



NIH PUBLIC ACCESS

Author Manuscript

*Pharmacoepidemiol Drug Saf.* Author manuscript; available in PMC 2014 January 01.

Published in final edited form as:

*Pharmacoepidemiol Drug Saf.* 2013 January ; 22(1): 77–85. doi:10.1002/pds.3356.

## Variable Selection for Propensity Score Models When Estimating Treatment Effects on Multiple Outcomes: a Simulation Study

Richard Wyss<sup>1</sup>, Cynthia J. Girman<sup>1,2</sup>, Robert J. LoCasale<sup>2</sup>, M. Alan Brookhart<sup>1</sup>, and Til Stürmer<sup>1</sup><sup>1</sup>Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill<sup>2</sup>Department of Epidemiology, Merck Research Laboratories, Merck Sharp & Dohme Corp

### Abstract

**Purpose**—It is often preferable to simplify the estimation of treatment effects on multiple outcomes by using a single propensity score (PS) model. Variable selection in PS models impacts the efficiency and validity of treatment effects. However, the impact of different variable selection strategies on the estimated treatment effects in settings involving multiple outcomes is not well understood. The authors use simulations to evaluate the impact of different variable selection strategies on the bias and precision of effect estimates to provide insight into the performance of various PS models in settings with multiple outcomes.

**Methods**—Simulated studies consisted of dichotomous treatment, two Poisson outcomes, and eight standard-normal covariates. Covariates were selected for the PS models based on their effects on treatment, a specific outcome, or both outcomes. The PSs were implemented using stratification, matching, and weighting (IPTW).

**Results**—PS models including only covariates affecting a specific outcome (outcome-specific models) resulted in the most efficient effect estimates. The PS model that only included covariates affecting either outcome (generic-outcome model) performed best among the models that simultaneously controlled measured confounding for both outcomes. Similar patterns were observed over the range of parameter values assessed and all PS implementation methods.

**Conclusions**—A single, generic-outcome model performed well compared with separate outcome-specific models in most scenarios considered. The results emphasize the benefit of using prior knowledge to identify covariates that affect the outcome when constructing PS models and support the potential to use a single, generic-outcome PS model when multiple outcomes are being examined.

### Keywords

propensity scores; simulation; variable selection; instrumental variables; pharmacoepidemiology

## INTRODUCTION

The propensity score (PS), defined as the conditional probability of treatment given a set of observed covariates, has been shown to effectively balance measured covariates across

---

Address for correspondence: Til Stürmer, MD, MPH, PhD, Department of Epidemiology, UNC Gillings School of Global Public Health, University of North Carolina at Chapel Hill, McGavran-Greenberg, CB # 7435, Chapel Hill, NC 27599-7435, Phone: +1 919 966 7433, Fax: +1 919 966 2089, [til.sturmer@post.harvard.edu](mailto:til.sturmer@post.harvard.edu).

**Conflicts of interest:** none declared.

treatment groups through methods such as matching, stratification, and weighting.<sup>1,2</sup> Variable selection in PS analysis has an impact on the efficiency and validity of the estimated treatment effect.<sup>3–5</sup> Both theory and simulations have shown that including variables that affect treatment but not the outcome (instrumental variables or IVs) decreases the efficiency of effect estimates and, in the presence of unmeasured confounding, can increase the bias of those estimates.<sup>6–9</sup> Studies have further shown that the inclusion of variables affecting only the outcome (risk factors) can increase the precision of estimated treatment effects by controlling for random or chance imbalances of the risk factors across treatment groups.<sup>3</sup> These studies indicate that the ideal PS model, with respect to efficiency and bias, is a model that includes all variables affecting the outcome of interest while excluding IVs.

Although the potential negative effects of controlling for IVs are known, the magnitude of these effects is not well understood. In a recent study conducted by Myers, et al.,<sup>10</sup> the authors explored this issue by evaluating the effects of controlling for IVs in various settings. Their results showed that the increase in variance and potential increase in bias of effect estimates due to conditioning on IVs was relatively small in most of the settings examined. Although these findings provide valuable insight into the magnitude of the impact that conditioning on IVs can have on effect estimates, their results are specific to the scenarios assessed in their study.<sup>11</sup> Uncertainty remains regarding the performance of various PS variable selection strategies in settings specific to multiple outcomes.

Evaluating the effect of a treatment or exposure on multiple outcomes is common in many areas of epidemiologic research, particularly in the area of drug safety where monitoring multiple health events is often the focus.<sup>12–14</sup> In these settings, it is often preferable to simplify the causal analysis by fitting a single, higher-dimensional PS model to simultaneously balance measured covariates across treatment groups instead of fitting separate PS models for each outcome.<sup>12</sup> In practice, however, identifying the ideal or optimal set of covariates for inclusion in the PS model(s) may be unclear and can become increasingly difficult in studies involving multiple outcomes. Even if the true relations among covariates are known, when estimating the effect of a treatment on multiple outcomes with non-identical risk factors, it is not always possible to include the optimal set of variables using a single, high-dimensional PS model. This is because the same covariate may be a confounder for one outcome but an IV for another.

