only related to A. Denote \mathcal{R} as indices of covariates that are only related to M. Also, denote \mathcal{S} as indices of covariates that are not related to A, M, and Y. Suppose we want to estimate the following propensity score models:

$$\text{logit}\{\pi_{A_i}(\mathbf{X}_i, \boldsymbol{\alpha}_A)\} = \text{logit}\{P(A_i = 1 | \mathbf{X}_i, \boldsymbol{\alpha}_A)\} = \sum_{j \in \mathcal{C} \cup \mathcal{C}_A} \alpha_{A_j} X_{ij} + \sum_{k \in \mathcal{P}} \alpha_{A_k} X_{ik} \quad (3)$$

and

$$\begin{aligned} \text{logit}\{\pi_{M_i}(A_i, \mathbf{X}_i, \boldsymbol{\alpha}_M, \eta)\} &= \text{logit}\{P(M_i = 1 | A_i, \mathbf{X}_i, \boldsymbol{\alpha}_M, \eta)\} \\ &= \sum_{j \in \mathcal{C} \cup \mathcal{C}_M} \alpha_{M_j} X_{ij} + \sum_{k \in \mathcal{P}} \alpha_{M_k} X_{ik} + \eta A_i \end{aligned} \quad (4)$$

For logistic regression, the outcome-adaptive LASSO estimate has similar form as the adaptive LASSO estimate. The outcome-adaptive LASSO estimates, $\hat{\boldsymbol{\alpha}}_A = (\hat{\alpha}_{A_1}, \dots, \hat{\alpha}_{A_p})$ and $(\hat{\boldsymbol{\alpha}}_M, \hat{\eta}) = (\hat{\alpha}_{M_1}, \dots, \hat{\alpha}_{M_p}, \hat{\eta})$, which are defined as

$$\hat{\boldsymbol{\alpha}}_A = \underset{\boldsymbol{\alpha}_A}{\text{argmin}} \sum_{i=1}^n \{-A_i(\mathbf{X}_i^T \boldsymbol{\alpha}_A) + \log(1 + e^{\mathbf{X}_i^T \boldsymbol{\alpha}_A})\} + \lambda_A \sum_{j=1}^p \hat{w}_j |\alpha_{A_j}| \quad (5)$$

and

$$(\hat{\boldsymbol{\alpha}}_M, \hat{\eta}) = \underset{\boldsymbol{\alpha}_M, \eta}{\text{argmin}} \sum_{i=1}^n \{-M_i(\mathbf{X}_i^T \boldsymbol{\alpha}_M + A_i \eta) + \log(1 + e^{\mathbf{X}_i^T \boldsymbol{\alpha}_M + A_i \eta})\} + \lambda_M \left(\eta + \sum_{j=1}^p \hat{w}_j |\alpha_{M_j}| \right)$$

where $\hat{w}_j = |\tilde{\beta}_j|^{-\gamma}$ and $(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\eta}}) = \underset{\boldsymbol{\beta}, \boldsymbol{\eta}}{\text{argmin}} l_n(\boldsymbol{\beta}, \boldsymbol{\eta}; \mathbf{Y}, \mathbf{X}, \mathbf{A}, \mathbf{M})$. l_n is the negative log-likelihood of \mathbf{Y} given $\mathbf{X}, \mathbf{A}, \mathbf{M}$ parametrized by $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ for a sample size of n . $\tilde{\boldsymbol{\beta}}$ are the unpenalized coefficient estimates corresponding to all the covariates and $\tilde{\boldsymbol{\eta}}$ are the unpenalized coefficient estimates corresponding to treatment and mediator, respectively. Furthermore, $\tilde{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\eta}}$ can be obtained by fitting a linear regression of \mathbf{Y} on $\mathbf{X}, \mathbf{A}, \mathbf{M}$. Compared to AdaLASSO, OAL utilizes the additional information about the outcome variable while computing the weight for each coefficient, which AdaLASSO does not take account of. This is the major difference between AdaLASSO and OAL.

2.2 Selecting tuning parameters by optimizing balance

According to Rosenbaum and Rubin,²² the propensity score is one type of balancing score, which means that the covariates will be balanced among treatment groups conditioning on the propensity scores. This balancing property, along with the strongly ignorable treatment assignment (SITA) assumption, implies that unbiased causal treatment effects can be obtained if sufficient covariate balancing has been achieved.²³ Following this idea,

Shorttree and Ertefaie⁴ proposed that λ_A in equation (5) can be tuned by minimizing weighted absolute mean difference (wAMD), which is a balance statistic instead of a deviance statistic.

wAMD is a measure of how similar different exposure groups are and can be computed using the fitted treatment propensity score model as

$$wAMD(\lambda_A) = \sum_{j=1}^p |\tilde{\beta}_j| \left| \frac{\sum_{i=1}^n \hat{\tau}_{A_i, \lambda_A} X_{ij} A_i}{\sum_{i=1}^n \hat{\tau}_{A_i, \lambda_A} A_i} - \frac{\sum_{i=1}^n \hat{\tau}_{A_i, \lambda_A} X_{ij} (1 - A_i)}{\sum_{i=1}^n \hat{\tau}_{A_i, \lambda_A} (1 - A_i)} \right|$$

where $\hat{\tau}_{A_i, \lambda_A}$ is the inverse probability-weighted estimator for subject i , computed using the fitted treatment propensity score model of subject i with AdaLASSO or OAL method, $\hat{\pi}_{A_i, \lambda_A}(\mathbf{X}_i, \hat{\alpha}_A)$. Specifically, $\hat{\tau}_{A_i, \lambda_A}$ is defined as the following

$$\hat{\tau}_{A_i, \lambda_A} = \frac{A_i}{\hat{\pi}_{A_i, \lambda_A}(\mathbf{X}_i, \hat{\alpha}_A)} + \frac{1 - A_i}{1 - \hat{\pi}_{A_i, \lambda_A}(\mathbf{X}_i, \hat{\alpha}_A)}$$

