

SREE Title: The effect of differential measurement error on numeric covariate balance

Authors: Heather Harris, S. Jeanne Horst

Background/Context:

Educational researchers interested in estimating the impact of an educational program must contend with unaddressed threats to the internal validity of their study, due to selection bias. Propensity score analyses are a popular method for accounting for confounding variables related to self-selected participation (Kim & Steiner, 2015; Pearl, 2010). However, recommendations on the use of propensity score analyses typically assume covariates are measured *without error* (Guo & Fraser, 2014). That is, only a handful of studies have discussed the impact of measurement error on treatment effect estimation following propensity score analyses (e.g., Millimet, 2011; Rudolph & Stuart, 2016; Steiner, Cook, & Shadish., 2011).

The purpose of the current study was to evaluate the impact of varying types (e.g., measurement error that is similar across treatment groups versus measurement error that is systematically differential by group) and degrees of measurement error (e.g., 10% versus 30% measurement error or unreliability) on a researcher’s ability to create balanced treatment-comparison groups. (See Table 1 for the simulated conditions in the current study.)

Research Questions

Research Question 1: How do differing levels of covariate score measurement error (e.g., 10%, 20%, 30% unreliability) affect the quality of matches created when using common propensity score matching techniques (e.g., nearest neighbor matching within a caliper distance, optimal matching, Mahalanobis distance matching)?

Research Question 2: How do different types of covariate score measurement error (i.e., measurement error that is systematically similar across groups versus measurement error that is differential by group) affect the quality of matches created when using common propensity score matching techniques (e.g., nearest neighbor matching within a caliper distance, optimal matching, Mahalanobis distance matching)?

Research Design and Data Generation

Data were generated in R version 3.4.3 (R Core Team, 2017) for 1,000 simulees across 1,000 replications using the mvtnorm package (Genz et al., 2015). Figure 1 displays the conceptual relationship among variables in the simulated data set (adapted from Harris et al., 2018) set via the correlations displayed in Table 1. Twelve measurement error conditions were simulated for the error-prone covariates, which are displayed in Table 2. Table 3 displays the correlations among simulated covariates in a validation data set, illustrating that the data were simulated correctly.

The performance of two sets of covariates in the context of propensity score matching were evaluated: a) a covariate set including X1, X2 (i.e., the two covariates measured with error) and X3-X5 (i.e., three covariates measured without error), versus b) a covariate set including L1, L2 (i.e., substituting the error-free versions of X1 and X2 in the covariate set), and X3-X5. Henceforth, propensity scores estimated using the error-prone covariate (i.e., X1-X5) set are referred to as “naïve propensity scores,” and propensity scores estimated using the error-free covariate set (i.e., L1, L2, and X3-X5) are referred to as “true propensity scores.”

The MatchIt package in R (Ho, Imai, King, & Stuart, 2013) was used to conduct propensity score matching for each of the 1,000 replications across the twelve conditions of the simulation study. Treatment-comparison groups were constructed using nearest neighbor

matching, nearest neighbor matching with a 0.2 caliper width, optimal matching, and Mahalanobis distance matching (Austin, 2009; Ho et al., 2011; Stuart, 2010).

The numeric diagnosis of matches involved evaluation of the following: the standardized mean difference between propensity score matched treatment-comparison groups on key covariates (Austin, 2011; Stuart, 2010), the variance ratio between groups' propensity score distributions (Rubin, 2001), and the percent in bias reduction from before matching to after matching (Pan & Bai, 2015). Matched groups are considered balanced across the set of covariates if the standardized mean difference on each individual covariate falls below a specified benchmark (e.g., 0.2; Austin, 2011). Similarly, for percent of bias reduction (PBR), an 80% reduction in bias from before matching to after matching is viewed as sufficient (Pan & Bai, 2015).

Results

Tables 4 through 7 display the percent in bias reduction (PBR), standardized mean difference (d), and variance ratio (Var Ratio) for each of the matching methods for the same measurement error conditions. Similarly, Tables 8 through 11 display the same information for the different measurement error conditions.

The percent in bias reduction (displayed as a proportion in the tables) met or exceeded the 80% improvement (or 0.8) recommended in the literature (Pan & Bai, 2015) for most of the twelve simulated conditions. Figures 2 and 3 display the average percent in bias reduction for L1 (dark grey bars) and X1 (light grey bars) by condition for each matching method for the same measurement error conditions (Figure 2) and for the different measurement error conditions (Figure 3). Figures 4 and 5 display the average standardized mean difference for L1 (dark grey bars) and X1 (light grey bars) for the same measurement error conditions (Figure 4) and for the different measurement error conditions (Figure 5). Notice the identical pattern was present for both the same measurement error and the different measurement error conditions. That is, as the amount of measurement error increased (i.e., as unreliability goes from 10% in Condition 1 to 60% in Condition 6), matches created using the naïve propensity score did not appropriately balance the true covariate (i.e., L1).

Conclusions

When groups were matched using the naïve propensity score, the final matched treatment-comparison groups were adequately balanced only on the error-prone covariates (i.e., naïve covariates, X1 and X2). That is, the final naïve propensity score matched treatment-comparison groups appeared to be well-balanced on the error-prone covariates after evaluating the percent bias reduction, standardized mean differences, and variance ratios. However, the error-free covariates (i.e., true covariates, L1 and L2) were not well balanced – particularly in the high measurement error conditions.

One way in which researchers can improve the accuracy of treatment effect estimates when using propensity score matching is to administer and include measures that result in reliable covariates scores. Future research studies should include conditions with differing – and larger – covariate sets that include both continuous and categorical variables.

