Published in final edited form as:

Paediatr Perinat Epidemiol. 2013 May ; 27(3): 263–265. doi:10.1111/ppe.12049.

Marginal Structural Models, Doubly Robust Estimation, and Bias Analysis in Perinatal and Pediatric Epidemiology

Onyebuchi A. Arah^{a,b}, Madhuri Sudan^a, Jørn Olsen^{a,c}, and Leeka Kheifets^a

^aDepartment of Epidemiology, The Fielding School of Public Health, University of California, Los Angeles, California, USA ^bDepartment of Public Health, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands ^cInstitute of Public Health, University of Aarhus, Aarhus, Denmark

We thank Ahrens and Schisterman (henceforth, A&S) for their commentary¹ on our article.² Although it was not our original intention, we are grateful for the invited discussion on the place of causal inference in perinatal and pediatric epidemiology. In response, we briefly offer some clarifications and extensions.

A&S claim that we did not adjust for reduced hearing at age 18 months (Y_1) in our analysis of the impact of postnatal cellphone exposure (X_2) on hearing loss at age 7 years (Y_2). In fact, we adjusted for Y_1 and other variables listed in the footnotes of Tables 3 and 4 in our article.² In applying directed acyclic graphs (DAGs),^{3,4} A&S rightly caution against grouping variables into one node (e.g., B) in the DAG (Figure below or in our article²). This grouping would imply that all arrows pointing into and out of B apply to every variable in B. We used the grouping to avoid clutter and were always mindful of it. A&S raised the issue of a variable in a grouping also being a collider. This is applicable to every collider on a backdoor path and which is selected for confounding control (e.g., A is simultaneously a confounder and a collider with respect to $X_2 \rightarrow Y_2$ in the DAG below). To eliminate the collider bias introduced by conditioning on A for confounding control or on X_2 (which by being a consequence of the collider A induces conditioning on A),⁴ we need to have measured variable(s) that can be used to close the open bidirected path between A and Y_2 or A and X_2 . This issue leads us to an important but overlooked result that should be part of the existing causal assumptions: there should be no uncontrolled collider bias before or following confounding control. That is, one must control for any collider bias that arises from using a collider on an open backdoor path to close that backdoor or when the exposure under study is caused by a collider that also lies on an open backdoor path.

We agree with A&S on the need for multiple bias modeling.⁵ We expect to see more of it in the literature as probabilistic bias analysis is increasingly accepted by journals, large data become more available, and investigators routinely use bias formulas^{6,7} and simulation techniques. We disagree with A&S that bias analysis must be preceded by “placement of the unmeasured confounder in the DAG”¹ and that such placement can reveal when “the

potential bias is no longer a concern”¹. A known but unmeasured variable should be part of the working DAG from the outset, and not left out until bias analysis. Adding a dashed bidirected arc between the exposure X_2 and outcome Y_2 in our DAG at the bias analysis stage implies the suspicion of, at least, an unmeasured, possibly unknown, common cause of X_2 and outcome Y_2 .

In our article, we triangulated our effect estimate using conventional logistic regression, inverse-probability-weighted (IPW) fitting of marginal structure models (MSM) and doubly robust estimation despite the differences in the qualitative meaning of their effect estimates. This is useful because finding conflicting quantitative results such as reversed effect directions can send warning signals. Importantly, in our article, the different estimates were in the same direction and of similar magnitude. A&S claim that this similarity in magnitude was simply due to minimal confounding in our study. Using hypothetical data with more confounding, we show this claim to be incorrect (see the first three models in the second column of the Table below). A&S then claim that we did not specify how we implemented DRE. Please see the text and Tables 3 and 4 of our article.²

Given journal space limitations, commentaries can sometimes confusingly oversimplify complex issues. First, A&S incorrectly state that inverse-probability-weighting (IPW) can “further be stabilized to increase precision”¹. It is “unstabilized” IPW (e.g., $1/P(X_2 = x_2|X_1 = x_1, Y_1 = y_1, A = a, B = b)$) that yields unduly high precision. “Stabilized” IPW (e.g., $P(X_2 = x_2)/P(X_2 = x_2|X_1 = x_1, Y_1 = y_1, A = a, B = b)$), as used in our article,² reduces excessively high weights assigned to subjects with very low probabilities of exposure.⁸ Second, A&S’s reference to exchangeability as “no residual confounding or selection bias” conflates selection bias and uncontrolled confounding, and makes no explicit reference to conditional exchangeability.^{4,8} Third, calling an exposed subject’s observed outcome a “counterfactual outcome” in defining consistency is problematic. Under consistency, their observed outcome is their *potential outcome* had they been exposed (which they were). This outcome is, thus, not counterfactual. Fourth, A&S rightly state that DRE may be inappropriate in the presence of effect modification. However, not all types of DRE have this limitation. See the Table below for more DRE and other union models. Union models combine different estimators or sub-models into one, can be mono-, doubly or multiply robust with respect of the chances to control for confounding, and have types that are appropriate for effect modification. G-estimation of structural nested models^{8–10} can be doubly or multiply robust and can be used for effect modification. Fifth, we suspect that A&S were referring to g-estimation of structural nested models when they wrote “...structural nested models, with parameters estimated using g-computation”¹. G-computation (g-formula) is one of key g-methods of Robins; the others are IPW fitting of MSM, and g-estimation of structural nested models.^{8,9}

