[Working Paper: Please Do Not Cite or Circulate Without the Author's Permission]

Marginal Mean Weighting Adjustment for Selection Bias

Guanglei Hong

Ontario Institute for Studies in Education of the University of Toronto

Toronto, Ontario M5S 1V6, Canada

ghong@oise.utoronto.ca

SUMMARY

Defining causal effects as comparisons between marginal population means, I develop a method of using marginal mean weighting (MMW) to adjust for selection bias in observational studies. I show the inherent connection between MMW and inverse-probability-of-treatment weighting (IPTW). Both weighting methods are suitable for evaluating multiple concurrent treatments and time-varying treatments in observational data, hence have much broader applications than matching, stratification, or covariance adjustment for the propensity score of a binary treatment. Mathematical derivation and simulations reveal the advantages of the MMW adjustment method in reducing bias, in comparison with IPTW, especially when the functional form of the propensity model is misspecified.

KEYWORDS

Bias Reduction; Concurrent Treatments; Inverse-Probability-of-Treatment Weighting; Potential Outcomes; Propensity Score, Subclassification; Time-Varying Treatments

1. Introduction

The effect of one treatment is likely contingent upon some other concurrent treatments; later treatments are often built upon the earlier ones. The optimal combinations of concurrent treatments or the ideal sequences of time-varying treatments can be identified through factorial randomized experiments or sequential randomization, respectively. Without randomization, we may resort to methods of statistical adjustment if the observed data contain useful information about selection bias. Although well-designed quasi-experimental studies can be promising in removing a large amount if not all of the selection bias, statistical adjustment is challenging especially when there are multiple treatments or when it is necessary to adjust for time-varying covariates (Joffe and Rosenbaum, 1999; Robins, 1986).

In observational studies, the assignment to a treatment may be predicted by a vast number of covariates. An effective way to simplify the statistical adjustment for all those covariates is to summarize their information in a propensity score for each treatment when the strong ignorability assumption holds (Rosenbaum and Rubin, 1983), that is, when the treatment assignment is independent of the unmeasured covariates given the observed covariates. However, propensity score matching, stratification, or covariance adjustment, initially developed for assessing the effect of a binary treatment in observational data, are not suitable for evaluating a relatively large number of concurrent treatments or for handling time-varying covariates. Inverse-probability-of-treatment weighting (IPTW) (Robins, 2000; Robins, Hernan, and Siebert, 2003) is a seemingly

viable option for causal inference for multiple treatments or time-varying treatments. Yet its robustness to propensity model misspecification has not been closely examined.

Imbens (2000) proposed computing multiple propensity scores for multi-valued treatments and using each propensity score to estimate its corresponding population average potential outcome. Following this logic and extending the propensity score stratification method, I develop a strategy of using marginal mean weighting (MMW) to adjust for selection bias in observational studies. MMW combines some important strengths of the propensity score stratification method and the IPTW method, and therefore is expected to outperform the above two methods in many circumstances. With a crucial built-in step of identifying the empirical support for causal inference, MMW prevents the evaluation from being distorted by observations of units not belonging to the set of units at risk of receiving each treatment. Moreover, a comparison between MMW and IPTW reveals that, when the propensity model is correctly specified or when the omitted covariates are linear predictors of the logit of propensity, the MMW estimate of treatment effect converges to the IPTW estimate as the number of strata increases. However, when the propensity model fails to represent a nonlinear relationship, the IPTW adjustment leads to a bias proportional to the amount of confounding associated with the nonlinearity. In contrast, the MMW method effectively approximates a nonlinear relationship in a piecemeal manner and is hence robust to misspecification of the functional form of the propensity model.

This paper is organized as follows. Section 2 contrasts two different rationales for causal inference which, when applied to observational data, correspond to propensity

score stratification and IPTW. Section 3 compares the strengths and limitations between these two approaches. Section 4 introduces the marginal mean weighting method. I reveal its inherent connections with propensity stratification and IPTW, and explicate its applications to evaluations of multiple concurrent treatments and time-varying treatments. Section 5 compares the MMW estimates and the IPTW estimates in bias reduction when the propensity model is correctly or incorrectly specified. Section 6 summarizes the strengths and limitations of the MMW method and suggests issues for further investigation.

2. Two Different Rationales for Causal Inference

We consider an infinitely large population of interest. Our task is to compare the effect of treatment z versus z' on some outcome Y. According to Rubin's causal model (Holland, 1986; Rubin, 1978), the causal estimand,

$$\delta = E\{Y(z) - Y(z')\},$$

is, by definition, the population average of the unit-specific treatment effect. This is equivalent to the difference between the two respective population means associated with the two treatments, $E\{Y(z)\} - E\{Y(z')\}$. Causal inference is necessary because, in this most basic setup, only one potential outcome per unit can be observed.

Once we draw a random sample from the population, the first rationale for causal inference is to estimate the average conditional treatment effect, typically through a randomized block design in which levels of pretreatment covariates constitute blocks and treatment effects are estimated within blocks. Data can be analyzed with standard

methods such as multiple regression or analysis of covariance. The second rationale for causal inference is to estimate the marginal population mean of each treatment from the observed outcomes of units actually assigned to that treatment. This is straightforward in a completely randomized design as the average observed outcome of each treatment group provides an unbiased estimate of the population mean associated with that treatment.

In observational studies, treatment selection is typically associated with a large number of pretreatment covariates, denoted with **X**. Causal inference is nonetheless possible if the strong ignorability assumption holds,

$$Z \perp Y(z), Y(z') \mid \mathbf{X}.$$

Once the pretreatment information has been summarized in a propensity score $\theta = pr(Z = z \mid \mathbf{X})$ for each treatment, one can proceed in two ways corresponding to the above two rationales. Propensity score matching, stratification, and covariance adjustment estimate the conditional treatment effect within levels of the propensity score and then take an average over the distribution of the propensity score. Adjusting for the propensity score as a covariate is not highly recommended because it requires assumptions about the functional relationship between the covariate and the outcome (Drake, 1993). Here I focus on propensity score stratification (Rosenbaum and Rubin, 1984) and view matching as stratification at the most fine-grained level. Once we divide the propensity score distribution into K strata ($s = 1, \ldots, K$) such that the treatment groups z and z' have the same pretreatment composition, we have that,

$$\delta = E\left[E\{Y(z) \mid Z = z\} - E\{Y(z') \mid Z = z'\} \mid \theta\right]$$
$$\approx \sum_{s=1}^{K}\left[E\{Y(z) \mid Z = z, s\} - E\{Y(z') \mid Z = z', s\}\right] pr(s).$$

Alternatively, if we follow the second rationale for causal inference, selection in non-random treatment assignment can be adjusted through assigning a weight to each unit that is inverse to the unit's propensity of having the treatment actually received (Robins, 2000; Rosenbaum, 1987). As a result, the weighted sample of treatment group z will be representative of the population. So will the weighted sample of treatment group z'.

$$\delta = E\{Y(z)\} - E\{Y(z')\} = E\{WY(z) \mid Z = z\} - E\{WY(z') \mid Z = z'\},$$

where W is equal to

$$\begin{cases} \dfrac{pr(Z = z)}{\theta} \text{ if } Z = z, \text{ and} \\ \dfrac{pr(Z = z')}{1 - \theta} \text{ if } Z = z'. \end{cases} \quad (1)$$

Equation (1) defines the IPTW.