Similarly, we can define wAMD in a way that utilizes the fitted mediator propensity score model, as the following

$$wAMD(\lambda_M) = \sum_{j=1}^p |\tilde{\beta}_j| \left| \frac{\sum_{i=1}^n \hat{\tau}_{M_i, \lambda_M} X_{ij} M_i}{\sum_{i=1}^n \hat{\tau}_{M_i, \lambda_M} M_i} - \frac{\sum_{i=1}^n \hat{\tau}_{M_i, \lambda_M} X_{ij} (1 - M_i)}{\sum_{i=1}^n \hat{\tau}_{M_i, \lambda_M} (1 - M_i)} \right|$$

where $\hat{\tau}_{M_i, \lambda_M}$ is the inverse probability of mediator weighted estimator for subject i , computed using the fitted mediator propensity score model of subject i with AdaLASSO or OAL method, $\hat{\pi}_{M_i, \lambda_M}(A_i, \mathbf{X}_i, \hat{\alpha}_M, \hat{\eta})$. Specifically, $\hat{\tau}_{M_i, \lambda_M}$ is defined as the following

$$\hat{\tau}_{M_i, \lambda_M} = \frac{M_i}{\hat{\pi}_{M_i, \lambda_M}(A_i, \mathbf{X}_i, \hat{\alpha}_M, \hat{\eta})} + \frac{1 - M_i}{1 - \hat{\pi}_{M_i, \lambda_M}(A_i, \mathbf{X}_i, \hat{\alpha}_M, \hat{\eta})}$$

Small wAMD indicates that, after applying the weighting procedure, the means of the covariates between different exposure groups are close to each other. Therefore, optimal λ_A and λ_M should be chosen by minimizing wAMD values. We will illustrate this idea in Section 3.3.

3 A simulation study

In this section, a simulation study was conducted to investigate the finite sample performance of the proposed methods. Results of multiple benchmark models and regularized models are presented. These methods are evaluated based on the relative bias, standard deviation, and mean squared error (MSE) of estimated NDE, NIE, and TE odds ratios. Assuming there is no interaction between the treatment and the mediator, we simply denote the natural direct effect as NDE and the natural indirect effect as NIE, since $NDE_1 = NDE_0$ and $NIE_1 = NIE_0$.

3.1 Simulation design

In the simulation study, 500 subjects ($n = 500$) and 200 continuous covariates ($p = 200$) are generated for each data set. Covariates are generated from a multivariate normal distribution with zero mean vector and identity covariance matrix, i.e. $\mathbf{X}_i \sim \text{MVN}(\boldsymbol{\mu} = \mathbf{0}, \boldsymbol{\Sigma} = \mathbf{I})$.

After the covariates are created, binary variables including the treatment, mediator, and outcome variables are generated using the covariates. The treatment of each subject is generated using a Bernoulli distribution with probability of being treated produced by a logistic regression model, i.e. $A_i \sim \text{Bernoulli}(P(A_i = 1))$, where $\text{logit}\{P(A_i = 1)\} = \sum_{j=1}^{200} \alpha_j X_{ij}$. Similarly, the mediator of each subject is generated using a Bernoulli distribution with probability produced by a logistic regression model, i.e. $M_i \sim \text{Bernoulli}(P(M_i = 1))$ where $\text{logit}\{P(M_i = 1)\} = \sum_{j=1}^{200} \beta_j X_{ij} + \eta_1 A_i$. The outcome of each subject is generated using a Bernoulli distribution

with probability produced by a logistic regression model, i.e. $Y_i \sim \text{Bernoulli}(P(Y_i = 1))$ where $\text{logit}\{P(Y_i = 1)\} = \sum_{j=1}^{200} \delta_j X_{ij} + \eta_2 A_i + \eta_3 M_i$.

We set $\boldsymbol{\alpha} = (1, 0, 1, 0, 0, \dots, 0)$, $\boldsymbol{\beta} = (1, 0, 0, 1, 0, \dots, 0)$ and $\boldsymbol{\eta} = (\eta_1, \eta_2, \eta_3) = (1, 1.5, 2)$ when generating the data. To consider different levels of correlation, we consider three different scenarios for the true values of $\boldsymbol{\delta}$

- Scenario A (weekly related): $\boldsymbol{\delta} = (0.2, 0.2, 0, 0, 0, \dots, 0)$
- Scenario B (moderately related): $\boldsymbol{\delta} = (0.6, 0.6, 0, 0, 0, \dots, 0)$
- Scenario C (strongly related): $\boldsymbol{\delta} = (1, 1, 0, 0, 0, \dots, 0)$

In the simulation setup, X_{i1} ($i = 1, \dots, n$) is a confounder of A_i and M_i as well as Y_i ; X_{i2} is only related to the outcome; X_{i3} is only related to the treatment; X_{i4} is only related to the mediator. The other 196 covariates are not related to the treatment, mediator, or outcome. Relations among the different simulated variables can be visualized using Figure 1. The theoretical NDE odds ratio is calculated as e^{η_2} and the theoretical NIE odds ratio needs to be calculated using a numerical approximation. The detailed derivation is provided in Web Appendix A in the Supporting Information. The theoretical TE odds ratio is the product of NDE and NIE odds ratio.