Risking oversimplification, we conclude that causal analysis involves estimating well-defined causal effects using (i) (possibly untestable, qualitative) causal assumptions (e.g., no uncontrolled confounding), and (ii) appropriate statistical estimation techniques (e.g. doubly or multiply robust estimation) to remove existing bias without introducing new bias (e.g. handling time-varying confounding, mediation or effect modification without introducing collider bias).^{3,4,8–10} Even the most sophisticated estimation technique, *sans* causal assumptions, cannot endow an estimate with causal meaning. Conversely, the simplest

conventional regression model coupled with appropriate causal assumptions can be used for causal effect estimation. For details, we defer to our and A&S's references. Clear and defined research questions guided our analysis and presentation of the results of the difficult, yet important, pursuit of the role of environmental exposures in the health of children.

Acknowledgments

This work was supported by the Lundbeck Foundation (Grant Number 195/04), the Danish Medical Research Council (Grant Number SSVF 0646), the National Institutes of Health/National Institute of Environmental Health Sciences (Grant Number R21ES016831), and a Veni career grant from the Netherlands Organization for Scientific Research (NWO) to O. A. A. (Grant Number 916.96.059).

References

1. Ahrens KA, Schisterman EF. A time and place for causal inference methods in perinatal and pediatric epidemiology. *Pediatric and Perinatal Epidemiology*. 2013 IN PRESS.
2. Sudan M, Kheifets L, Arah OA, Olsen J. Cell phone exposures and hearing loss in children in the Danish National Birth Cohort. *Pediatric and Perinatal Epidemiology*. 2013 IN PRESS.
3. Arah OA. The role of causal reasoning in understanding Simpson's paradox, Lord's paradox, and the suppression effect: covariate selection in the analysis of observational studies. *Emerging Themes in Epidemiology*. 2008; 5:5. [PubMed: 18302750]
4. Pearl, J. *Causality: Models, Reasoning and Inference*. 2. Cambridge: Cambridge University Press; 2009.
5. Greenland S, Kheifets L. Leukemia attributable to residential magnetic fields: results from analyses allowing for study biases. *Risk Analysis*. 2006; 26:471–482. [PubMed: 16573634]
6. Arah OA, Chiba Y, Greenland S. Bias formulas for external adjustment and sensitivity analysis of unmeasured confounders. *Annals of Epidemiology*. 2008; 18:637–646. [PubMed: 18652982]
7. VanderWeele TJ, Arah OA. Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, exposures and confounders. *Epidemiology*. 2011; 22(1):42–52. [PubMed: 21052008]
8. Robins, JM.; Hernan, MA. Estimation of the causal effects of time-varying exposures. In: Fitzmaurice, G.; Davidian, M.; Verbeke, G.; Molenberghs, G., editors. *Longitudinal Data Analysis*. Boca Raton: CRC Press; 2009. p. 553-599.
9. Robins, JM. Marginal Structural Models versus Structural Nested Models as Tools for Causal Inference. In: Halloran, ME.; Berry, D., editors. *Statistical Models in Epidemiology: The Environment and Clinical Trials*. New York: Springer-Verlag; 1999.
10. Vansteelandt S, VanderWeele TJ, Tchetgen EJ, Robins JM. Multiply robust inference for statistical interactions. *Journal of the American Statistical Association*. 2008; 103(484):1693–1704. [PubMed: 21603124]

Figure.

Directed acyclic graph modified from Sudan *et al*² to incorporate uncontrolled confounding between X_2 and Y_2 , unmeasured common causes of A and X_2 , and of A and Y_2 , and non-differential independent misclassification of Y_2 .

Dashed bidirected arc or double-headed arrow: presence of unmeasured common cause, thus uncontrolled confounding.