3. Propensity Score Stratification and IPTW: Strengths and Limitations

Despite the distinction between the two rationales, when the effect of a binary treatment is generalized to the same population, both propensity score stratification and IPTW results estimate the same parameter δ. However, we shall find that the first rational for causal inference has limitations in dealing with multiple treatments and time-varying treatments.

3.1. Multiple concurrent treatments

When following the first rationale, as the number of treatments goes up, it becomes increasingly impractical to estimate the average conditional treatment effects in an observational study. For example, let Z_1 and Z_2 denote two concurrent binary treatments without any additional constraints. Our goal is to simultaneously estimate their main effects and interaction effect: $\delta_1 = E\{Y(Z_1 = 1) - Y(Z_1 = 0)\}$;

$\delta_2 = E\{Y(Z_2 = 1) - Y(Z_2 = 0)\}$; and

$\delta_{1\times 2} = E\{Y(Z_1 = 1, Z_2 = 1) - Y(Z_1 = 1, Z_2 = 0) - Y(Z_1 = 0, Z_2 = 1) + Y(Z_1 = 0, Z_2 = 0)\}$.

When the strong ignorability assumption holds in the observational data, that is, $Z_1, Z_2 \perp Y(1,1), Y(0,1), Y(1,0), Y(0,0) \mid \mathbf{X}$, to approximate a factorial randomized design within levels of **X**, we will stratify the sample such that the compositions of all the four treatment groups are balanced within each stratum. Every unit will have a vector of four propensity scores: $\theta_{11} = pr(Z_1 = 1, Z_2 = 1 \mid \mathbf{X})$; $\theta_{10} = pr(Z_1 = 1, Z_2 = 0 \mid \mathbf{X})$;

$\theta_{01} = pr(Z_1 = 0, Z_2 = 1 \mid \mathbf{X})$; and $\theta_{00} = pr(Z_1 = 0, Z_2 = 0 \mid \mathbf{X}) = 1 - \theta_{11} - \theta_{10} - \theta_{01}$. Stratifying the sample on three of these four propensity scores, with five strata in each dimension, will generate as many as $5^3 = 125$ strata. Unless the sample is sufficiently large, data in many strata will be too sparse to allow for within-stratum estimation of all the causal estimands. If we want to study three binary treatments Z_1, Z_2, and Z_3 all at once, then a unit will have a vector of up to eight propensity scores. An evaluation of the interactions among the three treatments will require dividing the sample into $5^7 = 78{,}125$ strata, making causal inferences impossible in most cases.

3.2. Time-varying treatments

The first rational fails when the effects of time-varying treatments are confounded by time-varying covariates. Let Z_1 and Z_2 denote binary treatments at time 1 and time 2, respectively. Let $Y_1(z_1)$ and $Y_2(z_1, z_2)$ denote the potential outcomes at time 1 and time 2, respectively. Suppose that the causal effects of interest are $\delta_{1.1} = E\{Y_1(1) - Y_1(0)\}$, $\delta_{2.1} = E\{Y_2(1,0) - Y_2(0,0)\}$, $\delta_{2.2} = E\{Y_2(0,1) - Y_2(0,0)\}$, and $\delta_{2.1\times2} = E\{Y_2(1,1) - Y_2(1,0) - Y_2(0,1) + Y_2(0,0)\}$. To approximate a sequential randomized experiment, the treatment at each time needs to be independent of the subsequent potential outcomes given the observed baseline and time-varying covariates, that is, $Z_1 \perp Y_1(1), Y_1(0), Y_2(1,1), Y_2(0,1), Y_2(1,0), Y_2(0,0) \mid \mathbf{X}_1$, and $Z_2 \perp Y_2(1,1), Y_2(0,1), Y_2(1,0), Y_2(0,0) \mid Z_1, \mathbf{X}_1, \mathbf{X}_2, Y_1$. The key problem is that $\mathbf{X}_2$ and Y_1 are plausibly outcomes of Z_1, and in the meantime are confounders for Z_2. To estimate the effect of Z_2 on Y_2 would require statistical adjustment for $\mathbf{X}_2$ and Y_1. However, because these quantities are post-treatment covariates with regard to Z_1, an estimate of the association between Z_1 and Y_2 within levels of $\mathbf{X}_2$ and Y_1 would yield biased estimates of $\delta_{2.1}$ and $\delta_{2.1\times2}$ (Rosenbaum 1984; Robins, 1986).

The second rationale, represented by the IPTW method, makes possible causal inferences for multiple concurrent treatments and time-varying treatments. This is because, instead of estimating the treatment effects by directly conditioning on the pretreatment covariates, the strategy is to estimate the population average potential outcome of each treatment as if the whole population has been assigned to that treatment. However, applications of the IPTW method may introduce bias when units that have no

matches in another treatment group receive a nonzero weight, or when the functional form of the propensity model is mis-specified. In contrast, the nonparametric procedure of propensity score stratification effectively prevents these problems. The marginal mean weighting method, to be introduced in the next section, incorporates this nonparametric procedure in implementing the second rationale. I will compare the relative performance of MMW and IPTW in Section 5.

4. Marginal Mean Weighting Adjustment

4.1. Weighting on the basis of propensity score stratification

The concept of estimating the marginal mean potential outcome associated with each treatment was further developed by Imbens (2000) for evaluating the effects of multi-valued treatments in observational data. Let $D_i(z) = 1$ if unit i is assigned to treatment z, $D_i(z) = 0$ otherwise. The propensity of having treatment z is

$$\theta_z = pr\{D(z) = 1 | \mathbf{X}\}.$$

Under the weak ignorability assumption, introduced by Imbens (2000), that is,

$$D(z) \perp Y(z) | \theta_z,$$

the marginal mean potential outcome associated with treatment z can be represented as a function of the observed data,

$$E\{Y(z)\} = E[E\{Y(z) | \theta_z\}] = E[E\{Y(z) | D(z) = 1, \theta_z\}]. \quad (2)$$

Similarly, we can represent the marginal mean potential outcome of any other treatment z' in terms of the observed outcome and thereby evaluate the causal effect of treatment z versus z'. As indicated in Equation (2), to take the expectation of the conditional mean of

Y associated with treatment *z* requires computing an integral over the distribution of θz. Imbens (2000) provided no practical measures for such computation.

One possibility is to approximate the integration through stratification. To proceed, one can divide a sample into *K* strata. The marginal mean potential outcome of *Y*(*z*) is approximately equal to the within-stratum mean outcome of the treated units multiplied by the proportion of units in the stratum and summed over all the *K* strata. Huang, Frangakis, Dominici, Diette, and Wu (2005) applied this strategy to an evaluation of the effectiveness of 18 physicians.

