Published in final edited form as:

Stat Med. 2014 November 20; 33(26): 4577–4589. doi:10.1002/sim.6252.

Covariance Adjustment on Propensity Parameters for Continuous Treatment in Linear Models

Wei Yang, Marshall M. Joffe, Sean Hennessy, and Harold I. Feldman

Department of Biostatistics and Epidemiology, Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19104, USA

Wei Yang: weiyang@mail.med.upenn.edu

Summary

Propensity scores are widely used to control for confounding when estimating the effect of a binary treatment in observational studies. They have been generalized to ordinal and continuous treatments [1–3] in recent literature. Following the definition of propensity function and its parameterizations (called the propensity parameter in this paper) in Imai and van Dyk [1], we explore sufficient conditions for selecting propensity parameters to control for confounding for continuous treatments in the context of regression based adjustment in linear models. Typically, investigators make parametric assumptions about the form of the dose-response function for a continuous treatment. Such assumptions often allow the analyst to use only a subset of the propensity parameters to control confounding. When the treatment is the only predictor in the structural, i.e., causal model, it is sufficient to adjust only for the propensity parameters that characterize the expectation of the treatment variable or its functional form. When the structural model includes selected baseline covariates other than the treatment variable, those baseline covariates, in addition to the propensity parameters, must also be adjusted in the model. We demonstrate these points with an example estimating the dose-response relationship for the effect of erythropoietin on hematocrit level in patients with end stage renal disease.

Keywords

confounding; generalized propensity score; propensity parameter; treatment covariate interaction

1. Introduction

The propensity score was originally proposed by Rosenbaum and Rubin [4] for estimating the effects of binary treatments. It is defined as the probability of receiving treatment given the covariates. An important advantage of the propensity score is that it allows a high-dimensional vector of confounders to be characterized by a scalar quantity. The propensity score has recently been generalized to non-binary treatments [1–3, 5]. A generalization of the propensity score, a.k.a., generalized propensity score, is defined as the conditional probability or density of the treatment given the covariates. Imai and Van Dyk [1] extended the definition further by using finite-dimensional parameters to characterize the treatment distribution, called the propensity function. In the generalized definition, the propensity function is the probability (or density) of the treatment a subject actually received given the covariates. If the treatment is binary, it is the same as the propensity score for treated

subjects, and it is one minus the propensity score for untreated subjects. Following Imai and Van Dyk [1], we consider the parameters characterizing the treatment distribution which are also functions of the baseline covariates as the definition of generalized propensity scores and call them the propensity parameter. This definition is consistent with the original definition of propensity scores for binary treatment, i.e., the probability of receiving the treatment.

If we do not make any distributional assumptions about the treatment variable, the propensity parameter has the same dimension as the number of possible values of the treatment minus one; thus, for a continuous-valued treatment, there is potentially an infinite number of propensity parameters. The dimension reduction property of propensity scores for binary treatments may be reduced or lost for non-binary treatments. Further dimension reduction is desirable where warranted, but ways to do this have not been fully explored in the literature. As we shall see, this can be achieved in two ways: through restrictions on the distribution of treatment given covariates, and through restrictions on the functional form of the effect of treatment on outcome. A similar idea for dimension reduction has been used in other scenarios [6].

In this paper, we consider the use of propensity parameters in parametric estimation of dose-response functions and treatment by covariate interactions, neither of which has received adequate consideration in the literature. Consideration of dose-response is necessary for interval-scaled treatments, and interactions will be of interest when we try to personalize treatment to individuals by their measured characteristics. To do this, we focus on sufficient conditions for selecting propensity parameters to estimate the causal effects of general treatment regimes. We focus on the scenarios where the structural model is pre-specified and in which there are limited number of covariates other than the treatment. We explore first the sufficiency of using propensity parameters when there are no covariates other than the treatment in the structural model, and then extend the result to general forms of structural models, e.g., when the causal effect of treatment is modified by baseline covariates or effects are modeled as functions of change in treatment.

The rest of the paper is organized as follows. Section two reviews the propensity parameter framework and provides definitions. Section three provides sufficient conditions for the selection of propensity parameters when the effect of treatment is not modified by other covariates. Section four extends the theory of propensity parameter selection for general structural models, followed by simulation studies in section five. Section six illustrates these points using data from a study of the effect of erythropoietin on hematocrit levels in subjects with end-stage renal disease. Section seven provides further discussion.

2. The propensity parameter

Let A denote the treatment or a vector of treatments, Y the observed outcome, and X the covariates. Let $f(A; \varphi)$ denote the probability density function of A given X , and is parameterized by $\varphi = (\varphi_{1X}, \varphi_2)^t$. φ_{1X} is the subset of φ that is dependent upon X , and φ_2 is variationally independent of X . Following Imai and Van Dyk [1], $f(A; \varphi)$ is the propensity

function for treatment A and depends on X only through φ_{1X} . We call φ_{1X} the propensity parameter.

We provide a few examples for illustration.

Binary treatment—The propensity function for a treatment variable with a Bernoulli distribution is $\{P(A = 1 | X)\}^A \{1 - P(A = 1 | X)\}^{1-A}$ and can be uniquely characterized by the propensity parameter $p = P(A = 1 | X)$, the probability of receiving the treatment given covariates, which is the same as the original propensity score definition [4].

Treatment with multiple levels—The propensity function for a treatment variable with t

levels is $\prod_{j=1}^t P(A=j|X)^{I(A=j)} = \left\{ \prod_{j=1}^{t-1} P(A=j|X)^{I(A=j)} \right\} \left\{ 1 - \sum_{j=1}^{t-1} P(A=j|X) \right\}^{I(A=t)}$ and can be characterized by the propensity parameters $P(A = j | X), j = 1, \dots, t-1$. A common modeling approach is to specify parametric models for $P(A = j | X), j = 1, \dots, t-1$ as

$g \left\{ \frac{P(A=j|X)}{P(A=t|X)} \right\} = \alpha_j + \beta_j X, j = 1, \dots, t-1$. Given that $P(A = j | X), j = 1, \dots, t-1$ depends on X only through the vector $\beta X \equiv (\beta_1 X, \beta_2 X, \dots, \beta_{t-1} X)^t$, where $\beta \equiv \{\beta_1, \dots, \beta_{t-1}\}$ is the vector obtained by stacking the level-specific regression vectors β_j . βX may be considered as the propensity parameter φ_{1X} instead and has the same dimension as $P(A = j | X), j = 1, \dots, t-1$. φ_2 is comprised of $\alpha_j, j = 1, \dots, t-1$. Under different parameterizations of the dependence of A on X , the propensity parameters can sometimes be reduced to be a scalar quantity [3].

