

RESEARCH

Open Access



A doubly robust estimator for continuous treatments in high dimensions

Qian Gao¹, Jiale Wang¹, Ruiling Fang¹, Hongwei Sun² and Tong Wang^{1*}

Abstract

Background Generalized propensity score (GPS) methods have become popular for estimating causal relationships between a continuous treatment and an outcome in observational studies with rich covariate information. The presence of rich covariates enhances the plausibility of the unconfoundedness assumption. Nonetheless, it is also crucial to ensure the correct specification of both marginal and conditional treatment distributions, beyond the assumption of unconfoundedness.

Method We address limitations in existing GPS methods by extending balance-based approaches to high dimensions and introducing the Generalized Outcome-Adaptive LASSO and Doubly Robust Estimate (GOALDeR). This novel approach integrates a balance-based method that is robust to the misspecification of distributions required for GPS methods, a doubly robust estimator that is robust to the misspecification of models, and a variable selection technique for causal inference that ensures an unbiased and statistically efficient estimation.

Results Simulation studies showed that GOALDeR was able to generate nearly unbiased estimates when either the GPS model or the outcome model was correctly specified. Notably, GOALDeR demonstrated greater precision and accuracy compared to existing methods and was slightly affected by the covariate correlation structure and ratio of sample size to covariate dimension. Real data analysis revealed no statistically significant dose-response relationship between epigenetic age acceleration and Alzheimer's disease.

Conclusion In this study, we proposed GOALDeR as an advanced GPS method for causal inference in high dimensions, and empirically demonstrated that GOALDeR is doubly robust, with improved accuracy and precision compared to existing methods. The R package is available at <https://github.com/QianGao-SXMU/GOALDeR>.

Keywords Causal inference, Doubly robust, High-dimensional data, Generalized propensity score

Introduction

The advent of omics data and health care data makes it possible to draw causal conclusions from observational studies because a substantial number of covariates makes

the assumption of unconfoundedness plausible [1]. The propensity score (PS) method is a common statistical tool for performing such causal inference in observational studies. The PS method was originally developed to estimate the causal effects of a binary treatment, exposure, or intervention (hereafter referred to as 'treatment') on an outcome [2]. Recently, extensions of PS methods to the context of continuous treatment have been developed and are collectively known as generalized PS (GPS) methods. GPS methods are focused on estimating the dose-response function (DRF) describing the relationship between a continuous treatment and an outcome [3–5]. Similar to PS methods, GPS methods estimate

*Correspondence:

Tong Wang
tongwang@sxmu.edu.cn

¹ Department of Health Statistics, School of Public Health, MOE Key Laboratory of Coal Environmental Pathogenicity and Prevention, Shanxi Medical University, No.56 Xinjian South Road, Taiyuan 030001, China

² Department of Health Statistics, School of Public Health, Binzhou Medical University, Yantai, China



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

the DRF through regression adjustment [5], matching [6], stratification [7], and inverse probability weighting (IPW) [8]. Additionally, the doubly robust approach has been proposed and has received increasing attention as a robust method to model misspecification of either the GPS model or the outcome model [9, 10].

GPS is a probability density function of the treatment conditional on observed covariates [5]. The validity of GPS methods relies on the assumption that both the conditional mean and the conditional distribution of the treatment, given the covariates, must be correctly specified [8]. To relax this assumption, several balancing approaches have been proposed under a weighting or doubly robust framework [11–15]. The balancing approaches are focused on directly estimating weights under the balance constraints, including covariate balance; specifically, the weighted cross-moments between the treatment and each covariate are 0. Recent methods include the nonparametric covariate balancing generalized propensity score (npCBGPS) of Fong et al. [11], entropy balancing weights [12, 13], and covariate association eliminating weights of Yiu et al. [14]. Whereas these methods are appealing in terms of robustness to GPS model misspecification, the orders of the moment of both the covariates and the treatment to decorrelate must be carefully chosen. A higher moment may be helpful when there are nonlinear correlations between the treatment and the covariates, but this may violate the positivity assumption [13, 15]. To our knowledge, there is still a lack of guidance for specifying correct orders of moment, which is necessary to mitigate confounding bias. To address the issue of what moments to decorrelate, Huling et al. proposed distance covariance optimal weights (DCOWs) [15]. However, the abovementioned methods do not consider variable selection, which is another important factor influencing the performance of the estimated DRF [16–21]; therefore, their application is limited in the case of high-dimensional covariates.

The GPS methods are susceptible to the covariates being balanced. For example, the inclusion of instrumental variables (IVs) that can only predict the treatment in the GPS model could inflate variance without reducing bias in the estimates [16–21]. It has been well documented that an optimal GPS method should balance or control for all confounders and prognostic covariates that can only predict the outcome [16–21]. Doing so can not only remove confounding bias but also improve the efficiency of the estimates [16–21]. Hence, it is necessary to introduce variable selection techniques into GPS methods in high-dimensional context. In the doubly robust framework, Su et al. [22] and Colangelo et al. [23] used machine learning methods, and Antonelli et al. [24] used a Gaussian process to estimate nuisance parameters related to the GPS model and the outcome model. Unfortunately, these studies failed to address the

adverse influence of IVs. Under adaptive lasso-based shrinkage, our previously proposed generalized outcome-adaptive LASSO (GOAL) approach discourages the selection of IVs by strongly penalizing covariates that are not associated with the outcome [25]. The GOAL method is robust to the GPS model misspecification. However, its validity depends on the assumption that the outcome model is linear.

Here, we retained the idea of variable selection from the GOAL method and proposed a generalized outcome-adaptive LASSO and doubly robust estimation (GOALDeR) method. Unlike the GOAL method, our proposed method constructs a penalty function that is independent of the outcome model. Consequently, we can estimate the DRF in the doubly robust framework [21]. In recognizing that the correlation between the treatment and confounders is a source of confounding bias [15], our method uses a distance correlation coefficient as a measure to assess covariate balance. The distance correlation coefficient is zero if and only if the variables are independent of each other [26]. With a simulation, we show that the GOALDeR method is doubly robust, provides more precise and accurate estimates than existing methods, and is scarcely affected by the covariate correlation structure and ratio of sample size to covariate dimension. We also applied the GOALDeR method to investigate potential causality between epigenetic age acceleration and Alzheimer's disease (AD).

Generalized outcome-adaptive LASSO and doubly robust estimation

Notations and assumptions

We let $D_{i=1}^n = (T_i, Y_i, Z_i)$ denotes an independent and identically distributed sample drawn from a common joint distribution $f(T, Y, Z)$. Each subject $i \in \{1, \dots, n\}$ has a continuous treatment T_i whose support is $T \subseteq \mathcal{R}$, and an outcome Y_i . We characterize causal DRF using potential outcome notation [27] and define $Y_i(t)$ as the potential outcome for subject $i \in \{1, \dots, n\}$ given treatment level $T_i = t$ ($t \in T$). Our target estimand is $\mathbb{E}(Y_i(t))$. The observed $Z_i \in \mathcal{R}^p$ denotes pre-treatment covariates, where p is the dimension. Each available covariate belongs to one of four mutually exclusive covariate sets:

- confounders (Z_c): covariates that contribute to both the treatment and the outcome;
- prognostic covariates (Z_p): covariates that contribute to the outcome only;
- instrumental variables (Z_I): covariates that contribute to the treatment only;
- spurious covariates (Z_S): covariates that contribute to neither the treatment nor the outcome.

Under the potential outcome framework, we established the following assumptions to identify the DRF from

the observed data, and we maintained these assumptions throughout this work.

Assumption 1 (Consistency): For subject $i \in \{1, \dots, n\}$, $T_i = (t \in \mathcal{T})$ implies $Y_i = Y_i(t)$.

Assumption 2 (Positivity): The GPS or conditional probability density function of the treatment $f_{T|Z}(T_i = t|Z_i)$ is positive for any $t \in \mathcal{T}$ and for any $Z_i \in \mathcal{R}^p$.

Assumption 3 (Unconfoundedness): $Y_i(t) \perp T_i | Z_i, \forall t \in \mathcal{T}$ means that for any treatment level, the potential outcome $Y_i(t)$ is conditionally independent of the treatment given the covariates. Note that this assumption is untestable from the observed data.

Assumption 4 (Stable unit treatment assumption): This assumption indicates that there is no interference among subjects.

Variable selection based on outcome-adaptive LASSO

We retained the idea of variable selection from the GOAL method [25] and started by assuming the GPS model as follows:

$$E(T|Z) = \alpha_0 + \sum_{j=1}^p Z_j \alpha_j \tag{1}$$

As mentioned in the introduction, an optimal GPS method should control for or balance covariates that are associated with the outcome (including Z_c and Z_p). The covariate selection mechanism should be free from the outcome model for a doubly robust estimator. We borrowed the idea from the adaptive LASSO and achieved a covariate selection procedure by solving:

$$\hat{\alpha} = \arg \min_{\alpha} \|T - \alpha_0 - \sum_{j=1}^p Z_j \alpha_j\|_2^2 + \lambda_n \sum_{j=1}^p \hat{w}_j |\alpha_j| \tag{2}$$

where \hat{w}_j denotes penalty weight and is inversely proportional to the influence of covariates Z_j on the outcome. Here, the GOALDeR method defines an outcome model-free penalty weight as $\hat{w}_j = \left| \frac{dcor(Z_j, Y|T)}{\max_j dcor(Z_j, Y|T)} \right|^\gamma$, where $dcor(Z_j, Y|T)$ is the conditional distance correlation coefficient between Z_j and the outcome Y , given treatment T , measuring any kind of correlations [28]. $\gamma > 1$ is a tuning parameter. $\lambda_n > 0$ is another tuning parameter satisfying $\lambda_n / \sqrt{n} \rightarrow 0$ and $\lambda_n n^{\gamma/2-1} \rightarrow \infty$ for consistency in variable selection, as with the GOAL method [25, 29]. On the contrary, the GOAL method utilizes coefficients from a separate linear outcome model to create penalty weights, which means that the validity of the GOAL method depends on the correct specification of the outcome model.