Through algebraic manipulation, I derive a weighted mean of the observed outcome *Y* as an estimate of the marginal mean potential outcome *Y*(*z*), where the weight is the ratio of the number of units in stratum *s*, denoted with n_s, to the number of units in stratum *s* who were actually assigned to treatment *z*, denoted with $n_{z,s}$:

$$\begin{aligned} E\{Y(z)\} &\approx E\left[E\{Y(z) \mid D(z)=1, s\}\right] \\ &= \sum_{s=1}^{K} E\{Y(z) \mid D(z)=1, s\} pr(s) \\ &= \sum_{s=1}^{K} \left(\sum_{i \in z,s} \frac{Y_i}{n_{z,s}} \right) \frac{n_s}{N} \\ &= \frac{1}{N} \sum_{s=1}^{K} \left\{ \sum_{i \in z,s} \left(\frac{n_s}{n_{z,s}} \right) Y_i \right\} . \end{aligned} \tag{3}$$

This ratio of the two observed sample sizes, $n_s / n_{z,s}$, is an estimate of the true marginal mean weight,

$$\text{MMW} = \frac{pr(\theta_z)}{pr\{\theta_z, \text{D(z)}=1\}}.$$

The marginal mean weight is inherently connected to IPTW. To be specific, as $K \to \infty$, the estimated MMW converges to a multiple of IPTW as shown below.

$$\begin{aligned}\text{MMW} &= \frac{pr(\theta_z)}{pr\{\theta_z, \text{D(z)} = 1\}} \\ &= \frac{pr(\theta_z)}{pr\{D(z) = 1 \mid \theta_z\} pr(\theta_z)} \\ &= \frac{pr\{D(z) = 1\}}{pr\{D(z) = 1 \mid \theta_z\} pr\{D(z) = 1\}} \\ &= \text{IPTW} \times \frac{1}{pr\{D(z) = 1\}},\end{aligned}$$

where the inverse of $pr\{D(z) = 1\}$ is a constant. This result is important because the consistency of IPTW estimator, as proved by Robins (2000), can readily be applied to MMW estimator of treatment effect.

4.2. MMW for evaluating multiple concurrent treatments

The MMW adjustment method is convenient to apply in an evaluation of multiple concurrent treatments. For example, with two binary treatments Z_1 and Z_2, after estimating the four propensity scores for each unit—$\theta_{z_1 z_2}$ for $z_1, z_2 \in \{0,1\}$—corresponding to the four alternative treatment conditions denoted with $D(z_1, z_2)$, we identify the set of sampled units that have a nonzero probability of being assigned to each of these four treatment conditions and use them subsequently as the analytic sample. We then work with one treatment condition at a time. Take $D(1,1) = 1$ as an example. We divide the analytic sample into a number of strata on the basis of the estimated propensity score $\hat{\theta}_{11}$. Once within-stratum balance has been achieved between this treatment group and the rest of the analytic sample, following Equation (3), we compute a weight for each unit in this

treatment group. We then obtain a weighted mean outcome of the treated units as an estimate of the marginal mean potential outcome $E\{Y(1,1)\}$. The same procedure is to be repeated for the other three treatment conditions. A comparison between any pair of the estimated marginal mean potential outcomes provides an estimate of the relative effect of the two treatment conditions under consideration. The main effects and the interaction effect of Z_1 and Z_2 can be estimated simultaneously by analyzing a weighted general linear model. In general, in an evaluation of R concurrent treatments taking values z_r, $r = 1, \cdots, R$, under the assumption

$$D(z_1, \ldots, z_R) \perp Y(z_1, \ldots, z_R) \mid \theta_{z_1 \ldots z_R},$$

the marginal mean weight for unit i assigned to concurrent treatments $z_1, \ldots, z_R$ is

$$MMW_{i.z_1 \ldots z_R} = \frac{pr(\theta_{i.z_1 \ldots z_R})}{pr\{\theta_{i.z_1 \ldots z_R}, D_i(z_1, \ldots, z_R) = 1\}} \approx \frac{n_{s.z_1 \ldots z_R}}{n_{D(z_1 \ldots z_R), s.z_1 \ldots z_R}}, \qquad (4)$$

where $n_{s.z_1 \ldots z_R}$ is the number of units in stratum s as a result of stratifying the sample on the basis of $\theta(z_1 \ldots z_R)$; and $n_{D(z_1 \ldots z_R), s.z_1 \ldots z_R}$ is the number of units in stratum s who were actually assigned to the treatment condition $D_i(z_1 \ldots z_R) = 1$. When the number of strata goes to infinity, Equation (4) is equivalent to its corresponding IPTW multiplied by a constant $\{pr(z_1, \ldots, z_R)\}^{-1}$.

4.3. MMW for time-varying treatments

Because MMW follows the second rationale for causal inference, for the reasons stated in Section 3.2, we could apply this method to evaluations of time-varying treatments as well. In the case of a study of binary treatments Z_1 at time 1 and Z_2 at time

2, the propensity scores are $\theta_{z_1} = pr\{D(z_1)=1|\mathbf{X}_1\}$ and $\theta_{z_2} = pr\{D(z_2)=1|Z_1,\mathbf{X}_1,\mathbf{X}_2,Y_1\}$. Assuming $D(z_1) \perp Y_1(z_1)|\theta_{z_1}$ and $D(z_1,z_2) \perp Y_2(z_1,z_2)|\theta_{z_1},\theta_{z_2}$, we can obtain unbiased estimates of $E\{Y_1(z_1)\}$ and $E\{Y_2(z_1,z_2)\}$. The marginal mean weights for time-1 and time-2 observations, respectively, are

$$MMW_1 = \frac{pr(\theta_{z_1})}{pr\{\theta_{z_1}, D(z_1)=1\}}, \text{ and}$$

$$MMW_2 = \frac{pr(\theta_{z_1})}{pr\{\theta_{z_1}, D(z_1)=1\}} \times \frac{pr(\theta_{z_2})}{pr\{\theta_{z_2}, D(z_2)=1\}}.$$

In general, in a study over *T* time periods with *R* treatments in each time period, we assume that, for $t=1,\cdots,T$,

$$D(z_{11},\ldots,z_{1R},\cdots,z_{t1},\ldots,z_{tR}) \perp Y_t(z_{11},\ldots,z_{1R},\cdots,z_{t1},\ldots,z_{tR})|\theta_{z_{11},\ldots,z_{1R}},\cdots,\theta_{z_{t1},\ldots,z_{tR}}.$$

In order to estimate the marginal mean potential outcome at time *t* associated with the treatment sequence $z_{h1},\ldots,z_{hR}$, for $h=1,\cdots,t$, below is the general form of the marginal mean weight for unit *i* assigned to this treatment sequence:

$$\begin{aligned} MMW_{i.z_{t1}\ldots z_{tR}} &= \prod_{h=1}^{t} \frac{pr(\theta_{i.z_{h1}\ldots z_{hR}})}{pr\{\theta_{i.z_{h1}\ldots z_{hR}}, D_i(z_{h1},\ldots,z_{hR})=1\}} \\ &\approx \prod_{h=1}^{t} \frac{n_{s.z_{h1}\ldots z_{hR}}}{n_{D(z_{h1},\ldots,z_{hR}),s.z_{h1}\ldots z_{hR}}}, \end{aligned} \quad (5)$$

where $n_{s.z_{h1}\ldots z_{hR}}$ is the number of units in stratum *s* as a result of stratifying the sample on the basis of $\theta(z_{h1}\ldots z_{hR})$ at time *h*, $h=1,\cdots,t$; and $n_{D(z_{h1}\ldots z_{hR}),s.z_{h1}\ldots z_{hR}}$ is the number of units in stratum *s* who were actually assigned to the treatment condition $D_i(z_{h1}\ldots z_{hR})=1$ at

time h, $h = 1, \cdots, t$. When the number of strata goes to infinity, Equation (5) is equivalent to its corresponding IPTW multiplied by a constant $\{pr(z_{11}, \ldots, z_{1R}, \cdots, z_{t1}, \ldots, z_{tR})\}^{-1}$.

