Invited Commentary

Invited Commentary: Composite Outcomes as an Attempt to Escape From Selection Bias and Related Paradoxes

Miguel A. Hernán*, Enrique F. Schisterman, and Sonia Hernández-Díaz

* Correspondence to Dr. Miguel Hernán, Department of Epidemiology, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115 (e-mail: miguel_hernan@post.harvard.edu).

Initially submitted August 29, 2013; accepted for publication September 20, 2013.

This commentary reviews the recent history of explanations to crossover paradoxes such as the birth weight and gestational age paradoxes, with a special emphasis on the current proposal by Kramer et al. in this issue of the Journal (Am J Epidemiol. 2014;179(3):361–367). We contend that the causal structure of these paradoxes is essentially identical to that of several well-known selection biases. We then consider the pros and cons of using composite outcomes to circumvent these selection biases.

causal diagrams; composite outcomes; paradox; selection bias

At first sight, Kramer et al. (1) have written yet another article on the birth weight paradox. However, their paper has broader applicability because it ties together several conceptual threads that underlie much of epidemiologic research, including selection bias, competing risks, and composite outcomes. That Kramer et al. happened to write about perinatal epidemiology was just the result of their research interests.

This commentary has 2 parts. First, we review the recent history of explanations to the birth weight and gestational age paradoxes and argue that the causal structure of these paradoxes is essentially identical to that of several well-known selection biases. Second, we consider the pros and cons of using composite outcomes, as proposed by Kramer et al., to circumvent these selection biases.

THE STRUCTURE OF THE PARADOX

As Kramer et al. (1) remind us, and Wilcox (2) reviewed in detail, the birth weight paradox was originally defined as the apparently surprising intersection of the birth weight–specific mortality curves. Analogously, the gestational age paradox can be defined as the intersecting of the gestational age–specific mortality curves. These 2 paradoxes are distinct phenomena because the birth weight paradox might occur even in a population in which all fetuses were born alive at the same gestational age.

A decade ago, we proposed the following simple explanation for these 2 crossover paradoxes: conditioning on a collider (3). We then explored the birth weight paradox (4, 5) under the simplifying assumption that all infants were born at a fixed gestational age. Wilcox and Basso (6) and Whitcomb et al. (7) described hypothetical, but realistic, settings that are compatible with this explanation of the birth weight paradox. To describe the gestational age paradox (3), we relaxed the simplifying assumption of fixed gestational age at birth and considered the time-varying nature of birth itself (i.e., different fetuses are born at different times). To do so, we proposed a causal diagram like that in Figure 1, which includes time-varying indicators for birth and death status. Figure 1 suggests that the gestational age paradox, like the birth weight paradox, may be viewed as the result of conditioning on a collider. The collider—being born—is time varying; selection bias will occur when the exposure of interest affects (or shares causes with) the timing of birth.

Two features of Figure 1 are worth noting. First, to focus our attention on the gestational age paradox rather than the birth weight paradox, Figure 1 does not include the time-varying weight (sometimes used as a proxy for gestational age, as Kramer et al. point out). Second, Figure 1 represents the causal null setting of no effect of the prenatal exposure on mortality through any pathway. However, it might be argued that birth itself may have a direct effect on mortality (e.g., through the premature birth of an otherwise healthy fetus). If that is indeed the case, then conditioning on being born will not only introduce selection bias, but will also make it impossible to estimate the total effect of the exposure.
Under our explanation, the paradox has a structure similar to other forms of selection bias in epidemiology, such as the built-in bias in interval-specific hazard ratios (8), the apparent bias in estimating the effect of smoking on dementia risk (9), and the more recently described obesity paradox (10, 11). Kramer et al. (1), however, argue that the gestational age paradox may be explained by selection bias that does not arise from conditioning on a collider. They illustrate their argument with a causal diagram, which we now analyze.

First, note that the causal diagram from Kramer et al. (1) can be simplified by combining the nodes SB (stillbirth) and LB (livebirth) into 1. Conditional on birth, these 2 are the same variable with reversed coding. The simplified causal diagram is then Exposure → LB/SB → Outcome.

Second, because causal graphs are only causal if they include all common causes of any 2 variables in the graph, this diagram implies that stillbirth and the outcome (i.e., early neonatal death) share no causes, which some would argue is biologically implausible. If stillbirth and the outcome had common causes, conditioning on livebirth would amount to conditioning on a collider.

Third, Kramer et al. (1) acknowledge that their diagram includes “the hidden aspect of time at risk,” which makes it difficult to understand what variables are really being conditioned on. Figure 1 brings that hidden aspect to the surface, which makes the conditioning on the collider “birth” explicit. Interestingly, Kramer et al. (1) start out their paper by stating that conditioning on birth causes the paradox, which is precisely the implication of Figure 1. A remaining discussion point is whether birth is or is not a collider. We hope that Figure 1 will facilitate a conversation on this issue.

We have previously defined selection bias as the bias arising from conditioning on a collider under the causal null hypothesis of no effect of exposure on the outcome (12). However, when the exposure affects the outcome, selection on noncolliders affected by either treatment or outcome may result in biased estimates of even a nonnull effect (13).