References

- Austin, P. C. (2009). Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in Medicine*, 28, 3083-3107.
- Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies, *Multivariate Behavioral Research*, 46(3), 399-424.
- Genz, A., & Bretz, F. Miwa, T., Mi, X., Leisch, F., Scheipl, F., & Hothorn, T. (2015). mvtnorm: Multivariate normal and t distributions. R package version 1.0-3. URL <http://CRAN.R-project.org/package=mvtnorm>
- Guo, S., & Fraser, M. W. (2014). *Propensity score analysis: Statistical methods and applications* (2nd ed.). Thousand Oaks, CA: SAGE Publications Inc.
- Harris, H., Horst, S. J., & DeMars, C. E. (2018). Investigating the effect of covariate choice on treatment estimation when using propensity score matching techniques: A simulation study. Unpublished manuscript.
- Ho, D., Imai, K., King, G., & Stuart, E. (2011). MatchIt: Nonparametric preprocessing for parametric casual inference. *Journal of Statistical Software*, 42, 1-28.
- Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2013). MatchIt: Nonparametric preprocessing for parametric causal inference. Software for using matching methods in R. Available at <http://gking.harvard.edu/matchit/>
- Kim, J., & Steiner, P. M. (2015). Multilevel propensity score methods for estimating causal effects: A latent class modeling strategy. *Quantitative Psychology Research*, 140, 293-306.
- Millimet, D. L. (2011). The elephant in the corner: A cautionary tale about measurement error in treatment effect models. In T. B. Fomby, R. C. Hill, I. Jeliazkov, J. C. Escanciano, and E. Hillebrand (Eds.), *Book Series: Advances in Economics*. Bingley, West Yorkshire: Emerald Insights Publications, Inc.
- Pan, W., & Bai, H. (2015). Propensity score analysis: Concepts and issues. In W. Pan & H. Bai (Eds.), *Propensity Score Analysis: Fundamentals and Developments* (pp. 3-19). New York, NY: Guilford Publications, Inc.
- Pearl, J. (2010). The foundations of causal inference. *Sociological Methodology*, 40, 75-149.
- R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Rubin, D. B. (2001). Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services & Outcomes Research Methodology*, 2, 169-188.
- Rudolph, K. E., & Stuart, E. A. (2016). Using sensitivity analyses for unobserved confounding to address covariate measurement error in propensity score methods. *Johns Hopkins University Dept. of Biostatistics Working Papers*. Working Paper 283.
- Steiner, P. M., Cook, T. D., & Shadish, W. R. (2011). On the importance of reliable covariate measurement in selection bias adjustments using propensity scores. *Journal of Educational and Behavioral Statistics*, 36, 213-236.
- Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science*, 25(1), 1-21.

Table 1
Correlations Among Covariates Measured Without Error

Covariates	L1	L2	X3	X4	X5
L1	1.00				
L2	0.58	1.00			
X3	0.42	0.56	1.00		
X4	0.49	0.44	0.59	1.00	
X5	0.60	0.41	0.44	0.51	1.00

Note. As measurement error is imposed on covariates L1 and L2, the correlations between these covariates and other covariates in the model is attenuated. Correlations among covariates were simulated to represent relationships found in educational and psychological research studies.

Table 2
Simulated Conditions and Measurement Error Levels on X1 and X2

	CTT Measurement Error Treatment & Control		Differential by Group			
	% Error	Reliability	Treatment		Control	
	% Error	Reliability	% Error	Reliability	% Error	Reliability
Condition 1	10	0.90	20	0.80	10	0.90
Condition 2	20	0.80	20	0.80	20	0.80
Condition 3	30	0.70	20	0.80	30	0.70
Condition 4	40	0.60	20	0.80	40	0.60
Condition 5	50	0.50	20	0.80	50	0.50
Condition 6	60	0.40	20	0.80	60	0.40

Note. For each of the six conditions, nearest neighbor matching, nearest neighbor matching with a 0.2 caliper, optimal matching, and Mahalanobis matching were conducted. Levels of reliability were simulated to represent typical reliability estimates seen for scores from educational and psychological research measures.

Table 3
Simulated Correlations Between Simulated Error-free and Error-prone Covariates

Same Measurement Error				Different Measurement Error							
Treatment and Control Groups				Treatment Group				Control Group			
Square Root of				Square Root of				Square Root of			
Reliability	Reliability	L1_X1	L2_X2	Reliability	Reliability	L1_X1	L2_X2	Reliability	Reliability	L1_X1	L2_X2
0.90	0.95	0.95	0.95	0.80	0.89	0.89	0.90	0.90	0.95	0.94	0.94
0.80	0.89	0.90	0.90	0.80	0.89	0.90	0.88	0.80	0.89	0.88	0.88
0.70	0.84	0.84	0.84	0.80	0.89	0.88	0.87	0.70	0.84	0.81	0.83
0.60	0.77	0.76	0.76	0.80	0.89	0.88	0.89	0.60	0.77	0.76	0.78
0.50	0.71	0.69	0.69	0.80	0.89	0.86	0.88	0.50	0.71	0.70	0.71
0.40	0.63	0.66	0.62	0.80	0.89	0.90	0.88	0.40	0.63	0.66	0.60

Note. L1_X1 indicates the correlation between the first error-free simulated covariate (L1) and the corresponding error-prone covariate (X1). L2_X2 indicates the correlation between the second error-free simulated covariate (L2) and the corresponding error-prone covariate (X2).

Table 4

Numeric Diagnostic Information for Same Measurement Error Conditions After Nearest Neighbor Matching Using the True and Naïve Propensity Scores

	L1		L2		X1		X2		X3		X4		X5		Propensity Score Var Ratio	
	PBR	<i>d</i>	PBR	<i>d</i>	PBR	<i>d</i>	PBR	<i>d</i>	PBR	<i>d</i>	PBR	<i>d</i>	PBR	<i>d</i>	Ratio	<i>d</i>
True																
Cond 1	0.94	0.04	0.95	0.04	0.95	0.03	0.95	0.03	0.95	0.03	0.95	0.03	0.95	0.03	1.25	0.06
Cond 2	0.95	0.04	0.95	0.03	0.94	0.03	0.95	0.03	0.95	0.03	0.95	0.03	0.95	0.03	1.10	0.02
Cond 3	0.95	0.03	0.95	0.03	0.95	0.03	0.95	0.03	0.95	0.03	0.95	0.03	0.95	0.03	1.11	0.02
Cond 4	0.95	0.03	0.95	0.03	0.95	0.03	0.96	0.02	0.95	0.03	0.95	0.03	0.95	0.03	1.26	0.06
Cond 5	0.95	0.03	0.95	0.03	0.94	0.03	0.95	0.02	0.95	0.03	0.95	0.03	0.95	0.03	1.14	0.03
Cond 6	0.95	0.04	0.95	0.03	0.94	0.03	0.94	0.02	0.95	0.03	0.95	0.03	0.95	0.03	1.10	0.02
Naïve																
Cond 1	0.92	0.05	0.91	0.06	0.95	0.03	0.95	0.03	0.95	0.03	0.95	0.04	0.95	0.03	1.23	0.06
Cond 2	0.90	0.07	0.88	0.08	0.95	0.03	0.95	0.03	0.95	0.03	0.95	0.03	0.95	0.03	1.12	0.02
Cond 3	0.87	0.09	0.84	0.10	0.95	0.03	0.96	0.02	0.96	0.03	0.96	0.03	0.95	0.03	1.12	0.03
Cond 4	0.85	0.10	0.81	0.12	0.96	0.02	0.96	0.02	0.95	0.03	0.95	0.03	0.96	0.03	1.20	0.05
Cond 5	0.82	0.12	0.78	0.14	0.95	0.02	0.95	0.02	0.95	0.03	0.96	0.03	0.96	0.03	1.14	0.03
Cond 6	0.80	0.13	0.76	0.16	0.95	0.02	0.96	0.02	0.96	0.03	0.96	0.03	0.95	0.03	1.10	0.02

Note. PBR indicates the average percent in bias reduction across 1,000 replications for each condition, *d* indicates the average standardized mean difference across 1,000 replications for each condition (control group subtracted from treatment group divided by the pooled standard deviation), and Var Ratio indicates the average variance ratio across 1,000 replications for each condition (treatment group's propensity score variance divided by the control group's propensity score variance). True in the table refers to the true propensity score, which included the error-free covariates (L1 and L2) in the covariate set. Naïve in the table refers to the naïve propensity score, which included the error-prone covariates (X1 and X2) in the covariate set. Cond indicates the corresponding measurement error condition.