Dashed single-headed arrow: connects unobserved true Y_2 to observed but misclassified Y_2^* .

Conditioning on the collider A closes the backdoor path between X_2 and Y_2 running through A (via X_1 , B or Y_1) but opens up the closed path $X_2 \rightarrow A \leftarrow Y_2$. Conditioning on X_2 which is an “offspring” of A also opens up that biasing path.

Table

Odds ratios (95% confidence limits) for the effect of X_2 on Y_2 obtained from conventional outcome regression, inverse probability weighted fitting of marginal structural models, doubly robust estimation, and union models using hypothetical data generated under the DAG structure (but without the dashed bidirected arcs) in the Figure, and the original DNBC data used in Sudan *et al*'s article²

	Hypothetical study	DNBC study data ²
Model for analyzing the $X_2 \rightarrow Y_2$ total effect	Odds ratio (95% confidence limits)	Odds ratio (95% confidence limits)
Unadjusted for confounding by X_1 , A, B, and Y_1	3.09 [2.85, 3.35]	1.23 [1.07, 1.41]
Correctly specified outcome model (OM), adjusting for X_1 , A, B, and Y_1	1.85 [1.68, 2.04]	1.21 [0.99, 1.46]
Marginal structural model (MSM) with correctly specified IPW, with robust variance estimation	1.83 [1.63, 2.06]	1.23 [1.01, 1.49]
MSM with correctly specified IPW, <i>without</i> robust variance estimation	1.83 [1.69, 1.98]	1.23 [1.02, 1.48]
DRE: correctly specified OM, plus correctly specified IPW of MSM, with robust variance estimation	1.86 [1.65, 2.10]	1.22 [1.00, 1.49]
DRE: correctly specified OM, plus correctly specified IPW of MSM, without robust variance estimation	1.86 [1.72, 2.02]	1.22 [1.01, 1.47]
DRE: correctly specified OM, but misspecified IPW of MSM (omitting Y_1), with robust variance estimation	1.87 [1.68, 2.09]	1.22 [1.00, 1.49]
DRE: misspecified OM (omitting Y_1), but correctly specified IPW of MSM, with robust variance estimation	1.85 [1.64, 2.08]	1.23 [1.01, 1.49]
DRE: correctly specified OM, but conditioned on misspecified propensity score $E(X_2 X_1, A, B)$ (i.e., omitting Y_1)	1.86 [1.69, 2.05]	1.20 [0.99, 1.46]
DRE: misspecified OM (omitting Y_1), but conditioned on correctly specified propensity score $E(X_2 X_1, A, B, Y_1)$	1.86 [1.69, 2.05]	1.21 [1.00, 1.47]
'Mono-robust' union model: misspecified OM (omitting X_1 and B), plus misspecified IPW of MSM (omitting Y_1), with robust variance estimation	1.84 [1.65, 2.05]	1.22 [1.00, 1.49]
'Mono-robust' union model: misspecified OM (omitting A), plus misspecified IPW of MSM (omitting Y_1), with robust variance estimation	1.86 [1.67, 2.07]	1.23 [1.01, 1.50]
'Mono-robust' union model: misspecified OM (omitting X_1 and B), but also conditioned on misspecified propensity score $E(X_2 X_1, A, B)$ (i.e., omitting Y_1)	1.84 [1.67, 2.03]	1.20 [0.99, 1.46]
'Mono-robust' union model: misspecified OM (omitting A and B), but also conditioned on misspecified propensity score $E(X_2 X_1, A, B)$ (i.e., omitting Y_1)	1.85 [1.68, 2.04]	1.21 [1.00, 1.47]
'Mono-robust' union model: misspecified OM (omitting all except Y_1) plus misspecified IPW of MSM (omitting Y_1), also conditioned on misspecified propensity score $E(X_2 X_1, Y_1)$ (i.e., omitting A and B)	1.83 [1.64, 2.04]	1.23 [1.01, 1.50]

Hypothetical data available from authors on request.

DNBC: Danish National Birth Cohort.

OM: Outcome model fit using conventional logistic regression.

MSM: marginal structural model.

IPW: inverse probability weighting (i.e., inverse probability of exposure X_2 weighting)

DRE: doubly robust estimation

'Mono-robust' union model: combines at least two sub-models, each with a covariate control set that is insufficient to control for confounding on their own. The union of the different covariate control sets is sufficient within one model (e.g., in outcome regression model), but each covariate

control set can “pool” with another within a union of at least two sub-models (e.g., outcome regression model plus propensity score covariate adjustment) to allow for consistent effect estimation by eliminating residual confounding.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript