Vol. 191, No. 1
https://doi.org/10.1093/aje/kwab228
Advance Access publication:
August 28, 2021

Amplification of Bias Due to Exposure Measurement Error

David B. Richardson*, **Alexander P. Keil, and Stephen R. Cole**

* Correspondence to Dr. David B. Richardson, Department of Environmental and Occupational Health, Program in Public Health, University California, Irvine, 653 East Peltason Drive, Anteater Instruction and Research Building, Room 2084, Irvine, CA 92697-3957 (e-mail: david.richardson@uci.edu).

Initially submitted November 9, 2020; accepted for publication August 26, 2021.

Observational epidemiologic studies typically face challenges of exposure measurement error and confounding. Consider an observational study of the association between a continuous exposure and an outcome, where the exposure variable of primary interest suffers from classical measurement error (i.e., the measured exposures are distributed around the true exposure with independent error). In the absence of exposure measurement error, it is widely recognized that one should control for confounders of the association of interest to obtain an unbiased estimate of the effect of that exposure on the outcome of interest. However, here we show that, in the presence of classical exposure measurement error, the net bias in an estimate of the association of interest may increase upon adjustment for confounders. We offer an analytical expression for calculating the change in net bias in an estimate of the association of interest upon adjustment for a confounder in the presence of classical exposure measurement error, and we illustrate this problem using simulations.

bias; cohort studies; epidemiologic methods; regression analysis

Observational epidemiologic studies typically face challenges of exposure measurement error and confounding. In the absence of exposure measurement error, it is widely recognized that we should control for well-measured confounders of the association of interest if we wish to obtain an unbiased estimate of an effect of the exposure on the outcome of interest. In the presence of exposure measurement error, one might assume that the same logic holds. However, if we have an imperfect measure of the exposure of interest that suffers from classical measurement error and we wish to obtain the least biased estimate of the association of interest, this logic may not hold.

Here, focusing on the context of covariate-conditional outcome regression modeling of the association between an error-prone exposure variable and an outcome, we posit that upon adjustment for a confounder, a reduction in bias due to confounding may be accompanied by an amplification of bias due to classical exposure measurement error. Such a situation is not necessarily one of a bias-variance tradeoff. To the contrary, there may be situations where, upon adjustment for a confounder, net bias may increase, and, as a consequence of increasing the number of parameters to be estimated, the variance of the estimate of the association of interest may increase as well.

In this paper, we consider the impact of adjusting for confounders on net bias in an estimate of an exposure-disease association in the presence of classical exposure measurement error. First, we consider net bias in an estimate upon adjustment for a single covariate. Second, we consider the change in net bias as the number of covariates adjusted for in a regression analysis increases. We describe the problem, offer an analytical expression for the change in net bias in an estimate of the association of interest upon adjustment for a confounder in the presence of classical exposure measurement error, and illustrate it with simulations.

METHODS

Consider the setting of an epidemiologic study in which there is a continuous exposure of primary interest, X, a continuous outcome, Y, and a continuous potential confounder, Z. However, we do not observe X; rather, we observe an imperfect exposure measure, X^*, that suffers from classical measurement error of the form $X^* = X + U, U \sim N\left(0, \sigma_U^2\right)$.

Figure 1 includes a structural representation of measurement error in the exposure of primary interest, X. The figure also includes a structural representation of confounding of the association between X and the outcome of primary

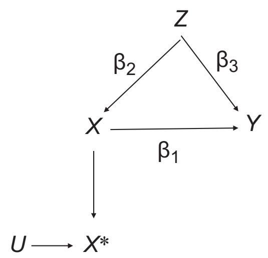

Figure 1. Simple diagram of a study setting in which there is exposure measurement error and confounding.

interest, Y, by a third variable Z. The parameter β_1 describes a conditional effect of X on Y, holding Z constant; the parameter β_2 describes an effect of Z on X; and the parameter β_3 describes a direct effect of Z on Y, holding X constant.

Suppose our interest is in the conditional effect of X on Y, given by

$$E(Y|X = x + 1, Z) - E(Y|X = x, Z) = \beta_1.$$

However, we do not observe X but rather observe the imperfect proxy, X^*. Assuming correct model specification, there are 2 options that an investigator could consider. First, fit a regression model for Y on X^* that is adjusted for Z:

$$E(Y|X^* = x^* + 1, Z) - E(Y|X^* = x^*, Z) = \gamma_1 \quad \text{(model 1)}$$

Second, fit a regression model for Y on X^* that is not adjusted for Z:

$$E(Y|X^* = x^* + 1) - E(Y|X^* = x^*) = \delta_1 \qquad \text{(model 2)}$$

Using the term "bias" to describe the difference between the parameter of interest, β_1, and the parameters given under models 1 and 2, bias is given by the expressions

$$\beta_1 - \gamma_1 = \beta_1\left(1 - \frac{\mathrm{Var}(X|Z)}{\mathrm{Var}(X|Z) + \sigma_U^2}\right)$$

and

$$\beta_1 - \delta_1 = \beta_1\left(1 - \frac{\mathrm{Var}(X)}{\mathrm{Var}(X) + \sigma_U^2}\right) - \beta_3[E(Z|X^* = x^* + 1) - E(Z|X^* = x^*)].$$

These expressions follow from general formulas for bias in linear models due to classical measurement error (1) and for bias due to confounding (2).

Should an investigator adjust for Z if she wishes to obtain the least biased estimate of the association of interest? There are settings in which bias in the estimate obtained under model 1 exceeds bias in the estimate obtained under model 2. This occurs when

$$\left|\beta_1\left(1 - \frac{\mathrm{Var}(X|Z)}{\mathrm{Var}(X|Z) + \sigma_U^2}\right)\right| > \left|\beta_1\left(1 - \frac{\mathrm{Var}(X)}{\mathrm{Var}(X) + \sigma_U^2}\right) - \beta_3[E(Z|X^* = x^* + 1) - E(Z|X^* = x^*)]\right|.$$

From the expression above, it can be seen that an increase in bias upon conditioning on Z will not occur when $\beta_1 = 0$ (i.e., in the absence of a true exposure effect); nor will there be an increase in bias upon conditioning on Z when $\sigma_U^2 = 0$ (i.e., in the absence of classical measurement error); nor will it occur when $\mathrm{Var}(X|Z) = \mathrm{Var}(X)$, which happens when $\beta_2 = 0$ (i.e., in the absence of a confounder-exposure association). However, an increase in net bias may occur when $\beta_3 = 0$. More generally, an increase in bias upon conditioning on Z will tend to occur when there is a true exposure effect and classical measurement error (i.e., β_1 and σ_U^2 diverge from 0) and when Z is strongly associated with the exposure of interest but weakly associated with the outcome of interest (i.e., β_2 diverges from 0 and β_3 approaches 0); such covariates are sometimes referred to as near-instruments or "instrument-like" variables (3).

We use simulations to illustrate results implied by this analytical expression for a range of scenarios.

Example 1: simple case

We simulated data where exposure was associated with the outcome, under the structural model illustrated in Figure 1, and there was classical measurement error. We generated a single covariate, Z, by sampling from a normal distribution with zero mean and unit variance. We generated the exposure variable of interest under the model $X = \beta_2 Z + N(0, 1)$, the outcome variable of interest under the model $Y = \beta_1 X - \beta_3 Z + N(0, 1)$, and an imperfect proxy exposure measure, X^*, under the model $X^* = X + N(0, \sigma_U^2)$. Cohorts of 20,000 subjects were generated under the conditions $\beta_1 = 1, \beta_2 = \{0.25, 0.5, 1\}, \beta_3 = \{0, 0.25, 0.5, 1\}$, and $\sigma_U^2 = \{0.5, 1\}$. For each simulated cohort, we performed a linear regression of Y on X^* adjusted for Z and a regression of Y on X^* not adjusted for Z, and we calculated bias as the difference between the true value of β_1 and each estimate. We note that the measure of association from this model is collapsible, so that in the absence of confounding and effect-measure modification, Z-conditional estimates of the Y-X association would be identical to the unconditional estimate of the Y-X association. The Web Appendix illustrates bias in settings where X is a binary exposure variable that is misclassified.

Example 2: multiple potential confounders

We simulated data for 5 covariates, Z_1–Z_5, by sampling from the multivariate normal distribution with zero means, unit variances, and zero covariance (Figure 2). We generated the exposure variable of interest under the model $X = \beta_2 Z_1 + \beta_2 Z_2 + \beta_2 Z_3 + \beta_2 Z_4 + \beta_2 Z_5 + N(0, 1)$, the outcome

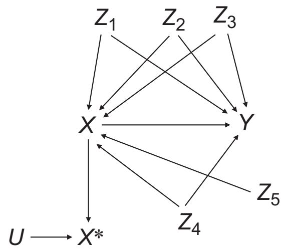

Figure 2. Diagram of a study setting in which there is exposure measurement error and multiple confounders.