3.2 Modeling procedure

The treatment and mediator propensity score models need to be fitted before we can estimate NDE and NIE odds ratios by MSMs. After normalizing covariates to have mean zero and standard deviation one, treatment and mediator propensity score models are fitted using logistic regression with either LASSO, AdaLASSO, or OAL. For OAL, a linear regression model of Y on X, A, M is fitted and the regression coefficients are used to compute penalty weights for both the treatment and mediator propensity score models. For AdaLASSO, two linear regression models are fitted; one model regresses A on X and the other model regresses M on X, A . The regression coefficients of A on X are used to compute penalty weights for the treatment propensity score models and the regression coefficients of M on X, A are used to compute penalty weights for the mediator propensity score models. After we obtain penalty weights for the treatment and mediator propensity score models, treatment and mediator propensity score models regularized by AdaLASSO or OAL methods can then be fitted using the corresponding penalty weights.

In addition, four types of benchmark models are also fitted using logistic regression based on different subsets of covariates for comparison. For example, the true treatment benchmark model is fitted using X_{i1} and X_{i3} ; the true mediator benchmark model is fitted using X_{i1} and X_{i4} as well as the treatment variable. Other benchmark model specifications are included in Web Appendix B in the Supporting Information. The purpose of the benchmark models is two-fold: (1) to investigate which set of covariates should be included in the propensity models for calculating the inverse weights; (2) to provide benchmarks for examining the performance of the regularization methods.

Following the modeling procedure proposed by Lange et al.,¹³ NDE and NIE odds ratios are estimated by MSMs (equation (1)) with weights computed by equation (2). In addition, the original data set needs to be replicated twice to create a new data set before we can calculate W_i^S . A new variable A^* is created by setting it equal to A for first copy of the original data and equal to $1 - A$ for the second copy of the original data. Then, W_i^S can be computed using the fitted value of the treatment and mediator propensity score models on the new data set. Finally, MSMs are fitted with W_i^S on the new data set; the natural direct effect odds ratio is estimated as $e^{\hat{c}_1}$ and the natural indirect effect odds ratio is estimated as $e^{\hat{c}_2}$, where \hat{c}_1 is the estimated coefficient of a and \hat{c}_2 is the estimated coefficient of a^* in the fitted MSM (1) as mentioned in Section 1. The total effect odds ratio is estimated as $e^{\hat{c}_1} \times e^{\hat{c}_2}$.

3.3 Parameter tuning process

When fitting regularized treatment and mediator propensity score models, tuning parameters in the penalty term need to be chosen carefully to achieve reasonable coefficient estimates. In the simulation study, the LASSO regularization parameter (λ) is chosen by minimizing deviance with 10-fold cross-validation. Similarly, for each pre-specified γ value from a list: $\{0.5, 1, 2, \dots\}$, AdaLASSO and OAL regularization parameters (λ_n) are also chosen by minimizing deviance with 10-fold cross-validation.

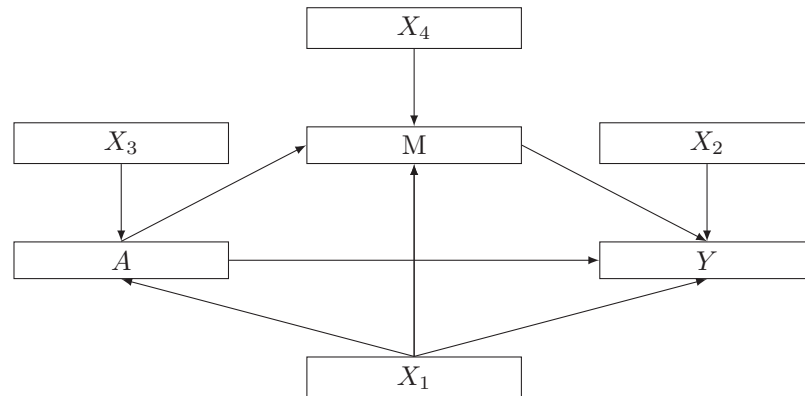


Figure 1. Relations among different simulated variables.

In addition, for each pre-specified power parameter γ value: $\{0.5, 1, 2, \dots\}$, AdaLASSO and OAL regularization parameters are chosen by minimizing both $wAMD(\lambda_A)$ and $wAMD(\lambda_M)$, from a set of regularization parameter values: $\{n^{-10}, n^{-5}, n^{-1}, n^{-0.75}, n^{-0.5}, n^{-0.25}, n^{0.25}, n^{0.49}\}$. This set of regularization parameter values proposed by Shortreed and Ertefaie,⁴ satisfies conditions of the asymptotic properties. We search for the optimal λ_A from this set based on $wAMD(\lambda_A)$, and then the optimal λ_M from the set based on $wAMD(\lambda_M)$. Therefore, the optimal regularization parameter values for AdaLASSO and OAL have the smallest $wAMD(\lambda_A)$ and $wAMD(\lambda_M)$ values.

Simulation results show that the performance of different adaptive methods is not sensitive to the value of γ and there is a slight increasing trend in the bias and standard deviation as the value of γ increases. Therefore, in the following, we only show the simulation results for $\gamma = 0.5$.

3.4 Simulation results

The simulation results based on 1000 replications are displayed in Table 1 for NDE, Table 2 for NIE, and Table 3 for TE. Additional graphs used to visualize simulation results are available in Web Appendix C in the Supporting Information.