Treatment with normal distribution—Suppose that the distribution of treatment A given X follows a normal distribution $A \sim N(\mu_X, \sigma_X^2)$. The propensity function is the

Gaussian density function. The propensity parameters are (μ_X, σ_X^2) . If instead σ^2 is not a function of X , the only propensity parameter is the mean parameter for the treatment distribution μ_X (and φ_2 is comprised of σ^2).

Treatment with bivariate normal distribution—Suppose that the distribution of treatment A given X follows a bivariate normal distribution

$A = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} \sim N \left(\begin{pmatrix} \mu_{1X} \\ \mu_{2X} \end{pmatrix}, \begin{pmatrix} \sigma_{1X}^2 & \rho_X \sigma_{1X} \sigma_{2X} \\ \rho_X \sigma_{1X} \sigma_{2X} & \sigma_{2X}^2 \end{pmatrix} \right)$. The propensity function is the bivariate Gaussian density function. Assuming all parameters are functions of X , the propensity parameters are: two mean propensity parameters μ_{1X}, μ_{2X} , two variance propensity parameters $\sigma_{1X}^2, \sigma_{2X}^2$ and one correlation propensity parameter ρ_X . The form of the propensity parameter is not unique, as we can substitute $\sigma_{12X} = \rho_X \sigma_{1X} \sigma_{2X}$, a covariance term, for ρ_X in the propensity parameter. While these formulations are equivalent nonparametrically, in practice they may be different depending on whether one chooses to model ρ_X or σ_{12X} directly.

3. Propensity parameter selection when the treatment is the only predictor in the structural model

We showed in the previous section that the number of propensity parameters can be reduced substantially by making parametric assumptions about the treatment distribution. In this section, we provide sufficient conditions to control for confounding by using a subset of the propensity parameters in the setting of linear models. In particular, by making further parametric assumptions on the structural model, we show in theorem one that only a subset of the propensity parameters is required to control for confounding when the treatment is the only predictor in the structural model. The next section extends the theorem to the case when there are covariates other than the treatment in the structural model.

3.1 Potential outcomes and strongly ignorable treatment assignment

The potential outcomes framework [7, 8] is useful for defining causal effects. Let Y^a denote the outcome that would be seen in a subject were s/he to receive treatment level $a \in \Omega_a$, where Ω_a denotes a set of potential treatment values. Let $Y = \{Y^a\}$, $a \in \Omega_a$ denote the set of all potential outcomes. The effect of treatment is defined in terms of contrasts of the distributions or expectations of Y^a and $Y^{a'}$, $a \neq a'$.

The usual assumption justifying estimation of causal effects from observational data is the strongly ignorable treatment assignment assumption [4], which includes independence and positivity components. The independence component states that treatment assignment is independent of all potential outcomes conditional on covariates X , i.e., $f(A | Y^a, X) = f(A | X)$ for all $a \in \Omega_a$ [1]. The positivity assumption is that each treatment level a has a positive probability at each level of X , i.e., $\Pr(A = a | X) > 0$ for all a . The strongly ignorable treatment assignment assumption leads to nonparametric identification of marginal expectations of all potential outcomes when A is not continuous. For continuous A , we replace $\Pr(A = a | X) > 0$ by $f_{A|X}(a | x) > 0$ for all a in the support of A .

Let $e(X)$ denote the full set of propensity parameters. It is trivial to show that the independence assumption above implies independence between Y and A conditional on the full set of propensity parameters; i.e., $f\{A | Y^a, e(X)\} = f\{A | e(X)\}$ for all $a \in \Omega_a$. [1, 4]. This implies that, if adjustment for X is sufficient to control for confounding, so is adjustment for the lower dimensional summary $e(X)$.

3.2 The structural model and unbiased estimating equation conditioning on the full set of propensity parameters

Before we show the theorem for the selection of propensity parameters, we first provide an unbiased estimating equation for the causal treatment effect by conditioning on the full set of propensity parameters under the assumption of strongly treatment assignment. Let Y^0 denote the potential outcome for a subject were s/he to receive no treatment. We suppose that

$$Y^0 = Y - g(A; \varphi), \quad (1)$$

where $g(A; \phi)$ specifies the causal treatment effect; we refer to (1) as the structural model. Following Robins *et al.* [9], it requires that $g(0; \phi) = 0$ (and typically also that $g(A; 0) = 0$). We consider inference when we use the propensity parameters $e(X)$ as regressors and fit a linear regression model to estimate the causal parameter ϕ .

Consider the estimating equation for the causal parameter ϕ

$$\sum \left[Y^0 - E \left\{ Y^0 | e(X) \right\} \right] g'(A) = \sum \left[Y - g(A; \phi) - E \left\{ Y^0 | e(X) \right\} \right] g'(A) = 0$$

where $g'(A)$ is the first order derivative of $g(A; \phi)$ with respect to ϕ . When $g(A; \phi)$ is linear in A , these equations are equivalent to standard least squares linear regression. The estimating equation provides an unbiased estimate of the causal parameter ϕ because:

$$\begin{aligned} & E \left(\left[Y^0 - E \left\{ Y^0 | e(X) \right\} \right] g'(A) | e(X) \right) \\ & \stackrel{1}{=} E \left\{ Y^0 g'(A) | e(X) \right\} - E \left[E \left\{ Y^0 | e(X) \right\} g'(A) | e(X) \right] \\ & \stackrel{2}{=} E \left\{ Y^0 g'(A) | e(X) \right\} - E \left\{ Y^0 | e(X) \right\} E \left\{ g'(A) | e(X) \right\} \\ & \stackrel{3}{=} E \left\{ Y^0 | e(X) \right\} E \left\{ g'(A) | e(X) \right\} - E \left\{ Y^0 | e(X) \right\} E \left\{ g'(A) | e(X) \right\} \\ & \stackrel{4}{=} 0 \end{aligned}$$

Step 2 follows because $E \{ Y^0 | e(X) \}$ is a function of $e(X)$ and is constant conditional on $e(X)$.