Choosing λ_n

We propose dual-weight distance correlation (DWDC) as a measure for selecting the optimal λ_n , and the rule is minimizing DWDC. Similar to dual-weight correlation (DWC) in the GOAL method, the standpoint of DWDC is covariate balance for unbiased, efficient estimation. However, unlike DWC which only captures linear correlations between covariates and both the treatment and the outcome, DWDC uses distance correlation to capture all types of correlations between covariates and both the treatment and the outcome.

$$DWDC = \sum_{j=1}^p |dcor(Z_j, Y|T)|^2 |dcor_{w^{\lambda_n}}(Z_j, T)| \tag{3}$$

where $dcor_{w^{\lambda_n}}(Z_j, T)$ refers to the weighted distance correlation between covariate Z_j and the treatment, serving as a measure of covariate balance. The smaller the $|dcor_{w^{\lambda_n}}(Z_j, T)|$, the better the covariate balance achieved after weighting. Recall that $dcor(Z_j, Y|T)$ is the conditional distance correlation between Z_j and Y given T . Multiplying these two components implies that the DWDC is more affected by the imbalance of Z_c and Z_p and less affected by the imbalance of Z_I and Z_S . Hence, a smaller DWDC could further encourage the selection of Z_c and Z_p .

The balance weights w^{λ_n} in the DWDC are estimated using the DCOWs method with covariates selected according to Eq. (2) with λ_n , without requiring the specification of moment orders for both the covariates and the treatment to achieve decorrelation. The DCOWs method uses weighted distance covariance between the treatment

and covariates as a loss function and directly estimates balance weights under the following constraints: (1) the marginal distributions of the treatment and the covariates are preserved after weighting; (2) the weights are positive and sum to the sample size. The authors showed that the balance weights estimated by the DCOWs could enhance a doubly robust estimator. Further details are provided in the article by Huling et al. [15]. On the contrary, the balance weights w^{λ_n} in the DWC are estimated using npCBGPS [11], which requires the specification of moment orders for both the covariates and the treatment to achieve decorrelation of nonlinearities.

Estimating DRF using a doubly robust estimator

Based on balance weights estimated using covariates selected by optimal λ_n , the GOAL method uses the IPW

method to estimate DRF. The GOAL method cannot estimate DRF using the “doubly robust” method as variable selection in the GPS model hinges on the outcome model being correct, which undermines the “doubly robust” nature of the method [21]. In contrast, the variable selection in GOALDeR is independent of the outcome model; therefore, we ultimately use the doubly robust estimator of Kennedy et al. [10] to estimate DRF. The doubly robust estimator of Kennedy et al. [10] consists of two steps. In the first step, a pseudo-outcome is constructed, and in the second step, the pseudo-outcome is regressed on the treatment to estimate DRF. The pseudo-outcome can be estimated as [15]:

$$\hat{\theta}(T_i) = \frac{1}{n} \sum_{i=1}^n \hat{\mu}(\bar{\mathbf{Z}}, T_i) + (Y_i - \hat{\mu}(\mathbf{Z}_i, T_i))w_i \tag{4}$$

where $\hat{\mu}(\mathbf{Z}, T)$ denotes an estimate of the outcome model $\mu(\mathbf{Z}, T)$. Here, the Super Learner method (SL) which combines LASSO, XGBoost, Random Forest, and Support vector machines is applied to estimate $\hat{\mu}(\bullet)$ [30]. w_i denotes balance weights estimated by the DCOWs method with covariates selected by optimal λ_n . Subsequently, the DRF is estimated using a linear or non-linear regression model of the treatment on the pseudo-outcome. In this work, we used a linear regression model for comparison purposes.

Simulations

Simulation setup

We modeled simulations to assess the performance of GOALDeR and compare it with existing approaches when there are a large number of covariates. Following our previous studies [25, 31], we developed simulations

by adapting the research conducted by Tan et al. [32] and Shortreed et al. [29]. For each replicated dataset, p covariates and n individuals were drawn independently from a multivariate standard Gaussian distribution with varying correlations of 0, 0.2, and 0.5. We generated a continuous treatment and outcome from models given by:

$$\text{GPS model : } T = \sum_{j=1}^p m(Z_j)\alpha_j + \zeta, \zeta \sim N(0,1) \tag{5}$$

$$\text{Outcome model : } Y = \eta T + \sum_{j=1}^p g(Z_j)\beta_j + \xi, \xi \sim N(0,1) \tag{6}$$

where $\eta = 0$ or 2 .

We used two data-generating scenarios to compare the GOALDeR method with existing methods, which were summarized in Table 1. In the first scenario, we assumed that both the GPS model and the outcome model are linear, that is, $g(Z_j) = Z_j$ and $m(Z_j) = Z_j, j = 1, \dots, p$, and we conducted simulations in three settings by varying the strength of the relationship between confounders and outcome, and treatment. We considered varying levels of confounding because the strength of the confounders affects the bias, variance, and mean-squared error (MSE) of an estimate [19]. For all three settings, the first two covariates, Z_1 and Z_2 , are true confounders; the third and fourth covariates, Z_3 and Z_4 , are prognostic covariates; the fifth and sixth covariates, Z_5 and Z_6 , are IVs; and the other $p-6$ covariates are spurious covariates. The first setting (SoSt) sets $\alpha = (1,1,0,0,1,1,0, \dots, 0)$ and $\beta = (1,1,1,1,0,0,0, \dots, 0)$. The second setting (SoWt) sets $\alpha = (0.5,0.5,0,0,1,1,0, \dots, 0)$ and $\beta = (1,1,1,1,0,0,0, \dots, 0)$. The third setting (WoSt) sets $\alpha = (1,1,0,0,1,1,0, \dots, 0)$ and $\beta = (0.5,0.5,1,1,0,0,0, \dots, 0)$. The coefficients of 1

Table 1 Simulation scenarios. Treatment T is generated as $N(m(\mathbf{Z}), 1)$, and outcome Y is generated as $N(\eta T + g(\mathbf{Z}), 1)$ where $\eta = 0$ or 2

Scenarios	Covariates (Z)	(n, p)	Settings	$m(\mathbf{Z})$ (treatment)	$g(\mathbf{Z})$ (outcome)
1	$Z_p \sim N(\mathbf{0}_p, \Sigma)$ $\Sigma_{ij} = 1 (i = j)$ $\Sigma_{ij} = \rho (i \neq j)$ $\rho = 0, 0.2, 0.5$	(200,100) (500,200)	SoSt	$Z_1 + Z_2 + Z_5 + Z_6$	$Z_1 + Z_2 + Z_3 + Z_4$
			SoWt	$0.5 * Z_1 + 0.5 * Z_2 + Z_5 + Z_6$	$Z_1 + Z_2 + Z_3 + Z_4$
			WoSt	$Z_1 + Z_2 + Z_5 + Z_6$	$0.5 * Z_1 + 0.5 * Z_2 + Z_3 + Z_4$
2	$Z_p \sim N(\mathbf{0}_p, \Sigma)$ $\Sigma_{ij} = 1 (i = j)$ $\Sigma_{ij} = \rho (i \neq j)$ $\rho = 0, 0.2, 0.5$	(200,100) (500,200)	CoMt	$[\exp(Z_1/2)]$ $+ [(Z_2/(1 + \exp(Z_1))) + 10]$ $+ [(0.04 * Z_1 * Z_3 + 0.6)^3]$ $+ [(Z_2 + Z_4)^2] + Z_7 + Z_8$	$Z_1 + Z_2 + Z_3 + Z_4 + Z_5 + Z_6$
			MoCt	$Z_1 + Z_2 + Z_3 + Z_4 + Z_7 + Z_8$	$[\exp(Z_1/2)]$ $+ [(Z_2/(1 + \exp(Z_1))) + 10]$ $+ [(0.04 * Z_1 * Z_3 + 0.6)^3]$ $+ [(Z_2 + Z_4)^2] + Z_5 + Z_6$
			MoMt	$[\exp(Z_1/2)]$ $+ [(Z_2/(1 + \exp(Z_1))) + 10]$ $+ [(0.04 * Z_1 * Z_3 + 0.6)^3]$ $+ [(Z_2 + Z_4)^2] + Z_7 + Z_8$	$[\exp(Z_1/2)]$ $+ [(Z_2/(1 + \exp(Z_1))) + 10]$ $+ [(0.04 * Z_1 * Z_3 + 0.6)^3]$ $+ [(Z_2 + Z_4)^2] + Z_5 + Z_6$

and 0.5 for confounders are commonly used in epidemiology [31, 33–35].

Under the second scenario, we introduced model misspecification via a nonlinear transformation of confounders and conducted simulations under three settings by varying whether the GPS model or the outcome model was misspecified. The data-generating processes were similar to those in the simulations by Tan et al. [32] and Kang et al. [36], which explored the impact of model misspecification on DR and non-DR estimators. The first setting correctly specified the outcome model, and misspecified the GPS model (CoMt) given $m(Z_1) = \exp(Z_1/2)$, $m(Z_2) = (Z_2/(1 + \exp(Z_1))) + 10$, $m(Z_3) = (0.04 * Z_1 * Z_3 + 0.6)^3$, $m(Z_4) = (Z_2 + Z_4)^2$, $m(Z_j) = Z_j$ for $j > 4$, and $g(Z_j) = Z_j$ for $j = 1, \dots, p$. The second setting (MoCt) used a nonlinear data-generating process for the outcome, and linear for the treatment given $m(Z_j) = Z_j$ for $j = 1, \dots, p$, and $g(Z_1) = \exp(Z_1/2)$, $g(Z_2) = (Z_2/(1 + \exp(Z_1))) + 10$, $g(Z_3) = (0.04 * Z_1 * Z_3 + 0.6)^3$, $g(Z_4) = (Z_2 + Z_4)^2$, $g(Z_j) = Z_j$ for $j > 4$. The third setting (MoMt) used a nonlinear data-generating process for both the outcome and treatment, as with CoMt and MoCt. For all three settings, the coefficients were set to $\alpha = (1, 1, 1, 1, 0, 0, 1, 1, 0, \dots, 0)$ and $\beta = (1, 1, 1, 1, 1, 0, 0, 0, \dots, 0)$.

