Published in final edited form as:
*Am J Epidemiol*. 2006 June 15; 163(12): 1149–1156.

# Variable selection for propensity score models.

**M. Alan Brookhart**[1], **Sebastian Schneeweiss**[1], **Kenneth J. Rothman**[1,2], **Robert J. Glynn**[1,3], **Jerry Avorn**[1], and **Til Stürmer**[1]

1 *Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital & Harvard Medical School, Boston, MA*

2 *Departments of Epidemiology and Medicine, Boston University Medical Center, Boston, MA*

3 *Division of Preventive Medicine, Department of Medicine, Brigham and Women's Hospital & Harvard Medical School, Boston, MA*

## Abstract

Despite the growing popularity of propensity score (PS) methods in epidemiology, relatively little has been written in the epidemiologic literature about the problem of variable selection in PS models. The authors present the results of two simulation studies designed to help epidemiologists gain insight into the variable selection problem in a PS analysis. The simulation studies illustrate how the choice of variables that are included in a PS model can affect the bias, variance and mean-squared error of an estimated exposure effect. The results suggest that variables that are unrelated to the exposure but related to the outcome should always be included in a PS model. The inclusion of these variables will increase the precision of the estimated exposure effect without increasing bias. In contrast, including variables that are related to the exposure but not the outcome will decrease the precision of the estimated exposure effect without decreasing bias. In small studies, the inclusion of variables that are strongly related to the exposure but only weakly related to the outcome can be detrimental to an estimate in a mean-squared error sense. The addition of these variables removes only a small amount of bias but can strongly decrease the precision of the estimated exposure effect. These simulation studies and other analytical results suggest that standard model building tools designed to create good predictive models of the exposure will not necessarily lead to optimal PS models, particularly in the setting of small samples.

### Keywords

confounding; propensity scores; stepwise regression; variable selection; subset selection; simulation study

---

Confounding in non-experimental studies can occur when baseline covariates that predict the exposure under study are independently related to the outcome of interest. In the presence of confounding, any marginal association between the exposure and outcome can at least partly be attributed to the confounder. When the exposure is dichotomous, one approach that can be used to control confounding is the method of propensity scores (PS) as formalized by Rosenbaum and Rubin [1]. PS methods depend on a model of the conditional probability of exposure given the confounders. Ideally, specification of the PS model will be driven by subject matter knowledge, e.g., a detailed understanding of how a particular treatment is assigned to a patient. Typically, however, the researcher does not have the benefit of such knowledge and instead is confronted with a large collection of pre-treatment covariates and many derived

---

Address Correspondence to: M. Alan Brookhart, Ph.D., Division of Pharmacoepidemiology and Pharmacoeconomics, 1620 Tremont Street, Suite 3030, Boston, MA 02120 Phone: (617) 278-0335.

Support: This project was funded by a grant (R01 AG023178) from the National Institute on Aging.

functions of these covariates (e.g., interactions, quadratic terms, log transformations) and must decide which of these terms to enter into a regression model of the exposure. The bias and variance of the estimated exposure effect can depend strongly on which of these candidate variables are included in the PS model.

Despite their growing popularity, relatively little has been written about variable selection strategies for PS models. In the context of multivariate normal confounders, Rubin and Thomas derived approximations for the reduction in the bias and variance of an estimated exposure effect from a PS matched analysis [2]. In this paper, the authors suggest including in a PS model all variables thought to be related to the outcome, whether or not they are related to exposure. In a later paper, Rubin suggests that including variables that are strongly related to exposure, but unrelated to the outcome, can decrease the efficiency of an estimated exposure effect; but he argues that if such a variable had even a weak effect on the outcome, the bias resulting from its exclusion would dominate any loss of efficiency for a reasonably sized study [3]. Some of these guidelines are repeated in Perkins et. al. [4]. Robins et al. derived analytic results showing that the asymptotic variance of an estimator based on an exposure model is not increased and often decreased as the number of parameters in the exposure model is increased [5]. These results suggest that the size of a PS model should increase with the study size. Hirano and Imbens proposed a variable selection strategy for use with a multivariate outcome model employing propensity score weighting [6].

In practice, variables are often selected in data-driven ways, for example, by using stepwise variable selection algorithms to develop good predictive models of the exposure [7], [8]. Furthermore, many PS analyses report the *AUC* or *c* statistic (area under the receiver operating characteristic curve) of the final PS model as a means of assessing the model's adequacy [7]. Implicit in this practice is the assumption that PS models that are better predictors or discriminators of the exposure status result in superior estimators of exposure effect. According to this criterion, any variable that increases the *c* statistic or predictive ability of the PS model should be selected for inclusion in the model. Despite the widespread use of such variable selection strategies, there has been little discussion of their appropriateness. In a recent editorial, Rubin expressed doubt over the usefulness of such diagnostics in a PS analysis [9].

