Revisiting Overadjustment Bias

To the Editor:

In epidemiology, overadjustment bias is defined as adjusting for a variable that increases rather than decrease bias, while unnecessary adjustment is referred to as control for a variable that adversely affects precision without introducing bias.[1,2] The term of overadjustment bias is used to refer to different scenarios. Here, we propose a unified definition of overadjustment, with four types.

Type 1 overadjustment is the classic adjustment for an intermediate variable or a descendant of an intermediate variable (see Table). Schisterman et al[1] explained type 1 overadjustment bias in detail. If the total effect of the exposure on the outcome is of interest, adjusting for an intermediate variable, or a downstream proxy of an intermediate variable may lead to overadjustment bias because (in ideal settings) such adjustment results in an estimate of the controlled direct effect instead of the total effect.

Type 2 overadjustment is adjustment for a collider or a descendant of a collider.[3] Collider bias is defined as a bias that occurs when conditioning on a common effect (or a descendant of a common effect) of two variables, of which one is the exposure or a cause of the exposure, and the other is the outcome or a cause of the outcome.[4] Here, we make a distinction between collider restriction bias and collider stratification bias. Collider restriction bias due to restricting to one level of a collider is a form of selection bias; however, collider stratification bias through (for example) the unnecessary inclusion of a collider in a regression model may be best understood as a type of overadjustment bias instead of selection bias. However, additionally blocking another covariate that lies on a backdoor path that is opened by collider adjustment, if possible, will obviate the bias.

Type 3 overadjustment is adjustment for an instrumental variable in the presence of unmeasured confounding. An instrumental variable is a cause of the exposure that has no relation to the outcome except through the exposure.[5] It has been shown that in a regression model of the outcome on the exposure, adjusting for a strong instrumental variable (more broadly, strong predictors of the exposure), in the presence of unmeasured confounding, has the potential to amplify bias as well as affect precision, which we term an overadjustment bias.[6,7] Amplification occurs because, within strata of the exposure, the instrumental variable is associated with the outcome via the unmeasured confounder. Thus, in practice, it is best to exclude pure instrumental variables from covariate control.

Type 4 overadjustment is adjustment for a descendant of the outcome when the exposure has a causal effect on the outcome. As shown in the Table, adjusting for a descendant of the outcome D is in fact adjusting for a descendant of the collider, when there is an another cause of outcome D (that is L) that makes the outcome D a collider.[8] As a result, it induces a spurious association between E and L and hence between E and D, therefore resulting in biased effect estimates of exposure E on outcome D on risk difference and risk ratio scales. Thus, adjustment for a descendant of the outcome can also be regarded as a special type of collider bias.

Here, we summarized four types of covariate adjustment that may induce bias, and, hence, we call these overadjustment biases. Consciously avoiding these types of overadjustment bias, together with adjustment for a sufficient set of confounders (both of which rely on substantive knowledge), is necessary to satisfy the assumption of correct causal model specification in the standard set of identification conditions.[9] Correct causal model specification, followed by correct statistical model specification (which determines the functional form and relationships between exposure, covariates, and outcome), along with other standard assumptions (e.g., no measurement error), constitute an important set of sufficient conditions for the identification of causal effects in epidemiology and beyond.[10,11] In some circumstances, a variable may play more than one role, and while this is beyond the scope of the present paper, whether or not to adjust for such variable may well become an issue of minimizing overall bias.

ACKNOWLEDGMENTS

The authors thank the anonymous reviewer of this letter for their helpful comments.

Haidong Lu
Public Health Modeling Unit and
Department of Epidemiology of
Microbial Diseases
Yale School of Public Health
New Haven, CT

Stephen R. Cole
Department of Epidemiology
Gillings School of Global Public Health
University of North Carolina at Chapel Hill
Chapel Hill, NC