For other settings, including the true causal parameter in the DRF ($\eta = 0$ or 2), sample size, and the dimension of covariates, we followed Shortreed et al. [29] and our previous study [25]. For each setting, we generated 100 simulated datasets each for dimensionality (n/p ratio): $n=200, p=100$ and $n=500, p=200$. We searched over several possible λ_n values $\{n^{-10}, n^{-5}, n^{-2}, n^{-1.25}, n^{-1}, n^{-0.75}, n^{-0.5}, n^{-0.25}, n^{0.25}, n^{0.49}\}$ for each dataset and chose γ such that $\lambda_n n^{\gamma/2-1} = n^2$.

Furthermore, to investigate the impact of effect size on statistical testing, we also explored the performance of each method when the DRF parameter was set to 0.4 and 0.7. The data-generating processes are the same as those described in Table 1, with the only difference being $\eta=0.4$ or 0.7 . To examine the performance of the GOALDeR method as the sample size increases, we let $p=20$ and $n=200, 500, 1000$. The data-generating processes are the same as those described in Table 1, with the only difference being the values of (n, p) .

Comparing methods

We compared the following methods for estimating DRF: (1) GOAL [25], whose processes are similar to those of the GOALDeR method. The main differences between the GOAL method and the GOALDeR method are described in Sect. 2. A detailed implementation of the GOAL method can be found in the Supplementary Materials; (2) SL-DR, which estimates the DRF in the DR

framework of Kennedy et al. [10] (described in subsection 2.4). Briefly, the SL-DR method fits the GPS model and the outcome model using the SL method to estimate the pseudo-outcome. The SL method combines the results of LASSO, XGBoost, Random Forest, and Support vector machines. The balance weights used to estimate the pseudo-outcome are given by $w_i = f_T(T_i)/f_{T|Z}(T_i|Z_i)$ where the numerator is the marginal density of the treatment, and $f_{T|Z}(T_i|Z_i)$ is the GPS. In this study, we normally approximated both $f_T(T_i)$ and $f_{T|Z}(T_i|Z_i)$. The R packages used to implement the GOALDeR, SL-DR, and GOAL methods are available at <https://github.com/QianGao-SXMU/GOALDeR> and <https://github.com/QianGao-SXMU/GOAL>, respectively.

Results

The results of data-generating with $\eta = 2$ are shown following. The others are in the Supplementary Materials.

Estimation under scenarios 1 and 2 with a modest $p=20$

We performed simulations to evaluate GOALDeR and compare it with existing methods. For illustrating the performance of GOALDeR as the sample size (n) increases, we plotted the distribution of the causal parameter estimates using a boxplot and the proportion of times each covariate was selected for simulation with a modest number of covariates ($p=20$). For Scenario 1, where both the outcome and GPS models are linear, we present the results for SoSt (the confounders are strongly correlated with both the treatment and the outcome), with a true causal parameter equal to 2. The remaining results are provided in the Supplementary Materials. The boxplot of causal parameter estimates is presented in Fig. 1 (Supplementary Figs. S1 and S8). GOALDeR produced nearly unbiased estimates across all sample sizes, and the precision of the estimates was enhanced as n increased. In Scenario 2, when either the GPS model or outcome model was nonlinear (CoMt and MoCt), GOALDeR could still yield nearly unbiased estimates across all sample sizes (Fig. 2 and Supplementary Fig. S12) despite having unsatisfactory performance in variable selection (Supplementary Figs. S5 to S7 and S13 to S15). Under the setting where the outcome model is nonlinear and the GPS model is linear (MoCt), the variability of the estimates became smaller as the sample size increased. Not surprisingly, the estimates became biased when both the GPS and outcome models were nonlinear (MoMt; Fig. 2 and Supplementary Fig. S12).

The percentage of each covariate being selected under Scenario 1 is shown in Fig. 3 (Supplementary Figs. S2 to S4 and S9 to S11). We present the results for SoSt, with a true causal parameter equal to 2. The remaining results are presented in the Supplementary Materials. In

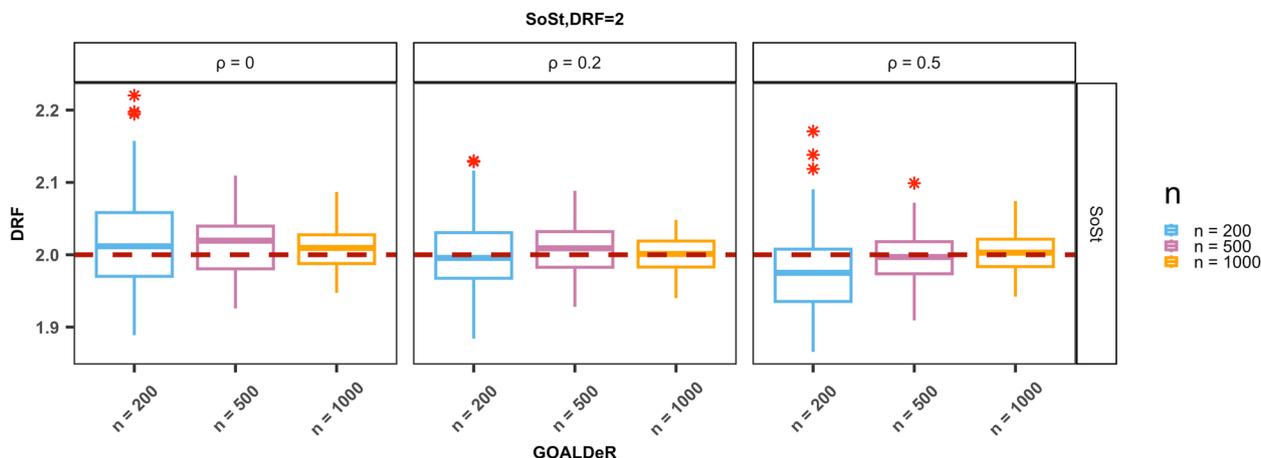


Fig. 1 Illustrations with a modest $\rho=20$. Boxplot of parameters for the dose–response function (DRF) under the setting where confounders are strongly correlated with both the treatment and outcome (SoSt) and $\eta=2$. The true causal parameter of 2 is indicated by a dotted line, and the asterisks represent outliers

general, the likelihood of selecting IVs decreased sharply with n increased. To illustrate no correlation between covariates, the average proportion of selecting IVs was 30% when $n=200$, decreasing to 1.5% when $n=500$, and further decreasing to 0 when $n=1000$. The selection of IVs and spurious covariates increased as the correlation between covariates increased. Although GOALDeR may underselect confounders that are weakly correlated with outcome (Supplementary Figs. S4 and S11), it still yielded nearly unbiased estimates (Supplementary Figs. S1 and S8). Additionally, the GOALDeR showed a similar variable selection pattern when there was a large number of covariates.

Estimation and testing under scenario 1 with a large number of covariates

In Scenario 1, we compared the accuracy and precision of causal estimates between GOALDeR, GOAL, and SL-DR under varying strengths between confounders and both the treatment and outcome. The bias of parameter estimates in the DRF was used to evaluate accuracy. The bias distribution is shown in Fig. 4, and the summary statistics for Scenario 1 are listed in Table 2. GOALDeR showed nearly unbiased estimates across all three settings (Fig. 4; Table 2). The root mean squared error (RMSE) and the empirical standard error of the estimates were used to assess precision. The precision of GOALDeR was slightly enhanced when $n=500$ compared to when $n=200$. Interestingly, the accuracy and precision of GOALDeR were slightly impacted by the correlations between covariates (Fig. 4; Table 2). In contrast, as previously observed, the bias and variability (RMSE and empirical standard error) of

GOAL increased as the correlations between covariates increased and the n/p ratio decreased. Compared with GOALDeR, SL-DR provided similar estimation accuracy, but the precision was significantly worse than that of GOALDeR (Fig. 4; Table 2). The reason is presumably owing to ignoring the negative effects of IVs when fitting the GPS model.

The standard deviation (SD) was estimated using the regression of the treatment on pseudo-outcome for the GOALDeR and SL-DR methods, and the sandwich-type variance estimator for the GOAL method [25]. The coverage probability of the 95% confidence interval (CI) and power were used to assess statistical testing. As shown in Table 2, for GOALDeR, the estimated SDs were lower than the empirical standard errors in most cases, resulting in the coverage of the 95% CI being less than 95% (ranging from 67 to 90%). For SL-DR, the estimated SDs were significantly lower than the empirical standard errors, and the coverage of the 95% CI (ranging from 48 to 81%) was consistently lower than that of the GOALDeR method. This implied that the estimated SDs for GOALDeR and SL-DR methods were underestimated. For GOAL, the estimated SDs were larger than the empirical standard errors, resulting in the coverage of the 95% CI tending to be conservative (ranging from 95 to 100%). We also estimated the SD using the bootstrap method for the GOALDeR method and found that the bootstrap SD was slightly higher than the empirical standard errors, and the corresponding coverage was around 95% in most cases.

