



Use of Generalized Propensity Scores for Assessing Effects of Multiple Exposures

Kecheng Li¹ · Tugba Akkaya-Hocagil¹ · Richard J. Cook¹ · Louise M. Ryan² · R. Colin Carter³ · Khue-Dung Dang⁴ · Joseph L. Jacobson⁵ · Sandra W. Jacobson⁵

Received: 3 April 2023 / Revised: 23 September 2023 / Accepted: 12 October 2023 /
Published online: 1 December 2023

© The Author(s) under exclusive licence to International Chinese Statistical Association 2023

Abstract

When interest lies in causal analysis of the effects of multiple exposures on an outcome, one may be interested in investigating the interaction between the exposures. In such settings, causal analysis requires modeling the joint distribution of exposures given pertinent confounding variables. In the most general setting, this may require modeling the effect of confounding variables on the association between exposures via a second-order regression model. We consider joint modeling of exposures for causal analysis via regression adjustment and inverse weighting. In both frameworks, we also investigate the asymptotic bias of estimators when the dependence model for the generalized propensity score incorrectly assumes conditional independence of exposures or is based on a naive dependence model which does not accommodate the effect of confounders on the conditional association of exposures. We also consider the problem of a semi-continuous bivariate exposure and propose a two-stage estimation technique to study the effects of prenatal alcohol exposure, and the effects of drinking frequency and intensity on childhood cognition.

Keywords Causal analysis · Dependence modeling · Generalized propensity score · Inverse weighting · Multiple exposures · Two-stage analysis

1 Introduction

1.1 Overview

Causal analysis is concerned with mitigating the effect of confounding variables on inferences regarding the exposure–disease relationship. With binary exposures, propensity scores capturing the association between confounders and the exposure of interest can be used for matching, stratification, or regression adjustment [1]. Inverse

probability of exposure weighting offers another framework for the estimation of causal effects. For continuous exposure variables, generalized propensity scores have been developed [2, 3].

The complex etiology of many chronic conditions motivates joint consideration of multiple exposure variables and their possible interacting effects [4]. Examples include the study of smoking and alcohol consumption on pancreatic cancer [5], the interacting effect of smoking and drinking on dietary intake of antioxidants on oral, pharyngeal, or laryngeal squamous cell carcinoma [6], and environmental studies of several different air pollutants and their effects on lung cancer risk [7]. When the joint effects of multiple exposures are under consideration, multivariate propensity scores can play a useful role. Vegetabile et al [8] describe a joint model for multivariate binary exposures through the specification of a model for a latent multidimensional Gaussian variable. Williams and Crespi [9] develop a generalized propensity score and use it to construct inverse multivariate density-weighted estimating functions. Chen and Zhou [10] propose the use of entropy balancing for multiple continuous treatments as a framework for causal analysis; this approach is best suited for settings in which the dimensions of the exposure and confounding variables are modest.

In settings with a single toxin, interest may lie in the causal effect of different exposure patterns defined in terms of the frequency of the exposure, the dose of the exposure on exposure occasions, and their interaction. Di Credico et al [11], for example, study the effect of the duration and intensity of smoking on the risk of head and neck cancers. This is the setting of our motivating study assessing the effect of maternal drinking patterns on childhood cognition. Here, even though previous studies show that high levels of prenatal alcohol exposure (PAE) can cause serious developmental problems in children [12–15], the impact of low levels of exposure is not well understood. Still less is known about the impact of “dose rate” which refers to the possibility that infants exposed to the same total volume of alcohol over the course of a pregnancy may have very different cognitive deficits depending on whether the mother drinks small amounts relatively frequently, or whether she drinks large amounts but less frequently (chronic versus binge drinking). This motivates the development and study of the methods to better understand possible complex causal effects of PAE on cognition outcomes.

When considering confounding in the causal analysis of multiple exposures, the effect of confounders is usually modeled on each marginal exposure variable with dependencies accommodated between the exposures through joint models - conditional associations are typically assumed to be functionally independent of any confounders. We consider a more general framework in which confounders may affect second moments (i.e., the relation between different features of exposures to a single toxin). Specifically, we develop methods to study the joint effect of the frequency of drinking and the dose of alcohol per drinking occasion in expectant mothers on the cognitive development of their child [16]; we then apply the methods we develop to the motivating study.

The remainder of the paper is as follows. In Sect. 2, we define a response model involving the main effects of two binary exposures and their interaction, along with a generalized propensity score. We formulate a general model for the exposures in

which confounders affect the first and second-order moments so that they explain how the two aspects of exposure vary together. We then describe how this can be used for regression adjustment and inverse weighting. The importance of modeling the effect of confounders on the association between exposures is investigated in Sect. 3 by evaluating the bias of causal estimands from regression adjustment and inverse probability of exposure weighting when naive propensity scores only deal with the marginal effects of confounders. In Sect. 4, we consider the case of continuous exposures including semi-continuous exposures and carry out an analogous investigation for analyses based on regression adjustment and inverse density weighting for estimating causal effects. A two-stage approach for estimating the causal effects of semi-continuous exposures is proposed in Sect. 5. An application to the Detroit Cohort Study [12] is given in Sect. 6 and concluding remarks and topics for future research are given in Sect. 7. We focus on the case of a bivariate exposure but as suggested by the notation in Sect. 2.1, the methods can be naturally generalized to accommodate higher dimensional exposures.

2 Causal Analysis with Multiple Exposures

2.1 Specification of the Full Model

Let Y denote a continuous response variable of interest, $\mathbf{A} = (A_1, \dots, A_p)'$ a $p \times 1$ vector of exposure variables with $p \geq 2$, and $\mathbf{B} = (A_1A_2, \dots, A_{p-1}A_p)'$ a $q \times 1$ vector of all pairwise products where $q = p(p - 1)/2$. Higher-order interaction terms can be considered, but here we focus on two-way interactions and so let $\mathbf{X} = (\mathbf{A}', \mathbf{B}')'$ contain the exposure data of interest. Furthermore, let $\mathbf{Z} = (Z_1, \dots, Z_r)'$ be an $r \times 1$ vector of confounding variables obscuring the effects of \mathbf{X} on the response Y . The data-generating model for the response is taken to have the form

$$Y = \mathbf{X}'\boldsymbol{\theta} + \mathbf{Z}'\boldsymbol{\zeta} + \epsilon, \tag{2.1}$$

where $\boldsymbol{\theta} = (\boldsymbol{\alpha}', \boldsymbol{\beta}')'$ with $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \dots, \alpha_p)'$ and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_q)'$, $\boldsymbol{\zeta} = (\zeta_1, \dots, \zeta_r)'$, and ϵ is a mean zero error term with $\mathbf{X}, \mathbf{Z} \perp \epsilon$. Let $\mathbf{A}_{(-j)} = (A_1, \dots, A_{j-1}, A_{j+1}, \dots, A_p)'$ be the vector of all exposure variables excluding A_j , let $\mathbf{B}_{(-j)}$ be the vector of all products of A_j with the $p - 1$ other exposures, and let $\boldsymbol{\beta}_{(-j)}$ be the corresponding vector of coefficients. The causal effect of a one unit change in A_j from $a_{j\circ}$ to $a_{j\circ} + 1$ given $\mathbf{A}_{(-j)} = \mathbf{a}_{(-j)}$ and \mathbf{Z} is $\alpha_j + \mathbf{a}'_{(-j)}\boldsymbol{\beta}_{(-j)}$. This model therefore enables one to study the causal interacting effects of the two exposures.

When $\dim\{\mathbf{Z}\} = r$ is large or one is concerned about getting the functional form of the effects of confounders correct in the response model, regression adjustment on the propensity score mitigates the effects of confounders. Let $f_{\mathbf{a}|\mathbf{z}}(\mathbf{a} | \mathbf{z})$ be the conditional joint probability distribution or mass function for multivariate continuous or discrete exposures given confounders. Imai and Van Dyk [17] defined the generalized propensity score (GPS) as $G(\mathbf{A}, \mathbf{Z}) = f_{\mathbf{a}|\mathbf{z}}(\mathbf{A} | \mathbf{Z})$ so that $\mathbf{Z} \perp I(\mathbf{A} = \mathbf{a}) | G(\mathbf{A}, \mathbf{Z}) = g$. Additionally, they stated that if there is a finite dimensional quantity $S(\mathbf{Z})$ such that $G(\mathbf{A}, \mathbf{Z})$ depends on \mathbf{Z} only through $S(\mathbf{Z})$, matching

or stratification on $S(\mathbf{Z}) = s$ or $G(\mathbf{A}, \mathbf{Z})$ is equivalent. For a single Bernoulli exposure variable, the generalized propensity score is the conventional propensity score, namely the conditional probability of being exposed given confounders (i.e., $P(A_1 = 1 | \mathbf{Z})$), whereas for a scalar continuous exposure it is the conditional density of the exposure $f_{a_1|z}(A_1 | \mathbf{Z})$; based on Assumption 3 of [17], it will often be sufficient to condition on a linear function of confounders such as $\mathbb{E}(A_1 | \mathbf{Z})$. For multivariate exposures, the generalized propensity score is the multivariate joint distribution of exposures given confounders, $G(\mathbf{A}, \mathbf{Z}) = f_{a|z}(\mathbf{A} | \mathbf{Z})$. When the dependence of \mathbf{A} on \mathbf{Z} in $G(\mathbf{A}, \mathbf{Z})$ can be expressed through a low dimensional function $S(\mathbf{Z})$, this generalized propensity score $S(\mathbf{Z})$ can be considered the effective propensity score; we consider this in what follows.

2.2 Propensity Scores for Two Binary Exposures

2.2.1 Regression Adjustment

Here, we aim to determine the minimal requirements for a potential vector-valued propensity score $S(\mathbf{Z})$ to ensure valid causal inference regarding the exposure effects conditional on $S(\mathbf{Z})$. We assume that data are available from independent individual processes and consider properties of a single process in much of what follows. From (2.1), note

$$\mathbb{E}_{Y|X,S}(Y) = \mathbb{E}_{Z|X,S}\{\mathbb{E}_{Y|X,S,Z}(Y)\} = \mathbf{X}'\boldsymbol{\theta} + \mathbb{E}_{Z|X,S}(\mathbf{Z}')\boldsymbol{\zeta}. \quad (2.2)$$

Using $P(\cdot)$ to informally denote a joint probability mass function or density, we write

$$P(\mathbf{Z} | \mathbf{X}, S) = \frac{P(\mathbf{X}, \mathbf{Z}, S)}{\int_{\mathbf{Z}} P(\mathbf{X}, \mathbf{Z}, S) d\mathbf{z}} = \frac{P(\mathbf{X} | \mathbf{Z}, S)P(\mathbf{Z}, S)}{\int_{\mathbf{Z}} P(\mathbf{X} | \mathbf{Z}, S)P(\mathbf{Z}, S) d\mathbf{z}}, \quad (2.3)$$

so if $S(\mathbf{Z})$ contains enough information to construct the conditional distribution for the exposure variables given \mathbf{Z} (i.e., if $\mathbf{X} \perp \mathbf{Z} | S$), then following Assumption 3 of [17], S serves as a valid propensity score. In this case, the final term in (2.2) is independent of \mathbf{X} and a working regression model of the form

$$\mathbb{E}(Y | \mathbf{X}, S) = \mathbf{X}'\boldsymbol{\lambda} + S'\boldsymbol{\eta} \quad (2.4)$$

will give an estimator $\hat{\boldsymbol{\lambda}}$ which is consistent for the causal parameter $\boldsymbol{\theta}$.

Here, we give an illustrative example involving a bivariate binary exposure $\mathbf{A} = (A_1, A_2)'$ where A_k indicates exposure status with respect to exposure k , $k = 1, 2$. Let \mathbf{Z} be a set of all potential confounders. Suppose the status for exposure k can be characterized by the marginal model $P(A_k | \mathbf{Z}) = P(A_k | \mathbf{Z}_k)$, where \mathbf{Z}_k is a subvector of \mathbf{Z} , $k = 1, 2$. We may assume, for example, that $\text{logit}P(A_k = 1 | \mathbf{Z}_k) = \mathbf{Z}_k'\boldsymbol{\gamma}_k$. While such marginal models are typically used when considering scalar exposures, they may not be sufficient to fully characterize the joint distribution of exposures if some confounding variables exert a second-order effect. Let \mathbf{Z}_3 be another set of confounders characterizing the dependence between A_1 and A_2 given \mathbf{Z}_1 and \mathbf{Z}_2 through a dependence regression model

$$\log \text{OR}(A_1, A_2 \mid \mathbf{Z}) = \mathbf{Z}'_3 \boldsymbol{\gamma}_3, \tag{2.5}$$

where

$$\text{OR}(A_1, A_2 \mid \mathbf{Z}) = \frac{P(A_1 = 1, A_2 = 1 \mid \mathbf{Z})/P(A_1 = 0, A_2 = 1 \mid \mathbf{Z})}{P(A_1 = 1, A_2 = 0 \mid \mathbf{Z})/P(A_1 = 0, A_2 = 0 \mid \mathbf{Z})}.$$

Based on the marginal models for $A_k \mid \mathbf{Z}_k, k = 1, 2$, and (2.5),

$$P(A_1 = 1, A_2 = 1 \mid \mathbf{Z}) = \begin{cases} \frac{a-[a^2-4\text{OR}(\text{OR}-1)P(A_1=1|\mathbf{Z}_1)P(A_2=1|\mathbf{Z}_2)]^{1/2}}{2(\text{OR}-1)}, & \text{OR} \neq 1 \\ P(A_1 = 1 \mid \mathbf{Z}_1)P(A_2 = 1 \mid \mathbf{Z}_2), & \text{OR} = 1 \end{cases}$$

where $a = 1 - (1 - \text{OR}) \sum_{k=1}^2 P(A_k = k \mid \mathbf{Z}_k)$ [18] and we let OR denote $\text{OR}(A_1, A_2 \mid \mathbf{Z})$. The generalized propensity score is obtained by noting that $P(A_1 = 1, A_2 = 0 \mid \mathbf{Z}) = P(A_1 = 1 \mid \mathbf{Z}_1) - P(\mathbf{A} = \mathbf{1} \mid \mathbf{Z})$ and $P(A_1 = 0, A_2 = 1 \mid \mathbf{Z}) = P(A_2 = 1 \mid \mathbf{Z}_2) - P(\mathbf{A} = \mathbf{1} \mid \mathbf{Z})$, from which $P(\mathbf{A} = \mathbf{0} \mid \mathbf{Z})$ can be obtained. A sufficient propensity score would be $\mathbf{S}(\mathbf{Z}) = (S_1(\mathbf{Z}_1), S_2(\mathbf{Z}_2), S_3(\mathbf{Z}_3))'$ where $S_k(\mathbf{Z}_k) = P(A_k = 1 \mid \mathbf{Z}_k), k = 1, 2$, and $S_3(\mathbf{Z}) = P(\mathbf{A} = \mathbf{1} \mid \mathbf{Z})$; alternatively, one could define $S_k(\mathbf{Z}_k) = \mathbf{Z}'_k \boldsymbol{\gamma}_k, k = 1, 2, 3$.

