Approaches for analyzing the risks of adverse pregnancy outcomes have been the source of much debate and many publications. Much of the problem, in our view, is the conflation of time at risk with gestational age at birth (or birth weight, a proxy for gestational age). We consider the causal questions underlying such analyses with the help of a generic directed acyclic graph. We discuss competing risks and populations at risk in the context of appropriate numerators and denominators, respectively. We summarize 3 different approaches to quantifying risks with respect to gestational age, each of which addresses a distinct etiological or prognostic question (i.e., cumulative risk, prospective risk, or instantaneous risk (hazard)) and suggest the appropriate denominators for each. We show how the gestational age–specific risk of perinatal death (PND) can be decomposed as the product of the gestational age–specific risk of birth and the risk of PND conditional on birth at a given gestational age. Finally, we demonstrate how failure to consider the first of these 2 risks leads to selection bias. This selection bias creates the well-known crossover paradox, thus obviating the need to posit common causes of early birth and PND other than the study exposure.

birth weight; birth weight paradox; early neonatal death; gestational age; perinatal mortality; preterm birth; selection bias; stillbirth

Abbreviation: PND, perinatal death.

Editor's note: An invited commentary on this article appears on page 368, and the authors' response appears on page 371.

More than 30 years ago, Yerushalmy (1, 2) observed that neonatal mortality rates among infants born to smokers were lower than among infants of nonsmokers at birth weights of 2,500 g or less but were higher at higher birth weights. In fact, birth weight or gestational age at birth has been interpreted as modifying the effects of many risk factors for fetal and infant mortality, including plurality, parity, altitude, and race. This phenomenon has been called the “birth weight paradox” and has led to much debate and many publications (3–17).

This paradox has been discussed recently in the context of causal diagrams (directed acyclic graphs) (12–17). Gestational age at birth (or birth weight) has been considered a collider in that context, and thus “adjusting” for it can bias the estimate of association between a study exposure and perinatal mortality, owing to confounding by a common cause of early birth (or low birth weight) and perinatal mortality. In fact, such collider stratification bias can even change the direction of that estimate and produce the well-documented paradox.

In this conceptualization of perinatal mortality, the observed conventional birth weight–specific or gestational age–specific mortality curve crossover is considered a statistical artifact. But the statistical artifact depends on 1 or more unknown common causes of birth weight (or gestational age) and perinatal mortality. Basso and Wilcox (13) postulate 2 rare, unknown factors, both of which increase neonatal mortality; 1 causes a large increase in birth weight, and the other causes a large decrease in birth weight. Although the authors are able to simulate the birth weight paradox, the existence and identification of these factors remain hypothetical.

A simpler, less hypothetical, and hence more plausible, explanation is that conditioning on birth at a given gestational age creates a selection bias by ignoring the risk of birth (both livebirth and stillbirth) at that gestational age. (Perinatal death
(PND) is often analyzed as a function of birth weight rather than gestational age. Because birth weight increases in parallel with advancing gestational age, it is also a proxy for time at risk.) In our view, much of the difficulty in analyzing risk associated with an exposure of interest stems from failure to consider birth itself as a pregnancy outcome. Thus, conditioning on birth when analyzing the risk of other outcomes like PND fails to account for the first of these outcomes. Because early (preterm) birth may itself be caused by the exposure, we will show how conditioning on the first outcome (birth) when examining the association between exposure and the second outcome (e.g., neonatal mortality or morbidity) systematically distorts that association (i.e., causes selection bias).

THE CAUSAL FRAMEWORK

Figure 1 shows a “generic” directed acyclic graph for the relationship between E, an exposure or treatment of interest, and the outcomes of stillbirth (SB) and early neonatal death (END). PND is a composite outcome of stillbirth or early neonatal death. Livebirth (LB) is also shown as a required proximal outcome preceding early neonatal death. Stillbirths and livebirths comprise total births.

All of the 1-way (directed) causal arrows shown in Figure 1 include the hidden aspect of time at risk. During pregnancy, time at risk is measured on the gestational age scale in completed days or weeks. It is strange to consider study exposure as a potential “cause” of livebirth. Rather, it is the early (preterm) timing of the livebirth that is on the causal path to early neonatal death.

Many exposures studied in perinatal epidemiology are individual characteristics, such as maternal age, parity, race, smoking, and prepregnancy body mass index (weight (kg)/height (m)^2); for such exposures, time at risk for adverse pregnancy outcomes can often be assumed to start at conception. Other exposures vary over the course of pregnancy by trimester or even by week; examples include nutrient intake, rate of gestational weight gain, and medication use. Obstetrical interventions such as amniocentesis, stress tests, or tocolysis are short-acting exposures that can have immediate or longer-term effects (causal or preventive) on pregnancy outcomes that occur later in gestation. For these types of interventions, time at risk begins at the time of intervention. It is useful to consider the heuristic of the randomized controlled trial, wherein time at risk begins at the time of randomization, usually close to the time when treatment is initiated.