Because uncertainty exists regarding the performance of various PS variable selection strategies and since it is not always possible to exclude IVs when using a single PS model in settings involving multiple outcomes, it is important to understand the relative loss in precision and potential increase in bias that can occur when using a single or “generic” PS model to simultaneously control measured confounders for multiple outcomes. In this study, we conducted Monte Carlo simulations in an effort to better understand the relative performance of various PS models with respect to bias and precision of effect estimates in settings involving multiple outcomes.

## METHODS

### Simulation setup

We simulated a causal structure consisting of one dichotomous treatment (T), two Poisson distributed outcomes ( $Y_1$ ,  $Y_2$ ), six independent standard-normal measured covariates ( $X_1$ ,  $X_2$ ,  $X_3$ ,  $X_4$ ,  $X_5$ ,  $X_6$ ), and two independent standard-normal unmeasured confounders ( $U_1$ ,  $U_2$ ). Figure 1 provides a diagram of the described causal structure. In Figure 1,  $X_1$  is an IV for both outcomes,  $X_2$  an IV for outcome  $Y_1$  and a confounder for outcome  $Y_2$ ,  $X_3$  a confounder for outcome  $Y_1$  and an IV for outcome  $Y_2$ ,  $X_4$  a confounder for both outcomes,

$X_5$  a risk factor for outcome  $Y_1$ , and  $X_6$  a risk factor for outcome  $Y_2$ . Poisson outcomes were chosen since modeling counts is common in epidemiologic research and estimating rate ratios avoids issues with the non-collapsibility of the odds ratio.<sup>15,16</sup> The functional relations among the covariates are summarized in equations 1–3.

$$E(T|X_1-X_4, U_1, U_2) = (1 + \exp\{\alpha_0 + \alpha_1(X_1) + \alpha_2(X_2) + \alpha_3(X_3) + \alpha_4(X_4) + \alpha_5(U_1) + \alpha_6(U_2)\})^{-1} \quad [1]$$

$$E(Y_1|T, X_3-X_5, U_1) = \exp\{\beta_{10} + \beta_{11}(T) + \beta_{13}(X_3) + \beta_{14}(X_4) + \beta_{15}(X_5) + \beta_{1U}(U_1)\} \quad [2]$$

$$E(Y_2|T, X_2, X_4, X_6, U_2) = \exp\{\beta_{20} + \beta_{21}(T) + \beta_{22}(X_2) + \beta_{24}(X_4) + \beta_{26}(X_6) + \beta_{2U}(U_2)\} \quad [3]$$

### Parameter values and sensitivity analysis

We created a variety of scenarios where the strength and direction of the causal relations among the confounders both increased and decreased the crude or unadjusted effect estimate. The parameters and values used in the simulations are shown in Table 1. The relations among the variables were first simulated using two basic sets of parameter values. The first set (Scenario A) involved no unmeasured confounding while the second (Scenario B) included unmeasured confounders for both outcomes,  $Y_1$  and  $Y_2$ . Since unmeasured confounding is a fundamental obstacle in pharmacoepidemiology and non-experimental research in general, unmeasured confounders were included to assess their impact on the precision and bias of effect estimates for each of the PS models (Table 1).

As with any simulation study, the generalizability of the results is limited to the specific scenarios assessed. To make the results of this study more generally applicable, we conducted multiple sensitivity analyses by varying selected parameter values over a range of causal scenarios (alternative scenarios in Table 1). For outcome  $Y_1$ , parameter values for the effect of a confounder on treatment ( $\alpha_3$ ), a confounder on outcome ( $\beta_{13}$ ), an IV on treatment ( $\alpha_2$ ), and a risk factor on outcome ( $\beta_{15}$ ) were varied one at a time while holding all other parameters constant at the Scenario A parameter values and again while holding all other parameters constant at the Scenario B parameter values (Table 1). The selected parameters for the alternative scenarios, or sensitivity analyses, were chosen to represent each type of covariate-treatment or covariate-outcome relation in the simulated causal structure (i.e. effects between a confounder on treatment, confounder on outcome, IV on treatment, and risk factor on outcome). A similar process was repeated for outcome  $Y_2$ .

The parameter values shown in Table 1 are not specific to any particular study and were simply chosen to provide a general understanding of the issues being addressed and provide a general representation of the strength of relations common to many epidemiologic settings. We simulated 5,000 studies for each scenario with sample sizes of  $N=10,000$  and 1,000 for each simulated study. With 5,000 simulations, the Monte Carlo standard error (MCSE) was less than 0.002 for each of the scenarios evaluated in this study. The sample sizes of 10,000 and 1,000 were chosen since automated databases, which often result in very large sample sizes, are primarily used in pharmacoepidemiology and are increasingly used in other areas of epidemiologic research.<sup>17</sup>

### Propensity score estimation

We used a variety of logistic regression models to estimate the PS. Each model was conditioned on a separate set of covariates.

