American Journal of Epidemiology
© The Author 2011. Published by Oxford University Press on behalf of the Johns Hopkins Bloomberg School of Public Health. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com.

Vol. 174, No. 11
DOI: 10.1093/aje/kwr352
Advance Access publication:
October 27, 2011

Invited Commentary

Invited Commentary: Understanding Bias Amplification

Judea Pearl*

* Correspondence to Prof. Judea Pearl, Department of Computer Science, University of California, Los Angeles, 4532 Boelter Hall, Los Angeles, CA 90095-1596 (e-mail: judea@cs.ucla.edu).

Initially submitted June 29, 2011; accepted for publication August 11, 2011.

In choosing covariates for adjustment or inclusion in propensity score analysis, researchers must weigh the benefit of reducing confounding bias carried by those covariates against the risk of amplifying residual bias carried by unmeasured confounders. The latter is characteristic of covariates that act like instrumental variables—that is, variables that are more strongly associated with the exposure than with the outcome. In this issue of the *Journal* (*Am J Epidemiol.* 2011;174(11):1213–1222), Myers et al. compare the bias amplification of a near-instrumental variable with its bias-reducing potential and suggest that, in practice, the latter outweighs the former. The author of this commentary sheds broader light on this comparison by considering the cumulative effects of conditioning on multiple covariates and showing that bias amplification may build up at a faster rate than bias reduction. The author further derives a partial order on sets of covariates which reveals preference for conditioning on outcome-related, rather than exposure-related, confounders.

bias (epidemiology); confounding factors (epidemiology); epidemiologic methods; instrumental variable; precision; simulation; variable selection

Abbreviation: IV, instrumental variable.

THE PHENOMENON OF BIAS AMPLIFICATION

This commentary deals with a class of variables that, if conditioned on, tend to amplify confounding bias in the analysis of causal effects. This class, independently discovered by Bhattacharya and Vogt (1) and Wooldridge (2), includes instrumental variables (IVs) and variables that have greater influence on exposure than on the outcome (3).

I am pleased to see that the phenomenon of *bias amplification*, which until recently was practically unknown to researchers in the health sciences, has received a thorough and comprehensive treatment by Myers et al. (4), confirming and qualifying several theoretical predictions derived by Pearl (3) and White and Lu (5). I am particularly struck by Myers et al.'s description of the hip fracture study by Patrick et al. (6), in which "the strength of the IV-exposure relation in this example makes the IV easy to identify and remove by investigators" (4, p. 1218). This awareness that strong predictors of exposure may be a source of troublesome bias is perhaps the most significant impact that the theory of bias amplification has had thus far, because, as Myers et al. point out, it goes against conventional wisdom. Hirano and Imbens (7), for example, devote a major effort to choosing the strongest possible predictors for propensity score inclusion, and Rubin (8) regards the very idea of leaving an observed covariate unconditioned on as "nonscientific ad hockery." (See my previous article (9) for an explanation.)

In this commentary, I supplement the discussion of Myers et al. (4) with several observations that might shed additional light on their conclusions, especially as they pertain to the cumulative effect of multiple near-IV confounders, and the problem of selecting a reasonable set of covariates from a massive host of promising candidates.

BIAS AMPLIFICATION WITH MULTIPLE COVARIATES

Let us examine the simple IV model depicted in Figure 1A, assuming a zero-mean, unit-variance standardization. If we retrace the derivation of the association between X and Y conditional on Z,

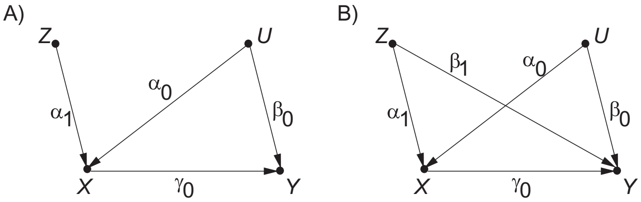

Figure 1. A) A linear model with instrumental variable *Z* and confounder *U*. B) A near-instrumental variable *Z* that is also a confounder.

$$E(Y|X=x+1,Z=z)-E(Y|X=x,Z=z)$$

$$=\gamma_0+\frac{\alpha_0\beta_0}{1-\alpha_1^2},\tag{1}$$

we find that this formula holds not only for a perfect IV but also for a near-IV, as the one depicted in Figure 1B (see my previous article (3)). Allowing a confounding path to extend from *Z* to *Y* will only change the crude association, which will increase from $\gamma_0+\alpha_0\beta_0$ to $\gamma_0+\alpha_0\beta_0+\alpha_1\beta_1$ to reflect the added confounding path $X \leftarrow Z \rightarrow Y$.

