Published in final edited form as:

Paediatr Perinat Epidemiol. 2013 May ; 27(3): 258–262. doi:10.1111/ppe.12048.

A Time and Place for Causal Inference Methods in Perinatal and Paediatric Epidemiology

Katherine A. Ahrens and Enrique F. Schisterman

Epidemiology Branch, Division of Epidemiology, Statistics and Prevention Research, Eunice Kennedy Shriver National Institute of Child Health & Human Development, Bethesda, MD

In this issue of *Paediatric and Perinatal Epidemiology*, Sudan *et al.*¹ describe their analysis of cell phone exposure and hearing loss among children enrolled in the Danish National Birth Cohort. Sudan and colleagues have provided a thoughtful analysis and incorporated several relatively new analytic methods. These include the use of directed acyclic graphs (DAGs),² marginal structural models (MSM),³ and doubly robust estimators (DRE).⁴ In addition, the authors also present results from a sensitivity analyses for unmeasured confounding and outcome misclassification.⁵ While we applaud their efforts, we would like to provide a rationale for the use – and misuse – of these and other methods for causal inference.

To illustrate the causal relationships assumed in their analysis, Sudan and colleagues provided a directed acyclic graph. DAGs (also known as causal diagrams) are useful tools for inferring statistical associations from assumed underlying causal relationships². Previous research, subject-matter knowledge and relationships observed within the empirical data can inform the relationships in a DAG. Confounders can be identified from the DAG as factors along a biasing or confounding pathway. From a faithful DAG, one can discern the smallest set of covariates necessary to control for bias, which is referred to as the minimal sufficient adjustment set.⁶ DAGs can be used to identify colliders too (factors that are also along a pathway from the exposure to the outcome but are caused by two other factors along the pathway⁶). The presence of a collider blocks the pathway without need for covariate adjustment, and in fact, adjustment for a collider can induce bias. Critics of the use of DAGs have argued that causal reality is more complicated than the graph can represent and it is impossible to distinguish between all the different pathways, which renders them of little utility in practice. Fortunately, a freely-available useful tool for graphing DAGs and identifying the minimal sufficient set of confounders for adjustment has been developed, DAGitty (<http://www.dagitty.net>)⁷, and even very complicated scenarios can be evaluated with ease.

Using the DAG proposed by Sudan *et al.*, we investigated two aspects of their analysis that may have led to a bias in their findings. First, although reduced hearing at age 18 months (Y_1) was along its own biasing pathway from X_2 to Y_2 (Fig 1), they did not include this variable as a confounder in their models, which could have led to residual confounding. However, as evaluated in the first part of their analysis, there appeared to be no statistical association between Y_1 and X_2 ; the bias due to residual confounding therefore was likely minimal. Nonetheless, if the association between Y_1 and Y_2 were strong and had the prevalence of Y_1 been high, the bias could have been greater. Second, Sudan *et al.* sensibly

Correspondence: Enrique F. Schisterman, Epidemiology Branch, DESPR, Eunice Kennedy Shriver National Institute of Child Health & Human Development, 6100 Executive Boulevard, 7B03, Bethesda, MD, 20892, schistee@mail.nih.gov.

Conflict of interest:

None declared.

grouped factors that they presumed occurred together causally in the DAG, which reduced its complexity. For example, gestational age, breastfeeding and ear infection up to 18 months were grouped together as factor “B.” Before grouping factors, care should be taken to determine that grouped factors are all affected by the exact same set of causes and, similarly, all factors cause the exact same set of effects. One way to cross-check if such a bias may have been introduced as a result of grouping factors is to separate these variables in the DAG and add in hypothetical biasing paths. For example, if an unmeasured confounder (U_1) affected both the outcome (Y_2) and gestational age (B_2), and another unmeasured confounder (U_2) affected both the exposure (X_2) and gestational age (B_2), then either one of those confounders must then be included in the minimal sufficient adjustment set to control for bias (Fig 2).

