Abstract Title Page
Not included in page count.

Title:

On the Bias-Amplifying Effect of Near Instruments in Observational Studies

Authors and Affiliations:

Peter M. Steiner, University of Wisconsin-Madison
Yongnam Kim, University of Wisconsin-Madison

Abstract Body

Background / Context:
In contrast to randomized experiments, the estimation of unbiased treatment effects from observational data requires an analysis that conditions on all confounding covariates. Conditioning on covariates can be done via standard parametric regression techniques or non-parametric matching like propensity score (PS) matching. The regression or matching estimators are causally unbiased (or at least consistent) if the selection mechanism is strongly ignorable, i.e., if all confounding covariates are reliably measured (Rosenbaum & Rubin, 1983). However, for practitioners, the strong ignorability assumption is not very informative because it does not tell them which covariates should actually be included in a regression or PS analysis. Moreover, researchers frequently presume that they might not completely succeed in removing selection bias because their data set might lack some potentially confounding covariates; Or they might only have unreliable measures or proxies of some crucial constructs. Thus, in order to remove as much of the selection bias as possible, it is common advice to condition on all or at least a big set of covariates. The general credo is "the more covariates, the less bias": Conditioning on more covariates cannot do harm, i.e., increase the bias in the estimated treatment effect.

However, recent studies have shown that conditioning on certain types of covariates such as *instrumental variables* (IVs) or *collider variables* can actually amplify or induce bias (e.g., Bhattacharya & Vogt, 2007; Wooldridge, 2009). Neither IVs nor colliders are confounders because, with respect to the data-generating model, they do not simultaneously determine the outcome (*Y*) and the treatment (*Z*). IVs causally determine the treatment but are causally unrelated to the outcome except for its indirect relation via treatment *Z* (see Figure 1). Colliders are typically unrelated to both the outcome and the treatment but are themselves determined by other variables that might affect treatment Z or outcome Y. Interestingly, though IVs and colliders are unrelated to the outcome, including them in a regression or PS model may result in a dramatically increased bias—the bias might be much larger than the bias of the naïve estimate (i.e., the simple mean difference between the treatment and control group, without any covariate adjustments). Thus, if one would know which variables are instruments or colliders they should not be conditioned on when estimating the treatment effect. Though the prevalence of pure instruments and collider variables is very likely low in practice, they clearly demonstrate the counterintuitive fact that conditioning on some variables might actually increase bias.

Even if IVs might be rare in practice, we are much more likely confronted with covariates that are almost like IVs, that is, covariates that strongly determine treatment *Z* but are only weakly related to the outcome (other than via *Z*). Due to their resemblance to pure IVs they are called *near instrumental variables* (near-IVs, Myers et al, 2011). Thus, as Figure 2 suggests, near-IVs are clearly confounders and should definitely be conditioned on when estimating the treatment effect, given the strong ignorability assumption is met. However, if the strong ignorability assumption is not met, near-IVs partially behave like pure IVs and amplify the remaining bias. This is so, because conditioning on an observed confounder reduces *overt bias* but it simultaneously amplifies any remaining *hidden bias* (Pearl, 2010, 2011). Particularly for near-IVs, bias amplification might dominate bias reduction, in particular if the near-IV is strongly related to treatment Z and if the extent of remaining bias is not negligibly small. Given that conditioning on near-IVs might actually increase bias some authors suggest excluding near-IVs from causal analyses (e.g., Pearl, 2011). Similar arguments can be made for near-colliders, but

this is not the focus of this paper.