The present work was conducted to illuminate this issue and to help researchers gain some practical insight into the variable selection problem in a propensity score analysis. We present the results of two Monte-Carlo simulation experiments designed to evaluate how different specifications of a PS model affect the bias, variance and resulting mean squared error of an estimated exposure effect under a variety of assumptions about the data generating process.

## METHODS

### Overview of Propensity Score Methods in Non-experimental Cohort Studies

Propensity score methods are designed to estimate the effect of a dichotomous exposure $A$ on an outcome $Y$ that is not confounded by a set of measured covariates $X = (X_1, X_2, \ldots, X_p)$. As potential confounders, the elements of $X$ can be both predictors of the exposure and independent risk factors for the outcome. As an illustration, we can consider a cohort study in which the exposure of interest is the use of a particular cholesterol lowering drug at the start of the study and the outcome is a myocardial infarction (MI) within one year. Baseline confounders could include age, gender, history of MI, previous drug exposures, and various comorbid conditions.

A propensity score is the conditional probability that a subject receives a treatment or exposure under study given all measured confounders, i.e., $Pr[A = 1|X_1, X_2, \ldots, X_p]$. The propensity score has been termed a balancing score, meaning that among subjects with the same propensity to be exposed, treatment is conditionally independent of the covariates [1]. This property

suggests that estimates of the exposure effect that are not confounded by any of the measured covariates can be obtained by estimating the effect of exposure within groups of people with the same propensity score. Within such a group, any difference in the outcome between the exposed and unexposed subjects is not attributable to the measured confounders. When treatment assignment is strongly ignorable and other specific assumptions hold, estimates derived from a propensity score analysis can be interpreted causally [1].

In non-experimental research the true PS will not usually be known and, therefore, will need to be estimated, typically according to an assumed model. How the model for $Pr[A|X]$ is specified has the potential to affect the bias and variance of the estimated exposure effect. Given an estimated PS, exposure effects are usually estimated by either matching on the PS to create two comparable groups, including the PS and the exposure in a multivariate model of the outcome under study, or conducting an analysis stratified on the PS. It is also possible to fit a weighted regression using inverse-probability of exposure weights generated from the estimated PS [10]. A more detailed discussion and review of PS methods can be found elsewhere [1], [11], [12].

## Monte-Carlo Simulation Study

We performed two Monte-Carlo simulation experiments. The first examined how the inclusion of three different types of covariates in a PS model affected the estimated exposure effect (see figure 1):

1. a variable related to both outcome and exposure, a true confounder ($X_1$),
2. a variable related to the outcome but not the exposure ($X_2$),
3. and a variable related to the exposure but not the outcome ($X_3$).

In the second experiment, we considered how the addition of a single confounder to a PS model changes the bias and variance of an estimated exposure effect under varying assumptions about the strength of the confounder-outcome and confounder-exposure relations.

Both simulation experiments employ the same basic data generating mechanism. The simulated data consisted of realizations of a dichotomous exposure, a Poisson distributed count outcome, and continuous confounders. The data were generated in the following order according to the specified probability models:

- The covariates $X_1, X_2, X_3$ are independent standard normal random variables with mean 0 and unit variance.
- The conditional distribution of the dichotomous exposure $A$ given $X_1, X_2, X_3$ follows a Bernoulli distribution with a conditional mean given by the function

$$Pr[A = 1 \mid X_1, X_2, X_3] = \Phi(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3).$$

where $\Phi$ is the standard Normal distribution function.

- The conditional distribution of $Y$ given $X$ and $A$ follows a Poisson distribution with two possible specifications of the mean. The first specification (used in the first simulation experiment) is given by

$$E[Y \mid A, X_1, X_2, X_3] = \exp\{\alpha_0 + \alpha_1((1 + \exp(-3 * X_1))^{-1} - 0.5) + \alpha_2 X_2 + \alpha_3 X + \alpha_4 A\}.$$

This specification creates a non-linear (S-shaped) relationship between the confounder $X_1$ and the log of the expected value of the outcome. The second specification is given by

$$E[Y \mid A,\ X_1,\ X_2,\ X_3] = \exp\left\{\alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \alpha_4 A\right\}.$$

This model specifies a standard log-linear relationship between the covariates and the expected value of the outcome.

Within both simulation experiments, the effect of exposure is held constant ($\alpha_4 = 0.5$). The simulations differ in how the covariates are related to the exposure and outcome.

We considered two approaches to controlling for the propensity score. In the first, the exposure effects were estimated by adjusting for the PS in a multivariate outcome model in which the effect of the estimated PS was flexibly modeled through a cubic regression spline with three interior knot points placed at quartiles of the estimated propensity score. The fitted model is given by

$$E[Y \mid \hat{PS},\ A] = \exp\left\{\lambda + \sum_k \psi_k B_k(\hat{PS}) + \gamma A\right\}.$$

where $\lambda$ is the baseline rate, the $B_k$ are the $B$-spline basis functions [13], and $\gamma$ is the treatment effect. The second approach that we employed was based on sub-classification. Exposure effects were estimated within strata defined by quintiles of the estimated propensity score.