Table 5
Numeric Diagnostic Information for Same Measurement Error Conditions After Nearest Neighbor Matching with a Caliper Using the True and Naïve Propensity Scores

True	L1		L2		X1		X2		X3		X4		X5		Propensity Score	
	PBR	<i>d</i>	PBR	<i>d</i>	PBR	<i>d</i>	PBR	<i>d</i>	PBR	<i>d</i>	PBR	<i>d</i>	PBR	<i>d</i>	Var Ratio	<i>d</i>
Cond 1	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	0.99	0.00	1.00	0.00
Cond 2	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
Cond 3	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.01	0.00
Cond 4	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	0.99	0.00	1.00	0.00	1.00	0.01
Cond 5	1.00	0.00	0.99	0.00	0.99	0.00	0.99	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
Cond 6	0.99	0.01	1.00	0.00	0.98	0.01	0.99	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
Naïve																
Cond 1	0.96	0.03	0.96	0.03	0.99	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
Cond 2	0.94	0.04	0.92	0.05	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
Cond 3	0.91	0.06	0.88	0.08	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
Cond 4	0.88	0.08	0.85	0.10	1.00	0.00	1.00	0.00	1.00	0.00	0.99	0.00	1.00	0.00	1.00	0.00
Cond 5	0.86	0.10	0.82	0.12	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.01	0.00
Cond 6	0.83	0.11	0.79	0.14	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00

Note. PBR indicates the average percent in bias reduction across 1,000 replications for each condition, *d* indicates the average standardized mean difference across 1,000 replications for each condition (control group subtracted from treatment group divided by the pooled standard deviation), and Var Ratio indicates the average variance ratio across 1,000 replications for each condition (treatment group's propensity score variance divided by the control group's propensity score variance). True in the table refers to the true propensity score, which included the error-free covariates (L1 and L2) in the covariate set. Naïve in the table refers to the naïve propensity score, which included the error-prone covariates (X1 and X2) in the covariate set. Cond indicates the corresponding measurement error condition.

Table 6

Numeric Diagnostic Information for Same Measurement Error Conditions After Optimal Matching Using the True and Naïve Propensity Scores

	L1		L2		X1		X2		X3		X4		X5		Propensity Score		
	True	PBR	<i>d</i>	PBR	<i>d</i>	PBR	<i>d</i>	PBR	<i>d</i>	PBR	<i>d</i>	PBR	<i>d</i>	PBR	<i>d</i>	Var Ratio	<i>d</i>
Cond 1	0.95	0.95	0.03	0.95	0.03	0.95	0.03	0.95	0.03	0.95	0.03	0.96	0.03	0.95	0.03	1.24	0.06
Cond 2	0.95	0.95	0.03	0.95	0.03	0.95	0.03	0.95	0.03	0.96	0.03	0.96	0.03	0.95	0.03	1.06	0.01
Cond 3	0.95	0.95	0.03	0.96	0.03	0.95	0.03	0.96	0.02	0.96	0.03	0.96	0.03	0.95	0.03	1.09	0.02
Cond 4	0.96	0.96	0.03	0.96	0.03	0.96	0.02	0.96	0.02	0.96	0.03	0.96	0.03	0.95	0.03	1.26	0.06
Cond 5	0.96	0.95	0.03	0.95	0.03	0.94	0.03	0.95	0.02	0.95	0.03	0.95	0.03	0.96	0.03	1.13	0.03
Cond 6	0.95	0.95	0.03	0.95	0.03	0.95	0.02	0.95	0.02	0.96	0.03	0.95	0.03	0.95	0.03	1.09	0.02
Naïve																	
Cond 1	0.93	0.92	0.05	0.92	0.05	0.95	0.03	0.95	0.03	0.95	0.03	0.96	0.03	0.96	0.03	1.22	0.06
Cond 2	0.90	0.88	0.06	0.88	0.08	0.96	0.03	0.95	0.03	0.96	0.03	0.96	0.03	0.96	0.03	1.11	0.01
Cond 3	0.87	0.85	0.09	0.85	0.10	0.96	0.02	0.96	0.02	0.96	0.03	0.96	0.03	0.96	0.03	1.11	0.02
Cond 4	0.85	0.82	0.10	0.82	0.12	0.96	0.02	0.96	0.02	0.96	0.03	0.96	0.03	0.96	0.03	1.19	0.04
Cond 5	0.83	0.79	0.11	0.79	0.14	0.96	0.02	0.96	0.02	0.96	0.03	0.96	0.03	0.96	0.02	1.14	0.03
Cond 6	0.81	0.76	0.13	0.76	0.16	0.96	0.02	0.96	0.02	0.96	0.03	0.96	0.03	0.96	0.03	1.09	0.02

Note. PBR indicates the average percent in bias reduction across 1,000 replications for each condition, *d* indicates the average standardized mean difference across 1,000 replications for each condition (control group subtracted from treatment group divided by the pooled standard deviation), and Var Ratio indicates the average variance ratio across 1,000 replications for each condition (treatment group's propensity score variance divided by the control group's propensity score variance). True in the table refers to the true propensity score, which included the error-free covariates (L1 and L2) in the covariate set. Naïve in the table refers to the naïve propensity score, which included the error-prone covariates (X1 and X2) in the covariate set. Cond indicates the corresponding measurement error condition.