First, we observe that the four benchmark models yield fairly close relative bias, standard deviation, and MSE of the estimated odds ratio. Across all benchmark models, the outcome benchmark model yields the smallest standard deviation and MSE of all the estimated NDE, NIE, and TE odds ratios. Adding outcome-related variables into the propensity score model improve the efficiency of the causal estimators, which is a common phenomenon found in the causal inference literature.¹⁹

Comparing OAL with AdaLASSO, i.e. OAL (deviance) vs. AdaLASSO (deviance) or OAL (wAMD) vs. AdaLASSO (wAMD), we found that by incorporating outcome information in the penalty weights, the standard deviation of the estimated odds ratios is reduced in most cases, indicating more efficient estimators. It also leads to smaller MSEs. This is again consistent with the finding from the benchmark models.

By comparing the criteria for choosing the tuning parameters, i.e. AdaLASSO (wAMD) vs. AdaLASSO (deviance) or OAL (wAMD) vs. OAL (deviance), we found in Scenarios B and C, where the outcome-related covariates are at least moderately related to the outcome variable, wAMD leads to much less biased estimates of NDE and TE odds ratios. This indicates the bias of the causal estimates can be reduced if the balance in the covariates is optimized.^{24–26}

Overall, all regularization methods we implement here outperform the benchmark models where the relationships among the variables are completely known. The underlying reason is that the propensity scores are nuisance parameters and by over-fitting the propensity scores, we can correct for the randomness and obtain better causal estimates. The phenomenon that the true propensity score model may perform worse than the estimated propensity scores has been shown in the literature, both theoretically and empirically, e.g. Lunceford and Davidian²⁷ and Brookhart et al.¹⁹ In general, OAL (deviance) tends to outperform the other regularization methods.

In addition, we examine which variables are selected by different regularization methods. Under Scenario B, we record the percentage of 1000 replications that each variable of $X_1 - X_4$ is chosen for both the treatment and mediator propensity scores and the results are shown in Table 4. As shown in the table, LASSO and AdaLASSO always choose the “true” variables (that is, X_1 and X_3 for the treatment propensity score model; X_1 and X_4 for the

Table 1. Performance measures of natural direct effect odds ratio estimators.

Scenario A	Relative bias (in %)	SD	MSE
LASSO	11.922	1.699	3.171
AdaLASSO (deviance)	-2.375	1.653	2.742
AdaLASSO (wAMD)	9.667	2.007	4.210
OAL (deviance)	-2.553	1.504	2.273
OAL (wAMD)	8.843	1.907	3.790
Benchmark (True)	-6.598	1.816	3.382
Benchmark (Outcome)	-10.864	1.380	2.139
Benchmark (True+Outcome)	-6.670	1.816	3.383
Benchmark (Full)	-6.237	1.835	3.440
Scenario B	Relative bias (in %)	SD	MSE
LASSO	2.298	1.360	1.857
AdaLASSO (deviance)	-13.191	1.305	2.052
AdaLASSO (wAMD)	-1.795	1.647	2.716
OAL (deviance)	-13.941	1.186	1.795
OAL (wAMD)	-3.337	1.566	2.471
Benchmark (True)	-23.052	1.346	2.877
Benchmark (Outcome)	-26.377	0.967	2.331
Benchmark (True+Outcome)	-23.187	1.323	2.829
Benchmark (Full)	-22.948	1.325	2.812
Scenario C	Relative bias (in %)	SD	MSE
LASSO	-10.953	1.084	1.414
AdaLASSO (deviance)	-24.810	1.018	2.272
AdaLASSO (wAMD)	-13.611	1.333	2.146
OAL (deviance)	-26.250	0.909	2.210
OAL (wAMD)	-15.832	1.257	2.081
Benchmark (True)	-35.270	1.039	3.577
Benchmark (Outcome)	-38.227	0.714	3.445
Benchmark (True+Outcome)	-35.634	0.996	3.542
Benchmark (Full)	-35.405	0.996	3.508

mediator propensity score model). On the other hand, OAL sometimes chooses the additional outcome-related variable X_2 . In the last column of Table 4, we record the average number of selected covariates for the treatment and mediator propensity scores, respectively. It is shown that the LASSO-based methods always select more variables than the true set, and thus lead to over-fitting. In addition, we find the wAMD criterion tends to select more covariates as we aim to achieve balance in all the available covariates $X_1 - X_{200}$. A separate research problem is whether we should achieve balance in all the available covariates or a subset of the covariates by variable selection. This requires further investigation. There is some preliminary study that shows by achieving balance in the covariates that are related to the outcome, both finite bias and variance of the causal estimates can be reduced,²¹ but to our best knowledge, no comprehensive simulations or theoretical investigations have been conducted in the literature to answer this question.

4 Data analysis: Transthoracic echocardiography and mortality

TTE has been widely adopted in hospitals during the past 10 years and some research has been conducted to examine the effectiveness of TTE on patients' mortality. Feng et al.⁹ found that the use of TTE can significantly reduce the 28-day mortality for ICU patients with sepsis and suggested that secondary analyses might be helpful to conduct, in order to understand the underlying causal mechanism. We employ the proposed methods to study whether the use of TTE affects patients' mortality indirectly. More specifically, we are interested in knowing whether the use of TTE would affect 28-day mortality by changing how the doctors treat the patients. Our analysis is based on the same data set used in the article of Feng et al.⁹ and this data set is extracted from the MIMIC-III database introduced in Section 1.