The simulation studies presented in this paper compare the performance of various specifications of PS models. To evaluate each PS model, we use the simulation results to determine the variance, bias and mean-squared error of the corresponding estimators. Because we have used a log-linear model of the outcome, the parameter estimate $\hat{\gamma}$ is consistent for the parameter $\alpha_4$ from our data generating distribution at the true propensity score [14]. Therefore, we can estimate the bias of a given estimator with

$$\hat{BIAS} = \frac{1}{S}\sum_{s=1}^{S}(\hat{\gamma}(s) - \alpha_4),$$

and its mean-squared error with

$$\hat{MSE} = \frac{1}{S}\sum_{s=1}^{S}(\hat{\gamma}(s) - \alpha_4)^2,$$

where $\hat{\gamma}(s)$ is the estimated effect of exposure in the $s^{th}$ simulated data set according to a particular PS model and $S$ is the total number of simulations.

**Simulation experiment 1—**For this experiment, exposure was confounded through $X_1$, $X_3$ predicted treatment but was unrelated to the outcome, and $X_2$ predicted the outcome but was unrelated to treatment ($\alpha_0 = 0.5$, $\alpha_1 = 4$, $\alpha_2 = 1$, $\alpha_3 = 0$, $\beta_0 = 0$, $\beta_1 = 0.5$, $\beta_2 = 0$, $\beta_3 = 0.75$). This scenario is depicted graphically in figure 1.

We simulated 1000 data sets for both n=500 and n=2500. For each simulated data set, we estimated seven different propensity scores corresponding to all possible combinations of ($X_1$, $X_2$, $X_3$) in a probit regression model. These models are given by

- **PS Model 1: Pr**$[A = 1|X] = \Phi(\beta_0 + \beta_1 X_1)$.
- **PS Model 2: Pr**$[A = 1|X] = \Phi(\beta_0 + \beta_1 X_2)$.
- **PS Model 3: Pr**$[A = 1|X] = \Phi(\beta_0 + \beta_1 X_3)$.
- **PS Model 4: Pr**$[A = 1|X] = \Phi(\beta_0 + \beta_1 X_1 + \beta_2 X_2)$.

- **PS Model 5: Pr**$[A = 1|X] = \Phi(\beta_0 + \beta_1 X_1 + \beta_2 X_3)$.
- **PS Model 6: Pr**$[A = 1|X] = \Phi(\beta_0 + \beta_1 X_2 + \beta_2 X_3)$.
- **PS Model 7: Pr**$[A = 1|X] = \Phi(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3)$.

We also report the estimated bias, variance, and MSE of an estimator corresponding to the crude log relative rate (RR) and the average area under the receiver operating characteristic (ROC) curve (*AUC* or *c* statistic) for each PS model considered.

We conducted a variety of sensitivity analyses with $n = 500$. These were done by holding all parameters at their default value while a single parameter was altered. The following sensitivity analyses were performed: standard deviation of each covariate was both increased and decreased by 50%; the treatment effect was decreased to $\alpha_4 = 0.25$ and increased to $\alpha_4 = 1$; and the baseline prevalence of the exposure was decreased from approximately 50% to approximately 20% ($\beta_0 = -1$).

**Simulation experiment 2—**The second simulation experiment examined how the inclusion of a single true confounder in a PS model affected the bias and variance of an estimated exposure effect under varying assumptions about the strength of association between the single confounder and both the outcome and exposure. For each simulated data set two estimators were considered: the first was derived from the crude log relative rate and the second was derived from a PS adjusted estimate of the effect of $A$ on $Y$ in which the PS model contained only the confounder $X_1$. In this simulation experiment, the adjustment for the PS used the spline approach. We denote the crude estimator of the log relative rate with $\hat{\gamma}_0$ and the PS adjusted estimator with $\hat{\gamma}_1$.

The parameter $\alpha_1$, the strength of association between $X_1$ and $Y$, took values in $\{0, 0.01, \ldots, 0.20\}$ corresponding to relative rates ranging from 1.00 to 1.28. The parameter $\beta_1$, the strength of association between $X_1$ and $A$, took values in $\{0.00, 0.05, \ldots, 1.25\}$. For all possible combinations of these values of $\alpha_1$ and $\beta_1$, we simulated 1000 data sets of $n = 500$ and $n = 2500$. In this simulation, the covariates $X_2$ and $X_3$ are not used. For each set of 1000 data sets we computed the estimated bias, variance, and MSE of each of the two estimators.

**Computation—**All simulations were performed in R version 1.9.1 [15], [16] running on a Windows XP platform using software created by the authors.