variable of interest under the model $Y = \beta_1 X + \beta_3 Z_1 - \beta_3 Z_2 + \beta_3 Z_3 - \beta_3 Z_4 + N(0,1)$, and an imperfect proxy exposure measure, X^*, under the model $X^* = X + N\left(0, \sigma_U^2\right)$. Cohorts of 20,000 subjects were generated under the conditions $\beta_1 = 1, \beta_2 = \{0.5, 1\}, \beta_3 = \{0.5, 1\}$, and $\sigma_U^2 = \{0.5, 1\}$. For each simulated cohort, we performed an ordinary linear regression of Y on X^* adjusted for Z_1–Z_5, a regression of Y on X^* adjusted for Z_1–Z_4, a regression of Y on X^* adjusted for Z_1–Z_3, a regression of Y on X^* adjusted for Z_1 and Z_2, a regression of Y on X^* adjusted for Z_1, and an unadjusted regression of Y on X^*. We calculated bias as the difference between β_1 and each estimate.

RESULTS

Example 1: simple case

As expected from the analytical expression, when β_3 was set equal to 0, an increase in bias occurred in all simulations upon conditioning on Z (Table 1). In such cases, the unconditional estimate of association between X^* and Y was closer to β_1 than the Z-conditional estimate of association. The net bias upon conditioning on Z increased with the degree of classical measurement error, σ_U^2, and with the magnitude of the association between Z and X, β_2. As β_3 diverged from 0, bias in the crude estimate of association increased, and it eventually exceeded that of the Z-adjusted estimate (Table 1). When β_3 was greater than 0 (i.e., Z was associated with the outcome of interest), an increase in net bias upon conditioning on Z was observed in some, but not all, simulation scenarios; it tended to be observed when the degree of classical measurement error, σ_U^2, was largest and when the magnitude of association between Z and X, β_2, was greatest. Web Table 1 (available at https://doi.org/10.1093/aje/kwab228) provides results derived over the range of scenarios considered and illustrates that the Z-conditional estimate of association may have a larger standard error than the unconditional estimate of association as well. The Web Appendix provides results for simulations in which X is a binary variable that is misclassified.

Example 2: multiple covariates

We fitted a regression model for Y on X^* that was unadjusted for Z_1–Z_5, and we fitted a regression model for Y on X^* under a stepwise evaluation of the impact of adjustment for covariates Z_1–Z_5. The unadjusted estimate suffered from the least amount of bias (Table 2); and upon adjustment for each additional covariate, there was a marked change in the estimate of the X^*-Y association. Under the simulation setup, there was no net confounding of the X-Y association; the confounding effects of Z_1 and Z_2 canceled out, as did the effects of Z_3 and Z_4. Consequently, the estimates of association obtained under models 1, 3, 5, and 6 were unconfounded; nonetheless, bias was greater under model 3 than under model 1, greater under model 5 than under model 3, and greater under model 6 than under model 5. Notably, adjustment for Z_5 led to a change in estimate and an increase in bias, despite Z_5's not being a confounder in the underlying data-generating model (Figure 2). The estimate obtained upon fitting model 6 was half as large as β_1, the specified association between X and Y under the simulation setup. Models 2 and 4 were confounded, in addition to suffering from bias due to exposure measurement error. Web Table 2 provides results derived over a range of simulation scenarios.

DISCUSSION

Confounding is a routine challenge in observational epidemiology. Assessing confounding is often an important step in decisions regarding whether or not a factor should be adjusted for in statistical analysis and regarding how one may interpret unadjusted associations. Here we show that problems of confounding often are entangled with problems of exposure measurement error, reinforcing observations made much earlier regarding confounding and misclassification of a binary exposure variable (4). This work also builds upon prior literature on exposure measurement error which has shown that the degree of attenuation due to classical exposure measurement error tends to increase as variables that are predictors of exposure are introduced into a linear regression model (1). However, we are not aware of either examples of or discussions about the implications for epidemiologic analyses in which the exposure of interest suffers from measurement error. The fact that, upon adjustment for a confounder, a reduction in bias due to confounding may be accompanied by an even more substantial amplification in bias due to classical exposure measurement error bias amplification has received relatively little attention. This should lead to careful consideration of the net bias that arises upon conditioning on a covariate, which encompasses confounding as well as potential exacerbation of bias due to measurement error.