Purpose / Objective / Research Question / Focus of Study:
In this paper we focus on the bias-amplifying effect of near-IVs—but not only of a single near-IV in the context of a simple data-generating model (e.g., Myers et al., 2011; Pearl, 2010), but of multiple near-IVs in the context of more complex and realistic data-generating models. While for the single near-IV case with only one or two confounding covariates closed analytic formulas for the extent of bias amplification can be derived (Pearl, 2010), this is essentially impossible for most of the data-generating models that involve more than two confounders. Thus, with the exemption of some formal derivations, we investigate the effect of conditioning on near-IVs on bias reduction using simulated data that involve many (interrelated) confounders. In our simulations we vary the confounders' (i) degree of being a near instrument, (ii) correlation structure, (iii) heterogeneity with respect their relation to treatment Z and outcome Y (i.e., the strength and direction of the relation), and (iv) measurement reliability. The main research questions can be summarized as follows:

(A) Given strong ignorability is not met, how strong can bias-amplifying effects of confounders, particularly near-IVs be?
(B) Given the bias-amplifying effect of near-IVs, should on deliberately exclude potential near-IVs from a causal analysis?

Research Design:
We based our simulation study on a data generating model with two continuous endogenous variables—the outcome *Y* and the treatment variable *Z*—and one hundred exogenous confounders $\mathbf{U} = (U_1, \ldots, U_k, \ldots, U_{100})'$. Figure 3 shows the data generating model. In simulating the data we modeled the outcome *Y* as a linear function of the confounder vector **U** and treatment *Z*, $Y = \mathbf{U}'\boldsymbol{\beta} + \tau Z + \varepsilon_Y$, where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_{100})'$ is the column vector of path coefficients. The continuous treatment variable *Z* was generated according to $Z = \mathbf{U}'\boldsymbol{\alpha} + \varepsilon_Z$, where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_{100})'$ is the column vector of corresponding path coefficients. For each simulated scenario, we changed the values of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and ρ. By setting $\alpha_k = .9$ and $\beta_k = .1$, we constructed *near-IVs* because it shows a stronger relation with the treatment *Z* rather than the outcome *Y*. *Outcome-related confounders* were generated by setting $\alpha_k = .5$ and $\beta_k = .5$. In the scenarios with homogeneous confounders, all confounders come from a single confounder type, i.e., $\alpha_1 = \cdots = \alpha_{100}$ and $\beta_1 = \cdots = \beta_{100}$. But for heterogeneous confounder scenarios we generated two different confounder types which we evenly split into 50 near-IVs and 50 outcome-related confounders. For each simulated scenario we run 101 regressions in order to estimate the treatment effect. The first model was always estimated without the inclusion of any confounders. Then we continuously increased the number of included confounders from *j*=1 to 100 such that the estimated models are given by $\hat{Y} = b_0 + \hat{\tau}_j X + b_1 U_1 + \ldots + b_j U_j$. The plots in the results section show the average remaining bias in percent, with the *x*-axis representing the number of included confounders, from 0 to 100. (More details are given in the Appendix.)

Findings / Results:
Types of Covariates (Near-Instrumental Variables and Outcome-Related Confounders).
Figure 4 shows how the two different sets of uncorrelated covariates affect bias-reduction. Conditioning on outcome-related confounders, i.e., confounders that are strongly related to the

outcome but only weakly related to the treatment, shows an almost linearly pattern of bias reduction (dashed line): each confounder removes approximately the same amount of bias. For example, if one conditions on 80% of the confounders (i.e., 20% are assumed to be unobserved), almost 80% bias is removed from the initially biased treatment effect (20% is remaining). The result is quite different for near-IVs (solid line in Figure 4). Conditioning on 80% of confounders removes only 40% of the bias (60% remaining). The reduced bias reduction of the near-IVs as compared to the outcome-related confounders is due to bias-amplification. The difference between the dashed line of the outcome-related covariate and the solid line for the near-IVs indicates the degree of bias amplification (as compared to the outcome-related confounders).