## RESULTS

### Simulation Experiment 1

For the simulations controlling for the PS through a spline, we report the estimated bias, variance, and MSE of all estimators in Table 1. We also report the average *c*-statistic for each candidate PS model. The sole confounder was the covariate $X_1$, therefore any estimator that did not contain $X_1$ in the PS model was biased. For both study sizes, the unbiased estimator with the smallest variance was the one that contained the covariates $X_1$ and $X_2$. This estimator had more than 40% less variance than the estimator containing just the confounder $X_1$. Adding $X_3$, the covariate related only to exposure, increased the variance of the estimated effect. The estimator with all covariates in the PS model had a variance that was approximately 40% greater (for both study sizes) than the estimator with just the covariates $X_1$ and $X_2$. The *c*-statistic of the PS model with $X_1$ and $X_2$ was smaller (0.67) than *c*-statistic of the less efficient PS model with all covariates (0.80). For both study sizes, the PS models with the highest average *c* statistic contained all variables related to the exposure.

In Table 2, we report the results when this simulation experiment was repeated using sub-classification instead of spline adjustment. The results are qualitatively similar. In this simulation experiment, all estimators admit some bias due to some residual confounding within strata of the propensity score. However, the variance and MSE of all estimators was smaller than the corresponding estimator based on spline adjustment.

The results of the sensitivity analysis are presented in Table 3. In all of the sensitivity analyses the same essential pattern prevailed: the inclusion of the variable related only to exposure increased the variance of the estimator without altering bias, inclusion of the variable related only to the outcome decreased variance without affecting bias, and failure to include the confounder yielded a biased estimator. However, the perturbation of simulation parameters changed absolute and, in some cases, relative numbers.

### Simulation Experiment 2

In figure 2, we plot the estimated variance of the PS adjusted estimator $\hat{\gamma}_1$ and the unadjusted estimator $\hat{\gamma}_0$ across values of $\beta_1$ for both $n = 500$ and $n = 2500$. We transform the value of $\beta_1$ into a risk difference. This is done by computing the probability of treatment difference between $X_1 = 1$ and $X_1 = -1$. In others words the probability of treatment for someone with a moderately large value of $X_1$ (at the $84^{th}$ quantile) minus the probability of treatment for someone with a moderately small value of $X_1$ (at the $16^{th}$ quantile). For both sample sizes, increasing the value of $\beta_1$ (i.e., increasing the strength of association between $X_1$ and $A$) increased the variability of the estimated exposure effect $\hat{\gamma}_1$ (the PS adjusted estimator). The increase in variance did not depend on the strength of association between $X_1$ and $Y$ (data not presented). The bias of $\hat{\gamma}_0$ increased as the association between either $X_1$ and $Y$ or $X_1$ and $A$ increased, unless there was no association between either $X_1$ and $A$ or between $X_1$ and $Y$.

In figure 3, we plot contours of the MSE of $\hat{\gamma}_1$ relative to the MSE of $\hat{\gamma}_0$ on a grid of values of $\alpha_1$ and $\beta_1$. The values of $\beta_1$ are transformed into a risk difference as described previously. This plot indicates values of $\alpha_1$ and $\beta_1$ for which the addition of the confounder $X_1$ to a PS model is detrimental in a MSE sense, i.e., the MSE of $\hat{\gamma}_1$ is greater than $\hat{\gamma}_0$. The region between the contour lines at 0.95 and 1.05 represents an indifference zone for which the analyst concerned with minimizing the MSE might be indifferent about adding $X_1$ to a PS model since the effect on MSE would be small. The region above and to the left of the contour line at 1.05 indicates the region where the analyst might chose to exclude $X_1$ from the PS as it would increase the MSE of the estimated exposure effect by more than 5%. This region is characterized by large values of $\beta_1$ (strong association between $X_1$ and $A$) and small values of $\alpha_1$ (weak association between $X_1$ and $Y$ ). Here the increase in variance of $\hat{\gamma}_1$ is not offset by a large enough decrease in bias to reduce the MSE of $\hat{\gamma}_1$ relative to $\hat{\gamma}_0$. Similarly, the region below and to the right of the contour line at 0.95 would represents the region where the analyst would want to add the confounder to the PS as it would decrease the MSE by more than 5%. Here the bias of an estimator excluding $X_1$ overwhelms any resulting increase in variance. For $n = 2500$ the same pattern prevailed, but the region for which $\hat{\gamma}_0$ yielded a smaller MSE than $\hat{\gamma}_1$ was reduced.

## DISCUSSION

Our first simulation experiment revealed that the model that best predicted exposure (as measured by a *c*-statistic) did not yield the optimal PS model (in terms of MSE). The optimal model was the one that included the confounder and the variable related only to the outcome. This finding is consistent with the advice of Rubin and Thomas [2], i.e., that one should include

in a PS model variables that are thought to be related to the outcome, regardless of whether they are related to the exposure. This result may run counter to intuition for many people. One might wonder why a PS model should include a variable that is unrelated to exposure. The answer is that even if a covariate is theoretically unassociated with exposure, there can be some slight chance relation between the covariate and the exposure for any given realization of a data set. If that covariate is also related to the outcome, then it is a empirical confounder for that particular data set. Including such a covariate in a PS model corrects for small amounts of chance bias or empirical confounding existing within each realization of the data set, thereby improving the precision of the estimator. This finding is related to the result that it is better to use an estimated rather than a known PS [17], [5].