Here, \mathbf{Z}_1 and \mathbf{Z}_2 are sets of confounders influencing the marginal causal effects of A_1 and A_2 on Y , and conditioning on (S_1, S_2) is sufficient if A_1 does not interact with A_2 (i.e., if $\beta_1 = 0$). If $\beta_1 \neq 0$, then it is necessary to model the association between A_1 and A_2 given $(\mathbf{Z}'_1, \mathbf{Z}'_2)'$ as a function of possibly different sets of confounders \mathbf{Z}_3 as suggested in (2.5). We investigate the implications of not doing this adequately in Section 3.

2.2.2 Inverse Probability Weighting

Here, we consider the use of a generalized propensity score as weight in inverse probability weighting analysis for two binary exposures. Let $P(\mathbf{A} \mid \mathbf{Z}; \boldsymbol{\gamma})$ denote the conditional joint probability of $\mathbf{A} = (A_1, A_2)'$ given confounders under the positivity condition [1] where $P(\mathbf{A} = \mathbf{a} \mid \mathbf{Z}; \boldsymbol{\gamma}) > 0$ for all possible values of \mathbf{Z} . The inverse ‘‘joint probability’’-weighted estimating function [19] is expressed as

$$U_1(\boldsymbol{\varphi}; \boldsymbol{\gamma}) = \frac{I(\mathbf{A} = \mathbf{a})}{P(\mathbf{A} = \mathbf{a} \mid \mathbf{Z}; \boldsymbol{\gamma})} D(\mathbf{X}) [Y - \mathbb{E}^R(Y \mid \mathbf{X}; \boldsymbol{\varphi})], \tag{2.6}$$

where $\mathbb{E}^R(\cdot)$ represents an expectation in the setting where $\mathbf{X} \perp \mathbf{Z}$ and $\mathbb{E}^R(Y \mid \mathbf{X}; \boldsymbol{\varphi}) = \mathbf{X}' \boldsymbol{\varphi}$ defines a marginal structural model. Here,

$$D(\mathbf{X}) = \frac{\partial}{\partial \boldsymbol{\varphi}} \mathbb{E}^R(Y \mid \mathbf{X}; \boldsymbol{\varphi}) = (1, A_1, A_2, A_1 A_2)'$$

To show that (2.6) is an unbiased estimating function, we take the expectation of $U_1(\boldsymbol{\varphi}; \boldsymbol{\gamma})$ with respect to $Y \mid \mathbf{X}, \mathbf{Z}$ to obtain

$$\frac{I(\mathbf{A} = \mathbf{a})}{P(\mathbf{A} = \mathbf{a} \mid \mathbf{Z}; \boldsymbol{\gamma})} D(\mathbf{X}) [\mathbb{E}(Y \mid \mathbf{X}, \mathbf{Z}; \boldsymbol{\Psi}) - \mathbb{E}^R(Y \mid \mathbf{X}; \boldsymbol{\varphi})], \tag{2.7}$$

where $\Psi = (\theta', \zeta')'$. The expectation of (2.7) with respect to $Z | A = a$ gives

$$\sum_Z \frac{I(A = a)}{P(A = a | Z; \gamma)} D(X) [\mathbb{E}(Y | X, Z; \Psi) - \mathbb{E}^R(Y | X; \varphi)] P(Z | A = a). \tag{2.8}$$

Since

$$P(Z | A = a) = \frac{P(A = a | Z; \gamma)P(Z; \tau)}{P(A = a)},$$

we can rewrite (2.8) as

$$\sum_Z \frac{I(A = a)}{P(A = a)} D(X) [\mathbb{E}(Y | X, Z; \Psi) - \mathbb{E}^R(Y | X; \varphi)] P(Z; \tau),$$

which effectively results in the expectation being taken in the setting where $Z \perp A$, equivalent to how this would be done in the hypothetical randomized trial [19]. Since (2.6) has expectation zero, solving $U_1(\varphi; \gamma) = 0$ defines an estimator that is consistent for the causal estimand of interest. In practice, γ must be estimated but provided a \sqrt{n} -consistent estimator is used in its place, consistency of $\hat{\varphi}$ is still achieved.

3 Exposure Dependence Model Misspecification

3.1 Regression Adjustment

3.1.1 General Framework

Here, we derive the limiting bias of the estimators of exposure effects obtained through regression adjustment when the dependence model is misspecified through a) an inappropriate conditional independence assumption, or b) specification of an exchangeable dependence structure functionally independent of confounders; this is defined more formally in what follows.

Consider the following regression model with a generalized propensity score adjustment,

$$\mathbb{E}(Y | X, \tilde{S}) = X' \varphi + \tilde{S}' \psi, \tag{3.1}$$

where $\tilde{S} = \tilde{S}(Z)$ is a vector-valued propensity score which might be insufficient for adjustment under model misspecification, e.g., assuming conditional independence among exposures or using an inadequate dependence model. The contribution to the least squares equation for φ from a single individual based on (3.1) is $U(Y | X, \tilde{S}; \varphi, \psi) = (X', \tilde{S}')'(Y - X' \varphi - \tilde{S}' \psi)$. With a sample of n independent individuals yielding data $\{(Y_i, X_i, Z_i), i = 1, \dots, n\}$, the least squares estimator of the exposure effect is the solution to $\sum_{i=1}^n U(Y_i | X_i, \tilde{S}_i; \varphi, \psi) = 0$ and is denoted by $\hat{\varphi}$; this is consistent for the solution to $\mathbb{E}\{U(Y | X, \tilde{S}; \varphi, \psi)\} = 0$ where the expectation is taken with respect to the full data generation process [20, 21].

Again we consider the bivariate case with $X = (1, A_1, A_2, A_1A_2)'$ and the corresponding coefficient $\varphi = (\varphi_0, \varphi_1, \varphi_2, \varphi_3)'$. The explicit form of the limiting value of the estimator $\tilde{\varphi}$ for θ in (2.1) is

$$\tilde{\varphi}_1 = \alpha_1 + \zeta' \Delta_{11} + \psi' \Delta_{21}, \tag{3.2}$$

$$\tilde{\varphi}_2 = \alpha_2 + \zeta' \Delta_{12} + \psi' \Delta_{22}, \tag{3.3}$$

$$\begin{aligned} \tilde{\varphi}_3 = & \beta_1 + \zeta' \left(\frac{\Sigma_3}{\sigma_3^2} - \Delta_{11} \frac{\rho_{13}\sigma_1}{\sigma_3} - \Delta_{12} \frac{\rho_{23}\sigma_2}{\sigma_3} \right) \\ & - \psi' \left(\frac{\Lambda_3}{\sigma_3^2} + \Delta_{21} \frac{\rho_{13}\sigma_1}{\sigma_3} + \Delta_{22} \frac{\rho_{23}\sigma_2}{\sigma_3} \right), \end{aligned} \tag{3.4}$$

where ζ and ψ are given in (2.1) and (3.1),

$$\begin{aligned} \Delta_{1k} = & \frac{(\rho_{12} - \rho_{13}\rho_{23})\left(\frac{\sigma_k^2 \Sigma_{3-k}}{\sigma_1\sigma_2} - \frac{\sigma_k}{\sigma_3} \frac{\rho_{13}\rho_{23}}{\rho_{k3}} \Sigma_3\right) + \left(\frac{\rho_{13}^2 \rho_{23}^2}{\rho_{k3}^2} - 1\right)\left(\Sigma_k - \frac{\sigma_k}{\sigma_3} \rho_{k3} \Sigma_3\right)}{\sigma_k^2 \{(\rho_{12} - \rho_{13}\rho_{23})^2 - (1 - \rho_{13}^2)(1 - \rho_{23}^2)\}}, \\ \Delta_{2k} = & \frac{(\rho_{12} - \rho_{13}\rho_{23})\left(\frac{\sigma_k^2 \Lambda_{3-k}}{\sigma_1\sigma_2} - \frac{\sigma_k}{\sigma_3} \frac{\rho_{13}\rho_{23}}{\rho_{k3}} \Lambda_3\right) + \left(\frac{\rho_{13}^2 \rho_{23}^2}{\rho_{k3}^2} - 1\right)\left(\Lambda_k - \frac{\sigma_k}{\sigma_3} \rho_{k3} \Lambda_3\right)}{\sigma_k^2 \{(\rho_{12} - \rho_{13}\rho_{23})^2 - (1 - \rho_{13}^2)(1 - \rho_{23}^2)\}}, \end{aligned}$$

$\Sigma_k = \text{cov}(\mathbb{E}(A_k | Z), Z)$, $\Lambda_k = \text{cov}(\mathbb{E}(A_k | Z), \tilde{S})$, $\Sigma_3 = \text{cov}(\mathbb{E}(A_1A_2 | Z), Z)$, and $\Lambda_3 = \text{cov}(\mathbb{E}(A_1A_2 | Z), \tilde{S})$. We also have $\sigma_k^2 = \text{var}(A_k)$, $\sigma_3^2 = \text{var}(A_1A_2)$, $\rho_{12} = \text{corr}(A_1, A_2)$, and $\rho_{k3} = \text{corr}(A_k, A_1A_2)$, $k = 1, 2$; see Online Appendix A for more details.

With two binary exposures, $\sigma_k^2 = p_k(1 - p_k)$ and $\sigma_3^2 = p_{12}(1 - p_{12})$, where $p_k = P(A_k = 1)$, $k = 1, 2$, and $p_{12} = P(A_1A_2 = 1)$. The correlation ρ_{12} is

$$\rho_{12} = \frac{p_{12} - p_1p_2}{\sqrt{p_1(1 - p_1)p_2(1 - p_2)}}$$

[22], and

$$\rho_{k3} = \frac{p_{12} - p_kp_{12}}{\sqrt{p_k(1 - p_k)p_{12}(1 - p_{12})}}, \quad k = 1, 2.$$

3.1.2 A Misspecified Dependence Model

If exposure variables are treated as conditionally independent given confounders, one might specify \tilde{S} as $\tilde{S} = (S_1(Z_1), S_2(Z_2))'$ or $\tilde{S} = (S_1(Z_1), S_2(Z_2), S_1(Z_1)S_2(Z_2))'$, where $S_1(Z_1)$ and $S_2(Z_2)$ are the marginal propensity scores. A naive model using regression adjustment might be specified as

$$\mathbb{E}(Y \mid X, S) = X' \boldsymbol{\varphi} + \psi_1 S_1(\mathbf{Z}_1) + \psi_2 S_2(\mathbf{Z}_2), \tag{3.5}$$

or possibly

$$\mathbb{E}(Y \mid X, S) = X' \boldsymbol{\varphi} + \psi_1 S_1(\mathbf{Z}_1) + \psi_2 S_2(\mathbf{Z}_2) + \psi_3 S_1(\mathbf{Z}_1) S_2(\mathbf{Z}_2). \tag{3.6}$$

In the most general setting, such propensity score regression adjustments are insufficient to render $X \perp Z \mid \tilde{S}$ since $A_1 A_2$ depends on the joint model of A_1 and A_2 given Z . As a result, estimates of the causal effects φ_1, φ_2 and φ_3 are inconsistent with the asymptotic bias given by the second and third terms on the right-hand side of Eqs. (3.2)–(3.4). It can be seen that the bias depends on the second-order association between the exposures and confounders, the association between exposures and generalized propensity scores, and variances of exposures.

A more flexible exposure model may recognize the conditional association between exposures (given marginal confounders Z_1 and Z_2), but assume no role of confounders on the second-order dependence. In this case, one might consider a propensity score term based on

$$\mathbb{E}(A_1 A_2 \mid Z) = \text{cov}(A_1, A_2 \mid Z_1, Z_2) + S_1(\mathbf{Z}_1) S_2(\mathbf{Z}_2),$$

where $\text{cov}(A_1, A_2 \mid Z_1, Z_2) = \rho \sqrt{V(A_1 \mid Z_1) V(A_2 \mid Z_2)}$ and $V(A_k \mid Z_k)$ is the residual variance from the propensity score for exposure $k, k = 1, 2$. In this case, the regression model adjusted by the generalized propensity score may be,

$$\mathbb{E}(Y \mid X, S) = X' \boldsymbol{\varphi} + \psi_1 S_1(\mathbf{Z}_1) + \psi_2 S_2(\mathbf{Z}_2) + \psi_3 \mathbb{E}(A_1 A_2 \mid Z_1, Z_2). \tag{3.7}$$

We explore the impact of such misspecifications in Sect. 3.3.