GOALDeR, GOAL, and SL-DR had similar power, which was 1 in all three settings (Table 2). Furthermore, we also explored the performance of each method when

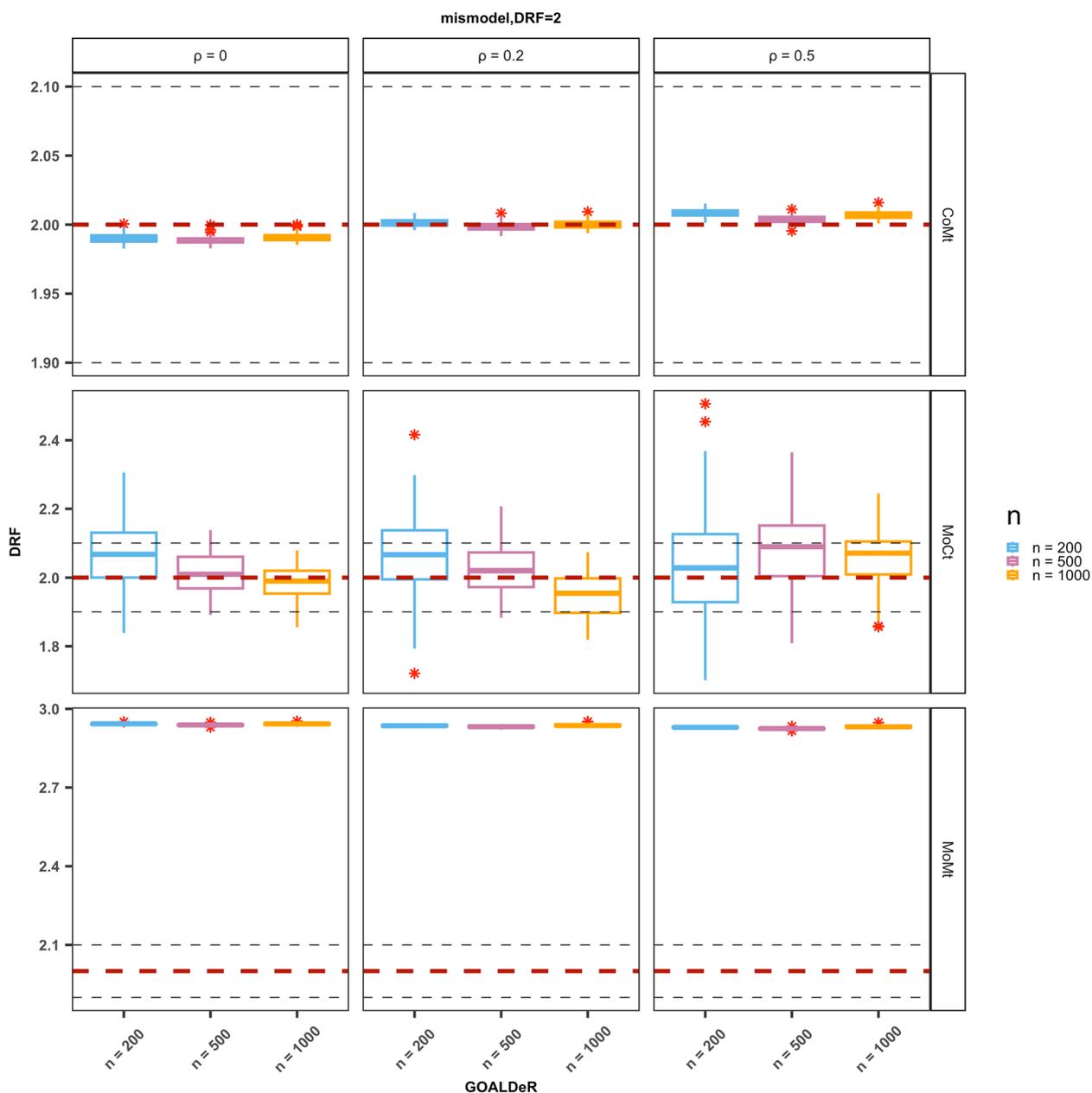


Fig. 2 Illustrations with a modest $\rho=20$. Boxplot of parameters for the dose–response function (DRF) under Scenario 2 with $\eta=2$. The true causal parameter of 2 is indicated by a red dotted line, and the asterisks represent outliers. The black dashed lines at 1.9 and 2.1 are primarily intended to aid in evaluation

the DRF parameter was set to 0.4 and 0.7 to investigate the impact of effect size on statistical testing. The results for $\eta = 0.4$ and 0.7 are presented in Supplementary Tables S3 and S4, which were similar to those for $\eta = 2$ except for the power. When the DRF parameter decreased from 2 to 0.4, the power for GOALDeR remained consistently at 1, while it slightly decreased

for SL-DR and significantly decreased for the GOAL method.

Estimation and testing under scenario 2 with a large number of covariates

In Scenario 2, we assessed the double robustness of GOALDeR, GOAL, and SL-DR. The bias distribution of parameter estimates in DRF is shown in Fig. 5, and the

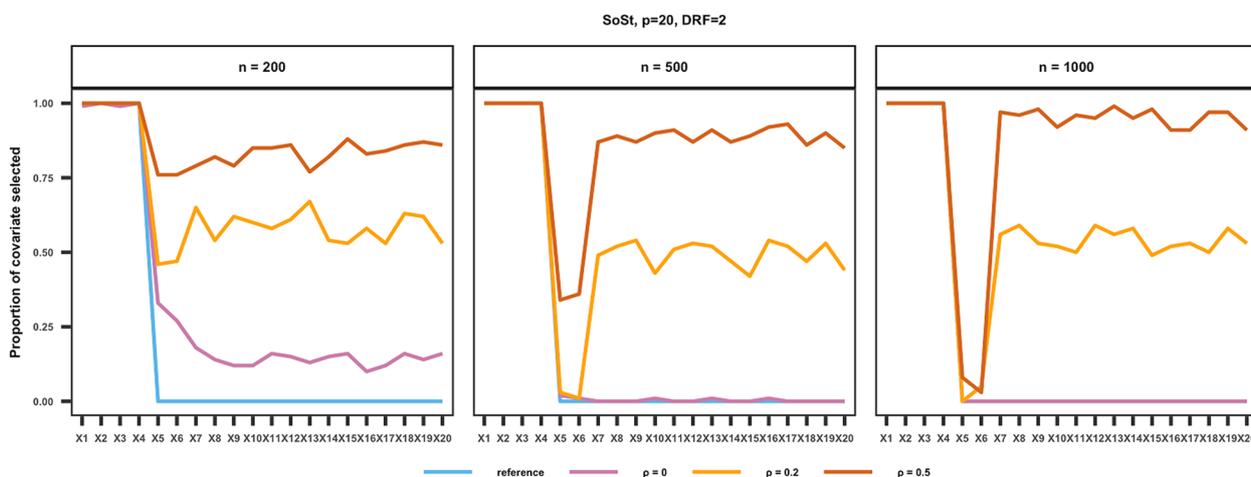


Fig. 3 Illustrations with a modest $\rho = 20$. The probability of covariate selection being balanced over 100 simulations under the setting where the confounders were strongly correlated with both the treatment and the outcome (SoSt) and $\eta = 2$

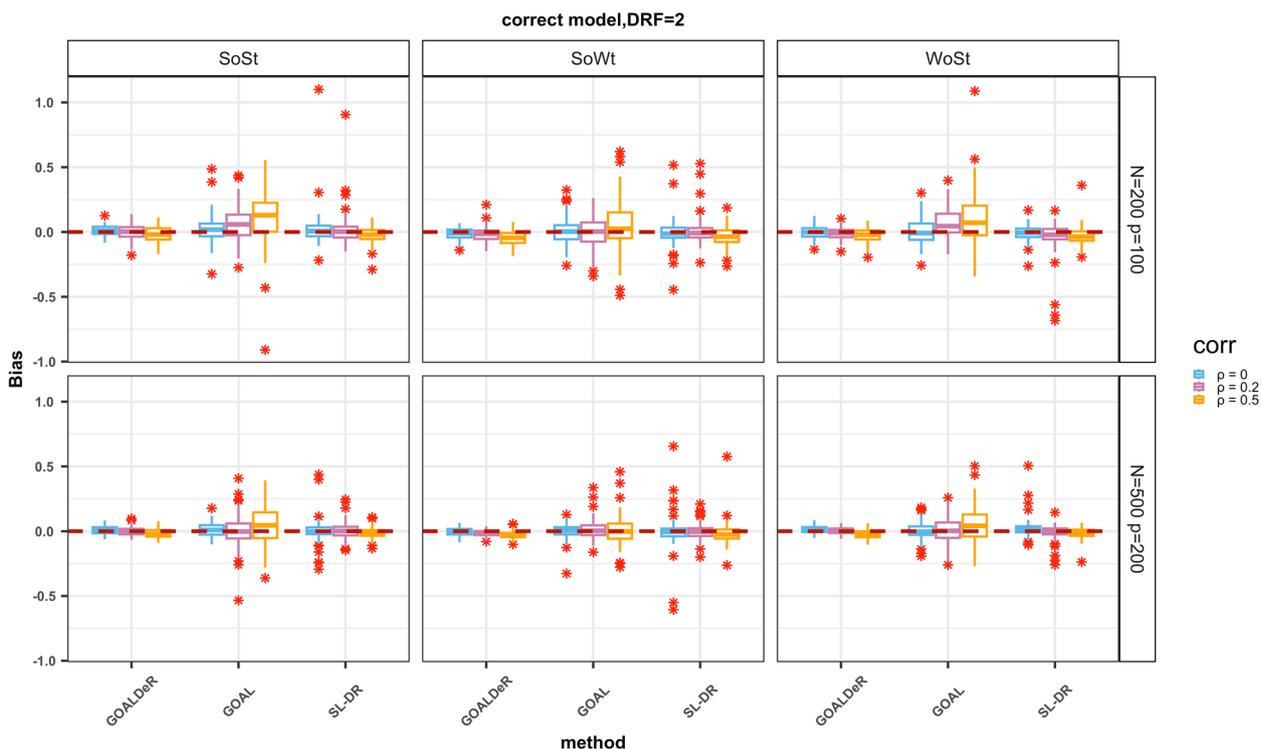


Fig. 4 Boxplot of the bias for causal parameters in the dose–response function (DRF) by our method and GOAL, SL-DR under Scenario 1 with $\eta = 2$. The zero reference line is indicated by a dotted line, and the asterisks represent outliers

summary statistics for Scenario 2 are listed in Table 3. GOALDeR yielded estimates that were close to the true value 2 as long as one of the outcome and the GPS models were correctly specified, and the biases were less impacted by the correlation between covariates and the n/p ratio. This indicated the double robustness of GOALDeR (Table 3; Fig. 5). In the setting of MoCt, the

variability (RMSE and empirical standard error) of the estimates by GOALDeR became large as the correlation between covariates increased, especially when the n/p ratio was small ($n/p = 200/100$). The SL-DR method also tended to be doubly robust. Compared to GOALDeR, SL-DR provided estimates with smaller biases and slightly higher RMSE in the MoCt setting. However,