This simulation study also revealed that if variables unrelated to the outcome but related to the exposure are added to a PS model they will increase the variance of an estimated exposure effect without decreasing its bias. Adding strong predictors of exposure to the PS model increases the variability of the estimated PS. If these added variables are unrelated to the outcome, then the variation they induce in the PS is not correcting confounding and is therefore only adding noise the estimated exposure effect. This result also suggests that there is little risk in adding a variable unrelated to exposure to a PS model. If the included covariate is unrelated to the outcome, it will affect neither the bias nor the variance of the estimator, but if it is related to the outcome, it can improve efficiency.

The second simulation experiment revealed that if one seeks to minimize the MSE of an estimate, then in small studies there are situations in which it might be advantageous to exclude true confounders from a PS model. This occurs when a covariate is only weakly related to the outcome, but very strongly related to the exposure. The loss in efficiency due to the inclusion of such a covariate is not offset by a large enough decrease in bias. However, as the study size increases, the variance of the estimator decreases at a rate proportional to $1 = n$, yet the bias due to an omitted confounder remains. Therefore, in large studies one would probably not want to exclude any covariate related exposure from a PS model, unless it was known to be completely unrelated to the outcome.

Although the results presented in this paper are consistent with theoretical results (e.g., [2]), the specific numbers are highly dependent on the specification of the data generating mechanism and the choice of parameter values considered. Through sensitivity analysis we varied the parameters that seemed to be the most relevant, however, the probability distributions and other structural elements of the study (e.g., using only three covariates, assuming a homogeneous exposure effect) remained unaltered. It is also important to point out that matching and other PS methods can be used in conjunction with standard multivariate regression models containing additional covariates [18]. The variable selection problem in these situations is more complex, as variables can appear in the PS model, the outcome model, or both. The results presented in this paper do not offer insight into the variable selection problem for these hybrid analytic methods.

Our findings and the analytical results in [2] and [5] raise questions about the appropriateness of standard model building strategies for the construction of PS models. Iterative stepwise model-building algorithms (e.g., forward stepwise regression) are designed to create good predictive models of exposure. Similarly, the *c* statistic, commonly used to asses the quality of a PS model, is a measure of the predictive ability of the model. The goal of a PS model is to efficiently control confounding, not to predict treatment or exposure. A variable selection criterion based on prediction of the exposure will miss variables related only to the outcome and could miss important confounders that have a weak relationship to the exposure, but a strong relationship to the outcome. Future work in this area should focus on identifying and evaluating practical strategies or rules of thumb that practitioners can use to help them select

variables for inclusion in a propensity score model with an aim of decreasing both the bias and variance of an estimated exposure effect.

## References


1. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika 1983;79:516–524.
2. Rubin DB, Thomas N. Matching using estimated propensity score: relating theory to practice. Biometrics 1996;52:249–264. [PubMed: 8934595]
3. Rubin DB. Estimating causal effects from large data sets using the propensity score. Ann Intern Med 1997;127:757–763. [PubMed: 9382394]
4. Perkins SM, Tu W, Underhill MG, Zhou XH, Murray MD. The use of propensity scores in pharmacoepidemiologic research. Phamacoepidemiolog Drug Saf 2000;9:93–101.
5. Robins JM, Mark SD, Newey WK. Estimating exposure effects by modelling the expectation of exposure conditional on confounders. Biometrics 1992;48:479–495. [PubMed: 1637973]
6. Hirano K, Imbens G. Estimation of causal effects using propensity score weighting: An application to data on right heart catheterization. Health Services and Outcome Research Methodology 2001;2:259–278.
7. Weitzen S, Lapane KL, Toledano AY, Hume AL, Mor V. Principles for modeling propensity scores in medical research: a systematic literature review. Phamacoepidemiol Drug Saf 2004;13:841–853.
8. Stürmer T, Joshi M, Glynn RJ, Avorn J, Rothman KJ, Schneeweiss S. A review of the application of propensity score methods yielded increasing use, advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods. J Clin Epidemiol, in press.
9. Rubin DB. On principles for modeling propensity score in medical research. Phamacoepidemiol Drug Saf 2005;14:227–238.
10. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology 2000;11:550–560. [PubMed: 10955408]
11. Joffe MM, Rosenbaum PR. Invited commentary: Propensity scores. Am J Epidemiol 1999;150:327–333. [PubMed: 10453808]
12. D'Agostino RB Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. Stat Med 1998;17:2265–2281. [PubMed: 9802183]
13. Hastie TJ, Tibshirani RJ. Generalized Additive Models. London: Chapman and Hall, 1996.
14. Gail MH, Wieand S, Piantadosi S. Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. Biometrika 1984;71:431–444.
15. Ihaka R, Gentleman RR. A Language for Data Analysis and Graphics. Journal of Computational and Graphical Statistics 1996;5:299–314.
16. R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2003; ISBN 3-900051-00-3, http://www.R-project.org
17. Rosenbaum PR. Model-based direct adjustment. J Am Stat Assoc 1987;82:387–394.
18. Cochran W, Rubin DB. Controlling bias in observational studies: A Review. Sankhya 1973; 417–46.