Step 3 follows because of independence assumption.

3.3 Sufficient set of the propensity parameters

If we only control for a subset of the propensity parameters, say $e^*(X)$, the estimating equation is generally biased because $Y^0 \not\perp\!\!\!\perp g'(A) | e^*(X)$ even if $Y^0 \perp\!\!\!\perp g'(A) | e(X)$, and so step 4 above does not follow. However, if we make a parametric assumption for the structural model $g(A; \phi)$, only a subset of the propensity parameters may be required to yield unbiased estimates of the causal parameters. We provide this sufficient condition in theorem one.

Theorem 1—Consider a structural model of form $Y^0 = Y - g(A; \phi)$, satisfying the conditions $g(A = 0; \phi) = 0$ and $g(A; \phi = 0) = 0$. Let $e^*(X)$ denote the propensity parameters that are involved in the characterization of the expectation of the first derivative of the structural model $g'(A)$ given covariates X , i.e., $E \{ g'(A) | X \}$. The estimating equations

$$\sum \left[Y^0 - E \left\{ Y^0 | e^*(X) \right\} \right] g'(A) = \sum \left[Y - g(A; \phi) - E \left\{ Y^0 | e^*(X) \right\} \right] g'(A) = 0$$

provides unbiased estimate of the causal parameter ϕ .

See proof in appendix 1.

For concreteness, we provide one example. Assume the treatment A given X follows a normal distribution, i.e., $A \sim N(\mu_X, \sigma_X^2)$. The two propensity parameters are μ_X and σ_X^2 .

Suppose the structural model is $g(A; \phi) = A\phi$, so that $g'(A) = A$ and $E\{g'(A) | X\} = E(A | X) = \mu_X$. Controlling for μ_X is sufficient to yield unbiased estimate of the causal parameter ϕ . If

$g(A; \phi) = A^2\phi$, $g'(A) = A^2$ and $E\{g'(A) | X\} = E(A^2 | X) = \mu_X^2 + \sigma_X^2$. In this case, both propensity parameters are necessary to yield unbiased estimate of the causal parameter ϕ .

In the proof of theorem 1, we did not specify a parametric form for $E\{Y^0 | e^*(X)\}$. One can include $e^*(X)$ as a regressor in a parametric regression model for Y to estimate the causal parameter. It is required that the parametric form for $E\{Y^0 | e^*(X)\}$ be correctly specified to obtain unbiased estimate. These results parallel those of Robins, Mark and Newey [10], which considers semiparametric inferential methods whose validity does not depend on correct specification of $E\{Y^0 | e^*(X)\}$.

4. Extension of the propensity parameter selection when there are covariates other than the treatment in the structural model

There are scenarios when one wants to include predictors other than the treatment variable in the structural model. For example, to explore the treatment effect heterogeneity, we can include the treatment by covariate interactions in the structural model. This will be of importance when we want to tailor treatment or treatment dosing to an individual's personal characteristics. Another example is when one is interested in estimating the effect of treatment change from the current assigned treatment on outcome rather than the treatment itself. Theorem two generalizes to the case when there are covariates other than the treatment in the structural model.

Theorem 2—Consider a structural model of form $Y^0 = Y - g(A^*, V; \phi)$, where $A^* \equiv k(A, W)$, W, V are subsets of the covariate X , $g(A^*, V; \phi)$ satisfies the conditions $g(A^* = 0, V; \phi) = 0$ and $g(A, V; \phi = 0) = 0$, and $g(\cdot)$ and $k(\cdot)$ are known functions of their arguments. Let $T(X) \equiv E\{g'(A^*, V) | X\}$ denote the expectation of the derivative of $g(A^*, V)$ given covariates X . The estimating equations

$$\sum \left[Y^0 - E\{Y^0 | T(X)\} \right] g'(A^*, V) = \sum \left[Y - g(A^*, V; \phi) - E\{Y^0 | T(X)\} \right] g'(A^*, V) = 0$$

provides unbiased estimate of the causal parameter ϕ .

See proof in appendix two.

We now connect theorem two with the use of propensity parameters. To this end, we provide two corollaries. In both corollaries we suppose that the causal model $g(A^*, V; \phi)$ can be expressed in terms of finite linear combinations of products of known functions of A^* and V , and so that $T(X)$ is a function of $e(X)$ only through a finite set of components $e^*(X)$. The first corollary applies to the scenario when there is modification of treatment effect by some covariates and the causal effect is parameterized in terms of the treatment itself, i.e., $A^* = A$.

Corollary 1—Assume $A^* = k(A, W) = A$ and the structural model is

$g(A^*, V; \varphi) = \sum_{j=1}^n h_{1j}(A) h_{2j}(V) \varphi_j$, where $h_{1j}(A), j = 1, \dots, n$ are arbitrary functions of A (with $h_{1j}(0) = 0$) and $h_{2j}(V), j = 1, \dots, n$ are arbitrary functions of V . It is sufficient to control for V and $e^*(X)$, the subset of propensity parameters for A involved in the characterization of $E\{h_{1j}(A) | X\}, j = 1, \dots, n$ to yield unbiased estimates of the causal parameters $\phi_j, j = 1, \dots, n$.

See proof in appendix 3.

To make it concrete, we provide one example here. Assume A follows a normal distribution, i.e., $A \sim N(\mu_X, \sigma_X^2)$. The two propensity parameters are μ_X and σ_X^2 . If the structural model is $g(A^*, V; \phi) = AV\phi$, then $E\{g'(A^*, V) | X\} = E(AV | X) = VE(A | X)$. To yield unbiased estimate of the causal parameter ϕ , it is sufficient to control for V and the propensity parameter for A involved in the characterization of $E(A | X)$, which is μ_X .