- Full PS model: Included all measured covariates  $X_1$  through  $X_6$ .
- Outcome  $Y_1$  specific PS model: Included only measured covariates affecting outcome  $Y_1$  ( $X_3$ ,  $X_4$ , and  $X_5$ ). The outcome  $Y_1$  specific model was only implemented when evaluating the treatment effect on outcome  $Y_1$ .
- Outcome  $Y_2$  specific PS model: Included only measured covariates affecting outcome  $Y_2$  ( $X_2$ ,  $X_4$ , and  $X_6$ ). The outcome  $Y_2$  specific model was only implemented when evaluating the treatment effect on outcome  $Y_2$ .
- Generic-outcome PS model: Included all measured covariates affecting either outcome  $Y_1$  or outcome  $Y_2$  ( $X_2$ ,  $X_3$ ,  $X_4$ ,  $X_5$ , and  $X_6$ ).
- Treatment-specific PS model: Included all measured covariates affecting treatment ( $X_1$ ,  $X_2$ ,  $X_3$ , and  $X_4$ ).

### Propensity score implementation and estimation of treatment effects

We implemented the estimated PSs with three commonly used techniques: matching, stratification, and inverse probability treatment weighting (IPTW).<sup>18,19</sup> Each of these methods can potentially result in different effect estimates. With no treatment effect heterogeneity, however, the treatment effect is constant across disparate populations and each of these estimation methods (i.e. matching, stratification, and IPTW) should result in the same treatment effect, albeit in different populations. In this study, we did not directly compare the relative performance of one method with another, but instead included each of these methods to provide insight into the performance of a variety of PS methods that are used in practice.

For PS Matching we matched each observation receiving treatment 1:1 without replacement with an untreated observation using a varying-width caliper matching algorithm (five to one digit matching).<sup>20</sup> From the matched set of observations, the treatment effect was then estimated using Poisson regression.

We next implemented the PS using PS Stratification. We trimmed the non-overlapping regions of the treated and untreated PS distributions. After dividing the sample into ten strata using deciles of the estimated PSs, Poisson regression was used to estimate the rate ratio while using indicator variables to control for the PS deciles.

For inverse probability treatment weighting (IPTW), we defined weights as the inverse probability of receiving the treatment actually received.<sup>21–24</sup> Stabilized weights were used by multiplying the previously defined weights by the marginal probability of receiving the treatment actually received. The weights were then used to create a pseudo-population in which the unconfounded treatment effect could be estimated using Poisson regression.<sup>22–24</sup>

We also estimated the treatment effects using the crude and true outcome models. The true outcome model contained all confounders and risk factors specific to each outcome in a Poisson regression model, while the crude model did not control for any covariates. The crude and true outcome models did not implement any PS analysis and were, therefore, not of primary interest, but serve for comparison with the performance of the PS methods.

### Measures of performance

We estimated the bias, standard error, and mean squared error (MSE) for each of the effect estimates. The bias, defined as the expected value of the difference between the effect estimate and the true effect, was calculated by taking the mean of this difference over all 5,000 simulation runs. To evaluate the precision of effect estimates, we estimated the standard error using the square root of the sample variance of the treatment effect estimates

across all 5,000 simulation runs. The MSE was calculated by taking the mean of the squared errors over all the simulation runs.

## RESULTS

The patterns among the results were very similar between outcome  $Y_1$  and outcome  $Y_2$  for all of the estimated measures. Therefore, we only present results for outcome  $Y_1$  as the conclusions drawn regarding the relative performance of the PS models were similar for outcome  $Y_2$ . Further, the patterns among the results were also very similar for each of the scenarios where the incidence of the outcomes and prevalence of treatment were varied. We do not present the results for these scenarios as the overall conclusions did not change. Finally, for the sensitivity analysis (alternative scenarios) we only present results when using IPTW to estimate treatment effects as the patterns and conclusions did not change when using matching or stratification.

### Basic scenarios

We present results for the bias, standard error, and mean squared error (MSE) of the estimated treatment effects for the basic scenarios (Scenario A and Scenario B) in Table 2. Similar patterns were found for each of these scenarios with respect to the precision of effect estimates. When evaluating the effect of the treatment on outcome  $Y_1$ , the outcome  $Y_1$  specific model resulted in effect estimates with the greatest precision followed by the generic-outcome model, the full model and finally the treatment-specific model (Table 2).

In the absence of unmeasured confounding (Scenario A), each of the PS models (i.e. full, treatment-specific, outcome  $Y_1$  or  $Y_2$  specific, and generic-outcome) reduced confounding bias of the estimated treatment effects to approximately zero (results for outcome  $Y_1$  are presented in Table 2). It should be noted, however, that Cochran<sup>25</sup> showed that effect estimates based on stratification methods are not asymptotically unbiased. This explains why there remains a small degree of residual confounding bias when estimating the treatment effect using stratification on PS deciles.

In the presence of unmeasured confounding (Basic Scenario B), the full and treatment-specific PS models resulted in effect estimates with the largest bias (Table 2). The full and treatment-specific models were followed by the generic-outcome model while the outcome  $Y_1$  specific model resulted in effect estimates with the smallest degree of bias (Table 2). Similar patterns and findings for both the precision and bias were found when evaluating the effect of treatment on outcome  $Y_2$  (results not shown).