Now consider a system of multiple confounders, such as the one depicted in Figure 2, where each covariate intercepts a distinct confounding path between *X* and *Y* and for which the crude bias (without any conditioning) is

$$B_0=\alpha_0\beta_0+\alpha_1\beta_1+\alpha_2\beta_2.\tag{2}$$

If we condition on Z_1, two modifications are required. First, the path containing Z_1 will no longer contribute to confounding and, second, whatever bias is contributed by the remaining paths, namely $\alpha_0\beta_0+\alpha_2\beta_2$, will be amplified by a factor of $\left(1-\alpha_1^2\right)^{-1}$, reflecting the decreased variance of *X* due to fixing Z_1. Overall, the bias remaining after conditioning on Z_1 will read

$$B(Z_1)=\frac{\alpha_0\beta_0+\alpha_2\beta_2}{1-\alpha_1^2}.\tag{3}$$

Further conditioning on Z_2 will remove the factor $\alpha_2\beta_2$ from the numerator (deactivating the path $X \leftarrow Z_2 \rightarrow Y$) and will replace the denominator by the factor $\left(1-\alpha_1^2-\alpha_2^2\right)$, representing the reduced variance of *X*, due to fixing both Z_1 and Z_2. The resulting bias will be

$$B(Z_1,Z_2)=\frac{\alpha_0\beta_0}{\left(1-\alpha_1^2-\alpha_2^2\right)}.\tag{4}$$

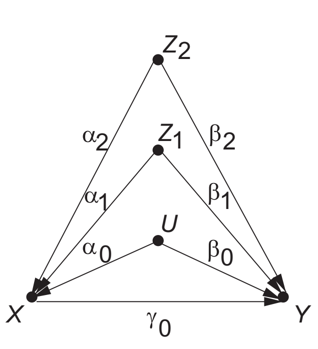

Figure 2. A linear model with multiple covariates (Z_1 and Z_2) and an unobserved confounder *U*.

We see the general pattern that characterizes sequential conditioning on sets of covariates, organized as in Figure 2. The bias *B*(*Z*) remaining after conditioning on a set $Z=(Z_1, Z_2, \ldots, Z_{k-1}, Z_k)$ is given by the formula

$$B(Z)=\frac{B_0-\alpha_1\beta_1-\alpha_2\beta_2-\ldots-\alpha_k\beta_k}{\left(1-\alpha_1^2-\alpha_2^2-\ldots-\alpha_k^2\right)},\tag{5}$$

which reveals 2 distinct patterns of progression, one representing confounding reduction (shown in the numerator) and one representing IV amplification (shown in the denominator). The latter increases monotonically while the former progresses nonmonotonically, since the signs of the added terms may alternate. Thus, the cumulative effect of sequential conditioning has a built-in slant towards bias amplification as compared with confounding reduction; the latter is tempered by sign cancellations, the former is not.

In deriving equation 5, we assumed that no Z_k is a collider, that each Z_k has a distinct path characterized by α_k, and that the Z_k's are not correlated. In a general graph, where multiple paths may traverse each Z_k, *B*(*Z*) will read

$$B(Z)=\frac{B_0^-(k)}{\left(1-{\alpha'_1}^2-{\alpha'_2}^2-\ldots-{\alpha'_k}^2\right)},\tag{6}$$

where $B_0^-(k)$ represents the crude bias B_0 modified by conditioning on $(Z_1, Z_2, \ldots, Z_{k-1}, Z_k)$, and α'_k is the coefficient of Z_k in the regression of *X* on $(Z_1, Z_2, \ldots, Z_{k-1}, Z_k)$. For example, in model 5 of Myers et al. (4) (shown in Figure 3), the crude bias is

$$B(0)=\alpha_2\gamma_1\beta_1+\alpha_1\beta_1,\tag{7}$$

while the bias remaining after conditioning on *Z* reads

$$B(Z)=\frac{\alpha_1\beta_1\left(1-\gamma_1^2\right)}{1-\left(\alpha_2+\gamma_1\alpha_1\right)^2}.\tag{8}$$

The numerator is obtained by setting $\alpha_2=0$ in equation 7 and multiplying the remaining term by $\left(1-\gamma_1^2\right)$, to account for the effect that conditioning on *Z* has on the path $X \leftarrow U \rightarrow Y$. The denominator invokes the factor $\alpha'=(\alpha_2+\gamma_1\alpha_1)$, which is the regression coefficient of *X* on *Z*.

We see that, in this model, γ_1 controls simultaneously the reduction of confounding bias and the amplification of residual bias, both caused by conditioning on *Z*. Myers et al. (4) assumed that γ_1 controls the former only.

In examining the extent to which these results are generalizable to nonlinear models, it has been shown (3) that, while in linear systems conditioning on an IV always amplifies confounding bias (if such exists), bias in nonlinear systems may be amplified as well as attenuated. Additionally, an IV may

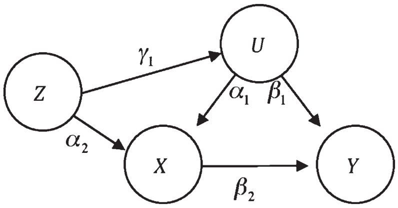

Figure 3. The model used by Myers et al. for studying near-instrumental variables. The parameter γ_1 contributes to confounding as well as to bias amplification.

introduce new bias where none exists. This can be demonstrated if we introduce an interaction term into the model of Figure 1A, to read

$$Y = \gamma_0 X + \beta_0 U + \delta XU + \varepsilon.$$

With this modification, equation 1 becomes

$$E\big(Y|X = x + 1, Z = z\big) - E\big(Y|X = x, Z = z\big)$$
$$= \gamma_0 + \frac{\alpha_0\big(\beta_0 + \delta(2x + 1 - \alpha_1 z)\big)}{1 - \alpha_1^2}, \tag{9}$$

while the crude association becomes

$$E(Y|X = x + 1) - E(Y|X = x)$$
$$= \gamma_0 + \alpha_0(\beta_0 + \delta(2x + 1)). \tag{10}$$

The resulting z-adjusted bias therefore reads

$$B\big(Z = z\big) = \frac{B_0 - \alpha_0\alpha_1\delta z}{1 - \alpha_1^2},$$

where B_0 is the unadjusted bias.

We see that, if $B_0 \geq 0$ and $\alpha_0\alpha_1\delta_z > 0$, we can get $|B_z| < |B_0|$. This means that conditioning on Z may reduce confounding bias, even though Z is a perfect instrument and both Y and X are linear in U. Note that, owing to the nonlinearity of $Y(x, u)$, the conditional bias depends on the value of Z and, moreover, for $Z = 0$ we obtain the same bias amplification as in the linear case (equation 1).

We also see that conditioning on Z can introduce bias where none exists. However, this occurs only for a specific value of X,

$$x = -(1 + \beta_0/\delta)/2,$$

a condition that yields $B_0 = 0$ and $|B_z| > 0$.

ON THE CHOICE BETWEEN EXPOSURE-RELATED AND OUTPUT-RELATED COVARIATES

Investigators are often faced with the need to adjust for a large number of potential confounders; some are strongly related to exposure, and some are more related to the output.

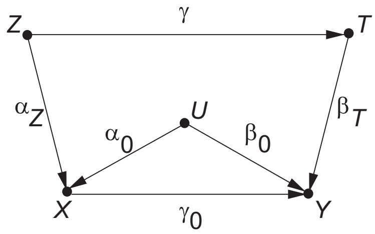

Figure 4. Adjustment for an output-related covariate (T) is preferred to adjustment for a treatment-related covariate (Z) or both (Z, T). The former covariate has a lower bias-amplification potential than the latter two when U is unobserved.

Since estimation efficiency usually deteriorates with the number of covariates involved, the question arises as to which subset of potential confounders to measure and control for (see discussions by Day et al. (10), Thomas and Greenland (11), Hill (12), Austin (13), Pearl (9), White and Lu (5), Patrick et al. (6), and Myers et al. (4)).

Figure 4 represents this choice formally, where T represents output-related covariates, Z represents exposure-related covariates, and U represents unmeasured confounders. We ask which set of variables should be chosen for adjustment: $\{Z\}$, $\{T\}$ or $\{Z, T\}$. Morgan and Winship (14) raise the same question, and they state a preference for $\{Z, T\}$.