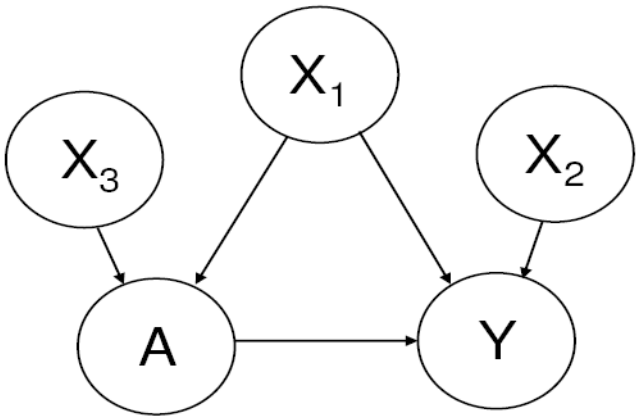


**Figure 1.**
The causal diagram for Simulation Experiment 1.

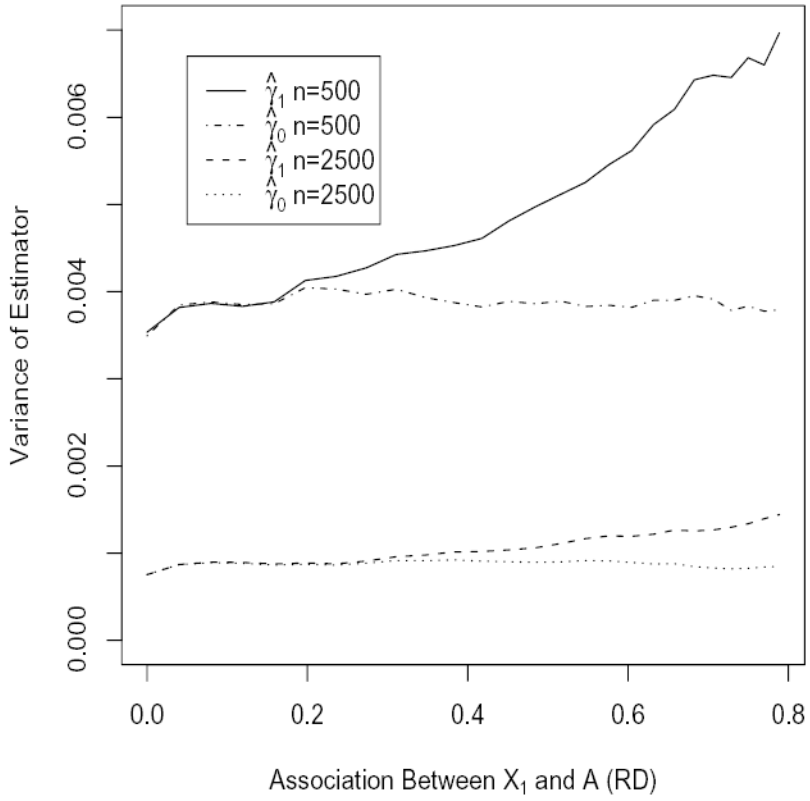


**Figure 2.**
Variance of unadjusted estimator $\hat{\gamma}_0$ and PS adjusted estimator $\hat{\gamma}_1$ for different values of $\beta_1$ for $n = 500$ and $n = 2500$.

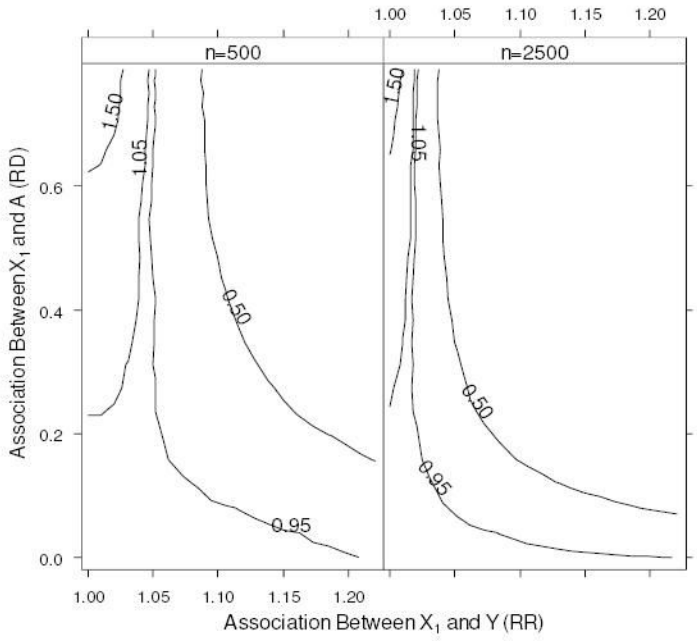


**Figure 3.**
Contours of the MSE of the PS adjusted estimator relative to the unadjusted estimator, $\text{MSE}(\hat{\gamma}_1) \,/\, \text{MSE}(\hat{\gamma}_0)$.