Thus, an explanation of the paradox that relies on conditioning on a noncollider (like the one that Kramer et al. propose) can apply only to settings in which the null does not hold, whereas an explanation that relies on collider conditioning applies to both cases in which the prenatal exposure is and is not expected to have an effect on the outcome (e.g., maternal smoking and altitude, respectively).

ESCAPING THE PARADOX

Figure 1 also shows that the structure of the gestational age paradox has much in common with that of competing risks for the outcome of interest (12). The role played in other settings by “death from other causes” is played here by “not having been born yet.” The common structure results in parallel methodological problems for both phenomena. In the presence of selection bias due to censoring by death, one needs to choose between 1) not adjusting for selection bias (obviously bad); 2) adjusting for selection bias, which is tantamount to attempting to estimate the effect if nobody had ever died from any cause other than the outcome of interest (obviously weird); and 3) changing the question in order to estimate the effect in a subset of individuals, a principal stratum or combination of principal strata, that cannot be identified in practice (not obviously interesting). None of the above analytical strategies is very appealing.

Strategy 2 aims at a causal effect that can be defined only if one is willing to postulate a hypothetical intervention to eliminate the competing events. Strategy 3 restricts the inference to an unidentifiable, and possibly ill-defined, subset of the population. In addition, effect estimation under both strategies relies on strong, unverifiable assumptions. Similar strategies, reviewed by VanderWeele et al. (14), are available to deal with the conditioning on birth status that underlies the gestational age paradox.

There is, however, a fourth strategy, which is to define a composite outcome that encompasses both the outcome of interest and the competing event. This strategy is the default approach to the handling of competing risks in many randomized trials. For example, the primary outcome of many trials is a composite endpoint that includes both cardiovascular disease and death from any cause. Kramer et al. (1) support the use of the composite outcome “perinatal death,” which includes both stillbirth and neonatal (postbirth) death, like many randomized trials do. By engulfing both the outcome of interest and the competing events, composite endpoints effectively eliminate the problem of competing risks, circumvent the selection bias, and allow us to escape the paradox.

Figure 1 naturally incorporates composite outcomes because it does not include separate nodes for stillbirth and neonatal death; death is represented by D, regardless of whether it occurs before or after birth. The difference between stillbirth and neonatal death is determined by the value of B. For example, $B_{210} = 1$ and $D_{210} = 1$ indicates birth on or before day 210 and subsequent neonatal death, whereas $B_{210} = 0$ and $D_{210} = 1$ indicates stillbirth.

But, in general, there may be a hefty price to pay for this strategy. Suppose the investigators’ original intent was to estimate the effect of lifetime smoking on dementia. Using the composite outcome dementia plus death does not necessarily address the investigators’ original question. Changing the outcome changes the question, which may leave us wondering. Composite endpoints can be viewed as weighted utility measures in which different events are assigned the same weight.
This approach may be reasonable for perinatal death because it might be argued that both clinicians and expecting parents are equally interested in stillbirths and neonatal deaths. However, more generally, the use of composite outcomes that conflate events with different etiology and/or significance does not always satisfactorily solve the methodological problems related to competing risks.

For example, suppose we want to estimate the effect of a prenatal exposure on risk of sudden infant death. Miscarriages and stillbirths are competing events. Do we then define a composite endpoint that includes miscarriages, stillbirths, neonatal death, and sudden infant death? Because the number of sudden infant deaths will be overwhelmed by that of miscarriages and stillbirths (but each of these events is assigned the same weight), the resulting composite endpoint bears little resemblance to the outcome in which we are really interested. In a more extreme case, what if one wants to estimate the effect of maternal smoking during pregnancy on the risk of Alzheimer’s disease in the offspring? Would we consider a composite outcome that includes Alzheimer’s disease plus miscarriages, stillbirths, neonatal deaths, infant deaths, and any other deaths during childhood and adult life? Such an outcome, dominated by events other than Alzheimer’s disease, would be unrecognizable for someone interested in the effect of maternal smoking on the risk of Alzheimer’s disease.

In summary, composite endpoints eliminate bias due to competing risks but at the expense of making it hard to interpret the effect estimates. One possible way forward would be the use of composite endpoints that assign different weights to different outcomes.

Finally, Kramer et al. (1) are to be commended, because they frame their discussion of the analysis of observational studies within the randomized trial paradigm, which follows the principle that the start of follow-up and the start of exposure should coincide. Neglecting this simple principle has been the source of much confusion not only in perinatal epidemiology but also in studies that attempted to estimate the effects of therapies (e.g., postmenopausal hormone therapy (15), statins (16)) by using prevalent users.

**ACKNOWLEDGMENTS**

Author affiliations: Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts (Miguel A. Hernán, Sonia Hernández-Díaz); Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts (Miguel A. Hernán); Harvard-MIT Division of Health Sciences and Technology, Boston, Massachusetts (Miguel A. Hernán); and Epidemiology Branch, Eunice Kennedy Shriver National Institute of Child Health and Development, National Institutes of Health, Bethesda, Maryland (Enrique F. Schisterman).

This work was supported by the National Institutes of Health (grants R01 HL080644 and R01 AI102634. E.F.S. was supported by the Intramural Research Program of the Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health.

Conflict of interest: none declared.

**REFERENCES**