4.4 Treatment effect on the treated

When comparing a focal treatment z with some other alternative treatment z', the treatment effect on the treated needs to be integrated over the propensity distribution of the treated units.

$$\underset{z}{E}\{Y(z)\} - \underset{z}{E}\{Y(z')\} = \int_{i \in z} \int_{y} \{y(z) - y(z') \mid \theta_z\} f(y) g(\theta_z) dy d\theta_z .$$

The marginal mean potential outcome associated with treatment z for the treated population can be estimated as below where the weight is 1 for every treated unit; $pr(s_z)$ represents the number of treated units in stratum s as a proportion of the treated population of size N_z.

$$\begin{aligned} \underset{z}{E}\{Y(z)\} &\approx \underset{z}{E}\left[E\{Y(z) \mid D(z) = 1, s\}\right] \\ &= \sum_{s=1}^{K} E\{Y(z) \mid D(z) = 1, s\} pr(s_z) \\ &= \sum_{s=1}^{K} \left(\sum_{i \in z,s} \frac{Y_i}{n_{z,s}} \right) \frac{n_{z,s}}{N_z} \\ &= \frac{1}{N_z} \sum_{s=1}^{K} \left(\sum_{i \in z,s} Y_i \right) . \end{aligned}$$

The marginal mean potential outcome associated with treatment z' for the treated population is to be estimated as follows. For every unit in stratum s who were assigned to treatment z', the estimated MMW is a ratio of the sample size of treatment group z to that of treatment group z' in stratum s, $n_{z,s} / n_{z',s}$.

$$
\begin{aligned}
\underset{z}{E}\{Y(z')\} &\approx \underset{z}{E}\left[E\{Y(z') \mid D(z') = 1, s\}\right] \\
&= \sum_{s=1}^{K} E\{Y(z') \mid D(z') = 1, s\} pr(s_z) \\
&= \sum_{s=1}^{K} \left(\sum_{i \in z', s} \frac{Y_i}{n_{z',s}} \right) \frac{n_{z,s}}{N_z} \\
&= \frac{1}{N_z} \sum_{s=1}^{K} \left[\sum_{i \in z', s} \left(\frac{n_{z,s}}{n_{z',s}} \right) Y_i \right] .
\end{aligned}
$$

Hence, MMW is applicable to an estimation of the treatment effect on the treated as well as for estimating the population average treatment effect.

5. Comparison Between MMW and IPTW

In this section, I compare the relative performance of MMW with that of IPTW in bias reduction. In particular, I show the additional bias introduced in IPTW adjustment when only a portion of a population provides support for causal inference, or when the functional form of the propensity model is misspecified. These problems can be avoided in MMW adjustment due to its nonparametric approach.

5.1. Empirical support for causal inference

When the treatment is binary, usually the first step in propensity score stratification is to compare the distributions of the estimated logit of θ between the two treatment groups. We thereby identify the empirical support for causal inference and exclude units that have no counterfactual information in the data. In contrast, when IPTW is applied, the units with no matches in the alternative treatment group typically receive weights approximately equal to the proportions of sampled units assigned to their

respective treatment groups. That is, the estimated weight will be equal to $pr(Z=1)$ for the treated units with no matches, and will be equal to $pr(Z=0)$ for the control units with no matches. This will lead to a bias in the treatment effect estimate if the unmatched units have an essentially zero probability of receiving the alternative treatment. Let Ω denote the set of units that constitute the support for causal inference, that is, $0 < pr(Z_i = 1) < 1$ for $i \in \Omega$. I use $i \in Z$ to represent a treated unit and $i \in Z'$ for a control unit. Let $\hat{W}_i$ denote the estimated IPTW for unit *i*. Suppose that the propensity model is correctly specified, that is, $\hat{W}_i = W_i$. The IPTW estimate of the effect of treatment *z* versus *z′* is

$$\hat{\delta}_{IPTW} = \left(\frac{\sum_{i \in Z \cap i \in \Omega} \hat{W}_i Y_i}{\sum_{i \in Z \cap i \in \Omega} \hat{W}_i} - \frac{\sum_{i \in Z' \cap i \in \Omega} \hat{W}_i Y_i}{\sum_{i \in Z' \cap i \in \Omega} \hat{W}_i} \right) + \left(\frac{\sum_{i \in Z \cap i \notin \Omega} Y_i}{\sum_{i \in Z \cap i \notin \Omega} 1} - \frac{\sum_{i \in Z' \cap i \notin \Omega} Y_i}{\sum_{i \in Z' \cap i \notin \Omega} 1} \right) = \delta + \left(\frac{\sum_{i \in Z \cap i \notin \Omega} Y_i}{\sum_{i \in Z \cap i \notin \Omega} 1} - \frac{\sum_{i \in Z' \cap i \notin \Omega} Y_i}{\sum_{i \in Z' \cap i \notin \Omega} 1} \right).$$

Hence, the bias due to the lack of support is

$$\frac{\sum_{i \in Z \cap i \notin \Omega} Y_i}{\sum_{i \in Z \cap i \notin \Omega} 1} - \frac{\sum_{i \in Z' \cap i \notin \Omega} Y_i}{\sum_{i \in Z' \cap i \notin \Omega} 1}. \qquad (6)$$

5.2. Misspecification of the propensity model

For simplicity, in my discussion below, I examine the consequences of misspecifying the propensity model for a binary treatment—with $z = 1$ indicating the experimental condition and $z = 0$ for the control condition—at one time point only, assuming that the sample size is sufficiently large.

Correctly specified propensity model. When the propensity model is correctly specified, that is, when the true propensity score is known and is normally distributed, according to Cochran (1968), stratifying a sample into five subclasses typically removes about 90% of the initial bias. The percent bias reduction increases as the number of strata increases, and converges to 100% as the number of strata goes to infinity. With IPTW adjustment, 100% of the bias will be removed if the true propensity score is known for each unit and if the bias defined in (6) is zero.

Misspecified propensity model with omission of linear covariates. If the assumption of ignorable treatment assignment does not hold, the propensity model will be incorrectly specified due to the omission of some pretreatment covariates. Suppose that an omitted covariate *U* is linearly associated with the logit of propensity η and with the outcome *Y*. Also suppose that the entire population provides support for the causal inference. With either propensity score stratification or IPTW adjustment, additional bias due to the omission of *U* is $\alpha\Delta_U^2$, where ($\alpha\Delta_U$) is the linear association between *U* and *Y*, and Δ_U is the mean difference in *U* between the experimental group and the control group (see Appendix for derivation). If there is an omitted vector $\mathbf{U} = \begin{pmatrix} U_1 & \dots & U_Q \end{pmatrix}^T$ with *Q* independent elements that are all linearly associated with η and *Y*, the additional bias due to the omission of **U** will be $\alpha_1\Delta_{U_1}^2 + \cdots + \alpha_Q\Delta_{U_Q}^2$.

Misspecification of the functional form of the propensity model. When the functional form of the propensity model is misspecified, in particular, when the estimated propensity model fails to represent the nonlinear relationship between a pretreatment

covariate and the treatment assignment, I hypothesize that the IPTW estimate of the treatment effect typically contains more bias than the MMW estimate. I test this hypothesis through three sets of simulation studies. In all cases, I generate two normally distributed potential outcomes $Y(1)$ and $Y(0)$ corresponding to treatments $Z = 1$ and $Z = 0$. Both $Y(1)$ and $Y(0)$ are linear functions of a standard normal covariate X. To be specific,

$$Y(1) = 6 + 0.7X + \varepsilon(1);$$

$$Y(0) = 5 + 0.7X + \varepsilon(0);$$

$$\varepsilon(1), \varepsilon(0) \sim N(0, 0.25).$$

Hence, the true treatment effect is $E\{Y(1) - Y(0)\} = 1$. Let η denote a unit's true propensity, converted onto a logit scale, of being assigned to treatment $z = 1$ rather than $z = 0$. The treatment assignment indicator Z, taking on values $z \in \{0, 1\}$, is a Bernoulli random variable with its parameter equal to the true propensity $\theta = \{1 + \exp(-\eta)\}^{-1}$. In all three sets of simulations, the true propensity model, with η as the outcome, is a nonlinear function of X; while the misspecified propensity model estimates η as a linear function of X:

$$\hat{\eta} = \beta_0 + \beta_1 X. \qquad (7)$$

Each simulation consists of 5,000 samples of size 2,000.

In the first set of simulations, the actual treatment assignment is based on a dichotomized version of X, denoted with *DM_X*. Let *DM_X* take a value of 1 for units whose X values are at or above the median, and 0 otherwise. The true propensity model is

$$\eta = \alpha_0 + \alpha_1(DM\_X).$$

Hence, the true propensity scores have only two values: $\theta_{X+} = pr(Z=1 \mid X \geq 0)$; $\theta_{X-} = pr(Z=1 \mid X < 0)$. From analyzing the misspecified propensity model (7), the IPTW adjustment computes a weight for each unit that is inverse to the estimated propensity score $\hat{\theta} = \{1+\exp(-\hat{\eta})\}^{-1}$. That is, for unit *i*, the weight is

$$\hat{W}_i = \left\{ \frac{pr(Z=1)}{\hat{\theta}_i} \right\}^{Z_i} \left\{ \frac{pr(Z=0)}{1-\hat{\theta}_i} \right\}^{1-Z_i}.$$

Due to the misspecification of the propensity model, $\hat{\theta}$ becomes continuous rather than dichotomous. For the same reason, the estimated IPTW is continuous, although the true weight is dichotomous. The MMW adjustment divides the sample equally into six strata on the basis of $\hat{\eta}$, and then computes a weight for each unit as the ratio of the number of units in the same stratum with the focal unit to the number of units in that stratum who have been assigned to the same treatment group with the focal unit. For unit *i*, the weight is

$$\hat{M}_i = \left(\frac{n_s}{n_{z,s}} \right)^{Z_i} \left(\frac{n_s}{n_s - n_{z,s}} \right)^{1-Z_i}.$$

I conduct a series of four simulations, and compare the consequent bias in the estimated treatment effects between the IPTW method and the MMW method as the nonlinear confounding represented by α_1 increases. Table 1 summarizes the results of these four simulations. The naïve estimate is the mean difference in the observed outcome between the treatment group and the control group with no adjustment for selection bias. In all the four simulations, IPTW over corrects for the positive bias and

produces a considerable amount of negative bias. Clearly, as the amount of nonlinear confounding increases, we observe an increasing amount of bias in the IPTW estimate of treatment effect, along with an increase in the standard error. In comparison, the MMW adjustment based on six strata generates relatively stable estimates of the treatment effect with a minimal amount of bias and a relatively high level of precision throughout the four simulations. Using data from one sample in the first of these simulations, I plot the true weight, the IPTW, and the MMW across the range of X. Figures 1.a, 1.b, and 1.c reveal the deviation of IPTW from the true weight in contrast with the close approximation of MMW to the true weight.

The second set of simulations assigns the treatment according to a polynomial propensity model specified as follows:

$$\eta = \alpha_0 + \alpha_1 X + \alpha_2 X^2.$$

The mis-specified propensity model omits the quadratic term, hence takes the same linear form as that in (7). In each random sample, I identify units with no counterfactual information and assign them a zero weight in both IPTW adjustment and MMW adjustment. The results consistently show that the MMW estimates based on six strata contain less bias than the IPTW estimates. For example, when $\alpha_0 = 1$, $\alpha_1 = 0.6$, and $\alpha_2 = -0.6$, the mean naïve estimate is 1.33 (standard error = 0.04). The mean IPTW estimate is 0.92 (standard error = 0.04) with 123.85% of the initial bias removed. In other words, the IPTW method introduces 23.85% of bias. In comparison, the mean MMW6 estimate is 1.00 (standard error = 0.03) with 98.69% of the initial bias removed.

The third set of simulations specifies a nonlinear propensity model as follows:

$$\eta = \alpha_0 + \alpha_1 X + \alpha_2 \log(X^2 + 1).$$

The mis-specified propensity model, taking the same form as Equation (7), does not include the nonlinear term $\log(X^2 + 1)$. Again, for both IPTW and MMW, I assign a zero weight to units who have no matches in the alternative treatment group. The MMW estimates based on six strata outperform the IPTW estimates in bias reduction. For example, when $\alpha_0 = -2$, $\alpha_1 = -1$, and $\alpha_2 = 1$, the mean naïve estimate is 0.30 (standard error = 0.05), the mean IPTW estimate is 1.27 (standard error = 0.10) with 139.24% of bias reduction, and the mean MMW6 estimate is 0.96 (standard error = 0.03), removing 93.56% of the initial bias.

6. Conclusion and Questions for Further Investigation

In sum, MMW estimate of treatment effect is consistent when the number of strata approaches infinity. MMW adjustment can be used to estimate both population average treatment effect and treatment effect on the treated. Through estimating the marginal mean potential outcome of each treatment, the MMW adjustment method is not only convenient for comparing the effects of multiple concurrent treatments but also capable of handling time-varying covariates. Hence, MMW has broader applications than propensity score matching, stratification, and covariance adjustment.

Furthermore, when we apply the MMW method, the built-in procedures of stratifying the sample and of examining within-stratum balance in pretreatment covariates between treatment groups ensure that, after excluding units that do not supply empirical support for causal inference, the weighted sample of the treated units and the control

units approximates a completely randomized experiment under the assumption of ignorable treatment assignment given the observed covariates. Even when a nonlinear propensity model is misspecified as a linear one, the MMW estimate of treatment effect remains robust. This is because MMW is not computed as a function of the estimated propensity score itself. Rather, after dividing the joint distribution of the covariates into segments, MMW is estimated as a ratio of the sample sizes within each segment. Hence, misspecifying the functional form of the propensity model has a comparatively little impact on MMW adjustment. In contrast, as I have shown in Section 5, IPTW estimation of treatment effect is biased when units with no counterfactual information in the data receive a nonzero weight, or when the propensity model fails to represent a nonlinear relationship between pretreatment covariates and treatment assignment. In both cases, because IPTW is prone to systematic errors in the estimated propensity score, the weighted sample of the treated units and the control units will remain imbalanced, leading to a considerable amount of bias in treatment effect estimation.