Correlation Structure. The pattern of bias reduction strongly changes once we allow for correlated confounders. If confounders are correlated with each other ($\rho = .3$) then the extent of bias reduction increases and bias-amplification diminishes (Figure 5). First, bias reduction increases because the confounders included the analysis partially pick up the selection bias due to the unobserved but correlated confounders. Thus, conditioning on only 20% of the confounders leaves about 35% and 10% of the bias for the case of near-IVs and outcome-related confounders, respectively. Bias reduction is weaker for near-IVs than for outcome-related confounders because, as before, the remaining bias is amplified. However, since the remaining bias is less than in the case of independent confounders, the extent of bias amplification is smaller for correlated confounders as the difference between the solid and dashed line indicates. In any case, for uncorrelated or correlated confounders, conditioning on an additional near-IV (or outcome-related confounder) always reduced bias, it never increased bias; Though the bias-reducing potential of near-IVs shrinks as the extent of remaining bias (i.e., violation the strong ignorability assumption) increases.

Heterogeneity of Covariates. When confounders are different (i.e., we now have two groups of confounders in our data-generating model: near-IVs and outcome-related confounders) the situation changes dramatically. The extent of bias reduction now depends on which covariates are included and which ones are omitted (i.e., unobserved) from the regression or PS model. Moreover, bias might even increase. Figure 6 shows two extreme cases. In the first case (solid line) the near-IVs are included first and only afterwards the outcome-related confounders are included. As Figure 6 shows, if we condition only on near-IVs, the bias-amplifying effect of near-IVs increases the bias in the treatment effect even more. Because bias-amplification dominates bias-reduction the overall bias increases. Only when outcome-related covariates are included, bias is actually reduced. In the second case (dashed lines) outcome-related covariates are included first, near-IVs afterwards. In this case, bias reduces as one conditions on more and more covariates. Near-IVs no longer cause an increasing bias because the remaining bias (after conditioning on the outcome-related confounders) is too small to be amplified beyond the bias-reducing effect on the near-IVs. These results suggest that one should not condition on near-IVs if there is still considerable remaining bias left. However, in practice we rarely know which variables are near-IVs and how much bias is remaining after conditioning on some covariates.

Reliability of Covariates. In general, covariates should be reliably measured because measurement error reduces a covariate's potential for removing selection bias (e.g., Steiner, Cook & Shadish, 2011). However, if a covariate, like a near-instrument, has a tendency to increase bias, measurement error actually attenuates the bias-amplifying effect (but also the bias

reducing effect). Figure 7 shows the remaining bias for near-IVs measured with different reliabilities. When conditioning on a highly reliable near-IV, bias amplification is stronger, resulting in a more heavily biased treatment effect. However, this pattern only occurs as long as bias amplification dominates bias reduction. Once the remaining bias is small enough (due to the inclusion of most confounders) the bias amplifying effect is dominated by the bias-reducing effect and, thus, bias decreases; But due to measurement error bias decreases at an attenuated rate. Thus, whether conditioning a causal analysis on a near-IV actually increases or decreases bias depends not only on the extent of remaining bias but also on how reliably the near-IV has been measured.

Conclusions:
This paper demonstrates the bias-amplifying effect of near-instrumental variables under different conditions. Besides reducing selection bias, near-IVs also amplify remaining bias (i.e., hidden bias that has not been removed by the covariates already in the regression or PS model). Bias amplification might exceed bias reduction and thus increase the bias instead of reducing it. The simulation results also show that the number of covariates included in a regression or PS model is not an indicator about how much biased is removed. Even with 50% or more confounders included bias might still exceed the bias of the naïve estimate. Moreover, the correlation between observed and unobserved confounders can considerably weaken bias-amplification. Given the potentially negative effects of near-IVs, in practice one should carefully think about (i) whether near-IVs are among the covariates on considers for including in a regression or PS model and (ii) whether the potential near-IVs might increase instead of reduce selection bias. If one presumes that a near-IV might increase selection bias one might decide not to condition on it when estimating a causal effect.