### Sensitivity analysis

We present the standard error of the effect estimates for each of the sensitivity analyses (alternative scenarios) with no unmeasured confounding in Figure 2 and with unmeasured confounding in Figure 3. Similar patterns for the precision of effect estimates were observed for each of the PS models and parameter settings evaluated in the sensitivity analyses. For settings with and without unmeasured confounding, the precision of the effect estimates was greatest for the correctly specified outcome  $Y_1$  or outcome  $Y_2$  specific PS models (results for outcome  $Y_2$  not shown), followed by the generic-outcome model, the full PS model, and finally, the treatment-specific model (Figures 2 and 3). When the sample size was decreased from 10,000 to 1,000, similar patterns were observed, although the standard error and bias of the effect estimates did increase for all PS models (results not shown).

With respect to the bias of the estimated treatment effect, the bias was reduced to approximately zero for scenarios involving no unmeasured confounding (results not shown). Figure 4 shows the mean bias of the effect estimates in the presence of unmeasured

confounding. Figure 4 illustrates that when unmeasured confounding was present, the full and treatment PS models resulted in effect estimates with larger bias than the generic-outcome PS model and outcome  $Y_1$  specific PS model. These patterns were observed for all parameter values (Figure 4). Again, similar patterns and findings were found when estimating the effect of treatment on outcome  $Y_2$  (results not shown).

### Relative performance of PS models

The correctly specified outcome  $Y_1$  specific model performed best in terms of precision and, in the presence of unmeasured confounding, bias of effect estimates. The generic-outcome model performed best among the PS models that simultaneously controlled measured confounding for both outcomes (i.e. full, generic-outcome, and treatment-specific PS models). When comparing the generic-outcome model with the separate outcome  $Y_1$  specific PS model, the generic-outcome model performed well with the increase in standard error being less than 15% for all scenarios except those involving a very strong IV-treatment association (approximately 40% increase in standard error when  $\alpha_2 = 1.609$  or  $-1.609$  and 20% increase when  $\alpha_2 = 1.1$  or  $-1.1$ ).

Similar patterns were found in settings with unmeasured confounding. The increase in standard error of the generic-outcome model compared to the outcome  $Y_1$  specific model was less than 12% for all scenarios except those involving a very strong IV-treatment association (approximately 30% increase in standard error when  $\alpha_2 = 1.609$  or  $-1.609$  and 17% increase when  $\alpha_2 = 1.1$  or  $-1.1$ ). For scenarios with unmeasured confounding, the increase in bias when using a single generic-outcome PS model compared to the outcome  $Y_1$  specific PS model was also small ( $< 8\%$ ) for almost all scenarios except those involving a very strong IV-treatment association (approximately 29% increase in bias when  $\alpha_2 = 1.609$  or  $-1.609$  and 18% increase when  $\alpha_2 = 1.1$  or  $-1.1$ ). Similar results and conclusions were observed for both bias and precision when evaluating the treatment effect on outcome  $Y_2$  (results not shown).

## DISCUSSION

Results for the parameter scenarios and causal structure assessed in this study demonstrate that separate outcome-specific models perform best in terms of bias and precision of effect estimates. Among the PS models that simultaneously control for measured confounding for both outcomes in a single PS model (full, generic-outcome, and treatment-specific PS models), the generic-outcome PS model produced effect estimates with the greatest precision and, in the presence of unmeasured confounding, least bias. These results indicate that, when using a single PS model to control for confounding in settings with more than one outcome, a generic-outcome PS model is preferred in terms of precision and bias of the effect estimates while the treatment-specific PS model is the least preferred. These generic-outcome models include all covariates affecting any of the outcomes, while excluding covariates that only affect treatment.

When comparing the performance of the generic-outcome model with the performance of the outcome  $Y_1$  or  $Y_2$  specific models, the generic-outcome PS model performed well in terms of precision and bias of effect estimates except for situations where a variable acting as an instrument for one outcome, but a confounder for the other had a strong effect on treatment (e.g.  $X_2$  acts as an instrument for  $Y_1$  but a confounder for  $Y_2$ ). However, when the effect of these variables on treatment was moderate or weak, the generic-outcome model performed well compared to the correctly specified outcome  $Y_1$  or  $Y_2$  specific models. Therefore, either in the presence or absence of unmeasured confounding, estimation of treatment effects using a single generic-outcome PS model across multiple outcomes may be practical for many applications.



The results of this study are consistent with both theory and previous studies that have shown that controlling for IVs can negatively impact treatment effects.<sup>3–5</sup> Although the findings of this study show that controlling for IVs increases the variability and can increase the bias of effect estimates, the increase in variation and bias amplification was relatively small in most of the parameter settings assessed. These findings are consistent with the study conducted by Myers, et al.<sup>10</sup> Based on their results, Myers, et al. assert that, “estimating an exposure effect conditional on a perfect instrument can increase the bias and standard error of the exposure effect estimate, but these increases were generally small.”

## CONCLUSION

Results from this study provide a basis to determine when cohorts balanced using an already estimated PS could be used to assess additional outcomes which were not pre-specified. If the estimated PS is conditioned on all the known risk factors for the additional outcomes, while not conditioning on strong predictors of treatment which do not affect the additional outcomes, cohorts balanced based on this existing PS will likely be close to optimal. Recognizing when an already estimated PS will perform well in subsequent studies has the benefit of avoiding the process of re-estimating multiple PSs for each study outcome.