**APPROPRIATE NUMERATORS AND DENOMINATORS**

In a population of fetuses, placental abruption can be a cause of both preterm stillbirth and preterm livebirth. But for an individual fetus, a placental abruption that causes a stillbirth at 25 weeks gestation cannot also cause a preterm livebirth at 28 weeks or subsequent neonatal death. In fact, all epidemiologic studies of adverse pregnancy outcomes are affected by competing risks for earlier pregnancy loss (e.g., miscarriage, stillbirth, or infant death following livebirth). Importantly, stillbirths are competing risks for adverse outcomes (including infant death) among subsequent livebirths. This is the major rationale for combining stillbirths with early neonatal deaths into the composite outcome of PND and explains why many randomized trials of obstetrical interventions assess perinatal mortality, rather than stillbirth alone or neonatal mortality alone, as their primary outcome. It is also 1 of the reasons etiological studies of preterm birth should include both preterm stillbirths and preterm livebirths (18).

Early neonatal death is, of course, a competing risk for subsequent morbidity in the late neonatal, postneonatal, or childhood periods. A preterm infant who dies on day 1 or 2 does not live long enough to develop, for example, bronchopulmonary dysplasia or cerebral palsy. A severely growth-restricted infant who dies of neonatal infection is no longer at risk for developmental disabilities. These types of competing risks are well known to neonatal clinical trialists. As with the outcome of PND (a composite of stillbirth and early neonatal death) often studied in randomized trials of obstetrical interventions, neonatal trialists use composite primary outcomes such as death or severe developmental disability by age 2 years when comparing randomized treatment groups.

Controversy about the appropriate denominator for estimating risks of adverse pregnancy outcomes has usually revolved around the nature of the outcome, that is, stillbirth (surviving fetuses) versus infant death (livebirths) (19, 20). However, as in other epidemiologic studies, the denominator should include all individuals exposed who were thus at risk of the outcome because of exposure. If the mortality outcome is being studied as a potential consequence of prenatal exposure, then it is exposure of the fetus that may have caused the death rather than postnatal exposure of the liveborn infant. Thus, even if infant death depends (by definition) on a prior livebirth, all fetuses who were alive at the time of exposure are at increased risk of death, whether or not the death occurs before or after the first breath, that is, whether the risk is of 1 outcome (stillbirth) or 2 (livebirth followed by infant death). The use of PND as a composite outcome (stillbirth or early neonatal death) for the numerator and surviving fetuses as the denominator, therefore, solves both the problem of competing risks and that of the population at risk. Indeed, some authors have argued that fetuses-at-risk is the appropriate denominator even when comparing gestational exposures for the risk of mortality in the postneonatal period (e.g., sudden infant death syndrome) (21) or for outcomes diagnosed after infancy, such as cerebral palsy (22). It therefore seems reasonable to restrict the denominator to livebirths only when the relevant exposure occurs after a livebirth, such as a neonatal infection or therapeutic intervention (e.g., mechanical ventilation or postnatal corticosteroid administration). Once
WHAT RISK MEASURES CAN BE USED?

Risk can be measured cumulatively from onset of exposure (or since conception, recognized pregnancy, or 20 weeks’ gestation) until any later period during which outcomes can be caused by the exposure. It can also be measured “prospectively” beginning at any time during pregnancy after exposure onset until the end of the risk period (23) or “instantaneously” (in practice, over the ensuing week) (24). Cumulative incidence and prospective risk are equivalent when exposure and follow-up both begin at a specific time after cohort recruitment.

For a given gestational week $j$, we denote as $O_j$ the number of fetuses or infants in whom the outcome occurs at gestational week $j$. The cumulative incidence from the time of entry (e.g., 20 weeks) to week $j$ is the total number who experience the outcome at or before week $j$ as a proportion of the total recruited cohort, that is, $(\Sigma_{k \leq j} O_k) / N$, where $k$ indicates each gestational week and $N$ is the total number of fetuses in the recruited cohort. The prospective risk beginning at week $j$ is $(\Sigma_{k > j} O_k) / FAR_j$; $FAR_j$ is the number of fetuses at risk (i.e., still alive) at week $j$ and is calculated as $\Sigma_{k \geq j} B_k$, where $B_k$ is the number of births during week $j$ (24). For both cumulative and prospective risk, gestational age represents time at risk. For cumulative risk, it represents time since recruitment (initiation of follow-up); for prospective risk, it represents time remaining until the end of the risk period. The instantaneous risk (the risk during week $j$) of the outcome is simply $O_j / FAR_j$. For instantaneous risk, gestational age is the specific time point at which the hazard is ascertained.

