Harvard University
Harvard University Biostatistics Working Paper Series

Year 2014 *Paper* 172

Adjustment for Mismeasured Exposure using Validation Data and Propensity Scores

Danielle Braun* Malka Gorfine† Corwin Zigler‡
Francesca Dominici** Giovanni Parmigiani††

*Dana-Farber Cancer Institute and Harvard School of Public Health, dbraun@mail.harvard.edu
†Faculty of Industrial Engineering and Management, Technion- Israel Institute of Technology
‡Harvard School of Public Health
**Harvard School of Public Health
††Dana-Farber Cancer Institute and Harvard School of Public Health

http://biostats.bepress.com/harvardbiostat/paper172

Adjustment for Mismeasured Exposure using Validation Data and Propensity Scores

Danielle Braun, Malka Gorfine, Corwin Zigler, Francesca Dominici, and Giovanni Parmigiani

Abstract

Propensity score methods are widely used to analyze observational studies in which patient characteristics might not be balanced by treatment group. These methods assume that exposure, or treatment assignment, is error-free, but in reality these variables can be subject to measurement error. This arises in the context of comparative effectiveness research, using observational administrative claims data in which accurate procedural codes are not always available. When using propensity score based methods, this error affects both the exposure variable directly, as well as the propensity score. We propose a two step maximum likelihood approach using validation data to adjust for the measurement error. First, we use a likelihood approach to estimate an adjusted propensity score. Using the adjusted propensity score, we then use a likelihood approach on the outcome model to adjust for measurement error in the exposure variable directly. In addition, we show the bias introduced when using error-prone treatment in the inverse probability weighting (IPW) estimator and propose an approach to eliminate this bias. Simulations show our proposed approaches reduce the bias and mean squared error (MSE) of the treatment effect estimator compared to using the error-prone treatment assignment.

Adjustment for Mismeasured Exposure using Validation Data and Propensity Scores

Danielle Braun[1,2,*], Malka Gorfine[3], Corwin Zigler[2], Francesca Dominici[2], and Giovanni Parmigiani[1,2]

[1]Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute

[2]Department of Biostatistics, Harvard School of Public Health, Boston, MA, USA.

[3]Faculty of Industrial Engineering and Management, Technion- Israel Institute of Technology, Haifa, Israel.

Running title: Adjustment for Mismeasured Exposure using Validation Data and Propensity Scores.

Contact person:

Danielle Braun

Harvard School of Public Health

677 Huntington Avenue, SPH2, 4th Floor

Boston, MA 02115

United States

(617)-582-7228

dbraun@hsph.harvard.edu

Abstract

Propensity score methods are widely used to analyze observational studies in which patient characteristics might not be balanced by treatment group. These methods assume that exposure, or treatment assignment, is error-free, but in reality these variables can be subject to measurement error. This arises in the context of comparative effectiveness research, using observational administrative claims data in which accurate procedural codes are not always available. When using propensity score based methods, this error affects both the exposure variable directly, as well as the propensity score. We propose a two step maximum likelihood approach using validation data to adjust for the measurement error. First, we use a likelihood approach to estimate an adjusted propensity score. Using the adjusted propensity score, we then use a likelihood approach on the outcome model to adjust for measurement error in the exposure variable directly. In addition, we show the bias introduced when using error-prone treatment in the inverse probability weighting (IPW) estimator and propose an approach to eliminate this bias. Simulations show our proposed approaches reduce the bias and mean squared error (MSE) of the treatment effect estimator compared to using the error-prone treatment assignment.

Keywords

Measurement Error; Comparative Effectiveness Research; Propensity Score; Causal Inference; exposure; Likelihood Adjustment; Validation Data

1 Introduction

There is a lot of interest in estimating causal treatment effects. Ideally, randomized control studies would be used to study treatment effects, but not always these are feasible due to ethical reasons, cost, time constraints, and compliance (among other reasons). Observational studies are more widely available, however, involve some limitations. Since subjects are not randomized by treatment, their characteristics might not be balanced by treatment group. In order to overcome this limitation, propensity score based methods have been proposed (Rosenbaum and Rubin, 1983).

The propensity score is defined as the probability that an individual has been assigned to treatment given their covariates. Various propensity score methods have been introduced. Rosenbaum and Rubin (1984) introduce a method that stratifies individuals based on their propensity score, and take the average of the treatment effects across strata. Propensity scores can also be used to weigh individuals observations (Rosenbaum, 1987). Matching individuals by their propensity scores attempts to create treated and control groups that have similar covariate values. Propensity score have also been used as covariates in the outcome model.

We focus on the setting in which covariates are not balanced by treatment assignment, and propensity score methods are used to overcome this limitation. Although one could run a regression analysis, including all confounders in the outcome model to overcome this limitation, propensity scores are advantageous in two main settings. The first, if there are many confounders, and fewer than eight events are observed per confounder (Cepeda et al., 2003), propensity scores allow for a way to reduce the dimensionality, and preform better than standard regression. The second, in the case of model misspecification, standard regression will lead to bias. It has been shown that treatment effect estimates are more sensitive to outcome model misspecification compared to an incorrect propensity score model (Drake, 1993).

These propensity score methods assume that treatment assignment is measured without error, but in reality treatment assignment in observational studies could be measured with error. Such is the case in our motivating example of a comparative effectiveness study of surgical treatments among Medicare beneficiaries diagnosed with glioblastoma (brain cancer). We came across this

in comparative effectiveness research, where the exact procedural codes are not available and we have error prone codes. We are able to obtain Medicare Part A Hospital claims data, which is a large data set (41,971 individuals) containing information on our outcome of interest (mortality), the treatment assignment (the determination of whether surgery was preformed), as well as many confounders. However, treatment assignment in this data set, is based on ICD9 billing codes which are not always an accurate measure of the treatment. For a subset of individuals (5,463 individuals), we are able to obtain data from SEER-Medicare, a cancer surveillance database with detailed clinical information. For these individuals, we have more accurate treatment assignment information. The SEER-Medicare is our internal validation study.

Treatment assignment in this context, can be thought of as the exposure variable. Measurement error in exposures has been extensively studied in the measurement error literature. Various techniques to adjust for measurement error in this setting have been developed including likelihood based approaches, regression calibration, Bayesian approaches, among others (Carroll, 2006). A literature review conducted by Jurek et al. (2006), shows that exposure measurement error is often ignored in studies. Causal inferences about the effect of an exposure may be biased by errors in the exposure (Hernán and Cole, 2009). Thus, there is a clear need to adjust for measurement error in this setting.

Standard techniques adjusting for measurement error in exposures cannot be applied directly to this setting. Misclassification of treatment assignment will lead to both error in the exposure variable directly, as well as error in the propensity score estimates. Previous literature using propensity score methods has largely focused on measurement error in confounders and missing confounders. Various approaches have been proposed in this setting. Stürmer et al. (2005), propose a regression calibration approach using validation data to estimate an adjusted propensity score in the case of unmeasured confounders. McCandless et al. (2012) propose a flexible Bayesian procedure to adjust for missing confounders using external validation data. McCaffrey et al. (2013) provide a consistent inverse probability weighting estimator in the case where confounders are measured with error. Babanezhad et al. (2010) consider a similar setting of mismeasured exposures. Using estimating equations they derive the asymptotic bias for the IPW and doubly robust estimators. Although,

we also derive the bias for the IPW estimator, we do not approach the problem using estimating equations. In addition, they do not consider the likelihood based adjustment approaches we propose in this paper. We are not aware of any additional literature addressing measurement error in treatment assignment in this context.

We propose a two step likelihood approach to adjust for measurement error. First, we adjust for measurement error in the propensity score using validation data. Next, we use the adjusted propensity score and adjust for the measurement error in the treatment effect using external validation data. We compare our proposed approach using four different propensity score based methods; stratification, inverse probability weighing of the likelihood, matching, and covariate adjustment. In addition, we show the bias caused by the misclassification of treatment to the inverse probability weighting (IPW) estimator directly and propose an approach to eliminate this bias.

In section 2, we introduce general notation and formulate our model. Afterwards, in section 3 we describe our proposed likelihood adjustment using various propensity score techniques. In section 4, we propose a way to eliminate the bias in the IPW estimator. We perform simulations in section 5, and summarize the main results in section 6.

2 General Notation and Model Formulation

2.1 Notations

Let Y denote the true outcome (ex: binary disease status, or a continuous measure, etc), let X denote a true binary treatment (ex: SEER claims data medication use), let X^* denote the error-prone treatment (ex: Part A medicare medication use), let $\mathbf{C}$ denote a vector of confounders measured without error (ex: age, etc).

Let $i = 1, ..., N_m$ index individuals in the main study (Part A medicare enrollees). For these individuals $Y, X^*, \mathbf{C}$ is available. In addition, suppose you have a validation study with $j = 1, ..., N_v$ individuals (SEER-Medicare enrollees). For these individuals $X, X^*, \mathbf{C}, Y$ is available. Our work focuses on the scenario of an internal validation study with known Y, but can easily be extended to other scenarios.