However, the problem in practice is that we rarely know for sure which covariates are near-IVs and whether they actually would increase the bias if conditioned on. Moreover, lacking results from within-study comparison that compare randomized experiments to observational data, it is hard to say how prevalent the issue of bias-increasing near-IVs is (they definitely amplify bias but not necessarily increase bias). So what can we do in practice? First, the results make it very clear that substantive theories about the selection process and the outcome model are very important. They help in identifying potential near instruments and in selecting the confounding covariates for removing bias. If one suspects near-IVs in the observed set of confounders one can probe the treatment effects sensitivity to the inclusion/exclusion of these covariates in the model. However, frequently we do not have strong enough theories in order to make a good call about which variables might be near-IVs and how much bias we might have removed by the covariates included in the model. Given the lack of methodological evidence so far, we only can speculate: We believe that the bias-amplifying effect of near instruments is usually rather weak (i.e., does not increase the bias) given that one has a rich set of conditioning covariates, i.e., a large number of reliably measured covariates including pretest measures of the outcome. Thus, we are cautiously inclined to suggest that near-IVs should be included in a regression or PS model when estimating causal treatment effects—unless substantive theory suggests the opposite. In any case, as with randomized experiments, one should never rely on the results of a single observational data set. Replication matters!

Appendices

Not included in page count.

Appendix A. References

Barnow, B. S., Cain, G. G., & Goldberger, A. S. (1980). Issues in the analysis of selectivity bias. *Evaluation Studies Review Annual, 5,* 43-59.

Bhattacharya, J, & Vogt, W. (2007). *Do Instrumental Variables Belong in Propensity Scores?* (NBER Technical Working Paper no. 343). Cambridge, MA: National Bureau of Economic Research.

Myers, J. A., Rassen, J. A., Gagne, J. J., Huybrechts, K. F., Schneeweiss, S., Rothman, K. J., Joffe, M. M., & Glynn, R. J. (2011). Effects of adjusting for instrumental variables on bias and precision of effect estimates. *American Journal of Epidemiology, 174,* 1213-1222.

Pearl, J. (2009). *Causality: Models, Reasoning, and Inference.* 2nd ed. Cambridge University Press, New York.

Pearl, J. (2010). On a class of bias-amplifying variables that endanger effect estimates. *Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence (UAI 2010) Corvallis.* OR: *Association for Uncertainty in Artificial Intelligence; 2010:* 425–432.

Pearl, J. (2011). Invited commentary: Understanding bias amplification. *American Journal of Epidemiology, 174,* 1223-1227.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70,* 41-55.

Weinberg, C. (1993). Toward a clearer definition of confounding. *American Journal of Epidemiology, 137,* 1–8.

Wooldridge, J. (2009). *Should instrumental variables be used as matching variables?* East Lansing, MI: Michigan State University; 2009.

Appendix B. Simulation Design

We based our simulation study on a data generating model with two continuous endogenous variables—the outcome *Y* and the treatment variable *Z*—and one hundred exogenous confounders $\mathbf{U} = (U_1, \ldots, U_k, \ldots, U_{100})'$. Here we chose 100 confounders since it allows us to conveniently interpret results for the number of included covariates in terms of the percentage (%), which directly allows to generalize to data sets with less or more than 100 covariates. Figure 3 shows the causal diagram of the data generating model. In simulating the data we modeled the outcome *Y* as a linear function of the confounder vector **U** and treatment *Z*, $Y = \mathbf{U}'\boldsymbol{\beta} + \tau Z + \varepsilon_Y$, where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_{100})'$ is the column vector of path coefficients, τ is the treatment effect , which we assume to be a structural constant of .3, and ε_Y is the normally distributed error term. Similarly, the continuous treatment variable *Z*, which might indicate the exact dosage of a treatment (e.g., of a drug), linearly depends on confounders **U**, $Z = \mathbf{U}'\boldsymbol{\alpha} + \varepsilon_Z$, where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_{100})'$ is the column vector of corresponding path coefficients, and ε_Z is the normally distributed error term. The confounders **U** were generated from a multivariate normal distribution, $\mathbf{U} \sim N(\mathbf{0}, \Sigma)$, where Σ is a 100×100 variance-covariance matrix satisfying $Var(U_1) = \cdots = Var(U_{100})$ and $Var(U_1 + \cdots + U_{100}) = 1$, and has covariances equal to ρ.

