## 1 Proof for the two group setting

### 1.1 Estimand of Matching

This proof essentially follows the structure of the proof in the appendix of Li & Greene's 2013 paper[1]. The initial expression for the sample mean outcome in the matched treated group appears different from theirs, *i.e.*, $\frac{\sum_{k=1}^{K}\sum_{i=1}^{n} Y_i I(i \in \mathcal{S}_{1k})}{\sum_{k=1}^{K}\sum_{i=1}^{n} I(i \in \mathcal{S}_{1k})}$ where $k$ is an index over discrete values of propensity scores, however both are the equivalent sample marginal mean outcome in the matched treated group. Instead of the explicit sum over $k$, we define a specific structure for the matched set.

The usual causal inference assumptions[2] are all required. The first is conditional exchangeability (unconfoundedness) given a function of the covariate vector $\mathbf{X}_i$ including the vector itself (finest balancing score) or the propensity score (coarsest balancing score). The latter requires no model misspecification for the propensity score model. The second is consistency, *i.e.,* $Y_i = Z_i Y_{1i} + (1 - Z_i) Y_{0i}$. That is, the observed outcome is the counterfactual potential outcome corresponding to the treatment received. This requires well-defined treatment and non-interference among individuals' potential outcomes. The third is positivity, *i.e.*, at any level of $\mathbf{X}_i$ (and thus propensity score), both treatment choices have non-zero (positive) probability. In this setting, this implies a perfect common support, *i.e.*, any propensity score values present in one of the treatment groups are also present in the other group.

Additional assumptions are required for the propensity score matching process. Matching has to be 1:1 matching without replacement. It also has to be exact matching on propensity scores (no calipers are allowed). This necessarily requires discrete propensity scores taking on a finite set of values because there has to be a positive probability of finding an exact match across two treatment groups[1]. The set of values can be arbitrarily large as long as its size is bounded and does not grow with the sample size $n$. When multiple untreated candidates are available for matching a treated individual at a given propensity score ($< 0.5$), one is selected at random. The same should apply when there are more treated individuals than untreated individuals at a given propensity score ($> 0.5$).

**Proof**: Let $l \in \{1, 2, ..., L\}$ be the index for the propensity score matched pairs. Let $\mathcal{S}_{1l}$ be the single member set of the treated subject from the $l$-th matched pair and the $\mathcal{S}_{0l}$ be the corresponding set of the untreated subject. Thus, $\mathcal{S}_1 = \bigcup\limits_{l=1}^{L} \mathcal{S}_{1l}$ is the set of matched treated subjects, $\mathcal{S}_0 = \bigcup\limits_{l=1}^{L} \mathcal{S}_{0l}$ is the set of matched untreated subjects, and $\mathcal{S} = \mathcal{S}_0 \cup \mathcal{S}_1$ is the set of the entire matched cohort. This matched cohort is balanced, *i.e.*, both groups contain the same number ($L$) of matched subjects. Index $n$ is over the entire dataset before matching, thus, it includes subjects who do not match. The group mean in the matched treated group is expressed as follows. The selection indicator is effectively acting as a 0, 1 weight.

$$\frac{\sum_{i=1}^{n} Y_i I(i \in \mathcal{S}_1)}{\sum_{i=1}^{n} I(i \in \mathcal{S}_1)}$$

The numerator is examined first. The expression is multiplied by $\frac{1}{n}$, but it cancels out in the original expression as we do the same for the denominator. $Y_i$ is the observed outcome of the $i$-th subject, whereas $Y_{1i}$ is the treated counterfactual potential outcome of the $i$-th subject.

By consistency, the treated counterfactual is observed among the treated.

Only the treated contribute to the expression, thus, $Y_i = Y_{1i}$.

$$\frac{1}{n}\sum_{i=1}^{n} Y_i I(i \in \mathcal{S}_1) = \frac{1}{n}\sum_{i=1}^{n} Y_{1i} I(i \in \mathcal{S}_1)$$

Asymptotically, by the Weak Law of Large Number

$$\xrightarrow{p} E[Y_{1i} I(i \in \mathcal{S}_1)]$$

Rewrite as an iterative expectation.

$$= E[E[Y_{1i} I(i \in \mathcal{S}_1)|\mathbf{X}_i]]$$

Break the indicator into selection and treatment.

$$= E[E[Y_{1i} I(i \in \mathcal{S}) I(Z_i = 1)|\mathbf{X}_i]]$$

$\because$ only the treated subjects contribute to the inner expectation,

and otherwise it is zero, expectation can be taken

in the treated and weighted by its prevalence.

$$= E[E[Y_{1i} I(i \in \mathcal{S})|Z_i = 1, \mathbf{X}_i] P(Z_i = 1|\mathbf{X}_i)]$$

$\because$ given $Z_i = 1$ and within levels of $\mathbf{X}_i$, selection $(i \in \mathcal{S})$ is random,

$Y_{1i}$ and selection indicator are conditionally independent.

$$= E[E[Y_{1i}|Z_i = 1, \mathbf{X}_i] E[I(i \in \mathcal{S})|Z_i = 1, \mathbf{X}_i] P(Z_i = 1|\mathbf{X}_i)]$$

By conditional exchangeability, $E[Y_{1i}|Z_i = 1, \mathbf{X}_i] = E[Y_{1i}|Z_i = 0, \mathbf{X}_i] = E[Y_{1i}|\mathbf{X}_i]$.

$$= E[E[Y_{1i}|\mathbf{X}_i] E[I(i \in \mathcal{S})|Z_i = 1, \mathbf{X}_i] P(Z_i = 1|\mathbf{X}_i)]$$

$\because$ expectation of a 0,1 selection indicator is the selection probability.

$$= E[E[Y_{1i}|\mathbf{X}_i] P(i \in \mathcal{S}|Z_i = 1, \mathbf{X}_i) P(Z_i = 1|\mathbf{X}_i)]$$

The last term is the propensity score by definition.

$$= E[E[Y_{1i}|\mathbf{X}_i] P\left(i \in \mathcal{S}|Z_i = 1, \mathbf{X}_i\right) e_i]$$

At a given $\mathbf{X}_i$, only the smaller group can match fully.

$e_i$ is the fraction of the treated group at a given $\mathbf{X}_i$.

$\min(e_i, 1 - e_i)$ is the fraction of the smaller group at $\mathbf{X}_i$.

$\therefore$ among the treated group, only $\frac{\min(e_i, 1 - e_i)}{e_i}$ can match.

As this is a function of $\mathbf{X}_i$, conditioning is implicit.

$$= E\left[E[Y_{1i}|\mathbf{X}_i] \frac{\min(e_i, 1 - e_i)}{e_i} e_i\right]$$

$$= E\left[E[Y_{1i}|\mathbf{X}_i]\min(e_i, 1-e_i)\right]$$

The denominator is a simplified version of the above proof.

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} I(i \in \mathcal{S}_1) &= \frac{1}{n}\sum_{i=1}^{n} I(i \in \mathcal{S}_1)\\
&\overset{p}{\to} E[I(i \in \mathcal{S}_1)]\\
&= E[E[I(i \in \mathcal{S}_1)|\mathbf{X}_i]]\\
&= E[E[I(i \in \mathcal{S})I(Z_i = 1)|\mathbf{X}_i]]\\
&= E[E[I(i \in \mathcal{S})|Z_i = 1, \mathbf{X}_i]P\left(Z_i = 1|\mathbf{X}_i\right)]]\\
&= E[P(i \in \mathcal{S}|Z_i = 1, \mathbf{X}_i)P\left(Z_i = 1|\mathbf{X}_i\right)]]\\
&= E[P\left(i \in \mathcal{S}|Z_i = 1, \mathbf{X}_i\right)e_i]\\
&= E\left[\frac{\min(e_i, 1-e_i)}{e_i}e_i\right]\\
&= E\left[\min(e_i, 1-e_i)\right]
\end{aligned}$$

Therefore, the estimand of the group mean of the matched treated cohort is asymptotically the following.

$$\frac{E\left[E[Y_{1i}|\mathbf{X}_i]\min(e_i, 1-e_i)\right]}{E\left[\min(e_i, 1-e_i)\right]}$$

Similarly, the estimand of the group mean of the matched untreated cohort is asymptotically the following.

$$\frac{E\left[E[Y_{0i}|\mathbf{X}_i]\min(e_i, 1-e_i)\right]}{E\left[\min(e_i, 1-e_i)\right]}$$

Using these, the estimand of the group mean difference is

$$\begin{aligned}
\hat{\Delta}_M &= \frac{\sum_{i=1}^{n} Y_i I(i \in \mathcal{S}_1)}{\sum_{i=1}^{n} I(i \in \mathcal{S}_1)} - \frac{\sum_{i=1}^{n} Y_i I(i \in \mathcal{S}_0)}{\sum_{i=1}^{n} I(i \in \mathcal{S}_0)}\\
&= \frac{\sum_{i=1}^{n} Y_{1i} I(i \in \mathcal{S}_1)}{\sum_{i=1}^{n} I(i \in \mathcal{S}_1)} - \frac{\sum_{i=1}^{n} Y_{0i} I(i \in \mathcal{S}_0)}{\sum_{i=1}^{n} I(i \in \mathcal{S}_0)}\\
&\overset{p}{\to} \frac{E\left[E[Y_{1i}|\mathbf{X}_i]\min(e_i, 1-e_i)\right]}{E\left[\min(e_i, 1-e_i)\right]} - \frac{E\left[E[Y_{0i}|\mathbf{X}_i]\min(e_i, 1-e_i)\right]}{E\left[\min(e_i, 1-e_i)\right]}\\
&= \frac{E\left[E[Y_{1i}|\mathbf{X}_i]\min(e_i, 1-e_i) - E\left[E[Y_{0i}|\mathbf{X}_i]\min(e_i, 1-e_i)\right]\right]}{E\left[\min(e_i, 1-e_i)\right]}\\
&= \frac{E\left[(E[Y_{1i}|\mathbf{X}_i] - E[Y_{0i}|\mathbf{X}_i])\min(e_i, 1-e_i)\right]}{E\left[\min(e_i, 1-e_i)\right]}\\
&= \frac{E\left[E[Y_{1i} - Y_{0i}|\mathbf{X}_i]\min(e_i, 1-e_i)\right]}{E\left[\min(e_i, 1-e_i)\right]}
\end{aligned}$$

$$= \frac{E\left[\Delta_i \min(e_i, 1-e_i)\right]}{E\left[\min(e_i, 1-e_i)\right]}$$

where $\Delta_i$ is the causal effect given covariates.

### 1.2 Estimand of Matching Weight

The corresponding matching weight estimator of the mean outcome in the treated is the following. The same causal inference assumptions are required except for the additional assumptions required for the matching algorithm.

$$\frac{\sum_{i=1}^{n} Y_i Z_i W_i}{\sum_{i=1}^{n} Z_i W_i}$$

where

$$W_i = \frac{\min(e_i, 1-e_i)}{Z_i e_i + (1-Z_i)(1-e_i)}$$

i.e., $W_i$ is a function of covariates $\mathbf{X}_i$ and treatment $Z_i$.

The numerator has the following asymptotic characteristic.

By consistency, the treated counterfactual is observed among the treated.

Only the treated contribute to the expression, thus, $Y_i = Y_{1i}$.

$$\frac{1}{n}\sum_{i=1}^{n} Y_i Z_i W_i = \frac{1}{n}\sum_{i=1}^{n} Y_{1i} Z_i W_i$$

Asymptotically, by the Weak Law of Large Number

$$\overset{p}{\to} E[Y_{1i} Z_i W_i]$$

Rewrite as an iterative expectation.

$$= E[E[Y_{1i} Z_i W_i | \mathbf{X}_i]]$$

$\because (Y_{1i}, Y_{0i}) \perp\!\!\!\perp Z_i | \mathbf{X}_i$ implies $(Y_{1i}, Y_{0i}) \perp\!\!\!\perp f(\mathbf{X}_i, Z_i) | \mathbf{X}_i$,

the following holds $(Y_{1i}, Y_{0i}) \perp\!\!\!\perp Z_i W_i | \mathbf{X}_i$

$$= E[E[Y_{1i} | \mathbf{X}_i] E[Z_i W_i | \mathbf{X}_i]]$$

$\because$ only the treated units contribute to the second term,

and otherwise it is zero, expectation can be taken

in the treated and weighted by its prevalence.

$$= E[E[Y_{1i} | \mathbf{X}_i] E[W_i | Z_i = 1, \mathbf{X}_i] P(Z_i = 1 | \mathbf{X}_i)]$$

The last term is the propensity score by definition.

Also expand the weight.

$$
\begin{aligned}
&= E\left[E[Y_{1i}|\mathbf{X}_i]E\left[\frac{\min(e_i, 1-e_i)}{Z_i e_i + (1-Z_i)(1-e_i)}\middle|Z_i = 1, \mathbf{X}_i\right] e_i\right] \\
&\quad \because Z_i = 1 \text{ for the second term} \\
&= E\left[E[Y_{1i}|\mathbf{X}_i]\frac{\min(e_i, 1-e_i)}{e_i} e_i\right] \\
&= E\left[E[Y_{1i}|\mathbf{X}_i]\min(e_i, 1-e_i)\right]
\end{aligned}
$$

Similarly, the denominator has the following asymptotic characteristic.

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} Z_i W_i \xrightarrow{p} E[Z_i W_i]& \\
&= E[E[Z_i W_i|\mathbf{X}_i]] \\
&= E[E[W_i|Z_i = 1, \mathbf{X}_i]P(Z_i = 1|\mathbf{X}_i)] \\
&= E\left[E\left[\frac{\min(e_i, 1-e_i)}{Z_i e_i + (1-Z_i)(1-e_i)}\middle|Z_i = 1, \mathbf{X}_i\right] e_i\right] \\
&= E\left[\frac{\min(e_i, 1-e_i)}{e_i} e_i\right] \\
&= E\left[\min(e_i, 1-e_i)\right]
\end{aligned}
$$

Therefore, the estimand of matching weight estimator for the treated group mean has the same form as the corresponding matching estimator asymptotically.

$$
\begin{aligned}
\frac{\sum_{i=1}^{n} Y_i Z_i W_i}{\sum_{i=1}^{n} Z_i W_i} &= \frac{\sum_{i=1}^{n} Y_{1i} Z_i W_i}{\sum_{i=1}^{n} Z_i W_i} \\
&\xrightarrow{p} \frac{E\left[E[Y_{1i}|\mathbf{X}_i]\min(e_i, 1-e_i)\right]}{E\left[\min(e_i, 1-e_i)\right]}
\end{aligned}
$$

Because this holds similarly for the untreated group, the estimand of the treatment effect is also asymptotically equivalent.

## 2 Extension to 3+ group settings

In the previous proof following Li & Greene 2013, the effect estimate was compared between the matching method and the matching weight method. Proving the asymptotic equivalence of the estimand of an arbitrary group-specific mean outcome in 3+ group setting will generalize the proof. The same assumptions are required on all the treatment groups under study.