The problem may be avoided by recognizing the conditions under which the problem tends to arise. As we have shown via an analytical expression for net bias and have illustrated via some simple simulations, this problem tends to occur when there is substantial classical exposure measurement error and an investigator conditions upon covariates that are strongly associated with the exposure of interest but weakly associated with the outcome of interest. Therefore, efforts to reduce classical measurement error, as well as efforts to avoid conditioning on instrument-like variables, should remedy the problem. Alternatively, to avoid this problem, we

Table 1. Simulation Results Illustrating Bias in the *Z*-Conditional and Crude Estimates of an *X*-*Y* Association Under Varying Degrees of Classical Measurement Error and Varying Values of the Parameters β_2 and β_3 Which Determine the Degree of Confounding of the *X*-*Y* Association by *Z*

Simulation Setup				Regression Estimate		Absolute Bias	
β_1	σ^2_U	β_2	β_3	*Z*-Conditional ($\hat{\gamma}_1$)	Crude ($\hat{\delta}_1$)	*Z*-Conditional ($\mid \beta_1 - \gamma_1 \mid$)	Crude ($\mid \beta_1 - \delta_1 \mid$)
1.00	1.00	1.00	0.00	0.50	0.66	0.50	0.34
			0.25	0.50	0.58	0.50	0.42
			0.50	0.50	0.50	0.50	0.50
			1.00	0.50	0.33	0.50	0.67
		0.50	0.00	0.50	0.55	0.50	0.45
			0.25	0.50	0.50	0.50	0.50
			0.50	0.50	0.45	0.50	0.55
			1.00	0.50	0.34	0.50	0.66
		0.25	0.00	0.50	0.51	0.50	0.49
			0.25	0.50	0.49	0.50	0.51
			0.50	0.50	0.46	0.50	0.54
			1.00	0.50	0.40	0.50	0.60
	0.50	1.00	0.00	0.67	0.80	0.33	0.20
			0.25	0.67	0.70	0.33	0.30
			0.50	0.67	0.60	0.33	0.40
			1.00	0.67	0.40	0.33	0.60
		0.50	0.00	0.67	0.71	0.33	0.29
			0.25	0.67	0.64	0.33	0.36
			0.50	0.67	0.57	0.33	0.43
			1.00	0.67	0.43	0.33	0.57
		0.25	0.00	0.67	0.68	0.33	0.32
			0.25	0.67	0.64	0.33	0.36
			0.50	0.67	0.60	0.33	0.40
			1.00	0.67	0.53	0.33	0.47

recently illustrated an approach to obtaining a marginal estimate of association that allows for control of confounding without conditioning on covariates and does not suffer from this problem of measurement error bias amplification (5). In that prior work, we proposed a method for estimating a marginal (i.e., standardized) exposure-outcome association that will suffer from less bias due to measurement error than a covariate-conditional estimate of association (5); we did not address the implications for variable selection or the net bias that may arise when a reduction in confounding is accompanied by amplification of bias due to classical exposure measurement error.

Variable selection often is a needed part of model-building. In regression modeling, controlling for too many potential confounders can lead to problems of data sparsity, particularly when the number of covariates is large in relation to the study size. Strategies such as the "change-in-estimate" approach offer a simple approach to assessing confounding by comparing estimated associations derived with and without adjustment for a potential confounder. Prior authors have cautioned against this approach, pointing out that conditional and marginal measures of association can differ absent confounding (and that confounding can occur even when conditional and marginal effect measures are equal) due to noncollapsibility (6). Here, we show that when the exposure variable of interest is measured with classical error, the "change-in-estimate" approach may fail because adjustment for a covariate associated with exposure may exacerbate bias due to classical measurement error. Absent further information, in a change-in-estimate approach we cannot distinguish a reduction in confounding bias from amplification of exposure measurement error bias. We illustrate that even if individually no single covariate is a strong predictor of the exposure of interest, adjustment for an ensemble of covariates may explain a substantial proportion of the variation in the exposure of interest, resulting in

Table 2. Simulation Results Illustrating an Increase in Bias with Stepwise Increasing Adjustment for Covariates[a]

Model	Regression Model	Regression Estimate	Absolute Bias
Unconditional	$E[Y\|X^*] = \alpha_0 + \gamma X^*$	0.86	0.14
Adjusted for Z_1	$E[Y\|X^*, Z_1] = \alpha_0 + \alpha_1 Z_1 + \gamma X^*$	0.66	0.34
Adjusted for Z_1 and Z_2	$E[Y\|X^*, Z_1, Z_2] = \alpha_0 + \alpha_1 Z_1 + \alpha_2 Z_2 + \gamma X^*$	0.80	0.20
Adjusted for Z_1–Z_3	$E[Y\|X^*, Z_1, Z_2, Z_3] = \alpha_0 + \alpha_1 Z_1 + \alpha_2 Z_2 + \alpha_3 Z_3 + \gamma X^*$	0.50	0.50
Adjusted for Z_1–Z_4	$E[Y\|X^*, Z_1, Z_2, Z_3, Z_4] = \alpha_0 + \alpha_1 Z_1 + \alpha_2 Z_2 + \alpha_3 Z_3 + \alpha_4 Z_4 + \gamma X^*$	0.67	0.33
Adjusted for Z_1–Z_5	$E[Y\|X^*, Z_1, Z_2, Z_3, Z_4, Z_5] = \alpha_0 + \alpha_1 Z_1 + \alpha_2 Z_2 + \alpha_3 Z_3 + \alpha_4 Z_4 + \alpha_5 Z_5 + \gamma X^*$	0.50	0.50