Propensity score stratification has other desirable features including its convenience in examining heterogeneity of treatment effect and improved efficiency through precision weighting. Future research needs to investigate possibilities of incorporating some of these features into the MMW adjustment. More research is also needed in comparing properties of MMW estimator of treatment effect with those of IPTW estimator under a variety of circumstances. In general, there are three possible reasons why an estimated propensity score may deviate from the true propensity score: (1) misspecification of the propensity model; (2) measurement error in pretreatment

covariates; and (3) sampling error. The current study has focused on the robustness of treatment effect estimate in the presence of (1). Although researchers can manage to reduce (2) and (3) by improving measurement tools and increasing sample size, respectively, these improvements may not always be feasible. Further research issues therefore include: How do MMW and IPTW compare when there are measurement errors in pretreatment covariates? How do MMW and IPTW compare in analyses of small samples? And finally, which method is relatively more efficient in estimation? These properties are to be examined in evaluations of both time-invariant and time-varying treatments, and of treatments conducted at the individual level or at the organization level in multi-level settings.

ACKNOWLEDGEMENTS

This work was supported by a major research grant from the Spencer Foundation and a National Academy of Education/Spencer Postdoctoral Fellowship. The author owes thanks to Colm O'Muircheartaigh, Stephen Raudenbush, Tyler VanderWeele, and Kazuo Yamaguchi for helpful comments and suggestions.

APPENDIX

Bias Due to the Omission of Pretreatment Linear Covariate *U*

Let us assume that two pretreatment covariates *X* and *U* are linearly independent:

$X_{(z=1)} \sim N(\Delta_X,1)$; $X_{(z=0)} \sim N(0,1)$; $U_{(z=1)} \sim N(\Delta_U,1)$; $U_{(z=0)} \sim N(0,1)$.

Below I use $\{z=1\}$ and $\{z=0\}$ to denote the sets of units assigned to the experimental condition and the control condition, respectively. The average probability of being assigned to $z = 1$ rather than $z = 0$ is denoted by p.

Let the true propensity model be:

$$\eta = (X - p\Delta_X)\beta + (U - p\Delta_U)\gamma.$$

Hence,

$$\eta_{(z=1)} \sim N\left((1-p)(\beta\Delta_X + \gamma\Delta_U), \beta^2 + \gamma^2\right);$$
$$\eta_{(z=0)} \sim N\left(-p(\beta\Delta_X + \gamma\Delta_U), \beta^2 + \gamma^2\right).$$

$$E\{\eta_{(z=1)} - \eta_{(z=0)}\} = \beta\Delta_X + \gamma\Delta_U.$$

Alternatively, by definition, we have that,

$$E\{\eta_{(z=1)}\} = \lim_{X\to\Delta_X, U\to\Delta_U}\left\{\log\frac{\Pr(Z=1\mid X=\Delta_X, U=\Delta_U)}{\Pr(Z=0\mid X=\Delta_X, U=\Delta_U}\right\} = \log f_{\eta_{(z=1)}}(\Delta_X,\Delta_U) - \log f_{\eta_{(z=0)}}(\Delta_X,\Delta_U);$$
$$E\{\eta_{(z=0)}\} = \lim_{X\to 0, U\to 0}\left\{\log\frac{\Pr(Z=1\mid X=0, U=0)}{\Pr(Z=0\mid X=0, U=0]}\right\} = \log f_{\eta_{(z=1)}}(0,0) - \log f_{\eta_{(z=0)}}(0,0).$$

Substituting the normal density functions for $f_{\eta_{(z=1)}}(\Delta_X,\Delta_U)$, $f_{\eta_{(z=0)}}(\Delta_X,\Delta_U)$, $f_{\eta_{(z=1)}}(0,0)$, and $f_{\eta_{(z=0)}}(0,0)$, we can derive that,

$$E[\eta_{(z=1)} - \eta_{(z=0)}] = \frac{(\beta\Delta_X + \gamma\Delta_U)^2}{\beta^2 + \gamma^2}.$$

Solving $\beta\Delta_X + \gamma\Delta_U = \left(\beta\Delta_X + \gamma\Delta_U\right)^2 / \left(\beta^2 + \gamma^2\right)$, we find that $\beta = \Delta_X$ and $\gamma = \Delta_U$.

The *outcome model* is specified as follows:

$$Y = \delta Z + \alpha\eta + \varepsilon, \qquad \varepsilon \sim N(0, \sigma^2).$$

Because

$$E\left\{Y_{(z=1)}\right\} = \delta + \alpha E\left\{\eta_{(z=1)}\right\} = \delta + (1-p)\alpha\left(\Delta_X^2 + \Delta_U^2\right),$$

$$E\left\{Y_{(z=0)}\right\} = \alpha E\left\{\eta_{(z=0)}\right\} = -p\alpha\left(\Delta_X^2 + \Delta_U^2\right),$$

the *initial bias* in treatment effect estimate is

$$E\left\{Y_{(z=1)}\right\} - E\left\{Y_{(z=0)}\right\} - \delta = \alpha\left(\Delta_X^2 + \Delta_U^2\right).$$

A misspecified propensity model omits *U*, that is, $\hat{\eta} = \left(X - p\Delta_X\right)\beta$, where $\beta = \Delta_X$. With IPTW adjustment for the estimated propensity score, the expected value of the treatment effect estimate is

$$\begin{aligned}
&E\left(\sum_{i\in\{z=1\}} \frac{\hat{W}_i Y_i}{\sum_{i\in\{z=1\}} \hat{W}_i} - \sum_{i\in\{z=0\}} \frac{\hat{W}_i Y_i}{\sum_{i\in\{z=0\}} \hat{W}_i}\right) \\
&= E\left[\sum_{i\in\{z=1\}} \frac{\hat{W}_i\left\{\delta + \alpha\Delta_X\left(X_i - p\Delta_X\right) + \alpha\Delta_U\left(U_i - p\Delta_U\right) + \varepsilon_i\right\}}{\sum_{i\in\{z=1\}} \hat{W}_i}\right] \\
&\qquad - E\left[\sum_{i\in\{z=0\}} \frac{\hat{W}_i\left\{\alpha\Delta_X\left(X_i - p\Delta_X\right) + \alpha\Delta_U\left(U_i - p\Delta_U\right) + \varepsilon_i\right\}}{\sum_{i\in\{z=0\}} \hat{W}_i}\right] \\
&= \delta + \alpha\Delta_X E\left(\frac{\sum_{i\in\{z=1\}} \hat{W}_i X_i}{\sum_{i\in\{z=1\}} \hat{W}_i} - \frac{\sum_{i\in\{z=0\}} \hat{W}_i X_i}{\sum_{i\in\{z=0\}} \hat{W}_i}\right) + \alpha\Delta_U E\left(\frac{\sum_{i\in\{z=1\}} \hat{W}_i U_i}{\sum_{i\in\{z=1\}} \hat{W}_i} - \frac{\sum_{i\in\{z=0\}} \hat{W}_i U_i}{\sum_{i\in\{z=0\}} \hat{W}_i}\right) + E\left(\frac{\sum_{i\in\{z=1\}} \hat{W}_i \varepsilon_i}{\sum_{i\in\{z=1\}} \hat{W}_i} - \frac{\sum_{i\in\{z=0\}} \hat{W}_i \varepsilon_i}{\sum_{i\in\{z=0\}} \hat{W}_i}\right).
\end{aligned}$$