In this data set, patients who had TTE performed less than 24 h before their ICU admission or during their ICU stay are considered as the treatment group and the remaining patients are considered as the control group.

Table 2. Performance measures of natural indirect effect odds ratio estimators.

Scenario A	Relative bias (in %)	SD	MSE
LASSO	-11.586	0.105	0.038
AdaLASSO (deviance)	-1.205	0.137	0.019
AdaLASSO (wAMD)	-4.249	0.149	0.026
OAL (deviance)	2.075	0.152	0.024
OAL (wAMD)	2.595	0.164	0.028
Benchmark (True)	-2.107	0.141	0.021
Benchmark (Outcome)	-2.364	0.129	0.018
Benchmark (True+Outcome)	-2.112	0.142	0.021
Benchmark (Full)	-2.233	0.154	0.025
Scenario B	Relative bias (in %)	SD	MSE
LASSO	-12.883	0.100	0.043
AdaLASSO (deviance)	-3.335	0.130	0.019
AdaLASSO (wAMD)	-6.328	0.137	0.027
OAL (deviance)	-2.333	0.129	0.018
OAL (wAMD)	-5.344	0.145	0.027
Benchmark (True)	-4.698	0.127	0.021
Benchmark (Outcome)	-4.885	0.117	0.018
Benchmark (True+Outcome)	-4.727	0.127	0.021
Benchmark (Full)	-4.824	0.137	0.023
Scenario C	Relative bias (in %)	SD	MSE
LASSO	-14.446	0.088	0.050
AdaLASSO (deviance)	-6.710	0.113	0.022
AdaLASSO (wAMD)	-9.542	0.115	0.032
OAL (deviance)	-6.365	0.110	0.020
OAL (wAMD)	-8.686	0.119	0.029
Benchmark (True)	-8.090	0.112	0.026
Benchmark (Outcome)	-8.237	0.101	0.024
Benchmark (True+Outcome)	-8.115	0.111	0.026
Benchmark (Full)	-8.226	0.118	0.028

Moreover, 28-day mortality from the date of ICU admission is the binary outcome variable. There are 39 variables in the covariate space, including demographics, interventions, comorbidities, vital signs, laboratory results, and admission information. Regularization methods discussed in this paper are used to select important variables from the covariate space, for both treatment and mediator propensity score models.

The data set contains 6361 patients and 3262 of them are in the treatment group, whereas 3099 of them are in the control group. In our data analysis, we focus on the four potential mediators with no missing values: ventilation-free days, vasopressor-free days, the use of dobutamine, and the maximum dose of norepinephrine, which are mentioned as secondary outcomes in Feng et al.⁹ We first use the *mediation* package in R²⁸ to conduct an exploratory data analysis. The estimated natural direct, indirect, and total effects (on the linear scale) by the bootstrap approach are summarized in Table 5. According to Table 5, vasopressor-free days, the use of dobutamine, and the maximum dose of norepinephrine have significant total effects and indirect effects. Since ventilation-free days do not have a significant indirect effect, we removed it from the list of potential mediators. Moreover, we found that the maximum dose of norepinephrine has a significant interaction between the treatment and the mediator using the *testTMint* function in the *mediation* package. We did not observe a significant interaction term for the other three potential mediators.

It is worth noticing that more than half of the values are missing for several covariates including CVP values and laboratory results for BNP, troponin, and creatinine kinase. In Feng et al.,⁹ indicators are created to indicate whether patients are missing these covariates and we use the same set of indicators in our data analysis as well. Given that there are a large number of other variables that contain missing values, using imputation might create a tremendous amount of uncertainty. Therefore, we handle the other missing values using complete case analysis and patients who have any other covariates missing are excluded in our analysis; our reduced sample has 3021 patients with no missingness. In addition, six dummy variables are created to indicate which day of the week

Table 3. Performance measures of total effect odds ratio estimators.

Scenario A	Relative bias (in %)	SD	MSE
LASSO	-1.064	2.209	4.880
AdaLASSO (deviance)	-3.760	2.383	5.733
AdaLASSO (wAMD)	5.000	2.841	8.165
OAL (deviance)	-0.594	2.296	5.268
OAL (wAMD)	5.693	2.716	7.499
Benchmark (True)	-8.325	2.645	7.268
Benchmark (Outcome)	-12.788	2.047	4.848
Benchmark (True+Outcome)	-8.391	2.650	7.299
Benchmark (Full)	-8.176	2.652	7.297
Scenario B	Relative bias (in %)	SD	MSE
LASSO	-10.854	1.832	3.829
AdaLASSO (deviance)	-16.269	1.806	4.330
AdaLASSO (wAMD)	-8.200	2.301	5.562
OAL (deviance)	-15.836	1.800	4.254
OAL (wAMD)	-8.859	2.155	4.959
Benchmark (True)	-26.638	1.912	6.525
Benchmark (Outcome)	-29.984	1.384	5.555
Benchmark (True+Outcome)	-26.784	1.886	6.459
Benchmark (Full)	-26.685	1.873	6.388
Scenario C	Relative bias (in %)	SD	MSE
LASSO	-23.829	1.361	4.150
AdaLASSO (deviance)	-29.900	1.394	5.562
AdaLASSO (wAMD)	-21.989	1.743	4.994
OAL (deviance)	-30.976	1.261	5.473
OAL (wAMD)	-23.310	1.667	4.975
Benchmark (True)	-40.431	1.424	8.647
Benchmark (Outcome)	-43.306	0.986	8.567
Benchmark (True+Outcome)	-40.770	1.373	8.615
Benchmark (Full)	-40.642	1.373	8.574

Table 4. Variables selection results in scenario B.