2.2 Model Formulation

2.2.1 Outcome Model

In general, we assume the outcome model is a generalized linear model. The true, error-free, outcome model can be written as $E(Y|X,C) = g^{-1}(\beta_{0x} + \beta_{1x}X + \beta_{2x}\mathbf{C})$, where g is known. The error-prone outcome model can be written as $E(Y|X^*,C) = g^{-1}(\beta_{0x^*} + \beta_{1x^*}X^* + \beta_{2x^*}\mathbf{C})$. In the *logit* case, which is the outcome model chosen for simulations in this paper, the true outcome model can be written as: $P(Y=1|X,\mathbf{C}) = \frac{1}{1+e^{-(\beta_{0x}+\beta_{1x}X+\beta_{2x}\mathbf{C})}}$ and the error-prone outcome model can be written as: $P(Y=1|X^*,\mathbf{C}) = \frac{1}{1+e^{-(\beta_{0x^*}+\beta_{1x^*}X^*+\beta_{2x^*}\mathbf{C})}}$.

2.2.2 Propensity Score Model

The true propensity score model can be written as: $PS_{true} = logit(P(X=1|\tilde{\mathbf{C}})) = \gamma_{\mathbf{x}}\tilde{\mathbf{C}}$, , where $\tilde{\mathbf{C}} = (1,\mathbf{C})$. The error-prone propensity score model can be written as: $PS_{ep} = logit(P(X^* = 1|\tilde{\mathbf{C}})) = \gamma_{\mathbf{x}^*}\tilde{\mathbf{C}}$.

2.2.3 Measurement Error Model

The measurement error model can be written as $E(X^*|X,C) = h^{-1}(\eta_0 + \eta_1 X + \eta_2\mathbf{C})$, where h is known.

3 Likelihood Adjustment Approach

Our interest is in estimating the true treatment effect on the outcome Y; $\frac{1}{N}\sum_{i=1}^{N}(y_{1i} - y_{0i})$, where y_{1i} is the value of the outcome for an individual i if he is treated and y_{0i} is the outcome for an individual i if he is not treated. We consider using different propensity score methods; stratification, weighted likelihood, matching and covariate adjustment. For each method, we propose a two step likelihood correction approach to correct for measurement error. First, use a likelihood approach to correct for measurement error in the propensity score model and estimate an adjusted propensity score. Next, use a likelihood approach on the outcome model to adjust for measurement error in the misclassified treatment variable directly using the adjusted propensity score.

For some propensity score methods, such as stratification by propensity score; the two step approach addresses the two possible sources of error (error in the propensity score, and error in the treatment assignment used as a covariate). For other methods, such as weighted likelihood using propensity scores or matching by propensity score, the error in the treatment assignment introduces an additional classification problem which we will illustrate when discussing the specific method.

3.1 Correcting for Measurement Error in Propensity Score

The likelihood of the propensity score model can be written as the product of the likelihood in the main study and in the validation study. In the main study, only X^* is observed. The likelihood in the main study can be rewritten using the law of total probability by summing over all possible values of the true X and multiplying by the measurement error $P(X^*|X, \mathbf{C})$. The overall likelihood (Equation (1)) is then maximized to obtain maximum likelihood estimates for γ.

$$\begin{aligned} L(\gamma) &= \prod_i^{N_m} P(x_i^*|\mathbf{c}_i, \gamma) \prod_j^{N_v} P(x_j|\mathbf{c}_j, \gamma) = \prod_i^{N_m} \sum_x P(x_i^*|x, \mathbf{c}_i, \gamma) P(x|\mathbf{c}_i, \gamma) \prod_j^{N_v} P(x_j|\mathbf{c}_j, \gamma) \\ &= \prod_i^{N_m} \sum_x P(x_i^*|x, \mathbf{c}_i) P(x|\mathbf{c}_i, \gamma) \prod_j^{N_v} P(x_j|\mathbf{c}_j, \gamma) \end{aligned} \quad (1)$$

The second equality follows from the assumption that $P(x_i^*|x, \mathbf{c}_i, \gamma) = P(x_i^*, \mathbf{c}_i|x)$, this is reasonable since we condition on both x and c. $P(X^*|X, \mathbf{C})$ is estimated using the validation data with N_v individuals. Equation (1) is maximized to obtain maximum likelihood estimates $\hat{\gamma}$. Using $\hat{\gamma}$ we can obtain an adjusted propensity score, $\widehat{PS_{adj}} = \hat{\gamma}\mathbf{C}$.

Assuming the propensity score follows a *logit* model, the likelihood can be written as:

$$\begin{aligned} L(\gamma) &= \prod_i^{N_m} [P(x_i^* = 1|x_i = 1, \mathbf{c}_i) \frac{1}{1 + \exp(-(\gamma \mathbf{c}_i))} + P(x_i^* = 1|x_i = 0, \mathbf{c}_i)(1 - \frac{1}{1 + \exp(-(\gamma \mathbf{c}_i))})]^{x_i^*} \\ &\quad * [P(x_i^* = 0|x_i = 1, \mathbf{c}_i) \frac{1}{1 + \exp(-(\gamma \mathbf{c}_i))} + P(x_i^* = 0|x_i = 0, \mathbf{c}_i)(1 - \frac{1}{1 + \exp(-(\gamma \mathbf{c}_i))})]^{(1-x_i^*)} \end{aligned}$$

$$* \prod_{j}^{N_v} [\frac{1}{1+\exp(-(\gamma \mathbf{c}_j))}]^{x_j} [1 - \frac{1}{1+\exp(-(\gamma \mathbf{c}_j))}]^{(1-x_j)} \quad (2)$$

3.2 Correcting for Measurement Error in Outcome Model

The propensity score model has now been corrected, but the outcome model still contains the misclassified treatment assignment. We use a likelihood approach on the outcome model to correct for this misclassification directly. The likelihood is different for each of the propensity score methods, therefore we discuss each method individually.

3.2.1 Stratification

In this propensity score method, individuals are stratified by their propensity scores into K groups with N_K individuals in each group. Treatment effect estimates are estimated for each of the K groups and averaged to obtain the overall treatment effect estimate. Ignoring the measurement error, we would stratify individuals by PS_{ep} and estimate treatment effects in each of the K strata using X^* in the outcome model. The proposed adjustment includes two steps; first stratifying individuals into K groups by PS_{adj} instead of PS_{ep}. Next, we rewrite the likelihood for the outcome model in each of the K strata using the law of total probability by summing over all possible values of the true X and weighing by the measurement error $P(X|X^*, \mathbf{C})$ which is estimated in the validation study (Equation (3)).We maximize $\beta_{\mathbf{k}}$ in each strata, to obtain treatment effect estimates for each strata. The overall treatment effect is the average of treatment effects across the strata.

In the case of an internal validation study, or an external validation study for which Y is known, the likelihood can be written as follows.

$$\begin{aligned} L^*(\beta_k) &= \prod_{i \in N_k} P(y_i|x_i^*, \mathbf{c}_i, \beta_k) \prod_{j \in N_{vk}} P(y_j|x_j, \mathbf{c}_j, \beta_k) \\ L(\beta_k) &= \prod_{i \in N_k} \sum_x P(y_i|x, x_i^*, \mathbf{c}_i, \beta_k) P(x|x_i^*, \mathbf{c}_i, \beta_k) \prod_{j \in N_{vk}} P(y_j|x_j, \mathbf{c}_j, \beta_k) \\ &= \prod_{i \in N_k} \sum_x P(y_i|x, \mathbf{c}_i, \beta_k) P(x|x_i^*, \mathbf{c}_i) \prod_{j \in N_{vk}} P(y_j|x_j, \mathbf{c}_j, \beta_k) \end{aligned} \quad (3)$$

The last equality follows from two assumptions, the first that $P(x_i|x_i^*, \mathbf{c}_i, \beta_k) = P(x_i|x_i^*, \mathbf{c}_i)$, which is reasonable since we condition on both X and $\mathbf{C}$. The second, that X is a surrogate for X^*, in other words that $P(y_i|x, x_i^*, \mathbf{c}_i, \beta_k) = P(y_i|x, \mathbf{c}_i, \beta_k)$. This is a very common assumption in the measurement error field.

$L(\beta_k)$ is maximized to obtain maximum likelihood estimates $\hat{\beta}_k$. We estimate the treatment effects within each strata, and report an average treatment effect estimate.

The likelihood in the case of a *logit* outcome model is shown in the Appendix Equation (A.1).

3.2.2 Weighted Likelihood

A weighted likelihood approach involves weighing each treated individual by the inverse of their propensity score and each untreated individual by the inverse of one minus their propensity score. β estimates are then obtained by maximizing the weighted likelihood. Ignoring measurement error, each individual would be weighted according to their error-prone treatment X^* using PS_{ep}. In other words, the decision of whether to use inverse PS_{ep} or inverse $1 - PS_{ep}$ as weights would be based on the error-prone treatment X^*. The proposed adjustment in this setting includes two steps; first using PS_{adj} as weights in the likelihood. Next, the likelihood is rewritten with the adjusted weights using the law of total probability by summing over all possible values of the true X and weighing by the measurement error $P(X|X^*, \mathbf{C})$ (Equation (4)). We maximize $L(\beta)$ to obtain maximum likelihood estimates $\hat{\beta}$ and estimate the treatment effects. We note as a possible limitation, that the proposed adjustment does not classify the use of inverse PS_{adj} or inverse $1 - PS_{adj}$ based on the true value X, but rather based on the error-prone treatment X^*.