(A1)

In large samples, we have that

$$E\left(\sum_{i\in\{z=1\}}\hat{W}_i X_i\right)=\sum_{i\in\{z=1\}}E\left[p\{1+\exp(-\hat{\eta}_i)\}X_i\right]$$
$$=p\sum_{i\in\{z=1\}}\left[E(X_i)+E\left\{\exp\left(-X_i\Delta_X+p\Delta_X^2\right)X_i\right\}\right]$$
$$=p\sum_{i\in\{z=1\}}\left[\Delta_X+\exp\left(p\Delta_X^2\right)E\left\{\exp(-X_i\Delta_X)X_i\right\}\right].$$

By expanding the power series for $E\left\{\exp(-X_i\Delta_X)X_i\right\}$, we derive the result

$$E\left(\sum_{i\in\{z=1\}}\hat{W}_i X_i\right)\cong p\Delta_X\sum_{i\in\{z=1\}}1.$$

Similarly, we can show that

$$E\left(\sum_{i\in\{z=0\}}\hat{W}_i X_i\right)\cong p\Delta_X\sum_{i\in\{z=0\}}1.$$
$$E\left(\sum_{i\in\{z=0\}}\hat{W}_i\right)\cong p\sum_{i\in\{z=0\}}1.$$

Hence, the second term in (A1) is

$$\alpha\Delta_X\left(\frac{\sum_{i\in\{z=1\}}\hat{W}_i X_i}{\sum_{i\in\{z=1\}}\hat{W}_i}-\frac{\sum_{i\in\{z=0\}}\hat{W}_i X_i}{\sum_{i\in\{z=0\}}\hat{W}_i}\right)\cong 0.$$

When U is omitted from the propensity model,

$$E\left(\sum_{i\in\{z=1\}}\hat{W}_i U_i\right)=\sum_{i\in\{z=1\}}E\left[p\{1+\exp(-\hat{\eta}_i)\}U_i\right]$$
$$=p\sum_{i\in\{z=1\}}\left[E(U_i)+E\left\{\exp\left(-X_i\Delta_X+p\Delta_X^2\right)U_i\right\}\right]$$
$$=p\sum_{i\in\{z=1\}}\left[\Delta_U+\exp\left(p\Delta_X^2\right)E\left\{\exp(-X_i\Delta_X)U_i\right\}\right]$$
$$\cong p\Delta_U\sum_{i\in\{z=1\}}1.$$

$$
\begin{aligned}
&E\left(\sum_{i\in\{z=0\}}\hat{W}_i U_i\right)\\
&= p\sum_{i\in\{z=0\}}\left[E(U_i)+E\left\{\exp\left(-X_i\Delta_X+p\Delta_X^2\right)U_i\right\}\right]\\
&= p\sum_{i\in\{z=0\}}\left[0+\exp\left(p\Delta_X^2\right)E\left\{\exp\left(-X_i\Delta_X\right)U_i\right\}\right]\\
&\cong 0.
\end{aligned}
$$

Hence, the third term in (A1) is

$$
\alpha\Delta_U\left[\frac{\sum_{i\in\{z=1\}}\hat{W}_i U_i}{\sum_{i\in\{z=1\}}\hat{W}_i}-\frac{\sum_{i\in\{z=0\}}\hat{W}_i U_i}{\sum_{i\in\{z=0\}}\hat{W}_i}\right]=\alpha\Delta_U{}^2.
$$

Finally,

$$
\begin{aligned}
&E\left(\sum_{i\in\{z=1\}}\hat{W}_i\varepsilon_i\right)\cong 0;\\
&E\left(\sum_{i\in\{z=0\}}\hat{W}_i\varepsilon_i\right)\cong 0.
\end{aligned}
$$

Hence, the fourth term in (A1) is

$$
\frac{\sum_{i\in\{z=1\}}\hat{W}_i\varepsilon_i}{\sum_{i\in\{z=1\}}\hat{W}_i}-\frac{\sum_{i\in\{z=0\}}\hat{W}_i\varepsilon_i}{\sum_{i\in\{z=0\}}\hat{W}_i}\cong 0.
$$

In conclusion,

$$
\sum_{i\in\{z=1\}}\frac{\hat{W}_i Y_i}{\sum_{i\in\{z=1\}}\hat{W}_i}-\sum_{i\in\{z=0\}}\frac{\hat{W}_i Y_i}{\sum_{i\in\{z=0\}}\hat{W}_i}\cong\delta+\alpha\Delta_U^2.
$$

Therefore, in large samples with IPTW adjustment, the bias due to the omission of U is $\alpha\Delta_U^2$.

MMW based on *K* strata, when *K* goes to infinity, is equivalent to IPTW in bias reduction. When *K* is relatively small such as having five strata, there will be up to 10% bias remaining within strata (Cochran, 1968). Other than that, the additional amount of bias attributable to *U* is $\alpha\Delta_U^2$.

REFERENCES

Cochran, W. G. (1968). The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics*, **24**, 295-313.

Drake, C. (1993). Effects of misspecification of the propensity score on estimators of treatments effects. *Biometrics*, **49**, 1231-1236.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, **81**, 945-960.

Huang, I-C., Frangakis, C., Dominici, F., Diette, G. B., and Wu, A. W. (2005). Approach for risk adjustment in profiling multiple physician groups on asthma care. *Health Services Research*, **40**, 253-278.

Imbens, G. W. (2000). The role of the propensity score in estimating dose-response functions. *Biometrika*, **87**, 706-710.

Joffe, M. M. and Rosenbaum, P. R. (1999). Invited commentary: Propensity scores. *American Journal of Epidemiology*, **150**, 327-333.

Robins, J. M. (1986). A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical Modeling*, **7**, 1393-1512.

Robins, J. M. (2000). Marginal structural models versus structural nested models as tools for causal inference. In M. Elizabeth Halloran and Donald Berry (Eds.), *Statistical Models in Epidemiology, the Environment, and Clinical Trials*. New York: Springer, 95-134.

Robins, J. M., Hernan, M., and Siebert, U. (2003). Effects of multiple interventions. *Population Health Metrics*, **2**, 2191-2230.

Rosenbaum, P. R. (1984), The consequences of adjustment for a concomitant variable that has been affected by the treatment. *Journal of the Royal Statistical Society, Series A*, **147**, 656-666.

Rosenbaum, P. R. (1987). Model-based direct adjustment. *Journal of the American Statistical Association*, **82**, 387-394.

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

Rosenbaum, P. R., and Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association*, **79**, 516-524.

Rubin, D. B. (1978), Bayesian inference for causal effects: The role of randomization. *The Annals of Statistics*, **6**, 34-58.

Table 1

Treatment Effect Estimate and Percent Bias Reduction: Simulations of Nonlinear Propensity Models[a]

Simulations	Parameters		Naive	IPTW[b]		MMW6[c]	
	α_0	α_1	Mean Estimate (*SE*)	Mean Estimate (*SE*)	% bias reduction	Mean Estimate (*SE*)	% bias reduction
I.1	-0.5	1	1.27 (0.04)	0.99 (0.02)	104.23%	1.00 (0.02)	98.29%
I.2	-1	2	1.52 (0.04)	0.92 (0.03)	115.91%	1.00 (0.03)	99.40%
I.3	-1.5	3	1.71 (0.03)	0.66 (0.10)	147.94%	1.00 (0.03)	99.30%
I.4	-2	4	1.85 (0.03)	0.03 (0.29)	213.40%	1.01 (0.04)	99.37%

Notes:

[a] The true propensity model is $\eta = \alpha_0 + \alpha_1(DM\_X)$. The treatment effect is $\delta = 1$.