	Model	X ₁ (%)	X ₂ (%)	X ₃ (%)	X ₄ (%)	# of X's selected
LASSO	Treatment PS	100	6.3	100	6.9	14.774
	Mediator PS	100	8.0	7.9	100	17.685
AdaLASSO (deviance)	Treatment PS	100	8.6	100	12.1	22.774
	Mediator PS	100	12.0	8.4	100	23.821
AdaLASSO (wAMD)	Treatment PS	100	50.2	100	50.1	100.757
	Mediator PS	100	45.9	45.2	100	95.578
OAL (deviance)	Treatment PS	100	30.3	98.2	7.8	15.985
	Mediator PS	100	34.6	5.0	98.2	16.777
OAL (wAMD)	Treatment PS	100	76.5	98.5	54.8	108.148
	Mediator PS	99.7	75.0	49.7	97.5	107.687

patients are admitted to ICU for the sake of computation efficiency. For mediators, the vasopressor-free days and the maximum dose of norepinephrine are dichotomized at their median for the applicability of the proposed method. For example, we replace each patient's maximum dose of norepinephrine with 1 if the patient's maximum dose of norepinephrine exceeds the median and 0, otherwise.

After handling the missing data issues and creating dummy variables, we then estimated the NDE and NIE log odds ratios using MSMs with different regularization methods for each of the three mediators: ventilation-free days, vasopressor-free days, and the maximum dose of norepinephrine. Here, we assume the potential mediators

Table 5. Exploratory data analysis results of four potential mediators (p-value is the number in parentheses and * is used to indicate a significant effect at $\alpha = 0.05$).

	NDE	NIE	TE
Ventilation-free days	-0.0001 (0.148)	-0.0173 (0.148)	-0.0175 (0.106)
Vasopressor-free days	0.0018 (0.730)	-0.0473 (<0.001*)	-0.0455 (0.010*)
Dobutamine use	-0.0687 (<0.001*)	0.0021 (0.028*)	-0.0666 (<0.001*)
Norepinephrine	-0.0884 (<0.001*)	0.0113 (<0.001*)	-0.0771 (<0.001*)

Table 6. Data analysis results of different regularization methods for three different mediators (p-value is the number in parentheses; only NIE estimates of dobutamine are non-significant and all other NDE/NIE estimates in this table are significant with a p-value < 0.001).

Vasopressor-free days	NDE log odds ratio	SE	NIE log odds ratio	SE
LASSO	-0.512	0.0890	0.105	0.0026
AdaLASSO (deviance, $\gamma = 0.5$)	-0.548	0.0896	0.138	0.0034
AdaLASSO (wAMD, $\gamma = 0.5$)	-0.560	0.0914	0.152	0.0040
OAL (deviance, $\gamma = 0.5$)	-0.549	0.0899	0.145	0.0036
OAL (wAMD, $\gamma = 2$)	-0.545	0.0921	0.151	0.0040
Full	-0.562	0.0926	0.152	0.0040
Dobutamine	NDE log odds ratio	SE	NIE log odds ratio	SE
LASSO	-0.416	0.0891	0.011 (0.086)	0.0061
AdaLASSO (deviance, $\gamma = 0.5$)	-0.426	0.0899	0.016 (0.071)	0.0087
AdaLASSO (wAMD, $\gamma = 0.5$)	-0.422	0.0915	0.014 (0.086)	0.0084
OAL (deviance, $\gamma = 0.5$)	-0.428	0.0917	0.015 (0.110)	0.0092
OAL (wAMD, $\gamma = 0.5$)	-0.423	0.0925	0.013 (0.120)	0.0084
Full	-0.422	0.0926	0.012 (0.130)	0.0078
Norepinephrine	NDE log odds ratio	SE	NIE log odds ratio	SE
LASSO	-0.445	0.0888	0.039	0.0109
AdaLASSO (deviance, $\gamma = 0.5$)	-0.455	0.0895	0.043	0.0126
AdaLASSO (wAMD, $\gamma = 0.5$)	-0.453	0.0911	0.044	0.0131
OAL (deviance, $\gamma = 0.5$)	-0.457	0.0911	0.043	0.0127
OAL (wAMD, $\gamma = 1$)	-0.453	0.0921	0.042	0.0125
Full	-0.457	0.0925	0.046	0.0136

are conditionally independent given the treatment and baseline covariates, so natural direct and indirect effects can be estimated separately for each mediator. The treatment and mediator propensity score models were fitted using different regularization methods discussed in Sections 1 and 2, to compute weights for MSMs. NDE and NIE log odds ratio estimates can then be obtained from those MSMs. All MSMs for the three mediators suggest that the interaction term between the treatment and the mediator is not significant. Therefore, we only present results of the MSMs without the interaction term in this section. Treatment and mediator propensity score models with no variable selection were also fitted for comparison; we refer to these propensity score models as the full models. Data analysis results of the different regularization methods for the three mediators are summarized in Table 6. As in the simulation study, for AdaLASSO and OAL, we tried different values of the power parameter γ in the penalty weight. We only report results of the different regularization methods with the γ value that leads to the smallest standard error of the estimates.

According to Table 6, all the estimated NDE log odds ratios obtained from the different regularization methods are significant for all three potential mediators. All these estimated NDE log odds ratios have negative signs indicating that the use of TTE can reduce the 28-day mortality directly. Moreover, the standard error of the estimated NDE log odds ratios obtained from the regularization methods is smaller than the standard error obtained from the full propensity score models with no variable selection conducted, which indicates that variable selection in propensity score models can lead to efficiency improvement.