In the case of an internal validation study, or an external validation study for which Y is known, the likelihood can be written as follows.

$$
\begin{aligned}
L^*(\beta) &= \prod_i^{N_m} [P(y_i|x_i^*, \mathbf{c}_i, \beta)]^{\frac{1}{P(x_i^*|\mathbf{c}_i, \gamma_{x^*})}} \prod_j^{N_v} [P(y_j|x_j, \mathbf{c}_j, \beta)]^{\frac{1}{P(x_j|\mathbf{c}_j, \gamma_x)}} \\
L(\beta) &= \prod_i^{N_m} [\sum_x P(y_i|x, x_i^*, \mathbf{c}_i, \beta) P(x|x_i^*, \mathbf{c}_i, \beta)]^{\frac{1}{P(x_i^*|\mathbf{c}_i, \gamma_{adj})}} \prod_j^{N_v} [P(y_j|x_j, \mathbf{c}_j, \beta)]^{\frac{1}{P(x_j|\mathbf{c}_j, \gamma_x)}}
\end{aligned}
$$

$$= \prod_{i}^{N_m}[\sum_{x} P(y_i|x,\mathbf{c}_i,\beta)P(x|x_i^*,\mathbf{c}_i)]^{\frac{1}{P(x_i^*|\mathbf{c}_i,\gamma_{adj})}} \prod_{j}^{N_v}[P(y_j|x_j,\mathbf{c}_j,\beta)]^{\frac{1}{P(x_j|\mathbf{c}_j,\gamma_x)}} \quad (4)$$

The likelihood in the case of a *logit* outcome model is shown in the Appendix Equation (A.2).

3.2.3 Matching

Various approaches for matching are used in the literature (Harder et al., 2010; Stuart, 2010).These methods include, one-to-one matching which matches each treated individual to an untreated individual based on their propensity score (un-matched individuals are not analyzed). Nearest neighbor matching which matches each treated individual to r untreated individuals. Variable ratio matching which matches different treated individuals to varying numbers of untreated individuals. Matching with replacement which matches untreated individuals to treated individuals more than once, thus each untreated individual can be matched multiple times. Exact matching which matches a treated individual to all possible un-treated individuals who have exactly the same values on all covariates. Full matching which forms matched sets, where each set has at least one treated and untreated individual in it.

After matching, regular outcome analysis on the data should be preformed (Ho et al., 2007; Stuart, 2010). There is some debate in the literature about whether or not the analysis needs to account for the matched pairs. Schafer and Kang (2008); Stuart and Green (2008) mention two reasons why this is not necessary. The first, that it is enough to condition on the variables used for the matching. The second, that one should pool all matched treated and untreated individuals and run analysis on the entire group, since matching based on propensity score, does not necessarily mean that each individual pair will be similar across all covariates.

In the case of matching with replacement or variable ratio matching, weights need to be incorporated into the outcome analysis (Dehejia and Wahba, 1999; Hill et al., 2004; Stuart, 2010). In the case of matching with replacement, once the matched sets are formed, each treated individual receives a weight of one, and each untreated individual receives a weight proportional to the number of times they were matched (Stuart, 2010). In the case of variable ratio matching, once

the matched sets are formed, each treated individual receives a weight of one, and each untreated individual receives a weight proportional to the number of untreated individuals in the matched set (Stuart, 2010).

We propose a method to adjust for measurement error in these types of matching (matching with replacement or variable ratio matching). Ignoring measurement error, each individual would be matched according to their treatment X^* based on their propensity score PS_{ep}. We will refer to these weights obtained from this error-prone matching as $\widehat{w_{ep}}$, and weights obtained from matching based on the true treatment X using the true propensity score PS_{true} as $\widehat{w_{true}}$.

The proposed adjustment in this setting includes two steps; first using PS_{adj} to match individuals based on their treatment X^*. We will refer to these weights obtained from this error-prone matching as $\widehat{w_{adj}}$. Next, the likelihood is rewritten with the adjusted weights using the law of total probability by summing over all possible values of the true X and weighing by the measurement error $P(X|X^*, \mathbf{C})$ (Equation (5)).We maximize $L(\beta)$ to obtain maximum likelihood estimates $\hat{\beta}$ and estimate the treatment effects. We note as a possible limitation that the proposed adjustment does not match individuals based on their true X but rather matches them based on X^*.

In the case of an internal validation study, or an external validation study for which Y is known, the likelihood can be written as follows.

$$\begin{aligned} L^*(\beta) &= \prod_i^{N_m} P(y_i|x_i^*, \mathbf{c}_i, \beta)^{\widehat{w_{epi}}} \prod_j^{N_v} P(y_j|x_j, \mathbf{c}_j, \beta)^{\widehat{w_{truei}}} \\ L(\beta) &= \prod_i^{N_m} \sum_x [P(y_i|x_i, x_i^*, \mathbf{c}_i, \beta) P(x|x_i^*, \mathbf{c}_i, \beta)]^{\widehat{w_{adji}}} \prod_j^{N_v} P(y_j|x_j, \mathbf{c}_j, \beta)^{\widehat{w_{truei}}} \\ &= \prod_i^{N_m} \sum_x [P(y_i|x, \mathbf{c}_i, \beta) P(x|x_i^*, \mathbf{c}_i)]^{\widehat{w_{adji}}} \prod_j^{N_v} P(y_j|x_j, \mathbf{c}_j, \beta)^{\widehat{w_{truei}}} \end{aligned} \quad (5)$$

The likelihood in the case of a *logit* outcome model is shown in the Appendix Equation (A.3).

3.2.4 Propensity Score Covariate Adjustment

Covariate adjustment involves including the propensity score as a covariate in the outcome model. Ignoring measurement error, PS_{ep} and X^* would be used as covariates in the outcome model. The proposed adjustment includes two steps, first using PS_{adj} as a covariate in the outcome model. Next, the likelihood is rewritten using the adjusted weights using the law of total probability by summing over all possible values of the true X and weighing by the measurement error $P(X|X^*, \mathbf{C})$ (Equation (6)).We maximize $L(\beta)$ to obtain maximum likelihood estimates $\hat{\beta}$.

In the case of an internal validation study, or an external validation study for which Y is known, the likelihood can be written as follows.

$$\begin{aligned} L^*(\beta) &= \prod_i^{N_m} P(y_i|x_i^*, \mathbf{c}_i, PS_{ep}, \beta) \prod_j^{N_v} P(y_j|x_j, \mathbf{c}_j, PS_{true}, \beta) \\ L(\beta) &= \prod_i^{N_m} \sum_x P(y_i|x, \mathbf{c}_i, \widehat{PS}_{adj}, x_i^*, \beta) P(x|x_i^*, \widehat{PS}_{adj}, \mathbf{c}_i, \beta) \prod_j^{N_v} P(y_j|x_j, \mathbf{c}_j, PS_{true}, \beta) \\ &= \prod_i^{N_m} \sum_x P(y_i|x, \mathbf{c}_i, \widehat{PS}_{adj}, \beta) P(x|x_i^*, \mathbf{c}_i) \prod_j^{N_v} P(y_j|x_j, \mathbf{c}_j, PS_{true}, \beta) \end{aligned} \quad (6)$$

The likelihood in the case of a *logit* outcome model is shown in the Appendix Equation (A.4).

4 IPW Estimator Adjustment

In addition to the likelihood approaches discussed in the previous section, we consider adjusting the IPW estimator directly. The IPW estimator gives an estimate of the treatment effect, by weighing treated individuals by their inverse propensity score and untreated individuals by the inverse of one minus their propensity score (Rosenbaum, 2005). Ignoring the measurement error the IPW estimator is shown in Equation (7).

$$\hat{\Delta}_{IPW_{ep}} = N_m^{-1} \sum_{i=1}^{N_m} \frac{X_i^* Y_i}{\widehat{PS}_{epi}} - N_m^{-1} \sum_{i=1}^{N_m} \frac{(1 - X_i^*) Y_i}{1 - \widehat{PS}_{epi}} \quad (7)$$

We first show that this estimator is biased, and then propose a way to adjust for this bias. We

begin by writing Y as $Y = XY_1 + (1-X)Y_0$, $X^*Y = X^*XY_1 + X^*(1-X)Y_0$ and $(1-X^*)Y = (1-X^*)XY_1 + (1-X^*)(1-X)Y_0$, where Y_0 is the outcome an individual would have had if he/she were untreated, and Y_1 is the outcome an individual would have had if he/she were treated. We look at each of the two components of the IPW estimator separately. The first component can be written as:

$$\begin{aligned}
\mathrm{E}(\frac{X^*Y}{PS_{ep}}) &= \mathrm{EE}(\frac{X^*Y}{PS_{ep}}|Y_1, Y_0, C) = \mathrm{EE}(\frac{X^*XY_1 + X^*(1-X)Y_0}{PS_{ep}}|Y_1, Y_0, C) \\
&= \mathrm{EE}(\frac{X^*XY_1}{PS_{ep}}|Y_1, C) + \mathrm{EE}(\frac{X^*(1-X)Y_0}{PS_{ep}}|Y_0, C) \\
&= E(\frac{Y_1}{PS_{ep}}P(X^*=1, X=1|Y_1, C)) + E(\frac{Y_0}{PS_{ep}}P(X^*=1, X=0|Y_0, C)) \\
&= E(\frac{Y_1}{PS_{ep}}P(X=1|X^*=1, C)P(X^*=1|C)) + E(\frac{Y_0}{PS_{ep}}P(X=0|X^*=1, C)P(X^*=1|C)) \\
&= E(Y_1P(X=1|X^*=1, C)) + E(Y_0P(X=0|X^*=1, C)) \qquad (8)
\end{aligned}$$

