A Simulation Study Using Inverse Probability Weighting to Adjust for Multiple Types of Bias in Observational Studies

Diqiong Xie, Ph.D.

Food and Drug Administration/Center for Drug Evaluation and Research/Office of Translational Sciences/Office of Biostatistics/Division of Biometrics 7

Abstract

Inverse probability weighting (IPW) has proven to be an effective tool to reduce biases from selective treatment assignment and missing at random, one type of bias at a time. Many studies have since employed IPW to adjust for multiple types of bias simultaneously, where separate weights were estimated for treatment assignment, loss of follow-up and censoring, and a total weight was estimated by the product of multiple weights.

This simulation study compares the efficiency of four PS methods, involving IPW and PS matching, in adjusting for biases of treatment selection and missing outcome.

We find that 1 : 1 PS matching for treatment assignment and IPW for missing outcome data has better estimates overall, even though it only uses part of the observed data.

Introduction

Propensity score (PS) matching has proven to be efficient in estimating the treatment effect adjusting for treatment selection bias. The advantages of PS matching are:

- Easy to interpret.
- Creates pseudo-randomization of treatment assignment.
- Has adjustable quality using different matching calipers.

The disadvantages of matching are:

- Does not use all the observed data.
- Result does not apply to subjects with missing outcome.

IPW is increasingly used to adjust for the selection biases caused by not only treatment assignment, but also missing outcome data. The advantages of IPW method are:

- It utilizes all observed data.
- It simulates pseudo-randomization of treatment assignment and missing outcome.
- Result can be generalized to those with missing outcome.

However, no study that we are aware of has shown the performance of combining multiple IPWs to adjust for multiple types of bias in estimating the treatment effect.

Notation

We define the following for subject i:

- Y_i continuous outcome.
- M_i indicator for missing outcome, 1 if Y_i is missing, 0 if Y_i isobserved.
- T_i assigned treatment, 1 if treated, 0 if control.
- X_i observed covariates.
- X_{M_i} subset of X_i to estimate the PS for missing.
- X_{T_i} subset of X_i to estimate the PS for treatment.

Methods

We simulate observational data with treatment selection bias and missing outcome, and compare the efficiency of four PS methods in estimating a binary treatment effect on a continuous outcome:

- Double IPW without weight trimming.
- Double IPW with 95% weight trimming.
- $1:1$ PS matching for treatment assignment and IPW for missing outcome.
- $1:R$ ($R \leq 4$) PS matching from treatment assignment and IPW for missing outcome.

For each method, two PSs are estimated using logistics regressions on the probability of being treated and having missing outcome.

$$PS_{Ti} = \text{Prob}(T_i = 1 | X_{Ti}), \quad PS_{Mi} = \text{Prob}(M_i = 1 | X_{Mi}, T_i).$$

In the double IPW method without weight trimming, two stabilized inverse probability weights are generated using the corresponding PS for treatment assignment and missing outcome.

$$IPW_{Ti} = \begin{cases} P(T_i = 1)/PS_{Ti} & \text{if } T_i = 1 \\ P(T_i = 0)/(1 - PS_{Ti}) & \text{if } T_i = 0, \end{cases} \quad IPW_{Mi} = \begin{cases} P(M_i = 1 | T_i = t_i)/PS_{Mi} & \text{if } M_i = 1 \\ P(M_i = 0 | T_i = t_i)/(1 - PS_{Mi}) & \text{if } M_i = 0, \end{cases}$$

The final weight is $IPW_{TMi} = IPW_{Ti} \times IPW_{Mi}$. The treatment effect is then estimated by a linear regression $Y \sim T$ weighted by IPW_{TM}. In the double IPW with 95% weight trimming, we use cutpoints of the 2.5^{th} and 97.5^{th} percentile of the IPW_{TM} distribution.

In the PS matching and IPW methods, we generate the 95% trimmed IPW_{Mi} using the above mentioned method. Then, treated and control subjects are matched with a $1:1$ or $1:R$ ratio without replace using the logit PS_T and the nearest neighbor matching method. A matching caliper of $0.2 \times \widehat{sd}(\text{logit } PS_T)$ is used. In $1:R$ matching, one subject can be matched up to four subjects within the matching caliper. Using the matched data, the treatment effect is estimated by a linear regression $Y \sim T$ weighted by IPW_M not conditioning on matching.

The standard deviation (sd) of $\widehat{\tau}$ is estimated by the conventional and the sandwich estimators. The 95% CI is calculated by $\widehat{\tau} \pm 1.96 \times \widehat{sd}(\widehat{\tau})$.

Simulation Dataset

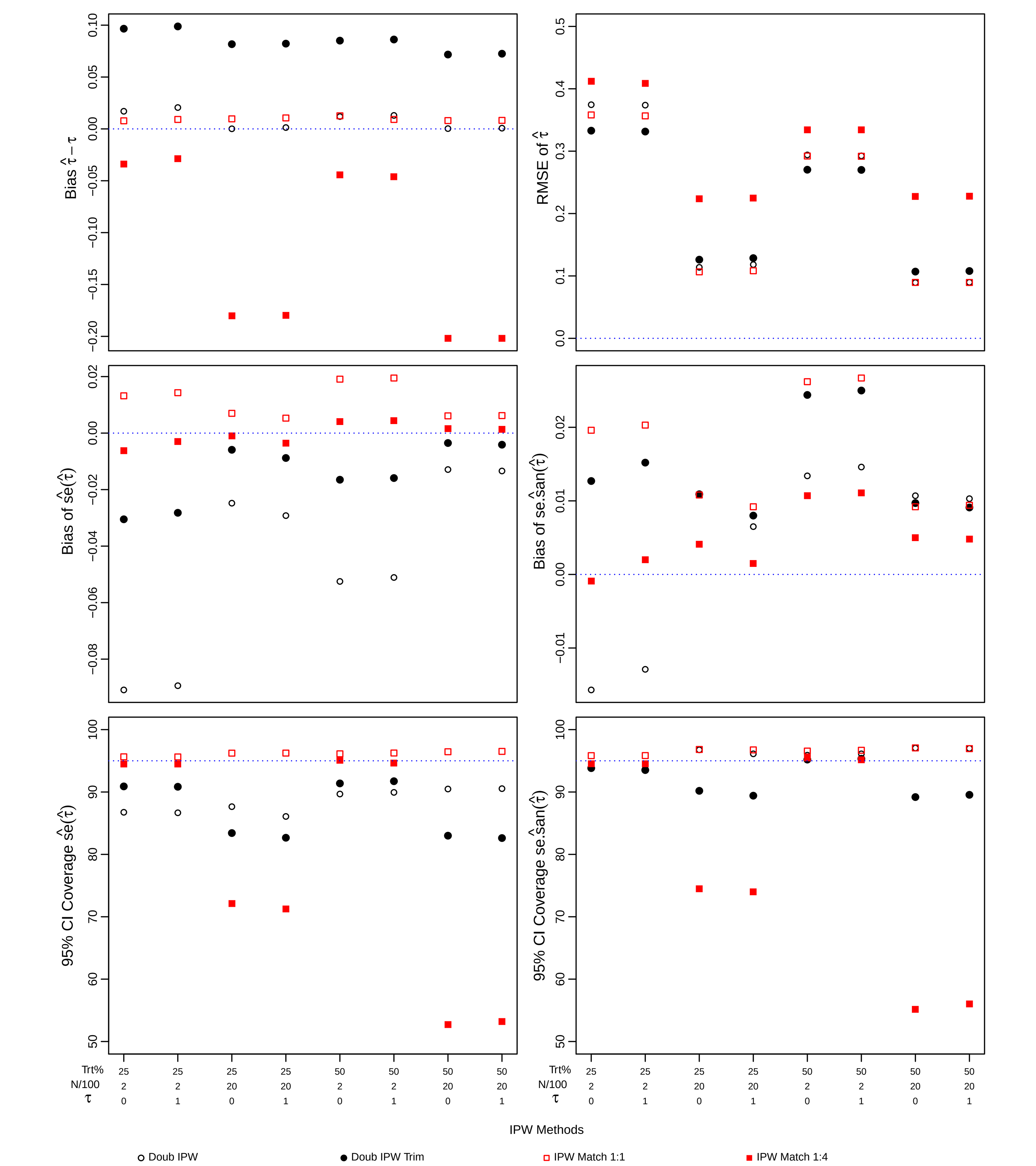
We simulate 10,000 datasets for all combinations of the following scenarios:

- Small and large sample size: $N = 200$ and 2000.
- Absent and present treatment effects: $\tau = 0$ and 1.
- Small and medium treatment assignment proportions: $p = 25\%$ and 50%.
- Proportion of missing outcomes: $m = 20\%$.

To mimic an observational study with selection biases,

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	T	M
Missing Outcome M	✓	✓	✓			✓				
Treatment T					✓	✓	✓	✓		
Outcome Y			✓	✓	✓	✓			✓	✓

Results



Conclusions

- Bias($\widehat{\tau}$): double IPW w.o. trimming and $1:1$ matching with IPW are better.
- RMSE($\widehat{\tau}$): double IPW w. trimming is better with small N, double IPW w.o. trimming and $1:1$ matching with IPW are better with large N.
- 95% CI coverage: double IPW and $1:R$ matching with IPW have low coverage, especially in large samples.
- In the double IPW method, sandwich variance estimator should be used.
- In the matching plus IPW method,
  - When trt. group sizes are similar, sandwich variance estimator is better.
  - Otherwise, conventional variance estimators is better.

$1:1$ PS matching with IPW has better estimates overall: less bias in $\widehat{\tau}$ and $\widehat{se}(\widehat{\tau})$, and better 95% CI coverage.