3.2 Inverse Probability Weighting

A misspecified propensity score will have a different impact in the framework of the inverse propensity score-weighted estimating function. Let $P_0(A = a \mid Z; \delta)$ denote the misspecified (naïve) conditional joint density or mass function indexed by δ , and let $U_{10}(\boldsymbol{\varphi}; \delta)$ and $U_{20}(\delta)$ be the estimating functions as follows:

$$U_{10}(\boldsymbol{\varphi}; \delta) = \frac{I(A = a)}{P_0(A = a \mid Z; \delta)} D(X) [Y - \mathbb{E}^R(Y \mid X; \boldsymbol{\varphi})], \tag{3.8}$$

$$U_{20}(\delta) = \frac{\partial \log P_0(A \mid Z; \delta)}{\partial \delta}. \tag{3.9}$$

Let δ^* be the solution to $\mathbb{E}_{A,Z}\{U_{20}(\delta); \boldsymbol{\gamma}, \boldsymbol{\tau}\} = 0$, the limiting value to which the estimator $\hat{\delta}$ converges when fitting the misspecified model. The limiting value of $\boldsymbol{\varphi}$ is obtained by setting the equation below to zero and solving $\hat{\boldsymbol{\varphi}}$,

$$\sum_A \sum_Z \frac{I(A = a) P(A = a \mid Z; \boldsymbol{\gamma})}{P_0(A = a \mid Z; \delta^*) P(Z; \boldsymbol{\tau})} D(X) [\mathbb{E}(Y \mid X, Z; \boldsymbol{\Psi}) - \mathbb{E}^R(Y \mid X; \boldsymbol{\varphi})]. \tag{3.10}$$

A misspecified propensity score $P_0(A = a \mid Z; \delta)$ is the product of two marginal propensity scores $P(A_1 = a_1 \mid Z; \delta)P(A_2 = a_2 \mid Z; \delta)$ under the conditional independence assumption. In this case, under the present formulation, we have

$$U_{20}(\delta) = (U'_{20,1}(\gamma_1), U'_{20,2}(\gamma_2)),$$

where

$$U_{20,k}(\gamma_k) = \frac{\partial \log P(A_k = a_k \mid Z_k; \gamma_k)}{\partial \gamma_k}, \quad k = 1, 2,$$

and $\delta^* = (\gamma'_1, \gamma'_2)'$ so $\hat{\delta}$ is consistent for the target parameters γ_1 and γ_2 but is of insufficient dimension to characterize the second-order dependence.

When the association is assumed functionally independent of confounders, the misspecified propensity score $P_0(A = a \mid Z; \delta)$ is an inadequately modeled conditional joint density or mass function $P(A_1 = a_1, A_2 = a_2 \mid Z_1, Z_2; \delta)$. In this case, we have $U_{20}(\delta) = \partial \log P(A = a \mid Z_1, Z_2; \delta) / \partial \delta$, and δ^* will not be consistent for the target parameters γ_1 or γ_2 .

3.3 Empirical Bias and Efficiency

Here, we report on simulation studies investigating the limiting values of the exposure effects subject to model misspecification. We consider a bivariate binary exposure and assess the relative asymptotic bias of estimates of exposure effects under two types of model misspecification of Sect. 3.1.2 under both regression adjustment and IPW analyses.

Suppose a sample of n independent replicates $\{(Y_i, X_i, Z_i), i = 1 \dots n\}$ are generated with $n = 500$. Let Z contain binary confounding variables with $Z_1 \perp Z_2$, $P(Z_1 = 1) = P(Z_2 = 1) = 0.5$, and $P(Z_3 = 1 \mid Z_1, Z_2) = \text{expit}\{-0.896 + (\log 2)Z_1 + (\log 3)Z_2\}$ giving $P(Z_3 = 1) = 0.5$. We let $P(A_1 = 1 \mid Z_1) = \text{expit}(\gamma_{10} + \gamma_{11}Z_1)$ where $\gamma_1 = (\gamma_{10}, \gamma_{11})' = (-0.764, \log 2)'$ so $\mathbb{E}(A_1) = 0.4$, and $P(A_2 = 1 \mid Z_2) = \text{expit}(\gamma_{20} + \gamma_{21}Z_2)$ with $\gamma_2 = (\gamma_{20}, \gamma_{21})' = (-0.112, \log 1.25)'$ so $\mathbb{E}(A_2) = 0.5$. Finally, we let $\log \text{OR}(A_1, A_2 \mid Z_3) = \gamma_{30} + \gamma_{31}Z_3$ with $\gamma_3 = (\gamma_{30}, \gamma_{31})'$; we consider $\gamma_{31} \in [-5, 5]$ and for each value of γ_{31} , we solve for γ_{30} in (2.5) such that $\mathbb{E}\{\text{OR}(A_1, A_2)\} = 4$. For response Y , we generate data according to (2.1) with $\theta = (\alpha_0, \alpha_1, \alpha_2, \beta_1)' = (0, 2, 2, 4)'$ and $\zeta = (0.5, 0.5, 0.5)'$.

Under the assumption of conditional independence between exposures, the regression model adjusted by the generalized propensity scores can be fitted as in (3.5) or (3.6), which is denoted as Reg1 and Reg2 in Fig. 1, respectively. The IPW approach involves solving the exposure effect in (3.10). We denote the estimated effect as $\hat{\phi} = (\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3)'$ for $X = (A_1, A_2, A_1A_2)'$, and the relative bias of the estimated exposure effect for X is shown in Fig. 1 as $(\hat{\phi}_1 - \alpha_1)/\alpha_1$, $(\hat{\phi}_2 - \alpha_2)/\alpha_2$, and $(\hat{\phi}_3 - \beta_1)/\beta_1$. In the bottom right panel of Fig. 1, we show the relative bias of the causal effect of a one unit change in A_2 given $A_1 = a_1 = 1$ or 0 and Z , and denote it as $(\hat{\Delta}_2(A_1) - \Delta_2(A_1))/\Delta_2(A_1)$ where $\Delta_2(A_1) = \alpha_2 + a_1\beta_1$ and $\hat{\Delta}_2(A_1) = \hat{\phi}_2 + a_1\hat{\phi}_3$.

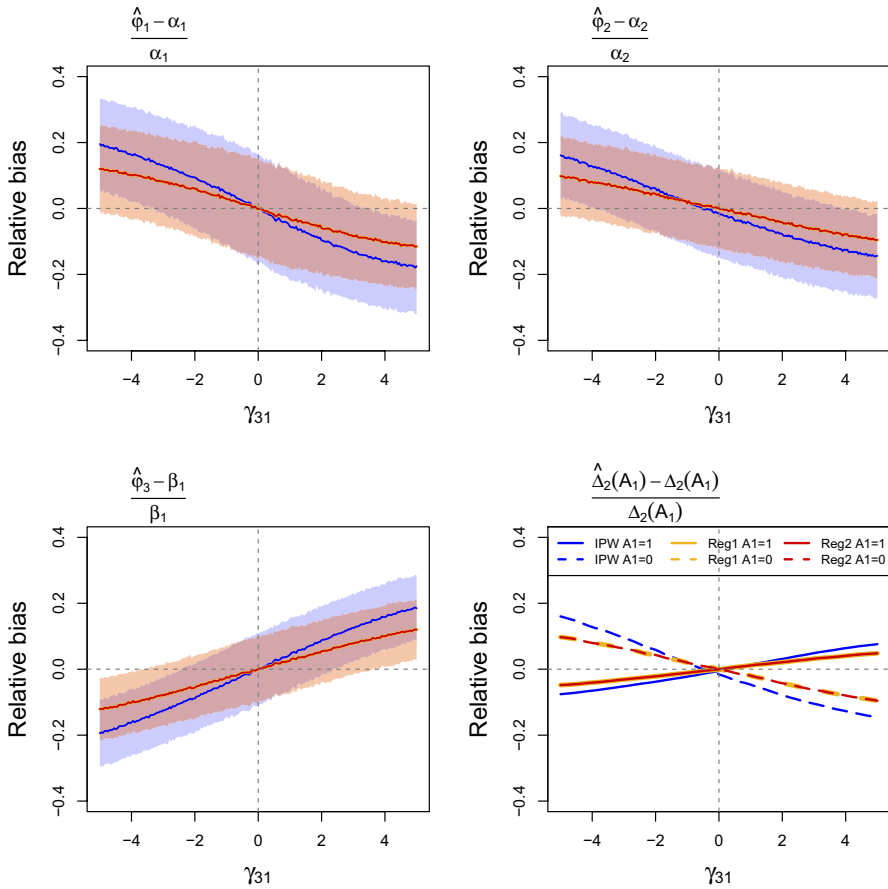


Fig. 1 Empirical relative bias with standard errors (reflected by shaded bands) of estimators of exposure effects from inadequate propensity score modeling assuming conditional independence between two binary exposures when the second-order confounder has the effect γ_{31} ; regression adjustment (Reg) and inverse weighting (IPW) are considered; the marginal association between exposures is OR = 4

The absolute value of relative bias of the exposure effect increases as the covariate effect on the exposure dependence increases. Moreover, the regression adjustment leads to less bias in effect estimation than the IPW approach in these settings. Adjusting for $(S_1(\mathbf{Z}), S_2(\mathbf{Z}))'$ or $(S_1(\mathbf{Z}), S_2(\mathbf{Z}), S_1(\mathbf{Z})S_2(\mathbf{Z}))'$ produces similar results. When $\gamma_{31} = 0$ (i.e., the conditional association between exposure is functionally independent of confounders), the bias of the estimated exposure effect is negligible when conditional independence is assumed between two binary exposures.

We follow (3.7) and (3.10) to conduct the estimation for the second type of misspecification, that is, using an exchangeable dependence structure which assumes no second-order effects of confounders on the exposure. The results are shown in Fig. 2. When there is no confounder affecting the dependence (i.e., $\gamma_{31} = 0$), the curves pass through zero points on both axes, indicating both the regression adjustment and IPW

approaches produce unbiased estimates as the models are correctly specified under the assumption of no second-order effects of confounders on the exposure. Similarly as in Fig. 1, the absolute value of relative bias of the exposure effect would increase as the covariate effect on the exposure dependence (γ_{31}) increases. The regression adjustment leads to less bias in effect estimation compared to the IPW approach.

4 Multivariate Continuous Exposures

In this section, we consider multiple continuous exposures and semi-continuous exposures.

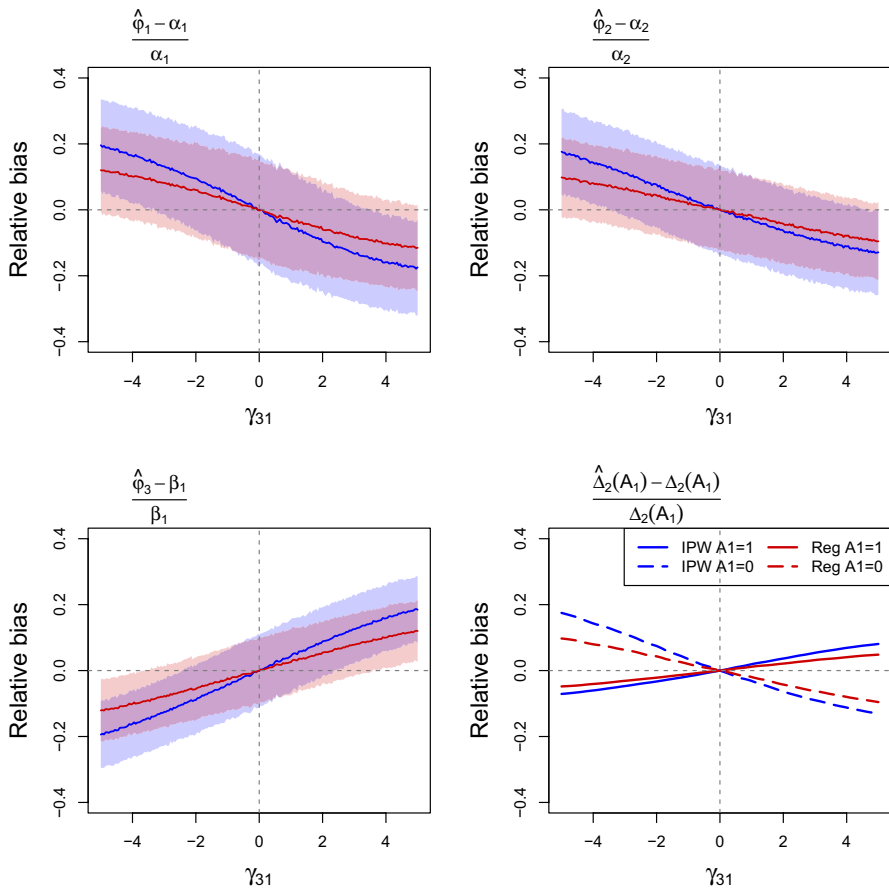


Fig. 2 Empirical relative bias with standard errors (reflected by shaded bands) of estimators of exposure effects from inadequate propensity score modeling due to the omission of a second-order confounder with effect γ_{31} ; regression adjustment (Reg) and inverse weighting (IPW) are considered; the marginal association between two binary exposures is OR = 4

