Biometrika (2018), **00**, 0, *pp.* 1–14
Printed in Great Britain
doi: 10.1093/biomet/asy015

Covariate association eliminating weights: a unified weighting framework for causal effect estimation

BY SEAN YIU AND LI SU

Medical Research Council Biostatistics Unit, School of Clinical Medicine, University of Cambridge, Robinson Way, Cambridge CB2 0SR, U.K.

sean.yiu@mrc-bsu.cam.ac.uk li.su@mrc-bsu.cam.ac.uk

SUMMARY

Weighting methods offer an approach to estimating causal treatment effects in observational studies. However, if weights are estimated by maximum likelihood, misspecification of the treatment assignment model can lead to weighted estimators with substantial bias and variance. In this paper, we propose a unified framework for constructing weights such that a set of measured pretreatment covariates is unassociated with treatment assignment after weighting. We derive conditions for weight estimation by eliminating the associations between these covariates and treatment assignment characterized in a chosen treatment assignment model after weighting. The moment conditions in covariate balancing weight methods for binary, categorical and continuous treatments in cross-sectional settings are special cases of the conditions in our framework, which extends to longitudinal settings. Simulation shows that our method gives treatment effect estimates with smaller biases and variances than the maximum likelihood approach under treatment assignment model misspecification. We illustrate our method with an application to systemic lupus erythematosus data.

Some key words: Causal inference; Confounding; Continuous treatment; Covariate balance; Inverse probability weighting; Propensity function.

1. INTRODUCTION

Weighting methods are widely used to estimate causal treatment effects. The propensity function, the conditional probability of receiving treatment given a set of measured pretreatment covariates (Imai & van Dyk, 2004), features prominently in weighting methods. A natural choice of weights is a ratio of the marginal probability of treatment assignment and the propensity function, henceforth referred to as the stabilized inverse probability of treatment weights (Robins, 2000; Robins et al., 2000). Despite the appeal of weighting methods, problems arise when the propensity function is unknown. Hence weights are usually constructed using the stabilized inverse probability of treatment weight structure with an estimated propensity function, often obtained by maximum likelihood (Imbens, 2000; Robins et al., 2000), although other methods have been proposed (Lee et al., 2010). However, because these estimation procedures do not directly aim at the goal of weighting, which is to eliminate the association between a set of measured pretreatment covariates satisfying the conditions in § 2·1 and treatment assignment after weighting, a slightly misspecified propensity function model can result in badly biased treatment effect estimates (Kang & Schafer, 2007). Recently, this problem has motivated new

weighting methods that optimize covariate balance, the covariate balancing weights (Graham et al., 2012; Hainmueller, 2012; McCaffrey et al., 2013; Imai & Ratkovic, 2014, 2015; Zhu et al., 2015; Zubizarreta, 2015; Chan et al., 2016; Fong et al., 2018). These weights can dramatically improve the performance of weighting methods, but there is a lack of a framework to generalize them to complex treatment types, such as semicontinuous or multivariate treatments, and even to longitudinal settings.

In this paper, we introduce covariate association eliminating weights, a unified framework for constructing weights with the goal being that a set of measured pretreatment covariates will be unassociated with treatment assignment after weighting. Our method can be used to estimate causal effects for semicontinuous, count, ordinal, or even multivariate treatments, and it extends to longitudinal settings. An example of estimating the direct effect of a time-varying treatment on a longitudinal outcome is provided in § 8. Utilizing the generality of the propensity function and its capacity to characterize covariate associations with treatment assignment, we derive conditions for weight estimation by eliminating the association between the set of measured pretreatment covariates and treatment assignment specified in a chosen propensity function model after weighting, i.e., by solving the weighted score equations of the propensity function model at parameter values which indicate that the covariates are unassociated with treatment assignment.

Our method has several attractive characteristics. First, it encompasses existing covariate balancing weight methods and provides a unified framework for weighting with treatments of any distribution; see § 4. By eliminating the associations between the covariates and treatment assignment after weighting, our method can provide some robustness against misspecification of the functional forms of the covariates in a propensity function model, particularly if they are predictive of the outcome; see § 6. Second, it is clear from our framework what type of covariate associations are eliminated after weighting. For example, the covariate balancing weight method proposed in Fong et al. (2018) will only eliminate the associations between the covariates and the mean of a continuous treatment; see § 4·2. Our method can also eliminate the associations between the covariates and the variance of the continuous treatment. Third, our method extends to longitudinal settings; see § 8. In particular, apart from handling treatments of any distribution, it can accommodate unbalanced observation schemes and can incorporate a variety of stabilized weight structures. In contrast, to the best of our knowledge, the only available covariate balancing weight method for longitudinal settings, proposed by Imai & Ratkovic (2015), focuses on binary treatments in a balanced observation scheme, and it is not clear how to incorporate arbitrary stabilized weight structures in their approach. Finally, our method can be implemented with standard statistical software by solving a convex optimization problem that identifies minimum-variance weights subject to our conditions (Zubizarreta, 2015). This is especially appealing for nonbinary treatments with outliers (Naimi et al., 2014), because it protects against extreme weights which often lead to unstable treatment effect estimates in practice.

2. THE PROPENSITY FUNCTION

2·1. *Definition and assumptions*

Let X_i, T_i and Y_i be respectively a set of measured pretreatment covariates, the possibly multivariate treatment variable and the outcome for the ith unit ($i = 1, \ldots, n$) in a simple random sample of size n. Following Imai & van Dyk (2004), we define the propensity function as the conditional probability of treatment given the set of measured pretreatment covariates, i.e., $\text{pr}(T_i \mid X_i; \beta_{\text{true}})$, where β_{true} parameterizes this distribution. The parameter β_{true} is assumed to be unique and finite-dimensional, and is such that $\text{pr}(T_i \mid X_i)$ depends on X_i only through a subset of β_{true};

this is the uniquely parameterized propensity function assumption of Imai & van Dyk (2004). For example, if $\text{pr}(T_i \mid X_i)$ follows a regression model, then β_{true} would include regression coefficients that characterize the dependence of T_i on X_i and intercept terms that describe the baseline distribution of T_i. In addition, we make the strong ignorability of treatment assignment assumption (Rosenbaum & Rubin, 1983), also known as the unconfoundedness assumption, $\text{pr}\{T_i \mid Y_i(t^{\text{P}}), X_i\} = \text{pr}(T_i \mid X_i)$ where $Y_i(t^{\text{P}})$ is a random variable that maps a potential treatment t^{P} to a potential outcome, and the positivity assumption (Imai & van Dyk, 2004), $\text{pr}(T_i \in \mathcal{A} \mid X_i) > 0$ for all X_i and any set $\mathcal{A}$ with positive measure. Finally, the distribution of potential outcomes for one unit is assumed to be independent of the potential treatment value of another unit given the set of pretreatment covariates; this is the stable unit treatment value assumption. Throughout the paper, we make the above assumptions; otherwise our method may result in severely biased causal effect estimates, even compared with an unadjusted analysis. For example, when unconfoundedness holds without conditioning on X_i, adjusting for X_i can induce M-bias (Ding & Miratrix, 2015).