The second corollary extends to the scenario when the causal effect of interest is parameterized as a function of the treatment variable and some covariates. For example, we may wish to model the effect of treatment as the effect of the change of current treatment from previous one rather than the treatment itself, i.e., $A^* = k(A, W) = A - W$, where W denotes the previous treatment.

Corollary 2—Suppose the causal model is

$g(A^*, V; \varphi) = g\{k(A, W), V; \varphi\} = \sum_{j=1}^n \{h_{1j}(A) - h_{2j}(W)\} h_{3j}(V) \varphi_j$, where $h_{1j}(A), j = 1, \dots, n$, $h_{2j}(W), j = 1, \dots, n$ and $h_{3j}(V), j = 1, \dots, n$ are arbitrary known functions of A, W and V respectively. It is sufficient to control for W, V and $e^*(X)$, the subset of propensity parameters for A involved in the characterization of $E\{h_{1j}(A) | X\}, j = 1, \dots, n$ to get unbiased estimates of the causal parameters $\phi_j, j = 1, \dots, n$.

See proof in appendix 4.

In the example above, if the structural model is $g(A^*, V; \phi) = (A - W)\phi$, then $E\{g'(A^*, V) | X\} = E(A - W | X) = E(A | X) - W$. To yield an unbiased estimate of the causal parameter ϕ , it is sufficient to control for W and the propensity parameter for A involved in the characterization of $E(A | X)$, which is μ_X .

Figure 1 illustrates why it is necessary to control for W in addition to the propensity parameters for A using a Directed Acyclic Graph (DAG) [11]. In the DAG, to figure out the causal effect of A^* on Y , both W and $e(X)$ are necessary to block all back door paths from A^* to Y : the propensity parameter $e(X)$ blocks the paths $A^* \leftarrow A \leftarrow e(X) \leftarrow X^* \rightarrow Y$ and $A^* \leftarrow A \leftarrow e(X) \leftarrow W \leftrightarrow X^* \rightarrow Y$, and the path of $A^* \leftarrow W \rightarrow Y$ can only be blocked by W .

In the proof of theorem two, we did not specify the parametric forms for $E\{Y^0 | T(X)\}$. To estimate the causal parameter by regressing on $T(X)$ in a linear regression model, it is required that the parametric form for $E\{Y^0 | T(X)\}$ is correctly specified. Similarly, it

requires that the model for $E\{Y^0 | V, e^*(X)\}$ in corollary one and the model for $E\{Y^0 | W, V, e^*(X)\}$ in corollary two are correctly specified.

5. Simulation study

In this section, we evaluate through simulations the bias of controlling for confounding using the propensity parameters in linear regression models under different scenarios. In each simulation, the sample size is 10000 with 1000 replications.

In the first two simulations, we consider the scenarios when the treatment follows a normal distribution and the causal effect is linear (simulation I) and quadratic (simulation II). We fit four models: the unadjusted model (with treatment variable only and no adjustment of covariates) (model 1); the models adjusted for the mean propensity parameter estimated in two different ways (models 2 and 3) and the model adjusted for both mean and variance propensity parameters (model 4). Table 1 show the results when the causal effect is a linear function of treatment. The results from all three models (models 2, 3 and 4) with propensity parameter adjustment give unbiased estimates of the causal parameter, suggesting that controlling for the mean propensity parameter alone (model 2) is sufficient to control for confounding. There is some efficiency gain by controlling for the mean propensity parameter estimated using all covariates (model 3) and by controlling for both the mean and variance propensity parameters (model 4). Table 2 shows the results when the causal effect is quadratic in A . The estimates of both causal parameters were biased when only the mean parameter was adjusted in the model (models 2 and 3). They became unbiased when both the mean and variance propensity parameters were adjusted in the model (model 4).

Simulation III explores the scenario when the effect of treatment is modified by the covariate X_1 . Table 3 shows the parameter estimates for both ϕ_1 and ϕ_2 were biased when controlling for the mean propensity parameter (model 2). When X_1 was added in the model (model 3), the estimates for both ϕ_1 and ϕ_2 became unbiased.

In simulation IV, we explore the scenario when the effect of interest is parameterized in terms of change of treatment ($A - A_0$). Table 4 shows the estimate for the effect of treatment change was biased when controlling for the mean propensity parameter for the treatment alone (model 2). It became unbiased when the previous treatment A_0 was added in the model (model 3).

In all simulations, the estimated standard errors (which, as is conventional in propensity-score based estimation, do not account for estimation of the parameters) were nearly the same as the empirical standard deviation of the estimates, and the coverage was at the nominal level.

6. Data analysis

In this section, we illustrate the use of propensity parameters when estimating the effect of erythropoietin on hematocrit level in patients with end stage renal disease (ESRD). Erythropoietin is a glycoprotein hormone that controls red blood cell production and is often prescribed to treat anemia in dialysis patients. In this paper, we used Medicare claims data

from the United State Renal Disease System [12] in 2004. Similar to the paper by Cotter et al. [13], the analysis was restricted to incident dialysis patients who were erythropoietin naïve at the time of dialysis. Other inclusion criteria included age of 65 or older, starting erythropoietin treatment within ninety days of initiating dialysis, and having hematocrit information available at the first month of dialysis. A total of 10,797 subjects were included in this analysis. We estimated the effect of the average erythropoietin dose in the first three months (initiation phase) starting dialysis on the achieved hematocrit level at month four. We were also interested in the effect of the average erythropoietin dose change from the first three months to month four to six (maintenance phase) on the achieved hematocrit level at month seven.

Through the propensity parameter adjustment, we controlled for confounding by the following covariates in both analyses: age, race, ethnicity, sex, body mass index, glomerular filtration rate (GFR), primary cause of ESRD, comorbidity (including atherosclerotic heart disease, congestive heart disease, cardiovascular accident, peripheral vascular disease, other cardiac disease, chronic obstructive pulmonary disease, gastrointestinal bleeding, liver disease, dysrhythmia, cancer and diabetes), dialysis center chain and profit status. In addition, erythropoietin dose at month one and hemoglobin level before initiation of dialysis were adjusted in the initiation phase analyses and average erythropoietin dose and hematocrit level at initiation phase were adjusted in the maintenance phase analyses.

