Published in final edited form as:

Paediatr Perinat Epidemiol. 2009 September ; 23(5): 421–423. doi:10.1111/j.1365-3016.2009.01052.x.

Reply to Commentaries: Biology and methodology – the quest for parsimonious models of a complex reality

Enrique F. Schisterman^a and Brian W. Whitcomb^{a,b}

^aEpidemiology Branch, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institute of Health, Bethesda, MD

^bDivision of Biostatistics and Epidemiology, School of Public Health and Health Sciences, University of Massachusetts, Amherst, MS, USA

We appreciate the comments of Dr Kramer and Dr Joseph regarding our work on causal inference in studies of perinatal mortality and methodological considerations.^{1,2} Many explanations for the paradox have been offered, and many remain plausible. The paradox has been observed with many risk factors including race, smoking and parity; as such, investigators must consider not only a specific scenario, but also what appears to be a general phenomenon. The hypothesis regarding collider-stratification to explain the paradox has only recently been considered and, in our view, our work demonstrates it to be among the most plausible explanations.^{3–5}

We wholeheartedly concur with the sentiments offered by Dr Kramer – all methodological research must retain a connection to biology.² One of the values of using directed acyclic graphs (DAG) paired with statistical analysis is their focus on causal relationships (in contrast with associations) among conceptual factors (in contrast with observed variables).⁶

In this light our methodological work adds biological context to the method of risk factor standardisation of birthweight.^{7–9} We have demonstrated the effects of standardisation of birthweights on risk factor effect estimates, and the potential for collider-stratification to explain the crossing of the curves.^{4,5} In regard to the former, use of z-scores of birthweight has previously been shown to remove the crossing of the curves, but the reason for this was not understood except in an intuitive sense. We have shown that adjustment for z-scores of birthweight yields identical estimates to those without adjustment, rendering the adjustment without clear use.⁴ While this does not advance biological understanding, it does suggest reconsideration of z-score-adjusted findings in a new context. Adjustment for z-scores does not decompose the estimates into direct and indirect effects, but rather estimates the total effect, an estimate already available from the crude analysis. Standardisation may negatively influence precision of estimates and, given that statistical significance is determined by consideration of estimates relative to their precision, the reduced precision can affect inference of individual study results.

Our findings regarding bias that may result from birthweight adjustment serve as a reminder to keep postulated causal relationships at the base of all study design and analysis and to consider the implications of analytical approaches, which sometimes are not easily predicted. Dr. Kramer suggests that adjustment for intermediates (i.e. birthweight,

Correspondence: Enrique F. Schisterman, PhD, Epidemiology Branch, Division of Epidemiology, Statistics and Prevention Research, Eunice Kennedy Shriver National Institute of Child Health and Human Development, 6100 Executive Blvd, 7803, Bethesda, MD20892, USA. schistee@mail.nih.gov.

gestational age at delivery) may be driven by an interest in birthweight and gestational age-specific mortalities for the sake of assessing interventions. Certainly, the efficacy of neonatal treatments will vary by gestational age at delivery, illustrating the heterogeneity in interventions even under a single effect of risk factors like smoking on neonatal mortality. Although we have not directly demonstrated it, the observation that adjustment for birthweight can result in biased estimates should also extend to gestational age, so that adjustment for gestational age may result in collider-stratification bias under certain assumptions. This does not preclude clinical or public health utility of gestational age stratification, but provides guidance to aetiological research into causal relationships between exposures and perinatal mortality. When methodological research identifies potential artifactual sources to explain observations, these must be weighed against alternatives. The importance of methodological research in this area can be described as follows: if the crossover is a methodological artifact then this suggests shifting attention away from finding biological explanation for the cause of the crossover, and instead focusing on causes of neonatal mortality.

Critically, all methods based upon DAGs are sensitive to the assumptions made regarding the causal relationships among the factors under consideration. As Dr Kramer notes, effect estimates derived under the assumption that restricted fetal growth *per se* affects risk of neonatal death may be affected if this assumption is false. In our papers, we have shown that collider-stratification bias affects estimates regardless of the presence of a direct effect between birthweight and mortality. Instead, it is the presence of confounders of the birthweight–mortality relationship that drive the bias. The list of candidate confounders is a long one, so that collider-stratification remains a viable explanation. Nevertheless, the caution regarding the sensitivity of analysis based on DAGs is a key one. The strength of DAGs is in forcing investigators to be explicit in stating the hypothesised causal relationships, whereas the limitation results from the potential that these assumptions may be wrong.^{4-6,10,11}

Another limitation of DAGs is in illustrating effect modification, despite recent advances.¹² While our work has provided a thorough exposition on the possibility of collider-stratification bias in contrast with z-scores and birthweight adjustment, we have not focused explicitly on the connection between this bias and the concept of effect modification.¹ First, we have described the effect of collider-stratification on effect estimates for neonatal mortality risk factors such as smoking, and shown the conditions that result in bias. Second, we have shown how this bias manifests in stratum-specific mortality curves, and thereby may explain the crossing of the curves.

Dr Joseph notes the connection between these stratum-specific curves and the concept of effect modification, where the effects of one factor (e.g. birth-weight) are unequal in different strata of another factor (e.g. smoking). The notion of spurious effect modification illustrates another limitation of stratified mortality curves for causal inference. To be considered effect modification, stratum-specific estimates must not only be different, but also be statistically unequal. A spurious effect modification can thus arise when stratified analysis results in large differences between stratum-specific estimates or when stratified analysis results in small differences in very large strata.