4.1 Multivariate Normal Exposures

4.1.1 Regression Adjustment

Here, we consider $A \mid Z \sim \text{BVN}(\mu(Z), \Sigma(Z))$ with $A = (A_1, A_2)'$, $Z = (Z'_1, Z'_2, Z'_3)'$, $\mu(Z) = (\mu_1(Z), \mu_2(Z))'$, $\mu_k(Z) = \mathbb{E}(A_k \mid Z) = \mathbb{E}(A_k \mid Z_k)$, $\text{var}(A_k \mid Z) = \sigma_k^2$, $k = 1, 2$, and $\text{corr}(A_1, A_2 \mid Z) = \rho(Z_3)$. We suppose $\mathbb{E}(A_k \mid Z) = Z'_k \gamma_k$ with a dependence model

$$\log \left\{ \frac{1 + \rho(Z_3)}{1 - \rho(Z_3)} \right\} = Z'_3 \gamma_3. \tag{4.1}$$

Since the variance parameters are functionally independent of Z , a sufficient propensity score is $S(Z) = (S_1(Z_1), S_2(Z_2), S_3(Z_3))'$ where $S_k(Z_k) = Z'_k \gamma_k$, $k = 1, 2$ and $S_3(Z) = Z'_3 \gamma_3$ [17] since

$$f(a_1, a_2 \mid Z, S) = \frac{1}{2\pi \sqrt{|\Sigma(S_3)|}} \exp\{- (a - S)' \Sigma^{-1}(S_3) (a - S) / 2\},$$

where $(a - S) = (a_1 - S_1, a_2 - S_2)'$ and

$$\Sigma(S_3) = \begin{bmatrix} \sigma_1^2 & \rho(Z_3) \sigma_1 \sigma_2 \\ \rho(Z_3) \sigma_1 \sigma_2 & \sigma_2^2 \end{bmatrix}.$$

An alternative to (4.1) would be to model $B_1 = A_1 A_2$ given Z directly to define $S_3(Z) = \mathbb{E}(B_1 \mid Z_3)$ and note that in this case, we could replace the covariance entries of Σ with $S_3(Z) - S_1(Z_1) S_2(Z_2)$; the key point is that $S(Z) = (S_1(Z_1), S_2(Z_2), S_3(Z_3))'$ contains the relevant information on confounders in the exposure model.

4.1.2 Inverse Joint Density Weighting

Inverse density weighting can be used when exposures are continuous [23]. With conditional joint density $f(a \mid z; \gamma)$ as in (2.6), we write the inverse joint density-weighted estimating function as

$$U_1(\varphi; \gamma) = \frac{I(A = a)}{f(a \mid z; \gamma)} D(X) [Y - \mathbb{E}^R(Y \mid X; \varphi)]. \tag{4.2}$$

The expectation of $U_1(\varphi; \gamma)$ with respect to $Y \mid X, Z$ and $Z \mid X$ gives

$$\int \frac{I(A = a)}{f(a \mid z; \gamma)} D(X) [\mathbb{E}(Y \mid X, Z; \theta, \xi) - \mathbb{E}^R(Y \mid X; \varphi)] f(z \mid a) dz, \tag{4.3}$$

where

$$f(z \mid a) = \frac{f(a \mid z; \gamma) f(z; \tau)}{f(a)}.$$

Thus, we can rewrite (4.3) as

$$\int_{\mathbf{Z}} \frac{I(\mathbf{A} = \mathbf{a})}{f(\mathbf{a})} D(\mathbf{X}) [\mathbb{E}(Y | \mathbf{X}, \mathbf{Z}; \boldsymbol{\theta}, \boldsymbol{\zeta}) - \mathbb{E}^R(Y | \mathbf{X}; \boldsymbol{\varphi})] f(\mathbf{z}; \boldsymbol{\tau}) d\mathbf{z}, \tag{4.4}$$

which results in the expectation being taken in the setting where $\mathbf{Z} \perp \mathbf{A}$ and gives consistent causal estimate $\hat{\boldsymbol{\varphi}}$ from solving $U_1(\boldsymbol{\varphi}; \boldsymbol{\gamma}) = 0$.

4.2 Multivariate Semi-continuous Exposures

4.2.1 Regression Adjustment

Here we suppose A_1 and A_2 are non-negative semi-continuous exposure variables with a mass at zero and we adopt a two-part generalized propensity score model [24] to address the effects of confounding variables. While these two exposures could in general represent exposures to different agents, here we consider the case where they represent different aspects of the exposure to a single agent. Specifically for motivating study, we may let A_1 and A_2 represent the drinking frequency (proportion of drinking days per week) and drinking intensity (dose per drinking occasion) – among non-drinkers, these degenerate to zero. We can therefore let $A^+ = \mathbb{I}(A_1 A_2 > 0)$ indicate positive exposure, $L_k = \log(A_k | A^+ = 1)$, $k = 1, 2$, and $\pi_1(\mathbf{Z}_0) = \mathbb{E}(A^+ | \mathbf{Z}_0)$.

We consider the true response model is (2.1) with $\mathbf{X} = (1, A^+, A^+L_1, A^+L_2, A^+L_1L_2)'$. We let

$$\pi_1(\mathbf{Z}_0) = \text{expit}(\mathbf{Z}'_0 \boldsymbol{\gamma}_0),$$

where $\mathbf{Z}_0 = (1, \mathbf{Z}'_1, \mathbf{Z}'_2)'$, and given $A^+ = 1$, we suppose $L_k = \mathbf{Z}'_k \boldsymbol{\gamma}_k + R_k$, $k = 1, 2$, and $\mathbf{R} = (R_1, R_2)'$ is bivariate normal with $\text{var}(R_k) = \sigma_k^2$, $k = 1, 2$, $\rho(\mathbf{Z}) = \text{corr}(R_1, R_2 | \mathbf{Z})$, and $\log\{(1 + \rho(\mathbf{Z}) / (1 - \rho(\mathbf{Z})))\} = \mathbf{Z}'_3 \boldsymbol{\gamma}_3$. As such, we have

$$\begin{aligned} \mathbb{E}(L_k | \mathbf{Z}_0, \mathbf{Z}_k) &= \pi_1(\mathbf{Z}_0) \mathbb{E}(L_k | A^+ = 1, \mathbf{Z}_k), \quad k = 1, 2, \\ \mathbb{E}(L_1 L_2 | \mathbf{Z}) &= \pi_1(\mathbf{Z}_0) \mathbb{E}(L_1 L_2 | A^+ = 1, \mathbf{Z}) \end{aligned}$$

to form the generalized propensity score as $\mathbf{S} = (\pi_1(\mathbf{Z}_0), \mathbb{E}(L_1 | \mathbf{Z}_0, \mathbf{Z}_1), \mathbb{E}(L_2 | \mathbf{Z}_0, \mathbf{Z}_2), \mathbb{E}(L_1 L_2 | \mathbf{Z}))'$ or $\mathbf{S} = (S_0(\mathbf{Z}_0), S_1(\mathbf{Z}_1), S_2(\mathbf{Z}_2), S_3(\mathbf{Z}_3))'$ and use it for adjustment in the regression model (2.4) as

$$\mathbb{E}(Y | \mathbf{X}, \mathbf{S}) = \mathbf{X}' \boldsymbol{\lambda} + \mathbf{S}' \boldsymbol{\eta}, \tag{4.5}$$

where $\mathbf{X} = (1, A^+, A^+L_1, A^+L_2, A^+L_1L_2)'$.

4.2.2 Inverse Joint Density Weighting

The inverse density of exposure weighted estimating function is given here by

$$U_0(Y | \mathbf{X}; \boldsymbol{\varphi}, \boldsymbol{\Gamma}, \boldsymbol{\gamma}_0) = w(\mathbf{X}, \mathbf{Z}) D(\mathbf{X}) [Y - \mathbb{E}^R(Y | \mathbf{X}; \boldsymbol{\varphi})], \tag{4.6}$$

where $\boldsymbol{\Gamma} = (\boldsymbol{\gamma}'_1, \sigma_1^2, \boldsymbol{\gamma}'_2, \sigma_2^2, \boldsymbol{\gamma}'_3)'$,

$$w(\mathbf{X}, \mathbf{Z}) = \frac{\mathbb{I}(\mathbf{X} = \mathbf{x})}{\{\pi_1(\mathbf{Z}_0; \boldsymbol{\gamma}_0) f(l_1, l_2 | A^+ = 1, \mathbf{Z})\}^{A^+} \{1 - \pi_1(\mathbf{Z}_0; \boldsymbol{\gamma}_0)\}^{1-A^+}},$$

and $D(\mathbf{X}) = (1, A^+, A^+L_1, A^+L_2, A^+L_1L_2)'$.

Stabilization can be used to address extreme weights [19, 25] where

$$w^s(\mathbf{X}, \mathbf{Z}) = \frac{\mathbb{I}(\mathbf{X} = \mathbf{x}) \{\pi_2(\boldsymbol{\gamma}_0^\dagger) f(l_1, l_2 | A^+ = 1)\}^{A^+} \{1 - \pi_2(\boldsymbol{\gamma}_0^\dagger)\}^{1-A^+}}{\{\pi_1(\mathbf{Z}_0; \boldsymbol{\gamma}_0) f(l_1, l_2 | A^+ = 1, \mathbf{Z})\}^{A^+} \{1 - \pi_1(\mathbf{Z}_0; \boldsymbol{\gamma}_0)\}^{1-A^+}},$$

with $\pi_2(\boldsymbol{\gamma}_0^\dagger)$ and $f(l_1, l_2 | A^+ = 1)$ unconditional quantities with respect to the conditional quantities $\pi_1(\mathbf{Z}_0; \boldsymbol{\gamma}_0)$ and $f(l_1, l_2 | A^+ = 1, \mathbf{Z})$ for weights stabilization. The inverse weighted estimating function for $\boldsymbol{\varphi}$ with stabilized weights is given by (4.6) with $w^s(\mathbf{X}, \mathbf{Z})$ in place of $w(\mathbf{X}, \mathbf{Z})$ giving

$$U_0(Y | \mathbf{X}; \boldsymbol{\varphi}, \boldsymbol{\Gamma}, \boldsymbol{\Gamma}^\dagger, \boldsymbol{\gamma}_0, \boldsymbol{\gamma}_0^\dagger) = w^s(\mathbf{X}, \mathbf{Z}) D(\mathbf{X}) [Y - \mathbb{E}(Y | \mathbf{X}; \boldsymbol{\varphi})]. \tag{4.7}$$

The estimate of $\boldsymbol{\varphi}$ is the solution from solving $U_0(Y | \mathbf{X}; \boldsymbol{\varphi}, \boldsymbol{\Gamma}, \boldsymbol{\gamma}_0) = 0$ or $U_0(Y | \mathbf{X}; \boldsymbol{\varphi}, \boldsymbol{\Gamma}, \boldsymbol{\Gamma}^\dagger, \boldsymbol{\gamma}_0, \boldsymbol{\gamma}_0^\dagger) = 0$; See Online Appendix B and C, for more details on parameter estimation through the second-order generalized estimating equations (GEE2) [26, 27] and derivation of the asymptotic covariance of the estimator.

4.3 Misspecification of Propensity Scores

4.3.1 Regression Adjustment and Inverse Density Weighting

If we treat the bivariate continuous or semi-continuous exposures as conditionally independent, the generalized propensity score can be adjusted similarly as in the regression model (3.5). That is,

$$\mathbb{E}(Y | \mathbf{X}, \tilde{\mathbf{S}}) = \mathbf{X}' \boldsymbol{\varphi} + \tilde{\mathbf{S}}' \boldsymbol{\psi}, \tag{4.8}$$

where $\mathbf{X} = (1, A_1, A_2, A_1A_2)'$ contains continuous exposures, $\tilde{\mathbf{S}} = (\mathbb{E}(A_1 | Z_1), \mathbb{E}(A_2 | Z_2), \mathbb{E}(A_1 | Z_1)\mathbb{E}(A_2 | Z_2))'$. For semi-continuous exposures, we have $\mathbf{X} = (1, L_1, L_2, L_1L_2)'$, and $\tilde{\mathbf{S}} = (\pi_1(Z_0), \mathbb{E}(L_1 | Z_0, Z_1), \mathbb{E}(L_2 | Z_0, Z_2), \mathbb{E}(L_1 | Z_0, Z_1)\mathbb{E}(L_2 | Z_0, Z_2))'$ in (4.8).

When exposures are continuous, the inverse density of exposure weighted estimating equation used for exposure effects estimation is given by

$$U_{10}(\boldsymbol{\varphi}; \boldsymbol{\delta}) = \frac{\mathbb{I}(\mathbf{X} = \mathbf{x}) f(a_1) f(a_2)}{f(a_1 | Z_1; \boldsymbol{\delta}) f(a_2 | Z_2; \boldsymbol{\delta})} D(\mathbf{X}) [Y - \mathbb{E}^R(Y | \mathbf{X}; \boldsymbol{\varphi})] = 0, \tag{4.9}$$

and when exposures are semi-continuous, the inverse density of exposure weighted estimating equation is given by

$$w^s(\mathbf{X}, \mathbf{Z}) D(\mathbf{X}) [Y - \mathbb{E}^R(Y | \mathbf{X}; \boldsymbol{\varphi})] = 0, \tag{4.10}$$

where

$$w^s(\mathbf{X}, \mathbf{Z}) = \frac{\mathbb{I}(\mathbf{X} = \mathbf{x})\{\pi_2(\gamma_0^\dagger)f(l_1 | A^+ = 1)f(l_2 | A^+ = 1)\}^{A^+} \{1 - \pi_2(\gamma_0^\dagger)\}^{1-A^+}}{\{\pi_1(\mathbf{Z}_0; \gamma_0)f(l_1 | A^+ = 1, \mathbf{Z}_1)f(l_2 | A^+ = 1, \mathbf{Z}_2)\}^{A^+} \{1 - \pi_1(\mathbf{Z}_0; \gamma_0)\}^{1-A^+}},$$

and $D(\mathbf{X}) = (1, A^+, A^+L_1, A^+L_2, A^+L_1L_2)'$.