2·2. *Covariate selection*

We briefly review some methods for covariate selection. When the causal structure is known and represented by a directed acyclic graph, Shpitser et al. (2010) gave a complete graphical criterion, the adjustment criterion, to determine whether adjusting for a set of covariates ensures unconfoundedness. The adjustment criterion generalizes the back-door criterion of Pearl (1995), which is sufficient but not necessary for unconfoundedness. In the absence of knowledge about how covariates are causally related to each other, VanderWeele & Shpitser (2011) proposed the disjunctive cause criterion. This says that if any subset of pretreatment covariates suffices to ensure unconfoundedness, then the subset of pretreatment covariates that are causes of the treatment assignment and/or the outcome will also suffice.

Given that an adjustment set that ensures unconfoundedness has been identified, many researchers have proposed dimension reduction procedures to increase efficiency while maintaining unconfoundedness (de Luna et al., 2011; VanderWeele & Shpitser, 2011), or to minimize mean squared error (Vansteelandt et al., 2012). Broadly, these methods tend to remove from the adjustment set covariates that are unassociated with the outcome.

2·3. *Stabilized inverse probability of treatment weighting*

A popular approach to causal effect estimation is to weight each unit's data by stabilized inverse probability of treatment weights $W_i = W_i(T_i, X_i) = \text{pr}(T_i)/\text{pr}(T_i \mid X_i)$ (Robins et al., 2000). The idea is that if the propensity function is known, the propensity function after weighting by W_i, $\text{pr}^*(T_i \mid X_i)$, will be equivalent to $\text{pr}(T_i)$ and hence does not depend on X_i, as shown in the Supplementary Material. Here * denotes the pseudo-population after weighting. Under the assumptions in § 2·1, weighting by W_i also preserves the causal effect of t^{P} on $E\{Y_i(t^{\text{P}})\}$ in the original population (Robins, 2000; Zhang et al., 2016), and so the causal effect can be consistently estimated without adjusting for X_i in the weighted data. For example, $E\{Y_i(t^{\text{P}})\}$ can be consistently estimated by modelling $E(Y_i \mid T_i)$ in the weighted data (Robins, 2000).

2·4. *Maximum likelihood estimation*

Estimating the weights by maximum likelihood involves specifying parametric models $\text{pr}(T_i; \alpha)$ and $\text{pr}(T_i \mid \tilde{X}_i; \beta)$, where $\tilde{X}_i$ are functionals of elements in X_i, and then estimating

the unknown parameters α and β by solving the score equations

$$S(\alpha) = \sum_{i=1}^{n} \frac{\partial}{\partial \alpha} \log \text{pr}(T_i; \alpha) = 0, \quad S(\beta) = \sum_{i=1}^{n} \frac{\partial}{\partial \beta} \log \text{pr}(T_i \mid \tilde{X}_i; \beta) = 0.$$

If $\text{pr}(T_i \mid \tilde{X}_i; \beta)$ and $\text{pr}(T_i; \alpha)$ are correctly specified, then the weights $W_i = W_i(T_i, \tilde{X}_i; \hat{\alpha}, \hat{\beta}) = \text{pr}(T_i; \hat{\alpha})/\text{pr}(T_i \mid \tilde{X}_i; \hat{\beta})$, where $\hat{\alpha}$ and $\hat{\beta}$ are the maximum likelihood estimates of α and β, are equivalent to $\text{pr}(T_i)/\text{pr}(T_i \mid \tilde{X}_i)$ asymptotically. Thus weighting by W_i will result in $\text{pr}^*(T_i \mid \tilde{X}_i)$ being asymptotically equivalent to $\text{pr}(T_i)$, i.e., T_i does not depend on $\tilde{X}_i$ after weighting. Here $\text{pr}^*(T_i \mid \tilde{X}_i)$ depends on $\hat{\alpha}$ and $\hat{\beta}$ through the estimated weights. However, when $\text{pr}(T_i \mid \tilde{X}_i; \beta)$ is misspecified, this estimation procedure not only will result in $\text{pr}^*(T_i \mid \tilde{X}_i)$ diverging from $\text{pr}(T_i)$ but also does not even guarantee that the association between $\tilde{X}_i$ and T_i is reduced in the weighted data relative to the observed data. Researchers are therefore encouraged to check for the absence of this association in the weighted data before proceeding to causal effect estimation. For nonbinary treatments, correctly specifying the propensity function model will generally also entail correct specification of the distribution and the dependence structure on covariates for higher-order moments of the treatment variable. Therefore, model misspecification for nonbinary treatments is likely to be worse.

3. METHODOLOGY

3·1. *General framework*

Maximum likelihood estimation indirectly aims to achieve the asymptotic equivalence of $\text{pr}^*(T_i \mid \tilde{X}_i)$ and $\text{pr}(T_i)$ by fitting a model $\text{pr}(T_i \mid \tilde{X}_i; \beta)$ for the propensity function. We instead propose to use weighting to directly eliminate the association between $\tilde{X}_i$ and T_i characterized by a chosen propensity function $\text{pr}(T_i \mid \tilde{X}_i; \beta)$ in the weighted data. When $\text{pr}(T_i \mid \tilde{X}_i; \beta)$ is misspecified, our method, in contrast to maximum likelihood estimation, will eliminate the association between $\tilde{X}_i$ and T_i as characterized by $\text{pr}(T_i \mid \tilde{X}_i; \beta)$ after weighting. This is necessary for T_i to be independent of $\tilde{X}_i$ after weighting. In the unlikely scenario that $\text{pr}(T_i \mid \tilde{X}_i; \beta)$ is correctly specified, maximum likelihood estimation will asymptotically eliminate the association between $\tilde{X}_i$ and T_i after weighting, while our method will eliminate their association in finite samples.

We now formalize our ideas. Given a set of known weights $W = (W_1, \ldots, W_n)$, we can fit a parametric propensity function model $\text{pr}\{T_i \mid \tilde{X}_i; \beta(W)\}$ to the data weighted by W by solving the score equations

$$\sum_{i=1}^{n} W_i \frac{\partial}{\partial \beta(W)} \log \text{pr}\{T_i \mid \tilde{X}_i; \beta(W)\} = 0, \tag{1}$$

where $\beta(W)$ is a vector of parameters. Here we write $\beta(W)$ as a function of W because the resulting maximum likelihood estimates, $\hat{\beta}(W)$, will depend on W. We use the uniquely parameterized propensity function assumption in § 2·1 to partition $\beta(W)$ into $\{\beta_{\text{b}}(W), \beta_{\text{d}}(W)\}$, where $\beta_{\text{d}}(W)$ are the unique parameters that characterize the dependence of T_i on $\tilde{X}_i$, e.g., regression coefficients, and $\beta_{\text{b}}(W)$ are parameters that characterize the baseline distribution, e.g., the intercept terms. Here the subscripts 'd' and 'b' stand for dependence and baseline, respectively. Without loss of generality, we assume $\text{pr}[T_i \mid \tilde{X}_i; \{\beta_{\text{b}}(W), \beta_{\text{d}}(W) = 0\}] = \text{pr}\{T_i; \beta_{\text{b}}(W)\}$. The conditions for covariate association eliminating weights are then derived by inverting (1) such that the weights

W satisfy the equations

$$\sum_{i=1}^{n} W_i \frac{\partial}{\partial \beta(W)} \log \mathrm{pr}\{T_i \mid \tilde{X}_i; \beta(W)\}\Big|_{\{\beta_\mathrm{b}(W)=\hat{\alpha},\, \beta_\mathrm{d}(W)=0\}} = 0, \tag{2}$$

where $\hat{\alpha}$ is obtained by fitting $\mathrm{pr}(T_i; \alpha)$ to the observed data. Our method therefore sets the goal of weighting as attaining $\hat{\beta}(W) = \{\hat{\beta}_\mathrm{b}(W) = \hat{\alpha},\ \hat{\beta}_\mathrm{d}(W) = 0\}$; that is, after weighting by W_i, (i) T_i is unassociated with $\tilde{X}_i$, as described by $\mathrm{pr}\{T_i \mid \tilde{X}_i; \beta(W)\}$, and (ii) the marginal distribution of T_i is preserved from the observed data, as described by $\mathrm{pr}(T_i; \alpha)$. Here (ii) is a statement concerning the projection function, the treatment assignment distribution in the weighted data. The choice of projection function, as long as it does not depend on $\tilde{X}_i$, does not affect the consistency of weighted estimators for causal treatment effects, although it does affect their efficiency (Peterson et al., 2010). In our method, the projection function may be altered by fixing $\beta_\mathrm{b}(W)$ in (2) at other values instead of $\hat{\alpha}$; an example is provided in § 4·1.

Our method is linked to the use of regression to assess covariate balance, for example in matched data (Lu, 2005). Specifically, if applying regression to matched data indicates small associations between the covariates and treatment assignment, it would be reasonable to assume that the covariate distributions are approximately balanced across treatment levels. Inspired by this, we propose to invert this covariate balance measure for weight estimation such that it implies no imbalances in covariate distributions across treatment levels, i.e., $\hat{\beta}_\mathrm{d}(W) = 0$, after weighting.

3·2. *Convex optimization for weight estimation*

We consider a convex optimization approach in the spirit of Zubizarreta (2015), intended to estimate minimum-variance weights subject to the conditions (2) being satisfied. We formulate the estimation problem as a quadratic programming task and use the lsei function in the R (R Development Core Team, 2018) package limSolve (Soetaert et al., 2009) to obtain the optimal weights. Specifically, we solve the following quadratic programming problem for W:

$$\text{minimize} \sum_{i=1}^{n} (W_i - 1)^2 \tag{3}$$

$$\text{subject to} \sum_{i=1}^{n} W_i \frac{\partial}{\partial \beta(W)} \log \mathrm{pr}\{T_i \mid \tilde{X}_i; \beta(W)\}\Big|_{\{\beta_\mathrm{b}(W)=\hat{\alpha},\, \beta_\mathrm{d}(W)=0\}} = 0, \tag{4}$$

$$\sum_{i=1}^{n} W_i = n, \qquad W_i \geqslant 0 \quad (i = 1, \ldots, n). \tag{5}$$

The constraints (4) are the equations (2), and so should eliminate the associations between covariates and treatment assignment after weighting. They also preserve the marginal distribution of the treatment variable in the observed data. The first constraint in (5) ensures equality of the numbers of units in the weighted and observed data. Since this also ensures that the mean of W is unity, (3) minimizes the variance of W. Interestingly, since the weights in the observed data are ones, (3) can also be interpreted as identifying the least extrapolated data as characterized by the L_2-norm. The second constraint in (5) requires that each element of W be nonnegative (Hainmueller, 2012; Zubizarreta, 2015). Allowing some elements of W to be zero can let the estimation problem be formulated as a convex quadratic programming problem. Then the estimation procedure could remove units which contribute greatly to the variability of the weights, while

forcing the remaining units to allow unbiased estimation of causal treatment effects (Crump et al., 2009). Following Zubizarreta (2015), it is possible to relax the strict equality constraints (4) to inequalities; the R function lsei has this option. This allows for less extrapolation at the expense of possibly introducing bias. For simplicity, we consider only the case where the strict equality constraints (4) are enforced.

4. RELATION TO PREVIOUS METHODS

4·1. *Binary treatment*

The moment conditions specified in existing covariate balancing weight methods (Imai & Ratkovic, 2014; Fong et al., 2018) are special cases of the conditions (2) after slight modifications. In this section we focus on binary treatments. The details for more general categorical treatments can be found in the Supplementary Material.

Unless stated otherwise in our derivations, we implicitly take $\beta_{\mathrm{d}}(W)$ and $\beta_{\mathrm{b}}(W)$ in (2) to be, respectively, the vector of regression coefficients of $\tilde{X}_i$ and the vector of the remaining parameters, including intercept terms, in the chosen propensity function model.

Let $X_i^* = (1, \tilde{X}_i^{\mathrm{T}})^{\mathrm{T}}$. When T_i is a binary treatment variable, i.e., $T_i = 1$ if the ith unit received treatment and $T_i = 0$ otherwise, the following covariate balancing conditions for the estimation of the propensity score have been proposed (Imai & Ratkovic, 2014):

$$\frac{1}{n}\sum_{i=1}^{n} W_i T_i X_i^* = \frac{1}{n}\sum_{i=1}^{n} W_i (1 - T_i) X_i^*. \tag{6}$$

Because X_i^* includes 1, these covariate balancing conditions constrain the number of units in the treated and control groups to be equal in the weighted data. The other conditions for $\tilde{X}_i$ constrain the weighted means of each element in $\tilde{X}_i$ in the treated and control groups to be equal.