Simulation studies are limited to the scenarios assessed. Our simulations do not address the trade-off between including an IV and excluding a confounder. Further, we assume IVs as well as risk factors can be identified based on prior knowledge. The specification of outcome PS models, including the generic-outcome model, requires a strong understanding of the causal relations among the covariates. Therefore, it is important to emphasize the use of study design and subject matter expertise to obtain an understanding of the underlying causal structure when estimating the PS.<sup>26</sup>

Our results provide insight into the relative loss in efficiency when using a single PS model to control for measured covariates for multiple outcomes instead of separate outcome-specific models. Such a setting is likely to arise in studies assessing drug safety such as the FDA sentinel initiative. We conclude that while separate outcome-specific PS models result in the most precise effect estimates, a single generic-outcome PS model that can be used for multiple outcomes performs well in many practical settings.

## Acknowledgments

This work was funded by a grant from Merck and a grant (R01 AG023178) from the National Institute on Aging.

## References

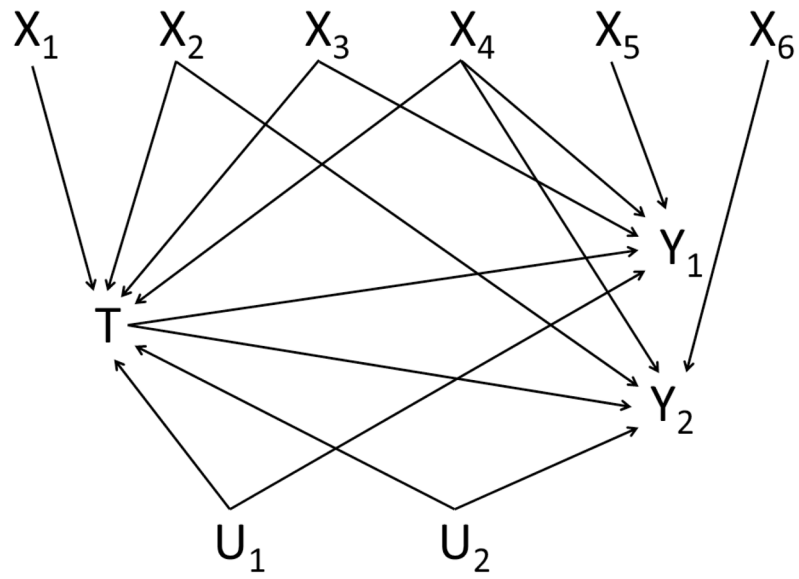
1. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983; 70(1):41–55.
2. D’Agostino RB. Tutorial in biostatistics: propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat Med*. 1998; 17:2265–2281. [PubMed: 9802183]
3. Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Stürmer T. Variable selection for propensity score models. *Am J Epidemiol*. 2006; 163(12):1149–1156. [PubMed: 16624967]
4. Austin PC, Grootendorst P, Anderson GM. A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: a Monte Carlo study. *Stat Med*. 2007; 26(4):734–753. [PubMed: 16708349]
5. Rubin DB. Estimating causal effects from large data sets using propensity scores. *Ann Intern Med*. 1997; 127(8):757–763. [PubMed: 9382394]
6. Pearl, J. On a class of bias-amplifying covariates that endanger effect estimates. In: Grünwald, P.; Spirites, P., editors. *Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial*