[b] IPTW as a function of the propensity score estimated from the misspecified propensity model $\hat{\eta} = \beta_0 + \beta_1 X$.

[c] MMW for six strata on the basis of the propensity score estimated from the misspecified propensity model $\hat{\eta} = \beta_0 + \beta_1 X$.

Figure 1.a

True Weight Based on True Propensity Model $\eta = \alpha_0 + \alpha_1(DM\_X)$

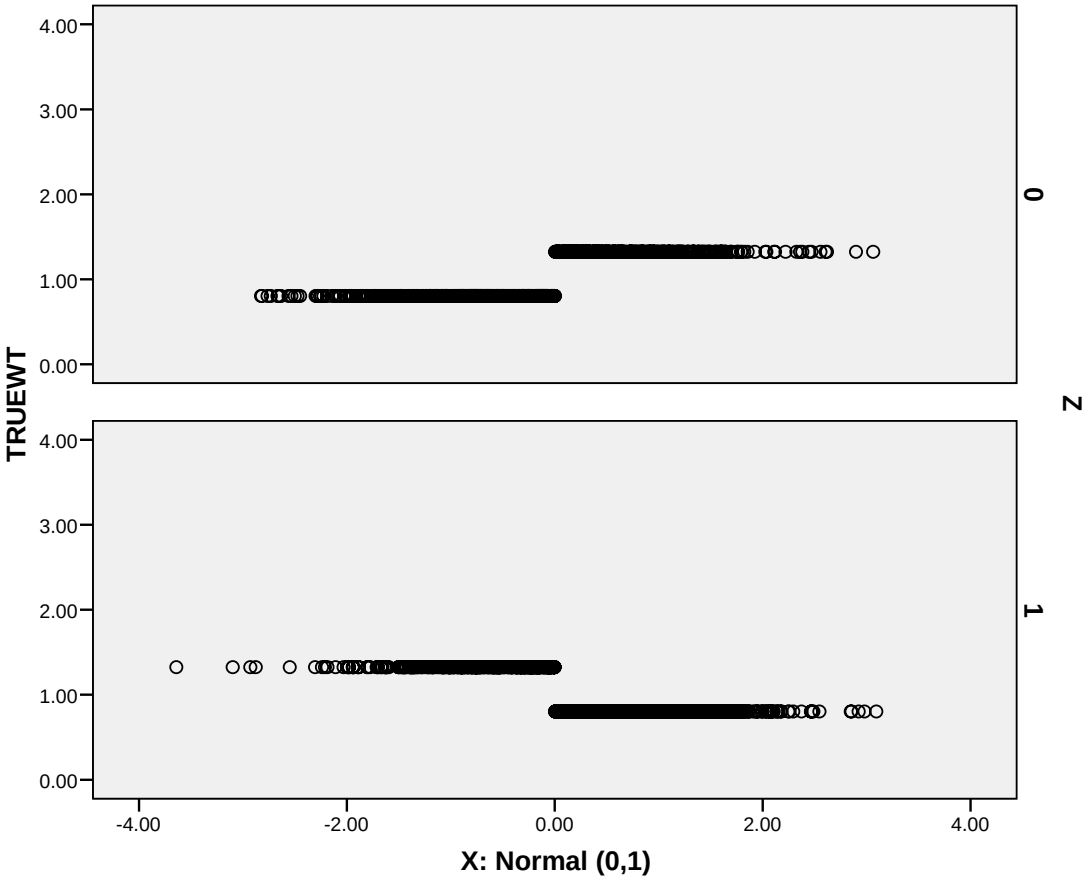

Figure 1.b

IPTW Based on Misspecified Propensity Model $\hat{\eta} = \beta_0 + \beta_1 X$

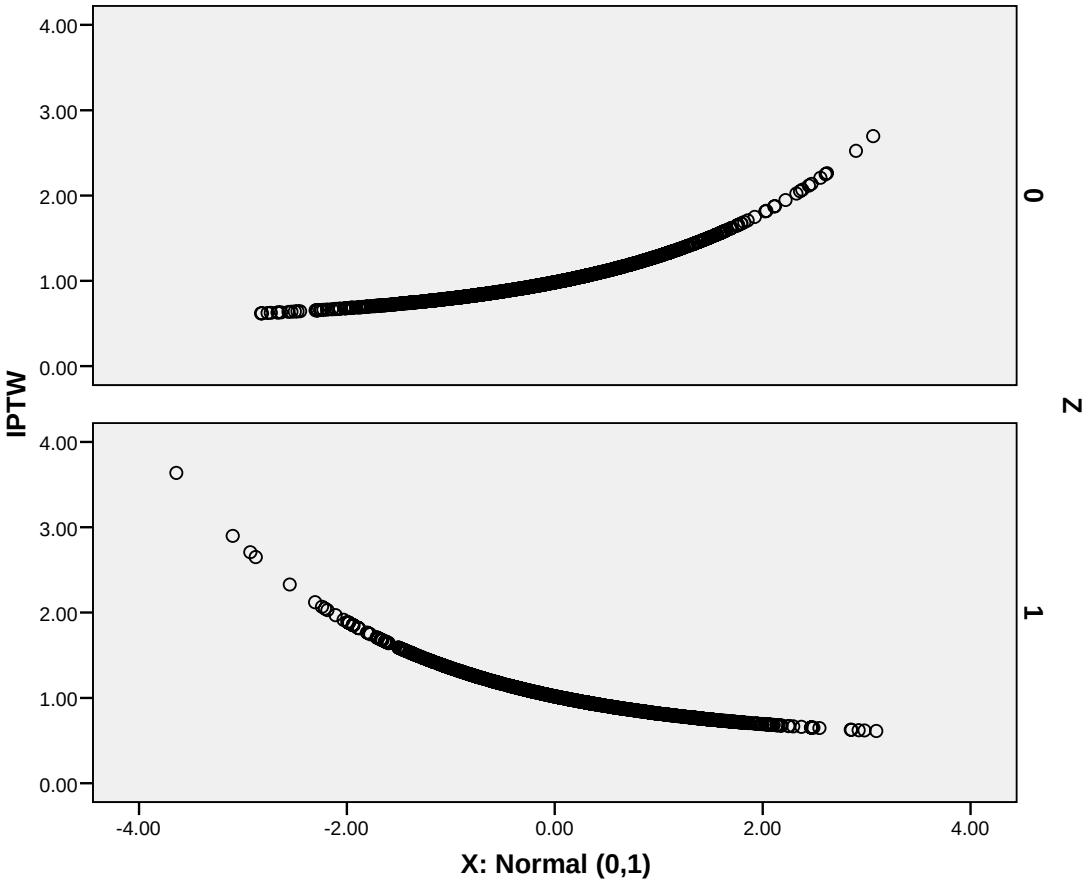

Figure 1.c

MMW Based on Misspecified Propensity Model $\hat{\eta} = \beta_0 + \beta_1 X$