To illustrate the differences among the 3 risk measures for perinatal mortality, we used linked birth–infant death birth cohort data files for 2000–2005 from the National Vital Statistics System of the Centers for Disease Control and Prevention (Atlanta, Georgia) to compare the risk of gestational age–specific perinatal mortality (stillbirth or neonatal death before 7 days) among non-Hispanic white twins versus singletons. These files are compiled by the US National Center for Health Statistics (Hyattsville, Maryland) and include information from birth certificates on maternal sociodemographic characteristics, birth weight, gestational age, plurality, and livebirth order that is linked to information from infant death certificates (25). We used the clinical gestational age estimate, because evidence suggests that it provides rates of preterm birth, postterm birth, and gestational age–specific absolute and relative risks of adverse pregnancy outcomes that are more consistent with those reported in other countries (26, 27). The National Center for Health Statistics data files are known to contain coding errors, missing data, and different gestational age and birth weight thresholds for registration of stillbirths (28, 29). Moreover, for stillbirths, the gestational age at death may precede the gestational age at birth by a considerable period of time. Nonetheless, we believe the data quality of the National Center for Health Statistics files is adequate to explore and illustrate the issues we address.

The overall risks of perinatal mortality were 27.1 per 1,000 twin births versus 6.3 per 1,000 singleton births. At or beyond term (≥37 weeks) gestational ages, the corresponding perinatal mortality risks were 2.7 versus 1.4 per 1,000, respectively. In preterm gestations, however, the perinatal mortality risk was lower (42.9 per 1,000) among twins than among singletons (62.7 per 1,000). As shown in Figure 2, the gestational age–specific instantaneous, prospective, and cumulative incidence perinatal mortality curves do not intersect. No “paradox” occurs with any of the 3 approaches, because all 3 use surviving fetuses as the denominator, thereby (as discussed below) avoiding the selection bias caused by conditioning on birth, which may itself be caused by exposure and is pathological when the birth is a stillbirth or a preterm livebirth.

What information do these different risk measures convey, and when and how should they be used? A comparison of cumulative incidence in exposed versus unexposed groups (Figure 2A) best conveys the overall risk associated with exposure, because it considers all prior gestational ages and, thus, all time at risk, at least for all fetuses who survive to 20 weeks. It is thus the preferred measure for expressing the magnitude of association between exposure and outcome (i.e., for etiological studies). Gestational age–specific cumulative incidence uses the same denominator at all gestational ages, that is, the number of pregnancies (or fetuses) or who entered the cohort at 20 weeks. Numerators continue to accumulate at each gestational age, but the denominator remains constant. The gestational age–specific graph of cumulative incidence can be used to compare the temporal pattern of cumulating risk over time (gestational age) in the 2 groups, starting with exposure (if all events are captured) or cohort recruitment. The widening risk difference in twins versus singletons shown in Figure 2A is a good example.

Prospective risk (Figure 2B) compares the total remaining risk among the exposed versus unexposed groups over the remainder of gestation. For prospective risk, the numerator comprises all future PNDs at each gestational age, and the denominator gradually diminishes over the remaining time of pregnancy as births occur and the pool of fetuses at risk diminishes. This measure is most useful for clinical prediction of outcome in individual pregnant women, that is, for comparing remaining risks in exposed versus unexposed women. For example, Figure 2B shows the small difference in risk of PND among twins versus singletons between 32 and 38 weeks. The temporal pattern of falling or rising prospective risks can also be used for timing clinical interventions on the basis of exposure, such as elective delivery at 37 or 38 weeks in twins versus 41 weeks in singletons.

For instantaneous risk, the denominator decreases weekly as it does for prospective risk, but the numerator includes only those PNDs that occur in the same gestational week. Once again, no crossover is observed. The absolute magnitude (y-axis) of the weekly risk of PND in Figure 2C is much lower than that of the cumulative risk (Figure 2A) or prospective risk (Figure 2B). The y-axis in Figure 2C is shown on a logarithmic scale; on a linear scale, the 2 exposure group risks would be indistinguishable until late in gestation. Figure 2C can be used to visually examine the proportionality assumption for a proportional hazards analysis; a constant...
Figure 2. Gestational age–specific A) cumulative, B) prospective, and C) “instantaneous” (weekly) perinatal mortality risk in twin versus singleton births among US non-Hispanic whites.
difference on the logarithmic y-axis scale indicates a stable ratio of the weekly hazards.

The instantaneous risk is useful for assessing the benefits and risks of obstetrical interventions when these effects are expected within hours or days (e.g., labor induction). The instantaneous risk approach is most commonly used, however, when assessing the temporal pattern of risks associated with longer-acting exposures at specific times (gestational ages) during gestation. For example, Figure 2C shows that the lowest weekly risk of PND for both twins and singletons occurs at weeks 28–33, probably reflecting the balance between the falling risk of early neonatal mortality and the rising risk of stillbirth with advancing gestation.