These conditions can be derived using our framework. Suppose that we specify a logistic regression model for the propensity function/score, $\mathrm{pr}\{T_i = 1 \mid \tilde{X}_i; \beta(W)\} = 1/[1+\exp\{-\beta(W)^{\mathrm{T}} X_i^*\}]$, in the weighted data induced by a set of known weights W. This model can be fitted to the weighted data by solving the score equations

$$\sum_{i=1}^{n} W_i X_i^* \left[T_i - \frac{1}{1 + \exp\{-\beta(W)^{\mathrm{T}} X_i^*\}} \right] = 0.$$

Conditions can be derived by fixing $1/[1 + \exp\{-\beta(W)^{\mathrm{T}} X_i^*\}] = \hat{\pi}_0$, i.e., letting $\hat{\beta}_{\mathrm{d}}(W) = 0$, in these weighted score equations, where $\hat{\pi}_0$ is the proportion of units that received treatment in the observed data, and then inverting these equations so that we are solving for W:

$$\sum_{i=1}^{n} W_i X_i^* (T_i - \hat{\pi}_0) = 0. \tag{7}$$

The correspondence between (6) and (7) can then be established by changing the projection function from $\hat{\pi}_0$ to $1/2$.

4·2. *Continuous treatment*

When T_i is continuous on the real line, Fong et al. (2018) proposed the following covariate balancing conditions for weight estimation,

$$\frac{1}{n}\sum_{i=1}^{n} W_i X_i^{*\mathrm{c}} T_i^{\mathrm{c}} = 0, \tag{8}$$

where the superscript c denotes the centred version of the variable and $X_i^* = (1, \tilde{X}_i^{\mathrm{T}})^{\mathrm{T}}$. We now derive these covariate balancing conditions using the proposed framework. First, we assume a simple normal linear model for the propensity function, $T_i \mid \tilde{X}_i; \beta(W) \sim N\{\beta(W)^{\mathrm{T}} X_i^*, \sigma^2\}$. The score equations for this model in the weighted data are

$$\sum_{i=1}^{n} W_i X_i^* \left\{ \frac{T_i - \beta(W)^{\mathrm{T}} X_i^*}{\sigma^2} \right\} = 0, \quad \sum_{i=1}^{n} W_i \left[-1 + \frac{\{T_i - \beta(W)^{\mathrm{T}} X_i^*\}^2}{\sigma^2} \right] = 0.$$

By inverting these score equations, we find weights W that satisfy

$$\sum_{i=1}^{n} W_i X_i^* \left(\frac{T_i - \hat{\mu}_0}{\hat{\sigma}_0^2} \right) = 0, \quad \sum_{i=1}^{n} W_i \left\{ -1 + \frac{(T_i - \hat{\mu}_0)^2}{\hat{\sigma}_0^2} \right\} = 0, \tag{9}$$

where $\hat{\mu}_0$ and $\hat{\sigma}_0^2$ are the sample mean and variance of T_i. The first set of conditions in (9) is equivalent to the conditions (8), except that the X_i^* are not necessarily centred. The usefulness of our framework can also be exemplified by the insight it gives into how conditions can be specified for the variance of T_i. Specifically, suppose that we specify an alternative propensity function model $T_i \mid \tilde{X}_i; \beta_\mu(W), \beta_\sigma(W) \sim N[\beta_\mu(W)^{\mathrm{T}} X_i^*, \exp\{2\beta_\sigma(W)^{\mathrm{T}} X_i^*\}]$; that is, we allow the variance of T_i, σ_i^2, to depend on $\tilde{X}_i$ with $\sigma_i = \exp\{\beta_\sigma(W)^{\mathrm{T}} X_i^*\}$. For this model, the conditions for weight estimation are derived by setting the regression coefficient elements in $\beta_\mu(W)$ and $\beta_\sigma(W)$ to zero in the score equations. This corresponds to solving the equations

$$\sum_{i=1}^{n} W_i X_i^* \left(\frac{T_i - \hat{\mu}_0}{\hat{\sigma}_0^2} \right) = 0, \quad \sum_{i=1}^{n} W_i X_i^* \left\{ -1 + \frac{(T_i - \hat{\mu}_0)^2}{\hat{\sigma}_0^2} \right\} = 0. \tag{10}$$

Thus, the additional conditions in (10) are designed to remove the association between $\tilde{X}_i$ and the variance of T_i. More details can be found in § 6.

5. Other treatment types

Having demonstrated that our framework encompasses previously proposed work, we now widen its applicability by considering semicontinuous treatments, motivated by our application in § 7. Details about count treatments can be found in the Supplementary Material.

Semicontinuous variables are characterized by a point mass at zero and a right-skewed continuous distribution with positive support (Olsen & Schafer, 2001). Semicontinuous treatments are common in clinical settings because only treated patients will be prescribed a continuous dose of treatment; otherwise their dose will be recorded as zero (Moodie & Stephens, 2010). A common

approach to modelling semicontinuous data is by using a two-part model, such as that in Olsen & Schafer (2001):

$$\text{pr}(T_i \mid \tilde{X}_i; \pi_i, \mu_i, \sigma_i) = (1-\pi_i)^{I(T_i=0)} \left[\frac{\pi_i}{\sigma_i} \phi \left\{ \frac{g(T_i) - \mu_i}{\sigma_i} \right\} \right]^{I(T_i>0)}, \tag{11}$$

where $\pi_i = 1/[1 + \exp\{-\beta_\pi(W)^{\mathrm{T}} X_i^*\}]$, $\mu_i = \beta_\mu(W)^{\mathrm{T}} X_i^*$ and $\sigma_i = \exp\{\beta_\sigma(W)^{\mathrm{T}} X_i^*\}$ with $X_i^* = (1, \tilde{X}_i^{\mathrm{T}})^{\mathrm{T}}$, $\phi(\cdot)$ is the standard normal density function, and $I(\cdot)$ is an indicator function. Here $g(\cdot)$ is a monotonic function included to make the normal assumption for the positive values of T_i more tenable. The likelihoods for the binary and continuous components of the two-part model are separable, so the results in § 4 imply that the conditions based on (11) are

$$\sum_{i=1}^{n} W_i X_i^* \{ I(T_i > 0) - \hat{\pi}_0 \} = 0, \tag{12}$$

$$\sum_{i:T_i>0} W_i X_i^* \left\{ \frac{g(T_i) - \hat{\mu}_0}{\hat{\sigma}_0^2} \right\} = 0, \quad \sum_{i:T_i>0} W_i X_i^* \left[-1 + \frac{\{g(T_i) - \hat{\mu}_0\}^2}{\hat{\sigma}_0^2} \right] = 0, \tag{13}$$

where $\hat{\pi}_0$, $\hat{\mu}_0$ and $\hat{\sigma}_0$ are maximum likelihood estimates of π_i, μ_i and σ_i obtained by fitting (11), but without covariates, to the observed T_i. The conditions (12) are derived from the score equations for the binary component and are equivalent to (7). The conditions (13) are derived from the score equations of the continuous component and are similar to (10). In our framework, the weights W are estimated by solving (12) and (13) simultaneously, whereas maximum likelihood estimation obtains weights $W_{\mathrm{b}i}$ and $W_{\mathrm{c}i}$ separately from the binary and continuous components, respectively, and then uses their unit-specific product $W_i = W_{\mathrm{b}i} W_{\mathrm{c}i}$ as the final weight.