The first three moments were used to characterize the skewed distribution of the average erythropoietin dose in months 1–3 in the initiation phase analyses (average erythropoietin dose in months 4–6 in the maintenance phase analyses). Three propensity parameters corresponding to the mean, variance and third central moment were estimated: $\hat{e}_1(x) = \hat{E}(A | X)$, $\hat{e}_2(x) = \hat{E}[\{A - \hat{E}(A | X)\}^2 | X]$ and $\hat{e}_3(x) = \hat{E}[\{A - \hat{E}(A | X)\}^3 | X]$, in which X includes the covariates listed above. The R^2 for the three models were 0.47, 0.11 and 0.01 respectively.

In the outcome models, the effect of erythropoietin dose or change in dose was modeled using natural cubic splines [14], with five knots located at the 5th, 25th, 50th, 75th and 95th percentiles of the erythropoietin dose distribution, and the estimated propensity parameters were entered as covariates in the model. We fit five models in the initiation phase analyses: 1) an unadjusted model; 2) a multivariable adjusted model; 3) a model adjusted for the mean propensity parameter; 4) a model adjusted for the mean and variance propensity parameters and 5) a model adjusted for the propensity parameters corresponding to the mean, variance and the third central moment for the erythropoietin dose. Figure 2 plots the estimated effects of erythropoietin dose at months 1–3 on hematocrit at month 4. The effects of erythropoietin dose estimated from the model adjusting for all three propensity parameters (model 5) were almost identical to the effects estimated from the multivariable model (model 2), suggesting the sufficiency of controlling for confounding using the three propensity parameters. Note that we relaxed the linearity assumption between the covariates and the outcome hematocrit level in the multivariable adjusted models. In the multivariable adjusted model (model 2), natural cubic spline was used for the two strongest confounders, i.e., hemoglobin level before dialysis and erythropoietin dose at month 1, with four knots (at the 20th, 40th, 60th and 80th percentiles of their distributions) and all interactions between the spline terms. In models 3–5, natural cubic spline was used for the mean propensity parameters with four

knots (at the 20th, 40th, 60th and 80th percentiles of the distribution). We calculated the adjusted dose response using model-based standardization [15]. At each erythropoietin dose level (a total of 100 values uniformly distributed between the minimum and maximum of the observed erythropoietin dose level), the hematocrit level is estimated as the average of all individual hematocrit levels estimated from the regression model. Figure 3 plots the adjusted dose response curve with 95% confidence intervals from model 5. It shows that mean hematocrit increases with increasing erythropoietin dose and plateaus at around 38%.

In the maintenance phase analyses, we fit four models to estimate the effect of the change in average erythropoietin dose from the first three months to months four to six on the achieved hematocrit level at month seven: 1) an unadjusted model; 2) a multivariable adjusted model; 3) a model adjusted for the propensity parameters corresponding to the mean, variance and the third central moment for the erythropoietin dose; and 4) a model adjusted for the three propensity parameters and the average erythropoietin dose in the first three months. In the multivariable adjusted model (model 2), natural cubic splines were used for the two strongest confounders, i.e., average hematocrit level and average erythropoietin dose at month 1–3, with four knots (at the 20th, 40th, 60th and 80th percentiles of their distributions) and all interactions between the spline terms. In models 3 and 4, a natural cubic spline was used for the mean propensity parameters with four knots (at the 20th, 40th, 60th and 80th percentiles of the distribution). The dose-response curve from model 4 with the adjustment of both propensity parameters and prior average erythropoietin dose at month 1–3 was very similar to the curve from the multivariable adjusted model 2 (Figure 4). However, the curve from model 3 with the adjustment for the propensity parameters alone was very different, indicating the insufficiency of controlling for confounding using the propensity parameters only. It showed that hematocrit level continues to increase with increasing erythropoietin dose at maintenance phase but starts to decrease when the average erythropoietin dose at maintenance phase increases more than 10,000 U/week compared to initiation phase.

7. Discussion

In this paper, we defined the propensity parameter as a subset of the parameters for the treatment distribution that are functions of baseline covariates. By specifying a parametric structural model, we showed that the set of propensity parameters sufficient to control for confounding are those involved in characterizing the expectation of the first derivative of the structural model with respect to the causal parameter in the setting of linear models. We also showed that the propensity parameters for the treatment variable are not sufficient to yield unbiased causal parameter estimates when there are covariates other than the treatment variable in the structural model.

We justified the typical use of the mean propensity parameter, i.e., the expectation of the treatment given covariates, as a scalar quantity to control for confounding when estimating an assumed linear dose-response causal relationship. However, when the dose-response relationship is non-linear, a single mean propensity parameter is not sufficient to control for confounding. Which additional propensity parameters may be required depends on the parametric form of the structural model.

We proved the theorems and illustrated the use of propensity parameters in the context of regression-based adjustment. Under the assumptions of no unmeasured confounding variables and correct specification of the model for each propensity parameter, a correct specification of the functional form of the association of the propensity parameters with the outcome is required to get unbiased estimation of the causal parameters. To guard against bias due to model misspecification, a more relaxed form in the outcome model for the dependence of outcome on the propensity parameters is desired. In the data example we provided, we relaxed the linearity assumption for the association between the mean propensity parameter and outcome by using natural cubic splines. One can further reduce dependence of findings by using semiparametric models in which the association of the propensity parameter and the outcome remains unspecified. Here, consistency of estimation does not require correct modeling of the association of the propensity parameter with the baseline potential outcome. Our causal models are special cases of structural nested distribution models and are also consistent with structural nested mean models (Robins et al., 1999); various semiparametric estimators have been proposed [10]. An advantage of our approach is that standard available software may be used for estimation.

Following the convention in the causal inference literature, the variance of the causal parameter is estimated from the linear regression model assuming the propensity parameters are known rather than being estimated. The inference is expected to be conservative, i.e., the estimated standard error of the causal parameter is smaller if properly taking into account the fact that the propensity parameters are estimated from the treatment model [10]. In our simulation, the standard error estimated from the regression model is almost identical to the empirical standard error, suggesting the variance estimation assuming the propensity parameters are known is reasonable.

