

# Environmental hazards, social inequality, and fetal loss: Implications of live-birth bias for estimation of disparities in birth outcomes

Dana E. Goin<sup>a</sup>, Joan A. Casey<sup>b</sup>, Marianthi-Anna Kioumourtoglou<sup>b</sup>, Lara J. Cushing<sup>c</sup>, Rachel Morello-Frosch<sup>d</sup>

**Background:** Restricting to live births can induce bias in studies of pregnancy and developmental outcomes, but whether this live-birth bias results in underestimating disparities is unknown. Bias may arise from collider stratification due to an unmeasured common cause of fetal loss and the outcome of interest, or depletion of susceptibles, where exposure differentially causes fetal loss among those with underlying susceptibility.

**Methods:** We conducted a simulation study to examine the magnitude of live-birth bias in a population parameterized to resemble one year of conceptions in California (N = 625,000). We simulated exposure to a non-time-varying environmental hazard, risk of spontaneous abortion, and time to live birth using 1000 Monte Carlo simulations. Our outcome of interest was preterm birth. We included a social vulnerability factor to represent social disadvantage, and estimated overall risk differences for exposure and preterm birth using linear probability models and stratified by the social vulnerability factor. We calculated how often confidence intervals included the true point estimate (CI coverage probabilities) to illustrate whether effect estimates differed qualitatively from the truth.

**Results:** Depletion of susceptibles resulted in a larger magnitude of bias compared with collider stratification, with larger bias among the socially vulnerable group. Coverage probabilities were not adversely affected by bias due to collider stratification. Depletion of susceptibles reduced coverage, especially among the socially vulnerable (coverage among socially vulnerable = 46%, coverage among nonsocially vulnerable = 91% in the most extreme scenario).

**Conclusions:** In simulations, hazardous environmental exposures induced live-birth bias and the bias was larger for socially vulnerable women.

**Keywords:** preterm birth; Environmental hazard; selection bias; fetal loss; miscarriage; spontaneous abortion; health disparities; perinatal health; live-birth bias

## Introduction

Many environmental exposures that occur in early pregnancy have been linked with increased risk of preterm delivery (live

birth before 37 completed weeks gestation), such as air pollutants, pesticides, persistent organic pollutants, toxic metals, and water contaminants.<sup>1–3</sup> Much of the evidence in support of these relationships utilizes administrative birth records—which only include live births and sometimes stillbirths—to capture outcome information. Because these records exclude fetal loss before 20 completed weeks gestation, they are subject to left truncation.<sup>4</sup> Given that harmful environmental exposures may also increase early fetal loss,<sup>5–8</sup> it is possible that studies which evaluate harm from exposures only among gestations that survive to live birth are biased. This issue has recently received increasing attention in the environmental and perinatal epidemiology and social sciences literature.<sup>9–12</sup>

Live-birth bias is a form of selection bias that can be induced by restricting only to live births when studying in utero exposures. Evidence suggests that up to 30% of conceptions do not result in a live birth<sup>13,14</sup>; therefore, the extent of selection is likely to be substantial. Live-birth bias may be larger or smaller in magnitude depending on the patterns of exposure, outcome, and selection across groups in society. For example, women of lower socioeconomic status and women of color are exposed to higher concentrations of environmental hazards such as air pollution, extreme temperatures, or polluted water, and several studies suggest they may also be more susceptible to those hazards compared with wealthier or white women; these disparities may be due to limited access to quality housing, prenatal care, healthy food, or higher levels of stress.<sup>15,16</sup> Socially or economically marginalized women may also have higher risk for preterm delivery due to other exposures and risk factors that are correlated with socioeconomic status or as a result of structural racism.<sup>17,18</sup> Our inability to capture fetal loss in population-level analyses, therefore, may result in an underestimation of disparities in the estimated effects of

<sup>a</sup>Program on Reproductive Health and the Environment, Department of Obstetrics, Gynecology, and Reproductive Sciences, School of Medicine, University of California, San Francisco, California; <sup>b</sup>Department of Environmental Health Sciences, Columbia University Mailman School of Public Health, New York, New York; <sup>c</sup>Department of Environmental Health Sciences, UCLA Fielding School of Public Health, Los Angeles, California; and <sup>d</sup>Department of Environmental Science, Policy, & Management and School of Public Health, University of California, Berkeley, Berkeley, California

This work was supported by the JPB Environmental Health Fellowship, NIEHS grant P30 ES009089, CARB grant 18RD018, NIEHS grants R00 ES027023 and R01 ES029943, and NIH grant UH3OD023272.

Data Availability: The code used to create this simulation study is available in the eAppendix; <http://links.lww.com/EE/A121>.

**SDC** Supplemental digital content is available through direct URL citations in the HTML and PDF versions of this article ([www.enviroepidem.com](http://www.enviroepidem.com)).

\*Corresponding Author. Address: Dana E. Goin, Department of OBGYN & RS, Box 0132, 480 16th Street, Floor 10, San Francisco, CA 94143. E-mail: [dana.goin@ucsf.edu](mailto:dana.goin@ucsf.edu) (D.E. Goin).

Copyright © 2021 The Authors. Published by Wolters Kluwer Health, Inc. on behalf of The Environmental Epidemiology. All rights reserved. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

Environmental Epidemiology (2021) 5:e131

Received: 21 October 2020; Accepted 29 December 2020

Published online 26 February 2021

DOI: 10.1097/EE9.000000000000131

environmental exposures on adverse birth outcomes between socioeconomic or racial/ethnic groups.