We simulated data corresponding to exaggerated versions of the scenarios we evaluated to assess the connection between collider-stratification bias, the crossing of the curves and the presence of statistically significant effect modification (Table 1). We found that under collider-stratification bias just large enough to result in the crossing of the curves, there was not significant effect modification. Under more extreme scenarios, we consistently observed the connection between colliderstratification bias and spurious effect modification. Thus,

through inappropriate treatment of a collider, causal inference is adversely affected for risk factor effect estimates and differences in effects may be falsely observed. The observations regarding spurious effect modification provide further caution to investigators.

The fact that the birthweight paradox is observed for so many factors suggests structural rather than biological explanations. Collider-stratification bias, as well as other methodological explanations, posit the involvement of additional factors to explain observations regarding the crossing of the curves.¹³ For example, for collider-stratification bias to apply, there must be very high net confounding of the birthweight relation with neonatal mortality. The pairing of biology with statistics is of paramount importance for identification and appropriate modelling of such a multifactorial process. Methodological considerations, such as the use of DAGs to illustrate colliders or application of the fetuses-at-risk paradigm,¹⁴ can help to contextualise and/or clarify complex biological relationships. It is critical for methodologists and substantive epidemiologists to maintain a dialogue so that the link between biology and the methods used to describe it is always clear.

References

1. Joseph KS. Exegesis of effect modification – biological or spurious? *Paediatric and Perinatal Epidemiology*. 2009; 23:418–421.
2. Kramer MS. Biology vs methodology in investigating causal pathways for infant mortality. *Paediatric and Perinatal Epidemiology*. 2009; 23:415–417.
3. Hernandez-Diaz S, Schisterman EF, Hernan MA. The birth weight ‘paradox’ uncovered? *American Journal of Epidemiology*. 2006; 164:1115–1120. [PubMed: 16931543]
4. Schisterman EF, Whitcomb BW, Mumford SL, Platt RW. Z-scores and the birthweight paradox. *Paediatric and Perinatal Epidemiology*. 2009; 23:404–414.
5. Whitcomb BW, Schisterman EF, Perkins NJ, Platt RW. Quantification of collider-stratification bias and the birthweight paradox. *Paediatric and Perinatal Epidemiology*. 2009; 23:394–403. [PubMed: 19689488]
6. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology*. 1999; 10:37–48. [PubMed: 9888278]
7. Wilcox AJ, Russell IT. Birthweight and perinatal mortality: III. Towards a new method of analysis. *International Journal of Epidemiology*. 1986; 15:188–196. [PubMed: 3721681]
8. Wilcox AJ. On the importance – and the unimportance – of birthweight. *International Journal of Epidemiology*. 2001; 30:1233–1241. [PubMed: 11821313]
9. Wilcox AJ. Commentary: on the paradoxes of birthweight. *International Journal of Epidemiology*. 2003; 32:632–633. [PubMed: 12913041]
10. Schisterman EF, Whitcomb BW, Louis GM, Louis TA. Lipid adjustment in the analysis of environmental contaminants and human health risks. *Environmental Health Perspectives*. 2005; 113:853–857. [PubMed: 16002372]
11. 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]
12. Weinberg CR. Can DAGs clarify effect modification? *Epidemiology*. 2007; 18:569–572. [PubMed: 17700243]
13. Basso O, Wilcox AJ, Weinberg CR. Birth weight and mortality: causality or confounding? *American Journal of Epidemiology*. 2006; 164:303–311. [PubMed: 16847040]
14. Joseph KS. Theory of obstetrics: the fetuses-at-risk approach as a causal paradigm. *Journal of Obstetrics and Gynaecology Canada*. 2004; 26:953–960. [PubMed: 15560858]

Table 1

Spurious effect modification in the presence of collider-stratification under varying causal relationships among a risk factor, birthweight, neonatal mortality and an unmeasured confounder of the birthweight-neonatal mortality relationship

Scenario	A ^a	B ^b	C ^c	D ^d	E ^e	% bias to the adjusted smoking effect	Significant effect modification ^f
1	-1000 g	1.6 (OR = 5)	0 (OR = 1)	0	-1000 g	N/A	7%
2	-800 g	3.2 (OR = 24)	0.4 (OR = 1.5)	-0.001	-800 g	395	15%
3	-800 g	3.2 (OR = 24)	0.8 (OR = 2.2)	-0.001	-800 g	127	38%
4	-200 g	3.2 (OR = 24)	0.4 (OR = 1.5)	0	-1000 g	959	89%
5	-200 g	3.2 (OR = 24)	0.8 (OR = 2.2)	0	-1000 g	180	96%

Notes, symbols and abbreviations:

^aA = E[bw|U = 1] - E[bw|U = 0].

^bB = log (p[NM = 1|U = 1]/p[NM = 1|U = 0]).

^cC @D = log (p[NM = 1|RF = 1]/p[NM = 1|RF = 0]) = 0.

^dD = log (p[NM = 1|bw = k + 1]/p[NM = 1|bw = k]) = 0.

^eE = E[bw|RF = 1] - E[bw|RF = 0] = 0.

^fThe percentage of simulations where the *P*-value for the interaction of the risk factor with birthweight was <0.05.