From Table 6, we observe that the estimated NIE log odds ratios are not significant for the use of dobutamine and the estimated NIE log odds ratios are significant for vasopressor-free days and the maximum dose of



Figure 2. Matrix plots of covariates selected in propensity score models for the use of dobutamine (indices at X-axis represent LASSO, AdaLASSO (deviance, $\gamma = 0.5$), AdaLASSO (wAMD, $\gamma = 0.5$), OAL (deviance, $\gamma = 0.5$), OAL (wAMD, $\gamma = 0.5$), respectively).

norepinephrine. The positive signs for the estimated NIE log odds ratios for all three potential mediators indicate that the use of TTE may increase 28-day mortality through an indirect pathway. There are two possible explanations for the indirect pathway. The first explanation is that the use of TTE leads to an increase in the mediator level and then the increased mediator level increases patients' mortality. An alternative explanation is that TTE reduces the mediator level and the reduced mediator level leads to a decrease in patients' mortality. To find which explanation is more plausible, we can implement MSMs with inverse probability weighting to assess the causal relations between the potential mediator and mortality. In fact, according to the results of MSMs summarized in Web Appendix D in the Supporting Information, the use of TTE increases patients' mortality rate by increasing the use of dobutamine or the maximum dose of norepinephrine; the use of TTE increases patients' mortality rate by reducing the vasopressor-free days. In particular, the positive NIE log odds ratio estimates for dobutamine and norepinephrine seem to be reasonable since both have some adverse effects reported in the literature.²⁹

In Figure 2, matrix plots are used to show which covariates are selected in both the treatment and mediator propensity score models for dobutamine. If the cell of a given covariate is red, then the covariate is excluded from the propensity score model; if the cell of a given covariate is yellow, then the covariate is included in the propensity score model. We observe that the wAMD-based methods select more covariates in both the treatment and mediator propensity score models, compared to the deviance-based methods. In terms of estimating the NDE log odds ratio, the wAMD-based methods are less efficient since the propensity score models include too many covariates. Moreover, compared to AdaLASSO, OAL tends to select more covariates in the treatment propensity score model and thus is less efficient when estimating NDE and NIE log odds ratios. Matrix plots for the other two mediators are included in Web Appendix D in the Supporting Information.

In conclusion, data analysis results of all 3 potential mediators suggest that the use of TTE can affect the 28-day mortality via a direct pathway and via an indirect pathway. It is worth noticing that when the power parameter $\gamma = 0.5$, the outcome-adaptive LASSO method achieves the smallest standard error compared to other power parameter values. This result is consistent with our simulation results described in Section 3.4. Furthermore, for vasopressor-free days and maximum dose of norepinephrine, LASSO-based methods improve the efficiency of both the NDE and NIE log odds ratio estimators significantly, compared to the case where no variable selection is conducted for both the treatment and mediator propensity score models.

5 Concluding remarks

In this paper, the idea of outcome adaptive LASSO introduced by Shortreed and Ertefaie⁴ is implemented and extended from a traditional causal inference setting to a mediation setting. By examining both oracle and regularized models, we observe that the efficiency of both NDE and TE odds ratio estimators can be improved significantly by incorporating outcome information into the propensity score models. Smaller relative bias is obtained by wAMD-based methods, which indicates bias reduction might be achieved by covariate balancing. In conclusion, based on our simulation results, estimating NDE and NIE odds ratios with propensity score models regularized by OAL (deviance) would be the best choice across all regularization methods, since this method yields the smallest MSEs. One limitation of the simulation study is that the logistic regression models we fit match the data-generating models. The setting is limited with respect to the number of covariates related to the treatment, mediator, and outcome variables, as well as to the zero relationship of the 196 out of the 200 covariates. The issue of model misspecification is not considered in our simulation study.

In the data illustration, the underlying mechanism of how TTE affects mortality is fully examined by the MSM approach. While previous studies show the use of TTE can reduce mortality rate, an interesting phenomenon is observed where the use of TTE can increase mortality by increasing either dobutamine use or norepinephrine dose; TTE can also increase mortality by reducing vasopressor-free days, according to our findings in Section 3.3.

While we focus on a binary treatment and a binary mediator variable in this article, the proposed methodology can be extended to other types of treatments and mediators. For example, if both the treatment and the mediator are continuous, the form of the inverse weights in equation (2) will remain the same but the conditional (marginal) probabilities should be replaced by the conditional (marginal) densities. While it appears to be a natural extension, the estimation of the conditional densities can be challenging because of the “curse of dimensionality”. Under the normality assumption, we can transform the estimation of the conditional density to the estimation of the conditional mean.^{2,30} For example, instead of estimating $f(A_i|X_i)$ where $f(\cdot|\cdot)$ refers to the conditional density, we can first estimate $E(A = A_i|X = X_i)$ and then use normal density to approximate the desired conditional density. To estimate the conditional mean, the traditional LASSO or AdaLASSO for regression can be used for this task.

Our methodological development and simulation results are based on a simple setting with only one mediator. However, it is possible that the indirect effect is carried by multiple mediators simultaneously and the mediators may interact with each other. This is often observed in real-world settings, such as in electronic health records. Advanced variable selection methods need to be developed to select the right subset of mediators and to estimate the mediator propensity score models. Regularization methods illustrated in this paper can be extended to this setting with multiple mediators.