Table 2 Summary statistics of the performance under scenario 1 with the true parameter of DRF = 2

		N = 500, P = 200										N = 200, P = 100											
		DRF	Est_Std	Est_Coverage	Est_Power	Boot_std	Boot_coverage	Boot_power	emp_std	emp_coverage	emp_power	RMSE	DRF	Est_Std	Est_Coverage	Est_Power	Boot_std	Boot_coverage	Boot_power	emp_std	emp_coverage	emp_power	RMSE
rho=0	SoSt	GOALDeR	2.009	0.025	0.85	1	0.041	0.98	1	0.033	0.034	0.034	2.01	0.034	0.9	1	0.061	1	1	0.038	0.04	0.038	0.04
	SL-DR	2.002	0.037	0.79	1	-	-	-	0.087	0.087	0.087	2.021	0.039	0.63	1	-	-	-	-	0.128	0.129	-	0.128
rho=0	SoWt	GOAL	2.012	0.1	1	-	-	-	0.052	0.053	0.053	2.025	0.162	1	1	-	-	-	-	0.106	0.109	-	0.106
	SL-DR	1.995	0.027	0.9	1	0.048	1	-	0.031	0.031	0.031	1.986	0.04	0.87	1	0.065	0.98	1	1	0.047	0.048	1	0.047
rho=0	SoSt	GOAL	1.997	0.047	0.8	-	-	-	0.125	0.124	0.124	1.994	0.049	0.79	1	-	-	-	-	0.102	0.102	-	0.102
	SL-DR	1.998	0.087	0.99	1	-	-	-	0.056	0.056	0.056	2.004	0.151	1	1	-	-	-	-	0.096	0.096	-	0.096
rho=0	SoWt	GOALDeR	2.012	0.022	0.87	1	0.041	1	0.028	0.031	0.031	1.998	0.033	0.81	1	0.061	0.98	1	1	0.051	0.051	1	0.051
	SL-DR	2.02	0.033	0.78	1	-	-	-	0.073	0.075	0.075	1.995	0.032	0.64	1	-	-	-	-	0.06	0.06	-	0.06
rho=0.2	SoSt	GOAL	2.002	0.087	0.96	1	-	-	0.064	0.063	0.063	2.002	0.133	0.97	1	-	-	-	-	0.095	0.095	-	0.095
	SL-DR	2.002	0.021	0.83	1	0.04	0.98	1	0.032	0.032	0.032	2.001	0.031	0.75	1	0.061	0.97	1	1	0.056	0.055	1	0.056
rho=0.2	SoWt	GOAL	2.004	0.033	0.57	1	-	-	0.063	0.063	0.063	2.017	0.043	0.66	1	-	-	-	-	0.119	0.12	-	0.119
	SL-DR	2.003	0.148	0.99	1	-	-	-	0.117	0.116	0.116	2.066	0.185	0.97	1	-	-	-	-	0.132	0.147	-	0.132
rho=0.2	SoSt	GOALDeR	1.985	0.025	0.9	1	0.046	1	0.027	0.031	0.031	1.982	0.038	0.78	1	0.066	0.98	1	1	0.052	0.055	1	0.052
	SL-DR	1.999	0.04	0.81	1	-	-	-	0.061	0.061	0.061	2.004	0.047	0.7	1	-	-	-	-	0.098	0.097	-	0.098
rho=0.2	SoWt	GOAL	2.01	0.13	0.99	1	-	-	0.072	0.073	0.073	1.998	0.2	1	1	-	-	-	-	0.117	0.116	-	0.117
	SL-DR	2.003	0.021	0.9	1	0.038	1	-	0.025	0.025	0.025	1.989	0.031	0.81	1	0.062	0.98	1	1	0.044	0.045	1	0.044
rho=0.2	SoSt	GOAL	1.992	0.033	0.8	1	-	-	0.055	0.055	0.055	1.968	0.04	0.61	1	-	-	-	-	0.123	0.126	-	0.123
	SL-DR	2.01	0.122	0.99	1	-	-	-	0.098	0.098	0.098	2.068	0.159	0.95	1	-	-	-	-	0.116	0.134	-	0.116
rho=0.5	SoSt	GOALDeR	1.983	0.024	0.81	1	0.041	0.96	0.034	0.038	0.038	1.984	0.031	0.67	1	0.066	0.98	1	1	0.059	0.061	1	0.059
	SL-DR	1.986	0.023	0.62	1	-	-	-	0.039	0.041	0.041	1.98	0.027	0.52	1	-	-	-	-	0.062	0.065	-	0.062
rho=0.5	SoWt	GOAL	2.046	0.183	0.99	1	-	-	0.146	0.152	0.152	2.112	0.223	0.9	0.99	-	-	-	-	0.213	0.24	-	0.213
	SL-DR	1.973	0.029	0.85	1	0.047	0.98	1	0.029	0.04	0.04	1.955	0.039	0.68	1	0.071	0.95	1	1	0.056	0.072	1	0.056
rho=0.5	SoSt	GOAL	1.983	0.034	0.64	1	-	-	0.08	0.082	0.082	1.963	0.037	0.59	1	-	-	-	-	0.07	0.079	-	0.07
	SL-DR	2.005	0.186	0.99	1	-	-	-	0.114	0.114	0.114	2.005	0.186	0.99	1	-	-	-	-	0.114	0.114	-	0.114
rho=0.5	SoWt	GOALDeR	1.979	0.024	0.74	1	0.04	0.96	0.034	0.04	0.04	1.977	0.032	0.74	1	0.065	0.97	1	1	0.056	0.06	1	0.056
	SL-DR	1.987	0.025	0.77	1	-	-	-	0.039	0.041	0.041	1.971	0.026	0.48	1	-	-	-	-	0.065	0.07	-	0.065
	GOAL	2.045	0.16	0.98	1	-	-	-	0.136	0.143	0.143	2.045	0.16	0.98	1	-	-	-	-	0.136	0.143	-	0.136

Notations: DRF = 2, the parameter of DRF is 2; rho = 0, rho = 0.2, and rho = 0.5, there are no, moderate and strong correlations between covariates; emp_std: the empirical standard deviation of the estimates; Est_Std: the mean of standard deviation estimates using sandwich-type estimator (GOAL) or regression of the treatment on pseudo-outcome (GOALDeR and SL-DR); Boot_std: the mean of bootstrapped standard deviation; coverage: coverage probability of the 95% confidence interval; RMSE: root mean square error calculated as $\sqrt{\frac{1}{n} \sum_{i=1}^n (e_i - \bar{e})^2}$

