



Response to Invited Commentary

Myers et al. Respond to “Understanding Bias Amplification”

**Jessica A. Myers*, Jeremy A. Rassen, Joshua J. Gagne, Krista F. Huybrechts,
 Sebastian Schneeweiss, Kenneth J. Rothman, and Robert J. Glynn**

* Correspondence to Dr. Jessica A. Myers, Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, 1620 Tremont Street, Suite 3030, Boston, MA 02120 (e-mail: jmyers6@partners.org).

Initially submitted September 6, 2011; accepted for publication September 9, 2011.

The threat of bias amplification extends to all epidemiologic methods that rely on conditioning on covariates, so understanding its mechanisms and implications is important. We thank Dr. Pearl (1) for initially bringing the issue to our attention and for now extending the analysis of bias amplification to the case of multiple covariates (2). Here, as in our original paper (3), we attempt to clarify the implications of these results in practical settings. In particular, we focus on secondary analysis of electronically collected data sets, where there may be hundreds or thousands of measured variables available for use in adjustment. Many of these covariates may be proxies of true confounders, even if their relations with outcome and exposure are uncertain.

In his commentary (2), Pearl showed that adjusting for a variable related to exposure may either increase or decrease the net confounding but will always reduce the unexplained variation in the exposure (thereby amplifying any remaining confounding bias). Thus, the potential for bias amplification compounds with each additional covariate adjustment, and large bias amplification is possible, even if each individual covariate has a relatively weak association with exposure. However, Pearl's analysis assumes that there is an important unmeasured confounder that cannot be controlled. In that scenario, the study is flawed even without consideration of bias amplification. On the other hand, if the residual confounding can be eliminated or nearly eliminated via additional covariate adjustments, then the bias will be small regardless of the amount of independent variation in exposure.

Pearl developed a partial preference order for selection of covariates that “are more related to the output,” interpreted as variables with a more direct causal path to the outcome. Unfortunately, such nuanced information about the causal structure is rarely available, especially in data sets with many variables. An investigator tasked with specifying a causal structure for a high-dimensional data set would probably specify a graph like Pearl's Figure 2 (2, p. 000), but with 1,000 confounders rather than 2. This structure alone, even

if known to be true, is not sufficient for identifying which of the potential covariates are bias reducers or bias amplifiers. For example, in Pearl's Figure 2, the parameter values $\alpha_0 = \alpha_1 = \alpha_2 = 0.6$ and $\beta_0 = \beta_1 = \beta_2 = 0.1$ imply that both Z_1 and Z_2 are bias amplifiers, based on his equations 2–4. With the same β values and only a slight reduction in the α values to 0.57, adjusting for Z_1 or Z_2 reduces bias. In both scenarios, the causal structure would indicate a need for adjustment.

We differ with Pearl's unqualified conclusion that “strong predictors of exposure should be excluded from the analysis” (2, p. 000). This suggestion would eliminate control of strong confounders, which are strong predictors of both exposure and outcome. Furthermore, in the analysis of the near-instrumental variable (Pearl's Figure 3), we disagree with his statement indicating that we assumed that γ_1 controls the reduction of confounding bias only. Our simulation studies were designed specifically to better understand the complex interactions between confounding reduction and bias amplification that were expected when Z was a near-instrumental variable.

The primary question of bias amplification that remains is how to distinguish in practice between the bias reducers and the bias amplifiers. As we have suggested previously (4, 5), screening covariates for strong associations with exposure but not outcome can identify potential bias amplifiers. However, as shown in the example above, even when a covariate has an association with exposure that is more than 5 times its association with the outcome, the variable can still reduce bias. Therefore, we maintain that bias amplification is a secondary concern compared with residual confounding and that investigators should err on the side of inclusion rather than exclusion of potential confounders.

ACKNOWLEDGMENTS

Author affiliations: Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham

and Women's Hospital and Harvard Medical School, Boston, Massachusetts (Jessica A. Myers, Jeremy A. Rassen, Joshua J. Gagne, Krista F. Huybrechts, Sebastian Schneeweiss, Kenneth J. Rothman, Robert J. Glynn); and RTI Health Solutions, Research Triangle Park, North Carolina (Kenneth J. Rothman).

Conflict of interest: none declared.

REFERENCES

1. Pearl J. On a class of bias-amplifying variables that endanger effect estimates. In: Grünwald P, Spirtes P, eds. *Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence (UAI 2010)*. Corvallis, OR: Association for Uncertainty in Artificial Intelligence; 2010:425–432.
2. Pearl J. Invited commentary: understanding bias amplification. *Am J Epidemiol*. 2011;00(0):000–000.
3. 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;00(0):000–000.
4. Schneeweiss S, Rassen JA, Glynn RJ, et al. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiology*. 2009;20(4):512–522.
5. Brookhart MA, Stürmer T, Glynn RJ, et al. Confounding control in healthcare database research: challenges and potential approaches. *Med Care*. 2010;48(6 suppl):S114–S120.