Acknowledgements

The authors thank Dr. Joel Dubin for critical readings of the original version of the paper and Sunny Huang for preparing the dataset. The content is solely the responsibility of the authors and does not necessarily represent the official views of NCI, OBSSR, or NIH.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Preparation of this article was supported by the National Institutes of Health (NIH) grant 1R01CA229542-01 (PI: Coffman) funded by the National Cancer Institute (NCI) and the Office of Behavioral and Social Science Research (OBSSR). Yeying Zhu's research is supported by the National Sciences and Engineering Research Council of Canada (grant number RGPIN- 2017-04064).

ORCID iD

Yeying Zhu  <https://orcid.org/0000-0002-9019-9716>

Supplemental material

Supplemental material for this article is available online.

References

1. Robins JM. Marginal structural models. In: *Proceedings of the section on Bayesian statistical science*, Anaheim, California, August 1997, pp.1–10. Alexandria, VA: American Statistical Association.
2. Robins JM, Hernan MA and Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology* 2000; **11**: 550–560.
3. Ghosh D, Zhu Y and Coffman DL. Penalized regression procedures for variable selection in the potential outcomes framework. *Stat Med* 2015; **34**: 1645–1658.
4. Shortreed SM and Ertefaie A. Outcome-adaptive lasso: variable selection for causal inference. *Biometrics* 2017; **73**: 1111–1122.
5. Johnson AE, Pollard TJ, Shen L, et al. MIMIC-iii, a freely accessible critical care database. *Sci Data* 2016; **3**: 160035.
6. Johnson, A., Pollard, T., & Mark, R. (2016). MIMIC-III Clinical Database (version 1.4). *PhysioNet*. <https://doi.org/10.13026/C2XW26>.
7. Goldberger, A., Amaral, L., Glass, L., Hausdorff, J., Ivanov, P. C., Mark, R., . . . & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation [Online]*. **101**(23), pp. e215–e220.
8. Johnson AE, Stone DJ, Celi LA, et al. The mimic code repository: enabling reproducibility in critical care research. *J Am Med Inform Assoc* 2017; **25**: 32–39.
9. Feng M, McSparron JI, Kien DT, et al. Transthoracic echocardiography and mortality in sepsis: analysis of the mimic-III database. *Intensive Care Med* 2018; **44**: 884–892.
10. Baron RM and Kenny DA. The moderator–mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J Personal Soc Psychol* 1986; **51**: 1173–1182.
11. VanderWeele TJ. Marginal structural models for the estimation of direct and indirect effects. *Epidemiology* 2009; **20**: 18–26.
12. Hong G, Deutsch J and Hill HD. Ratio-of-mediator-probability weighting for causal mediation analysis in the presence of treatment-by-mediator interaction. *J Educ Behav Stat* 2015; **40**: 307–340.
13. Lange T, Vansteelandt S and Bekaert M. A simple unified approach for estimating natural direct and indirect effects. *Am J Epidemiol* 2012; **176**: 190–195.
14. Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. *J Educ Psychol* 1974; **66**: 688.
15. Robins JM and Greenland S. Identifiability and exchangeability for direct and indirect effects. *Epidemiology* 1992; **3**: 143–155.
16. Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc* 1996; **58**: 267–288.
17. Zou H. The adaptive lasso and its oracle properties. *J Am Stat Assoc* 2006; **101**: 1418–1429.
18. Rubin DB. Estimating causal effects from large data sets using propensity scores. *Ann Intern Med* 1997; **127**: 757–763.
19. Brookhart MA, Schneeweiss S, Rothman KJ, et al. Variable selection for propensity score models. *Am J Epidemiol* 2006; **163**: 1149–1156.
20. Wyss R, Girman CJ, LoCasale RJ, et al. Variable selection for propensity score models when estimating treatment effects on multiple outcomes: a simulation study. *Pharmacoepidemiol Drug Saf* 2013; **22**: 77–85.
21. Zhu Y, Schonbach M, Coffman DL, et al. Variable selection for propensity score estimation via balancing covariates. *Epidemiology* 2015; **26**: e14–e15.
22. Rosenbaum PR and Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983; **70**: 41–55.
23. Sauppe JJ and Jacobson SH. The role of covariate balance in observational studies. *Naval Res Logist* 2017; **64**: 323–344.
24. Harder VS, Stuart EA and Anthony JC. Propensity score techniques and the assessment of measured covariate balance to test causal associations in psychological research. *Psychol Meth* 2010; **15**: 234.
25. Zhu Y, Savage JS and Ghosh D. A kernel-based metric for balance assessment. *J Causal Infer* 2018; **6**: 20160029.
26. Xie Y, Zhu Y, Cotton CA, et al. A model averaging approach for estimating propensity scores by optimizing balance. *Stat Meth Med Res* 2019; **28**: 84–101.

27. Lunceford JK and Davidian M. Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. *Stat Med* 2004; **23**: 2937–2960.
28. Tingley D, Yamamoto T, Hirose K, et al. Mediation: R package for causal mediation analysis. 2014; Available at the Comprehensive R Archive Network (CRAN), <http://CRAN.R-project.org/package=mediation> (accessed 19 February 2021).
29. Lietman PS, Perkin RM, Levin DL, et al. Dobutamine: a hemodynamic evaluation in children with shock. *J Pediatrics* 1982; **100**: 977–983.
30. Zhu Y, Coffman DL and Ghosh D. A boosting algorithm for estimating generalized propensity scores with continuous treatments. *J Causal Infer* 2015; **3**: 25–40.