### 2.1 Estimand of Matching in 3+ group setting

One propensity score is defined for each treatment group. For the $k$-th treatment group, $e_{ki}$ is the corresponding treatment-specific propensity score, *i.e.*, the probability of being assigned to the $k$-th treatment group for the $i$-th subject given covariates. The treatment-specific propensity scores must be formed in such a way that within an individual subject $\sum\limits_{k=1}^{K} e_{ki} = 1$ is met. This requires a single model be fit for estimation (*e.g.*, multinomial logistic regression).

The same assumptions as the two group setting are required. Regarding the matching process now it is a simultaneous $1 : 1 : ... : 1$ exact matching of $K$ treatment groups on their $K$ treatment-specific propensity scores without replacement. That is, $K$ individuals with the identical propensity scores (all of the treatment-specific propensity scores, $e_{1i}, \dots, e_{Ki}$ must match up across $K$ individuals) form a matched unit. If there are multiple candidates from a given treatment group $k$, the selection is random.

**Proof**: Let $\mathcal{S}_{kl}$ be the single member set of the subject in the $k$-th treatment group ($k \in \{1, 2, ..., K\}$) from the $l$-th propensity score matched unit ($l \in \{1, 2, ..., L\}$). Thus, $\mathcal{S}_k = \bigcup\limits_{l=1}^{L} \mathcal{S}_{kl}$ is the set of all matched subjects in the $k$-th treatment group, and $\mathcal{S} = \bigcup\limits_{k=1}^{K} \mathcal{S}_k$ is the set of entire matched cohort. This matched cohort is balanced, *i.e.*, each one of $K$ treatment groups contain the same number ($L$) of matched subjects. Index $n$ is still over all individuals in the dataset before matching. The treatment variable, $Z_i$ is now a nominal variable $1, 2, ..., K$ indicating the treatment group. The group mean in the $k$-th group is expressed as follows.

$$\frac{\sum_{i=1}^{n} Y_i I(i \in \mathcal{S}_k)}{\sum_{i=1}^{n} I(i \in \mathcal{S}_k)}$$

The numerator is examined first. The expression is multiplied by $\frac{1}{n}$, but it cancels in the original expression as we do the same for the denominator. For the most part the proof is almost identical to the previous one.

By consistency, the $k$-th counterfactual is observed in the $k$-th group

Also only the $k$-th group contributes to the expression, thus, $Y_i = Y_{ki}$

$$\frac{1}{n}\sum_{i=1}^{n} Y_i I(i \in \mathcal{S}_k) = \frac{1}{n}\sum_{i=1}^{n} Y_{ki} I(i \in \mathcal{S}_k)$$

Asymptotically, by the Weak Law of Large Number

$$\xrightarrow{p} E[Y_{ki} I(i \in \mathcal{S}_k)]$$

Rewrite as an iterative expectation.

$$= E[E[Y_{ki} I(i \in \mathcal{S}_k)|\mathbf{X}_i]]$$

Break the indicator into selection and treatment.

$$= E[E[Y_{ki} I(i \in \mathcal{S}) I(Z_i = k)|\mathbf{X}_i]]$$

$\because$ only the $k$-th group contributes to the inner expectation,

and otherwise it is zero, expectation can be taken

in the $k$-th group and weighted by its prevalence.

$$= E[E[Y_{ki} I(i \in \mathcal{S})|Z_i = k, \mathbf{X}_i] P(Z_i = k|\mathbf{X}_i)]$$

$\because$ given $Z_i = k$ and within levels of $\mathbf{X}_i$, selection ($i \in \mathcal{S}$) is random,

$Y_{ki}$ and selection indicator are conditionally independent.

$$= E[E[Y_{ki}|Z_i = k, \mathbf{X}_i] E[I(i \in \mathcal{S})|Z_i = k, \mathbf{X}_i] P(Z_i = k|\mathbf{X}_i)]$$

By conditional exchangeability, $E[Y_{ki}|Z_i = k, \mathbf{X}_i] = E[Y_{ki}|\mathbf{X}_i]$.

$$= E[E[Y_{ki}|\mathbf{X}_i] E[I(i \in \mathcal{S})|Z_i = k, \mathbf{X}_i] P(Z_i = k|\mathbf{X}_i)]$$

$\because$ expectation of a 0,1 selection indicator is the selection probability.

$$= E[E[Y_{ki}|\mathbf{X}_i] P(i \in \mathcal{S}|Z_i = k, \mathbf{X}_i) P(Z_i = k|\mathbf{X}_i)]$$

The last term is the PS for the $k$-th treatment by definition.

$$= E[E[Y_{ki}|\mathbf{X}_i] P\left(i \in \mathcal{S}|Z_i = k, \mathbf{X}_i\right) e_{ki}]$$

At a given $\mathbf{X}_i$, only the smallest group can match fully.

$e_{ki}$ is the fraction of $k$-th group at a given $\mathbf{X}_i$.

$\min(e_{1i}, e_{2i}, ..., e_{Ki})$ is the fraction of the smallest group at $\mathbf{X}_i$.

$\therefore$ Among the $k$-th group, only $\dfrac{\min(e_{1i}, e_{2i}, ..., e_{Ki})}{e_{ki}}$ can match.

As this is a function of $\mathbf{X}_i$, conditioning is implicit.

$$= E\left[E[Y_{ki}|\mathbf{X}_i] \frac{\min(e_{1i}, e_{2i}, ..., e_{Ki})}{e_{ki}} e_{ki}\right]$$

$$= E\left[E[Y_{ki}|\mathbf{X}_i] \min(e_{1i}, e_{2i}, ..., e_{Ki})\right]$$

Similarly,

$$\frac{1}{n}\sum_{i=1}^{n} I(i \in \mathcal{S}_k) = \frac{1}{n}\sum_{i=1}^{n} I(i \in \mathcal{S}_k)$$

$$\xrightarrow{p} E\left[\min(e_{1i}, e_{2i}, ..., e_{Ki})\right]$$

Therefore, the estimand of the group mean of the matched $k$-th group is asymptotically the following.

$$\frac{E\left[E[Y_{ki}|\mathbf{X}_i]\min(e_{1i}, e_{2i}, ..., e_{Ki})\right]}{E\left[\min(e_{1i}, e_{2i}, ..., e_{Ki})\right]}$$

### 2.2 Estimand of Matching Weight in 3+ group setting

The corresponding weighted estimator of the mean outcome in the treated is the following. The denominator of the weight picks the propensity score for the assigned treatment for the $i$-th unit.

$$\frac{\sum_{i=1}^n Y_i I(Z_i = k) W_i}{\sum_{i=1}^n I(Z_i = k) W_i}$$

where

$$W_i = \frac{\min(e_{1i}, e_{2i}, ..., e_{Ki})}{\sum_{k=1}^K I(Z_i = k) e_{ki}}$$

The numerator has the following asymptotic characteristic.

$$\begin{aligned}
& && \text{By consistency, the } k\text{-th counterfactual is observed in the } k\text{-th group} \\
& && \text{Also only the } k\text{-th group contributes to the expression, thus, } Y_i = Y_{ki} \\
\frac{1}{n}\sum_{i=1}^n Y_i I(Z_i = k) W_i &= \frac{1}{n}\sum_{i=1}^n Y_{ki} I(Z_i = k) W_i \\
& && \text{Asymptotically, by the Weak Law of Large Number} \\
& \xrightarrow{p} E[Y_{ki} I(Z_i = k) W_i] \\
& && \text{Rewrite as an iterative expectation.} \\
&= E[E[Y_{ki} I(Z_i = k) W_i|\mathbf{X}_i]] \\
& && \because Y_{ki} \perp\!\!\!\perp Z_i|\mathbf{X}_i \text{ implies } Y_{ki} \perp\!\!\!\perp f(\mathbf{X}_i, Z_i)|\mathbf{X}_i, \\
& && \quad \text{the following holds } Y_{ki} \perp\!\!\!\perp I(Z_i = k)W_i|\mathbf{X}_i \\
&= E[E[Y_{ki}|\mathbf{X}_i] E[I(Z_i = k) W_i|\mathbf{X}_i]] \\
& && \because \text{only the } k\text{-th group contributes to the second term,} \\
& && \quad \text{and otherwise it is zero, expectation can be taken} \\
& && \quad \text{in the } k\text{-th group and weighted by its prevalence.} \\
&= E[E[Y_{ki}|\mathbf{X}_i] E[W_i|Z_i = k, \mathbf{X}_i] P(Z_i = k|\mathbf{X}_i)] \\
& && \text{The last term is the propensity score for the } k\text{-th treatment.} \\
& && \text{Also expand the weight.}
\end{aligned}$$

$$= E\left[E[Y_{ki}|\mathbf{X}_i]E\left[\frac{\min(e_{1i}, e_{2i}, ..., e_{Ki})}{\sum_{k=1}^{K} I(Z_i = k)e_{ki}}\middle| Z_i = k, \mathbf{X}_i\right] e_{ki}\right]$$

$\because Z_i = k$ for the second term

$$= E\left[E[Y_{ki}|\mathbf{X}_i]\frac{\min(e_{1i}, e_{2i}, ..., e_{Ki})}{e_{ki}} e_{ki}\right]$$

$$= E\left[E[Y_{ki}|\mathbf{X}_i]\min(e_{1i}, e_{2i}, ..., e_{Ki})\right]$$

Similarly,

$$\frac{1}{n}\sum_{i=1}^{n} I(Z_i = k)W_i \overset{p}{\to} E[I(Z_i = k)W_i]$$

$$= E\left[\min(e_{1i}, e_{2i}, ..., e_{Ki})\right]$$

Therefore, the estimand of matching weight estimator has the same form as the matching estimator asymptotically.

$$\frac{\sum_{i=1}^{n} Y_i I(Z_i = k)W_i}{\sum_{i=1}^{n} I(Z_i = k)W_i} = \frac{\sum_{i=1}^{n} Y_{ki} I(Z_i = k)W_i}{\sum_{i=1}^{n} I(Z_i = k)W_i}$$

$$\overset{p}{\to} \frac{E\left[E[Y_{ki}|\mathbf{X}_i]\min(e_{1i}, e_{2i}, ..., e_{Ki})\right]}{E\left[\min(e_{1i}, e_{2i}, ..., e_{Ki})\right]}$$

Because this holds true for each treatment group, the estimand of any two group contrast effect is also asymptotically equivalent between the multi-way matching method and the matching weight method.

**References**

[1] L. Li and T. Greene, "A weighting analogue to pair matching in propensity score analysis," *The International Journal of Biostatistics*, vol. 9, no. 2, pp. 215–234, 2013.

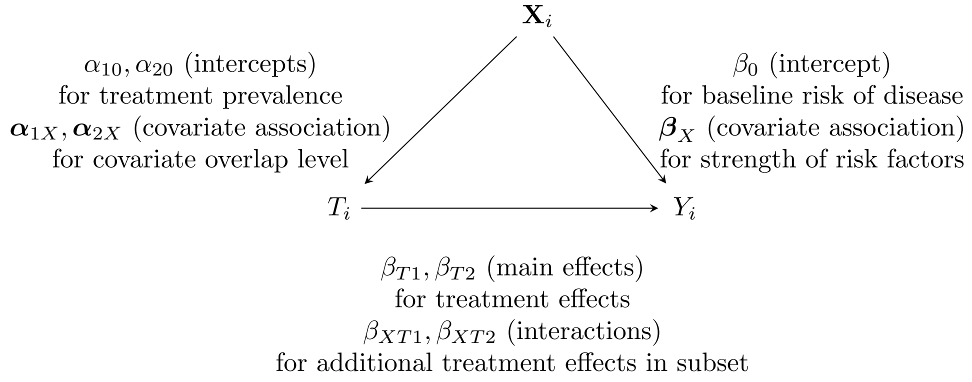
[2] M. A. Hernan and J. M. Robins, *Causal Inference.* Chapman & Hall/CRC, 2016.

## 1 Data Generation Mechanism DAG

The covariate were generated following the data generation process of Franklin *et al*[1]. The treatment assignment process also followed that of Franklin *et al*[1], but was extended to the three treatment group setting using a multinomial logistic model[2, 3]. The outcome model was a log-probability model to avoid non-collapsibility issues[4, 5].

### 1.1 Annotated Directed Acyclic Graph

$\mathbf{X}_i$ is a vector of ten covariates for the $i$-th individual, $T_i \in \{0, 1, 2\}$ is the treatment level, and $Y_i \in \{0, 1\}$ is the binary outcome.


$\mathbf{X}_i$
$\alpha_{10}, \alpha_{20}$ (intercepts)
for treatment prevalence
$\boldsymbol{\alpha}_{1X}, \boldsymbol{\alpha}_{2X}$ (covariate association)
for covariate overlap level
$\beta_0$ (intercept)
for baseline risk of disease
$\boldsymbol{\beta}_X$ (covariate association)
for strength of risk factors
$T_i$
$Y_i$
$\beta_{T1}, \beta_{T2}$ (main effects)
for treatment effects
$\beta_{XT1}, \beta_{XT2}$ (interactions)
for additional treatment effects in subset

### 1.2 Covariate Generation

The covariate vector for the $i$-th individual, $\mathbf{X}_i$ had the following random elements[1].

| Variable | Generation Process |
|---|---|
| $X_{1i}$ | Normal$(0, 1^2)$ |
| $X_{2i}$ | Log-Normal$(0, 0.5^2)$ |
| $X_{3i}$ | Normal$(0, 10^2)$ |
| $X_{4i}$ | Bernoulli$(p_i = e^{2X_{1i}}/(1 + e^{2X_{1i}}))$ where $E[p_i] = 0.5$ |
| $X_{5i}$ | Bernoulli$(p = 0.2)$ |
| $X_{6i}$ | Multinomial$(\mathbf{p} = (0.5, 0.3, 0.1, 0.05, 0.05)^T)$ |
| $X_{7i}$ | $\sin(X_{1i})$ |
| $X_{8i}$ | $X_{2i}^2$ |
| $X_{9i}$ | $X_{3i} \times X_{4i}$ |
| $X_{10i}$ | $X_{4i} \times X_{5i}$ |

### 1.3 Treatment Generating Model

As there were three treatment groups, two relative probabilities were jointly modeled by two simultaneous models (essentially multinomial logistic model).

$$\eta_{T1i} = \log\left(\frac{P(T_i = 1|\mathbf{X}_i = \mathbf{x}_i)}{P(T_i = 0|\mathbf{X}_i = \mathbf{x}_i)}\right) = \alpha_{10} + \boldsymbol{\alpha}_{1X}^T \mathbf{x}_i$$

$$\eta_{T2i} = \log\left(\frac{P(T_i = 2|\mathbf{X}_i = \mathbf{x}_i)}{P(T_i = 0|\mathbf{X}_i = \mathbf{x}_i)}\right) = \alpha_{20} + \boldsymbol{\alpha}_{2X}^T \mathbf{x}_i$$

where

$\alpha_{10}, \alpha_{20}$ determine treatment prevalence

$\boldsymbol{\alpha}_{1X}, \boldsymbol{\alpha}_{2X}$ determine covariate-treatment association

Importantly, the covariate-treatment association is inversely correlated with the covariate overlap in these model. This is because if patient characteristics play more important roles in treatment decision, the treatment assignment is less random.

