Published in final edited form as:

Epidemiology. 2017 May ; 28(3): e27–e28. doi:10.1097/EDE.0000000000000620.

Estimating the Causal Effect of an Exposure on Change from Baseline Using Directed Acyclic Graphs and Path Analysis

Cooper S Schumacher and

Department of Biostatistics, University of Washington, Seattle, WA, USA

Lianne Sheppard

Department of Environmental and Occupational Health Sciences, University of Washington, Seattle, WA, USA

Department of Biostatistics, University of Washington, Seattle, WA, USA

To the Editor

The recent article by Lepage et al.¹ considers estimation of the causal effect of an exposure on an outcome's change from baseline. The authors use directed acyclic graphs (DAGs) to explain how adjustment for an outcome's measured baseline in linear regression of change analyses will influence the bias in estimation of the exposure effect.

The authors provided a very helpful appendix with the code to replicate their analyses in the article's eAppendix (<http://links.lww.com/EDE/A839>). In reproducing their work we found minor errors in the simulation code. The fitted regression models were over-specified and coefficient values of zero were recorded in several iterations of the simulations where the exposure variable of interest was dropped. The Table presents the values reported in the article and the corrected values; our own eAppendix gives our corrections to the code. We applaud the authors for providing supplemental material and code that allowed for their work to be replicated. As is discussed in Peng², transparency of statistical procedures and code is a major component to reproducible research.

While the errors we uncovered somewhat affected the results, they did not change the conclusions of the paper. DAGs provide a useful guide for deciding whether linear regression of change models should be adjusted for the outcome's measured baseline. The setting of “non-randomized studies where exposure starts before the beginning of the study” is of particular interest to many environmental health studies that consider exposures such as air pollutants. Lepage et al. used DAGs to show that “unadjusted” models of change that don't condition on the baseline outcome measurement will provide less biased estimates in this setting. However, the analysis used in Lepage et al. was limited to regression of change as an outcome applied to data in which all subjects have two equally spaced measurement

Corresponding Author: Lianne Sheppard, Box 357232, University of Washington, Seattle, WA 98195-7232, 206-616-2722, sheppard@uw.edu.

Conflicts of Interest:
None declared

times. In many observational studies where a scientific interest is in rate of change or progression in an outcome measure a more general design may be required that allows for multiple unequally spaced follow-up times, and exposure/covariate frameworks that can vary over time. It is important to be able to generalize the outcome approach to account for a more general design framework, yet maintain the “unadjusted” for measured outcome quality recommended by Lepage et al. We propose one option is to fit a mixed model that controls for a modeled baseline, similar to the models used in several MESA Air study analyses^{3,4}. Controlling for a modeled baseline will not have the same bias as that caused by controlling for the measured baseline when there is outcome measurement error. Further work should consider how various plausible analysis approaches compare in order to determine how to best obtain an unbiased estimate of the effect of exposure on rate of change in a more general design setting.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Sources of Financial Support

This work was supported in part by the UW NIEHS sponsored Biostatistics, Epidemiologic and Bioinformatic Training in Environmental Health (BEBTEH) Training Grant, Grant #: NIEHS T32ES015459.

References

1. Lepage B, Lamy S, Dedieu D, Savy N, Lang T. Estimating the causal effect of an exposure on change from baseline using directed acyclic graphs and path analysis. *Epidemiology*. 2015; 26:122–129. [PubMed: 25401453]
2. Peng R. The reproducibility crisis in science: A statistical counterattack. *Significance*. 2015; 12(3): 30–32.
3. Gassett AJ, Sheppard L, McClelland RL, et al. Risk Factors for Long-Term Coronary Artery Calcium Progression in the Multi-Ethnic Study of Atherosclerosis. *J Am Heart Assoc*. 2015; 4:e001726. [PubMed: 26251281]
4. Kaufman JD, Adar SD, Barr RG, et al. Association between air pollution and coronary artery calcification within six metropolitan areas in the USA (the Multi-Ethnic Study of Atherosclerosis and Air Pollution): a longitudinal cohort study. *The Lancet*. 2016; doi: 10.1016/S0140-6736(16)00378-0

TABLE

Mean Bias and Standard Error of the Estimation of the Effect of the Exposure E on Blood Pressure Change (BP), Using Linear Regression Adjusted for measured baseline $BP^*(t_I)$ or Unadjusted for $BP^*(t_I)$ in Simulated Datasets of the 4 causal structures described in Lepage et al. In each scenario, 1050 samples of size 500 were simulated. SE indicates standard error of the estimated effect E on BP . The number of iterations (out of 1050) with erroneous zero coefficients produced are given in parentheses after the mean bias. The values obtained after correcting for the error are presented in bold.

Figures	Result	Regression Model	Mean Bias (# errors)	SE	Mean Bias (# errors)	SE	Mean Bias (# errors)	SE
Figure 1								
Randomized Study	Reported	Adjusted	Subfigure A		Subfigure B			
	Corrected		0.06 (4)	1.42	0.04 (3)	1.35	—	—
			0.02	1.28	0.01	1.24	—	—
	Reported	Unadjusted						
	Corrected		0.06 (5)	1.50	0.05 (4)	1.42	—	—
			0.01	1.34	0.01	1.28	—	—
Figure 2								
Nonrandomized study with unmeasured confounder C	Reported	Adjusted	Subfigure A		Subfigure C		Subfigure D	
	Corrected		1.24 (15)	1.81	0.77 (8)	1.62	1.04 (11)	1.75
			1.11	1.48	0.69	1.41	0.94	1.49
	Reported	Unadjusted						
	Corrected		0.93 (4)	1.56	-0.05 (7)	1.66	0.58 (8)	1.69
			0.91	1.46	-0.11	1.45	0.51	1.48
Figure 3								
Nonrandomized study $BP^*(t_I)$ influence E	Reported	Adjusted	Subfigure A		Subfigure B		Subfigure C	
	Corrected		0.19 (25)	1.99	0.18 (22)	1.95	1.17 (13)	1.65
			-0.05	1.27	-0.02	1.33	1.05	1.33
	Reported	Unadjusted						
	Corrected		-1.64 (2)	1.41	-1.46 (5)	1.57	0.40 (3)	1.48
			-1.66	1.32	-1.52	1.36	0.37	1.39
Figure 4								
Nonrandomized study E influences both $BP(t_I)$ and BP	Reported	Adjusted	Subfigure A		Subfigure B			
	Corrected		-8.75 (44)	4.08	-6.86 (68)	5.76	—	—
			-9.56	1.14	-8.32	1.30	—	—
	Reported	Unadjusted						
	Corrected		0.00 (0)	0.75	0.05 (0)	0.85	—	—
			0.00	0.75	0.05	0.85	—	—