Similarly, the second component of the sum can be rewritten as:

$$\begin{aligned}
\mathrm{E}(\frac{(1-X^*)Y}{PS_{ep}}) &= \mathrm{EE}(\frac{(1-X^*)Y}{1-PS_{ep}}|Y_1, Y_0, C) = \mathrm{EE}(\frac{(1-X^*)XY_1 + (1-X^*)(1-X)Y_0}{1-PS_{ep}}|Y_1, Y_0, C) \\
&= \mathrm{EE}(\frac{(1-X^*)XY_1}{1-PS_{ep}}|Y_1, C) + \mathrm{EE}(\frac{(1-X^*)(1-X)Y_0}{1-PS_{ep}}|Y_0, C) \\
&= E(\frac{Y_1}{1-PS_{ep}}P(X^*=0, X=1|Y_1, C)) + E(\frac{Y_0}{1-PS_{ep}}P(X^*=0, X=0|Y_0, C)) \\
&= E(\frac{Y_1}{1-PS_{ep}}P(X=1|X^*=0, C)P(X^*=0|C)) \\
&+ E(\frac{Y_0}{1-PS_{ep}}P(X=0|X^*=0, C)P(X^*=0|C)) \\
&= E(Y_1P(X=1|X^*=0, C)) + E(Y_0P(X=0|X^*=0, C)) \qquad (9)
\end{aligned}$$

Thus, overall, the expectation of the IPW estimator based on error-prone treatment assignment

is:

$$E(Y_1 P(X=1|X^*=1,C)) + E(Y_0 P(X=0|X^*=1,C))$$
$$-E(Y_1 P(X=1|X^*=0,C)) - E(Y_0 P(X=0|X^*=0,C)) \quad (10)$$

instead of $E(Y_1) - E(Y_0)$. Similarly, it can also be shown that simply replacing $\widehat{PS}_{ep}$ by an estimate of the true propensity score, $\widehat{PS}_{adj}$ does not eliminate the bias, and the expected value of such estimator would be:

$$E(Y_1 P(X^*=1|X=1,C)) + E(Y_0 P(X^*=1|X=0,C)\frac{1-PS_{adj}}{PS_{adj}})$$
$$-E(Y_0 P(X^*=0|X=0,C)) - E(Y_1 P(X^*=0|X=1,C)\frac{PS_{adj}}{1-PS_{adj}})$$

One proposed approach to eliminate the bias based caused by using the error-prone treatment, X^*, and error-prone propensity score PS_{ep}, shown in Equation (10), is to first divide the first component of the estimator in Equation (7) by $P(X=1|X^*=1,C)$ and the second component of the estimator by $P(X=0|X^*=0,C)$. The expected value of this new estimator would be:

$$E(\frac{Y_1 P(X=1|X^*=1,C)}{P(X=1|X^*=1,C)}) + E(\frac{Y_0 P(X=0|X^*=1,C)}{P(X=1|X^*=1,C)})$$
$$-E(\frac{Y_1 P(X=1|X^*=0,C)}{P(X=0|X^*=0,C)}) - E(\frac{Y_0 P(X=0|X^*=0,C)}{P(X=0|X^*=0,C)})$$
$$= E(Y_1) + E(\frac{Y_0 P(X=0|X^*=1,C)}{P(X=1|X^*=1,C)}) - E(\frac{Y_1 P(X=1|X^*=0,C)}{P(X=0|X^*=0,C)}) - E(Y_0)$$

The expected value above is biased. Two components $E(\frac{Y_0 P(X=0|X^*=1,C)}{P(X=1|X^*=1,C)})$ and $E(\frac{Y_1 P(X=1|X^*=0,C)}{P(X=0|X^*=0,C)})$ contribute to this bias in the expectation. In the case of an internal validation study, or an external validation study for which Y is known, we propose estimating these bias components in the validation data, and subtracting them from the overall estimator. Thus, our proposed adjusted estimator

would be:

$$\begin{aligned}\hat{\Delta}_{IPW_{adj}} &= N_m^{-1}\sum_{i=1}^{N_m}\frac{X_i^* Y_i}{\widehat{PS}_{epi}\hat{P}(X=1|X^*=1,C_i)} - N_m^{-1}\sum_{i=1}^{N_m}\frac{(1-X_i^*)Y_i}{(1-\widehat{PS}_{epi})\hat{P}(X=0|X^*=0,C_i)} \\ &- N_v^{-1}\sum_{j=1}^{N_v}\frac{(1-X_j)Y_j}{(1-\widehat{PS}_{truej})}\frac{\hat{P}(X=0|X^*=1,C_i)}{\hat{P}(X=1|X^*=1,C_i)} \\ &+ N_v^{-1}\sum_{j=1}^{N_v}\frac{X_j Y_j}{\widehat{PS}_{truej}}\frac{\hat{P}(X=1|X^*=0,C_i)}{\hat{P}(X=0|X^*=0,C_i)} \qquad (11)\end{aligned}$$

This would give an estimator for $E(Y_1) - E(Y_0)$. For the first two sums in the estimator in Equation (11), we propose estimating the measurement error model based on the validation study, and calculating the sum in the main study. The bias terms are estimated entirely in the validation study.

5 Simulations

In our simulations, we study the effects of measurement error in treatment assignment on the treatment effect estimator using various propensity score methods. For each of the likelihood based methods, we evaluate the performance of our proposed likelihood adjustment by comparing treatment effect estimates based on true treatment assignment, error-prone treatment assignment, and the likelihood adjustment for the error-prone treatment assignment. For each of these categories, we compare the propensity score methods based on the true propensity score, error-prone propensity score, and the proposed adjusted propensity score (based on Equation (1)).

5.1 Simulation Characteristics

For each simulation scenario we simulated two data sets in a similar manner, one to be used as the main study and one to be used as validation data. We simulated the main study with $N_m = 3,000$ individuals and the validation study with $N_v = 1,500$ individuals. We consider confounders $\mathbf{C} = (1, C_1, ...C_6)$ that include both continuous (C_1, C_2, C_3) and binary covariates (C_4, C_5, C_6). X was generated as Bernoulli according to the true propensity score $logit(P(X = 1|\mathbf{C}) = \gamma\mathbf{C}$. X^*

was generated as Bernoulli according to the measurement error model $logit(P(X^* = 1|X, \mathbf{C_s}) = \eta_0+\eta_1 X+\eta_s \mathbf{C_s}$, where $\mathbf{C_s}$ represents a subset of the confounders that were used for the measurement error model. We considered a *logit* outcome model, Y was generated as Bernoulli according to the outcome model $logit(P(Y = 1|X, \mathbf{C}) = \beta_0 + \beta_1 X + \beta_2 C_1 + \beta_3 C_2 + \beta_4 C_3 + \beta_5 C_4 + \beta_6 C_5 + \beta_7 C_6$. $\beta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7)^T = (0, -2, -1, 1, 1, -1, 1, 1)^T$, so that $\beta_1 = -2$. We consider two settings for $\gamma = (\gamma_0, \gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5, \gamma_6)^T = (0, 0.3, -0.3, -0.3, 0.3, -0.3, 0.3)^T$ representing a moderate association of X and $\mathbf{C}$, and $(0, 0.6, -0.6, -0.6, 0.6, -0.6, 0.6)^T$ representing a strong association for X and $\mathbf{C}$. For the measurement error model, we consider two settings. The first, $\eta = (\eta_0, \eta_1, \eta_s)^T = (1, -2, \mathbf{0})$ representing a moderate association between X^* and X without any dependence on $\mathbf{C}$. The second, $\mathbf{C_s}$ containing two covariates C_1 and C_4, for which we consider $\eta = (\eta_0, \eta_1, \eta_2, \eta_3)^T = (0.5, -0.4, -0.4, -0.4)$. After the data was generated, we applied the analysis strategies described in previous sections. 100 repetitions were preformed for each simulation. 100 repetitions were preformed for each simulation.

5.2 Simulation Results

For each simulation setting we evaluate the different approaches by evaluating the bias and mean squared error (MSE) in the treatment effect estimator. Results for three simulation settings are shown in Figures 1, 2, 3. Figure 1 corresponds to a simulation setting with a moderate association between X and $\mathbf{C}$ and with measurement error independent of $\mathbf{C}$. Figure 2 corresponds to a simulation setting with a strong association between X and $\mathbf{C}$ and with measurement error independent of $\mathbf{C}$. Figure 3 corresponds to a simulation setting with a moderate association between X and $\mathbf{C}$ and with measurement error dependent on $\mathbf{C}$.

In each figure, the top panel shows results when C's are included as covariates in the outcome model, whereas the bottom panel shows results when C's are not included in the outcome model. Results are grouped by propensity score methods, we also compare results to multivariate regression without using any propensity score methods. From left to right, we plot results based on using no propensity score method, stratifying by propensity score, weighted likelihood using inverse propensity scores, using propensity score as a covariate, and matching based on propensity score. For the

stratification method, individuals were ordered by propensity score, and stratified into four strata. For the weighted likelihood method using inverse propensity scores, the inverse propensity scores and one minus the inverse propensity scores weights were stabilized as suggested by Robins et al. (2000). These weights were stabilized by multiplying the weights for the treated individuals by the expected value of being treated, and the untreated individuals by the expected value of being untreated (Robins et al., 2000). For the matching method, we obtain weights for each individual by implementing matching using the MatchIt R-package (Ho, 2009).

For each of these methods, the squares represent outcome models which included the true treatment, X, as a covariate. The circles represent outcome models which included the error-prone treatment, X^*, as a covariate. The triangles represent outcome models which included the error-prone treatment, X^*, as a covariate, but for which a likelihood adjustment was conducted. Blue represents using the true propensity score, red represents using the error-prone propensity score, and green represents using the proposed adjusted propensity score (based on Equation (1)).