To obtain the three predicted probabilities (true propensity scores) from the two linear predictors, we conducted the following normalization process[2, 3].

$$
\begin{aligned}
e_{0i} &= P(T_i = 0|\mathbf{X}_i = \mathbf{x}_i) = \frac{1}{q_i} \\
e_{1i} &= P(T_i = 1|\mathbf{X}_i = \mathbf{x}_i) = \frac{\exp(\eta_{T1i})}{q_i} \\
e_{2i} &= P(T_i = 2|\mathbf{X}_i = \mathbf{x}_i) = \frac{\exp(\eta_{T2i})}{q_i} \\
&\text{where } q_i = 1 + \exp(\eta_{T1i}) + \exp(\eta_{T2i})
\end{aligned}
$$

Finally, the treatment level was assigned in a multinomial random number generating process.

$$T_i \sim \text{Multinomial}\left(n = 1, \mathbf{p} = (e_{0i}, e_{1i}, e_{2i})^T\right)$$

### 1.4 Outcome Generating Model

The log probability of disease was generated using a log-linear (log-probability) model to avoid the non-collapsibility issue of the logistic model.

$$\eta_{Yi} = \log(P(Y_i = 1|T_i = t_i, \mathbf{X}_i = \mathbf{x}_i)) = \beta_0 + \boldsymbol{\beta}_X^T \mathbf{x}_i + \beta_{T1} I(t_i = 1) + \beta_{T2} I(t_i = 2) + \beta_{XT1} x_{4i} I(t_i = 1) + \beta_{XT2} x_{4i} I(t_i = 2)$$

$$
\begin{aligned}
&\text{where} \\
t_i &= \text{ Assigned treatment} \\
\beta_0 &= \text{ Intercept determining baseline disease risk} \\
\boldsymbol{\beta}_X &= \text{ Effects of ten covariates (risk factors) on disease risk} \\
\beta_{T1} &= \text{ Main effect of Treatment 1 compared to Treatment 0} \\
\beta_{T2} &= \text{ Main effect of Treatment 2 compared to Treatment 0} \\
\beta_{XT1} &= \text{ Additional effect for Treatment 1 vs 0 among } X_{4i} = 1 \\
\beta_{XT2} &= \text{ Additional effect for Treatment 2 vs 0 among } X_{4i} = 1
\end{aligned}
$$

Using this linear predictor, the probability of disease was calculated as follows.

$$p_{Yi} = P(Y_i = 1|T_i = t_i, \mathbf{X}_i = \mathbf{x}_i) = \exp(\eta_{Yi})$$

Then the outcome was assigned using a Bernoulli random number generating process.

$$Y_i \sim \text{Bernoulli}\left(p_{Yi}\right)$$

The counterfactual probability of disease under each treatment was defined as follows.

$$
\begin{aligned}
p_{Yi}(0) &= P(Y_i = 1|T_i = 0, \mathbf{X}_i = \mathbf{x}_i) \\
p_{Yi}(1) &= P(Y_i = 1|T_i = 1, \mathbf{X}_i = \mathbf{x}_i) \\
p_{Yi}(2) &= P(Y_i = 1|T_i = 2, \mathbf{X}_i = \mathbf{x}_i)
\end{aligned}
$$

## 1.5 Parameter Settings

The parameters were assinged as follows.

### 1.5.1 Treatment Generating Model

All possible combinations of three treatment prevalences and two levels of covariate overlap (inverse of covariate-treatment association) were generated as follows (6 combinations).

| | Treatment Prevalence | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 33:33:33 | | | | 10:45:45 | | | | 10:10:80 | | | |
| | Covariate Overlap | | | | | | | | | | | |
| | Good | | Poor | | Good | | Poor | | Good | | Poor | |
| | $\boldsymbol{\alpha}_1$ | $\boldsymbol{\alpha}_2$ | $\boldsymbol{\alpha}_1$ | $\boldsymbol{\alpha}_2$ | $\boldsymbol{\alpha}_1$ | $\boldsymbol{\alpha}_2$ | $\boldsymbol{\alpha}_1$ | $\boldsymbol{\alpha}_2$ | $\boldsymbol{\alpha}_1$ | $\boldsymbol{\alpha}_2$ | $\boldsymbol{\alpha}_1$ | $\boldsymbol{\alpha}_2$ |
| Intercept | -0.13 | -0.26 | -0.75 | -3.75 | 1.30 | 1.18 | 1.55 | -0.65 | -0.10 | 1.87 | 0.60 | 1.70 |
| $X_1$ | 0.05 | 0.10 | 0.80 | 1.60 | 0.05 | 0.10 | 0.80 | 1.60 | 0.05 | 0.10 | 0.80 | 1.60 |
| $X_2$ | 0.00 | 0.01 | 0.06 | 0.12 | 0.00 | 0.01 | 0.06 | 0.12 | 0.00 | 0.01 | 0.06 | 0.12 |
| $X_3$ | 0.00 | 0.01 | 0.06 | 0.12 | 0.00 | 0.01 | 0.06 | 0.12 | 0.00 | 0.01 | 0.06 | 0.12 |
| $X_4$ | 0.09 | 0.19 | 1.50 | 3.00 | 0.09 | 0.19 | 1.50 | 3.00 | 0.09 | 0.19 | 1.50 | 3.00 |
| $X_5$ | 0.09 | 0.19 | 1.50 | 3.00 | 0.09 | 0.19 | 1.50 | 3.00 | 0.09 | 0.19 | 1.50 | 3.00 |
| $X_6$ | 0.03 | 0.05 | 0.40 | 0.80 | 0.03 | 0.05 | 0.40 | 0.80 | 0.03 | 0.05 | 0.40 | 0.80 |
| $X_7$ | 0.05 | 0.10 | 0.80 | 1.60 | 0.05 | 0.10 | 0.80 | 1.60 | 0.05 | 0.10 | 0.80 | 1.60 |
| $X_8$ | 0.00 | 0.01 | 0.04 | 0.08 | 0.00 | 0.01 | 0.04 | 0.08 | 0.00 | 0.01 | 0.04 | 0.08 |
| $X_9$ | 0.01 | 0.01 | 0.08 | 0.16 | 0.01 | 0.01 | 0.08 | 0.16 | 0.01 | 0.01 | 0.08 | 0.16 |
| $X_{10}$ | 0.06 | 0.12 | 1.00 | 2.00 | 0.06 | 0.12 | 1.00 | 2.00 | 0.06 | 0.12 | 1.00 | 2.00 |

where $\boldsymbol{\alpha}_1 = (\alpha_{10}, \boldsymbol{\alpha}_{1X}^T)^T$ and $\boldsymbol{\alpha}_2 = (\alpha_{20}, \boldsymbol{\alpha}_{2X}^T)^T$.

### 1.5.2 Outcome Generating Model

The outcome generating model parameters were the following.

$$
\begin{aligned}
&\text{Two types of baseline risks}\\
\beta_0 &\in \{\log(0.05), \log(0.20)\}\,,\ \textit{i.e.}\text{, 5\% and 20\% baseline risk}\\
\\
&\text{One type of covariate-outcome association}\\
\boldsymbol{\beta}_X &= (0.160, 0.012, 0.012, 0.300, 0.300, 0.080, 0.160, 0.008, 0.016, 0.200)^T\\
\\
&\text{Null or non-null treatment (main) effects}\\
\boldsymbol{\beta}_T &= (\beta_{T1}, \beta_{T2})^T \in \left\{(0,0)^T, (\log(0.9), \log(0.6))^T\right\}\\
&\text{For the non-null case:}\\
&\text{relative risk of 0.9 comparing Treatment 1 vs 0}\\
&\text{relative risk of 0.6 comparing Treatment 2 vs 0}\\
\Longrightarrow &\text{ relative risk of 6/9 comparing Treatment 2 vs 1}\\
\\
&\text{Null or non-null treatment effect modification}\\
\boldsymbol{\beta}_{XT} &= (\beta_{XT1}, \beta_{XT2})^T \in \left\{(0,0)^T, (\log(0.7), \log(0.5))^T\right\}\\
&\text{For the non-null case:}\\
&\text{additional } 0.7\times \text{ risk reduction among } X_{5i} = 1 \text{ for Treatment 1 vs 0}\\
&\text{additional } 0.5\times \text{ risk reduction among } X_{5i} = 1 \text{ for Treatment 2 vs 0}\\
\Longrightarrow &\text{ additional } 5/7\times \text{ risk reduction among } X_{5i} = 1 \text{ for Treatment 2 vs 1}
\end{aligned}
$$

There are thus, $2 \times 1 \times 2 \times 2 = 8$ combinations of the outcome generating model parameters

### 1.6 Simulation scenarios

There are $6 \times 8 = 48$ total simulation scenarios numbered as follows.

| Scenario | N | Effect modification | Main effects | Baseline risk | Group sizes | Covariate overlap |
|---|---|---|---|---|---|---|
| 1 | 6000 | Modification (-) | Null main effects | 0.05 | 33:33:33 | Good overlap |
| 2 | 6000 | Modification (-) | Null main effects | 0.05 | 33:33:33 | Poor overlap |
| 3 | 6000 | Modification (-) | Null main effects | 0.05 | 10:45:45 | Good overlap |
| 4 | 6000 | Modification (-) | Null main effects | 0.05 | 10:45:45 | Poor overlap |
| 5 | 6000 | Modification (-) | Null main effects | 0.05 | 10:10:80 | Good overlap |
| 6 | 6000 | Modification (-) | Null main effects | 0.05 | 10:10:80 | Poor overlap |
| 7 | 6000 | Modification (-) | Null main effects | 0.2 | 33:33:33 | Good overlap |
| 8 | 6000 | Modification (-) | Null main effects | 0.2 | 33:33:33 | Poor overlap |
| 9 | 6000 | Modification (-) | Null main effects | 0.2 | 10:45:45 | Good overlap |
| 10 | 6000 | Modification (-) | Null main effects | 0.2 | 10:45:45 | Poor overlap |
| 11 | 6000 | Modification (-) | Null main effects | 0.2 | 10:10:80 | Good overlap |
| 12 | 6000 | Modification (-) | Null main effects | 0.2 | 10:10:80 | Poor overlap |
| 13 | 6000 | Modification (-) | Non-null main effects | 0.05 | 33:33:33 | Good overlap |
| 14 | 6000 | Modification (-) | Non-null main effects | 0.05 | 33:33:33 | Poor overlap |
| 15 | 6000 | Modification (-) | Non-null main effects | 0.05 | 10:45:45 | Good overlap |
| 16 | 6000 | Modification (-) | Non-null main effects | 0.05 | 10:45:45 | Poor overlap |
| 17 | 6000 | Modification (-) | Non-null main effects | 0.05 | 10:10:80 | Good overlap |
| 18 | 6000 | Modification (-) | Non-null main effects | 0.05 | 10:10:80 | Poor overlap |
| 19 | 6000 | Modification (-) | Non-null main effects | 0.2 | 33:33:33 | Good overlap |
| 20 | 6000 | Modification (-) | Non-null main effects | 0.2 | 33:33:33 | Poor overlap |
| 21 | 6000 | Modification (-) | Non-null main effects | 0.2 | 10:45:45 | Good overlap |
| 22 | 6000 | Modification (-) | Non-null main effects | 0.2 | 10:45:45 | Poor overlap |
| 23 | 6000 | Modification (-) | Non-null main effects | 0.2 | 10:10:80 | Good overlap |
| 24 | 6000 | Modification (-) | Non-null main effects | 0.2 | 10:10:80 | Poor overlap |
| 25 | 6000 | Modification (+) | Null main effects | 0.05 | 33:33:33 | Good overlap |
| 26 | 6000 | Modification (+) | Null main effects | 0.05 | 33:33:33 | Poor overlap |
| 27 | 6000 | Modification (+) | Null main effects | 0.05 | 10:45:45 | Good overlap |
| 28 | 6000 | Modification (+) | Null main effects | 0.05 | 10:45:45 | Poor overlap |
| 29 | 6000 | Modification (+) | Null main effects | 0.05 | 10:10:80 | Good overlap |
| 30 | 6000 | Modification (+) | Null main effects | 0.05 | 10:10:80 | Poor overlap |
| 31 | 6000 | Modification (+) | Null main effects | 0.2 | 33:33:33 | Good overlap |
| 32 | 6000 | Modification (+) | Null main effects | 0.2 | 33:33:33 | Poor overlap |
| 33 | 6000 | Modification (+) | Null main effects | 0.2 | 10:45:45 | Good overlap |
| 34 | 6000 | Modification (+) | Null main effects | 0.2 | 10:45:45 | Poor overlap |
| 35 | 6000 | Modification (+) | Null main effects | 0.2 | 10:10:80 | Good overlap |
| 36 | 6000 | Modification (+) | Null main effects | 0.2 | 10:10:80 | Poor overlap |
| 37 | 6000 | Modification (+) | Non-null main effects | 0.05 | 33:33:33 | Good overlap |
| 38 | 6000 | Modification (+) | Non-null main effects | 0.05 | 33:33:33 | Poor overlap |
| 39 | 6000 | Modification (+) | Non-null main effects | 0.05 | 10:45:45 | Good overlap |
| 40 | 6000 | Modification (+) | Non-null main effects | 0.05 | 10:45:45 | Poor overlap |
| 41 | 6000 | Modification (+) | Non-null main effects | 0.05 | 10:10:80 | Good overlap |
| 42 | 6000 | Modification (+) | Non-null main effects | 0.05 | 10:10:80 | Poor overlap |
| 43 | 6000 | Modification (+) | Non-null main effects | 0.2 | 33:33:33 | Good overlap |
| 44 | 6000 | Modification (+) | Non-null main effects | 0.2 | 33:33:33 | Poor overlap |
| 45 | 6000 | Modification (+) | Non-null main effects | 0.2 | 10:45:45 | Good overlap |
| 46 | 6000 | Modification (+) | Non-null main effects | 0.2 | 10:45:45 | Poor overlap |
| 47 | 6000 | Modification (+) | Non-null main effects | 0.2 | 10:10:80 | Good overlap |
| 48 | 6000 | Modification (+) | Non-null main effects | 0.2 | 10:10:80 | Poor overlap |

**References**

[1] J. M. Franklin, J. A. Rassen, D. Ackermann, D. B. Bartels, and S. Schneeweiss, "Metrics for covariate balance in cohort studies of causal effects," *Statistics in Medicine*, vol. 33, pp. 1685–1699, May 2014.

[2] A. Linden, S. D. Uysal, A. Ryan, and J. L. Adams, "Estimating causal effects for multivalued treatments: a comparison of approaches," *Statistics in Medicine*, Oct. 2015.

[3] M. D. Cattaneo, D. M. Drukker, and A. D. Holland, "Estimation of multivalued treatment effects under conditional independence," vol. 13, no. 3, pp. 407–450, 2013.

[4] P. Cummings, "The relative merits of risk ratios and odds ratios," *Archives of Pediatrics & Adolescent Medicine*, vol. 163, pp. 438–445, May 2009.

[5] S. Greenland, J. M. Robins, and J. Pearl, "Confounding and Collapsibility in Causal Inference," *Statistical Science*, vol. 14, pp. 29–46, Feb. 1999.