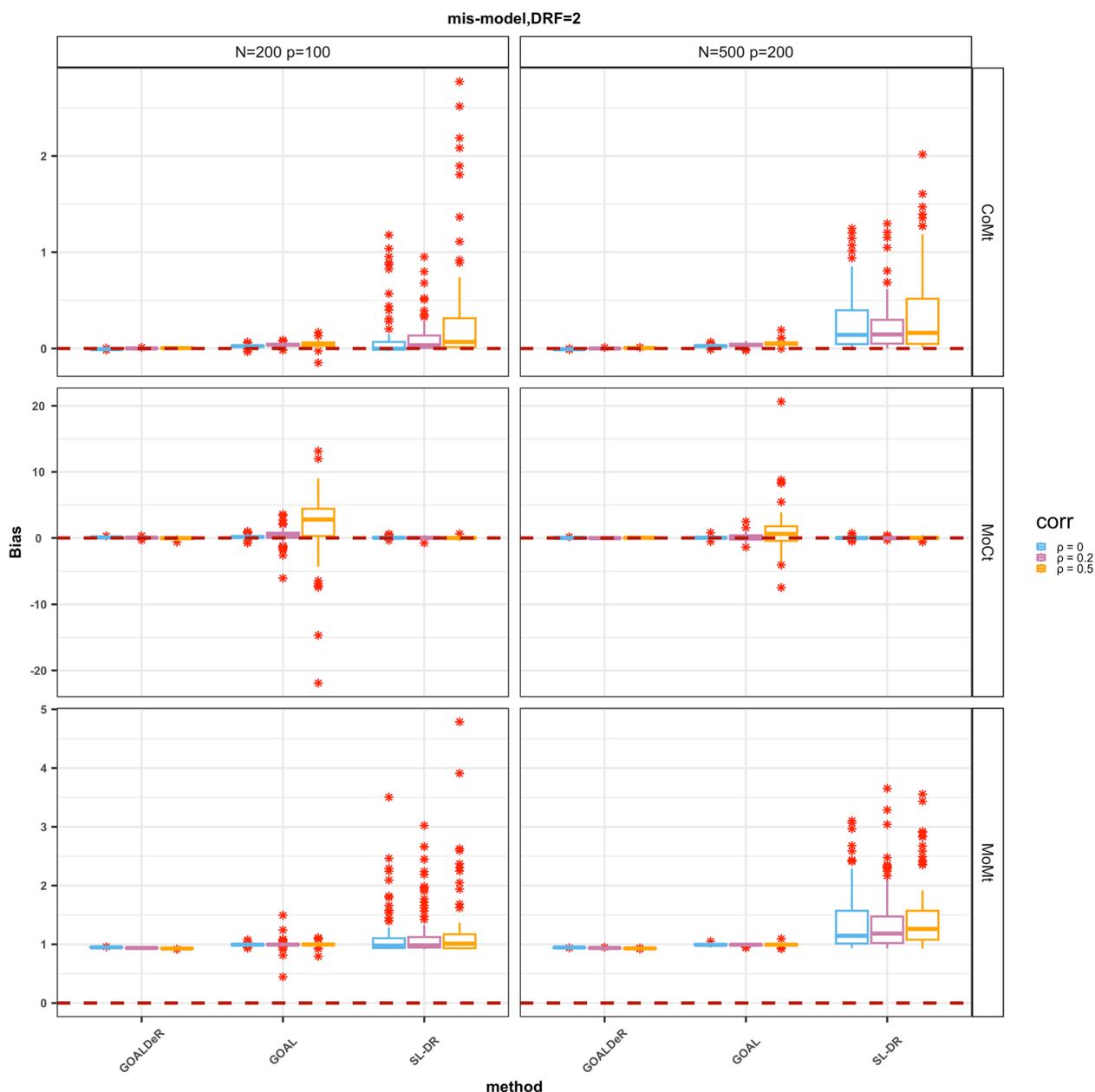


Fig. 5 Boxplot of the bias for the causal parameters in the dose–response function (DRF) using our method and GOAL, and SL-DR under Scenario 2 with $\eta=2$. The zero reference line is indicated by a dotted line, and the asterisks represent outliers

when only the GPS model was non-linear (CoMt), its biases and RMSE were significantly larger than those of the GOALDeR method, especially when there was a strong correlation among covariates. In contrast, GOAL became biased when the outcome model was incorrectly specified because it relies on the assumption that the outcome model is linear for variable selection. In the MoMt setting, all three approaches were biased, with SL-DR exhibiting the largest biases and RMSE.

As shown in Table 3, when one of the models was correctly specified (CoMt and MoCt), the estimated SDs of GOALDeR were less than or equal to the empirical standard errors, resulting in the coverage being less than 95% in most cases. The bootstrap SDs were higher than the empirical standard errors, and the corresponding coverage tended to be conservative (ranging from 93 to 100%). For the SL-DR method, the estimated SDs were significantly less than the empirical standard errors, and

Table 3 Summary statistics of the performance under scenario 2 with the true parameter of DRF = 2

		N = 500, P = 200										N = 200, P = 100												
		DRF	Est_Std	Est_Coverage	Est_Power	Boot_std	Boot_coverage	Boot_power	emp_std	emp_coverage	emp_power	RMSE	DRF	Est_Std	Est_Coverage	Est_Power	Boot_std	Boot_coverage	Boot_power	emp_std	emp_coverage	emp_power	RMSE	
CoMt rho=0	GOALDeR	1.993	0.002	0.12	1	0.006	0.93	1	0.002	0.99	1	0.007	1.993	0.004	0.58	1	0.0079	0.97	1	0.003	0.97	1	0.003	0.007
	SL-DR	2.257	0.054	0.08	1	-	-	-	0.297	-	-	0.392	2.094	0.032	0.14	1	-	-	-	0.242	-	-	0.259	
MoCt rho=0	GOAL	2.024	0.012	0.34	1	-	-	-	0.014	-	-	0.028	2.023	0.014	0.48	1	-	-	0.016	-	-	0.028		
	GOALDeR	2.022	0.055	0.96	1	0.079	0.99	1	0.049	0.99	1	0.054	2.078	0.078	0.83	1	0.116	0.97	1	0.083	0.97	1	0.114	
MoMt rho=0	SL-DR	2.007	0.077	0.77	1	-	-	-	0.138	-	-	0.137	2.039	0.074	0.61	1	-	-	0.149	-	-	0.153		
	GOAL	2.04	2.292	1	0	-	-	-	0.177	-	-	0.181	2.165	3.067	1	0	-	-	0.298	-	-	0.339		
CoMt rho=0.2	GOALDeR	2.946	0.003	0	1	0.009	0	1	0.003	0	1	0.946	2.948	0.004	0	1	0.012	0	0.004	0	1	0.004	0.948	
	SL-DR	3.388	0.09	0	1	-	-	-	0.546	-	-	1.49	3.129	0.058	0	1	-	-	0.391	-	-	1.194		
MoCt rho=0.2	GOAL	2.992	0.019	0	1	-	-	-	0.02	-	-	0.992	2.994	0.019	0	1	-	-	0.022	-	-	0.994		
	GOALDeR	2	0.002	0.73	1	0.0066	1	1	0.003	1	1	0.003	2.001	0.004	0.98	1	0.0096	1	0.003	1	1	0.003	0.003	
MoMt rho=0.5	SL-DR	2.227	0.044	0.01	1	-	-	-	0.261	-	-	0.345	2.109	0.031	0.12	1	-	-	0.171	-	-	0.202		
	GOAL	2.037	0.016	0.28	1	-	-	-	0.017	-	-	0.041	2.04	0.015	0.18	1	-	-	0.015	-	-	0.042		
CoMt rho=0.5	GOALDeR	1.986	0.048	0.79	1	0.099	0.99	1	0.072	0.99	1	0.073	2.027	0.059	0.65	1	0.1508	0.98	1	0.13	0.98	1	0.132	
	SL-DR	1.996	0.054	0.67	1	-	-	-	0.105	-	-	0.104	2.03	0.051	0.46	1	-	-	0.16	-	-	0.162		
MoMt rho=0.2	GOAL	2.074	2.673	1	0.01	-	-	-	0.598	-	-	0.599	2.329	3.717	1	0.01	-	-	1.182	-	-	1.221		
	GOALDeR	2.937	0.003	0	1	0.009	0	1	0.004	0	1	0.937	2.938	0.004	0	1	0.013	0	0.004	0	1	0.004	0.938	
CoMt rho=0.5	SL-DR	3.387	0.086	0	1	-	-	-	0.542	-	-	1.488	3.182	0.065	0	1	-	-	0.452	-	-	1.264		
	GOAL	2.99	0.014	0	1	-	-	-	0.016	-	-	0.99	2.997	0.03	0.01	1	-	-	0.084	-	-	1		
MoCt rho=0.2	GOALDeR	2.005	0.003	0.53	1	0.0084	1	1	0.002	0.84	1	0.005	2.004	0.004	0.8	1	0.012	1	0.004	0.84	1	0.004	0.006	
	SL-DR	2.352	0.062	0	1	-	-	-	0.434	-	-	0.557	2.31	0.074	0	1	-	-	0.561	-	-	0.638		
MoMt rho=0.5	GOAL	2.052	0.016	0.14	1	-	-	-	0.025	-	-	0.058	2.046	0.024	0.31	1	-	-	0.033	-	-	0.057		
	GOALDeR	2.023	0.041	0.52	1	0.142	0.99	1	0.11	0.99	1	0.111	1.963	0.058	0.41	1	0.209	0.96	0.189	0.96	1	0.189	0.191	
CoMt rho=0.2	SL-DR	2.014	0.04	0.37	1	-	-	-	0.132	-	-	0.132	2.009	0.044	0.32	1	-	-	0.183	-	-	0.182		
	GOAL	2.943	4.044	1	0.04	-	-	-	3.04	-	-	3.169	4.116	5.118	0.84	0.26	-	-	4.74	-	-	5.169		
MoMt rho=0.5	GOALDeR	2.929	0.003	0	1	0.0104	0	1	0.003	0	1	0.929	2.928	0.005	0	1	0.016	0	0.004	0	1	0.004	0.928	
	SL-DR	3.455	0.098	0	1	-	-	-	0.582	-	-	1.566	3.197	0.075	0	1	-	-	0.582	-	-	1.33		
CoMt rho=0.5	GOAL	2.991	0.016	0	1	-	-	-	0.022	-	-	0.992	2.995	0.022	0	1	-	-	0.035	-	-	0.996		

Notations: DRF = 2, the parameter of DRF is 2; rho = 0, rho = 0.2, and rho = 0.5, there are no, moderate and strong correlations between covariates; emp_std: the empirical standard deviation of the estimates; Est_Std: the mean of standard deviation estimates using sandwich-type estimator (GOAL) or regression of the treatment on pseudo-outcome (GOALDeR and SL-DR); Boot_Std: the mean of bootstrapped standard deviation; coverage: coverage probability of the 95% confidence interval; RMSE: root mean square error calculated as $\sqrt{\frac{1}{n} \sum_{i=1}^n (e_i - \bar{e})^2}$.

the coverage of the 95% CI (ranging from 0 to 77%) was consistently lower than that of the GOALDeR method. For the GOAL method, the estimated SDs were nearly equal to the empirical standard errors in the setting of CoMt, and the corresponding coverage was less than 95%. In the setting of MoCt, the estimated SDs of GOAL were significantly larger than the empirical standard errors, and the coverage was conservative. In the setting of MoMt, the coverage of all three methods was 0 because of the large bias.

The power of GOALDeR and SL-DR was always 1 in all three settings (Table 3). In contrast, the power of GOAL was significantly reduced when only the outcome model was incorrectly specified (MoCt). When the DRF parameter decreased from 2 to 0.4, the power for GOALDeR remained consistently at 1, while it slightly decreased for SL-DR. The coverage for the GOALDeR and SL-DR methods decreased.