When including C's in the outcome model (top panel), we see that across all methods and independent of the propensity score, outcome models which included the true treatment, X (squares in the plot), as a covariate, are unbiased and have a MSE close to 0. For example, for the stratification propensity score method in the first simulation setting, Figure 1, the bias and MSE are 0.00057 and 0.01324 using the true propensity scores. This is seen in all three figures. Across all methods and independent of propensity score, outcome models which included the error-prone treatment, X^* (circles in the plots) have a substantial amount of bias and MSE. For example, for the stratification propensity score method in the first simulation setting, Figure 1, the bias and MSE are 0.26676 and 0.26702 using the error-prone propensity scores. When we apply our likelihood adjustment to these outcome models which included the error-prone treatment, X^*, as a covariate, (triangles in the plots), the bias and MSE reduces substantially. For example, for the stratification propensity score method in the first simulation setting, Figure 1, the bias and MSE are 0.01146 and 0.01873 using the adjusted propensity scores. We note that in this setting where C's are included in the outcome model, there is no benefit of using propensity score methods (in blue, red, and green) compared to not using propensity score methods (in purple). For example, bias and MSE without

using propensity score methods, in the first simulation setting, Figure 1, are 0.00080 and 0.01317 based on X, 0.26666 and 0.26691 based on X^*, and 0.01079 and 0.01680 based on the likelihood adjustment. There is no model misspecification here, and we consider a model with a relatively small number of confounders, thus we expect standard multivariate regression to preform just as well as the propensity score methods.

In the bottom panel, we do not include C's as covariates in the outcome model. Propensity score methods are used as a way to reduce the dimensionality. Although in this specific simulation scenario there is no need to reduce the dimensionality, we evaluate the performance of our methods in this setting since propensity score methods are used for this purpose. In addition, propensity score methods are used in the case of model misspecification. Not including C's in the model is an extreme case of model misspecification. In this setting, the outcome model using X as a covariate without any propensity score methods has a substantial amount of bias and MSE (purple squares in the plots). Using X^* as a covariate without any propensity score methods shifts the bias in the opposite direction, in all three figures (purple circles), while the MSE increases in Figure 1, increases slightly in Figure 2, and decreases in Figure 3. Preforming the likelihood adjustment in this setting reduces the bias and MSE slightly in FIgures 1,2 while increasing the bias and MSE slightly in Figure 3. For example, in the first simulation setting, Figure 1, the bias and MSE are -0.32158 and 0.32195 based on X, 0.42043 and 0.42081 based on X^*, and -0.27235 and 0.27317 based on the likelihood adjustment. Thus, in this setting, without using any propensity score methods, preforming a likelihood adjustment does not eliminate the bias in the treatment effect estimators.

When stratifying by propensity scores, using X as a covariate preforms well in terms of bias and MSE. This is consistent across the three propensity scores, with better performance based on the true and adjusted propensity scores compared to the error-free propensity score seen in Figure 3. Using error-prone X^* increases the bias and MSE with all three propensity scores. While applying the likelihood adjustment reduces this bias and MSE, but only when using the true and adjusted propensity scores. For example, in the first simulation setting, Figure 1, the bias and MSE are -0.02437 and 0.02855 based on X using the true propensity score, 0.27680 and 0.27710 based on X^* using the error-prone propensity score, -0.24436 and 0.24503 based on the likelihood adjustment

using the error-prone propensity score, and -0.00849 and 0.01676 based on the likelihood adjustment using the adjusted propensity score.

For this method, we assess balance in our covariates by evaluating the standardized bias, which is defined as the difference in means in the covariate between the treatment and comparison group divided by the standard deviation. Ho et al. (2007) consider a covariate balance if the standardized bias is less than 0.25. We look at balance in covariates without using any propensity score methods, using the true propensity scores, using error-prone propensity scores, and using our adjusted propensity score (Table 1). We see that balance is improved using the adjusted propensity score compared to the error-prone propensity score, especially in the case where the error is dependent on $\mathbf{C}$. Thus, for stratification in these simulations we see that there is a benefit in a two step approach, first using an adjusted propensity score, and second preforming a likelihood adjustment on the outcome model.

The weighted likelihood approach is sensitive to propensity score selection. Using the true X as a covariate in the model, the weighted likelihood approach preforms best using the true propensity score, and worse with the error-prone propensity score, this is seen in all three figures. In all three figures, using the adjusted propensity score improves bias and MSE compared to the error-prone, and a substantial improvement is seen in Figure 3. For example, in the first simulation setting, Figure 1, the bias and MSE are 0.00044 and 0.01374 using the true propensity score, -0.42452 and 0.42487 using the error-prone propensity score, and -0.11787 and 0.12119 using the adjusted propensity score. Using the error-prone X^* as a covariate in the model, the weighted likelihood approach preforms best using the error-prone propensity scores, and worse with the true propensity scores, this is seen across all three figures. For example, in the first simulation setting, Figure 1, the bias and MSE are 0.81261 and 0.81292 using the true propensity score, 0.27284 and 0.27314 using the error-prone propensity score, and 0.68230 and 0.68275 using the adjusted propensity score.

The likelihood adjustment substantially improves the bias and MSE when using the error-prone propensity score. For example, in the first simulation setting, Figure 1, the bias and MSE are -0.34292 and 0.34366 using the true propensity score, -0.00286 and 0.01700 using the error-prone propensity score, and -0.23758 and 0.23836 using the adjusted propensity score. This is due to the

fact that the classification of using the inverse propensity score or the inverse of one minus the propensity score is based on X^*, therefore it is not surprising that the method preforms best using propensity score based on X^*. Thus, for the weighted likelihood approach in these simulations, we see that there is a benefit in a one step approach, using the error-prone propensity score, and preforming a likelihood adjustment on the outcome model.

Using the propensity score as a covariate in the model, including X as a covariate in the true model, bias and MSE are very low using all three propensity scores when the error is independent of the covariates (Figures 1, 2). Bias and MSE are better using the error-free and adjusted propensity scores when the error is dependent of the covariates (Figure 3). Including X^* as a covariate, increases the bias and MSE (with the exception of when the error is dependent on the covariates using the error-prone propensity score), and using the likelihood adjustment actually increases the bias and MSE (in all three figures). For example, in the first simulation setting, Figure 1, the bias and MSE are -0.00063 and 0.01378 based on X using the true propensity score, 0.26792 and 0.26818 based on X^* using the error-prone propensity score, -0.24436 and 0.24503 based on the likelihood adjustment using the error-prone propensity score, and-0.25158 and 0.25210 based on the likelihood adjustment using the adjusted propensity score. Thus, for this approach, in these simulations, no likelihood adjustment using the error-prone treatment X^* as a covariate in the model and using either of the propensity scores, preforms better than a likelihood adjustment on the outcome model.

For the matching, when including X as a covariate in the model, bias and MSE are very low. When the error is dependent on $\mathbf{C}$ (Figure 3), using the true propensity scores and adjusted propensity score preform much better than the error-prone propensity scores. For the other scenarios, we see similar performance across the three propensity scores (Figures 1, 2). Including X^* increases the bias and MSE substantially (with the exception of when the error is dependent on the covariates using the error-prone propensity score), and using the likelihood adjustment reduces the bias and MSE substantially across all three settings. For example, in the first simulation setting, Figure 1, the bias and MSE are 0.00980 and 0.01917 based on X using the true propensity score, 0.26951 and 0.26982 based on X^* using the error-prone propensity score, 0.00053 and 0.02011 based on the likelihood adjustment using the error-prone propensity score, and -0.00114 and 0.01993 based on

Table 1: Standardized bias; the difference in means in the covariate between the treatment and comparison group divided by the standard deviation, is shown for each covariate. Calculations are shown when stratifying by propensity scores, using various propensity scores, as well as without the stratification. Simulations settings correspond to Figures 1,2,3.

		C_1	C_2	C_3	C_4	C_5	C_6
Simulation Setting 1	No PS	0.23433	0.48912	0.75196	0.11354	0.12614	0.11939
	$\widehat{PS_{true}}$	0.03804	0.03990	0.04484	0.03626	0.03582	0.03803
	$\widehat{PS_{ep}}$	0.04682	0.04885	0.05275	0.04781	0.04625	0.04808
	$\widehat{PS_{adj}}$	0.04503	0.04996	0.04858	0.04517	0.04594	0.04944
Simulation Setting 2	No PS	0.33864	0.70049	1.13997	0.16356	0.17991	0.16622
	$\widehat{PS_{true}}$	0.05961	0.06415	0.07092	0.05835	0.05852	0.06003
	$\widehat{PS_{ep}}$	0.07591	0.07399	0.08303	0.06826	0.06945	0.07246
	$\widehat{PS_{adj}}$	0.07111	0.06894	0.07610	0.06637	0.06832	0.06703
Simulation Setting 3	No PS	0.23433	0.48912	0.75196	0.11354	0.12614	0.11939
	$\widehat{PS_{true}}$	0.03804	0.03990	0.04484	0.03626	0.03582	0.03803
	$\widehat{PS_{ep}}$	0.04582	0.19450	0.29517	0.04194	0.06043	0.05596
	$\widehat{PS_{adj}}$	0.04427	0.04773	0.04777	0.04630	0.04397	0.04714

the likelihood adjustment using the adjusted propensity score. For matching the likelihood adjustment preforms just as well independent of the propensity score method used. Thus, for matching in these simulations, we see that there is a benefit in a one step approach, preforming a likelihood adjustment on the outcome model. It does not seem to make a difference if one uses error-prone or adjusted propensity scores for this analysis.