## 1 Aim

This document provides a step-by-step guide for implementation of matching weight method in practice. The example is in the three-group setting. However, the essentially the same code can be used in the two-group setting or settings where there are more than three groups. The example is written in R, but it can be implemented in any statistical environment that has (multinomial) logistic regression and weighted data analysis capabilities.

## 2 Dataset

The `tutoring` dataset included in the `TriMatch` R package is used. The exposure is the `treat` variable, which takes one of `Treat1`, `Treat2`, and `Control`. These represent the tutoring method each student received. The outcome is the `Grade` ordinal variable, which takes one of 0, 1, 2, 3, or 4. Pre-treatment potential confounders include gender, ethnicity, military service status of the student, non-native English speaker status, education level of the subject's mother (ordinal), education level of the subject's father (ordinal), age of the student, employment status (no, part-time, full-time), household income (ordinal), number of transfer credits, grade point average. The dataset does not contain any missing values. See `?tutoring` for details. The employment categorical variable is coded numerically. Thus, it is converted to a factor.

```
## Load data
library(TriMatch)
data(tutoring)
summary(tutoring)

##      treat        Course              Grade         Gender    Ethnicity    Military
##  Control:918   Length:1142        Min.   :0.000   FEMALE:627   Black:211   Mode :logical
##  Treat1 :134   Class :character   1st Qu.:2.000   MALE  :515   Other:193   FALSE:783
##  Treat2 : 90   Mode  :character   Median :4.000                White:738   TRUE :359
##                                   Mean   :2.891                            NA's :0
##                                   3rd Qu.:4.000
##                                   Max.   :4.000
##     ESL             EdMother        EdFather          Age          Employment        Income
##  Mode :logical   Min.   :1.000   Min.   :1.000   Min.   :20.00   Min.   :1.000   Min.   :1.000
##  FALSE:1049      1st Qu.:3.000   1st Qu.:3.000   1st Qu.:30.00   1st Qu.:3.000   1st Qu.:3.000
##  TRUE :93        Median :3.000   Median :3.000   Median :37.00   Median :3.000   Median :5.000
##  NA's :0         Mean   :3.785   Mean   :3.684   Mean   :36.92   Mean   :2.667   Mean   :5.059
##                  3rd Qu.:5.000   3rd Qu.:5.000   3rd Qu.:43.00   3rd Qu.:3.000   3rd Qu.:7.000
##                  Max.   :8.000   Max.   :9.000   Max.   :65.00   Max.   :3.000   Max.   :9.000
##     Transfer          GPA         GradeCode           Level           ID
##  Min.   :  3.00   Min.   :0.000   Length:1142        Lower:988   Min.   :   1.0
##  1st Qu.: 36.66   1st Qu.:2.890   Class :character   Upper:154   1st Qu.: 286.2
##  Median : 48.31   Median :3.215   Mode  :character               Median : 571.5
##  Mean   : 52.12   Mean   :3.166                                  Mean   : 571.5
##  3rd Qu.: 65.00   3rd Qu.:3.518                                  3rd Qu.: 856.8
##  Max.   :126.00   Max.   :4.000                                  Max.   :1142.0

## Make employment categorical
tutoring$Employment <- factor(tutoring$Employment, levels = 1:3,
                              labels = c("no","part-time","full-time"))
```

## 3 Pre-weighting balance assessment

The `tableone` package can be utilized for covariate balance assessment using standardized mean differences (SMD). SMD greater than 0.1 is often regarded as a substantial imbalance. The SMD shown in the table is the average of all possible pairwise SMDs.

```
## Examine covariate balance
library(tableone)
covariates <- c("Gender", "Ethnicity", "Military", "ESL",
                "EdMother", "EdFather", "Age", "Employment",
                "Income", "Transfer", "GPA")
tab1Unadj <- CreateTableOne(vars = covariates, strata = "treat", data = tutoring)
print(tab1Unadj, test = FALSE, smd = TRUE)

##                         Stratified by treat
##                          Control       Treat1        Treat2        SMD
##   n                      918           134            90
##   Gender = MALE (%)      449 (48.9)     38 (28.4)     28 (31.1)   0.287
##   Ethnicity (%)                                                   0.095
##      Black                166 (18.1)     24 (17.9)     21 (23.3)
##      Other                157 (17.1)     23 (17.2)     13 (14.4)
##      White                595 (64.8)     87 (64.9)     56 (62.2)
##   Military = TRUE (%)     309 (33.7)     32 (23.9)     18 (20.0)   0.208
##   ESL = TRUE (%)           76 ( 8.3)      8 ( 6.0)      9 (10.0)   0.100
##   EdMother (mean (sd))  3.80 (1.49)   3.78 (1.51)   3.67 (1.54)   0.057
##   EdFather (mean (sd))  3.68 (1.65)   3.66 (1.73)   3.78 (1.73)   0.044
##   Age (mean (sd))      36.75 (8.95)  37.10 (9.41)  38.41 (9.49)   0.119
##   Employment (%)                                                  0.248
##      no                    95 (10.3)     24 (17.9)     18 (20.0)
##      part-time             75 ( 8.2)     20 (14.9)     11 (12.2)
##      full-time            748 (81.5)     90 (67.2)     61 (67.8)
##   Income (mean (sd))    5.10 (2.24)   5.04 (2.60)   4.69 (2.51)   0.111
##   Transfer (mean (sd)) 51.40 (24.38) 57.37 (25.10) 51.61 (26.39)  0.158
##   GPA (mean (sd))       3.16 (0.58)   3.16 (0.46)   3.24 (0.58)   0.097

## Examine all pairwise SMDs
ExtractSmd(tab1Unadj)

##                average      1 vs 2      1 vs 3     2 vs 3
## Gender     0.28718081 0.431825669 0.369462797 0.06025398
## Ethnicity  0.09475231 0.004540496 0.137619463 0.14209699
## Military   0.20773590 0.217301900 0.312032587 0.09387322
## ESL        0.09955894 0.089842245 0.059753148 0.14908142
## EdMother   0.05735067 0.014182489 0.086066827 0.07180268
## EdFather   0.04433253 0.007919139 0.059274560 0.06580389
## Age        0.11889003 0.038429226 0.179969129 0.13827175
## Employment 0.24838203 0.332479394 0.324590337 0.08807636
## Income     0.11113230 0.025003114 0.171951403 0.13644238
## Transfer   0.15777889 0.241327245 0.008454888 0.22355453
## GPA        0.09651297 0.009213587 0.128230886 0.15209444
```

## 4 Propensity score modeling

As the exposure is a three-category variable, the propensity score model can be modeled using multinomial logistic regression. In R, the VGAM (vector generalized linear and additive models) package provides a flexible framework for this. Because the sample size of the treatment 2 group is small, making flexible modeling difficult, the ordinal variables are used only as linear terms. Predicting the “response” gives predicted probabilities of each treatment as a (sample size) × 3 matrix, which then can be added to the dataset. The following AddGPS function can be used to ease this process. Three propensity scores (one for each treatment category) are added to the dataset.

```
## Function to add generalized PS to dataset
AddGPS <- function(data, formula, family = multinomial(), psPrefix = "PS_") {
    library(VGAM)
    ## Fit multinomial logistic regression
    resVglm <- vglm(formula = formula, data = data, family = family)
    ## Calculate PS
    psData <- as.data.frame(predict(resVglm, type = "response"))
    names(psData) <- paste0(psPrefix, names(psData))
    cbind(data, psData)
}

tutoring <- AddGPS(data = tutoring, # dataset
                   ## Propensity score model for multinomial regression
                   formula = treat ~ Gender + Ethnicity + Military +
                                 ESL + EdMother + EdFather + Age +
                                 Employment + Income + Transfer + GPA)
```

## 5 Weight creation

As mentioned in the text, the matching weight is defined as follows.

$$
\begin{aligned}
MW_i &= \frac{\text{Smallest PS}}{\text{PS of assigned treatment}} \\
&= \frac{\min(e_{1i}, ..., e_{Ki})}{\sum_{k=1}^{K} I(Z_i = k) e_{ki}}
\end{aligned}
$$

where $e_{ki}$ is the $i$-th individual's probability of being assigned to the $k$-th treatment category given the covariate pattern, $Z_i \in \{1, ..., K\}$ is the categorical variable indicating the $i$-th individual's treatment assignment.

The following function can be used to add matching weight to the dataset. Individuals' matching weights have a range of [0,1], where as the inverse probability treatment weights have a range of [1,∞].

```
## Function to add matching weight as mw to dataset
AddMwToData <- function(data, txVar, txLevels, psPrefix = "PS_") {
    ## Treatment indicator data frame (any number of groups allowed)
    dfAssign <- as.data.frame(lapply(txLevels, function(tx_k) {
        as.numeric(data[txVar] == tx_k)
    }))
    ## Name of PS variables
    psVars <- paste0(psPrefix, txLevels)
    ## Pick denominator (PS for assigned treatment)
    data$PS_assign <- rowSums(data[psVars] * dfAssign)
    ## Pick numerator
    data$PS_min <- do.call(pmin, data[psVars])
    ## Calculate the IPTW
    data$iptw <- 1 / data$PS_assign
    ## Calculate the matching weight
    data$mw <- exp(log(data$PS_min) - log(data$PS_assign))
    ## Return the whole data
    data
}

## Add IPTW and MW
tutoring <- AddMwToData(data = tutoring, # dataset
                        txVar = "treat", # treatment variable name
                        tx = c("Control", "Treat1", "Treat2")) # treatment levels

## Check how weights are defined
head(tutoring[c("treat","PS_Control","PS_Treat1","PS_Treat2","PS_assign","PS_min","iptw","mw")], 20)

##      treat PS_Control  PS_Treat1  PS_Treat2  PS_assign     PS_min      iptw         mw
## 3  Control  0.8192816 0.11440448 0.06631388 0.81928164 0.06631388  1.220581 0.08094149
## 4  Control  0.8313205 0.10516348 0.06351606 0.83132046 0.06351606  1.202906 0.07640383
## 11 Control  0.6346235 0.22597339 0.13940309 0.63462352 0.13940309  1.575737 0.21966266
## 12 Control  0.7203265 0.11853269 0.16114082 0.72032649 0.11853269  1.388259 0.16455412
## 14 Control  0.6759314 0.15931947 0.16474916 0.67593137 0.15931947  1.479440 0.23570361
## 16  Treat1  0.7278386 0.18054526 0.09161616 0.18054526 0.09161616  5.538777 0.50744155
## 17 Control  0.7963014 0.09228518 0.11141339 0.79630143 0.09228518  1.255806 0.11589227
## 18 Control  0.7963014 0.09228518 0.11141339 0.79630143 0.09228518  1.255806 0.11589227
## 19 Control  0.4011609 0.29293705 0.30590201 0.40116094 0.29293705  2.492765 0.73022327
## 23 Control  0.7980564 0.14170696 0.06023666 0.79805638 0.06023666  1.253044 0.07547920
## 28  Treat2  0.7696177 0.11208565 0.11829667 0.11829667 0.11208565  8.453323 0.94749620
## 31  Treat1  0.7876534 0.11912070 0.09322587 0.11912070 0.09322587  8.394847 0.78261688
## 32 Control  0.7602112 0.13218394 0.10760486 0.76021120 0.10760486  1.315424 0.14154600
## 34  Treat2  0.6994628 0.12694918 0.17358797 0.17358797 0.12694918  5.760768 0.73132478
## 38  Treat1  0.6359332 0.24401948 0.12004734 0.24401948 0.12004734  4.098034 0.49195804
## 39 Control  0.7523881 0.15006473 0.09754713 0.75238814 0.09754713  1.329101 0.12965001
## 40 Control  0.8281320 0.11921012 0.05265789 0.82813199 0.05265789  1.207537 0.06358635
## 49  Treat1  0.7963180 0.09950924 0.10417277 0.09950924 0.09950924 10.049318 1.00000000
## 50 Control  0.8929612 0.06199434 0.04504442 0.89296124 0.04504442  1.119869 0.05044387
## 51 Control  0.6910650 0.16455995 0.14437500 0.69106505 0.14437500  1.447042 0.20891666

## Check weight distribution
summary(tutoring[c("mw","iptw")])

##        mw                iptw
##  Min.   :0.01025   Min.   : 1.052
##  1st Qu.:0.05546   1st Qu.: 1.154
##  Median :0.09410   Median : 1.258
##  Mean   :0.21706   Mean   : 3.066
##  3rd Qu.:0.17721   3rd Qu.: 1.465
##  Max.   :1.00000   Max.   :46.446
```