Although we focus in this paper on comparing risks in 2 or more exposure groups, all 3 of the risk measures discussed above can also be used to summarize risk for an entire cohort, irrespective of exposure.

ONE OUTCOME OR 2? RESOLUTION OF THE CROSSOVER PARADOX

As shown in Figure 3, the crossover paradox occurs only when outcomes (here, PND) are compared for births at each gestational age and are conditional on birth at that gestational age. This type of analysis is tantamount to analyzing the risk of the following 2 outcomes instead of 1: 1) birth at a given gestational age, and 2) PND at the same gestational age. (Because stillbirth is a component of PND, the 2 outcomes become 1.) In the conditional risk, the absence from the denominator of unborn fetuses at each gestational age who are at risk for both outcomes leads to an overestimate of the absolute risk of the second outcome. However, because this occurs in both exposure groups (here, in twins and singletons), why should it matter? In other words, why does it create the crossover shown in Figure 3?

When analyzing the effects of exposure on PND, conditioning on birth at a given gestational age (or birth weight) fails to account for the fact that exposure can influence the gestational age at which birth occurs. Most exposures that increase the risk of PND also increase risk of earlier birth, and analysis of gestational age–specific PND by conditioning on birth at that gestational age is tantamount to ignoring the first (temporally precedent for early neonatal death) of these outcomes.

Birth is an inevitable outcome among pregnant women. It is analogous to death in studies of survival, but it occurs over a much shorter time period. Despite the short-term inevitability of birth, its timing is very important. Although late (post-term) birth (≥42 weeks’ gestation) is also associated with adverse consequences, early birth carries the highest risks of mortality and severe morbidity (30). Early neonatal mortality rises exponentially with decreasing gestational age (30), and approximately 80% of stillbirths occur prior to 37 weeks’ gestation (31). With stillbirth, however, it is the early death of the fetus that causes the birth, rather than the early birth causing the death.

The familiar smoking–lung cancer association provides a heuristic analogy. Suppose we were interested in studying the association between cigarette smoking and death from lung cancer. Analyzing the association by restricting the denominator to new cases of lung cancer would be equivalent to studying smoking as a cause of lung cancer case fatality—not as a cause of fatal lung cancer. Conditioning on lung cancer removes the smoking–lung cancer association and thus biases (reduces, removes, or even reverses) the association with fatal lung cancer.

Figure 4 compares the gestational age–specific birth rates (on a log scale) in twins and singletons; it is defined as the instantaneous (weekly) hazard of birth (stillbirth or livebirth).
Unlike the curves for gestational age–specific PND shown in Figure 2, the birth rate curves are not parallel (i.e., the hazards are clearly not proportional). Toward the end of pregnancy, the birth rates in twins and singletons approach each other, and by the end of pregnancy, they are near 100% for both twin and singleton fetuses who remain in utero.

We now conclude by showing how the crossover paradox arises by failing to consider the first of the 2 outcomes. The gestational age–specific risk of PND ($PND_j$) can be decomposed as the product of the risks of the 2 outcomes, that is, the risk of PND conditional on birth at a given gestational age and the risk of birth at that gestational age. Birth can be analyzed as a pregnancy outcome using the same notation presented earlier, as follows:

$$PND_j \text{FAR}_j = \left( \frac{PND_j}{B_j} \right) \times \left( \frac{B_j}{\text{FAR}_j} \right),$$

where $B_j$ is the number of births occurring during week $j$. Because $\text{FAR}_j = \sum B_k \geq j$,

$$\frac{PND_j}{\sum B_k \geq j} = \left( \frac{PND_j}{B_j} \right) \times \left( \frac{B_j}{\sum B_k \geq j} \right).$$

Analyzing exposure effects on PND conditionally on another outcome (birth) that is caused by exposure leads to selection bias (32). The crossover paradox can be entirely explained by selection bias due to conditioning on birth caused by exposure. The selection bias is increased in magnitude if the association with stillbirths or neonatal deaths is even stronger at earlier gestational ages, as suggested by the nonproportional hazards shown in Figure 4. No unknown common causes of

colliders and outcomes need be hypothesized to explain the paradox.

ACKNOWLEDGMENTS

Author affiliations: Department of Pediatrics, McGill University Faculty of Medicine, Montreal, Quebec, Canada (Michael S. Kramer, Xun Zhang, Robert W. Platt); and Department of Epidemiology, Biostatistics and Occupational Health, McGill University Faculty of Medicine, Montreal, Quebec, Canada (Michael S. Kramer, Robert W. Platt).

This work was supported by a grant from the Canadian Institutes of Health Research.

We thank K.S. Joseph and Sven Cnattingius for their valuable comments and suggestions based on previous versions of this manuscript.

Conflict of interest: none declared.

REFERENCES