Overall, we see that in these simulations, depending on the propensity score method, different approaches to handle the measurement error provide the least bias and MSE in the treatment effect estimator.

6 Discussion

In this paper we present various approaches to adjust for measurement error in treatment assignment in observational studies. Previous literature on measurement error in the setting of propensity scores has mostly focused on error in the confounders and not on error in the treatment assignment. Our simulations studies show that using error-prone treatment assignment introduces substantial bias in

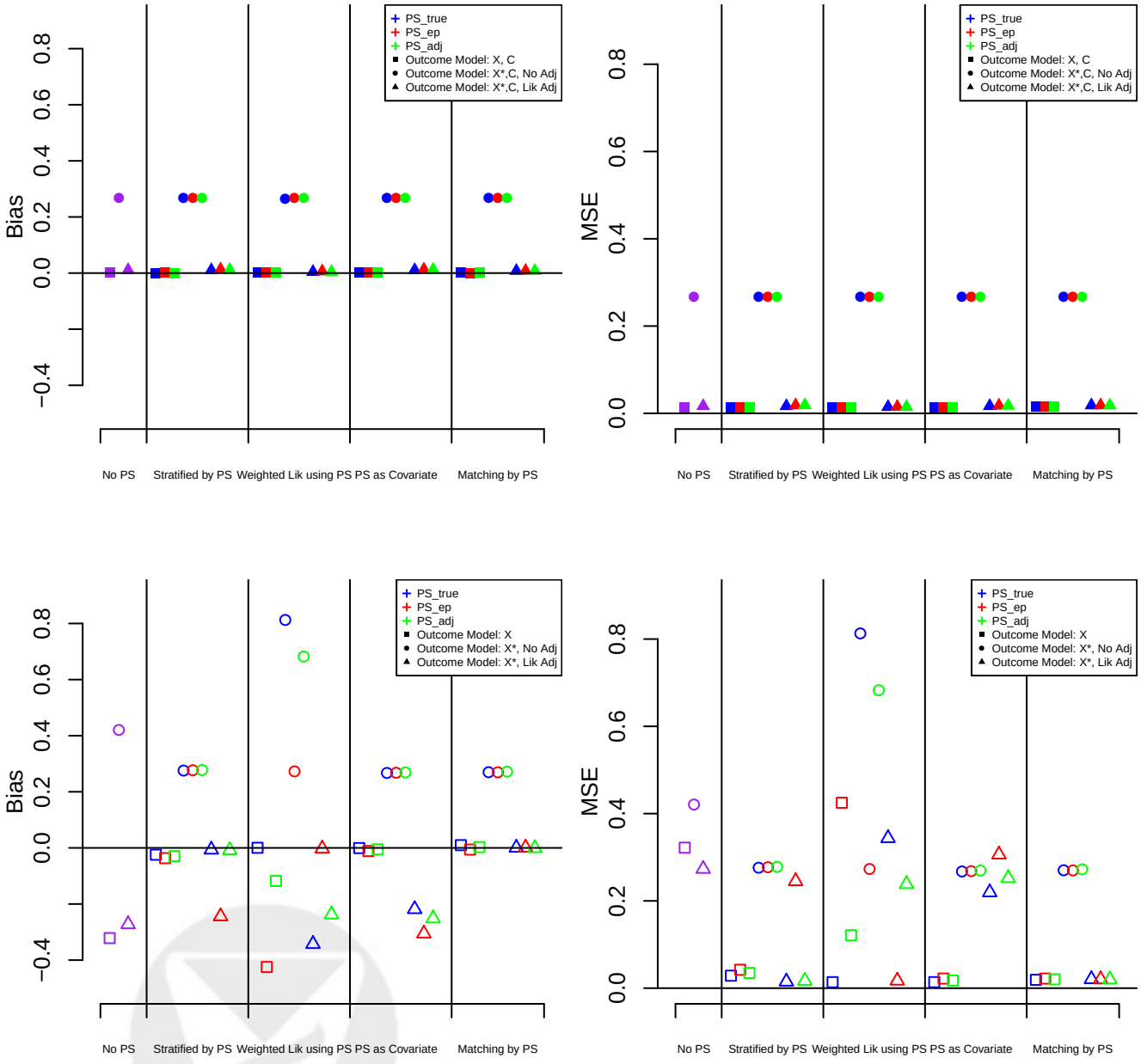

Figure 1: Simulation Setting 1: a moderate association between X and $\mathbf{C}$ and measurement error independent of $\mathbf{C}$.

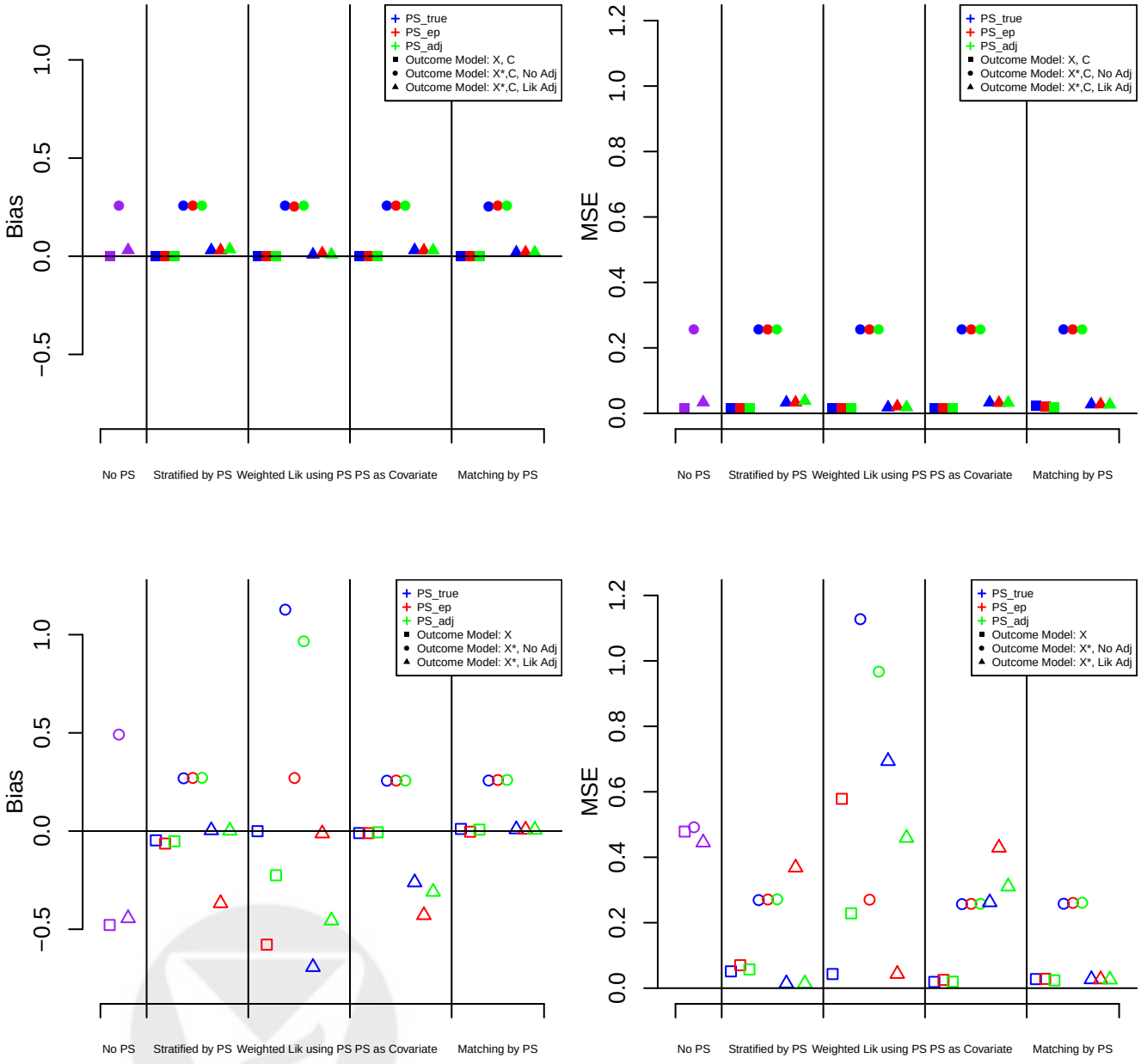

Figure 2: Simulation Setting 2: a strong association between X and $\mathbf{C}$ and measurement error independent of $\mathbf{C}$.