- Intelligence (UAI 2010). Corvallis, OR: Association for Uncertainty in Artificial Intelligence; 2010. p. 425-432.
7. Bhattacharya, J.; Vogt, WB. Do Instrumental Variables Belong in Propensity scores?. Cambridge, MA: National Bureau of Economic Research; 2007. (NBER Technical Working Paper no. 343)
  8. Wooldridge, J. Should Instrumental Variables Be Used As Matching Variables?. East Lansing, MI: Michigan State University; 2009. (Technical Working Paper)
  9. Brookhart MA, Stürmer T, Glynn RJ, Rassen JA, Schneeweiss S. Confounding control in healthcare database research: challenges and potential approaches. *Med Care*. 2010; 48(6):114–120.
  10. Myers JA, Rassen JA, Gagne JJ, Huybrechts KF, Schneeweiss S, Rothman KJ, Joffe MM, Glynn RJ. Effects of adjusting for instrumental variables on bias and precision of effect estimates. *Am J Epidemiol*. 2011; 174(11):1213–1222. [PubMed: 22025356]
  11. Pearl J. Invited Commentary: Understanding Bias Amplification. *Am J Epidemiol*. 2011; 174(11): 1223–1227. [PubMed: 22034488]
  12. LoCasale RJ, Girman CJ, Bortnichak EA, Wilson KA, Wyss R, Stürmer T. A Comparison of Covariate Selection Approaches for Propensity Score (PS) Derivation. *Pharmacoepidemiol Drug Saf*. 2011; 20(Suppl 1):S312.
  13. Observation Medical Outcomes Project (<http://omop.fnih.org/>)
  14. Sentinel Initiative (<http://www.fda.gov/Safety/FDAsSentinelInitiative/default.htm>).
  15. Rothman, KJ.; Greenland, S.; Lash, LL. *Modern Epidemiology*. 3. Philadelphia, PA: Lippincott Williams & Wilkins; 2008.
  16. Greenland S, Robins JM, Pearl J. Confounding and collapsibility in causal inference. *Statistical Science*. 1999; 14(1):29–46.
  17. Ray WA. Improving automated database studies. *Epidemiology*. 2011; 22(3):302–304. [PubMed: 21464650]
  18. 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*. 2006; 59:437–447. [PubMed: 16632131]
  19. Austin PC. The relative ability of different propensity score methods to balance measured covariates between treated and untreated subjects in observational studies. *Medical Decision Making*. 2009; 29:661–667. [PubMed: 19684288]
  20. Parsons, LS. SUGI 26 Proceedings. Cary, NC: SAS Institute Inc; 2001. Reducing bias in a propensity score matched-pair sample using greedy matching techniques. (Paper 214–26). <http://www2.sas.com/proceedings/sugi26/p214-26.pdf>
  21. Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000; 11:550–560. [PubMed: 10955408]
  22. Hernán MA, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the Survival of HIV-Positive Men. *Epidemiology*. 2000; 11(5):561–570. [PubMed: 10955409]
  23. Cole SR, Hernán MA. Constructing inverse probability weights for marginal structural models. *Am J Epidemiol*. 2008; 168:656–664. [PubMed: 18682488]
  24. Cole SR, Hernán MA, Robins JM, et al. Effect of highly active antiretroviral therapy on time to acquired immunodeficiency syndrome or death using marginal structural models. *Am J Epidemiol*. 2003; 158:687–694. [PubMed: 14507605]
  25. Cochran WG. The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics*. 1968; 24:205–213.
  26. Robins JM. Data, Design, and Background Knowledge in Etiologic Inference. *Epidemiology*. 2001; 12(3):313–320. [PubMed: 11338312]