6. Simulation study

We consider the set-up where there are three independent standard normal pretreatment covariates X_{1i}, X_{2i} and X_{3i}. The treatment T_i is semicontinuous. We first simulate a binary indicator for $T_i > 0$ with $\text{pr}(T_i > 0) = 1/[1 + \exp\{-(0{\cdot}5 + X_{1i} + X_{2i} + X_{3i})\}]$. Then if $T_i > 0$, T_i is drawn from a normal distribution with mean $1 + 0{\cdot}5X_{1i} + 0{\cdot}2X_{2i} + 0{\cdot}4X_{3i}$ and standard deviation $\exp(0{\cdot}3 + 0{\cdot}3X_{1i} + 0{\cdot}1X_{2i} + 0{\cdot}2X_{3i})$. The outcome Y_i follows a negative binomial distribution

$$\text{pr}(Y_i = y_i) = \frac{\Gamma(y_i + 1/\theta)}{\Gamma(1/\theta) y_i!} \left(\frac{\theta\lambda_i}{1+\theta\lambda_i} \right)^{y_i} \left(\frac{1}{1+\theta\lambda_i} \right)^{1/\theta} \quad (y_i = 0, 1, \ldots), \tag{14}$$

where $\theta = 1$ and $\lambda_i = E(Y_i \mid T_i, X_{1i}, X_{2i}, X_{3i}) = \exp[-1 + 0{\cdot}5T_i + 2/\{1 + \exp(-3X_{1i})\} + 0{\cdot}2X_{2i} - 0{\cdot}2\exp(X_{3i})]$. Using this set-up, we generate four sets of 2500 simulated datasets with $n = 500$, 1000, 2500 and 4000.

For each of the four sets of simulations, we use the proposed method for semicontinuous treatments in § 5, referred to as Approach 1, and maximum likelihood estimation, referred to as Approach 2, with the propensity function model (11) to obtain the weights. For both methods we consider two different model structures and two sets of covariates for the propensity function model. The correct model structure A allows the mean and variance of T_i conditional on $T_i > 0$ to depend on covariates, whereas the incorrect model structure B restricts σ_i to be a constant σ. The first set of covariates are the correct covariates, $\tilde{X}_i = (X_{1i}, X_{2i}, X_{3i})^{\mathrm{T}}$, and the second set

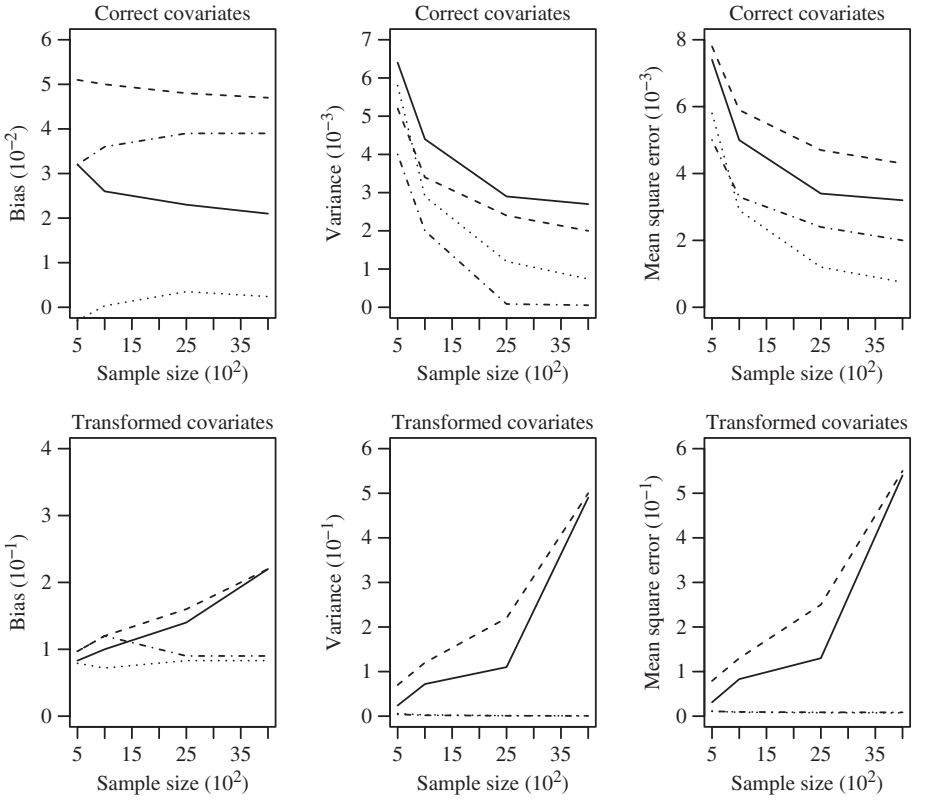

Fig. 1. Plots of the bias (left panels), empirical variance (middle panels) and mean squared error (right panels) of the treatment effect estimates as a function of sample size, for the correct (top panels) and transformed (bottom panels) covariate sets; in each panel the different line types represent Approach 1 with model structure A (dotted), Approach 1 with model structure B (dot-dash), Approach 2 with model structure A (solid), and Approach 2 with model structure B (dashed).

are transformed covariates of the form $\tilde{X}_i = (X_{1i}^{\mathrm{t}}, X_{2i}^{\mathrm{t}}, X_{3i}^{\mathrm{t}})^{\mathrm{T}}$, where $X_{1i}^{\mathrm{t}} = (1 + X_{1i} + X_{2i})^2$, $X_{2i}^{\mathrm{t}} = X_{2i}/\{1 + \exp(X_{1i})\}$ and $X_{3i}^{\mathrm{t}} = (X_{3i})^3$. The covariates X_{1i}^{t} and X_{3i}^{t} were chosen to be highly predictive of the outcome. In total, we fit eight models to each simulated dataset to estimate the weights, which are scaled by their averages so that they sum to n. Then, using the estimated weights, we fit a weighted negative binomial model (14) to the outcome with the marginal mean $\lambda_i = \exp(\gamma_0 + \gamma_1 T_i)$, where γ_0, γ_1 and θ are parameters to be estimated. The true causal treatment effect of interest is $\gamma_1 = 0{\cdot}5$.