### 6 Post-weighting balance assessment

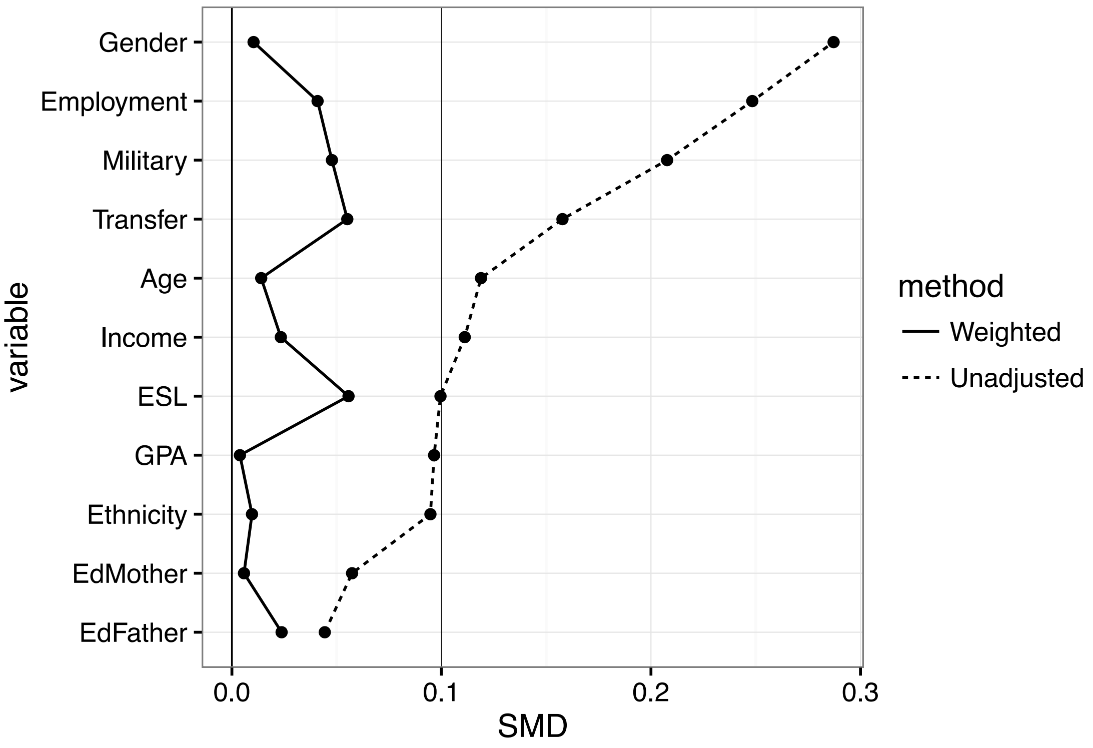
All analyses afterward should be proceeded as weighted analyses. In R, this is most easily achieved by using the survey package. Firstly, a survey design object must be created with svydesign function. The resulting object is then used as the dataset. The weighted covariate table can be constructed with the tableone package. All SMDs are less than 0.1 after weighting, indicating better covariate balance.

```
## Created weighted data object
library(survey)
tutoringSvy <- svydesign(ids = ~ 1, data = tutoring, weights = ~ mw)

## Weighted table with tableone
tab1Mw <- svyCreateTableOne(vars = covariates, strata = "treat", data = tutoringSvy)
print(tab1Mw, test = FALSE, smd = TRUE)

##                         Stratified by treat
##                          Control       Treat1        Treat2        SMD
##   n                        82.8          82.6          82.5
##   Gender = MALE (%)        24.9 (30.1)   25.0 (30.3)   24.4 (29.6)   0.010
##   Ethnicity (%)                                                    0.010
##      Black                 18.9 (22.9)   19.2 (23.3)   18.8 (22.8)
##      Other                 11.7 (14.1)   11.3 (13.7)   11.6 (14.1)
##      White                 52.2 (63.0)   52.1 (63.1)   52.0 (63.0)
##   Military = TRUE (%)      17.2 (20.8)   19.7 (23.8)   17.4 (21.1)   0.048
##   ESL = TRUE (%)            6.1 ( 7.4)    6.4 ( 7.7)    8.1 ( 9.8)   0.056
##   EdMother (mean (sd))     3.66 (1.49)   3.65 (1.47)   3.65 (1.55)   0.006
##   EdFather (mean (sd))     3.71 (1.70)   3.66 (1.75)   3.73 (1.70)   0.024
##   Age (mean (sd))         38.13 (9.68)  38.21 (9.63)  38.01 (9.38)   0.014
##   Employment (%)                                                   0.041
##      no                    16.3 (19.7)   15.6 (18.9)   15.2 (18.4)
##      part-time             10.2 (12.3)    9.2 (11.2)   10.5 (12.7)
##      full-time             56.3 (68.0)   57.7 (69.9)   56.8 (68.9)
##   Income (mean (sd))       4.76 (2.35)   4.72 (2.47)   4.80 (2.47)   0.023
##   Transfer (mean (sd)) 52.46 (24.04) 51.39 (25.02) 53.48 (26.19)  0.055
##   GPA (mean (sd))          3.21 (0.49)   3.21 (0.45)   3.21 (0.59)   0.004

## All pairwise SMDs
ExtractSmd(tab1Mw)

##                 average      1 vs 2       1 vs 3       2 vs 3
## Gender     0.010336859 0.004393687 0.0111115330 0.0155053556
## Ethnicity  0.009595945 0.013881066 0.0006174629 0.0142893048
## Military   0.047738733 0.071609306 0.0067821033 0.0648247896
## ESL        0.055666107 0.010019487 0.0834804231 0.0734984115
## EdMother   0.005765913 0.008755059 0.0082762793 0.0002663992
## EdFather   0.023721214 0.024874520 0.0107632204 0.0355259006
## Age        0.013982735 0.008033386 0.0128645704 0.0210502478
## Employment 0.040896810 0.043102022 0.0330741322 0.0465142771
## Income     0.023351441 0.019691181 0.0157469189 0.0346162234
## Transfer   0.055073782 0.043293809 0.0406028456 0.0813246930
## GPA        0.003834104 0.006104611 0.0018132523 0.0035844491
```

Visualizing the covariate balance before and after weighting can sometimes be helpful. Extracted SMD data can be fed to a plotting function (here ggplot2).

```
## Create SMD data frame
dataPlot <- data.frame(variable   = rownames(ExtractSmd(tab1Unadj)),
                       Unadjusted = ExtractSmd(tab1Unadj)[,"average"],
                       Weighted   = ExtractSmd(tab1Mw)[,"average"])
## Reshape to long format
library(reshape2)
dataPlotMelt <- melt(data          = dataPlot,
                     id.vars       = "variable",
                     variable.name = "method",
                     value.name    = "SMD")
## Variables names ordered by unadjusted SMD values
varsOrderedBySmd <- rownames(dataPlot)[order(dataPlot[,"Unadjusted"])]
## Reorder factor levels
dataPlotMelt$variable <- factor(dataPlotMelt$variable,
                                levels = varsOrderedBySmd)
```

```
dataPlotMelt$method <- factor(dataPlotMelt$method,
                              levels = c("Weighted","Unadjusted"))
## Plot
library(ggplot2)
ggplot(data = dataPlotMelt, mapping = aes(x = variable, y = SMD, group = method, linetype = method)) +
    geom_line() +
    geom_point() +
    geom_hline(yintercept = 0, size = 0.3) +
    geom_hline(yintercept = 0.1, size = 0.1) +
    coord_flip() +
    theme_bw() + theme(legend.key = element_blank())
```



## 7 Outcome analysis

The outcome analyses should also be proceeded as weighted analyses. Most functions in the survey package is named svy⋆ with ⋆ being the name of the unweighted counterpart.

The outcome was handled as a continuous outcome for simplicity. In weighted linear regression, both treatments appear superior to the control without tutoring regarding the course grade assuming the propensity score model was correctly specified. The mean difference was 0.45 [0.23, 0.67] for treatment 1 vs control and 0.67 [0.45, 0.89] for treatment 2 vs control.

```
## Weighted group means of Grade
svyby(formula = ~ Grade, by = ~ treat, design = tutoringSvy, FUN = svymean)

##           treat    Grade         se
## Control Control 2.792759 0.06648740
## Treat1   Treat1 3.244832 0.09179853
## Treat2   Treat2 3.463329 0.09070431

## Group difference tested in weighted regression
modelOutcome1 <- svyglm(formula = Grade ~ treat, design = tutoringSvy)
summary(modelOutcome1)

##
## Call:
## svyglm(formula = Grade ~ treat, design = tutoringSvy)
##
```

```
## Survey design:
## svydesign(ids = ~1, data = tutoring, weights = ~mw)
##
## Coefficients:
##              Estimate Std. Error t value      Pr(>|t|)
## (Intercept)  2.79276    0.06649  42.004       < 2e-16 ***
## treatTreat1  0.45207    0.11335   3.988 0.00007076303 ***
## treatTreat2  0.67057    0.11246   5.963 0.00000000331 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 1.394533)
##
## Number of Fisher Scoring iterations: 2

## ShowRegTable in tableone may come in handy
ShowRegTable(modelOutcome1, exp = FALSE)

##              coef [confint]    p
## (Intercept) 2.79 [2.66, 2.92] <0.001
## treatTreat1 0.45 [0.23, 0.67] <0.001
## treatTreat2 0.67 [0.45, 0.89] <0.001
```

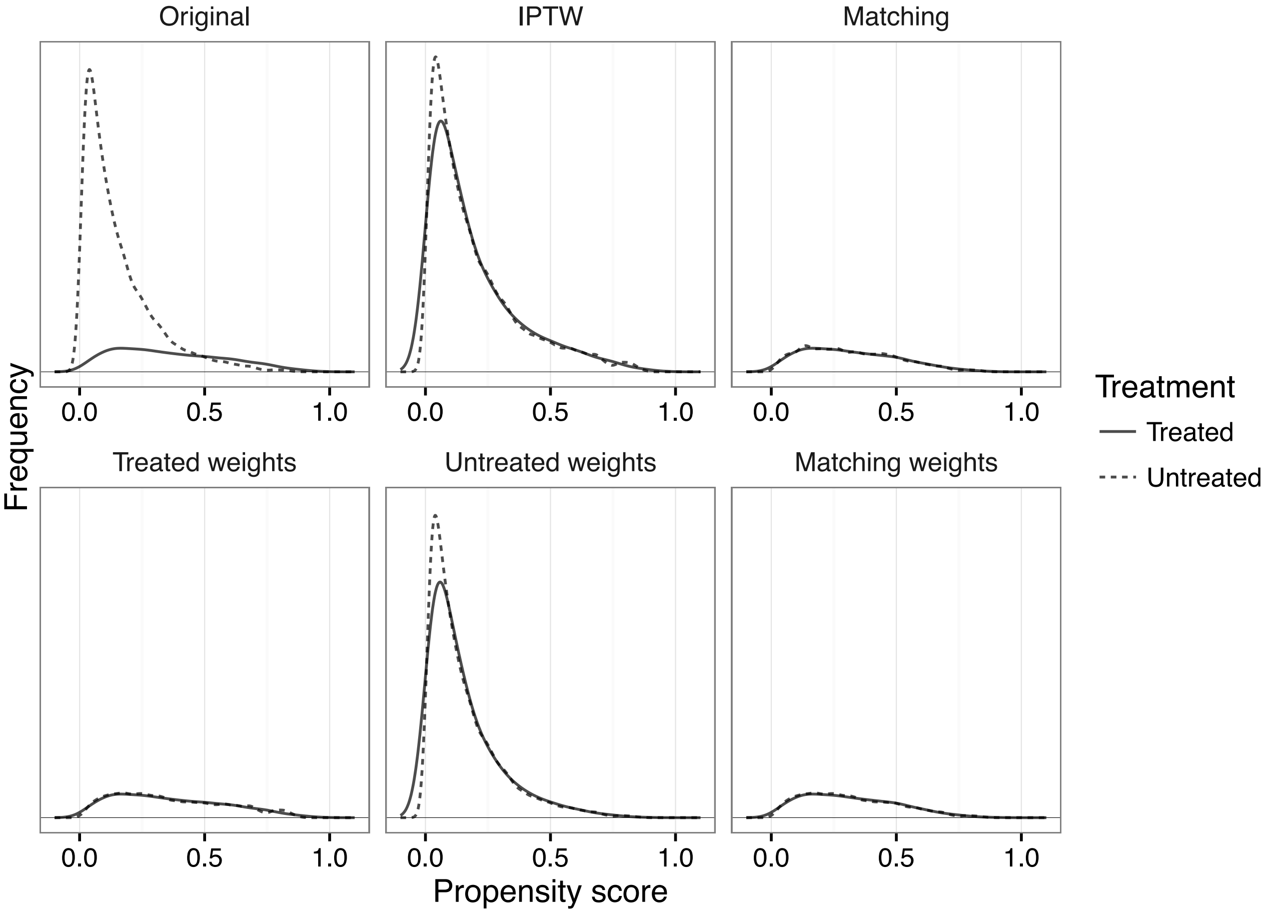
## 8 Bootstrapping

As discussed in the text, bootstrapping may provide better variance estimates than model-based inference. The boot package is a general purpose bootstrapping package. The following context-specific wrapper functions can be used to simplify the process. In this specific example, the bootstrap confidence intervals for the treatment effects were somewhat narrower.

```
## Define a function for each bootstrap step
BootModelsConstructor <- function(formulaPs, formulaOutcome, OutcomeRegFun, ...) {
    ## Obtain treatment variable name
    txVar <- as.character(formulaPs[[2]])
    ## Return a function
    function(data, i) {
        ## Obtain treatment levels
        txLevels <- names(table(data[,txVar]))
        ## Add generalized propensity scores
        dataB <- AddGPS(data = data[i,], formula = formulaPs)
        ## Add matching weight
        dataB <- AddMwToData(data = dataB, txVar = txVar, txLevels = txLevels)
        ## Weighted analysis (lm() ok as only the estimates are used)
        lmWeighted <- OutcomeRegFun(formula = formulaOutcome, data = dataB,
                                    weights = mw, ...)
        ## Extract coefs
        coef(lmWeighted)
    }
}

## Define a function to summarize bootstrapping
BootSummarize <- function(data, R, BootModels, level = 0.95, ...) {
    ## Use boot library
    library(boot)
    ## Run bootstrapping
    outBoot       <- boot(data = data, statistic = BootModels, R = R, ...)
    out           <- outBoot$t
    colnames(out) <- names(outBoot$t0)
    ## Confidence intervals
    lower <- apply(out, MARGIN = 2, quantile, probs = (1 - level) / 2)
    upper <- apply(out, MARGIN = 2, quantile, probs = (1 - level) / 2 + level)
    outCi <- cbind(lower = lower, upper = upper)
    ## Variance of estimator
    outVar <- apply(out, MARGIN = 2, var)
    outSe  <- sqrt(outVar)
    ## Return as a readable table
    cbind(est = outBoot$t0, outCi, var = outVar, se = outSe)
}
```