There are three other approaches of propensity score adjustment that are often used for binary treatments: stratification, matching, and inverse probability weighting. All of these can potentially help by avoiding the specification of the association between the propensity parameters and the outcome. However, these approaches have potential challenges. Stratification works well when the propensity parameter is a scalar quantity, for example, propensity score for binary treatment. It has been shown that 90% of the bias due to the observed covariates can typically be removed by stratifying on quintiles of the propensity scores [16]. Assuming we use five categories for each propensity parameter, the total number of bins in the analysis will be 5^n where n is the number of propensity parameters to control for. In a typical scenario where there are two propensity parameters for which we need to control, the analysis will stratify on 25 bins; as the number of propensity parameters increases, the strata will grow increasingly sparse, which will lead to problems with nonparametric inference and increasing reliance on modeling assumptions. Various matching algorithms based on the propensity score for binary treatment have been proposed in literature [17–19]. However, relative little research is done on the optimal matching criteria on two or more propensity parameters, or when one is matching with multiple doses [18].

Inverse probability weighting is often used to adjust for confounding for binary treatment [20] or treatments with finite discrete levels [21, 22]. For continuous treatments, one can use

as weights one over the probability density function. However, there is no upper bound on the weight a subject can have and so the inference can be unstable. For example, subjects whose treatment level is on the tails of the distribution can have extreme large weights and can completely dominate the inference in the weighted population. Stabilized weights proposed in Marginal Structural Models [23] may help to reduce the variability of the weights. Given the difficulties of the extension of stratification, matching and inverse probability weighting for continuous treatment, it seems most straightforward to adjust for confounding by including propensity parameters as covariates in the model. Other approaches that model the treatment outcome association directly without the propensity score model were also proposed in literature, e.g., [24].

The dimension reduction property of propensity parameters is achieved by making parametric assumptions about the treatment distribution. One can also characterize the treatment distribution using the moments without a particular parametric form. The question that arises then is how many moments are required to capture enough features of the treatment distribution given covariates to yield adequate control of confounding. In order to produce confounding not accounted for by the low-dimensional propensity parameters used in modeling the outcome, one would require the original covariates to substantially predict both treatment and outcome after adjusting for said propensity parameters. This could fail if the dimension of the propensity parameter vector used is too low. In the analysis of the USRDS data, we used the first three moments to characterize the treatment distribution. In the models for treatment, covariates did a poor job of predicting moments beyond the second ($R^2 = 0.01$ for the model for the third central moment). This suggests that confounding by baseline covariates can largely be captured through the lower moments. This is consistent with the result that the dose response curve after adjusting for the propensity parameter for the third moment is very similar to the model without adjusting for it (figure 2). Appropriate guidelines for determining the adequacy of confounding control and the degree of possible residual bias await formulation.

Determining whether the treatment model is correctly specified is also desired. For binary treatment, it is suggested to check the balance of the covariates across treatment groups after conditioning on the propensity score [4]. How to do that in the setting for continuous treatment has not yet been determined. Similar to the balance check for a binary treatment, one can divide the continuous treatment into a few categories and check the covariates balance within strata defined by the propensity parameters. Another option is to use regression models; one might use regression to check the association between each covariate and the treatment, adjusting for the propensity parameter. Further research is required on these topics.

We have assumed that the model for causal effects is correctly specified. Correct model specification might be checked by embedding the assumed causal model in a larger model and testing the significance of the additional terms. The larger model for the outcome might then need to include additional functions of the propensity parameters beyond those in the assumed more parsimonious model; this could lead to less dimension reduction, so reducing the benefits of using propensity parameters. This might also complicate model selection; exploration of this is beyond the scope of this manuscript.

In summary, we provided sufficient conditions of using propensity parameters to control for confounding. Further research is required on extending the current approaches for propensity score adjustment to the cases when there are two or more propensity parameters.

Acknowledgments

We thank Ben Hansen for his helpful discussions. The work of Drs. Yang and Joffe was supported in part by NIH grant R01-DK090385.

References

1. Imai K, van Dyk DA. Causal inference with general treatment regimes: Generalizing the propensity score. *Journal of the American Statistical Association*. 2004; 99:854–866.
2. Imbens GW. The role of the propensity score in estimating dose-response functions. *Biometrika*. 2000; 87:706–710.
3. Joffe MM, Rosenbaum PR. Invited commentary: Propensity scores. *Am J Epidemiol*. 1999; 150:327–333. [PubMed: 10453808]
4. Rosenbaum PR, Rubin DB. The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*. 1983; 70:41–55.
5. Hirano, K.; Imbens, GW. *The Propensity Score with Continuous Treatments*. Wiley; 2004. *The Propensity Score with Continuous Treatments*.
6. van der Laan MJ, Gruber S. Collaborative Double Robust Targeted Maximum Likelihood Estimation. *Int J Biostat*. 2010; 6:Article 17. [PubMed: 20628637]
7. Splawa-Neyman J, Dabrowska DM, Speed TP. On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9. *Statistical Science*. 1990; 5:465–472.
8. Rubin, D. *Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies*. City: 1974. *Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies*.
9. Robins, J.; Rotnitzky, A.; Scharfstein, D. *Sensitivity Analysis for Selection Bias and Unmeasured Confounding in Missing Data and Causal Inference Models*. 1999. *Sensitivity Analysis for Selection Bias and Unmeasured Confounding in Missing Data and Causal Inference Models*.
10. Robins JM, Mark SD, Newey WK. Estimating exposure effects by modelling the expectation of exposure conditional on confounders. *Biometrics*. 1992; 48:479–495. [PubMed: 1637973]
11. Pearl, J. *Causality: models, reasoning, and inference*. Cambridge University Press; Cambridge, U.K.; New York: 2000.
12. USRDS: the United States Renal Data System. *Am J Kidney Dis*. 2003; 42:1–230.
13. Cotter D, Zhang Y, Thamer M, Kaufman J, Hernan MA. The effect of epoetin dose on hematocrit. *Kidney Int*. 2008; 73:347–353. [PubMed: 18004296]
14. Hastie, T.; Tibshirani, R. *Generalized additive models*. 1. Chapman and Hall; London; New York: 1990.
15. Joffe MM, Greenland S. Standardized Estimates from Categorical Regression-Models. *Statistics in Medicine*. 1995; 14:2131–2141. [PubMed: 8552892]
16. Rosenbaum PR, Rubin DB. Reducing Bias in Observational Studies Using Subclassification on the Propensity Score. *Journal of the American Statistical Association*. 1984; 79:516–524.
17. Hansen BB. Full matching in an observational study of coaching for the SAT. *Journal of the American Statistical Association*. 2004; 99:609–618.
18. Lu B, Zanutto E, Hornik R, Rosenbaum PR. Matching with doses in an observational study of a media campaign against drug abuse. *Journal of the American Statistical Association*. 2001; 96:1245–1253.
19. Rosenbaum PR. Optimal Matching for Observational Studies. *Journal of the American Statistical Association*. 1989; 84:1024–1032.
20. Rosenbaum PR. Model-Based Direct Adjustment. *Journal of the American Statistical Association*. 1987; 82:387–394.