In each simulation scenario, we changed the values of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and ρ. By setting $\alpha_k = .9$ and $\beta_k = .1$, we constructed near-IVs because it shows a stronger relation with the treatment *Z* rather than the outcome *Y*. Likewise, *outcome-related confounders* were generated by setting $\alpha_k = .5$ and $\beta_k = .5$. In the scenarios with homogeneous confounders, all confounders come from a single confounder type, i.e., $\alpha_1 = \cdots = \alpha_{100}$ and $\beta_1 = \cdots = \beta_{100}$. But, for heterogeneous confounder scenario we generated two different confounder types which we evenly split in 50 near-IVs and 50 outcome-related confounders. Coefficients only varied across the two confounder types, that is, $\alpha_M \neq \alpha_L$ or $\beta_M \neq \beta_L$ and were identical within the given confounder type, i.e., $(\alpha_1 = \ldots = \alpha_{50}) \neq (\alpha_{51} = \cdots = \alpha_{100})$ and $(\beta_1 = \ldots = \beta_{50}) \neq (\beta_{51} = \cdots = \beta_{100})$. Also, to specify situations where the confonders **U** are independent or dependent of each other, we set $\rho = 0$ and $\rho = .3$, respectively. Finally, in order to investigate the effect of the reliability of confounders, we added measurement errors to the confounders **U**. Thus, fallibly measured confounders are generated according to $\mathbf{U}^* = \mathbf{U} + \boldsymbol{\varepsilon}_\mathbf{U}$, where $\boldsymbol{\varepsilon}_\mathbf{U}$ are independently normally distributed errors that have zero means and standard deviations such that reliabilities are 1.0, .9, .7, and .5.

In each scenario, we simulated one thousand data sets of **U**, *Z*, and *Y*. For each data set, we regressed *Y* on *Z* and a subset of **U**. In the first regression model we do not account for any confounders such that the estimated model is $\hat{Y} = b_0 + \hat{\tau}_0 X$, with $\hat{\tau}_0$ representing the estimated treatment effect (in this case the initially biased treatment effect). Next we estimated the model with one confounder included, then with two, and so on. More generally, with *j* covariates in the model, the estimated model is given by $\hat{Y} = b_0 + \hat{\tau}_j X + b_1 U_1 + \ldots + b_j U_j$, for *j* = 1, …, 100. We repeated this procedure basically 1,000 times (*k* = 1, …, 1000), averaged the estimated treatment effects $\hat{\tau}_{kj}$ across iterations and computed the remaining bias as follows:

$Bias(\hat{\tau}_k) = \frac{1}{1000}\sum_{j=1}^{1000}\left(\hat{\tau}_{kj} - \tau\right)$. The plots showing the results of the average remaining bias in percent, with the x-axis representing the number of included confounders, from 0 to 100.

Appendix B. Tables and Figures

Figure 1. Causal diagram illustrating an instrumental variable.

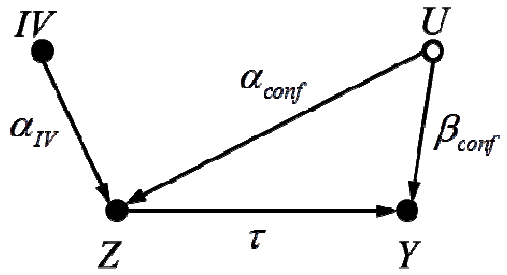

Figure 2. Causal diagram illustrating a near-instrumental variable.

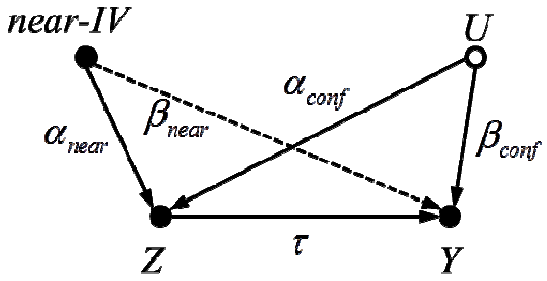

Figure 3. Causal diagram of the data-generating model.