Table 7

Numeric Diagnostic Information for Same Measurement Error Conditions After Mahalanobis Distance Matching Using the True and Naïve Propensity Scores

	L1		L2		X1		X2		X3		X4		X5		Propensity Score		
	True	PBR	d	PBR	d	PBR	d	PBR	d	PBR	d	PBR	d	PBR	d	Var Ratio	d
Cond 1	0.87	0.87	0.09	0.87	0.09	0.87	0.08	0.87	0.08	0.87	0.09	0.87	0.09	0.87	0.09	1.30	0.10
Cond 2	0.87	0.87	0.09	0.87	0.09	0.87	0.08	0.87	0.08	0.87	0.09	0.87	0.09	0.87	0.09	1.10	0.08
Cond 3	0.87	0.87	0.09	0.87	0.09	0.87	0.07	0.87	0.07	0.87	0.09	0.87	0.09	0.87	0.09	1.21	0.12
Cond 4	0.87	0.87	0.09	0.87	0.09	0.87	0.06	0.87	0.07	0.87	0.09	0.87	0.09	0.87	0.09	1.29	0.13
Cond 5	0.87	0.87	0.09	0.87	0.09	0.86	0.07	0.87	0.06	0.87	0.09	0.87	0.09	0.87	0.09	1.22	0.12
Cond 6	0.87	0.87	0.09	0.87	0.09	0.86	0.06	0.87	0.05	0.87	0.09	0.87	0.09	0.87	0.09	1.22	0.11
Naïve																	
Cond 1	0.84	0.84	0.11	0.84	0.11	0.87	0.08	0.87	0.08	0.87	0.09	0.87	0.09	0.87	0.09	1.31	0.11
Cond 2	0.82	0.82	0.12	0.80	0.13	0.87	0.08	0.87	0.08	0.87	0.09	0.87	0.09	0.87	0.09	1.14	0.09
Cond 3	0.80	0.80	0.14	0.77	0.15	0.87	0.07	0.87	0.07	0.87	0.09	0.87	0.09	0.87	0.09	1.23	0.09
Cond 4	0.78	0.78	0.15	0.75	0.17	0.87	0.07	0.88	0.06	0.87	0.09	0.87	0.09	0.87	0.09	1.29	0.13
Cond 5	0.75	0.75	0.17	0.72	0.19	0.87	0.06	0.87	0.06	0.87	0.09	0.87	0.09	0.87	0.09	1.24	0.10
Cond 6	0.73	0.73	0.18	0.69	0.21	0.87	0.05	0.87	0.05	0.87	0.09	0.87	0.09	0.88	0.09	1.23	0.13

Note. PBR indicates the average percent in bias reduction across 1,000 replications for each condition, d indicates the average standardized mean difference across 1,000 replications for each condition (control group subtracted from treatment group divided by the pooled standard deviation), and Var Ratio indicates the average variance ratio across 1,000 replications for each condition (treatment group's propensity score variance divided by the control group's propensity score variance). True in the table refers to the true propensity score, which included the error-free covariates (L1 and L2) in the covariate set. Naïve in the table refers to the naïve propensity score, which included the error-prone covariates (X1 and X2) in the covariate set. Cond indicates the corresponding measurement error condition.

Table 8

Numeric Diagnostic Information for Different Measurement Error Conditions After Nearest Neighbor Matching Using the True and Naïve Propensity Scores

	L1		L2		X1		X2		X3		X4		X5		Propensity Score		
	True	PBR	<i>d</i>	PBR	<i>d</i>	PBR	<i>d</i>	PBR	<i>d</i>	PBR	<i>d</i>	PBR	<i>d</i>	PBR	<i>d</i>	Var Ratio	<i>d</i>
Cond 1	0.95	0.95	0.03	0.95	0.03	0.95	0.03	0.95	0.03	0.95	0.03	0.95	0.03	0.95	0.03	1.27	0.08
Cond 2	0.95	0.95	0.04	0.95	0.03	0.95	0.03	0.95	0.03	0.95	0.03	0.95	0.03	0.95	0.03	1.19	0.04
Cond 3	0.95	0.95	0.03	0.95	0.03	0.95	0.03	0.95	0.03	0.95	0.03	0.95	0.03	0.95	0.03	1.25	0.06
Cond 4	0.95	0.95	0.03	0.95	0.03	0.95	0.03	0.95	0.03	0.95	0.03	0.95	0.03	0.95	0.03	1.17	0.04
Cond 5	0.95	0.95	0.03	0.95	0.03	0.95	0.03	0.95	0.03	0.95	0.03	0.95	0.03	0.95	0.03	1.06	0.01
Cond 6	0.95	0.95	0.03	0.95	0.03	0.95	0.02	0.95	0.02	0.95	0.03	0.95	0.03	0.95	0.03	1.17	0.04
Naïve																	
Cond 1	0.92	0.92	0.05	0.90	0.06	0.95	0.03	0.94	0.04	0.95	0.03	0.95	0.03	0.96	0.03	1.24	0.07
Cond 2	0.90	0.90	0.07	0.88	0.08	0.95	0.03	0.95	0.03	0.95	0.03	0.95	0.04	0.95	0.03	1.17	0.03
Cond 3	0.87	0.87	0.08	0.85	0.10	0.96	0.02	0.96	0.02	0.95	0.03	0.95	0.03	0.95	0.03	1.17	0.04
Cond 4	0.85	0.85	0.10	0.83	0.11	0.96	0.02	0.98	0.01	0.95	0.03	0.94	0.04	0.95	0.03	1.14	0.03
Cond 5	0.83	0.83	0.11	0.80	0.13	0.98	0.01	0.99	0.01	0.95	0.03	0.95	0.03	0.95	0.03	1.02	0.01
Cond 6	0.81	0.81	0.13	0.77	0.15	1.00	0.00	1.00	0.00	0.95	0.03	0.95	0.03	0.95	0.03	1.15	0.04

Note. PBR indicates the average percent in bias reduction across 1,000 replications for each condition, *d* indicates the average standardized mean difference across 1,000 replications for each condition (control group subtracted from treatment group divided by the pooled standard deviation), and Var Ratio indicates the average variance ratio across 1,000 replications for each condition (treatment group's propensity score variance divided by the control group's propensity score variance). True in the table refers to the true propensity score, which included the error-free covariates (L1 and L2) in the covariate set. Naïve in the table refers to the naïve propensity score, which included the error-prone covariates (X1 and X2) in the covariate set. Cond indicates the corresponding measurement error condition.

Table 9

Numeric Diagnostic Information for Different Measurement Error Conditions After Nearest Neighbor Matching with a Caliper Using the True and Naïve Propensity Scores

	L1		L2		X1		X2		X3		X4		X5		Propensity Score Var		
	True	PBR	d	PBR	d	PBR	d	PBR	d	PBR	d	PBR	d	PBR	d	Ratio	d
Cond 1	1.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	0.99	0.00	1.00	0.00	1.00	0.00	1.00	0.00
Cond 2	1.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
Cond 3	1.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	0.99	0.01	1.00	0.00
Cond 4	1.00	1.00	0.00	1.00	0.00	0.99	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.01	0.00
Cond 5	1.00	1.00	0.00	0.99	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
Cond 6	1.00	1.00	0.00	1.00	0.00	1.00	0.00	0.99	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
Naïve																	
Cond 1	0.96	0.96	0.03	0.95	0.04	0.99	0.00	1.00	0.00	1.00	0.00	0.99	0.01	1.00	0.00	1.00	0.00
Cond 2	0.94	0.94	0.04	0.92	0.05	0.99	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
Cond 3	0.92	0.92	0.06	0.89	0.08	1.00	0.00	1.00	0.00	0.99	0.00	1.00	0.00	1.00	0.00	1.00	0.00
Cond 4	0.89	0.89	0.08	0.86	0.09	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.01	0.00
Cond 5	0.86	0.86	0.10	0.83	0.11	0.99	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
Cond 6	0.84	0.84	0.11	0.80	0.13	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00

Note. PBR indicates the average percent in bias reduction across 1,000 replications for each condition, d indicates the average standardized mean difference across 1,000 replications for each condition (control group subtracted from treatment group divided by the pooled standard deviation), and Var Ratio indicates the average variance ratio across 1,000 replications for each condition (treatment group's propensity score variance divided by the control group's propensity score variance). True in the table refers to the true propensity score, which included the error-free covariates (L1 and L2) in the covariate set. Naïve in the table refers to the naïve propensity score, which included the error-prone covariates (X1 and X2) in the covariate set. Cond indicates the corresponding measurement error condition.