Real data applications

We applied GOALDeR and SL-DR to study causal relationships between epigenetic age acceleration and AD. The results of the GOAL method have been reported in our previous study [25]. We followed steps similar to those implemented in the GOAL method to collect datasets, calculate DNA methylation (DNAm) age, and process and identify potential confounders [25]. Briefly, we downloaded seven datasets from the Gene Expression Omnibus database according to the inclusion and exclusion criteria. The accession numbers are GSE105109 [37], GSE125895 [38], GSE134379 [39], GSE59685 [40], GSE66351 [41], GSE80970 [42], and GSE109627 [43], covering four brain regions: frontal cortex (FC), temporal cortex (TC), entorhinal cortex (ERC), and cerebellum (CRB). The ‘cortical DNAm clock’ was used to estimate DNAm age, which is a measure of biological age [44]. The residuals of the regression model of chronological age on DNAm age were defined as epigenetic age acceleration. We considered chronological age and gender to be recognized risk factors for AD, and the datasets with raw data also controlled for the proportion of neuronal cells. Additionally, we regarded whole-genome CpG sites as potential covariates, as they may contain confounders and prognostic covariates or act as surrogates for these two types of covariates. Initially, we selected potential adjustment CpG sites through epigenome-wide association study (EWAS) meta-analysis for each brain region, keeping the top K CpG sites with the smallest Bonferroni-adjusted P values. The value of K for each brain region was determined as follows: $K = \text{minimum sample size in the specific brain region} - (\text{number of known covariates} + 2)$, since GOALDeR is not directly applicable when $p > n$.

Table 4 shows the estimated causal DRF of the GOALDeR and SL-DR methods between epigenetic age acceleration and AD across four brain regions. For the GOALDeR method, the four brain regions showed consistent results that there was no statistically significant dose–response relationship between epigenetic age acceleration and AD ($P > 0.05$). For the SL-DR method, the results for the four regions were inconsistent. We therefore performed a meta-analysis with a random-effects model (Supplementary Fig. S18) because there was heterogeneity among datasets (TC: $I^2 = 96.4\%$, $Q = 111.76$, $P < 0.0001$; FC: $I^2 = 94.5\%$, $Q = 54.63$, $P < 0.0001$; ERC: $I^2 = 92.8\%$, $Q = 27.64$, $P < 0.0001$; CRB: $I^2 = 90.3\%$, $Q = 30.95$, $P < 0.0001$). The pooled odds ratios were 0.9985 (95% confidence interval: 0.9943–1.0027, $P = 0.4883$), 1.0006 (95% confidence interval: 0.9945–1.0067, $P = 0.8550$), 0.9643 (95% confidence interval: 0.8860–1.0496, $P = 0.4008$), and 0.9885 (95% confidence interval: 0.9510–1.0276, $P = 0.5601$), respectively.

In summary, the GOALDeR and SL-DR analyses found that there was no statistically significant dose-response association between epigenetic age acceleration and AD, which is consistent with the results of the GOAL method [25]. In addition, the results of the SL-DR method showed greater variability than those of GOALDeR, which is consistent with our simulation results.

Discussion

We developed a new approach, GOALDeR, to estimate the linear or nonlinear DRF in high dimensions. Our extensive simulation studies, conducted under both correct and incorrect model specifications, indicated that GOALDeR can produce nearly unbiased estimates as long as either the outcome or GPS model is correctly specified. Therefore, it shows doubly robust empirically. The performance of GOALDeR is less impacted by the n/p ratio and correlated covariates. GOALDeR can also achieve statistical power and 95%CI coverage that are comparable to those of other methods.

Our simulations show that GOAL requires a linear outcome model to produce unbiased estimates, but the accuracy and precision worsen when there are correlated covariates or the n/p ratio is small. These results are consistent with those of previous studies [25]. The SL-DR requires the user to specify the conditional and marginal distributions of the treatment [22–24]. In our simulations, we assumed normal distributions for the treatment and the GPS both in the data-generating process and in the estimation of balance weights for SL-DR. This setting may partly contribute to the nearly doubly robust performance of SL-DR and explain why SL-DR performs less accurately and precisely than GOALDeR when the GPS

Table 4 Summary statistics for the datasets and results of GOALDeR and SL-DR analyses across the four brain regions

Brain region	dataset	N (AD/Control)	Potential confounders	Results for GOALDeR			Results for SL-DR				
				DRF	sd(DRF)	t	P	DRF	sd(DRF)	t	P
TC	GSE59685	861 (489/372)	age, gender, 60 CPG sites	0.0021	0.0054	0.3997	0.6904	-0.0011	0.0016	-0.7056	0.4824
	GSE66351		age, gender, neuronal cell proportion, 60 CPG sites	-0.011	0.0089	-1.2425	0.2186	-0.007	0.0002	-30.746	<0.0001
	GSE80970		age, gender, 60 CPG sites	0.0032	0.0037	0.8758	0.3826	0.0042	0.0016	2.5909	0.0106
	GSE109627		age, gender, 59 CPG sites	0.0004	0.0033	0.1092	0.9132	-0.0001	0.0009	-0.1107	0.912
	GSE134379		age, gender, neuronal cell proportion, 60 CPG sites	0.0097	0.0056	1.7423	0.0822	-0.0044	0.0051	-0.8533	0.394
FC	GSE59685	357 (192/165)	age, gender, 58 CPG sites	0.0068	0.0042	1.6141	0.1103	-0.0042	0.0022	-1.8791	0.0638
	GSE66351		age, gender, neuronal cell proportion, 58 CPG sites	-0.0027	0.0052	-0.5117	0.6107	-0.0031	0.0014	-2.2978	0.025
	GSE80970		age, gender, 58 CPG sites	-0.0026	0.004	-0.6533	0.5146	-0.0002	0.0016	-0.1314	0.8956
	GSE125895		age, gender, neuronal cell proportion, 57 CPG sites	0.0034	0.003	1.1252	0.2646	0.0094	0.0013	7.0841	<0.0001
	GSE59685	336 (212/124)	age, gender, 64 CPG sites	-0.0084	0.0051	-1.6233	0.1086	-0.129	0.0254	-5.0823	<0.0001
ERC	GSE105109		age, gender, neuronal cell proportion, 62 CPG sites	-0.0009	0.0027	-0.3418	0.7329	0.0106	0.0079	1.3435	0.1808
	GSE125895		age, gender, neuronal cell proportion, 64 CPG sites	-0.0002	0.0023	-0.0991	0.9214	0.0001	0.001	0.1012	0.9197
	GSE59685	746 (445/301)	age, gender, 59 CPG sites	-0.0117	0.0069	-1.7066	0.0917	-0.0774	0.0164	-4.7099	<0.0001
	GSE105109		age, gender, neuronal cell proportion, 46 CPG sites	0.0022	0.0108	0.2058	0.8371	-0.0001	0.0062	-0.022	0.9824
	GSE125895		age, gender, neuronal cell proportion, 58 CPG sites	0.0011	0.0071	0.1581	0.8749	0.0087	0.0032	2.6944	0.009
GSE134379		age, gender, neuronal cell proportion, 44 CPG sites	0.0118	0.0078	1.52	0.1293	0.0133	0.0035	3.7385	0.0002	

Notations: DRF, the causal parameter of the DRF; N, sample size

model is misspecified and the outcome model is correctly specified. Additionally, the variability of estimates for SL-DR is greater than that of GOALDeR. This may be because SL-DR ignores the influence of IVs when estimating the GPS [16–21].

The outstanding performance of GOALDeR in estimation accuracy and precision may be attributed to the following: (1) GOALDeR uses a balance-based method to estimate balance weights, thereby avoiding the need to specify distributions for the treatment and GPS [15]; (2) GOALDeR uses a distance correlation coefficient as the measure to assess covariate balance, thereby avoiding the need to specify the orders of moment of both the covariates and the treatment to decorrelate [15]; (3) GOALDeR constructs penalty weights based on conditional correlation between the outcome and covariates without depending on the outcome model, thereby achieving exclusion IVs and estimation DRF in the doubly robust framework [21]. However, as with most existing methods [16, 21, 22, 25, 29], GOALDeR lacks a standard deviation estimator to guarantee a valid confidence interval. Here, GOALDeR uses the regression coefficient of the treatment on pseudo-outcome to obtain an inference of DRF. The corresponding power consistently equaled 1, while the coverage of the 95% CI was often less than the nominal value, suggesting that the SDs were underestimated. This underestimation may be due to the estimated SD failing to adequately capture the variability of variable selection. We also employed the bootstrap method to estimate the SD and found that (i) the bootstrap SDs were slightly higher than the empirical standard errors when both the GPS and outcome models were correctly specified, resulting in coverage probabilities around the nominal value in most cases; (ii) the bootstrap SDs were moderately higher than the empirical standard errors when either the GPS or the outcome model was correctly specified, leading to coverage probabilities that tended to be conservative (greater than the nominal value) in most cases. Although the bootstrap method tends to improve the statistical tests, it does not completely resolve the inference problem after variable selection [45]. Therefore, further research on the development of a valid and widely applicable variance estimator after variable selection is a possible topic in future work [46].

In summary, this study proposed a doubly robust estimator for continuous treatment and high-dimensional covariates. Within the framework of the doubly robust (DR) estimator, the proposed GOALDeR method combined a variable selection technique for causal inference to ensure unbiased and statistically efficient estimation, along with a balance-based method that was robust to misspecification of the distributions required for GPS methods. Simulation results and real data analyses

provided empirical evidence that GOALDeR achieved double robustness, offering improved accuracy and precision compared to existing methods. We also provided an R package for implementing the GOALDeR method, available at <https://github.com/QianGao-SXMU/GOALDeR>.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12874-025-02488-3>.

Supplementary Material 1

Acknowledgements

Not applicable.