When the dependence model assumes no role of confounders on the association between exposures, the regression model adjusted by the generalized propensity score is similar to (3.7),

$$\mathbb{E}(Y | \mathbf{X}, \tilde{\mathbf{S}}) = \mathbf{X}'\boldsymbol{\varphi} + \tilde{\mathbf{S}}'\boldsymbol{\psi}, \tag{4.11}$$

where $\mathbf{X} = (1, A_1, A_2, A_1A_2)'$, $\tilde{\mathbf{S}} = (\mathbb{E}(A_1 | \mathbf{Z}_1), \mathbb{E}(A_2 | \mathbf{Z}_2), \mathbb{E}(A_1A_2 | \mathbf{Z}_1, \mathbf{Z}_2))'$. For semi-continuous exposures, we have $\mathbf{X} = (1, L_1, L_2, L_1L_2)'$, $\tilde{\mathbf{S}} = (\pi_1(\mathbf{Z}_0), \mathbb{E}(L_1 | \mathbf{Z}_0, \mathbf{Z}_1), \mathbb{E}(L_2 | \mathbf{Z}_0, \mathbf{Z}_2), \mathbb{E}(L_1L_2 | \mathbf{Z}_0, \mathbf{Z}_1, \mathbf{Z}_2))'$.

The inverse density weighting approach using the misspecified joint density of exposures is to solve the following estimating equation for continuous exposures,

$$U_{10}(\boldsymbol{\varphi}; \boldsymbol{\delta}) = \frac{\mathbb{I}(\mathbf{X} = \mathbf{x})f(a_1, a_2)}{f(a_1, a_2 | \mathbf{Z}_1, \mathbf{Z}_2; \boldsymbol{\delta})} D(\mathbf{X}) [Y - \mathbb{E}^R(Y | \mathbf{X}; \boldsymbol{\varphi})] = 0. \tag{4.12}$$

For the semi-continuous exposures, the inverse density of exposure weighted estimating equation is to solve (4.10) with

$$w^s(\mathbf{X}, \mathbf{Z}) = \frac{\mathbb{I}(\mathbf{X} = \mathbf{x})\{\pi_2(\gamma_0^\dagger)f(l_1, l_2 | A^+ = 1)\}^{A^+} \{1 - \pi_2(\gamma_0^\dagger)\}^{1-A^+}}{\{\pi_1(\mathbf{Z}_0; \gamma_0)f(l_1, l_2 | A^+ = 1, \mathbf{Z}_1, \mathbf{Z}_2)\}^{A^+} \{1 - \pi_1(\mathbf{Z}_0; \gamma_0)\}^{1-A^+}}. \tag{4.13}$$

4.3.2 Empirical Bias and Efficiency

In this section, we report on simulation studies based on the random sample of size 500 with 1000 repetitions to estimate the exposure effect when exposures are continuous or semi-continuous. The estimated exposure effect through both regression adjustment and inverse density-weighting approach would be assessed in terms of the relative bias, i.e., $(\hat{\boldsymbol{\varphi}} - \boldsymbol{\theta})/\boldsymbol{\theta}$.

When exposures in the true response model (2.1) are bivariate continuous, we let $A_k = \gamma_{k0} + \gamma_{k1}Z_k + R_k, k = 1, 2$, where

$$\mathbf{R} = (R_1, R_2)' \sim \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}\right),$$

and $Z_k \perp R_k$. Additionally, we have $\log\{(1 + \rho)/(1 - \rho)\} = \gamma_{30} + \gamma_{31}Z_3$. We let $\gamma_{10} = 2$ and $\gamma_{11} = \log 2$ such that $\mathbb{E}(A_1) = 2$, let $\gamma_{20} = 0.5$ and $\gamma_{21} = \log 1.25$ such that $\mathbb{E}(A_2) = 0.5$. We let $\gamma_{31} \in [-5, 5]$ and solve for γ_{30} such that $\mathbb{E}(\rho) = 0.5$.

The relative biases of exposure effects estimated from (4.8), (4.9), (4.11), and (4.12) under two types of model misspecifications are shown in Figs. 3 and 4. The patterns of relative biases for continuous exposures are quite different from the case of binary exposures, but the main findings are similar in general. We again see the relative bias

of the exposure effect increases as the covariate effect on the exposure dependence increases. Compared to the case of binary exposures, the estimated relative bias is relatively lower for continuous exposures. Regression adjustment tends to produce a smaller bias than the inverse density weighting approach with higher efficiency.

For semi-continuous exposures, we let $\pi_1(\mathbf{Z}_0) = \text{expit}(\gamma_{00} + \gamma_{01}Z_1 + \gamma_{02}Z_2)$ where $\gamma_{00} = 1.484, \gamma_{01} = 1, \gamma_{02} = 2$ such that $E(\pi_1) = 0.7$ in the first part of the model. For the second part where given $A^+ = 1$, we suppose $\log(A_k | A^+ = 1) = \gamma_{k0} + \gamma_{k1}Z_k + R_k, k = 1, 2$, where

$$\mathbf{R} = (R_1, R_2)' \sim \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}\right).$$

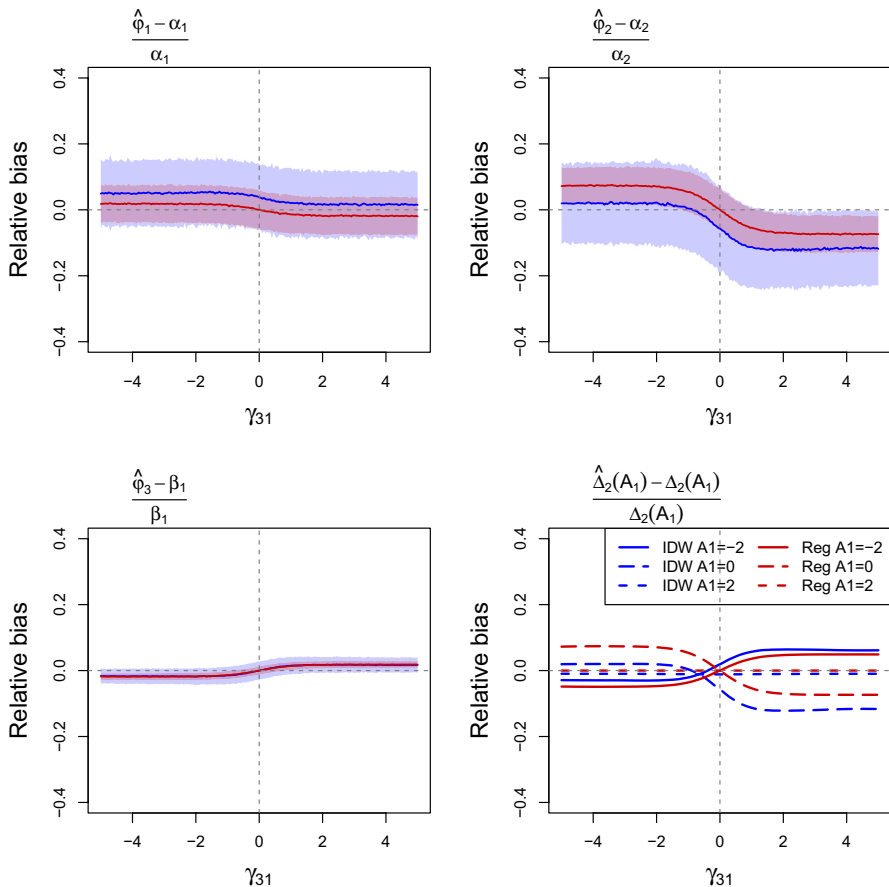


Fig. 3 Empirical relative bias with standard errors (reflected by shaded bands) of estimators of exposure effects from inadequate propensity score modeling assuming conditional independence between two continuous exposures when the second-order confounder has the effect γ_{31} ; regression adjustment (Reg) and inverse weighting (IPW) are considered; the marginal association between exposures is $\rho = 0.5$

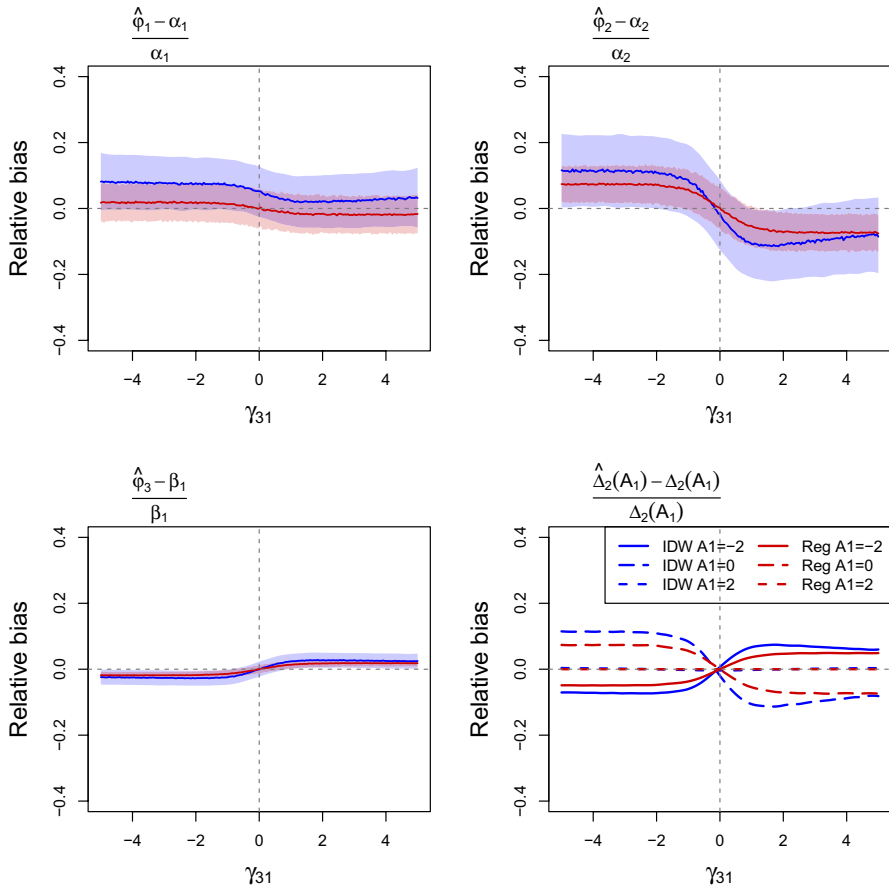


Fig. 4 Empirical relative bias with standard errors (reflected by shaded bands) of estimators of exposure effects from inadequate propensity score modeling due to the omission of a second-order confounder with effect γ_{31} ; regression adjustment (Reg) and inverse weighting (IPW) are considered; the marginal association between two continuous exposures is $\rho = 0.5$

We let $\gamma_{10} = 2$ and $\gamma_{11} = \log 2$ such that $\mathbb{E}(\log A_1 | A^+ = 1) = 2$, let $\gamma_{20} = 0.5$ and $\gamma_{21} = \log 1.25$ such that $\mathbb{E}(\log A_2 | A^+ = 1) = 0.5$. Additionally, we have $\rho = \sigma_{12}/(\sigma_1\sigma_2)$, and $\log\{(1 + \rho)/(1 - \rho)\} = \gamma_{30} + \gamma_{31}Z_3$, where we let $\gamma_{31} \in [-5, 5]$ and solve for γ_{30} such that $\mathbb{E}\{\rho(A_1, A_2)\} = 0.5$.

Figures D1 and D2 in Online Appendix D show the relative biases of semi-continuous exposure effects estimated via (4.8), (4.10), (4.11) and (4.13) under two model misspecifications. The patterns of the relative bias as a function of the covariate effect γ_{31} are similar to those of Figs. 3 and 4, while the case of semi-continuous exposures can lead to higher variability in estimates compared to binary and continuous exposures, especially under inverse density weighting and conditional independence. Again regression adjustment yields a smaller bias and greater precision compared to inverse density weighting in these simulation settings.