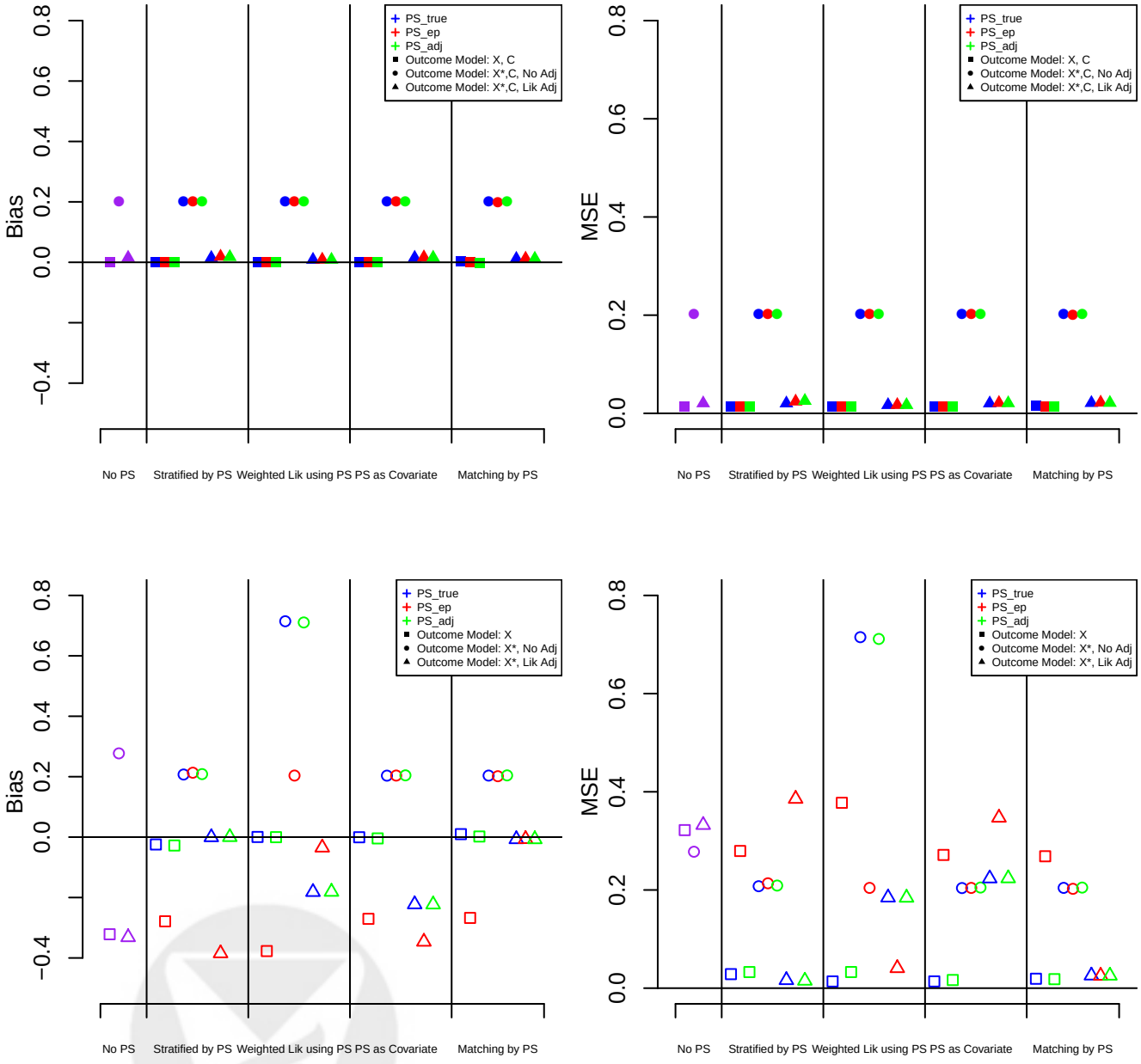

Figure 3: Simulation Setting 3: a moderate association between X and $\mathbf{C}$ and measurement error dependent on $\mathbf{C}$.

the treatment effect estimators. This bias is eliminated when preforming the likelihood adjustment on the outcome model; rewriting the likelihood using the law of total probability by summing over all possible values of the true treatment and weighing by the measurement error which is estimated in the validation study.

The benefit of using propensity score methods is apparent when confounders are not included in the outcome model. We realize that in these specific simulation scenarios, there is no reason not to include confounders in the outcome model. However, propensity score methods are often use when one is unable to include the confounders in the outcome model, thus it is relevant to examine this case. In addition, this case provides an example in which model misspecification has occurred. Conclusions, based on simulations, in the setting where confounders are not included in the model vary by propensity score method.

Overall, all four propensity score methods considered provide a benefit in terms of bias and MSE compared to not using any propensity score methods. For stratification, we see a clear benefit of using an adjusted propensity score and preforming a likelihood adjustment on the outcome model. For the weighted likelihood approach, we see a clear benefit of using an error-prone propensity score and preforming a likelihood adjustment on the outcome model. When using the propensity score as a covariate, we see that not preforming a likelihood adjustment preforms better than preforming a likelihood adjustment. For matching, we see a clear benefit in preforming a likelihood adjustment on the outcome model, but not on the choice of the propensity score.

It is important to note that in proposed weighted likelihood approach, weights are assigned based on the error-prone treatment assignment. For this reason, the weighted likelihood approach using error-prone propensity score preforms better than using the true propensity score. In the proposed matching approach, individuals are matched by their error-prone treatment assignment. Although this is the case, we see that this method preforms extremely well.

In this paper we consider a setting of an internal validation study. The likelihoods presented in Equations (3-6) can be easily modified to the setting of an external validation study with no outcome information. This paper covers a wide range of propensity score methods, leaving it to the reader to decide which propensity score method to use. This is a complex problem requiring

different approaches for the various settings. There are many ways in which the propensity score can be used to estimate treatment effects, and in this work we see that there is no single approach to measurement error adjustment. Depending on the propensity score method, a different approach to adjust for the measurement error should be applied.

Propensity score methods are widely used in the literature, and there is an increased use of them in recent years. Measurement error in exposure is often ignored. We provide various methods to address this issue. We cover a wide range of methods allowing the reader to choose which propensity score method to use based on their own preference.

References

Babanezhad, M., Vansteelandt, S., and Goetghebeur, E. (2010). Comparison of causal effect estimators under exposure misclassification. *Journal of Statistical Planning and Inference* **140,** 1306–1319.

Carroll, R. (2006). *Measurement error in nonlinear models: a modern perspective*, volume 105. CRC Press.

Cepeda, M. S., Boston, R., Farrar, J. T., and Strom, B. L. (2003). Comparison of logistic regression versus propensity score when the number of events is low and there are multiple confounders. *American Journal of Epidemiology* **158,** 280–287.

Dehejia, R. H. and Wahba, S. (1999). Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American statistical Association* **94,** 1053–1062.

Drake, C. (1993). Effects of misspecification of the propensity score on estimators of treatment effect. *Biometrics* **49,** 1231–1236.

Harder, V. S., Stuart, E. A., and Anthony, J. C. (2010). Propensity score techniques and the assessment of measured covariate balance to test causal associations in psychological research. *Psychological methods* **15,** 234.

Hernán, M. A. and Cole, S. R. (2009). Invited commentary: Causal diagrams and measurement bias. *American Journal of Epidemiology* **170,** 959–962.

Hill, J. L., Reiter, J. P., and Zanutto, E. L. (2004). A comparison of experimental and observational data analyses. *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives: An Essential Journey with Donald Rubin's Statistical Family* pages 49–60.

Ho, D. E. (2009). *MatchIt: Nonparametric preprocessing for parametric causal inference.* PhD thesis, Harvard University, Cambridge MA.

Ho, D. E., Imai, K., King, G., and Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political analysis* **15,** 199–236.

Jurek, A. M., Maldonado, G., Greenland, S., and Church, T. R. (2006). Exposure-measurement error is frequently ignored when interpreting epidemiologic study results. *European journal of epidemiology* **21,** 871–876.

McCaffrey, D. F., Lockwood, J. R., and Setodji, C. M. (2013). Inverse probability weighing with error-prone covariates. *Biometrika* **100,** 671–680.

McCandless, L. C., Richardson, S., and Best, N. (2012). Adjustment for missing confounders using external validation data and propensity scores. *Journal of the American Statistical Association* **107,** 40–51.

Robins, J. M., Hernán, M. Á., and Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology* **11,** 550–560.

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

Rosenbaum, P. R. (2005). Propensity score. *Encyclopedia of biostatistics* .

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.

Schafer, J. L. and Kang, J. (2008). Average causal effects from nonrandomized studies: a practical guide and simulated example. *Psychological methods* **13,** 279.

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical science: a review journal of the Institute of Mathematical Statistics* **25,** 1.

Stuart, E. A. and Green, K. M. (2008). Using full matching to estimate causal effects in nonexperimental studies: examining the relationship between adolescent marijuana use and adult outcomes. *Developmental Psychology* **44,** 395.

Stürmer, T., Schneeweiss, S., Avorn, J., and Glynn, R. J. (2005). Adjusting effect estimates for unmeasured confounding with validation data using propensity score calibration. *American journal of epidemiology* **162,** 279–289.