Authors' contributions

T.W. conceived the idea and contributed to the interpretation of the results. T.W. and Q.G. developed the model. Q.G. implemented the software; conducted analyses of simulation and real data with assistance from J.W.; interpreted the results with assistance from R.F. and H.S.; and drafted and revised the manuscript with input from all other authors. All authors approved the final manuscript.

Funding

This study was supported by the National Natural Science Foundation of China (grant numbers: 82204163, 82373692 and 82073674) and the Fundamental Research Program of Shanxi Province (grant number: 202203021212382).

Data availability

No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate

All data used in this study came from public databases, and the original studies had been approved by the relevant review boards.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 2 February 2024 Accepted: 3 February 2025

Published online: 13 February 2025

References

1. Athey S, Imbens GW, Wager S. Approximate residual balancing: debiased inference of average treatment effects in high dimensions. *J R Stat Soc Ser B-Stat Methodol.* 2018;80(4):597–623.
2. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika.* 1983;70(1):41–55.
3. Imbens G. The role of the propensity score in estimating dose-response functions. *Biometrika.* 2000;87(3):706–10.
4. Imai K, van Dyk DA. Causal inference with General Treatment regimes. *J Am Stat Assoc.* 2004;99(467):854–66.
5. Hirano K, Imbens GW. The propensity score with continuous treatments. In: *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives.* Chichester: John Wiley & Sons; 2004. p. 73–84.

6. Wu X, Mealli F, Kioumourtzoglou MA, Dominici F, Braun D. Matching on generalized propensity scores with continuous exposures. *J Am Stat Assoc.* 2022;119(545):757–72.
7. Zhang Z, Zhou J, Cao W, Zhang J. Causal inference with a quantitative exposure. *Stat Methods Med Res.* 2012;25(1):315–35.
8. Naimi AI, Moodie EE, Auger N, Kaufman JS. Constructing inverse probability weights for continuous exposures: a comparison of methods. *Epidemiology.* 2014;25(2):292–9.
9. Bang H, Robins JM. Doubly robust estimation in missing data and causal inference models. *Biometrics.* 2005;61(4):962–73.
10. Kennedy EH, Ma Z, McHugh MD, Small DS. Non-parametric methods for doubly robust estimation of continuous treatment effects. *J Royal Stat Soc Ser B.* 2017;79(4):1229–45.
11. Fong C, Hazlett C, Imai K. Covariate balancing propensity score for a continuous treatment: application to the efficacy of political advertisements. *Annals Appl Stat.* 2018;12(1):156–77.
12. Tübbicke S. Entropy balancing for continuous treatments. *J Econometric Methods.* 2022;11(1):71–89.
13. Vegetabile BG, Griffin BA, Coffman DL, Cefalu M, Robbins MW, McCaffrey DF. Nonparametric estimation of population average dose-response curves using entropy balancing weights for continuous exposures. *Health Serv Outcomes Res Method.* 2021;21(1):69–110.
14. Yu S, Su L. Covariate association eliminating weights: a unified weighting framework for causal effect estimation. *Biometrika.* 2018;105(3):709–22.
15. Huling JD, Greifer N, Chen G. Independence weights for causal inference with continuous treatments. *J Am Stat Assoc.* 2023;119(546):1657–70.
16. Koch B, Vock DM, Wolfson J. Covariate selection with group lasso and doubly robust estimation of causal effects. *Biometrics.* 2018;74(1):8–17.
17. Wilson A, Reich BJ. Confounder selection via penalized credible regions. *Biometrics.* 2014;70(4):852–61.
18. Ertefaie A, Asgharian M, Stephens DA. Variable selection in causal inference using a simultaneous penalization method. *J Causal Inference.* 2018;6(1):20170010.
19. Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Stürmer T. Variable selection for propensity score models. *Am J Epidemiol.* 2006;163(12):1149–56.
20. Zhu Y, Schonbach M, Coffman DL, Williams JS. Variable selection for propensity score estimation via balancing covariates. *Epidemiology.* 2015;26(2):e14–5.
21. Tang D, Kong D, Pan W, Wang L. Ultra-high dimensional variable selection for doubly robust causal inference. *Biometrics.* 2023;79(2):903–14.
22. Su L, Ura T, Zhang Y. Non-separable models with high-dimensional data. *J Econ.* 2019;212(2):646–77.
23. Colangelo K, Lee YY. Double debiased machine learning nonparametric inference with continuous treatments. *arXiv Preprint.* 2020. [arXiv:200403036](https://arxiv.org/abs/200403036).
24. Antonelli J, Papadogeorgou G, Dominici F. Causal inference in high dimensions: a marriage between bayesian modeling and good frequentist properties. *Biometrics.* 2022;78(1):100–14.
25. Gao Q, Zhang Y, Liang J, Sun H, Wang T. High-dimensional generalized propensity score with application to omics data. *Brief Bioinform.* 2021;22(6):bbab331.
26. Gábor JS, Maria LR, Nail KB. Measuring and testing dependence by correlation of distances. *Ann Stat.* 2007;35(6):2769–94.
27. Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. *J Educ Psychol.* 1974;66(5):688.
28. Edelman D, Fokianos K, Pitsillou M. An updated literature review of distance correlation and its applications to time series. *Int Stat Rev.* 2019;87(2):237–62.
29. Shortreed SM, Ertefaie A. Outcome-adaptive lasso: variable selection for causal inference. *Biometrics.* 2017;73(4):1111–22.
30. Van der Laan M, Polley E, Hubbard A. Super Learner, statistical applications in genetics and molecular biology. *Stat Appl Genet Mol Biol.* 2007;6(1):Article 25.
31. Gao Q, Zhang Y, Sun H, Wang T. Evaluation of propensity score methods for causal inference with high-dimensional covariates. *Brief Bioinform.* 2022;23(4):22727.
32. Tan Z. Regularized calibrated estimation of propensity scores with model misspecification and high-dimensional data. *Biometrika.* 2019;107(1):137–58.
33. Cepeda MS, Boston R, Farrar JT, Strom BL. Comparison of logistic regression versus propensity score when the number of events is low and there are multiple confounders. *Am J Epidemiol.* 2003;158(3):280–7.
34. Verbeek JH, Whaley P, Morgan RL, Taylor KW, Rooney AA, Schwingshackl L, Hoving JL, Vittal Katikireddi S, Shea B, Mustafa RA, et al. An approach to quantifying the potential importance of residual confounding in systematic reviews of observational studies: a GRADE concept paper. *Environ Int.* 2021;157:106868.
35. Schisterman EF, Cole SR, Platt RW. Overadjustment bias and unnecessary adjustment in epidemiologic studies. *Epidemiology.* 2009;20(4):488–95.
36. Kang JDY, Schafer JL. Demystifying double robustness: a comparison of alternative strategies for estimating a population mean from incomplete data. *Stat Sci.* 2007;22(4):523–39.
37. Smith AR, Smith RG, Pishva E, Hannon E, Roubroeks JA, Burrage J, Troakes C, Al-Sarraj S, Sloan C, Mill J. Parallel profiling of DNA methylation and hydroxymethylation highlights neuropathology-associated epigenetic variation in Alzheimer's disease. *Clin Epigenetics.* 2019;11(1):1–13.
38. Semick SA, Bharadwaj RA, Collado-Torres L, Tao R, Shin JH, Deep-Soboslay A, Weiss JR, Weinberger DR, Hyde TM, Kleinman JE. Integrated DNA methylation and gene expression profiling across multiple brain regions implicate novel genes in Alzheimer's disease. *Acta Neuropathol.* 2019;137:557–69.
39. Brokaw DL, Piras IS, Mastroeni D, Weisenberger DJ, Nolz J, Delvaux E, Serrano GE, Beach TG, Huentelman MJ, Coleman PD. Cell death and survival pathways in Alzheimer's disease: an integrative hypothesis testing approach utilizing-omic data sets. *Neurobiol Aging.* 2020;95:15–25.
40. Lunnon K, Smith R, Hannon E, De Jager PL, Srivastava G, Volta M, Troakes C, Al-Sarraj S, Burrage J, Macdonald R. Methylomic profiling implicates cortical deregulation of ANK1 in Alzheimer's disease. *Nat Neurosci.* 2014;17(9):1164–70.
41. Gasparoni G, Bultmann S, Lutsik P, Kraus TF, Sordon S, Vlcek J, Dietinger V, Steinmauer M, Haider M, Mulholland CB. DNA methylation analysis on purified neurons and glia dissects age and Alzheimer's disease-specific changes in the human cortex. *Epigenetics Chromatin.* 2018;11:1–19.
42. Smith RG, Hannon E, De Jager PL, Chibnik L, Lott SJ, Condliffe D, Smith AR, Haroutunian V, Troakes C, Al-Sarraj S. Elevated DNA methylation across a 48-kb region spanning the HOXA gene cluster is associated with Alzheimer's disease neuropathology. *Alzheimer's Dement.* 2018;14(12):1580–8.
43. Lardenoije R, Roubroeks JA, Pishva E, Leber M, Wagner H, Iatrou A, Smith AR, Smith RG, Eijssen LM, Kleideidam L. Alzheimer's disease-associated (hydroxy) methylomic changes in the brain and blood. *Clin Epigenetics.* 2019;11(1):1–15.
44. Shireby GL, Davies JP, Francis PT, Burrage J, Walker EM, Neilson GW, Dahir A, Thomas AJ, Love S, Smith RG. Recalibrating the epigenetic clock: implications for assessing biological age in the human cortex. *2020;143(12):3763–75.*
45. Dukes O, Vansteelandt S. How to obtain valid tests and confidence intervals after propensity score variable selection? *Stat Methods Med Res.* 2020;29(3):677–94.
46. Dukes O, Avagyan V, Vansteelandt S. Doubly robust tests of exposure effects under high-dimensional confounding. *Biometrics.* 2020;76(4):1190–200.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.