5 A Two-Stage Causal Analysis for Bivariate Semi-continuous Exposures

Here, we propose a two-stage approach for causal analysis involving semi-continuous exposures, again considering the special case that a single binary indicator A^+ indicates that an individual is exposed, and if $A^+ = 1$ then variables A_k , $k = 1, \dots, p$ record the different facets of exposure. It is helpful to let $A_k^c = A^+(A_k - \bar{A}_k)$ denote the centered value of A_k where $\bar{A}_k = \mathbb{E}(A_k | A^+ = 1)$, $k = 1, \dots, p$, and $\mathbf{A} = (A_1^c, \dots, A_p^c)'$ is the vector of exposures, $\mathbf{B} = (A_1^c A_2^c, \dots, A_{p-1}^c A_p^c)'$ is the vector of pairwise products of exposures, $\mathbf{X} = (A^+, \mathbf{A}', \mathbf{B}')$, and \mathbf{Z} is an $r \times 1$ vector of confounders. The data-generating model is

$$\mathbb{E}(Y | \mathbf{X}, \mathbf{Z}; \boldsymbol{\beta}) = \beta_0 + \beta_1 A^+ + \boldsymbol{\beta}'_2 \mathbf{A} + \boldsymbol{\beta}'_3 \mathbf{B} + \boldsymbol{\beta}'_4 \mathbf{Z}. \quad (5.1)$$

5.1 Stage I: The Dose–Response Surface

In stage I, we focus on the subgroup analysis modeling the dose–response surface among those with $A^+ = 1$ (the exposed). For a sample with $\{Y_i, \mathbf{X}_i, \mathbf{Z}_i, i = 1, \dots, n\}$, the estimating function is $U_1(\boldsymbol{\theta}) = \sum_{i=1}^n U_{i1}(\boldsymbol{\theta})$ for the covariate adjustment where

$$U_{i1}(\boldsymbol{\theta}) = A_i^+ D_1(\mathbf{A}_i, \mathbf{B}_i, \mathbf{Z}_i) \{Y_i - \eta_1(\mathbf{A}_i, \mathbf{B}_i, \mathbf{Z}_i; \boldsymbol{\theta})\}, \quad (5.2)$$

$D_1(\mathbf{A}_i, \mathbf{B}_i, \mathbf{Z}_i) = (1, \mathbf{A}'_i, \mathbf{B}'_i, \mathbf{Z}'_i)'$, and

$$\eta_1(\mathbf{A}_i, \mathbf{B}_i, \mathbf{Z}_i; \boldsymbol{\theta}) = \theta_1 + \boldsymbol{\theta}'_2 \mathbf{A}_i + \boldsymbol{\theta}'_3 \mathbf{B}_i + \boldsymbol{\theta}'_4 \mathbf{Z}_i.$$

Alternatively, if $\mathbf{S}_A = \mathbb{E}(\mathbf{A} | A^+ = 1, \mathbf{Z})$ and $\mathbf{S}_B = \mathbb{E}(\mathbf{B} | A^+ = 1, \mathbf{Z})$, we can set

$$U_{i2}(\boldsymbol{\theta}) = A_i^+ D_1(\mathbf{A}_i, \mathbf{B}_i, \mathbf{S}_i) \{Y_i - \eta_1(\mathbf{A}_i, \mathbf{B}_i, \mathbf{S}_i; \boldsymbol{\theta})\}, \quad (5.3)$$

where $D_1(\mathbf{A}_i, \mathbf{B}_i, \mathbf{S}_i) = (1, \mathbf{A}'_i, \mathbf{B}'_i, \mathbf{S}'_i)'$, and

$$\eta_1(\mathbf{A}_i, \mathbf{B}_i, \mathbf{S}_i; \boldsymbol{\theta}) = \theta_1 + \boldsymbol{\theta}'_2 \mathbf{A}_i + \boldsymbol{\theta}'_3 \mathbf{B}_i + \boldsymbol{\theta}'_4 \mathbf{S}_i.$$

We have $\hat{\boldsymbol{\theta}}$ from solving $\sum_{i=1}^n U_{i1}(\boldsymbol{\theta}) = 0$ or $\sum_{i=1}^n U_{i2}(\boldsymbol{\theta}) = 0$ which is the causal effect of the continuous part of exposure.

5.2 Stage II: The Effect of Exposure Status

The exposure effect of \mathbf{A} and \mathbf{B} is obtained from stage I and is used as an offset in stage II. Specifically we use $\tilde{\eta}_0(\mathbf{A}, \mathbf{B}; \boldsymbol{\theta}) = \hat{\boldsymbol{\theta}}'_2 \mathbf{A} + \hat{\boldsymbol{\theta}}'_3 \mathbf{B}$ as an offset taking the value zero for those with $A^+ = 0$. Stage II is directed at assessing the effect of the binary exposure A^+ ; many propensity score-based methods can be used to assess the effect of exposure at the reference level. We now discuss 4 possible approaches: simple

regression adjustment, propensity score regression adjustment, inverse probability weighting (IPW), and augmented inverse probability weighting (AIPW).

5.2.1 A. Simple regression adjustment

For the simple regression adjustment, we have the estimating function as

$$U_{i3}(\alpha^\circ, \theta) = D_2(A_i^+, Z_i) \{Y_i - \eta_2(A_i^+, Z_i; \alpha^\circ, \theta)\}, \tag{5.4}$$

where $\alpha^\circ = (\alpha_0, \alpha_1, \alpha'_5)'$, $D_2(A_i^+, Z_i) = (1, A_i^+, Z_i)'$, and

$$\eta_2(A_i^+, Z_i; \alpha^\circ, \theta) = \alpha_0 + \alpha_1 A_i^+ + \alpha'_5 Z_i + \text{offset}(\tilde{\eta}_0^Z(A, B; \theta)).$$

The superscript Z in $\tilde{\eta}_0^Z$ indicates that θ is estimated from (5.2) by adjusting for Z . If $\Psi = (\alpha^{\circ'}, \theta)'$, we solve the joint estimating equations,

$$\sum_{i=1}^n U_i(\Psi) = \sum_{i=1}^n \begin{pmatrix} U_{i1}(\theta) \\ U_{i3}(\alpha^\circ, \theta) \end{pmatrix} = \mathbf{0}$$

to get $\hat{\Psi}$.

5.2.2 B. Propensity score regression adjustment

If $S_0 = \mathbb{E}(A^+ | Z)$, replacing Z with S_0 in (5.4), we obtain the estimating function

$$U_{i4}(\alpha^\circ, \theta) = D_2(A_i^+, S_{i0}) \{Y_i - \eta_2(A_i^+, S_{i0}; \alpha^\circ, \theta)\}, \tag{5.5}$$

$$U_{i5}(\psi) = D_3(Z_i) \{A_i^+ - \mathbb{E}(A_i^+ | Z_i; \psi)\}, \tag{5.6}$$

where $D_2(A_i^+, S_{i0}) = (1, A_i^+, S_{i0})'$, $D_3(Z_i) = (1, Z_i)'$, $\mathbb{E}(A_i^+ | Z_i; \psi) = \text{expit}(\psi'Z_i)$, and

$$\eta_2(A_i^+, S_{i0}; \alpha^\circ, \theta) = \alpha_0 + \alpha_1 A_i^+ + \alpha_4 S_{i0} + \text{offset}(\tilde{\eta}_0^S(A, B; \theta)).$$

The superscript S in $\tilde{\eta}_0^S$ indicates that θ is estimated from (5.3) by adjusting with the generalized propensity score. Let $\Psi = (\alpha^{\circ'}, \theta', \psi)'$, we can consider solving the joint estimating equations to obtain $\hat{\Psi}$,

$$\sum_{i=1}^n U_i(\Psi) = \sum_{i=1}^n \begin{pmatrix} U_{i2}(\theta) \\ U_{i4}(\alpha^\circ, \theta) \\ U_{i5}(\psi) \end{pmatrix} = \mathbf{0}.$$

5.2.3 C. Inverse probability weighting (IPW)

We first consider a marginal structural model with the offset,

$$\eta_3(A^+; \phi, \theta) = \mathbb{E}(Y | A^+; \phi, \theta) = \phi_0 + \phi_1 A^+ + \text{offset}(\tilde{\eta}_0^S(A, B; \theta)).$$

Then we define the estimating function

$$U_{i6}(\boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\psi}) = \frac{I(A_i^+ = a_i^+)}{P(A_i^+ = a_i^+ | \mathbf{Z}; \boldsymbol{\psi})} D_4(A_i^+) \{Y_i - \eta_3(A_i^+; \boldsymbol{\phi}, \boldsymbol{\theta})\},$$

where $D_4(A_i^+) = (1, A_i^+)'$. Let $\boldsymbol{\Psi} = (\boldsymbol{\phi}', \boldsymbol{\theta}', \boldsymbol{\psi}')'$, so we can consider solving the joint estimating equations,

$$\sum_{i=1}^n U_i(\boldsymbol{\Psi}) = \sum_{i=1}^n \begin{pmatrix} U_{i2}(\boldsymbol{\theta}) \\ U_{i5}(\boldsymbol{\psi}) \\ U_{i6}(\boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\psi}) \end{pmatrix} = \mathbf{0}.$$

5.2.4 D. Augmented inverse probability weighting (AIPW)

The AIPW approach combines an outcome regression (OR) model and a propensity score (PS) model. The AIPW estimators are shown to be doubly robust, with estimators consistent if either the outcome regression model or the propensity score model is correctly specified [28]. To obtain this estimator, we begin by specifying the estimating functions for the outcome regression model through the G-computation approach [29, 30].

We consider

$$\begin{aligned} U_{i7}(\boldsymbol{\theta}, \lambda_1) &= A_i^+ D_3(\mathbf{Z}_i) \{Y_i - \lambda_1' \mathbf{Z} - \text{offset}(\tilde{\eta}_0^S(\mathbf{A}, \mathbf{B}; \boldsymbol{\theta}))\}, \\ U_{i8}(\boldsymbol{\theta}, \lambda_2) &= (1 - A_i^+) D_3(\mathbf{Z}_i) \{Y_i - \lambda_2' \mathbf{Z} - \text{offset}(\tilde{\eta}_0^S(\mathbf{A}, \mathbf{B}; \boldsymbol{\theta}))\}. \end{aligned}$$

Then the AIPW estimator $\hat{\mu}_1$ and $\hat{\mu}_0$ can be obtained by solving the estimating equation $\sum_{i=1}^n U_{i9}(\boldsymbol{\theta}, \boldsymbol{\psi}, \mu_1) = 0$ and $\sum_{i=1}^n U_{i,10}(\boldsymbol{\theta}, \boldsymbol{\psi}, \mu_2) = 0$, where

$$U_{i9}(\boldsymbol{\theta}, \boldsymbol{\psi}, \mu_1) = \frac{Y_i A_i^+ - \{A_i^+ - P(A_i^+ = 1 | \mathbf{Z}; \boldsymbol{\psi})\} \{\lambda_1' \mathbf{Z} + \text{offset}(\tilde{\eta}_0^S(\mathbf{A}, \mathbf{B}; \boldsymbol{\theta}))\}}{P(A_i^+ = 1 | \mathbf{Z}; \boldsymbol{\psi})} - \mu_1,$$

and

$$U_{i,10}(\boldsymbol{\theta}, \boldsymbol{\psi}, \mu_2) = \frac{Y_i (1 - A_i^+) - \{P(A_i^+ = 1 | \mathbf{Z}; \boldsymbol{\psi}) - A_i^+\} \{\lambda_2' \mathbf{Z} + \text{offset}(\tilde{\eta}_0^S(\mathbf{A}, \mathbf{B}; \boldsymbol{\theta}))\}}{1 - P(A_i^+ = 1 | \mathbf{Z}; \boldsymbol{\psi})} - \mu_2.$$

Let $\boldsymbol{\Psi} = (\boldsymbol{\theta}', \boldsymbol{\psi}', \lambda_1', \lambda_2')'$, we can solve the joint estimating equations $\sum_{i=1}^n U_i(\boldsymbol{\Psi}) = \mathbf{0}$, where

$$U_i(\boldsymbol{\Psi}) = (U'_{i2}(\boldsymbol{\theta}), U'_{i5}(\boldsymbol{\psi}), U'_{i7}(\boldsymbol{\theta}, \lambda_1), U'_{i8}(\boldsymbol{\theta}, \lambda_2), U'_{i9}(\boldsymbol{\theta}, \boldsymbol{\psi}, \mu_1), U'_{i,10}(\boldsymbol{\theta}, \boldsymbol{\psi}, \mu_2))'.$$

From standard estimating function theory, we have

$$\text{asvar}(\sqrt{n}(\hat{\boldsymbol{\Psi}} - \boldsymbol{\Psi})) = \mathcal{A}^{-1}(\boldsymbol{\Psi}) \mathcal{B}(\boldsymbol{\Psi}) [\mathcal{A}^{-1}(\boldsymbol{\Psi})]'$$

where $\mathcal{A}(\boldsymbol{\Psi}) = E\{-\partial U_i(\boldsymbol{\Psi})/\partial \boldsymbol{\Psi}'\}$ and $\mathcal{B}(\boldsymbol{\Psi}) = E\{U_i(\boldsymbol{\Psi})U_i'(\boldsymbol{\Psi})\}$, and the expectation is taken with respect to the true distribution. In practice, we use sample averages to

estimate $\mathcal{A}(\Psi)$ and $\mathcal{B}(\Psi)$, and evaluate the resulting expressions at the estimates to obtain $\mathcal{A}^{-1}(\hat{\Psi})\mathcal{B}(\hat{\Psi})[\mathcal{A}^{-1}(\hat{\Psi})]'$, where

$$\mathcal{A}(\hat{\Psi}) = -\frac{1}{n} \sum_{i=1}^n \left. \frac{\partial U_i(\Psi)}{\partial \Psi'} \right|_{\Psi=\hat{\Psi}},$$

and

$$\mathcal{B}(\hat{\Psi}) = \frac{1}{n} \sum_{i=1}^n \left. U_i(\Psi)U_i'(\Psi) \right|_{\Psi=\hat{\Psi}}.$$

5.3 Simulation Studies on the Two-Stage Approach

Here, we evaluate the performance of the two-stage approach for the analysis of a bivariate semi-continuous exposure under the data-generating process of Sect. 4.3.2. We consider the stage II estimator based on various combinations of methods in stages I and II. For stage I, confounder regression (5.2) and propensity score adjustment (5.3) are used. In stage II, we employ methods A) to D) of Sect. 5.2. We see that when all models are correctly specified, the empirical bias is negligible across all methods as shown in Table 1; use of IPW in stage II yields the estimates with the highest standard error, while the AIPW estimator improves the efficiency as expected.