The left panels of Fig. 1 show that the estimates from Approach 1 have smaller biases than those from Approach 2, particularly when the covariates are transformed. In this case, bias increases with sample size in Approach 2 but not in Approach 1. When the correct model structure A is used with the correct covariates, Approach 1 has smaller bias than Approach 2, perhaps because Approach 2 requires the law of large numbers to work well before the associations between covariates and treatment assignment can be eliminated after weighting.

The middle panels of Fig. 1 present the empirical variances of the treatment effect estimates. When the correct covariates are used, both approaches have small variances that decrease with sample size, although Approach 1 is better. Within each approach, estimates from the incorrect model structure B are less variable than those from model structure A, perhaps because the weights are less variable under model structure B as it has fewer degrees of freedom. The variances under Approach 1, but not under Approach 2, decrease with sample size when the transformed covariates

are used. This behaviour in Approach 2 with the transformed covariates is due to a few sets of estimated weights exacerbating the extremeness of the tails of the sampling distribution as the sample size increases; see Robins et al. (2007, pp. 553–4) for more details. Similar phenomena are observed with the mean squared error.

Because Approach 2 performs so poorly relative to Approach 1 when the transformed covariates are used, it is difficult to distinguish between the performances of Approach 1 under model structures A and B in Fig. 1, so we summarize the results here: within Approach 1, estimates from model structure A have smaller biases but larger variances than estimates from model structure B. Overall, estimates from model structure A have smaller mean squared errors for $n \geqslant 1000$.

In summary, in all examined scenarios, estimates based on the proposed method have smaller biases and variances than the maximum likelihood estimates.

7. APPLICATION

Steroids are effective and low-cost treatments for relieving disease activity in patients with systemic lupus erythematosus, a chronic autoimmune disease that affects multiple organ systems. However, there is evidence that steroid exposure could be associated with permanent organ damage that might not be attributable to disease activity. Motivated by a recent study (Bruce et al., 2015), we aim to estimate the causal dose-response relationship between steroid exposure and damage accrual shortly after disease diagnosis using data from the Systemic Lupus International Collaborating Clinics inception cohort.

We focus on 1342 patients who were enrolled between 1999 and 2011 from 32 sites and had at least two yearly assessments in damage accrual after disease diagnosis. The outcome of interest Y_i is the number of items accrued in the Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index during the period O_i, defined as the time interval between the two initial assessments in years. The semicontinuous treatment variable T_i is the average steroid dose per day in milligrams over the same time period. Based on clinical inputs, we consider the following pretreatment covariates X_i: the numerical scoring version of the British Isles Lupus Assessment Group disease activity index (Hay et al., 1993; Ehrenstein et al., 1995), which was developed according to the principle of physicians' intentions to treat; age at diagnosis in years; disease duration in years; and the groups of race/ethnicity and geographic region.

We consider the model (11) for T_i with $g(x) = \log(x + 1)$ to reduce right skewness in the positive steroid dose. For $\tilde{X}_i$, we include the main effects of X_i. We use the same four models as in the simulations to estimate weights. Further details can be found in the Supplementary Material.

We model Y_i with a weighted negative binomial model (14), where for examining the effect of presence of steroid use on damage accrual we specify $\lambda_i = E(Y_i \mid T_i; \gamma_0, \gamma_1) = O_i \exp\{\gamma_0 + \gamma_1 I(T_i > 0)\}$, and for examining the dose-response relationship we specify $\lambda_i = E(Y_i \mid T_i; \xi_0, \xi_1) = O_i \exp\{\xi_0 + \xi_1 \log(T_i + 1)\}$, with γ_0, γ_1, ξ_0, ξ_1 and θ parameters to be estimated. We construct 95% bootstrap percentile confidence intervals for parameter estimates using 1000 nonparametric bootstrap samples.

The results of the outcome regression models based on four sets of estimated weights are reported in Table 1. Consistent with current clinical evidence (Bruce et al., 2015), estimates from all weighted outcome regression models indicate that steroid use and the average steroid dose are positively and significantly associated with damage accrual in the initial period following diagnosis, although the estimated effect sizes from Approach 1 are larger than those from Approach 2.

Table 1. *Parameter estimates and* 95% *confidence intervals from fitting the weighted outcome regression models to the systemic lupus erythematosus data*

	Approach 1		Approach 2	
	Model structure A	Model structure B	Model structure A	Model structure B
		Binary treatment		
γ_0	−2·50 (−2·93, −2·19)	−2·50 (−2·93, −2·18)	−2·36 (−2·87, −1·98)	−2·36 (−2·87, −1·98)
γ_1	0·76 (0·41, 1·21)	0·73 (0·37, 1·19)	0·57 (0·10, 1·10)	0·57 (0·12, 1·10)
θ	1·60 (0·77, 2·53)	1·62 (0·83, 2·54)	1·30 (0·38, 2·48)	1·19 (0·23, 2·41)
		Semicontinuous treatment		
ξ_0	−2·57 (−2·93, −2·29)	−2·54 (−2·88, −2·27)	−2·47 (−2·89, −2·11)	−2·45 (−2·84, −2·12)
ξ_1	0·40 (0·24, 0·56)	0·37 (0·23, 0·52)	0·33 (0·16, 0·52)	0·32 (0·15, 0·49)
θ	1·41 (0·70, 2·27)	1·48 (0·75, 2·32)	1·17 (0·34, 2·22)	1·11 (0·28, 2·14)

Approach 1 also yields narrower confidence intervals than Approach 2. Within each approach, model structure B gives slightly smaller standard errors than model structure A.

We also fitted models with $I(T_i > 0)$ and $I(T_i > 0)\log(T_i + 1)$ included in the outcome regression. This provides the dose-response relationship after removing the patients unexposed to steroids (Greenland & Poole, 1995). The estimated dose-response functions were similar to those obtained from the models that include the semicontinuous treatment.

Overall, our results suggest that steroid dose is related to the damage accrual rate at the early disease stage of systemic lupus erythematosus. This suggests that clinicians might need to seek steroid-sparing therapies even at the early disease stage in order to reduce damage accrual.