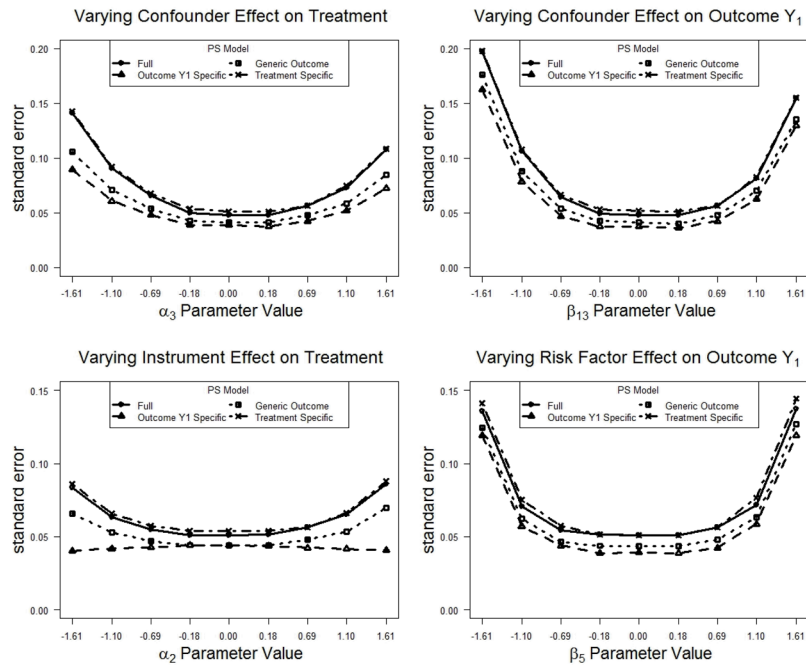


**KEY POINTS**

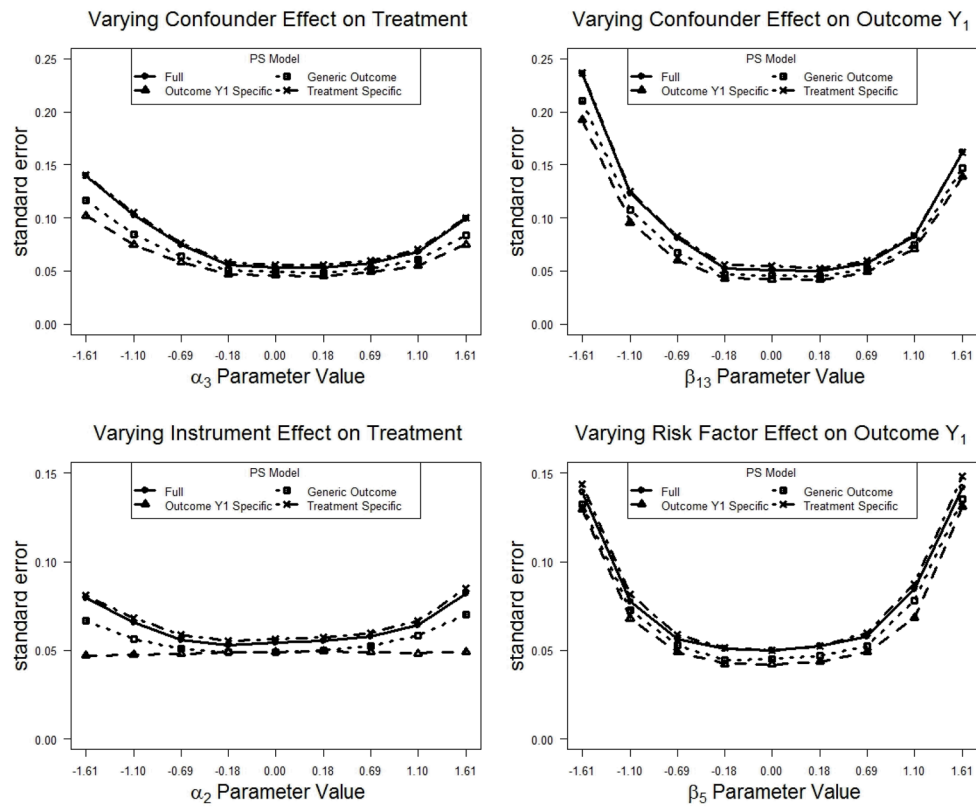
- Results from this study provide insight into situations where it is appropriate to use a single PS model to control for the measured confounding for multiple outcomes.
- When using a single PS model in settings with multiple outcomes, models that only include covariates affecting any of the outcomes of interest perform best in terms of the precision of effect estimates.
- A single, or generic, PS model that simultaneously controls for measured covariates for multiple outcomes performs well compared to separate outcome-specific models in many practical settings.



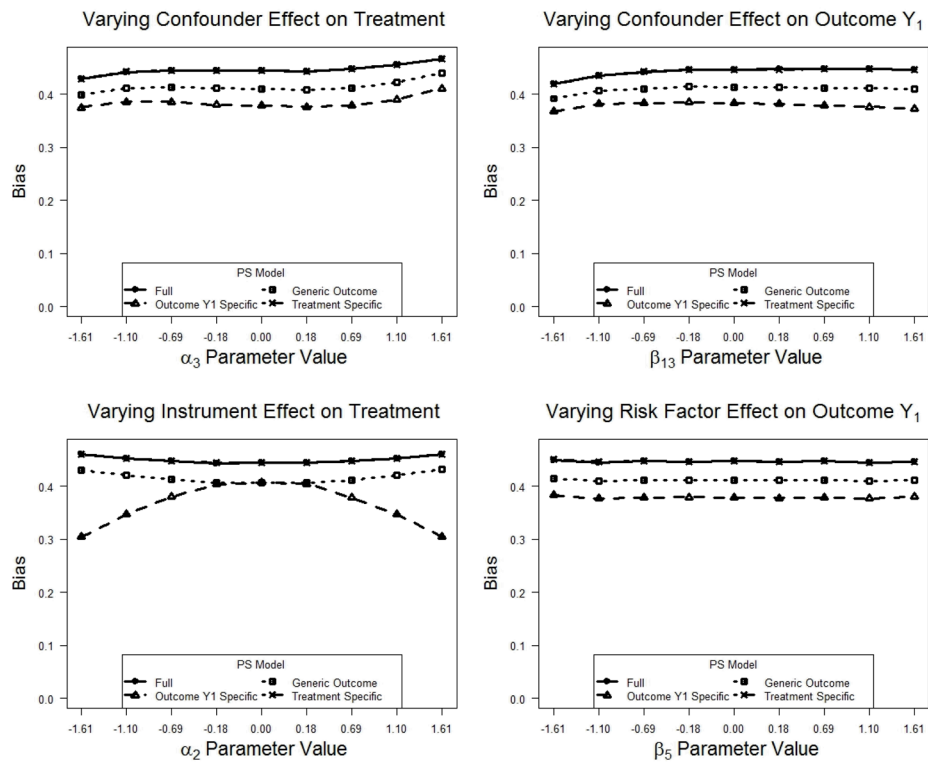
**Figure 1.** Simulated causal structure consisting of six standard-normal measured covariates ( $X_1$ – $X_6$ ), two standard-normal unmeasured covariates ( $U_1$ ,  $U_2$ ), two Poisson distributed outcomes ( $Y_1$ ,  $Y_2$ ) and one dichotomous treatment ( $T$ ).



**Figure 2.** Standard error of treatment effect estimates in settings with no unmeasured confounding and using IPTW to estimate treatment effects.



**Figure 3.** Standard error of treatment effect estimates in settings with unmeasured confounding and using IPTW to estimate treatment effects.



**Figure 4.** Bias of treatment effect estimates in settings with unmeasured confounding and using IPTW to estimate treatment effects.