21. McCaffrey DF, Griffin BA, Almirall D, Slaughter ME, Ramchand R, Burgette LF. A tutorial on propensity score estimation for multiple treatments using generalized boosted models. *Statistics in Medicine*. 2013; 32:3388–3414. [PubMed: 23508673]
22. Feng P, Zhou XH, Zou QM, Fan MY, Li XS. Generalized propensity score for estimating the average treatment effect of multiple treatments. *Statistics in Medicine*. 2012; 31:681–697. [PubMed: 21351291]
23. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000; 11:550–560. [PubMed: 10955408]
24. Hill JL. Bayesian Nonparametric Modeling for Causal Inference. *Journal of Computational and Graphical Statistics*. 2011; 20:217–240.

Appendix 1

Proof of theorem 1

Suppose that $e^*(X)$ is a subset of $e(X)$ that contains all propensity parameters that are involved in the characterization of the expectation of the first derivative of the causal model, i.e., $E\{g'(A) | X\}$. Conditioning on $e^*(X)$, the estimating equation is unbiased:

$$\begin{aligned}
 & E \left([Y^0 - E \{Y^0 | e^*(X)\}] g'(A) | e^*(X) \right) \\
 & \stackrel{1}{=} E \{Y^0 g'(A) | e^*(X)\} - E \left[E \{Y^0 | e^*(X)\} g'(A) | e^*(X) \right] \\
 & \stackrel{2}{=} E \{Y^0 g'(A) | e^*(X)\} - E \{Y^0 | e^*(X)\} E \{g'(A) | e^*(X)\} \\
 & \stackrel{3}{=} E \left[E \{Y^0 g'(A) | X\} | e^*(X) \right] - E \{Y^0 | e^*(X)\} E \{g'(A) | e^*(X)\} \\
 & \stackrel{4}{=} E \left[E \{Y^0 | X\} E \{g'(A) | X\} | e^*(X) \right] - E \{Y^0 | e^*(X)\} E \{g'(A) | e^*(X)\} \\
 & \stackrel{5}{=} E \{Y^0 | e^*(X)\} E \{g'(A) | e^*(X)\} - E \{Y^0 | e^*(X)\} E \{g'(A) | e^*(X)\} \\
 & \stackrel{6}{=} 0
 \end{aligned}$$

Step 5 above follows because $E\{g'(A) | X\}$ is a function of $e^*(X)$ alone so it is constant conditional on $e^*(X)$.

Appendix 2

Proof of theorem 2

By conditioning on $T(X) = E\{g'(A^*, V) | X\}$, we have

$$\begin{aligned}
 & E \left([Y^0 - E \{Y^0 | T(X)\}] g'(A^*, V) | T(X) \right) \\
 & \stackrel{1}{=} E \{Y^0 g'(A^*, V) | T(X)\} - E \left[E \{Y^0 | T(X)\} g'(A^*, V) | T(X) \right] \\
 & \stackrel{2}{=} E \{Y^0 g'(A^*, V) | T(X)\} - E \{Y^0 | T(X)\} E \{g'(A^*, V) | T(X)\} \\
 & \stackrel{3}{=} E \left[E \{Y^0 g'(A^*, V) | X\} | T(X) \right] - E \{Y^0 | T(X)\} E \{g'(A^*, V) | T(X)\} \\
 & \stackrel{4}{=} E \left[E \{Y^0 | X\} E \{g'(A^*, V) | X\} | T(X) \right] - E \{Y^0 | T(X)\} E \left[E \{g'(A^*, V) | X\} | T(X) \right] \\
 & \stackrel{5}{=} E \left[E \{Y^0 | X\} T(X) | T(X) \right] - E \{Y^0 | T(X)\} E [T(X) | T(X)] \\
 & \stackrel{6}{=} T(X) E \{Y^0 | T(X)\} - E \{Y^0 | T(X)\} T(X) \\
 & \stackrel{7}{=} 0
 \end{aligned}$$

Appendix 3

Proof of corollary 1

$$T(X)=E\left\{g'(A^*, V)|X\right\}=E\left\{\begin{array}{c} h_{11}(A)h_{21}(V) \\ \vdots \\ h_{1n}(A)h_{2n}(V) \end{array} \middle| X\right\}=\left(\begin{array}{c} E\{h_{11}(A)|X\}h_{21}(V) \\ \vdots \\ E\{h_{1n}(A)|X\}h_{2n}(V) \end{array}\right)$$

which is fully determined by V and $E\{h_{1j}(A) | X\}, j=1, \dots, n$.

If $T(X)$ is sufficient to control for confounding, so are V and $E\{h_{1j}(A) | X\}, j=1, \dots, n$.