**Table 1**

Simulation Experiment 1: results are based on an analysis in which the PS is controlled for in a multivariate model using a parametric spline. We report the estimated bias, variance, and mean-squared error (MSE) of all possible estimators and the average *C* statistic of the corresponding PS model.

| | | Variables in Propensity Score Model | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | $X_1$ | $X_2$ | $X_3$ | $X_1, X_2$ | $X_1, X_3$ | $X_2, X_3$ | $X_1, X_2, X_3$ | None |
| n=500 | $\hat{BIAS} \times 10^1$ | −0.03 | 5.97 | 7.34 | −0.03 | −0.07 | 7.36 | −0.06 | 5.94 |
| | $\hat{VAR} \times 10^1$ | 0.32 | 0.22 | 0.46 | 0.22 | 0.44 | 0.36 | 0.31 | 0.39 |
| | $\hat{MSE} \times 10^1$ | 0.32 | 3.79 | 5.85 | 0.22 | 0.44 | 5.77 | 0.31 | 3.92 |
| | Avg. *C*-stat. | 0.67 | 0.52 | 0.76 | 0.67 | 0.82 | 0.76 | 0.82 | |
| n=2500 | $\hat{BIAS} \times 10^1$ | 0.00 | 5.93 | 7.33 | −0.01 | −0.04 | 7.33 | −0.03 | 5.95 |
| | $\hat{VAR} \times 10^2$ | 0.66 | 0.53 | 0.96 | 0.49 | 0.89 | 0.79 | 0.69 | 0.80 |
| | $\hat{MSE} \times 10^2$ | 0.66 | 35.65 | 54.72 | 0.49 | 0.89 | 54.56 | 0.70 | 36.16 |
| | Avg. *C*-stat. | 0.67 | 0.51 | 0.76 | 0.67 | 0.81 | 0.76 | 0.81 | |

**Table 2**

Simulation Experiment 1: results are based on an analysis using subclassification in which strata are defined by quintiles of the estimated PS. We report the estimated bias, variance, and mean-squared error (MSE) of all possible estimators.

| | | Variables in Propensity Score Model | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | $X_1$ | $X_2$ | $X_3$ | $X_1, X_2$ | $X_1, X_3$ | $X_2, X_3$ | $X_1, X_2, X_3$ | None |
| n=500 | $\widehat{BIAS} \times 10^1$ | 0.29 | 6.07 | 7.96 | 0.24 | 0.24 | 7.93 | 0.24 | 5.94 |
| | $\widehat{VAR} \times 10^1$ | 0.22 | 0.14 | 0.62 | 0.16 | 0.71 | 0.43 | 0.69 | 0.39 |
| | $\widehat{MSE} \times 10^1$ | 0.23 | 3.82 | 6.95 | 0.17 | 0.71 | 6.71 | 0.70 | 3.92 |
| n=2500 | $\widehat{BIAS} \times 10^1$ | 0.28 | 5.96 | 7.61 | 0.29 | 0.55 | 7.60 | 0.56 | 5.95 |
| | $\widehat{VAR} \times 10^2$ | 0.43 | 0.31 | 1.02 | 0.27 | 1.12 | 0.87 | 0.96 | 0.80 |
| | $\widehat{MSE} \times 10^2$ | 0.51 | 35.82 | 58.90 | 0.35 | 1.43 | 58.63 | 1.27 | 36.16 |

**Table 3**

Sensitivity analysis of Simulation Study # 1. We consider nine different perturbations of the simulation parameters. Results are from 1000 simulations of data, n=500 using a parametric spline to adjust for the estimated PS. For each simulation, we report the estimated bias, variance, and mean-squared error (MSE) of the estimators corresponding to all possible specifications of the PS model.