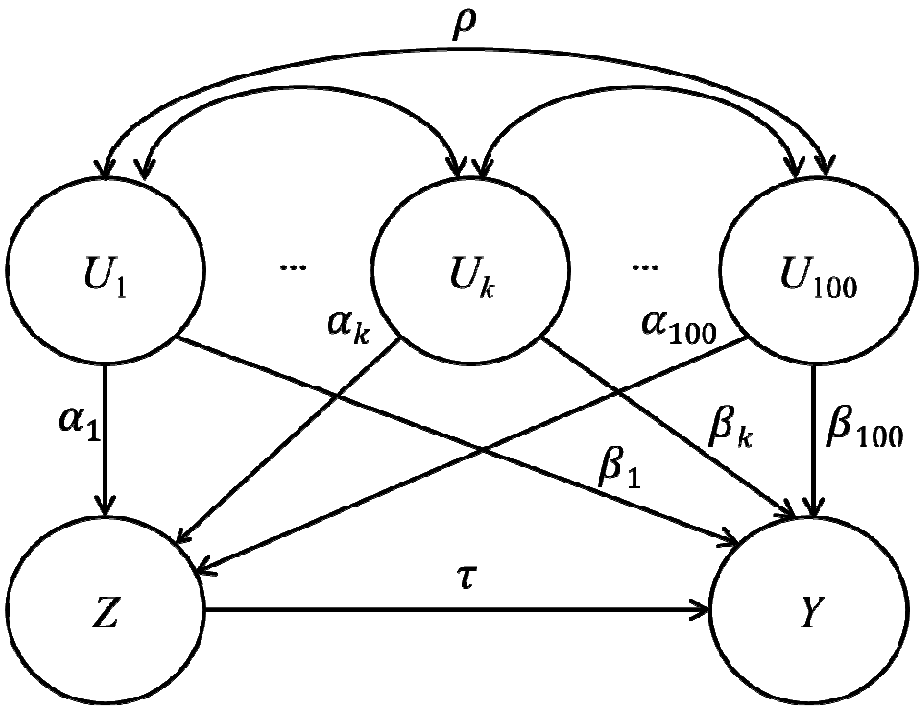

Figure 4. Remaining bias after conditioning on independent homogenous confounders.