8. Longitudinal setting

Our framework can be extended to the longitudinal setting. Here we give an example using a similar setting to that in Moodie & Stephens (2010). Suppose that for the ith unit ($i = 1, \ldots, n$), in each time interval $[s_{ij-1}, s_{ij})$ ($j = 1, \ldots, m_i; s_{i0} = 0$) of a longitudinal study, we observe in chronological order a vector of covariates X_{ij}, a time-varying treatment T_{ij} that can be of any distribution, and a longitudinal outcome Y_{ij}. The units are not necessarily followed up at the same time-points. Let the random variable $Y_{ij}(t_{ij}^{\mathrm{P}})$ be the potential outcome that would have arisen had treatment t_{ij}^{P} been administered in the time interval $[s_{ij-1}, s_{ij})$. We consider the setting where interest lies in estimating the direct causal effect of t_{ij}^{P} on $E\{Y_{ij}(t_{ij}^{\mathrm{P}})\}$, which may be confounded by histories of covariates, treatment assignment and response, $\bar{X}_{ij}$, $\bar{T}_{ij-1}$ and $\bar{Y}_{ij-1}$. Here an overbar represents the history of a process; for example, $\bar{X}_{ij} = (X_{i1}, \ldots, X_{ij})$.

In order to identify the direct causal effect of t_{ij}^{P} on $E\{Y_{ij}(t_{ij}^{\mathrm{P}})\}$, we make the sequential ignorability assumption, $\mathrm{pr}\{T_{ij} \mid Y_{ij}(t_{ij}^{\mathrm{P}}), \bar{X}_{ij}, \bar{T}_{ij-1}, \bar{Y}_{ij-1}\} = \mathrm{pr}(T_{ij} \mid \bar{X}_{ij}, \bar{T}_{ij-1}, \bar{Y}_{ij-1})$ for any time interval, and the positivity assumption, $\mathrm{pr}(T_{ij} \in \mathcal{A} \mid \bar{X}_{ij}, \bar{T}_{ij-1}, \bar{Y}_{ij-1}) > 0$ for all $\bar{X}_{ij}$, $\bar{T}_{ij-1}$ and $\bar{Y}_{ij-1}$ and for any set $\mathcal{A}$ with positive measure. Under these assumptions, the effect of t_{ij}^{P} on $E\{Y_{ij}(t_{ij}^{\mathrm{P}})\}$ can be consistently estimated by weighting the ith unit's data in the interval $[s_{ij-1}, s_{ij})$ with $W_{ij} = \mathrm{pr}(T_{ij})/\mathrm{pr}(T_{ij} \mid \bar{X}_{ij}, \bar{T}_{ij-1}, \bar{Y}_{ij-1})$ for all units and time intervals. The weights are typically estimated by $W_{ij} = \mathrm{pr}(T_{ij}; \hat{\alpha})/\mathrm{pr}(T_{ij} \mid \tilde{X}_{ij}; \hat{\beta})$, where $\tilde{X}_{ij}$ are functionals of $\bar{X}_{ij}$, $\bar{T}_{ij-1}$ and $\bar{Y}_{ij-1}$, and $\hat{\alpha}$ and $\hat{\beta}$ are the maximum likelihood estimates of α and β.

An alternative approach to weight estimation is to generalize our proposed method. Following the same strategy as in § 3·1, we assume that given known weights W_{ij}, the time-varying

propensity function in the time interval $[s_{ij-1}, s_{ij})$ follows a model $\text{pr}\{T_{ij} \mid \tilde{X}_{ij}; \beta(W)\}$. As in § 3·1, we partition $\beta(W)$ into $\{\beta_{\text{b}}(W), \beta_{\text{d}}(W)\}$, where $\beta_{\text{d}}(W)$ are regression coefficients that characterize the association between $\tilde{X}_{ij}$ and T_{ij} over time, and $\beta_{\text{b}}(W)$ are parameters that characterize the baseline distribution, e.g., the intercept terms. Similarly to before, conditions for weight estimation can be derived by inverting the weighted score equations

$$\sum_{i=1}^{n}\sum_{j=1}^{m_i} W_{ij}\frac{\partial}{\partial\beta(W)}\log \text{pr}\{T_{ij} \mid \tilde{X}_{ij}; \beta(W)\}\Big|_{\{\beta_{\text{b}}(W)=\hat{\alpha},\, \beta_{\text{d}}(W)=0\}} = 0, \tag{15}$$

where $\hat{\alpha}$ is obtained by fitting the model $\text{pr}(T_{ij}; \alpha)$ to the observed data for the time-varying treatment T_{ij}. Thus these conditions are designed to eliminate the association between $\tilde{X}_{ij}$ and T_{ij} and to preserve the observed marginal distribution of T_{ij} after weighting. Other projection functions that can help to further stabilize the weights, such as those that depend on some of the covariates, can also be considered in the proposed framework with only minor modifications. This would be useful when the interactions between these covariates and treatment are of interest in the outcome regression model. Finally, the convex optimization approach in § 3·2 can be used for weight estimation by replacing the equations (4) with (15). For large sample sizes, a parametric approach to solving the conditions (15) would be useful.

9. DISCUSSION

The proposed method has some limitations. First, both our method and existing covariate balancing weight methods treat covariates equally and balance them simultaneously. This can lead to poor performance in high-dimensional settings, so it would be of interest to incorporate different weights for the covariates when estimating the weights for the units. Second, it can be hard to detect near violations of the positivity assumption with our method, because it generally results in small variance of the causal effect estimates by exactly balancing the covariates. In such circumstances, e.g., when there is strong confounding, the results from our method can hide the fact that the observed data alone carry little information about the target causal effect and can have large bias under model misspecification because our method will almost entirely rely on extrapolation. It is therefore important to assess the positivity assumption when applying our framework. Third, our method does not necessarily estimate the causal effect for the population of interest, such as the target population of a health policy. This can be remedied by supplementing the conditions (4) with the additional conditions $\sum_{i=1}^{n} W_i\tilde{X}_i/n = c$, where c is the sample mean of $\tilde{X}_i$ with respect to the target population (Stuart et al., 2011).

ACKNOWLEDGEMENT

We thank two referees, the associate editor, the editor and Dr Shaun Seaman for helpful comments and suggestions. This work was supported by the U.K. Medical Research Council. We thank the Systemic Lupus International Collaborating Clinics for providing the data.

SUPPLEMENTARY MATERIAL

Supplementary material available at *Biometrika* online includes properties of stabilized inverse probability of treatment weights, derivations of conditions for other treatment distributions, simulation results and details of the application.

References

Bruce, I. N., O'Keeffe, A. G., Farewell, V., Hanly, J. G., Manzi, S., Su, L., Gladman, D. D., Bae, S.-C., Sanchez-Guerrero, J., Romero-Diaz, J. et al. (2015). Factors associated with damage accrual in patients with systemic lupus erythematosus: Results from the Systemic Lupus International Collaborating Clinics (SLICC) inception cohort. *Ann. Rheum. Dis.* **74**, 1706–13.