| Parameter Change | | Variables in Propensity Score Model | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | $X_1$ | $X_2$ | $X_3$ | $X_1, X_2$ | $X_1, X_3$ | $X_2, X_3$ | $X_1, X_2, X_3$ | None |
| Original | $\hat{BIAS} \times 10^1$ | −0.03 | 5.97 | 7.34 | −0.03 | −0.07 | 7.36 | −0.06 | 5.94 |
| | $\hat{VAR} \times 10^1$ | 0.32 | 0.22 | 0.46 | 0.22 | 0.44 | 0.36 | 0.31 | 0.39 |
| | $\hat{MSE} \times 10^1$ | 0.32 | 3.79 | 5.85 | 0.22 | 0.44 | 5.77 | 0.31 | 3.92 |
| # 1 | $\hat{BIAS}$ | 0.13 | 2.94 | 3.81 | 0.12 | 0.13 | 3.80 | 0.12 | 2.99 |
| ↓ $VAR[X_1]$ | $\hat{VAR} \times 10^1$ | 0.27 | 0.21 | 0.47 | 0.19 | 0.38 | 0.38 | 0.28 | 0.35 |
| | $\hat{MSE} \times 10^1$ | 0.27 | 1.07 | 1.92 | 0.19 | 0.38 | 1.82 | 0.28 | 1.24 |
| # 2 | $\hat{BIAS} \times 10^1$ | 0.06 | 8.46 | 10.1 | 0.02 | 0.07 | 10.04 | 0.02 | 8.49 |
| ↑ $VAR[X_1]$ | $\hat{VAR} \times 10^1$ | 0.38 | 0.28 | 0.50 | 0.27 | 0.51 | 0.41 | 0.38 | 0.44 |
| | $\hat{MSE} \times 10^1$ | 0.38 | 7.44 | 10.71 | 0.27 | 0.51 | 10.50 | 0.38 | 7.64 |
| # 3 | $\hat{BIAS} \times 10^1$ | 0.02 | 5.96 | 7.38 | 0.02 | 0.00 | 7.38 | 0.01 | 5.97 |
| ↓ $VAR[X_2]$ | $\hat{VAR} \times 10^1$ | 0.07 | 0.13 | 0.19 | 0.05 | 0.12 | 0.17 | 0.10 | 0.16 |
| | $\hat{MSE} \times 10^1$ | 0.07 | 3.69 | 5.64 | 0.05 | 0.12 | 5.62 | 0.10 | 3.72 |
| # 4 | $\hat{BIAS} \times 10^1$ | 0.23 | 6.16 | 7.59 | 0.19 | 0.24 | 7.53 | 0.18 | 6.16 |
| ↑ $VAR[X_2]$ | $\hat{VAR} \times 10^1$ | 1.20 | 0.60 | 1.57 | 1.00 | 1.62 | 1.32 | 1.32 | 1.25 |
| | $\hat{MSE} \times 10^1$ | 1.21 | 4.39 | 7.33 | 1.00 | 1.62 | 7.00 | 1.32 | 5.04 |
| # 5 | $\hat{BIAS} \times 10^1$ | 0.08 | 6.89 | 7.35 | 0.03 | 0.04 | 7.30 | 0.02 | 6.94 |
| ↓ $VAR[X_3]$ | $\hat{VAR} \times 10^1$ | 0.34 | 0.25 | 0.46 | 0.23 | 0.39 | 0.36 | 0.28 | 0.42 |
| | $\hat{MSE} \times 10^1$ | 0.35 | 5.00 | 5.86 | 0.23 | 0.39 | 5.70 | 0.28 | 5.24 |
| # 6 | $\hat{BIAS} \times 10^1$ | 0.10 | 5.07 | 7.53 | 0.07 | 0.05 | 7.49 | 0.01 | 5.1 |
| ↑ $VAR[X_3]$ | $\hat{VAR} \times 10^1$ | 0.29 | 0.23 | 0.55 | 0.21 | 0.49 | 0.46 | 0.39 | 0.38 |
| | $\hat{MSE} \times 10^1$ | 0.29 | 2.80 | 6.21 | 0.22 | 0.49 | 6.07 | 0.39 | 2.98 |
| # 7 | $\hat{BIAS} \times 10^1$ | 0.08 | 5.98 | 7.46 | 0.03 | 0.11 | 7.41 | 0.04 | 6.02 |
| ↓ $\alpha_4$ | $\hat{VAR} \times 10^1$ | 0.32 | 0.24 | 0.47 | 0.22 | 0.43 | 0.37 | 0.31 | 0.39 |
| ↓ tmt eff | $\hat{MSE} \times 10^1$ | 0.32 | 3.82 | 6.03 | 0.22 | 0.43 | 5.86 | 0.31 | 4.01 |
| # 8 | $\hat{BIAS} \times 10^1$ | −0.12 | 5.68 | 6.92 | −0.08 | −0.16 | 6.99 | −0.10 | 5.62 |
| ↑ $\alpha_4$ | $\hat{VAR} \times 10^1$ | 0.34 | 0.21 | 0.52 | 0.23 | 0.49 | 0.39 | 0.33 | 0.38 |
| ↑ tmt eff | $\hat{MSE} \times 10^1$ | 0.34 | 3.43 | 5.32 | 0.23 | 0.49 | 5.27 | 0.34 | 3.54 |
| # 9 | $\hat{BIAS} \times 10^1$ | 0.03 | 6.02 | 7.34 | 0.00 | 0.01 | 7.35 | 0.00 | 5.99 |
| ↓ $\beta_0$ | $\hat{VAR} \times 10^1$ | 0.32 | 0.27 | 0.51 | 0.23 | 0.46 | 0.41 | 0.34 | 0.43 |
| ↓ exp. prev | $\hat{MSE} \times 10^1$ | 0.32 | 3.90 | 5.89 | 0.23 | 0.46 | 5.81 | 0.34 | 4.01 |