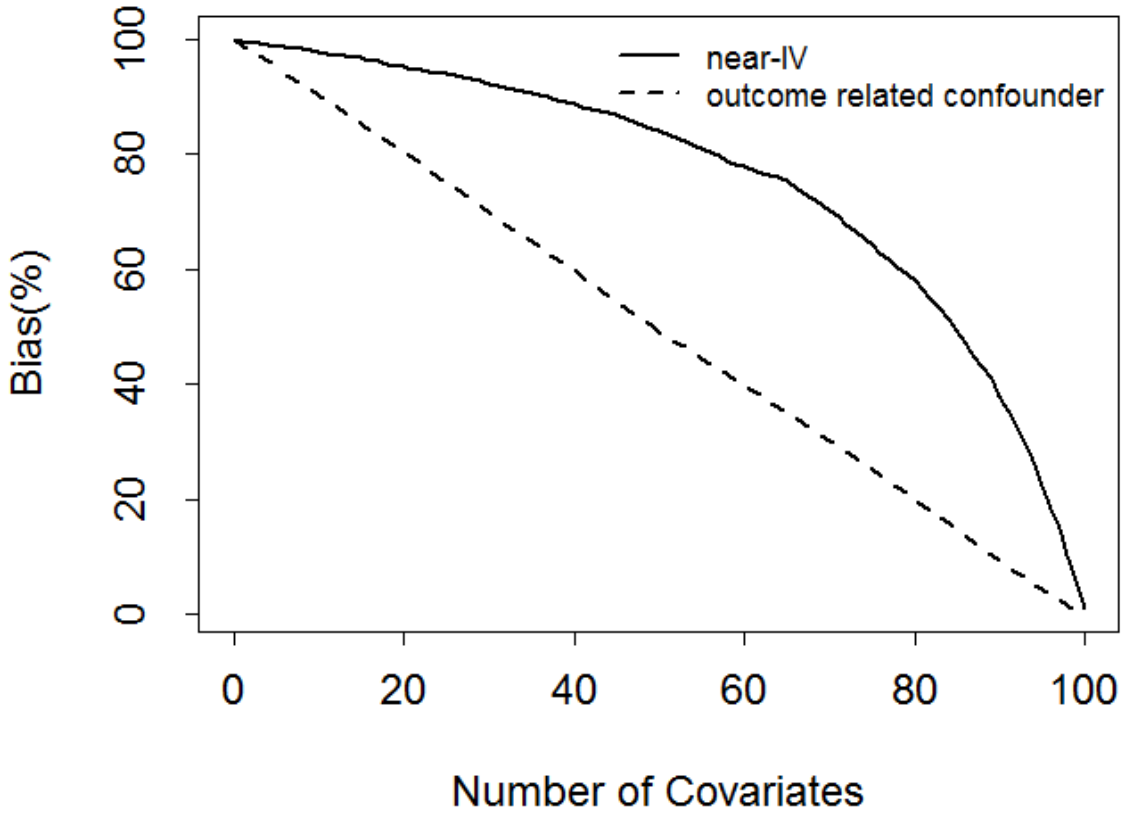

Figure 5. Remaining bias after conditioning on correlated homogenous confounders.