Another important application for DAGs is to lay the foundation for performing sensitivity analyses for unmeasured confounding and misclassification. Sudan *et al.* consider the effect of an unmeasured confounder on the unadjusted relationship between X_2 and Y_2 assuming different scenarios based on the prevalence of the unmeasured confounder and the strength of the confounding associations. They speculated that possible unmeasured confounders include the use of headphones or other sound-delivery devices. Adding this unmeasured confounder to the DAG illustrates that this type of confounder, presumed as proximal to both X_2 and Y_2 , would then need to be included in the minimal sufficient adjustment set. Therefore, adequately evaluating its potential for bias would be conditional on adjustment for the other variables in the minimal sufficient set. Probabilistic methods have been developed to allow for just such an adjusted bias analysis.^{5, 8, 9} Further, adjusted sensitivity analyses for unmeasured confounding, selection bias and information bias (i.e. misclassification), given that they often co-occur, should be performed in conjunction.⁵ By identifying the placement of the unmeasured confounder in the DAG, it is also possible that this factor is along a pathway between the exposure and the outcome that has already been blocked (by a collider or adjustment), in which case the potential bias is no longer a concern. For unmeasured confounders with an unknown placement in the DAG, the analyst must first consider possible placements of the factor within the DAG before embarking on a sensitivity analysis. The set of covariates for adjustment will depend on the hypothesized placement(s) of the unknown confounder.

DAGs can also be used to identify situations where the effects of the exposure on the outcome can't be estimated without bias using conventional methods. For example, had Sudan *et al.* been interested in the cumulative or joint effects of prenatal cell phone exposure (X_1) and cell phone use at age 7 (X_2) on hearing loss at age 7 (Y_2), there is no adjustment set of covariates that would estimate an unbiased effect using conventional methods with their presented data. In order to estimate the cumulative or joint effect, associations would first need to be “removed” from the DAG; in particular, the arrows between $B \rightarrow X_2$ and $Y_1 \rightarrow X_2$ would need to be removed. B and Y_1 are confounders that are affected by the prior exposure (X_1) (Fig 3). Including them as confounders in the model would adjust for confounding, but at the same time would block two causal pathways from X_1 to Y_2 and lead to over-adjustment bias.⁶ One way to handle this analytical problem is reweighting the data using the inverse of the probability of exposure (or treatment) weight, which can further be stabilised to increase precision.³ After this reweighting, the distribution of the time-varying confounders within each exposure group in the pseudo-population is the same as that for the total original population. Then the MSM (usually in the form of a conventional regression model) can be run using the reweighted data. This form of the structural model should be considered carefully, as different functional forms of the exposure metric may affect the fit of the model.¹⁰ Further information on implementing stabilised weights and MSM for time-varying confounding, including SAS and STATA code, is available.^{11–14}

Sudan *et al.* utilize MSM, with inverse probability of exposure weighting, in their analysis of cell phone use at age 7 (X_2) and hearing loss at age 7 (Y_2). Because the authors' research question concerned a time-fixed exposure, the use of MSM was not necessary to control for time-varying confounding.¹⁵ Use of MSM in this case was simply an alternative to traditional adjusted regression modeling. However, the effect estimate from MSM is different conceptually from that estimated using adjusted relative risk regression modeling. MSM estimates the marginal effect, which is similar to the effect estimate from a hypothetical randomized controlled trial while an effect estimate from an adjusted relative risk regression model is the effect conditional on a set of factors.^{14, 16} The effect estimate Sudan *et al.* obtained from the MSM was very similar to their crude effect estimate using traditional regression; this was because the bias due to confounding by A, B and X_1 was minimal. The penalty for using a MSM here was a slightly wider confidence interval compared with the crude analysis.

Doubly robust estimators, also employed by Sudan *et al.*, build on the weighting procedure implemented for MSM.⁴ The authors do not specify how DRE was implemented in their analysis, so here we provide a brief overview. In DRE, two models are specified: a treatment model and an outcome model. As long as at least one model is correctly specified and the necessary assumptions are met (see below) then the effect estimate is unbiased. There are many ways to implement DRE. One approach, as outlined by Funk *et al.*,¹⁷ is a three step process: (i) predicted outcomes are estimated for each individual under each exposure condition, (ii) a propensity score is calculated for each individual, and (iii) the data are combined using propensity score augmentation. When presenting methods and results from DRE, it is useful to describe the form of the treatment and outcome models and how covariates were selected for each. Excellent descriptions of DRE using SAS and STATA are available.^{14, 17–19}