**Table 1**

Parameter Values Used for Simulations

Parameter	Eq.	Meaning	Basic Scenario A <sup>b</sup>	Basic Scenario B <sup>b</sup>	Alternative Scenarios
$\text{expit}(\alpha_0)^a$	1	Prevalence of T	20%	20%	10%, 30%, 40%, 50%
$\text{exp}(\alpha_1)$	1	X <sub>1</sub> -T Odds Ratio	2.0	2.0	-----
$\text{exp}(\alpha_2)$	1	X <sub>2</sub> -T Odds Ratio	2.0	2.0	0.2, 0.3, 0.5, 0.83, 1.0, 1.2, 3.0, 5.0
$\text{exp}(\alpha_3)$	1	X <sub>3</sub> -T Odds Ratio	2.0	2.0	0.2, 0.3, 0.5, 0.83, 1.0, 1.2, 3.0, 5.0
$\text{exp}(\alpha_4)$	1	X <sub>4</sub> -T Odds Ratio	2.0	2.0	-----
$\text{exp}(\alpha_5)$	1	U <sub>1</sub> -T Odds Ratio	1.0	2.0	-----
$\text{exp}(\alpha_6)$	1	U <sub>2</sub> -T Odds Ratio	1.0	2.0	-----
$\text{exp}(\beta_{10})^a$	2	Incidence of Y <sub>1</sub>	20/100	20/100	5/100, 10/100, 30/100, 40/100
$\text{exp}(\beta_{11})$	2	T-Y <sub>1</sub> Rate Ratio	2.0	2.0	-----
$\text{exp}(\beta_{13})$	2	X <sub>3</sub> -Y <sub>1</sub> Rate Ratio	2.0	2.0	0.2, 0.3, 0.5, 0.83, 1.0, 1.2, 3.0, 5.0
$\text{exp}(\beta_{14})$	2	X <sub>4</sub> -Y <sub>1</sub> Rate Ratio	2.0	2.0	-----
$\text{exp}(\beta_{15})$	2	X <sub>5</sub> -Y <sub>1</sub> Rate Ratio	2.0	2.0	0.2, 0.3, 0.5, 0.83, 1.2, 2.0, 3.0, 5.0
$\text{exp}(\beta_{17})$	2	U <sub>1</sub> -Y <sub>1</sub> Rate Ratio	1.0	2.0	-----
$\text{exp}(\beta_{20})^a$	3	Incidence of Y <sub>2</sub>	20/100	20/100	5/100, 10/100, 30/100, 40/100
$\text{exp}(\beta_{21})$	3	T-Y <sub>2</sub> Rate Ratio	1.2	1.2	-----
$\text{exp}(\beta_{22})$	3	X <sub>2</sub> -Y <sub>2</sub> Rate Ratio	2.0	2.0	0.2, 0.3, 0.5, 0.83, 1.0, 1.2, 3.0, 5.0
$\text{exp}(\beta_{24})$	3	X <sub>4</sub> -Y <sub>2</sub> Rate Ratio	1.2	1.2	-----
$\text{exp}(\beta_{26})$	3	X <sub>6</sub> -Y <sub>2</sub> Rate Ratio	2.0	2.0	0.2, 0.3, 0.5, 0.83, 1.2, 2.0, 3.0, 5.0
$\text{exp}(\beta_{27})$	3	U <sub>2</sub> -Y <sub>2</sub> Rate Ratio	1.0	2.0	-----
N <sub>sim</sub>	----	# Simulations	5,000	5,000	-----
N <sub>study</sub>	----	Sample Size	10,000	10,000	1,000

<sup>a</sup>Intercept parameters refer to the baseline incidence/prevalence for a reference patient with covariate values set to 0.

<sup>b</sup>Basic Scenario A involves no unmeasured confounding while Basic Scenario B includes unmeasured confounding.



**Table 2**Results for Outcome  $Y_1$  With 5,000 Simulated Studies and Sample Size  $N_{\text{study}}=10,000$ 

PS Estimation Model	Outcome Model/PS Implementation	No Unmeasured Confounding <sup>a</sup>			With Unmeasured Confounding <sup>b</sup>		
		Bias	Standard Error	MSE	Bias	Standard Error	MSE
	Crude	0.71	0.045	0.5100	1.01	0.051	1.02
	True	0.00	0.029	0.0008	0.38	0.037	0.14
Outcome $Y_1$ Specific:	Deciles	0.05	0.036	0.0041	0.42	0.042	0.18
	Matching	0.00	0.041	0.0017	0.38	0.050	0.15
	IPTW	0.00	0.043	0.0018	0.38	0.049	0.15
Generic-Outcome:	Deciles	0.04	0.037	0.0030	0.43	0.044	0.19
	Matching	0.00	0.045	0.0021	0.41	0.055	0.17
	IPTW	0.00	0.048	0.0023	0.41	0.052	0.17
Full:	Deciles	0.03	0.039	0.0025	0.46	0.045	0.21
	Matching	0.00	0.049	0.0024	0.44	0.058	0.20
	IPTW	0.00	0.057	0.0032	0.44	0.057	0.20
Treatment-Specific:	Deciles	0.03	0.042	0.0028	0.46	0.048	0.21
	Matching	0.00	0.052	0.0027	0.44	0.062	0.20
	IPTW	0.00	0.059	0.0035	0.44	0.060	0.20

Abbreviations: IPTW, inverse probability treatment weighting; MSE, mean squared error; PS, propensity score

<sup>a</sup>Basic Scenario A parameter values shown in Table 1.<sup>b</sup>Basic Scenario B parameter values shown in Table 1.