Appendix

Stratification *logit* outcome model likelihood

In the *logit* outcome model case, the likelihood can be written as follows:

$$
\begin{aligned}
L(\beta) = \prod_{i \in N_k} & [P(x_i = 1 | x_i^* = 1, \mathbf{c}_i) \frac{1}{1 + \exp(-(\beta_{0k} + \beta_{1k} + \beta_{2k}\mathbf{c}_i))} \\
& + P(x_i = 0 | x_i^* = 1, \mathbf{c}_i) \frac{1}{1 + \exp(-(\beta_{0k} + \beta_{2k}\mathbf{c}_i))}]^{x_i^* y_i} \\
& * [P(x_i = 1 | x_i^* = 0, \mathbf{c}_i) \frac{1}{1 + \exp(-(\beta_{0k} + \beta_{1k} + \beta_{2k}\mathbf{c}_i))} \\
& + P(x_i = 0 | x_i^* = 0, \mathbf{c}_i) \frac{1}{1 + \exp(-(\beta_{0k} + \beta_{2k}\mathbf{c}_i))}]^{(1 - x_i^*) y_i} \\
& * [P(x_i = 1 | x_i^* = 1, \mathbf{c}_i)(1 - \frac{1}{1 + \exp(-(\beta_{0k} + \beta_{1k} + \beta_{2k}\mathbf{c}_i))}) \\
& + P(x_i = 0 | x_i^* = 1, \mathbf{c}_i)(1 - \frac{1}{1 + \exp(-(\beta_{0k} + \beta_{2k}\mathbf{c}_i))})]^{x_i^* (1 - y_i)}
\end{aligned}
$$

$$
\begin{aligned}
&*[P(x_i=1|x_i^*=0,\mathbf{c}_i)(1-\frac{1}{1+\exp(-(\beta_{0k}+\beta_{1k}+\beta_{2k}\mathbf{c}_i))})\\
&+P(x_i=0|x_i^*=0,\mathbf{c}_i)(1-\frac{1}{1+\exp(-(\beta_{0k}+\beta_{2k}\mathbf{c}_i))})]^{(1-x_i^*)(1-y_i)}\\
&*\prod_{j\in N_{vk}}[\frac{1}{1+\exp(-(\beta_{0k}+\beta_{1k}x_j+\beta_{2k}\mathbf{c}_j))}]^{y_j}[1-\frac{1}{1+\exp(-(\beta_{0k}+\beta_{1k}x_j+\beta_{2k}\mathbf{c}_j))}]^{1-y_j}
\end{aligned}
\tag{12}
$$

Weighted likelihood *logit* outcome model likelihood

In the *logit* outcome model case, the likelihood can be written as follows:

$$
\begin{aligned}
L(\beta)=&\prod_i^{N_m}[P(x_i=1|x_i^*=1,\mathbf{c}_i)\frac{1}{1+\exp(-(\beta_0+\beta_1+\beta_2\mathbf{c}_i))}\\
&+P(x_i=0|x_i^*=1,\mathbf{c}_i)\frac{1}{1+\exp(-(\beta_0+\beta_2\mathbf{c}_i))}]^{x_i^*y_i\frac{1}{P(x_i^*|\mathbf{c}_i,\gamma_{adj})}}\\
&+[P(x_i=1|x_i^*=0,\mathbf{c}_i)\frac{1}{1+\exp(-(\beta_0+\beta_1+\beta_2\mathbf{c}_i))}\\
&+P(x_i=0|x_i^*=0,\mathbf{c}_i)\frac{1}{1+\exp(-(\beta_{0+\beta_2\mathbf{c}_i}))}]^{(1-x_i^*)y_i\frac{1}{1-P(x_i^*|\mathbf{c}_i,\gamma_{adj})}}\\
&+[P(x_i=1|x_i^*=1,\mathbf{c}_i)(1-\frac{1}{1+\exp(-(\beta_0+\beta_1+\beta_2\mathbf{c}_i))})\\
&+P(x_i=0|x_i^*=1,\mathbf{c}_i)(1-\frac{1}{1+\exp(-(\beta_0+\beta_2\mathbf{c}_i))})]^{x_i^*(1-y_i)\frac{1}{P(x_i^*|\mathbf{c}_i,\gamma_{adj})}}\\
&+[P(x_i=1|x_i^*=0,\mathbf{c}_i)(1-\frac{1}{1+\exp(-(\beta_0+\beta_1+\beta_2\mathbf{c}_i))})\\
&+P(x_i=0|x_i^*=0,\mathbf{c}_i)(1-\frac{1}{1+\exp(-(\beta_0+\beta_2\mathbf{c}_i))})]^{(1-x_i^*)(1-y_i)\frac{1}{1-P(x_i^*|\mathbf{c}_i,\gamma_{adj})}}\\
&*\prod_j^{N_v}[\frac{1}{1+\exp(-(\beta_0+\beta_1x_j+\beta_2\mathbf{c}_j))}]^{y_jx_j\frac{1}{P(x_j|\mathbf{c}_j,\gamma_x)}}[1-\frac{1}{1+\exp(-(\beta_0+\beta_1x_j+\beta_2\mathbf{c}_j))}]^{(1-y_j)x_j\frac{1}{P(x_j|\mathbf{c}_j,\gamma_x)}}\\
&*[\frac{1}{1+\exp(-(\beta_0+\beta_2\mathbf{c}_j))}]^{y_j(1-x_j)\frac{1}{1-P(x_j|\mathbf{c}_j,\gamma_x)}}[1-\frac{1}{1+\exp(-(\beta_0+\beta_2\mathbf{c}_j))}]^{(1-y_j)(1-x_j)\frac{1}{1-P(x_j|\mathbf{c}_j,\gamma_x)}}
\end{aligned}
\tag{13}
$$

Matching *logit* model likelihood

In the *logit* outcome model case, the likelihood can be written as follows:

$$
L(\beta)=\prod_i^{N_m}\{[P(x_i=1|x_i^*=1,\mathbf{c}_i)\frac{1}{1+\exp(-(\beta_0+\beta_1+\beta_2\mathbf{c}_i))}
$$

$$
\begin{aligned}
&+P(x_i=0|x_i^*=1,\mathbf{c}_i)\frac{1}{1+\exp(-(\beta_0+\beta_2\mathbf{c}_i))}]^{x_i^*y_i}\\
&+[P(x_i=1|x_i^*=0,\mathbf{c}_i)\frac{1}{1+\exp(-(\beta_0+\beta_1+\beta_2\mathbf{c}_i))}\\
&+P(x_i=0|x_i^*=0,\mathbf{c}_i)\frac{1}{1+\exp(-(\beta_{0+\beta_2\mathbf{c}_i}))}]^{(1-x_i^*)y_i}\\
&+[P(x_i=1|x_i^*=1,\mathbf{c}_i)(1-\frac{1}{1+\exp(-(\beta_0+\beta_1+\beta_2\mathbf{c}_i))})\\
&+P(x_i=0|x_i^*=1,\mathbf{c}_i)(1-\frac{1}{1+\exp(-(\beta_0+\beta_2\mathbf{c}_i))})]^{x_i^*(1-y_i)}\\
&+[P(x_i=1|x_i^*=0,\mathbf{c}_i)(1-\frac{1}{1+\exp(-(\beta_0+\beta_1+\beta_2\mathbf{c}_i))})\\
&+P(x_i=0|x_i^*=0)(1-\frac{1}{1+\exp(-(\beta_0+\beta_2\mathbf{c}_i))})]^{(1-x_i^*)(1-y_i)}\}^{\widehat{w_{adji}}} \qquad (14)\\
&*\prod_j^{N_v}\{[\frac{1}{1+\exp(-(\beta_0+\beta_1x_j+\beta_2\mathbf{c}_j))}]^{y_j}[1-\frac{1}{1+\exp(-(\beta_0+\beta_1x_j+\beta_2\mathbf{c}_j))}]^{1-y_j}\}^{\widehat{w_{truei}}}
\end{aligned}
$$

PS Covariate Adjustment *logit* model likelihood

In the *logit* outcome model case, the likelihood can be written as follows:

$$
\begin{aligned}
L(\beta)=\prod_i^{N_m}&[P(x_i=1|x_i^*=1,\mathbf{c}_i)\frac{1}{1+\exp(-(\beta_0+\beta_1+\beta_2\mathbf{c}_i+\beta_3\widehat{PS}_{adj}))}\\
&+P(x_i=0|x_i^*=1,\mathbf{c}_i)\frac{1}{1+\exp(-(\beta_0+\beta_2\mathbf{c}_i+\beta_3\widehat{PS}_{adj}))}]^{x_i^*y_i}\\
&[P(x_i=1|x_i^*=0,\mathbf{c}_i)\frac{1}{1+\exp(-(\beta_0+\beta_1+\beta_2\mathbf{c}_i+\beta_3\widehat{PS}_{adj}))}\\
&+P(x_i=0|x_i^*=0,\mathbf{c}_i)\frac{1}{1+\exp(-(\beta_0+\beta_2\mathbf{c}_i+\beta_3\widehat{PS}_{adj}))}]^{(1-x_i^*)y_i}\\
&+[P(x_i=1|x_i^*=1,\mathbf{c}_i)(1-\frac{1}{1+\exp(-(\beta_0+\beta_1+\beta_2\mathbf{c}_i+\beta_3\widehat{PS}_{adj}))})\\
&+P(x_i=0|x_i^*=1,\mathbf{c}_i)(1-\frac{1}{1+\exp(-(\beta_0+\beta_2\mathbf{c}_i+\beta_3\widehat{PS}_{adj}))})]^{x_i^*(1-y_i)}\\
&+[P(x_i=1|x_i^*=0,\mathbf{c}_i)(1-\frac{1}{1+\exp(-(\beta_0+\beta_1+\beta_2\mathbf{c}_i+\beta_3\widehat{PS}_{adj}))})\\
&+P(x_i=0|x_i^*=0,\mathbf{c}_i)(1-\frac{1}{1+\exp(-(\beta_0+\beta_2\mathbf{c}_i+\beta_3\widehat{PS}_{adj}))})]^{(1-x_i^*)(1-y_i)} \qquad (15)
\end{aligned}
$$

$$* \prod_{j}^{N_v} [\frac{1}{1+\exp(-(\beta_0+\beta_1 x_j+\beta_2 \mathbf{c}_j+\beta_3 \widehat{PS_{true}}))}]^{y_j} [1-\frac{1}{1+\exp(-(\beta_0+\beta_1 x_j+\beta_2 \mathbf{c}_j+\beta_3 \widehat{PS_{true}}))}]^{1-y_j}$$