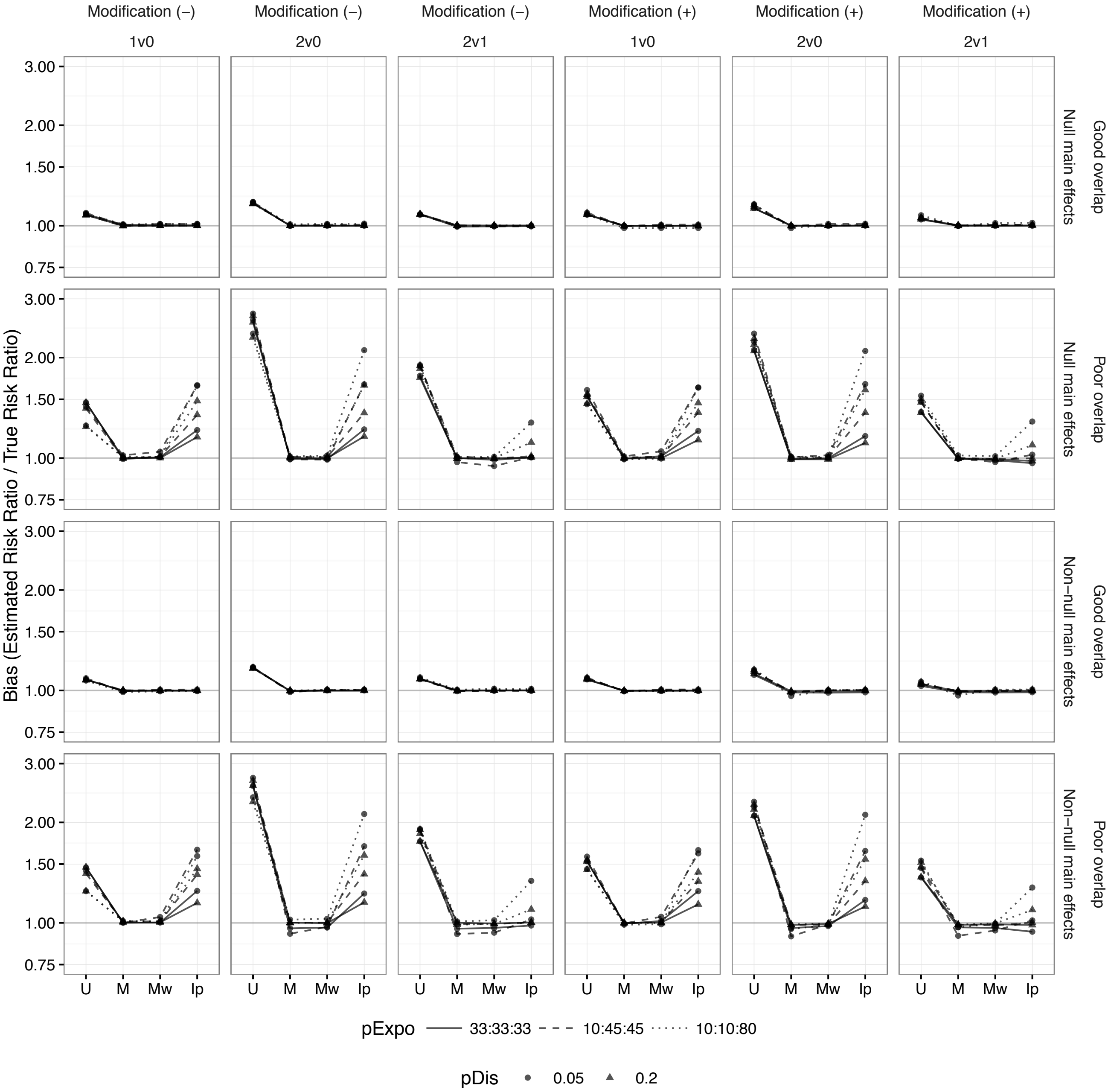
```
## Construct a custom bootstrap function with specific formulae
## formulaPs is propensity score model
BootModels <- BootModelsConstructor(formulaPs = treat ˜ Gender + Ethnicity + Military +
                                                ESL + EdMother + EdFather + Age +
                                                Employment + Income + Transfer + GPA,
                                    ## Outcome model
                                    formulaOutcome = Grade ˜ treat,
                                    ## Regression function for outcome model
                                    OutcomeRegFun = lm)

## Use a clean dataset without PS and weight variables
data(tutoring)
## Make employment categorical
tutoring$Employment <- factor(tutoring$Employment, levels = 1:3,
                              labels = c("no","part-time","full-time"))
## Run bootstrap
set.seed(201508131)
system.time(bootOut1 <- BootSummarize(data = tutoring, R = 2000, BootModels = BootModels))

##    user  system elapsed
## 191.394   5.377 208.653

bootOut1

##                   est     lower     upper         var         se
## (Intercept) 2.7927593 2.6130814 2.9872607 0.008972568 0.09472364
## treatTreat1 0.4520730 0.2325361 0.6577786 0.011831058 0.10877067
## treatTreat2 0.6705692 0.4626595 0.8484488 0.009776627 0.09887683

## Show naive confidence interval again
ShowRegTable(modelOutcome1, exp = FALSE, digits = 7)

##             coef [confint]                    p
## (Intercept) 2.7927593 [2.6624464, 2.9230722] <0.001
## treatTreat1 0.4520730 [0.2299169, 0.6742290] <0.001
## treatTreat2 0.6705692 [0.4501465, 0.8909920] <0.001
```

**eFigure 1.** Illustration of pre- and post-weighting or post-matching distributions of propensity score when the treatment prevalence is 20%. The solid line is the distribution of the propensity scores in the treated, and the dashed line is the distribution in the untreated. Matching and matching weight cohorts have a similar propensity score distribution, indicating that their estimands are similar. These cohorts are very similar to the original treated group (*i.e.*, their estimands approximate the average treatment effect on the treated) although there is a minor attrition in the cohort in the high propensity score range (propensity score > 0.5).



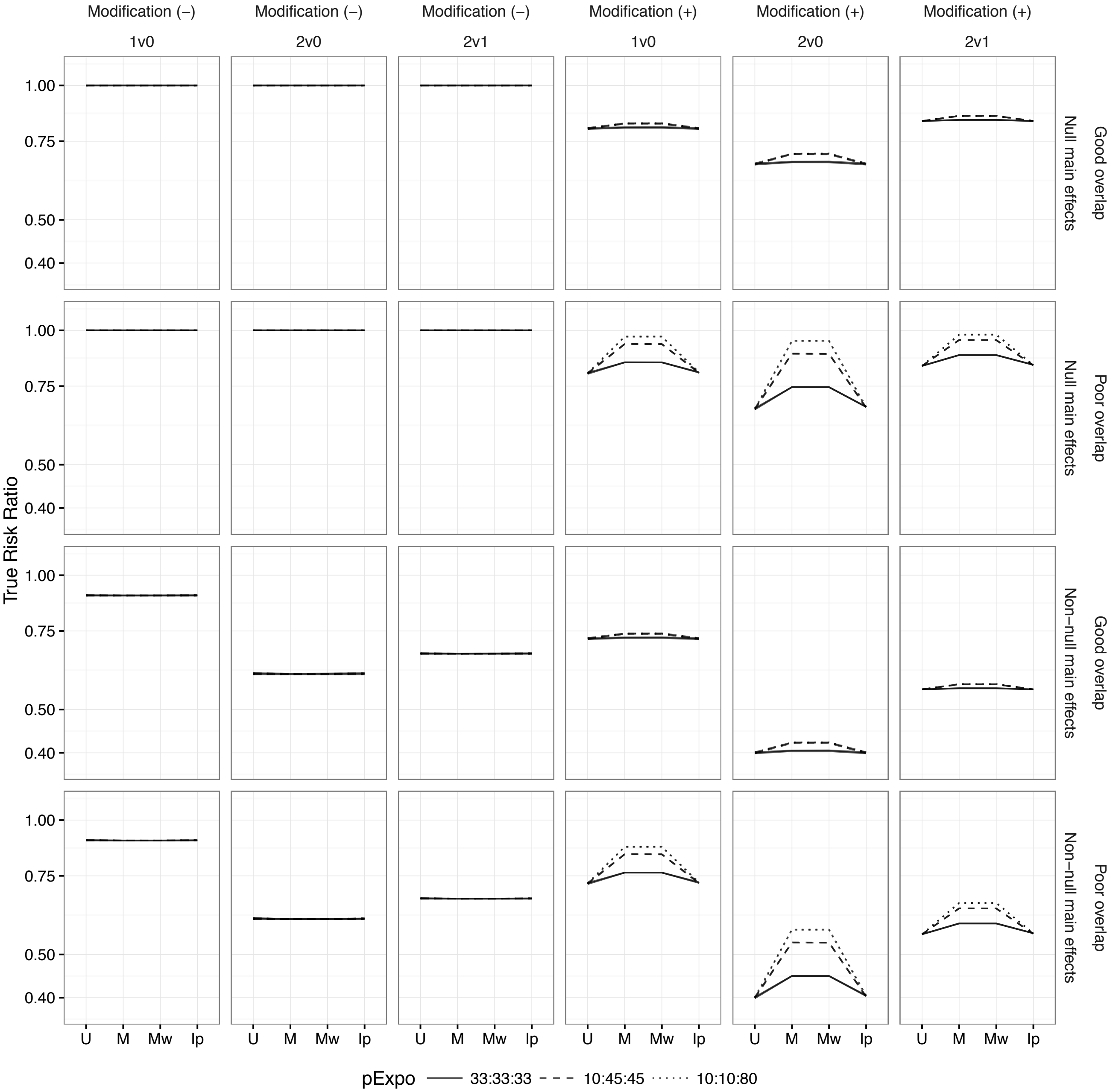
**Abbreviations:** IPTW: inverse probability of treatment weights.

**eFigure 2.** Comparison of bias (risk ratio / true risk ratio) between methods across 48 scenarios. The left half presents the constant treatment effect scenarios, whereas the right half presents treatment effect heterogeneity scenarios. Each three columns represent three treatment contrasts. Rows classify scenarios by good vs. poor covariate overlap levels and presence vs. absence of main effects. Each panel contains six lines classified by the exposure prevalence and the baseline risk. Matching weights and matching perform well in all scenarios; however, IPTW fails in the poor covariate overlap setting.



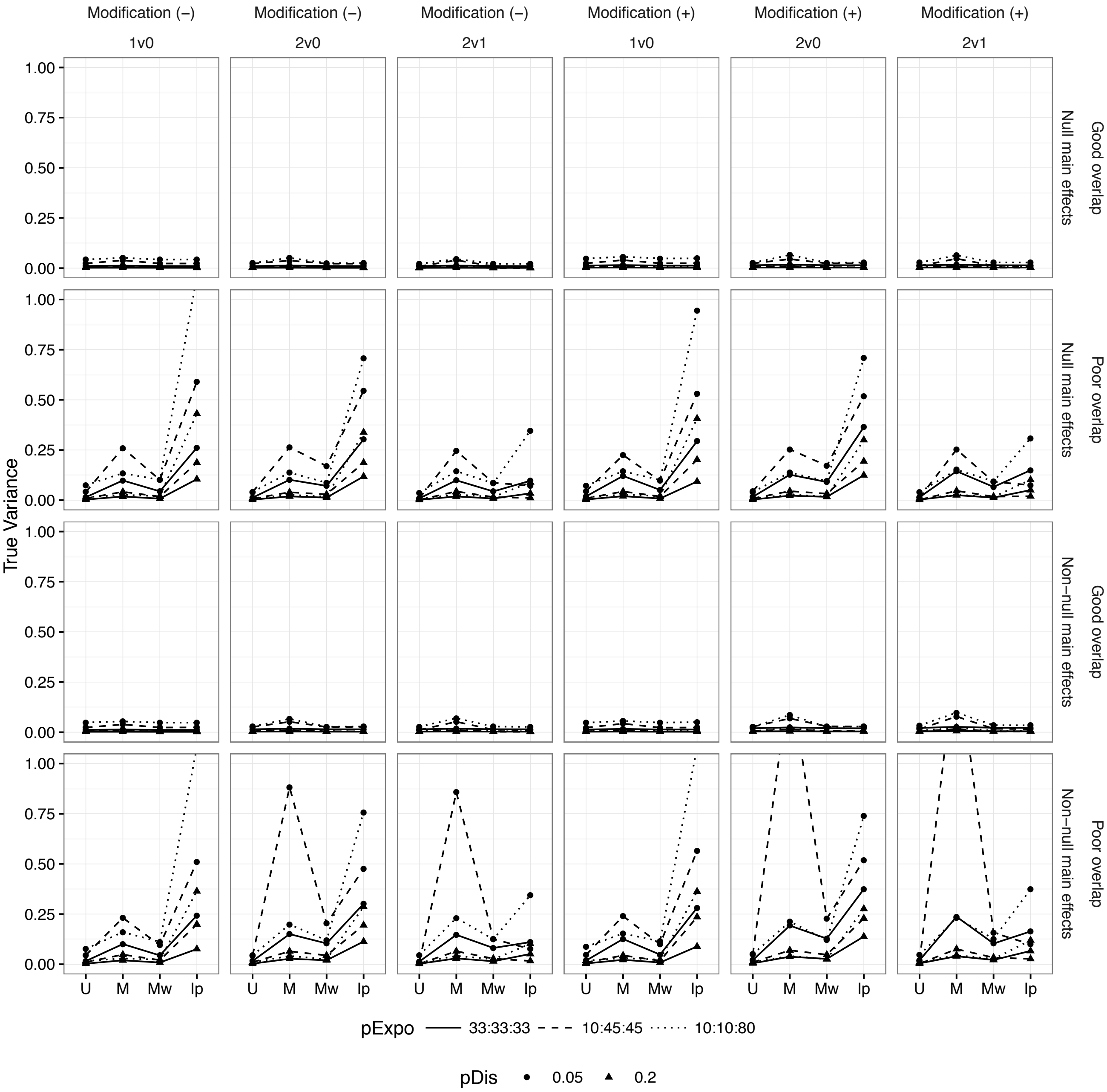
**Abbreviations:** U: Unmatched cohort, M: Matched cohort; Mw: matching weight cohort; Ip: Inverse probability of treatment weight cohort; pExpo: Exposure prevalence; pDis: Baseline risk of disease.

**eFigure 3.** Comparison of true risk ratios (estimands) between methods across 48 scenarios. Some scenarios have the same estimands and completely overlap. The left half presents the constant treatment effect scenarios, whereas the right half presents treatment effect heterogeneity scenarios. Each three columns represent three treatment contrasts. Rows classify scenarios by good vs. poor covariate overlap levels and presence vs. absence of main effects. Differences in estimands are only present in the treatment effect heterogeneity scenarios, particularly with poor covariate overlap and unbalanced treatment group sizes.



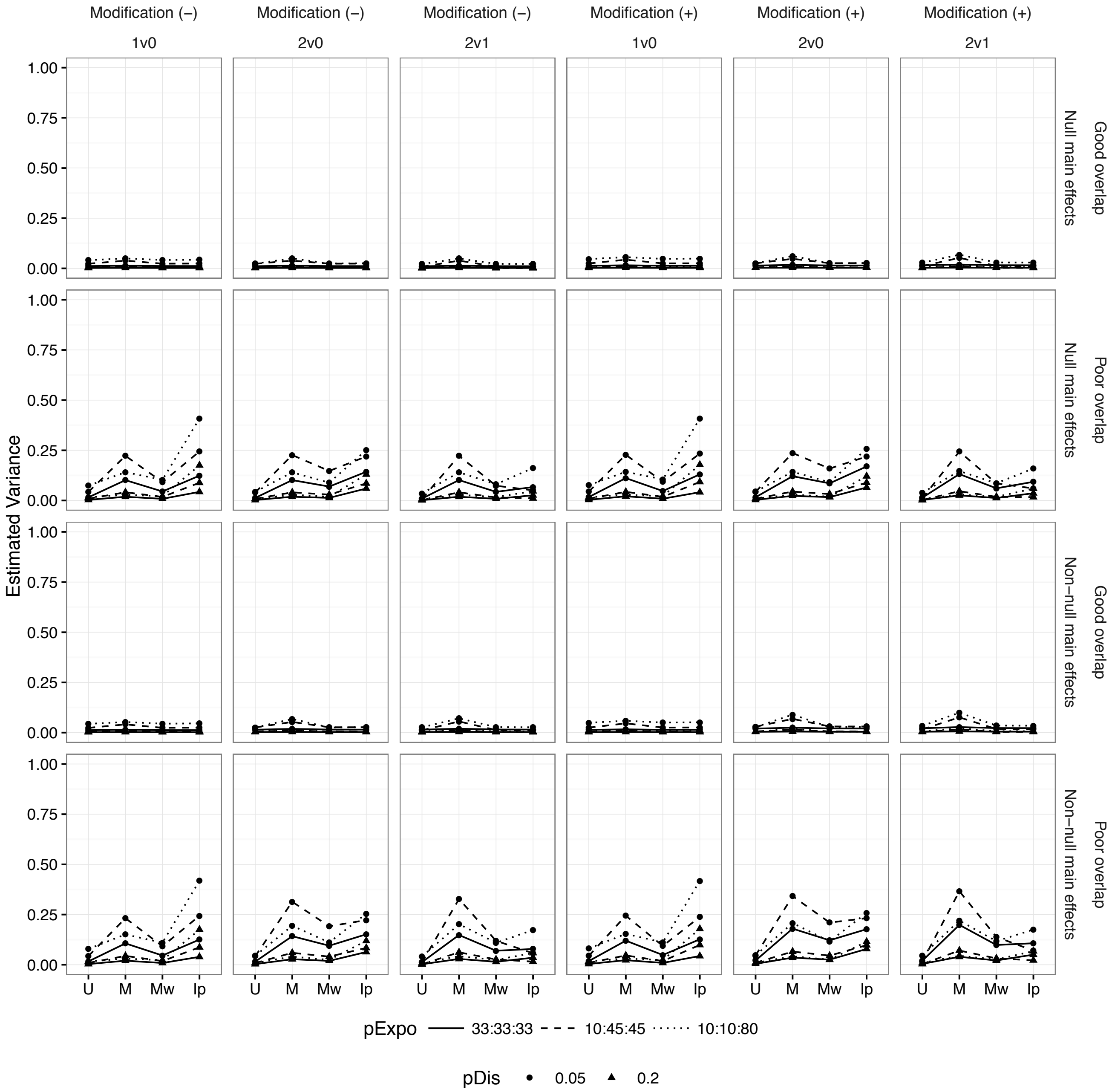
**Abbreviations:** U: Unmatched cohort, M: Matched cohort; Mw: matching weight cohort; Ip: Inverse probability of treatment weight cohort; pExpo: Exposure prevalence.