Table 2 displays the relative bias and empirical standard error of the stage II estimators when models are misspecified in stage II. Here, the confounders are transformed as $V_1 = \exp(Z_1/2) + Z_2$, $V_2 = Z_2/(1 + \exp(Z_1)) + 10$, $V_3 = (Z_1Z_2/10 + 0.6)^2$ and $V = (V_1, V_2, V_3)'$ is used instead of $Z = (Z_1, Z_2, Z_3)'$ in stage II for model misspecification. If either the outcome regression (OR) model or the propensity score (PS) model is misspecified in Stage II, biased estimates are obtained in methods reliant on the respective model. The AIPW estimator exhibits the double robustness property. When both models are misspecified in stage II, the AIPW approach yields an estimator with a large finite sample bias and high standard error.

Table 1 Simulation results with sample size γ_{31} , repetitions OR = 4: relative bias and standard error of stage II estimator of the binary exposure effect when models are correctly specified

Stage I	Stage II	Bias (%)	Empirical SE
Regression with Z	Regression with Z	-0.08	0.14
Regression with GPS	Regression with Z	-0.12	0.15
	Regression with PS	0.13	0.19
	IPW	0.16	0.23
	AIPW	-0.11	0.17

6 Prenatal Alcohol Exposure and Cognition

Here, we consider a large prospective cohort study comprised of 480 inner-city pregnant African–American mothers recruited between September 1986 and April 1989 [12]. Quantitative measures of alcohol consumption during pregnancy were obtained at each maternal prenatal visit using a timeline follow-back interview [31] and converted to three common metrics: the number of standard drinks per drinking occasion, frequency of drinking (days/month), and average oz absolute alcohol (AA/day). Additional variables collected included socioeconomic status (SES), maternal age, maternal years of education, marital status, and family income; maternal smoking (cigarettes/day) and marijuana, cocaine, and other illicit drug use (days/month), maternal history of alcohol abuse or dependence; and child’s sex, race, and whether s/he was raised by his/her biological mother and/or was in foster care; see Table D1 in Online Appendix D. The mothers and their children were followed for 19–21 years postpartum with a retention rate of 79.0% ($N = 377$). A variety of cognitive tests were administered to each child to assess various facets of cognitive function. The results of the various cognitive tests were synthesized by fitting a second-order confirmatory factor analysis giving a composite measure of cognitive function [32] which we adopted as the response, Y , in the analyses that follow.

Among the 377 mothers in this analysis, 16.2% reported no alcohol consumption during pregnancy resulting in the alcohol exposure variables being semi-continuous. For mothers who drank alcohol during pregnancy, two main exposures (oz prenatal absolute alcohol exposure per drinking occasion, and the proportion of drinking days during pregnancy) were log-transformed and centered. Here, we evaluate the causal effects in four ways as discussed in the previous sections: subgroup analysis involving drinkers, analyses based on the bivariate semi-continuous exposure using conventional regression adjustment and the two-part model, and full sample analysis with the two-stage approach. We aim to evaluate the causal effect of alcohol drinking, two main exposures, and their interaction on child cognitive function while controlling for selected measurements listed in Table 3.

Table 2 Simulation results with sample size γ_{31} , repetitions $OR = 4$: relative bias and standard error of stage II estimator of the binary exposure effect with model misspecification in stage II

Model misspecification	Stage I	Stage II	Bias (%)	Empirical SE
PS model	Regression with GPS	Regression with PS	0.91	0.18
		IPW	− 5.06	0.35
		AIPW	0.07	0.31
OR model	Regression with GPS	Regression with Z	4.48	0.18
		Regression with Z	4.45	0.18
		AIPW	− 0.12	0.22
PS+OR model	Regression with GPS	AIPW	5.95	0.43

6.1 Subgroup Analysis Involving Drinkers

We first restrict attention to drinkers and focus on the causal effects of alcohol exposure on children’s cognition. Let A_1 represent the number of standard drinks per drinking occasion, and A_2 represent the proportion of days in each month of drinking (days/month); we refer to these as the intensity and frequency of drinking, respectively. If $A^+ = \mathbb{I}(A_{i1}A_{i2} > 0)$ indicates any consumption of alcohol, to conduct the analyses among drinkers, we consider the regression model

$$\mathbb{E}(Y_i | A_i^+ = 1, \mathbf{L}_i, \mathbf{Z}_i) = \mu + \alpha_1 L_{i1} + \alpha_2 L_{i2} + \alpha_3 L_{i1}L_{i2} + \mathbf{Z}'_i \boldsymbol{\zeta}, \tag{6.1}$$

where $L_{ik} = A_i^+ \{ \log A_{ik} - \mathbb{E}(\log A_{ik} | A_i^+ = 1) \}$. When using propensity score regression adjustment, we consider

$$\mathbb{E}(Y_i | A_i^+ = 1, \mathbf{L}_i, \mathbf{S}_i) = \mu + \alpha_1 L_{i1} + \alpha_2 L_{i2} + \alpha_3 L_{i1}L_{i2} + \mathbf{S}'_i \boldsymbol{\zeta}, \tag{6.2}$$

Table 3 Estimates for the effects of each confounder

	Estimate (SE)			
	logit(π_{1i})	$\mathbb{E}(L_{1i} A_i^+ = 1)$	$\mathbb{E}(L_{2i} A_i^+ = 1)$	$\log \frac{(1+\rho_i)}{(1-\rho_i)}$
Intercept	-0.625 (1.794)	-1.499 (0.613)	-0.648 (0.636)	-0.184 (8.978)
Smoking during pregnancy (cigarettes/day)	0.090 (0.025)	0.007 (0.005)	0.009 (0.005)	0.018 (0.108)
Socioeconomic status at birth/infancy	0.012 (0.020)	-0.016 (0.007)	-0.001 (0.007)	0.040 (0.101)
Child’s gender (female)	-0.574 (0.310)	-0.042 (0.102)	-0.075 (0.105)	0.297 (2.169)
Mother’s age at delivery	0.048 (0.035)	0.020 (0.011)	0.036 (0.011)	-0.037 (0.315)
Prenatal Beck Depression Inventory	0.024 (0.022)	0.011 (0.007)	0.012 (0.007)	0.004 (0.168)
HOME score at infancy	0.023 (0.033)	0.003 (0.011)	-0.013 (0.012)	-0.023 (0.241)
Biological mother’s verbal IQ (PPVT score)	0.000 (0.013)	-0.003 (0.004)	0.000 (0.004)	0.020 (0.086)
Biological mother’s marital status (married)	-0.121 (0.498)	-0.001 (0.179)	0.098 (0.186)	-0.350 (3.102)
Prenatal cocaine exposure (days/month)	-0.140 (0.061)	-0.026 (0.019)	-0.031 (0.020)	0.828 (3.025)
Prenatal marijuana exposure (days/month)	0.226 (0.137)	-0.016 (0.019)	0.021 (0.020)	-0.002 (0.198)
Prenatal opiate exposure (days/month)	0.122 (0.234)	0.071 (0.044)	0.163 (0.046)	-0.010 (0.785)
Number of prenatal visits	-0.018 (0.059)	-0.002 (0.020)	-0.111 (0.021)	-0.076 (0.455)
Biological mother’s education (year)	-0.040 (0.116)	0.092 (0.038)	0.012 (0.040)	-0.011 (0.575)
Parity	-0.474 (0.213)	0.095 (0.059)	-0.082 (0.062)	-0.134 (1.357)
Gravidity	0.288 (0.144)	-0.019 (0.036)	0.040 (0.037)	0.064 (0.660)
Gestational age at screening	-0.011 (0.025)	0.007 (0.008)	0.014 (0.008)	0.001 (0.153)

where $\mathbf{S}_i = (S_{i1}, S_{i2}, S_{i3})'$, with $S_{ik} = \mathbb{E}(L_{ik} | A_i^+ = 1, \mathbf{Z}_i)$, $k = 1, 2$, $S_{i3} = \mathbb{E}(L_{i1}L_{i2} | A_i^+ = 1, \mathbf{Z}_i)$. The estimated effects of each confounder are given in columns 3–5 of Table 3 for each of the respective models.

6.2 Analyses Based on the Bivariate Semi-continuous Exposure Model

The conventional regression model adjusted by all the confounders is fitted as

$$\mathbb{E}(Y_i | A_i^+, \mathbf{L}_i, \mathbf{Z}_i) = \nu + \lambda_0 A_i^+ + \lambda_1 L_{i1} + \lambda_2 L_{i2} + \lambda_3 L_{i1}L_{i2} + \mathbf{Z}_i' \boldsymbol{\eta},$$

with \mathbf{Z}_i the set of confounders. Using a two-part model for the generalized propensity score as in [24], the regression model adjusted by generalized propensity score is

$$\mathbb{E}(Y_i | A_i^+, \mathbf{L}_i, \mathbf{S}_i) = \nu + \lambda_0 A_i^+ + \lambda_1 L_{i1} + \lambda_2 L_{i2} + \lambda_3 L_{i1}L_{i2} + \mathbf{S}_i' \boldsymbol{\eta},$$

where as in (4.5), $\mathbf{S}_i = (\pi_{i1}(\mathbf{Z}_i), \mathbb{E}(L_{i1} | \mathbf{Z}_i), \mathbb{E}(L_{i2} | \mathbf{Z}_i), \mathbb{E}(L_{i1}L_{i2} | \mathbf{Z}_i))'$ with

$$\begin{aligned} \pi_{i1}(\mathbf{Z}_i) &= \text{expit}(\bar{\mathbf{Z}}_i' \boldsymbol{\gamma}_0), \\ \mathbb{E}(L_{ik} | \mathbf{Z}_i) &= \pi_{i1}(\mathbf{Z}_i) \mathbb{E}(L_{ik} | A_i^+ = 1, \mathbf{Z}_i), \quad k = 1, 2, \\ \mathbb{E}(L_{i1}L_{i2} | \mathbf{Z}_i) &= \pi_{i1}(\mathbf{Z}_i) \mathbb{E}(L_{i1}L_{i2} | A_i^+ = 1, \mathbf{Z}_i). \end{aligned}$$

The results are shown in Table 3.

The inverse density weighting approach to estimating causal effects involves solving (4.7) with stabilized weights. We assume the exposures follow a bivariate normal distribution when we estimate the conditional joint density of the two exposures.

6.3 Full Sample Analysis with the Two-Stage Approach

We also address the effects of confounding variables through the two-stage approach as discussed in Sect. 5. We first define the offset from subgroup analysis on drinkers using estimates obtained from (6.1) or (6.2) which is the stage I. Let $\bar{\eta}_i = \hat{\alpha}_1 L_{i1} + \hat{\alpha}_2 L_{i2} + \hat{\alpha}_3 L_{i1}L_{i2}$ be the offset in stage II and use regression adjustment with covariates, regression adjustment with propensity score, IPW, and AIPW to differentiate the actual exposure effect as a drinker.

6.4 Results

We assess the covariate balance using Pearson correlation [23, 33]; see Table D2 in Online Appendix D. The Pearson correlations were computed between each exposure and each covariate in Table 3 before generalized propensity score adjustment or weighting adjustment. To evaluate the performance of each approach, we analyzed the covariate balance after conditioning on the generalized propensity score or applying inverse probability/density weighting. All the methods described above substantially reduce the maximum absolute correlation and average absolute

correlation compared to the conventional regression approach shown in the third row of Table 3, indicating improvement of covariate balance.

In the full sample analysis with the two-part model, we found that the regression approach with generalized propensity score adjustment reduced covariate imbalance more effectively than the inverse density weighting approach. In the full sample analysis with the two-stage approach, the propensity score-based methods in stage II (e.g., IPW) achieve better covariate balance than the regression approach. Overall, our results demonstrate the effectiveness of each analysis with different methods in improving covariate balance, with the optimal approach depending on the specific analysis context.

Table 4 presents the causal effects estimated from both subgroup analysis and full sample analysis with the various methods. All estimated effects are negative, including the effects of drinking at the reference level versus not drinking, prenatal absolute alcohol exposure per occasion, the proportion of drinking days, and their interaction, confirming that alcohol consumption by expectant mothers has a negative effect on children's cognitive development. For subgroup analysis among drinkers, the results from the regression with covariate adjustment and the regression model adjusted by generalized propensity scores are quite similar, indicating the cognitive function score of the child decreases as the mother increases the amount of alcohol consumed at each drinking occasion, or increases her drinking frequency. The interaction term is negative so the effect of increasing the intensity of drinking on each occasion is greater when the frequency of alcohol use during pregnancy is greater, and vice versa. For full sample analysis, the conventional regression adjustment and the regression model adjusted by the two-part generalized propensity scores also generate similar estimates except for the drinker indicator. The inverse density weighting approach results in a larger effect size of alcohol exposure per occasion and smaller effect sizes of drinking frequency and their interaction compared with estimates obtained from the regression methods. In full analysis with the two-stage approach, the effect of drinking obtained from the propensity score-based methods in stage II, that is regression with a propensity score adjustment, IPW, and AIPW method, is larger than the estimate obtained from regression approaches. The IPW approach in stage II has the largest standard error among all the methods, whereas the AIPW approach generates a similar, but more precise estimate of the causal effect.