Appendix 4

Proof of corollary 2

$$T(X)=E\{g'(A^*, V)|X\}=E\left\{\begin{array}{c} \{h_{11}(A)-h_{21}(W)\}h_{31}(V) \\ \vdots \\ \{h_{1n}(A)-h_{2n}(W)\}h_{3n}(V) \end{array} \middle| X\right\}=\left(\begin{array}{c} [E\{h_{11}(A)|X\}-h_{21}(W)]h_{31}(V) \\ \vdots \\ [E\{h_{1n}(A)|X\}-h_{2n}(W)]h_{3n}(V) \end{array}\right)$$

which is fully determined by W, V and $E\{h_{1j}(A) | X\}, j=1, \dots, n$.

If $T(X, V)$ is sufficient to control for confounding, so are W, V and $E\{h_{1j}(A) | X\}, j=1, \dots, n$.

Figure 1.

A DAG representation. X^* represents all covariates other than W . The causal effect of interest is parameterized in terms of A^* , which is a function of A and W .

Figure 2. Effect of average erythropoietin dose in months 1–3 on the achieved hematocrit level at month 4 from different models relaxing linear assumptions for covariates.

Figure 3.

Dose-response relationship of erythropoietin dose in months 1–3 with the achieved hematocrit level at month 4 adjusting for mean (using natural cubic spline), variance and the third central moment propensity parameters.

Figure 4.
Effect of change of average erythropoietin dose in months 1–3 to months 4–6 on the achieved hematocrit level at month 7.

Table 1Simulation I with $\phi = 0.1$.

Model	Bias	Model based SE	Empirical SE
Model 1: $Y = A$	0.0953	0.0084	0.0084
Model 2: $Y = A + \hat{e}_1(x)$	-0.0001	0.0083	0.0083
Model 3: $Y = A + \hat{e}_1'(x)$	-0.0002	0.0083	0.0080
Model 4: $Y = A + \hat{e}_1(x) + \hat{e}_2(x)$	-0.0002	0.0079	0.0080

1. X_1, X_2 and X_3 are independent and all follow uniform distribution $U(0,1)$.
2. The treatment variable A follows a normal distribution i.e., $A \sim N(0.5 + X_1 + X_2, 0.5 + X_3)$.
3. The outcome variable Y is $Y = A\phi + X_1 + X_2 + X_3 + N(0,1)$, in which ϕ is the causal parameter.
4. The propensity parameters are estimated as: $\hat{e}_1(x) = \hat{E}(A | X_1, X_2)$, $\hat{e}_1'(x) = \hat{E}(A | X_1, X_2, X_3)$ and $\hat{e}_2(x) = \hat{E}[\{A - \hat{E}(A | X_1, X_2)\}^2 | X_3]$.
5. SE: standard error.

Table 2

Simulation II with $\phi_1 = \phi_2 = 0.1$.

Model ¹	ϕ_1			ϕ_2		
	Bias	Model based SE ²	Empirical SE	Bias	Model based SE	Empirical SE
Model 1: $Y = A + A^2$	0.0452	0.0192	0.0195	0.0126	0.0043	0.0043
Model 2: $Y = A + A^2 + \epsilon_1(x)$	-0.0499	0.0182	0.0183	0.0126	0.0041	0.0040
Model 3: $Y = A + A^2 + \epsilon_1'(x)$	-0.0499	0.0182	0.0183	0.0125	0.0041	0.0041
Model 4: $Y = A + A^2 + \epsilon_1(x) + \epsilon_2(x)$	0.0008	0.0176	0.0175	-0.0001	0.0039	0.0039

1. X_1, X_2 and X_3 are independent and all follow uniform distribution $U(0,1)$.
2. The treatment variable A follows a normal distribution i.e., $A \sim N(0.5 + X_1 + X_2, 0.5 + X_3)$.
3. The outcome variable Y is $Y = A\phi_1 + A^2\phi_2 + X_1 + X_2 + X_3 + N(0,1)$
4. The propensity parameters are estimated as: $\hat{e}_1(x) = \hat{E}(A | X_1, X_2)$, $\hat{e}_1'(x) = \hat{E}(A | X_1, X_2, X_3)$ and $\hat{e}_2(x) = \hat{E}[(A - \hat{E}(A | X_1, X_2))^2 | X_3]$.
5. SE: standard error.

Table 3

Simulation III with $\phi_1 = \phi_2 = 1$.

Model ¹	ϕ_1			ϕ_2		
	Bias	Model based SE ²	Empirical SE	Bias	Model based SE	Empirical SE
Model 1: $Y = A + A * X$	0.068	0.017	0.017	0.330	0.022	0.021
Model 2: $Y = A + A * X + \epsilon(x)$	0.091	0.015	0.018	-0.184	0.021	0.030
Model 3: $Y = A + A * X + \epsilon(x) + X_1$	0.000	0.019	0.019	-0.001	0.032	0.032

1. X_1, X_2 and X_3 are independent and all follow uniform distribution $U(0,1)$.
2. The treatment variable A follows a normal distribution, i.e., $A \sim N(X_1 + X_2 + X_3, 1)$.
3. The outcome variable Y is $Y = A\phi_1 + AX_1\phi_2 + X_1 + 2X_2 + X_3 + N(0,1)$
4. The propensity parameter is estimated as $\hat{e}(x) = \hat{E}(A | X_1, X_2, X_3)$
5. SE: standard error.

Table 4Simulation IV with $\phi = 1$.

Model ¹	Bias	Model based SE ²	Empirical SE
Model 1: $Y = A^*$	0.142	0.012	0.012
Model 2: $Y = A^* + \hat{e}(x)$	-0.085	0.010	0.011
Model 3: $Y = A^* + \hat{e}(x) + A_0$	0.000	0.010	0.010

1. X_1, X_2 and A_0 are independent and all follow uniform distribution $U(0,1)$.
2. The treatment variable A follows a normal distribution, i.e., $A \sim N(X_1 + X_2 + 0.5A_0, 1)$.
3. The outcome variable Y is $Y = (A - A_0)\phi + X_1 + 2X_2 + 2A_0 + N(0,1)$
4. The propensity parameter for A is estimated as $\hat{e}(x) = \hat{E}(A | X_1, X_2, A_0)$
5. $A^* = A - A_0$
6. SE: standard error.