**eFigure 4.** Comparison of true variance of log risk ratios calculated across iterations between methods across 48 scenarios. The left half presents the constant treatment effect scenarios, whereas the right half presents treatment effect heterogeneity scenarios. Each three columns represent three treatment contrasts. Rows classify scenarios by good vs. poor covariate overlap levels and presence vs. absence of main effects. Each panel contains six lines classified by the exposure prevalence and the baseline risk. All methods performed well in the good covariate overlap scenarios; however, matching weights were most efficient in the poor covariate overlap scenarios (rows 2 and 4). Matching performed poorly in the poor covariate overlap with 10:45:45 exposure distribution, as there were often no events in Group 2 after matching.



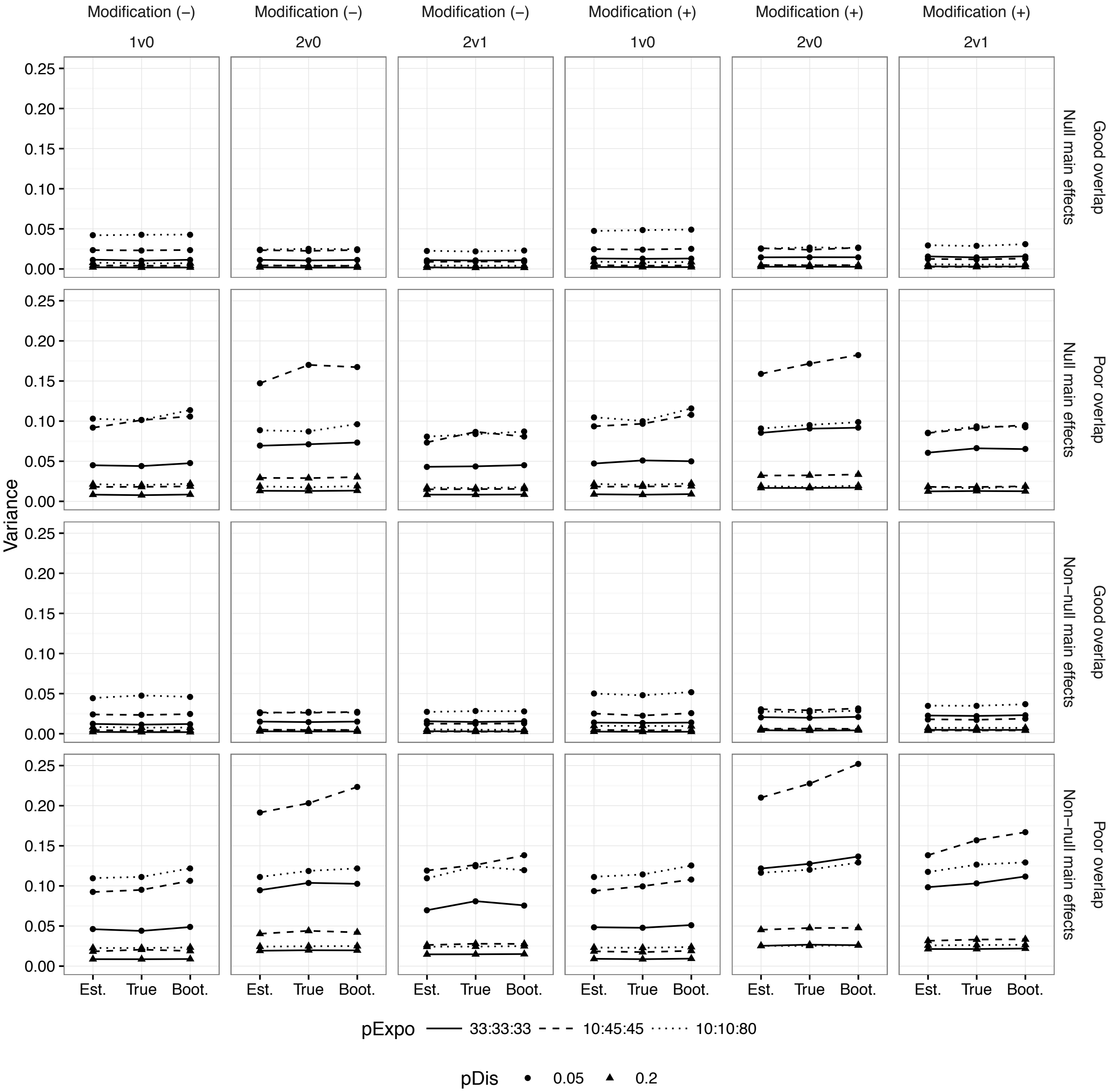
**Abbreviations:** U: Unmatched cohort, M: Matched cohort; Mw: matching weight cohort; Ip: Inverse probability of treatment weight cohort; pExpo: Exposure prevalence; pDis: baseline risk of disease

**eFigure 5.** Comparison of estimated variance of log risk ratios averaged across iterations between methods across 48 scenarios. The left half presents the constant treatment effect scenarios, whereas the right half presents treatment effect heterogeneity scenarios. Each three columns represent three treatment contrasts. Rows classify scenarios by good vs. poor covariate overlap levels and presence vs. absence of main effects. Each panel contains six lines classified by the exposure prevalence and the baseline risk. Results were similar to the true variance results.



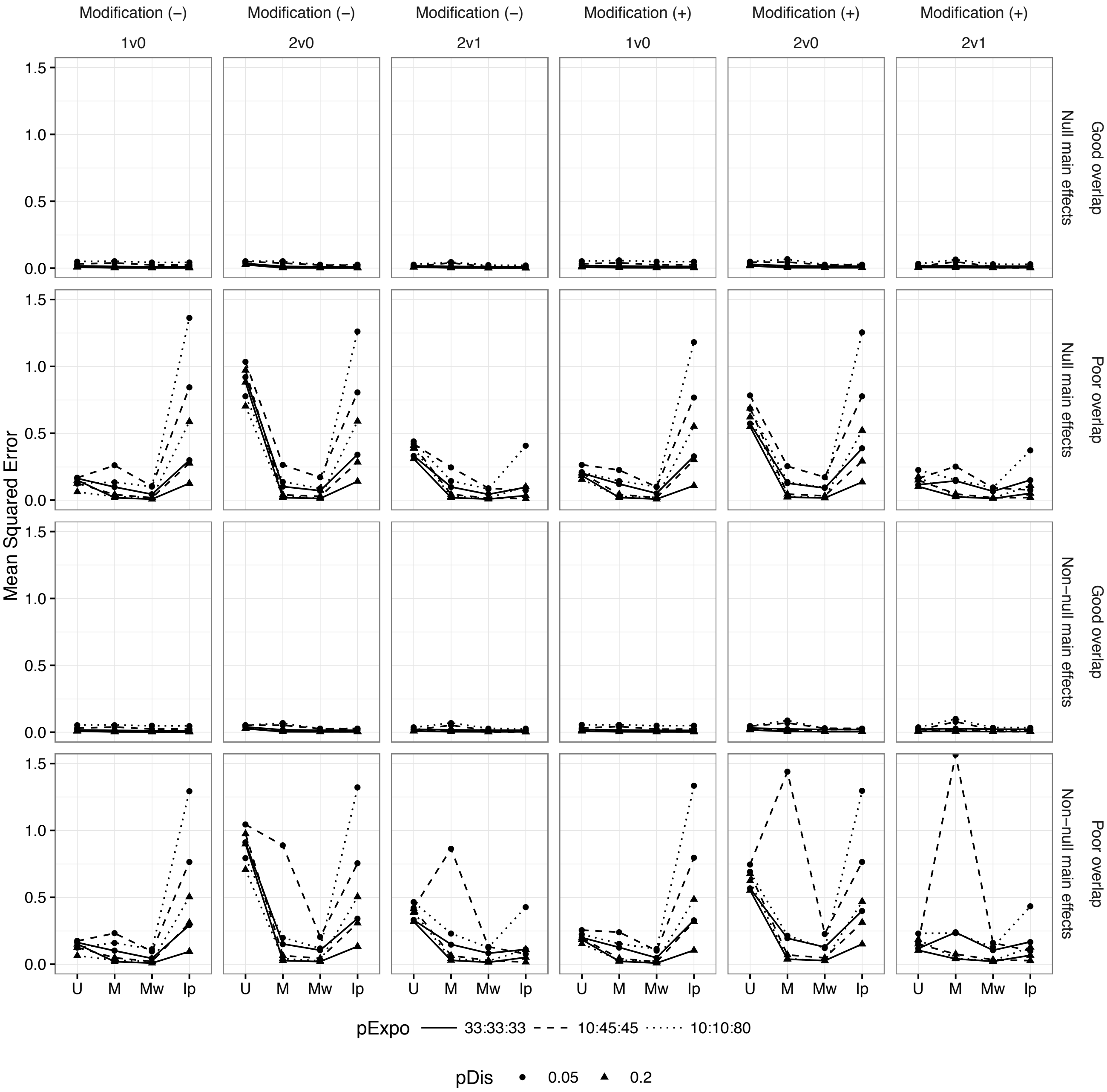
**Abbreviations:** U: Unmatched cohort, M: Matched cohort; Mw: matching weight cohort; Ip: Inverse probability of treatment weight cohort; pExpo: Exposure prevalence; pDis: baseline risk of disease

**eFigure 6.** Comparison of variance estimation methods for matching weights across 48 scenarios. The left half presents the constant treatment effect scenarios, whereas the right half presents treatment effect heterogeneity scenarios. Each three columns represent three treatment contrasts. Rows classify scenarios by good vs. poor covariate overlap levels and presence vs. absence of main effects. Each panel contains six lines classified by the exposure prevalence and the baseline risk. In good covariate overlap settings, the estimated variance and the bootstrap variance were both close to the true variance values. In the poor covariate overlap settings, however, the estimated variance was sometimes anti-conservative, whereas the bootstrap variance was more accurate or somewhat conservative.



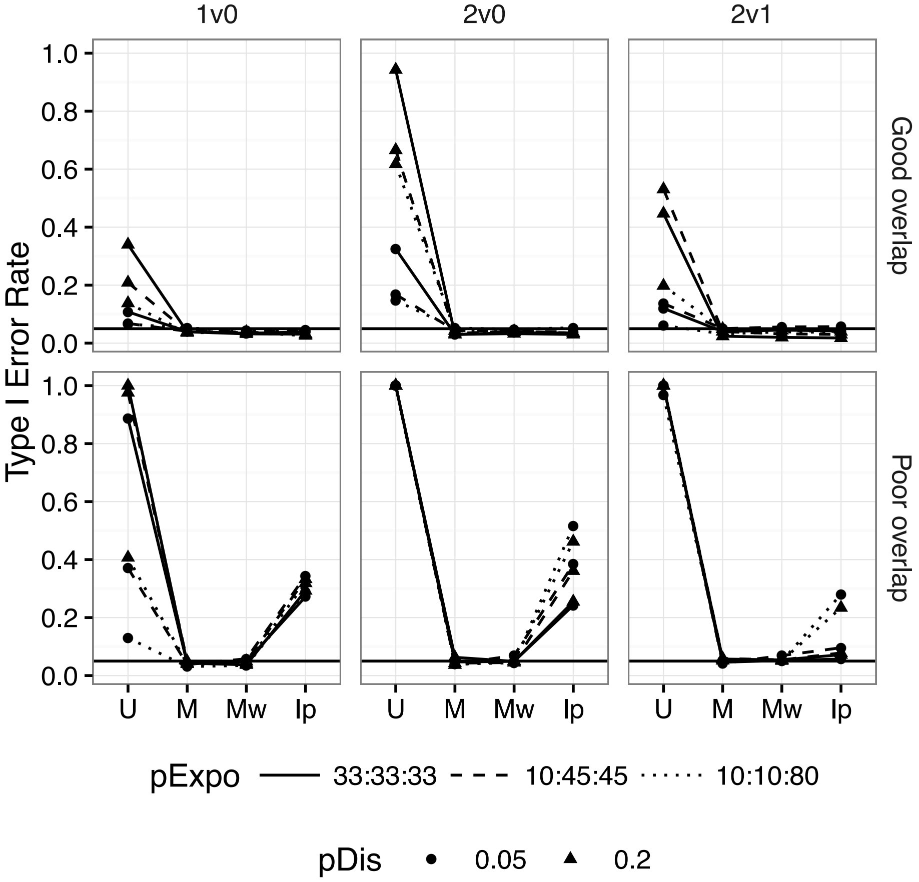
**Abbreviations:** Est.: Estimated variance; True: True variance calculated across iterations; Boot.: Bootstrap variance; pExpo: Exposure prevalence; pDis: baseline risk of disease

**eFigure 7.** Comparison of mean squared error of log risk ratios between methods across 48 scenarios. The left half presents the constant treatment effect scenarios, whereas the right half presents treatment effect heterogeneity scenarios. Each three columns represent three treatment contrasts. Rows classify scenarios by good vs. poor covariate overlap levels and presence vs. absence of main effects. Each panel contains six lines classified by the exposure prevalence and the baseline risk. All methods performed well in the good covariate overlap scenarios; however, matching weights were most robust in the poor covariate overlap scenarios (rows 2 and 4). Matching performed poorly in the poor covariate overlap with 10:45:45 exposure distribution, as there were often no events in Group 2 after matching.