We built a dose–response surface on the exposure estimates obtained from stage I using a regression model with generalized propensity score adjustment in the two-stage approach. The resulting surface is shown in the left panel of Fig. 5. We define the reference level of a child's cognition when the mother has an average level of intensity and density of alcohol use, that is having 1.81 oz absolute alcohol per occasion and spending approximately 4 days per month (0.13 months) consuming alcohol. The x -axis and y -axis show the increment levels of drinking frequency and intensity, respectively. This graphical depiction shows the expected reduction in child's cognitive function resulting from increased drinking frequency and intensity. The right panel of Fig. 5 displays a contour plot that shows the change in cognition associated with different combinations of changes in intensity and frequency of alcohol intake. For instance, spending approximately 9 more days per month (0.3

Table 4 Estimated effects of being a drinker versus non-drinker, the logarithm of prenatal absolute alcohol exposure per occasion, the logarithm of proportion of drinking days across pregnancy, and their interaction

Analysis	Method	Average causal effect (SE)			
		Drinker indicator	Alcohol per occasion	Proportion drinking days	Interaction
Subgroup analysis	Covariate adjustment	–	–1.80 (1.07)	–0.50 (0.94)	–1.39 (0.85)
	GPS adjustment	–	–1.80 (1.09)	–0.50 (0.96)	–1.39 (0.86)
	Covariate adjustment	–0.41 (2.08)	–2.09 (1.05)	–0.88 (0.91)	–1.62 (0.83)
Full sample analysis with conventional regression	GPS adjustment	–1.26 (2.22)	–1.84 (1.11)	–0.93 (0.95)	–1.39 (0.90)
	IDW	–1.43 (2.10)	–2.22 (1.18)	–0.14 (0.96)	–0.44 (0.84)
Full sample analysis with Two-part model	Reg Z + Reg Z	–0.35 (1.48)	–	–	–
	Reg GPS + Reg Z	–0.35 (1.96)	–	–	–
	Reg GPS + Reg PS	–2.90 (2.29)	–	–	–
	Reg GPS + IPW	–3.81 (2.35)	–	–	–
	Reg GPS + AIPW	–4.34 (1.98)	–	–	–

Reg Z Regression with covariate adjustment; Reg GPS Regression with generalized propensity score adjustment; Reg PS Regression with propensity score adjustment

months) drinking based on the reference level while maintaining the same drinking intensity, or spending 3 more days per month while also increasing the dose of drinks by 2 oz on each drinking occasion based on the reference level would both result in a 4-unit decrease in the child’s cognitive function score.

7 Discussion

In this article, we describe methods for assessing the effect of two or more exposure variables on a continuous response. Confounding is addressed through the use of a generalized propensity score. In this setting, the generalized propensity score is based on the joint density of exposures and in particular is the vector of linear functions of confounders necessary to construct this joint density. For inverse density weighting, we stabilize the weights via the marginal joint density of the exposures. A key finding is that when confounders affect the joint distribution of the exposure variables, they must be accounted for in the propensity score modeling to ensure consistent estimation and valid inference regarding the interactions between the exposures. Regression adjustment appears to be less sensitive to misspecification of the dependence structure and omission of the second-order effects of confounders, and moreover, it leads to more precise estimation of causal effects. We did not consider matching or stratification on the generalized propensity score but these represent alternative approaches for dealing with confounders [34].

Note that much of the literature on causal analysis includes the conceptualization of potential outcomes [35, 36] for each unit, here denoted by $\{Y(\mathbf{a}) \text{ for } \mathbf{a} \in \mathcal{A}\}$ where \mathcal{A} may be a multidimensional space for possible exposures. We do not make use of potential outcomes in this paper, but our methods could be reformulated and formalized within that framework.

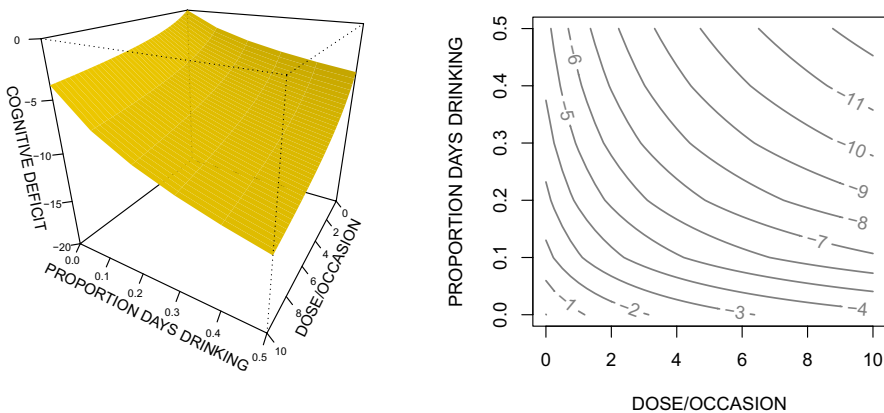


Fig. 5 Dose–response surface and contour plot generated from stage I in two-stage analysis: the effect of changes in drinking frequency (proportion days drinking) and intensity (dose/occasion) on child cognitive deficit; reference levels: 1.81 oz AA/occasion, 0.13 of the pregnancy drinking

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s12561-023-09403-8>.

Acknowledgements We thank Neil Dodge, Ph.D., for his assistance with data management support: This research was funded by grants to Sandra W. Jacobson and Joseph L. Jacobson from the National Institutes of Health/National Institute on Alcohol Abuse and Alcoholism (NIH/NIAAA; R01-AA025905) and the Lycaki-Young Fund from the State of Michigan. Richard J. Cook was supported by the Natural Sciences and Engineering Research Council of Canada through grants RGPIN 155849 and RGPIN 04207. Louise Ryan was supported by the Australian Research Council Centre of Excellence for Mathematical and Statistical Frontiers (ACEMS) CE140100049. Data collection for the Detroit Longitudinal Study was supported by grants from NIH/NIAAA (R01-AA06966, R01-AA09524, and P50-AA07606) and NIH/National Institute on Drug Abuse (R21- DA021034).


References

- Rosenbaum PR, Rubin DB (1983) The central role of the propensity score in observational studies for causal effects. *Biometrika* 70(1):41–55
- Imbens GW (2000) The role of the propensity score in estimating dose-response functions. *Biometrika* 87(3):706–710
- Hirano K, Imbens GW (2004) The propensity score with continuous treatments. *Appl Bayesian Model Causal Inference Incomplete-Data Perspect* 226164:73–84
- Dominici F, Peng RD, Barr CD, Bell ML (2010) Protecting human health from air pollution: shifting from a single-pollutant to a multi-pollutant approach. *Epidemiology* 21(2):187
- Duell EJ (2012) Epidemiology and potential mechanisms of tobacco smoking and heavy alcohol consumption in pancreatic cancer. *Mol Carcinog* 51(1):40–52
- Suzuki T, Wakai K, Matsuo K, Hirose K, Ito H, Kuriki K, Sato S, Ueda R, Hasegawa Y, Tajima K (2006) Effect of dietary antioxidants and risk of oral, pharyngeal and laryngeal squamous cell carcinoma according to smoking and drinking habits. *Cancer Sci* 97(8):760–767
- Beelen R, Hoek G, van den Brandt PA, Goldbohm RA, Fischer P, Schouten LJ, Armstrong B, Brunekreef B (2008) Long-term exposure to traffic-related air pollution and lung cancer risk. *Epidemiology* 19:702–710
- Vegetabile BG, Gillen DL, Stern HS (2020) Optimally balanced gaussian process propensity scores for estimating treatment effects. *J R Stat Soc Ser A* 183(1):355–377
- Williams JR, Crespi CM (2020a) Causal inference for multiple continuous exposures via the multivariate generalized propensity score. [arXiv:2008.13767](https://arxiv.org/abs/2008.13767)
- Chen J, Zhou Y (2023) Causal effect estimation for multivariate continuous treatments. *Biometr J*. <https://doi.org/10.1002/bimj.202200122>
- Di Credico G, Edefonti V, Polesel J, Pauli F, Torelli N, Serraino D, Negri E, Luce D, Stucker I, Matsuo K et al (2019) Joint effects of intensity and duration of cigarette smoking on the risk of head and neck cancer: a bivariate spline model approach. *Oral Oncol* 94:47–57
- Jacobson SW, Jacobson JL, Sokol RJ, Chiodo LM, Corobana R (2004) Maternal age, alcohol abuse history, and quality of parenting as moderators of the effects of prenatal alcohol exposure on 7.5-year intellectual function. *Alcoholism* 28(11):1732–1745
- Kodituwakku P (2007) Defining the behavioral phenotype in children with fetal alcohol spectrum disorders: a review. *Neurosci Biobehav Rev* 31(2):192–201
- Lewis CE, Thomas KG, Dodge NC, Molteno CD, Meintjes EM, Jacobson JL, Jacobson SW (2015) Verbal learning and memory impairment in children with fetal alcohol spectrum disorders. *Alcoholism* 39(4):724–732
- Jacobson JL, Dodge NC, Burden MJ, Klorman R, Jacobson SW (2011) Number processing in adolescents with prenatal alcohol exposure and ADHD: differences in the neurobehavioral phenotype. *Alcoholism* 35(3):431–442
- Jacobson JL, Jacobson SW, Sokol RJ, Martier SS, Ager JW, Kaplan-Estrin MG (1993) Teratogenic effects of alcohol on infant development. *Alcoholism* 17(1):174–183
- Imai K, Van Dyk DA (2004) Causal inference with general treatment regimes: generalizing the propensity score. *J Am Stat Assoc* 99(467):854–866

18. Lipsitz SR, Laird NM, Harrington DP (1991) Generalized estimating equations for correlated binary data: using the odds ratio as a measure of association. *Biometrika* 78(1):153–160
19. Hernán MA, Robins JM (2010) Causal inference
20. White H (1982) Maximum likelihood estimation of misspecified models. *Econometrica* 1–25
21. Rotnitzky A, Wypij D (1994) A note on the bias of estimators with missing data. *Biometrics* 1163–1170
22. Yule GU (1912) On the methods of measuring association between two attributes. *J R Stat Soc* 75(6):579–652
23. Williams JR, Crespi CM (2020b) Causal inference for multiple continuous exposures via the multivariate generalized propensity score. [arXiv:2008.13767](https://arxiv.org/abs/2008.13767)
24. Hocagil TA, Cook RJ, Jacobson SW, Jacobson JL, Ryan LM (2021) Propensity score analysis for a semi-continuous exposure variable: a study of gestational alcohol exposure and childhood cognition. *J R Stat Soc* 184:1390–1413
25. Robins JM (2000) Marginal structural models versus structural nested models as tools for causal inference. In: *Statistical models in epidemiology, the environment, and clinical trials*. Springer, New York, pp 95–133
26. Zhao LP, Prentice RL (1990) Correlated binary regression using a quadratic exponential model. *Biometrika* 77(3):642–648
27. Kalema G, Molenberghs G, Kassahun W (2016) Second-order generalized estimating equations for correlated count data. *Comput Stat* 31:749–770
28. Scharfstein DO, Rotnitzky A, Robins JM (1999) Adjusting for nonignorable drop-out using semiparametric nonresponse models. *J Am Stat Assoc* 94(448):1096–1120
29. Snowden JM, Rose S, Mortimer KM (2011) Implementation of g-computation on a simulated data set: demonstration of a causal inference technique. *Am J Epidemiol* 173(7):731–738
30. Williamson EJ, Forbes A, Wolfe R (2012) Doubly robust estimators of causal exposure effects with missing data in the outcome, exposure or a confounder. *Stat Med* 31(30):4382–4400
31. Jacobson SW, Chiodo LM, Sokol RJ, Jacobson JL (2002) Validity of maternal report of prenatal alcohol, cocaine, and smoking in relation to neurobehavioral outcome. *Pediatrics* 109(5):815–825
32. Jacobson JL (2022) Effects of prenatal alcohol exposure on cognitive development: a dose-response analysis. Under review
33. Austin PC (2019) Assessing covariate balance when using the generalized propensity score with quantitative or continuous exposures. *Stat Methods Med Res* 28(5):1365–1377
34. Wu X, Mealli F, Kioumourtzoglou MA, Dominici F, Braun D (2022) Matching on generalized propensity scores with continuous exposures. *J Am Stat Assoc* 1–29
35. Splawa-Neyman J, Dabrowska DM, Speed TP (1990) On the application of probability theory to agricultural experiments. *Essay on principles*. Section 9. *Stat Sci* 465–472
36. Rubin DB (1974) Estimating causal effects of treatments in randomized and nonrandomized studies. *J Educ Psychol* 66(5):688
37. Prentice RL, Zhao LP (1991) Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses. *Biometrics* 825–839

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Authors and Affiliations

Kecheng Li¹  · Tugba Akkaya-Hocagil¹ · Richard J. Cook¹ · Louise M. Ryan² · R. Colin Carter³ · Khue-Dung Dang⁴ · Joseph L. Jacobson⁵ · Sandra W. Jacobson⁵

✉ Kecheng Li
k285li@uwaterloo.ca

Tugba Akkaya-Hocagil
takkayah@uwaterloo.ca

Richard J. Cook
rjcook@uwaterloo.ca

Louise M. Ryan
louise.m.ryan@uts.edu.au

R. Colin Carter
rcc2142@cumc.columbia.edu

Khue-Dung Dang
kd.dang@unimelb.edu.au

Joseph L. Jacobson
joseph.jacobson@wayne.edu

Sandra W. Jacobson
sandra.jacobson@wayne.edu

- ¹ Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, ON, Canada
- ² School of Mathematical and Physical Sciences, University of Technology Sydney, Ultimo, NSW, Australia
- ³ Departments of Emergency Medicine and Pediatrics, Institute of Human Nutrition, Columbia University Vagelos College of Physicians and Surgeons, New York, NY, USA
- ⁴ School of Mathematics and Statistics, University of Melbourne, Parkville, VIC, Australia
- ⁵ Department of Psychiatry and Behavioral Neurosciences, School of Medicine, Wayne State University, Detroit, MI, USA