Under certain assumptions MSMs and DRE can yield effect estimates with a causal interpretation—that is, approximate the results from a hypothetical randomised controlled trial. These include: exchangeability, positivity, consistency, no model misspecification, and temporality—assumptions that are also necessary for an unbiased effect estimate using conventional methods.^{16, 20} Exchangeability refers to no residual confounding or selection bias. Although this is untestable, sensitivity analyses can help determine if there might be violations of this assumption. Positivity assumes there are exposed and unexposed individuals for every confounder combination in the data—which is testable empirically. Parametric regression methods can easily mask violations of this assumption. Consistency refers to the assumption that a subject's observed outcome reflects her counterfactual outcome given her observed exposure history. This assumption is untestable, but is fundamental to our understanding of the causal relationship. No model misspecification is very broad and refers to all models specified in the analysis, including the weight (or treatment) model and the final structural model.¹⁰ This assumption is untestable, but by specifying different models one can evaluate how robust the findings are to different model specifications. Unlike MSM, DRE relaxes this assumption by requiring that only either the treatment or outcome model be correctly specified.⁴ Correct temporal ordering is a key assumption for causal inference and one that has remained important since the early days of modern epidemiology.²¹ No analytical method can overcome fundamental design limitations from cross-sectional data; if the direction of the relationship between exposure and outcome cannot be determined, a causal relationship cannot be inferred.

In the presence of time-dependent confounding and effect measure modification, MSM and DRE may not be appropriate and several other methods are available to the analyst.³ To illustrate this situation, we return to the DAG proposed by Sudan *et al.*, and assume that within different levels of B the relationship between X_2 and Y_2 is modified. Methods to

handle this analytic scenario include, but are not limited to, structural nested models, with parameters estimated using g-computation,^{22–24} and artificially censoring subjects using inverse probability of exposure weighting.²⁵ Didactic explanations of analyses using g-computation are available.^{14, 26}

Longitudinal data with repeated measurements, such as those encountered by Sudan *et al* using the Danish National Birth Cohort Study, can lead to complex casual research questions. DAGs are a useful tool for identifying covariates necessary for adjustment to control for bias, laying the foundations for performing sensitivity analyses, and determining when conventional regression models may not be appropriate. Causal inference models, such as MSM and DRE, can be appropriate methods for analyses involving cumulative or joint effects of exposures over time. It is important to consider the assumptions necessary for causal inference from these models and situations where these models may be inappropriate. Clear and defined research questions are needed to guide the analysis and determine the time and place for causal inference methods.

Acknowledgments

This work was supported by the Intramural Research Program of the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, National Institutes of Health.

References

1. Sudan M, Kheifets L, Arah O, Olsen J. Cell phone exposures and hearing loss in children in the Danish National Birth Cohort. *Paediatric and Perinatal Epidemiology*. 2013
2. Glymour, M.; Greenland, S. Causal Diagrams. In: Rothman, K.; Greenland, S.; Lash, T., editors. *Modern Epidemiology*. 3. Philadelphia: Wolters Kluwer Lippincott Williams & Wilkins; 2008. p. 183–212.
3. Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000; 11:550–560. [PubMed: 10955408]
4. Bang H, Robins JM. Doubly robust estimation in missing data and causal inference models. *Biometrics*. 2005; 61:962–973. [PubMed: 16401269]
5. Lash, T.; Fox, M.; Fink, A. *Applying Quantitative Bias Analysis to Epidemiologic Data*. Springer; 2009.
6. Schisterman EF, Cole SR, Platt RW. Overadjustment bias and unnecessary adjustment in epidemiologic studies. *Epidemiology*. 2009; 20:488–495. [PubMed: 19525685]
7. Textor J, Hardt J, Knüppel S. DAGitty: a graphical tool for analyzing causal diagrams. *Epidemiology*. 2011; 22:745. [PubMed: 21811114]
8. Ahrens K, Lash TL, Louik C, Mitchell AA, Werler MM. Correcting for exposure misclassification using survival analysis with a time-varying exposure. *Annals of Epidemiology*. 2012; 22:799–806. [PubMed: 23041654]
9. Fox MP, Lash TL, Greenland S. A method to automate probabilistic sensitivity analyses of misclassified binary variables. *International Journal of Epidemiology*. 2005; 34:1370–1376. [PubMed: 16172102]
10. Platt RW, Alan Brookhart M, Cole SR, Westreich D, Schisterman EF. An information criterion for marginal structural models. *Statistics in Medicine*. 2012
11. Cole SR, Hernán MA, Robins JM, Anastos K, Chmiel J, Detels R, et al. Effect of highly active antiretroviral therapy on time to acquired immunodeficiency syndrome or death using marginal structural models. *American Journal of Epidemiology*. 2003; 158:687–694. [PubMed: 14507605]
12. Cole SR, Chu H. Effect of acyclovir on herpetic ocular recurrence using a structural nested model. *Contemp Clin Trials*. 2005; 26:300–310. [PubMed: 15911464]
13. Fewell Z, Hernan M, Wolfe F, Tilling K, Choi H, Sterne J. Controlling for time-dependent confounding using marginal structural models. *The Stata Journal*. 2004; 4:402–420.

