In American Journal of Epidemiology, 174 (11): 1223-1227. doi: 10.1093/aje/kwr352, 2011.

TECHNICAL REPORT
R-386
October 2011

Invited Commentary: Understanding Bias Amplification

Judea Pearl
Cognitive Systems Laboratory
Computer Science Department
University of California, Los Angeles, CA 90024 USA
judea@cs.ucla.edu

Abstract

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 (IV); that is, variables that are more strongly associated with the exposure than with the outcome (1). In this issue of the journal, Myers et al. (2) compare the bias amplification of a near-IV confounder with its bias-reducing potential and suggest that, in practice, the latter outweighs the former. 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. We further derive a partial order on sets of covariates which reveals preference for conditioning on outcome-related, rather than exposure-related confounders.

1 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 (3) and Wooldridge (4), includes instrumental variables (IV) and variables that have greater influence on exposure than on the outcome (1).

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. (2) confirming and qualifying several theoretical predictions derived in Pearl (1) and White and Lu (5).

I am particularly struck by Myers et al.'s description of the "hip fracture" study of Patrick et al. (6) in which "the strength of the IV-exposure relation in this example made the IV easy to identify and remove by investigators." 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 observed covariate unconditioned on as a "non-scientific ad hockery." (See (9) for an explanation.)

In this commentary, I supplement the discussion of Myers et al. with a couple of 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.

2 BIAS AMPLIFICATION WITH MULTIPLE COVARIATES

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

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

we find that this formula holds not only for a perfect IV but also for a near-IV as the one depicted in Figure 1(b) (see (1)). Allowing a confounding path

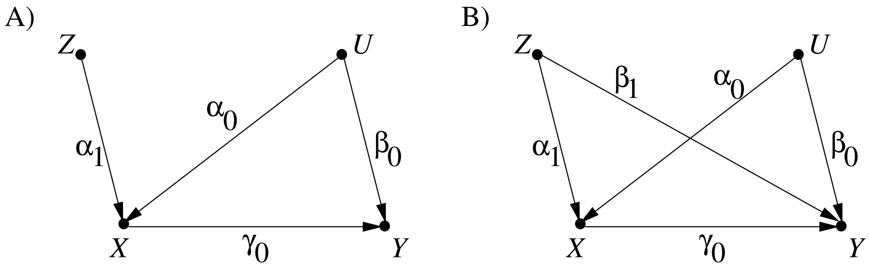

Figure 1: (a)Linear model with instrumental variable Z and confounder U. (b) A near-instrumental variable Z that is also a confounder.

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

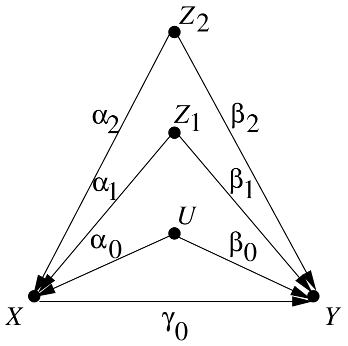

Figure 2: Illustrating a linear model with multiple covariate (Z_1, and Z_2) and an unobserved confounder U.

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

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 $(1 - \alpha_1^2)^{-1}$, 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 $(1 - \alpha_1^2 - \alpha_2^2)$ representing the reduced variance of X, due to fixing both Z_1 and Z_2. The resulting bias will be:

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

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:

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

which reveals two 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 to 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:

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

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

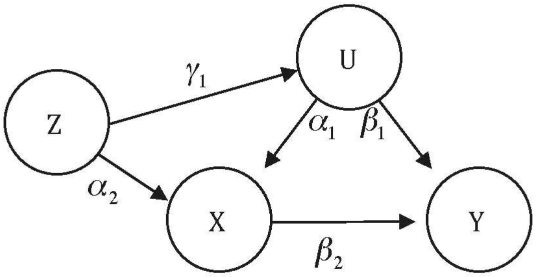

Figure 3: The model used by Myers et al. for studying near-IV's. The parameter γ_1 contributes to confounding as well as to bias amplification.

X on $(Z_1, Z_2, \ldots, Z_{k-1}, Z_k)$. For example, in model 5 of Myers et al. (2) (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:

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

The numerator is obtained by setting $\alpha_2 = 0$ in equation 7 and multiplying the remaining term by $(1-\gamma_1^2)$, 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. (2) assumed that γ_1 controls the former only.

In examining the extent to which these results generalize to non-linear models, it was shown (1) that, while in linear systems conditioning on an IV always amplifies confounding bias (if such exists), bias in non-linear systems may be amplified as well as attenuated. Additionally, an IV may introduce new bias where none exists. This can be demonstrated if we introduce an interaction term into the model of Figure 1(a), to read:

$$Y = \gamma_0 X + \beta_0 U + \delta X U + \epsilon.$$

With this modification, equation 1 becomes:

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

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

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

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, owed to the non-linearity 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. But this occurs only for a specific value of X,

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

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

3 ON THE CHOICE BETWEEN EXPOSURE-RELATED AND OUTPUT-RELATED CO-VARIATES

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. Since estimation efficiency usually deteriorates with the number of covariates involved, the question arises which subset of potential confounders should one choose to measure and control (see discussions in 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. (2)).

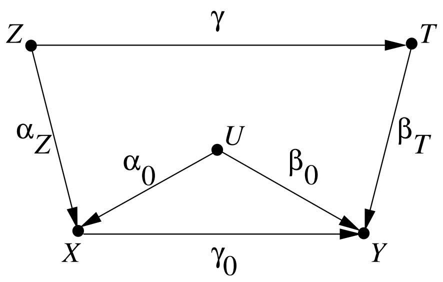

Figure 4: Adjustment for an output-related covariate (T) is preferred to adjustment for treatment-related covariate (Z) or both (Z, T). The former has a lower bias-amplification potential than the latter two when U is unobserved.

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, p. 83) 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:

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

and

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

with clear advantage to 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. These equations are not estimable of course 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's. 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 this logic, leading to the following rule:

> A set T is preferred to Z if
>
> (i) T blocks all paths between Y and Z that do not traverse X, and
>
> (ii) 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 two sets to be equally meritorious for bias reduction.

Thus far, our discussion was focused on adjustment and its effect on systematic bias, yet the harmful effects of overadjustment on *precision* is not less important and has been recognized by epidemiologists for at least three 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 X to swing away from its baseline value, the less 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) show 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.

4 CONCLUSIONS AND RELATED OBSERVATIONS

The study of Myers et al. confirms the general conclusions of Bhattacharya and Vogt (3), Wooldridge (4), Pearl (1) and White in Lu (5) that (i) strong predictors of exposure should be excluded from the analysis, (ii) factors affecting outcome (or their proxies) are safer and more effective bias reducers than those affecting exposure and (iii) consideration of covariate selection should be grounded in structural assumptions; they 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 re-evaluated 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 in Section 4 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 underly much of the propensity-score practice.

I would like to end this commentary with a related observation that may inspire additional investigations into the use of instrumental variables. In view of the amplification effect of IV's on confounding bias, one may surmise that a similar effect can be expected vis-à-vis selection bias, namely, bias caused by preferential selection of samples into the study dataset as in case-control studies. This however is not the case. Conditioning on Z has no effect whatsoever on selection-induced bias unless selection is determined by causes of X (1). Moreover, Bareinboim and Pearl (17) have shown that the use of an IV can, under certain weak conditions, eliminate selection bias altogether.

Acknowledgments

This note was supported in parts by grants from National Institutes of Health #1R01 LM009961-01, National Science Foundation #IIS-0914211 and #IIS-

1018922, and Office of Naval Research #N000-14-09-1-0665 and #N00014-10-1-0933 and has benefited from discussions with Peter Austin, Jay Bhattacharya, Antonio Forcina, Josh Gagne, Sander Greenland, Jennifer Hill, Jessica Myers, William Vogt, and Jeffrey Wooldridge.

References

[1] 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*, Corvallis, OR: AUAI. 2010;425–432. <http://ftp.cs.ucla.edu/pub/stat_ser/r356.pdf>.

[2] Myers JA, Rassen JA, Gagne JJ, et al. Effects of adjusting for instrumental variables on bias and precision of effect estimates. *American Journal of Epidemiology* 2011;:this issue.

[3] Bhattacharya J, Vogt W. Do instrumental variables belong in propensity scores? Tech. Rep. NBER Technical Working Paper 343, National Bureau of Economic Research, MA, 2007.

[4] Wooldridge J. Should instrumental variables be used as matching variables? Tech. Rep. <https://www.msu.edu/∼ec/faculty/wooldridge/current%20research/treat1r6.pdf>, Michigan State University, MI, 2009.

[5] White H, Lu X. Causal diagrams for treatment effect estimation with application to efficient covariate selection. *Review of Economics and Statistics* 2011;doi:10.1162/REST_a_00153:Forthcoming.

[6] Patrick AR, Schneeweiss S, Brookhart MA, et al. The implications of propensity score variable selection strategies in pharmacoepidemiology: An empirical illustration. *Pharmacoepidemioly and Drug Safety* 2011; doi: 10.1002/pds.2098:In press.

[7] Hirano K, Imbens G. Estimation of causal effects using propensity score weighting: an application to data on right heart catheterization. *Health Services and Outcomes Research Methodology* 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 group? *Statistics in Medicine* 2009;28:1420–1423.

[9] Pearl J. Myth, confusion, and science in causal analysis. Tech. Rep. R-348, University of California, Los Angeles, CA, 2009. <http://ftp.cs.ucla.edu/pub/stat_ser/r348.pdf>.

[10] Day N, Byar D, Green S. Overadjustment in case-control studies. *American Journal of Epidemiology* 1980;112(5):696–706.

[11] Thomas D, Greenland S. The relative efficiencies of matched and independent sample designs for case-control studies. *Journal of Chronic Diseases* 1983;36(10):685–697.

[12] Hill J. Comments on 'A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003' by Peter Austin, *Statistics in Medicine*. *Statistics in Medicine* 2008;27:2055–2061.

[13] Austin P. A critical appraisal of propensity-score matching in the medical literature from 1996 to 2003. *Statistics in Medicine* 2008; 27(12):2037–2049.

[14] Morgan S, Winship C. *Counterfactuals and Causal Inference: Methods and Principles for Social Research (Analytical Methods for Social Research).* New York, NY: Cambridge University Press, 2007.

[15] Pearl J, Paz A. Confounding equivalence in causal inference. In: *Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence*, Corvallis, OR: AUAI. 2010;433–441.

[16] Hahn J. Functional restriction and efficiency in causal inference. *Review of Economics and Statistics* 2004;86:73–76.

[17] Bareinboim E, Pearl J. Controlling selection bias in causal inference. Tech. Rep. R-381, <http://ftp.cs.ucla.edu/pub/stat_ser/r381.pdf>, University of California Los Angeles, Computer Science Department, CA, 2011.