[a] Unconditional and conditional estimates of association obtained upon stepwise adjustment for potential confounders, $\boldsymbol{Z}$. In the scenarios shown, $\beta_1 = 1$, $\beta_2 = 1$, $\beta_3 = 1$, and $\sigma_U^2 = 1$.

amplification of bias due to classical exposure measurement error (Table 2).

The problem of bias amplification due to covariate control with exposure measurement error discussed here shares some similarities with the problem of unmeasured confounder bias amplification discussed by Pearl (3) and others (7, 8). Similarly, bias amplification may arise upon adjustment for "instrument-like" variables that are strong predictors of the exposure (but have weak or no association with the outcome). While some have questioned how relevant unmeasured confounder bias amplification is in practice (9), it seems that exposure measurement error bias amplification is a reasonable concern in settings where an investigator may adjust for many potential confounders. Similarly, the problem may arise in regression models that adjust for coexposures that are correlated with the exposure of interest because they arise from a common source (as occurs in environmental epidemiologic studies of exposure mixtures).

While we have focused on net bias in the estimate of an association of interest, the problem that we address is not necessarily one of a bias-variance trade-off. To the contrary, our simple simulations illustrate that upon adjustment for a confounder, net bias may increase, and by increasing the number of parameters to be estimated, the variance of the estimate of the association of interest may increase as well (Web Table 1). We also illustrate simulation results for a misclassified binary exposure variable; while the measurement error problem is somewhat different from that of a continuous variable, similar conclusions hold about the potential for bias amplification upon conditioning for an instrument-like variable (Web Appendix and Web Table 3).

As with all statistical modeling, it is important to consider the structure of the association between variables, whether candidate covariates affect the outcome, and, in the setting of classical exposure measurement error, whether candidate covariates may be "instrument-like" variables in order to avoid simply including in outcome regression models covariates that are strong predictors of the exposure, that are weak confounders, and that upon adjustment amplify rather than diminish net bias.

ACKNOWLEDGMENTS

Author affiliations: Department of Epidemiology, School of Public Health, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States (David B. Richardson, Alexander P. Keil, Stephen R. Cole); and Department of Environmental and Occupational Health, Program in Public Health, University of California, Irvine, Irvine, California, United States (David B. Richardson).

D.B.R. received funding support from the National Institute for Occupational Safety and Health (grant R01OH011409).

Conflict of interest: none declared.

REFERENCES

1. Carroll RJ, Ruppert D, Stefanski LA. *Measurement Error in Nonlinear Models: A Modern Perspective*. 2nd ed. London, United Kingdom: Chapman & Hall Ltd.; 2006.
2. Vanderweele TJ, Arah OA. Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments, and confounders. *Epidemiology*. 2011;22(1):42–52.
3. Pearl J. Invited commentary: understanding bias amplification. *Am J Epidemiol*. 2011;174(11):1223–1227.
4. Greenland S, Robins JM. Confounding and misclassification. *Am J Epidemiol*. 1985;122(3):495–506.
5. Richardson DB, Keil AP, Cole SR, et al. Reducing bias due to exposure measurement error using disease risk scores. *Am J Epidemiol*. 2021;190(4):621–629.

6. Greenland S, Pearl J, Robins JM. Confounding and collapsibility in causal inference. *Stat Sci.* 1999;14(1):29–46.
7. Weisskopf MG, Seals RM, Webster TF. Bias amplification in epidemiologic analysis of exposure to mixtures. *Environ Health Perspect.* 2018;126(4):047003.
8. Brookhart MA, Schneeweiss S, Rothman KJ, et al. Variable selection for propensity score models. *Am J Epidemiol.* 2006; 163(12):1149–1156.
9. Evans D, Chaix B, Lobbedez T, et al. Combining directed acyclic graphs and the change-in-estimate procedure as a novel approach to adjustment-variable selection in epidemiology. *BMC Med Res Methodol.* 2012;12:156.