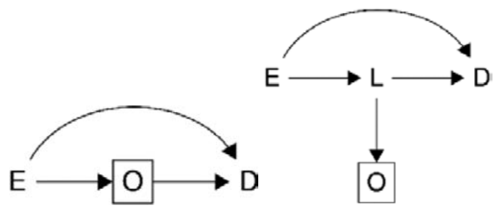
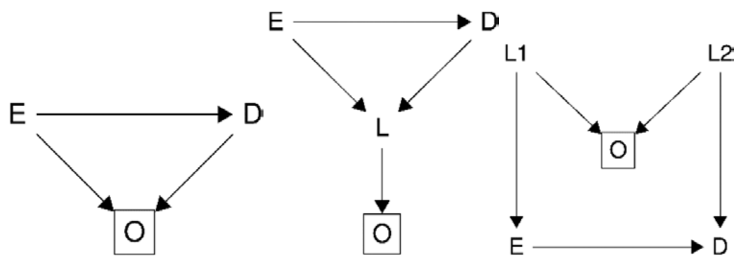
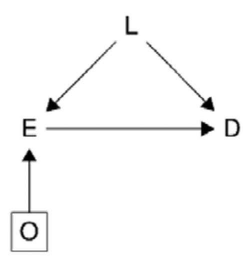
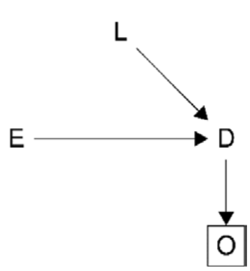
Robert W. Platt
Departments of Pediatrics and Epidemiology
Biostatistics and Occupational Health
McGill University
Montreal, QC, Canada

Enrique F. Schisterman
Epidemiology Branch
Division of Intramural Population Health
Research
Eunice Kennedy Shriver National Institute
of Child Health and Human Development
Bethesda, MD

The authors report no funding and conflicts of interest.

Correspondence: Haidong Lu, Public Health Modeling Unit and Department of Epidemiology of Microbial Diseases, Yale School of Public Health, 350 George Street, Ste 3rd Floor, New Haven, CT 06511. E-mail: haidong.lu@yale.edu

Table. Four Types of Overadjustment Bias

Type of Overadjustment Bias	Causal Diagrams	Explanation
Type 1—adjusting for an intermediate variable or a descendant proxy of an intermediate variable		Adjusting for an intermediate variable or a descendant of an intermediate variable can bias estimates of the total effect of exposore E on outcome D.
Type 2—adjusting for a collider or a descendant of a collider		Adjusting for a collider or a descendant of a collider will bias the effect estimate of exposure E on outcome D.
Type 3—adjusting for an instrumental variable		Adjusting for an instrumental variable in the presence of unmeasured confounding (when L is not measured) can lead to amplified bias.
Type 4—adjusting for a descendant of the outcome		Adjusting for a descendant of the outcome D can result in biased effect estimate on risk difference and risk ratio scales when there is a causal effect of exposure E on outcome D.

We consider only four types of variables: exposure E, outcome D, covariates L (L1, L2), and variable O, which can lead to overadjustment bias if adjusted for.

REFERENCES

1. Schisterman EF, Cole SR, Platt RW. Overadjustment bias and unnecessary adjustment in epidemiologic studies. *Epidemiology.* 2009;20:488–495.
2. VanderWeele TJ. On the relative nature of overadjustment and unnecessary adjustment. *Epidemiology.* 2009;20:496–499.
3. Cole SR, Platt RW, Schisterman EF, et al. Illustrating bias due to conditioning on a collider. *Int J Epidemiol.* 2010;39:417–420.
4. Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. *Epidemiology.* 2003;14:300–306.
5. Greenland S. An introduction to instrumental variables for epidemiologists. *Int J Epidemiol.* 2000;29:722–729.
6. Pearl J. Invited commentary: understanding bias amplification. *Am J Epidemiol.* 2011;174:1223–1227; discussion pg 1228.
7. Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Stürmer T. Variable selection for propensity score models. *Am J Epidemiol.* 2006;163:1149–1156.
8. Daniel RM, Kenward MG, Cousens SN, De Stavola BL. Using causal diagrams to guide analysis in missing data problems. *Stat Methods Med Res.* 2012;21:243–256.
9. Cole SR, Hernán MA. Constructing inverse probability weights for marginal structural models. *Am J Epidemiol.* 2008;168:656–664.
10. Keil AP, Mooney SJ, Jonsson Funk M, Cole SR, Edwards JK, Westreich D. Resolving an apparent paradox in doubly robust estimators. *Am J Epidemiol.* 2018;187:891–892.
11. Schomaker M. Regression and causality. [published online ahead of print June 21, 2021]. *Stat Theory.* Available at: https://arxiv.org/abs/2006.11754. Accessed 10 May 2021.