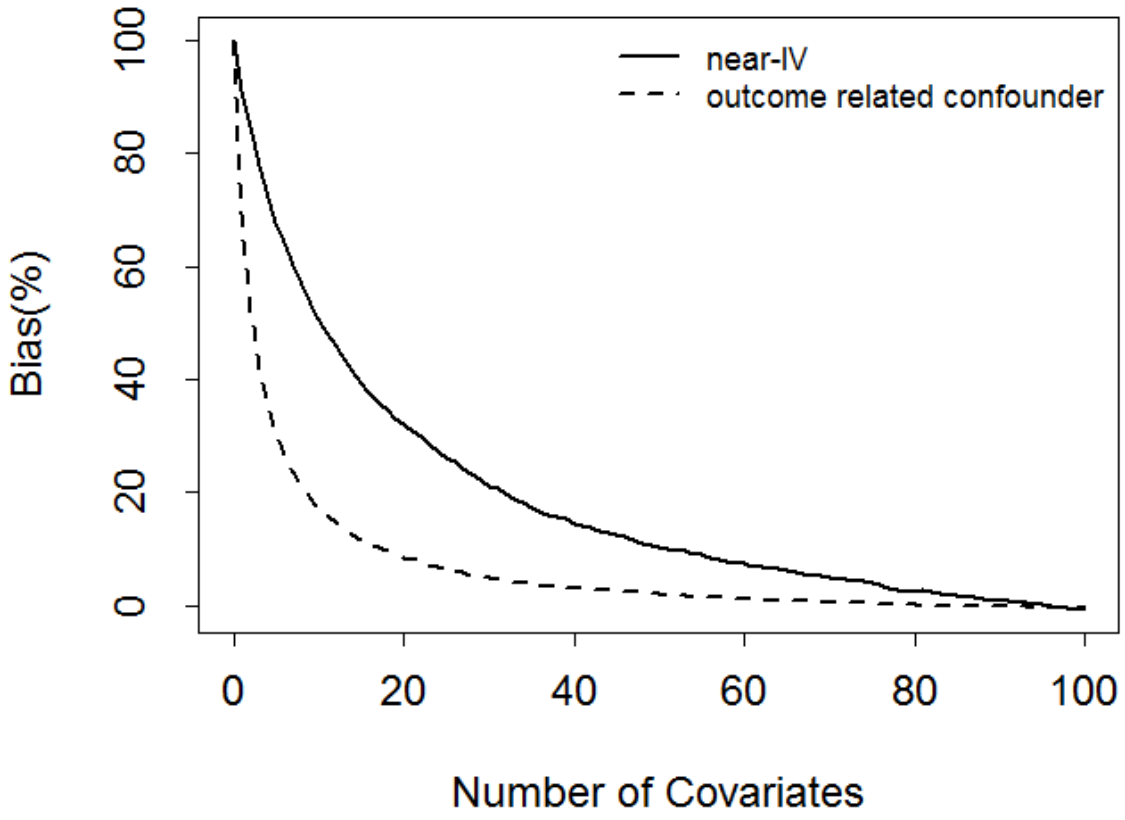

Figure 6. Remaining bias after conditioning on independent heterogeneous confounders.

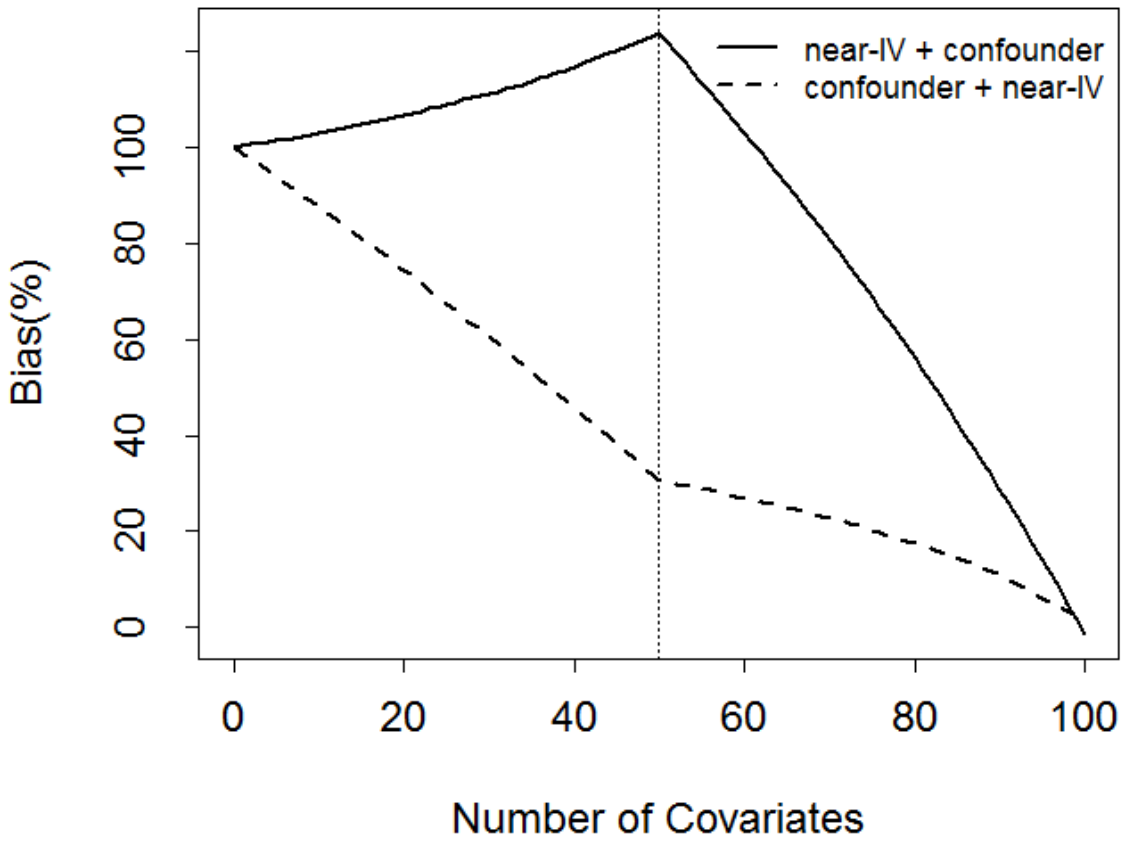

Figure 7. Remaining bias after conditioning on unreliably measured near-IVs and outcome-related confounders. All confounders are independent.