Table 10

Numeric Diagnostic Information for Different Measurement Error Conditions After Optimal Matching Using the True and Naïve Propensity Scores

	L1		L2		X1		X2		X3		X4		X5		Propensity Score		
	True	PBR	<i>d</i>	PBR	<i>d</i>	PBR	<i>d</i>	PBR	<i>d</i>	PBR	<i>d</i>	PBR	<i>d</i>	PBR	<i>d</i>	Var Ratio	<i>d</i>
Cond 1	0.95	0.95	0.03	0.96	0.03	0.95	0.03	0.96	0.03	0.95	0.03	0.95	0.03	0.96	0.03	1.27	0.08
Cond 2	0.95	0.95	0.03	0.95	0.03	0.95	0.03	0.95	0.03	0.96	0.03	0.95	0.03	0.95	0.03	1.18	0.03
Cond 3	0.95	0.95	0.03	0.95	0.03	0.95	0.03	0.95	0.03	0.95	0.03	0.96	0.03	0.96	0.03	1.25	0.06
Cond 4	0.95	0.95	0.03	0.95	0.03	0.96	0.02	0.95	0.03	0.96	0.03	0.96	0.03	0.96	0.03	1.16	0.04
Cond 5	0.95	0.95	0.03	0.96	0.03	0.95	0.02	0.95	0.02	0.95	0.03	0.95	0.03	0.95	0.03	1.05	0.01
Cond 6	0.96	0.96	0.03	0.96	0.03	0.96	0.02	0.96	0.02	0.96	0.03	0.96	0.03	0.96	0.03	1.17	0.04
Naïve																	
Cond 1	0.93	0.93	0.05	0.91	0.06	0.95	0.03	0.94	0.03	0.96	0.03	0.96	0.03	0.96	0.03	1.24	0.07
Cond 2	0.90	0.90	0.07	0.88	0.08	0.95	0.03	0.95	0.03	0.96	0.03	0.96	0.03	0.96	0.03	1.17	0.03
Cond 3	0.88	0.88	0.08	0.85	0.10	0.96	0.02	0.96	0.02	0.96	0.03	0.96	0.03	0.96	0.03	1.17	0.04
Cond 4	0.86	0.86	0.09	0.83	0.11	0.97	0.02	0.98	0.01	0.96	0.03	0.96	0.03	0.96	0.03	1.13	0.03
Cond 5	0.84	0.84	0.11	0.80	0.13	0.99	0.01	1.00	0.00	0.96	0.03	0.96	0.03	0.95	0.03	1.01	0.00
Cond 6	0.81	0.81	0.12	0.78	0.15	1.00	0.00	1.01	0.00	0.96	0.03	0.96	0.03	0.96	0.03	1.15	0.04

Note. PBR indicates the average percent in bias reduction across 1,000 replications for each condition, *d* indicates the average standardized mean difference across 1,000 replications for each condition (control group subtracted from treatment group divided by the pooled standard deviation), and Var Ratio indicates the average variance ratio across 1,000 replications for each condition (treatment group's propensity score variance divided by the control group's propensity score variance). True in the table refers to the true propensity score, which included the error-free covariates (L1 and L2) in the covariate set. Naïve in the table refers to the naïve propensity score, which included the error-prone covariates (X1 and X2) in the covariate set. Cond indicates the corresponding measurement error condition.

Table 11
Numeric Diagnostic Information for Different Measurement Error Conditions After Mahalanobis Distance Matching Using the True and Naïve Propensity Scores

	L1		L2		X1		X2		X3		X4		X5		Propensity Score	
	PBR	<i>d</i>	PBR	<i>d</i>	PBR	<i>d</i>	PBR	<i>d</i>	PBR	<i>d</i>	PBR	<i>d</i>	PBR	<i>d</i>	Var Ratio	<i>d</i>
True																
Cond 1	0.87	0.09	0.87	0.09	0.87	0.08	0.87	0.08	0.87	0.09	0.87	0.09	0.87	0.09	1.26	0.15
Cond 2	0.87	0.09	0.87	0.09	0.87	0.08	0.87	0.08	0.87	0.09	0.87	0.09	0.87	0.09	1.33	0.10
Cond 3	0.87	0.09	0.87	0.09	0.87	0.08	0.87	0.07	0.87	0.09	0.87	0.09	0.87	0.09	1.34	0.17
Cond 4	0.87	0.09	0.87	0.09	0.87	0.07	0.87	0.07	0.87	0.09	0.87	0.09	0.87	0.09	1.19	0.13
Cond 5	0.87	0.09	0.87	0.09	0.87	0.07	0.86	0.07	0.87	0.09	0.87	0.09	0.87	0.09	1.22	0.08
Cond 6	0.87	0.09	0.87	0.09	0.87	0.06	0.87	0.06	0.87	0.09	0.87	0.09	0.87	0.09	1.30	0.10
Naïve																
Cond 1	0.84	0.11	0.83	0.12	0.86	0.08	0.86	0.09	0.87	0.09	0.87	0.09	0.87	0.09	1.19	0.15
Cond 2	0.82	0.12	0.80	0.14	0.87	0.08	0.87	0.08	0.87	0.09	0.87	0.09	0.87	0.09	1.25	0.14
Cond 3	0.80	0.13	0.78	0.15	0.88	0.07	0.88	0.07	0.87	0.09	0.87	0.09	0.87	0.09	1.30	0.13
Cond 4	0.78	0.15	0.76	0.17	0.89	0.07	0.89	0.06	0.87	0.09	0.87	0.09	0.87	0.09	1.16	0.10
Cond 5	0.75	0.17	0.72	0.19	0.89	0.06	0.90	0.06	0.87	0.09	0.87	0.09	0.87	0.09	1.13	0.07
Cond 6	0.74	0.18	0.70	0.20	0.90	0.06	0.91	0.05	0.87	0.09	0.87	0.09	0.87	0.09	1.33	0.13

Note. PBR indicates the average percent in bias reduction across 1,000 replications for each condition, *d* indicates the average standardized mean difference across 1,000 replications for each condition (control group subtracted from treatment group divided by the pooled standard deviation), and Var Ratio indicates the average variance ratio across 1,000 replications for each condition (treatment group's propensity score variance divided by the control group's propensity score variance). True in the table refers to the true propensity score, which included the error-free covariates (L1 and L2) in the covariate set. Naïve in the table refers to the naïve propensity score, which included the error-prone covariates (X1 and X2) in the covariate set. Cond indicates the corresponding measurement error condition.