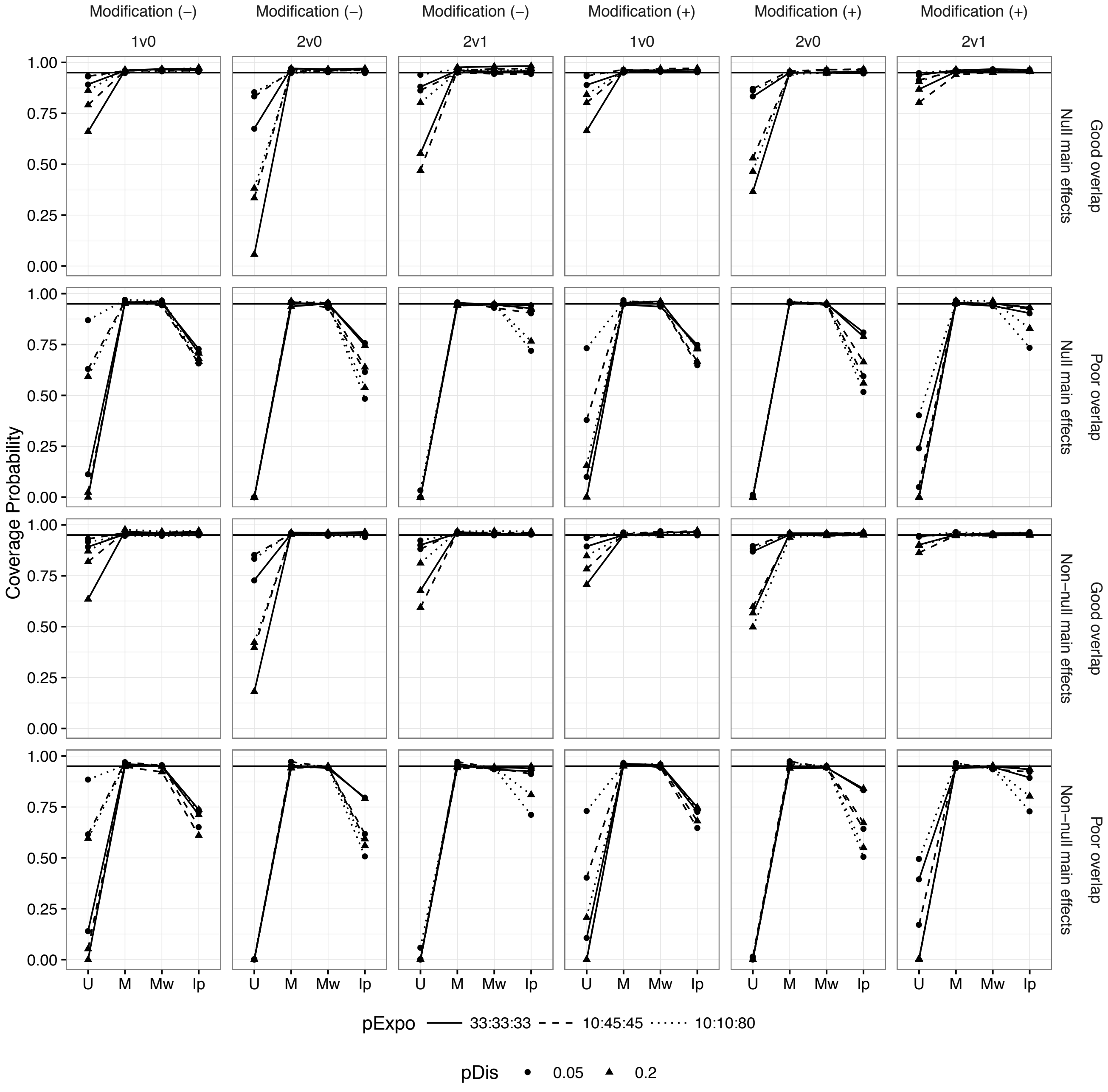
**Abbreviations:** U: Unmatched cohort, M: Matched cohort; Mw: matching weight cohort; Ip: Inverse probability of treatment weight cohort; pExpo: Exposure prevalence; pDis: Baseline risk of disease.

**eFigure 8.** Comparison of false positive probability in completely null treatment effect scenarios. Minor violation of the 0.05 expected false positive rate (false positive rates of 0.06-0.07) was seen in both matching weights and matching. IPTW made many false positives in the poor covariate overlap settings. These tests were based on the estimated variance.



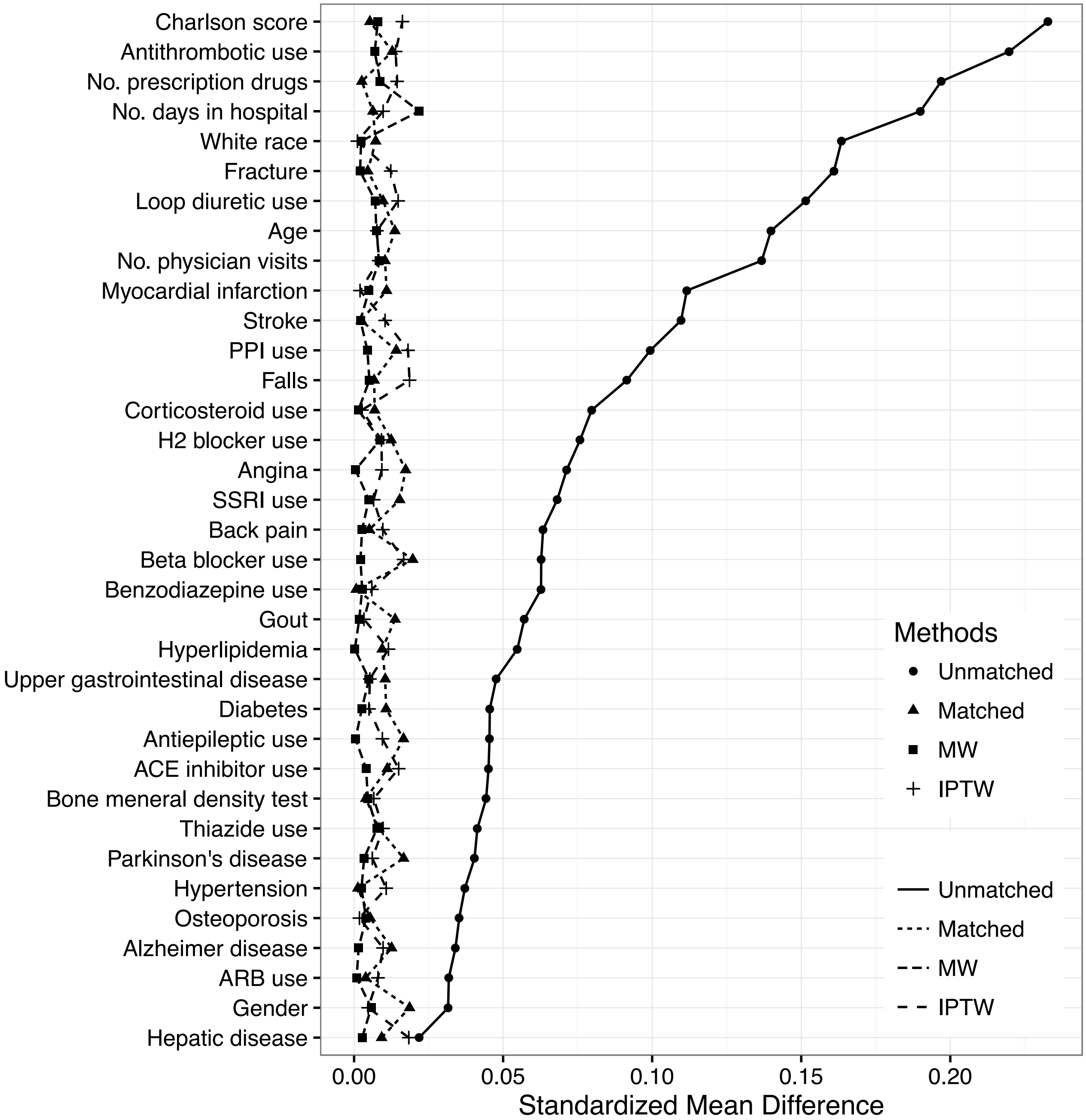
**Abbreviations:** U: Unmatched cohort, M: Matched cohort; Mw: matching weight cohort; Ip: Inverse probability of treatment weight cohort; pExpo: Exposure prevalence; pDis: baseline risk of disease

**eFigure 9.** Comparison of coverage probability of estimated confidence intervals between methods across 48 scenarios. The left half presents the constant treatment effect scenarios, whereas the right half presents treatment effect heterogeneity scenarios. Each three columns represent three treatment contrasts. Rows classify scenarios by good vs. poor covariate overlap levels and presence vs. absence of main effects. Each panel contains six lines classified by the exposure prevalence and the baseline risk. matching weights and matching performed similarly, whereas IPTW performed poorly in the poor covariate overlap settings. These confidence intervals were based on the estimated variance.



**Abbreviations:** U: Unmatched cohort, M: Matched cohort; Mw: matching weight cohort; Ip: Inverse probability of treatment weight cohort; pExpo: Exposure prevalence; pDis: baseline risk of disease

**eFigure 10.** Standardized mean differences for each covariate averaged across three treatment contrasts in the unmatched, weighted, and matched cohort. Matching weights achieved the best covariate balance most consistently (24 of the 35 covariates) compared to three-way matching (6 covariates) and IPTW (5 covariates).



**Abbreviations:** PPI: proton pump inhibitor; H2: histamine-2 receptor; SSRI: selective serotonin reuptake inhibitor; ACE: angiotensin converting enzyme; ARB: angiotensin receptor blocker; MW: matching weights; IPTW: Inverse probability of treatment weights.

**eTable 1.** Characteristics of unmatched, matched, and weighted cohorts for the variables that were least balanced (average standardized mean difference > 0.1). The MW and matched cohorts were similar in characteristics, confirming the notion that MW is a weighting analogue to matching. As expected from the definition of the common support (overlap area of all three groups), these two cohorts are most similar to the smallest group, *i.e.*, the NSAIDs group in the unmatched cohort. The IPTW cohort had somewhat different characteristics with higher morbidity levels, most closely resembling the largest group, *i.e.*, the opioids group.

| | nsNSAIDs | Coxibs | Opioids | SMD |
|---|---|---|---|---|
| ***Unmatched*** | | | | |
| **n** | 4874 | 6172 | 12601 | |
| **Charlson score, mean (SD)** | 1.59 (1.54) | 1.72 (1.53) | 2.17 (1.78) | 0.233 |
| **Antithrombotic use, %** | 14.4 | 17.6 | 27.7 | 0.220 |
| **No. prescription drugs, mean (SD)** | 8.28 (4.69) | 8.55 (4.76) | 9.76 (5.38) | 0.197 |
| **No. days in hospital, mean (SD)** | 1.85 (6.90) | 2.19 (6.86) | 4.18 (9.46) | 0.190 |
| **White race, %** | 84.6 | 88 | 92.4 | 0.164 |
| **Fracture, %** | 6.5 | 7.2 | 13.7 | 0.161 |
| **Loop diuretic use, %** | 21.3 | 25.8 | 31.3 | 0.152 |
| **Age, mean (SD)** | 79.67 (7.03) | 80.87 (6.99) | 81.15 (7.17) | 0.140 |
| **No. physician visits, mean (SD)** | 8.72 (6.32) | 8.80 (5.99) | 10.08 (7.14) | 0.137 |
| **Myocardial infarction, %** | 5.2 | 5.7 | 9.6 | 0.112 |
| **Stroke, %** | 15.2 | 16.1 | 21.5 | 0.110 |
| | | | | |
| ***Matched*** | | | | |
| **n** | 4611 | 4611 | 4611 | |
| **Charlson score, mean (SD)** | 1.62 (1.54) | 1.63 (1.52) | 1.61 (1.52) | 0.005 |
| **Antithrombotic use, %** | 15.1 | 15.5 | 15.8 | 0.013 |
| **No. prescription drugs, mean (SD)** | 8.34 (4.70) | 8.33 (4.69) | 8.32 (4.71) | 0.003 |
| **No. days in hospital, mean (SD)** | 1.89 (6.45) | 1.88 (6.54) | 1.94 (6.29) | 0.006 |
| **White race, %** | 86.9 | 86.7 | 86.6 | 0.007 |
| **Fracture, %** | 6.7 | 6.9 | 6.7 | 0.005 |
| **Loop diuretic use, %** | 22 | 22 | 22.6 | 0.010 |
| **Age, mean (SD)** | 79.97 (6.97) | 79.96 (6.93) | 80.11 (6.92) | 0.014 |
| **No. physician visits, mean (SD)** | 8.76 (6.08) | 8.76 (5.93) | 8.66 (5.84) | 0.010 |
| **Myocardial infarction, %** | 5.4 | 5.2 | 5.6 | 0.011 |
| **Stroke, %** | 15.5 | 15.6 | 15.7 | 0.002 |
| | | | | |
| ***Matching weights*** | | | | |
| **n** | 4633.49 | 4635.71 | 4618.71 | |
| **Charlson score, mean (SD)** | 1.62 (1.53) | 1.61 (1.52) | 1.63 (1.53) | 0.008 |
| **Antithrombotic use, %** | 14.9 | 14.8 | 15.2 | 0.007 |
| **No. prescription drugs, mean (SD)** | 8.32 (4.70) | 8.29 (4.67) | 8.35 (4.71) | 0.009 |
| **No. days in hospital, mean (SD)** | 1.87 (6.37) | 1.78 (6.18) | 2.00 (6.99) | 0.022 |
| **White race, %** | 86.3 | 86.4 | 86.4 | 0.002 |
| **Fracture, %** | 6.7 | 6.7 | 6.7 | 0.002 |
| **Loop diuretic use, %** | 22 | 21.8 | 22.3 | 0.007 |

| | | | | |
|---|---|---|---|---|
| **Age, mean (SD)** | 79.97 (6.95) | 79.95 (6.97) | 80.02 (6.95) | 0.007 |
| **No. physician visits, mean (SD)** | 8.72 (6.09) | 8.69 (6.01) | 8.76 (6.04) | 0.008 |
| **Myocardial infarction, %** | 5.3 | 5.2 | 5.4 | 0.005 |
| **Stroke, %** | 15.4 | 15.4 | 15.5 | 0.002 |
| | | | | |
| ***IPTW*** | | | | |
| **n** | 4926.58 | 6187.8 | 12585.04 | |
| **Charlson score, mean (SD)** | 1.98 (1.70) | 1.94 (1.68) | 1.94 (1.69) | 0.016 |
| **Antithrombotic use, %** | 23.3 | 22.5 | 22.4 | 0.014 |
| **No. prescription drugs, mean (SD)** | 9.27 (5.17) | 9.15 (5.15) | 9.17 (5.14) | 0.014 |
| **No. days in hospital, mean (SD)** | 3.48 (8.96) | 3.35 (8.78) | 3.39 (9.82) | 0.010 |
| **White race, %** | 89.7 | 89.7 | 89.7 | 0.001 |
| **Fracture, %** | 11.2 | 10.8 | 10.6 | 0.012 |
| **Loop diuretic use, %** | 28.9 | 27.9 | 27.9 | 0.015 |
| **Age, mean (SD)** | 80.89 (7.17) | 80.82 (7.11) | 80.81 (7.11) | 0.008 |
| **No. physician visits, mean (SD)** | 9.58 (6.82) | 9.49 (6.66) | 9.50 (6.75) | 0.008 |
| **Myocardial infarction, %** | 7.8 | 7.7 | 7.7 | 0.002 |
| **Stroke, %** | 19.4 | 18.8 | 18.9 | 0.010 |

**Abbreviations**: Matched: three-way matching; IPTW: inverse probability of treatment weights; Coxibs: COX-2 selective inhibitors; nsNSAIDs: non-selective nosteroidal anti-inflammatory drugs; SMD: standardized mean difference averaged across three pairwise contrasts