14. Hernán, MA.; Robins, JM. Causal Inference. <http://www.hsph.harvard.edu/faculty/miguel-hernan/causal-inference-book/ed>. Not yet published
15. Howards PP, Schisterman EF, Heagerty PJ. Potential confounding by exposure history and prior outcomes: an example from perinatal epidemiology. *Epidemiology*. 2007; 18:544–551. [PubMed: 17879426]
16. Mortimer KM, Neugebauer R, van der Laan M, Tager IB. An application of model-fitting procedures for marginal structural models. *American Journal of Epidemiology*. 2005; 162:382–388. [PubMed: 16014771]
17. Funk MJ, Westreich D, Wiesen C, Stürmer T, Brookhart MA, Davidian M. Doubly robust estimation of causal effects. *American Journal Epidemiology*. 2011; 173:761–767.
18. Funk, M.; Westreich, D.; Davidian, M.; Wiesen, C. Doubly robust estimation of treatment effects. 2010. [cited 2012 December 28]; Available from: <http://www.unc.edu/~mfunk/dr/>
19. Emsley R, Lunt R, Pickles A, Dunn G. Implementing double-robust estimators of causal effects. *The Stata Journal*. 2008; 8:334–353.
20. Cole SR, Hernán MA. Constructing inverse probability weights for marginal structural models. *American Journal of Epidemiology*. 2008; 168:656–664. [PubMed: 18682488]
21. Bradford-Hill A. The environment and disease: association or causation? *Proceedings of the Royal Society of Medicine*. 1965; 58:295–300. [PubMed: 14283879]
22. Cole SR, Hernán MA, Margolick JB, Cohen MH, Robins JM. Marginal structural models for estimating the effect of highly active antiretroviral therapy initiation on CD4 cell count. *American Journal of Epidemiology*. 2005; 162:471–478. [PubMed: 16076835]
23. Joffe MM. Structural nested models, g-estimation, and the healthy worker effect: the promise (mostly unrealized) and the pitfalls. *Epidemiology*. 2012; 23:220–222. [PubMed: 22317805]
24. Robins J. A new approach to causal inference in mortality studies with a sustained exposure period-application to control of the healthy worker survivor effect. *Mathematical Modeling*. 1986; 7:1393–1512.
25. Hernán MA, Lanoy E, Costagliola D, Robins JM. Comparison of dynamic treatment regimes via inverse probability weighting. *Basic & Clinical Pharmacology & Toxicology*. 2006; 98:237–242. [PubMed: 16611197]
26. Snowden JM, Rose S, Mortimer KM. Implementation of G-computation on a simulated data set: demonstration of a causal inference technique. *American Journal of Epidemiology*. 2011; 173:731–738. [PubMed: 21415029]

Figure 1.

Directed acyclic graph from Sudan *et al*¹, modified to identify exposure and outcome variable from their primary analysis with the biasing path $X2 \leftarrow Y1 \rightarrow Y2$ bolded

Minimal sufficient adjustment set for estimating the total effect of X on Y₂: {A, B, X₁, Y₁}

X₁: Prenatal cell phone exposure

X₂: Postnatal cell phone exposure (cell phone exposure at age 7)

Y₁: Reduced hearing at age 18 months

Y₂: Hearing loss at age 7 years

A: Mother's age, socio-occupational status, prenatal smoking, prenatal alcohol use, fever during pregnancy, sex of child

B: Gestational age at birth, breastfeeding, ear infection up to 18 months

C: Inner ear inflammation up to age 7 years

Figure 2.

Directed acyclic graph from Sudan *et al*¹, with factor B separated into B₁, B₂, and B₃ and with hypothetical U₁ and U₂ added

Minimal sufficient adjustment sets for estimating the total effect of X₂ on Y₂: {A, B₁, B₂, B₃, X₁, Y₁, U₁}, or {A, B₁, B₂, B₃, X₁, Y₁, U₂}

B₁: Gestational age at birth

B₂: Breastfeeding

B₃: Ear infection up to 18 months

Figure 3.

Directed acyclic graph from Sudan *et al*¹, with arrows from $B \rightarrow X_2$ and $Y1 \rightarrow X2$ removed and adjustment for A

Minimal sufficient adjustment set for estimating the total effect of X_1 and X_2 on Y_2 : {A}