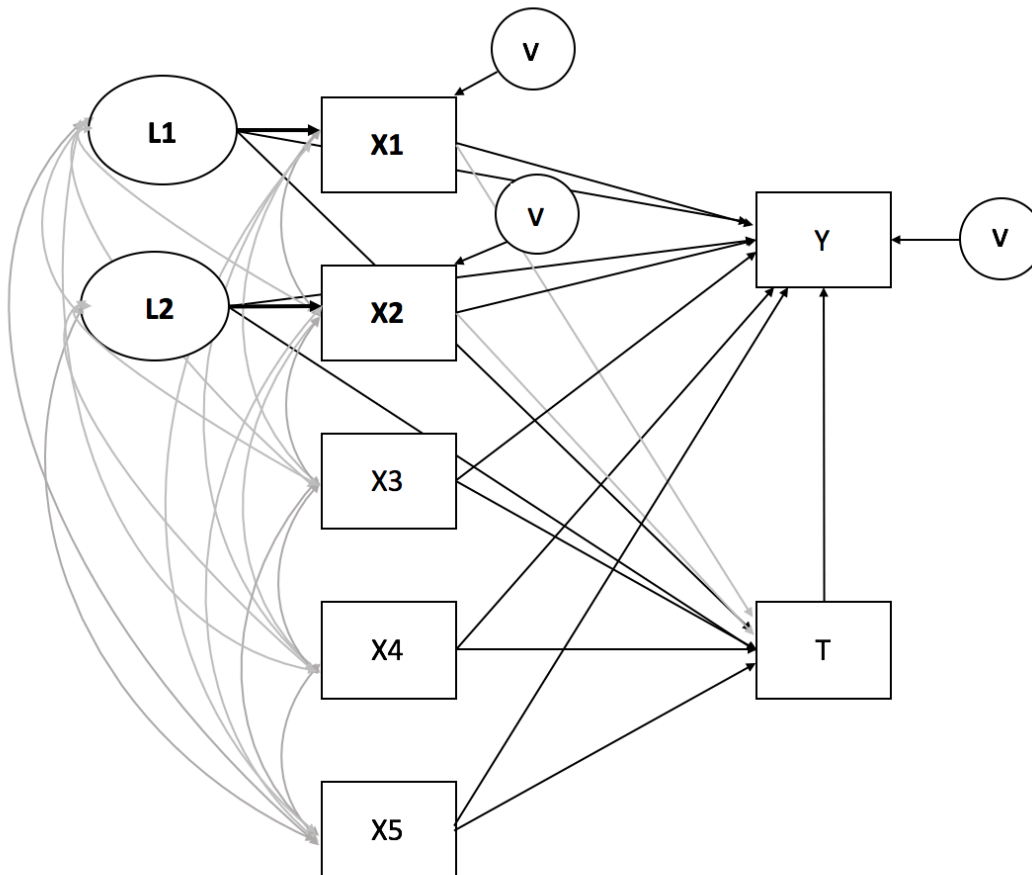


Figure 1. Conceptual diagram of latent and observed simulated variables where L1 and L2 are latent variables related to self-selection into treatment (T) and the outcome (Y), X1 and X2 are observed error-prone measures of L1 and L2, X3-X5 are observed error-free covariates, and V is a random disturbance term with a mean of zero and a standard deviation of one. Note the grey lines indicate correlations or indirect paths among the simulated covariates and black lines indicate the direct paths between simulated variables.

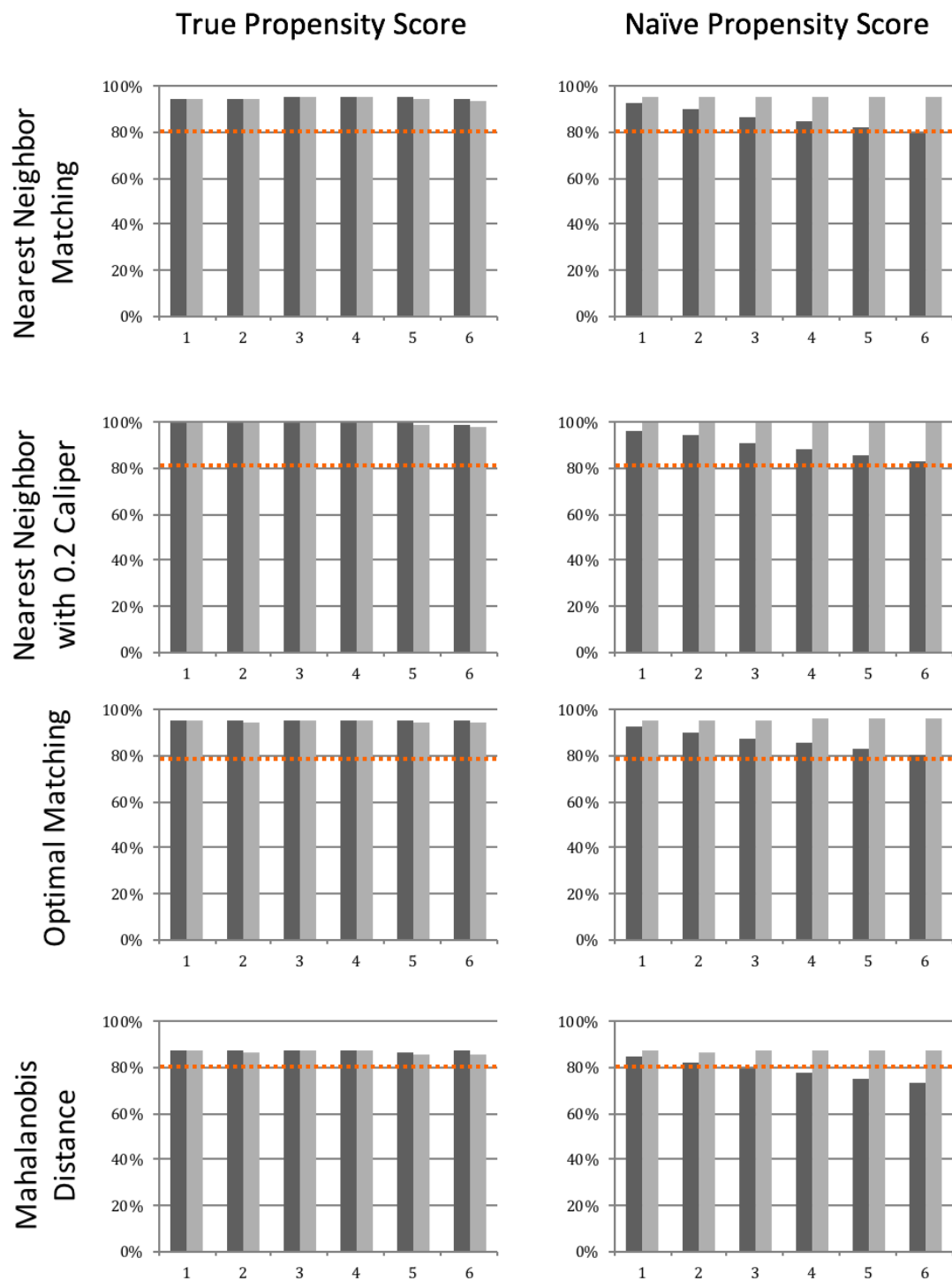


Figure 2. Plots of the average percent of bias reduction in L1 (dark grey) and X1 (light grey) for the same measurement error conditions. Note that the red dashed line indicates the 80% reduction in bias benchmark recommended in the literature (Pan & Bai, 2015).

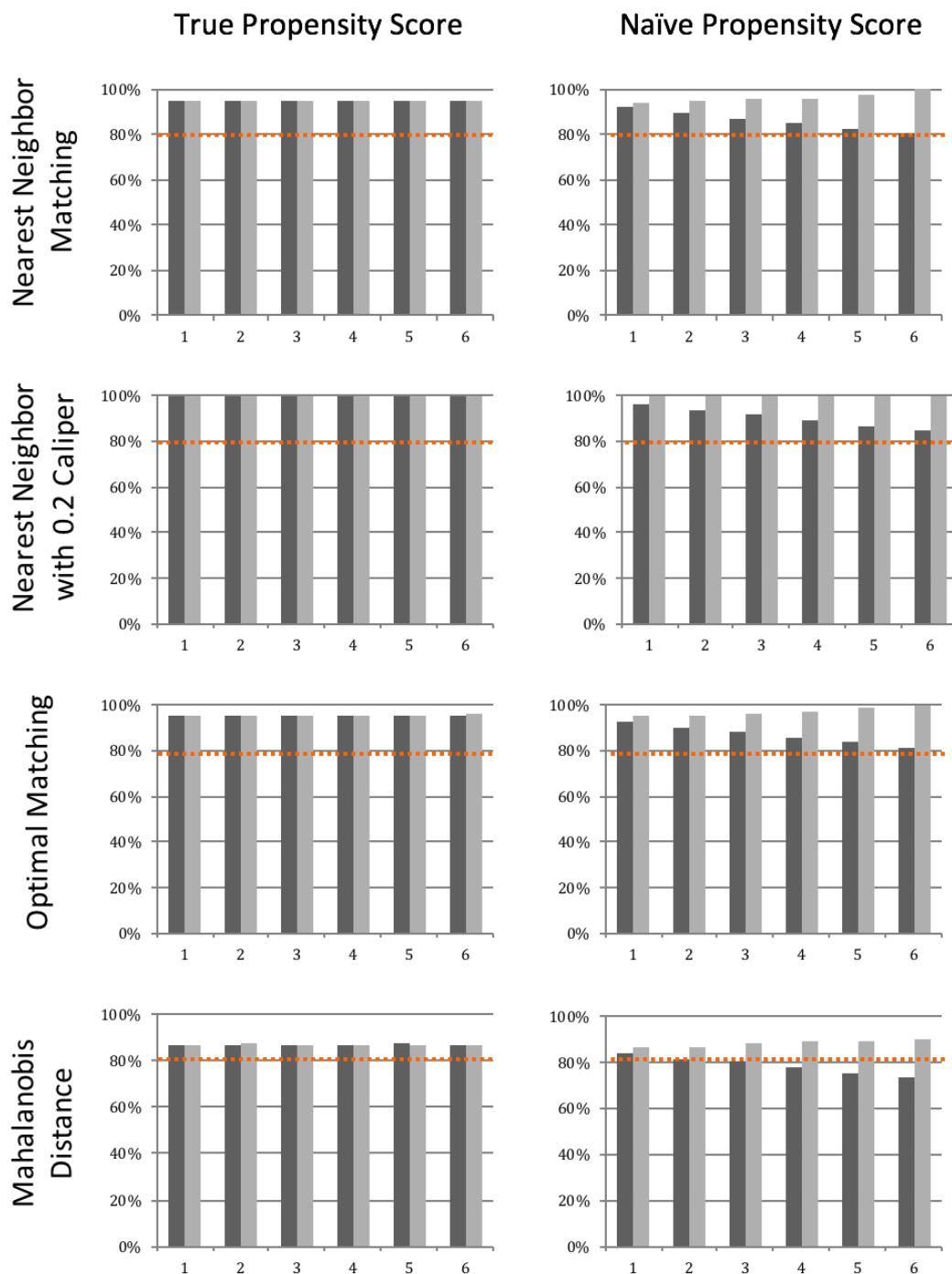


Figure 3. Plots of the average percent of bias reduction in L1 (dark grey) and X1 (light grey) for the different measurement error conditions. Note that the red dashed line indicates the 80% reduction in bias benchmark recommended in the literature (Pan & Bai, 2015).

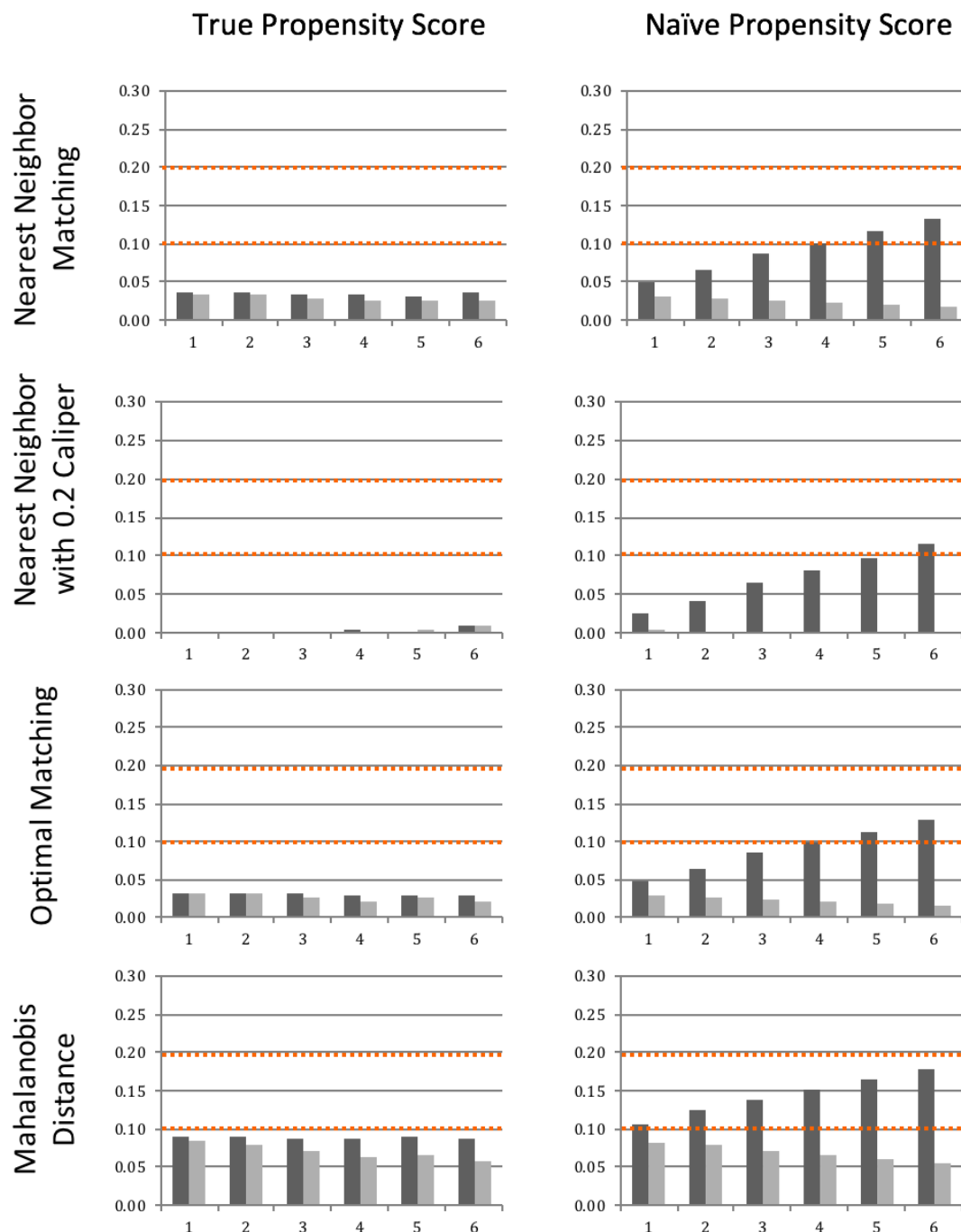


Figure 4. Plots of the standardized mean difference for L1 (dark grey) and X1 (light grey) for the same measurement error conditions. Note that the bottom red dashed line indicates the 0.10 benchmark (Ho et al., 2007) and the top red dashed line indicates the 0.2 benchmark (Austin, 2011) recommended in the literature.

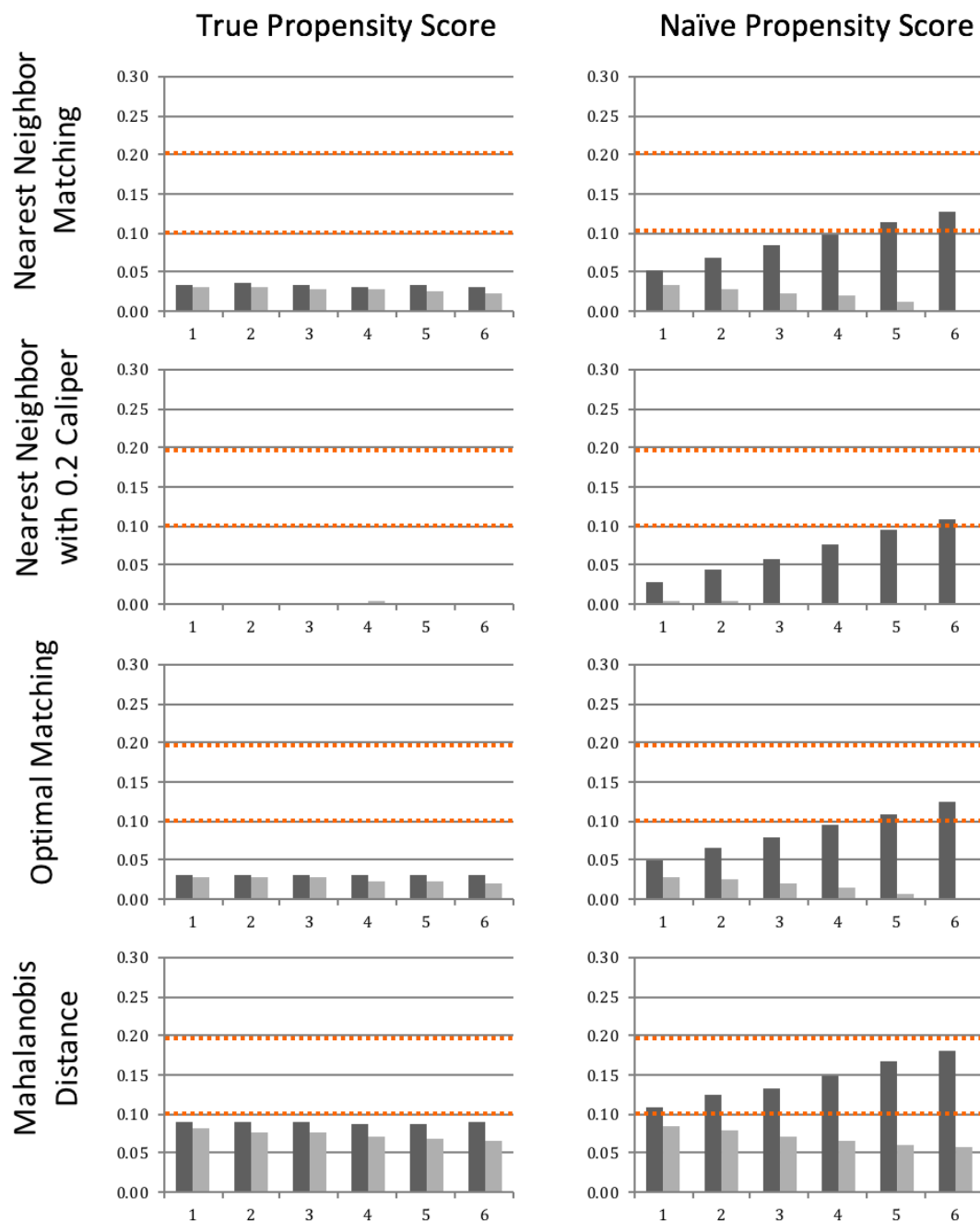


Figure 5. Plots of the standardized mean difference for L1 (dark grey) and X1 (light grey) for the different measurement error conditions. Note that the bottom red dashed line indicates the 0.10 benchmark (Ho et al., 2007) and the top red dashed line indicates the 0.2 benchmark (Austin, 2011) recommended in the literature.