Intuitively, since Z is "closer" to X, it acts more like an instrument than T, and one would expect T to yield a lower bias. Indeed, substituting the proper parameters for α_k and β_k in equation 5 confirms this preference; the biases obtained for Z and T are

$$B\big(Z\big) = \frac{\beta_0\alpha_0}{\big(1 - \alpha_Z^2\big)}, \tag{11}$$

and

$$B\big(T\big) = \frac{\beta_0\alpha_0}{\big(1 - \alpha_Z^2\gamma^2\big)}, \tag{12}$$

with a clear advantage of T over Z.

As to the set $\{Z, T\}$, from equation 6 and the fact that the coefficient of T in the regression of X on Z and T vanishes, we conclude that conditioning on $\{Z, T\}$ would have the same bias as conditioning on Z alone. This can also be seen from the theory of collapsibility and confounding-equivalence (15), since $X \perp\!\!\!\perp \{Z, T\}|Z$.

Equations 5 and 6 induce a total order on covariate sets, which in theory can be used to determine (in linear systems) which among several candidate sets of covariates will result, upon adjustment, in the lowest bias. Of course, these equations are not estimable from the data because, first, the residual bias $\alpha_0\beta_0$ is not estimable and, second, the graph structure is generally unknown. However, given a theoretically plausible graph structure, a partial order can be derived which is independent on the numerical values of the parameters. The idea is to compare sets that are known to give rise to the same numerator and for which one denominator is guaranteed to be greater than the other for all values of α_k. We have seen such a preference derived in equations 11 and 12, yet a more general condition for preferring set T over Z can be established by means of this logic, leading to the following rule:

A set T is preferred to Z if

1) T blocks all paths between Y and Z that do not traverse X, and
2) T does not block all paths between Z and X.

These two conditions are clearly satisfied in Figure 4. Complementing this partial order, Pearl and Paz (15) established a necessary and sufficient condition for 2 sets to be equally meritorious for bias reduction.

Thus far, our discussion has focused on adjustment and its effect on systematic bias, yet the harmful effect of overadjustment on *precision* is not less important and has been recognized by epidemiologists for at least 3 decades (10, 11). Remarkably, the ordering dictated by precision considerations coincides almost exactly with that dictated by consideration of bias amplification. Based on a result by Hahn (16) and assuming no unmeasured confounders, White and Lu (5) derived a partial order on covariates in terms of the asymptotic variance of the effect estimand. This ordering prefers covariates that do not constrain X—the more independent variation there is in the exposure, the more efficient the resulting estimator. The intuition is clear; the more latitude we allow for X to swing away from its baseline value, the fewer samples are needed to reveal the effect of that swing. Referring to Figure 4 with $\alpha_0 = 0$ (no measured confounders), White and Lu (5) showed that the asymptotic variance of the estimators of γ_0 obtained by conditioning on T alone is lower than that obtained by conditioning on both T and Z, and the latter is lower than that obtained by conditioning on Z alone. This further reinforces the idea that conditioning on factors affecting X (or their proxies) is to be avoided if possible.

CONCLUSIONS AND RELATED OBSERVATIONS

The study by Myers et al. (4) confirms the general conclusions of Bhattacharya and Vogt (1), Wooldridge (2), Pearl (3), and White in Lu (5) that 1) strong predictors of exposure should be excluded from the analysis, 2) factors affecting outcome (or their proxies) are safer and more effective bias reducers than those affecting exposure, and 3) consideration of covariate selection should be grounded in structural assumptions; it cannot be left at the mercy of conventional wisdom, however entrenched.

Myers et al.'s conclusions that, under conditions prevailing in practice, the bias-reducing potential of a near-IV outweighs its bias-amplification potential should be reevaluated in light of the way that bias accumulates in sequential conditioning over large sets of potential confounders. The fact that bias amplification increases monotonically while confounding reduction progresses nonmonotonically, moderated by cancellation of positive and negative confounding paths, may result in a more pronounced effect of bias amplification than the one revealed by studying a single covariate.

The partial preference order established above on subsets of candidate covariates, though requiring basic knowledge of the graph structure, should not be easily dismissed. The basic scientific knowledge required for this determination is often far more accessible than the knowledge needed for substantiating assumptions such as "strong ignorability," which underlie much of the propensity-score practice.

A few observations should be noted concerning the use of IVs in nonparametric models. First, IVs carry the unique (and rarely utilized) capability of detecting the presence of residual bias whenever a difference $B_0 \neq B_z$ is measured. Second, conditioning on Z has no effect whatsoever on selection-induced bias unless selection is determined by causes of X (3). Finally, Bareinboim and Pearl (17) have shown that the use of an IV can, under certain weak conditions, eliminate selection bias altogether.

ACKNOWLEDGMENTS

Author affiliation: Department of Computer Science, School of Engineering and Applied Science, University of California, Los Angeles, Los Angeles, California (Judea Pearl).

This article was supported in part by grants from the National Institutes of Health (grant 1R01 LM009961-01), the National Science Foundation (grants IIS-0914211 and IIS-1018922), and the Office of Naval Research (grants N000-14-09-1-0665 and N00014-10-1-0933).

The author benefited from discussions with Drs. Peter Austin, Jay Bhattacharya, Antonio Forcina, Josh Gagne, Sander Greenland, Jennifer Hill, Jessica Myers, William Vogt, and Jeffrey Wooldridge.

Conflict of interest: none declared.

REFERENCES

1. Bhattacharya J, Vogt W. *Do Instrumental Variables Belong in Propensity Scores?* (NBER Technical Working Paper no. 343). Cambridge, MA: National Bureau of Economic Research; 2007.
2. Wooldridge J. Should instrumental variables be used as matching variables? East Lansing, MI: Michigan State University; 2009. (https://www.msu.edu/~ec/faculty/wooldridge/current%20research/treat1r6.pdf). (Accessed July 2010).
3. Pearl J. On a class of bias-amplifying variables that endanger effect estimates. In: *Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence (UAI 2010)*. Corvallis, OR: Association for Uncertainty in Artificial Intelligence; 2010:425–432. (http://ftp.cs.ucla.edu/pub/stat_ser/r356.pdf). (Accessed September 2011).
4. Myers JA, Rassen JA, Gagne JJ, et al. Effects of adjusting for instrumental variables on bias and precision of effect estimates. *Am J Epidemiol*. 2011;174(11):1213–1222.
5. White H, Lu X. Causal diagrams for treatment effect estimation with application to efficient covariate selection. *Rev Econ Stat*. In press.
6. Patrick AR, Schneeweiss S, Brookhart MA, et al. The implications of propensity score variable selection strategies in pharmacoepidemiology: an empirical illustration. *Pharmacoepidemiol Drug Saf*. 2011;20(6):551–559.
7. Hirano K, Imbens G. Estimation of causal effects using propensity score weighting: an application to data on right heart

catheterization. *Health Serv Outcome Res Methodol*. 2001; 2(3-4):259–278.

8. Rubin D. Author's reply: Should observational studies be designed to allow lack of balance in covariate distributions across treatment groups? [letter]. *Stat Med*. 2009;28(9): 1420–1423.
9. Pearl J. *Myth, Confusion, and Science in Causal Analysis*. (Technical report R-348). Los Angeles, CA: Department of Computer Science, University of California, Los Angeles; 2009. (http://ftp.cs.ucla.edu/pub/stat_ser/r348.pdf). (Accessed May 2009).
10. Day NE, Byar DP, Green SB. Overadjustment in case-control studies. *Am J Epidemiol*. 1980;112(5):696–706.
11. Thomas DC, Greenland S. The relative efficiencies of matched and independent sample designs for case-control studies. *J Chronic Dis*. 1983;36(10):685–697.
12. Hill J. Discussion of research using propensity-score matching: comments on 'A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003' by Peter Austin, *Statistics in Medicine*. *Stat Med*. 2008;27(12): 2055–2061.
13. Austin PC. A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003. *Stat Med*. 2008;27(12):2037–2049.
14. Morgan SL, Winship C. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. New York, NY: Cambridge University Press; 2007:83.
15. Pearl J, Paz A. Confounding equivalence in causal inference. In: *Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence (UAI 2010)*. Corvallis, OR: Association for Uncertainty in Artificial Intelligence; 2010:433–441.
16. Hahn J. Functional restriction and efficiency in causal inference. *Rev Econ Stat*. 2004;86(1):73–76.
17. Bareinboim E, Pearl J. *Controlling Selection Bias in Causal Inference*. Los Angeles, CA: Department of Computer Science, University of California, Los Angeles; 2011. (http://ftp.cs.ucla.edu/pub/stat_ser/r381.pdf). (Accessed September 2011).