Chan, K. C. G., Yam, S. C. P. & Zhang, Z. (2016). Globally efficient nonparametric inference of average treatment effects by empirical balancing calibration weighting. *J. R. Statist. Soc.* B **78**, 673–700.

Crump, R. K., Hotz, V. J., Imbens, G. W. & Mitnik, O. A. (2009). Dealing with limited overlap in estimation of average treatment effects. *Biometrika* **96**, 187–99.

de Luna, X., Waernbaum, I. & Richardson, T. S. (2011). Covariate selection for the nonparametric estimation of an average treatment effect. *Biometrika* **98**, 861–75.

Ding, P. & Miratrix, L. W. (2015). To adjust or not to adjust? Sensitivity analysis of M-bias and butterfly-bias. *J. Causal Infer.* **3**, 41–57.

Ehrenstein, M. R., Conroy, S. E., Heath, J., Latchman, D. S. & Isenberg, D. A. (1995). The occurrence, nature and distribution of flares in a cohort of patients with systemic lupus erythematosus. *Br. J. Rheumatol.* **34**, 257–60.

Fong, C., Hazlett, C. & Imai, K. (2018). Covariate balancing propensity score for a continuous treatment: Application to the efficacy of political advertisements. *Ann. Appl. Statist.* **12**, 156–77.

Graham, B. S., Campos de Xavier Pinto, C. & Egel, D. (2012). Inverse probability tilting for moment condition models with missing data. *Rev. Econ. Stud.* **79**, 1053–79.

Greenland, S. & Poole, C. (1995). Interpretation and analysis of differential exposure variability and zero-exposure categories for continuous exposures. *Epidemiology* **6**, 326–8.

Hainmueller, J. (2012). Entropy balancing for causal effects: Multivariate reweighting method to produce balanced samples in observational studies. *Polit. Anal.* **20**, 24–46.

Hay, E., Bacon, P., Gordan, C., Isenberg, D. A., Maddison, P., Snaith, M. L., Symmons, D. P., Viner, N. & Zoma, A. (1993). The BILAG index: A reliable and validated instrument for measuring clinical disease activity in systemic lupus erythematosus. *Quart. J. Med.* **86**, 447–58.

Imai, K. & Ratkovic, M. (2014). Covariate balancing propensity score. *J. R. Statist. Soc.* B **76**, 243–63.

Imai, K. & Ratkovic, M. (2015). Robust estimation of inverse probability weights for marginal structural models. *J. Am. Statist. Assoc.* **110**, 1013–23.

Imai, K. & van Dyk, D. A. (2004). Causal inference with general treatment regimes: Generalizing the propensity score. *J. Am. Statist. Assoc.* **99**, 854–66.

Imbens, G. W. (2000). The role of the propensity score in estimating dose-response functions. *Biometrika* **87**, 706–10.

Kang, J. D. & Schafer, J. L. (2007). Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data (with Discussion). *Statist. Sci.* **22**, 523–39.

Lee, B. K., Lessler, J. & Stuart, E. A. (2010). Improving propensity score weighting using machine learning. *Statist. Med.* **29**, 337–46.

Lu, B. (2005). Propensity score matching with time-dependent covariates. *Biometrics* **61**, 721–8.

McCaffrey, D. F., Griffin, B. A., Almirall, D., Slaughter, M. E., Ramchard, R. & Burgette, L. F. (2013). A tutorial on propensity score estimation for multiple treatment using generalized boosted models. *Statist. Med.* **32**, 3388–414.

Moodie, E. M. & Stephens, D. A. (2010). Estimation of dose-response functions for longitudinal data using the generalised propensity score. *Statist. Meth. Med. Res.* **21**, 149–66.

Naimi, A. I., Moodie, E. E., Auger, N. & Kaufman, J. S. (2014). Constructing inverse probability weights for continuous exposures: A comparison of methods. *Epidemiology* **25**, 292–9.

Olsen, M. K. & Schafer, J. L. (2001). A two-part random effects model for semicontinuous longitudinal data. *J. Am. Statist. Assoc.* **96**, 730–45.

Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika* **82**, 669–88.

Peterson, M. L., Porter, K. E., Gruber, S. G., Wang, Y. & van der Laan, M. J. (2010). Diagnosing and responding to violations in the positivity assumption. *Statist. Meth. Med. Res.* **21**, 31–54.

R Development Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. http://www.R-project.org.

Robins, J., Sued, M., Lei-Gomez, Q. & Rotnitzky, A. (2007). Comment: Performance of doubly-robust estimators when "inverse probability" weights are highly variable. *Statist. Sci.* **22**, 544–59.

Robins, J. M. (2000). Marginal structural models versus structural nested models as tools for causal inference. In *Statistical Models in Epidemiology, the Environment and Clinical Trials*, M. E. Halloran & D. Berry, eds. New York: Springer, pp. 95–133.

Robins, J. M., Hernán, M. A. & Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology* **11**, 550–60.

Rosenbaum, P. R. & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**, 41–55.

SHPITSER, I., VANDERWEELE, T. J. & ROBINS, J. M. (2010). On the validity of covariate adjustment for estimating causal effects. In *Proceedings of the 26th Conference on Uncertainty and Artificial Intelligence*, P. Grünwald & P. Spirtes, eds. AUAI Press, pp. 527–36.

SOETAERT, K., VAN DEN MEERSCHE, K. & VAN OEVELEN, D. (2009). *limSolve: Solving Linear Inverse Models.* R-package version 1.5.1.

STUART, E. A., COLE, S. R., BRADSHAW, C. P. & LEAF, P. J. (2011). The use of propensity scores to assess the generalizability of results from randomized trials. *J. R. Statist. Soc.* A **174**, 369–86.

VANDERWEELE, T. J. & SHPITSER, I. (2011). A new criterion for confounder selection. *Biometrics* **67**, 1406–13.

VANSTEELANDT, S., BEKAERT, M. & CLAESKENS, G. (2012). On model selection and model misspecification in causal inference. *Statist. Meth. Med. Res.* **21**, 7–30.

ZHANG, Z., ZHOU, J., CAO, W. & ZHANG, J. (2016). Causal inference with a quantitative exposure. *Statist. Meth. Med. Res.* **25**, 315–35.

ZHU, Y., COFFMAN, D. L. & GHOSH, D. (2015). A boosting algorithm for estimating generalized propensity scores with continuous treatments. *J. Causal Infer.* **3**, 25–40.

ZUBIZARRETA, J. R. (2015). Stable weights that balance covariates for estimation with incomplete outcome data. *J. Am. Statist. Assoc.* **110**, 910–22.

[*Received on* 9 *January* 2017. *Editorial decision on* 26 *January* 2018]