The extent to which excluding fetal loss and restricting to live-birth biases the effect of interest depends on how that effect is defined and on which population researchers would like to make inferences.<sup>10,11,19</sup> Is the goal to estimate the total effect of the exposure on risk of preterm birth or fetal loss among all pregnancies? Or is the goal to estimate the effect only among live-born children? Often the latter is of higher public health importance, because future health outcomes for live-born children are defined, whereas spontaneous abortion is an endpoint in and of itself. It is possible, however, that exclusion of fetal loss as a competing risk or selection mechanism may result in significantly underestimating the harmful effects of social and environmental exposures on pregnancy outcomes.<sup>20</sup> Restricting to live births is less likely to induce bias if the target population is live births (rather than all conceptions), as is often the case for infant, child, or adolescent health outcomes such as autism spectrum disorder or schizophrenia that require a child to be alive, and to live to a certain age, to be diagnosed.<sup>21,22</sup> However, even if the parameter of interest is the effect among live births, there are still two potential sources of bias that can result from left-truncation. These include collider stratification, in which there is a common cause of fetal loss and the adverse pregnancy or child health outcome that is unknown or unadjusted for, or depletion of susceptibles, in which the exposure differentially causes those who have an underlying susceptibility to the outcome to experience a spontaneous abortion instead.<sup>9</sup> This latter example is a special case of collider stratification in which susceptibility, a common cause of selection and the outcome, interacts with the exposure in its effect on fetal loss. In this article, we are focusing on preterm birth as the outcome; however, these concepts also apply to other adverse pregnancy or developmental outcomes.

No real-life data exist on patterns or determinants of early fetal loss that can be generalized to the population at large. Studies that have examined early fetal loss have done so among people using assisted reproductive technology<sup>23</sup> or in small cohorts where people are attempting to become pregnant (which fails to capture early fetal loss among those with unintended pregnancies),<sup>14</sup> or in electronic health records among people seeking prenatal care, which excludes fetal loss that occurs before contact with the medical system.<sup>24</sup> Therefore, previous work in perinatal epidemiology has utilized simulations to illustrate how left truncation can bias estimates of risk for spontaneous abortion,<sup>25</sup> pregnancy complications,<sup>26,27</sup> and delayed child neurodevelopment,<sup>28</sup> as well as quantify bias due to conditioning on preterm birth or birthweight when estimating risk of neonatal death.<sup>29,30</sup>

In this study, we apply a simulation framework to quantify the magnitude of bias due to left truncation and selection in utero on hypothetical estimates of the relationship between exposure to an environmental hazard and risk of preterm birth in a population meant to represent 1 year of births in California. First, we show how accurately defining the parameter of interest and the target population will have important implications for bias calculations. Next, we show how the magnitudes of bias from collider stratification or depletion of susceptibles are shaped by the strength of the relationship between the environmental hazard and risk of spontaneous abortion. Finally, we illustrate that higher burden of exposure to environmental hazards among socially vulnerable population results in larger magnitudes of live-birth bias for these groups.

## Methods

### Simulation background

We conducted a simulation study that examined a range of possible relationships between exposure to an environmental hazard during early pregnancy and risk of spontaneous abortion and estimated the live-birth bias when the environmental

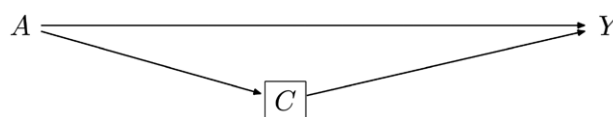
hazard increases risk of both spontaneous abortion and preterm birth.

We illustrate scenarios in which live-birth bias is induced by collider stratification or by depletion of susceptibles. We also included a social vulnerability factor, which represents low socioeconomic status or membership in a marginalized racial or ethnic group. In our simulation, this factor increases the probability of exposure to the environmental hazard and is associated with a higher risk for preterm delivery, independent of the environmental hazard. This is consistent with well-documented disparities in preterm birth along race and class lines and with higher burden of exposure to environmental hazards faced by women of color and lower socioeconomic status.<sup>31–33</sup> For example, Black women had a preterm birth prevalence of 14.1% compared with 9.1% among white women in the United States in 2018,<sup>34</sup> and Black women were exposed to higher average levels of air pollution and were more than twice as likely as white women to live in the most polluted counties during pregnancy.<sup>35</sup> In Michigan, Black children experienced higher lead exposure in-utero compared with white children,<sup>36</sup> and in Massachusetts, other environmental hazards, like power plants, industrial facilities, landfills, toxic waste sites, and incinerators were disproportionately located in neighborhoods with higher percentages of low-income and non-white populations.<sup>37</sup>

### Selection without selection bias

Our first simulation setting illustrates a scenario in which there is selection but no selection bias for a target population of live births (Figure 1). In this scenario, the binary baseline exposure *A* impacts the outcome *Y* (risk of preterm birth) and affects the risk of fetal loss, *C*. When we include only live births in the analysis, we are conditioning on *C*, but since *C* is not a collider, that is, there is no variable besides *A* that affects *C*, and there is no selection bias. To illustrate this scenario, we generated *A* to represent exposure to the environmental hazard using a *Bernoulli* ( $P = 0.5$ ) model among 625,000 individuals. This number approximates 1 year of births in California (about 500,000 births plus an extra 125,000 records [an increase of 25%] to account for spontaneous abortions not included in birth records).<sup>38,39</sup> For simplicity, we assumed the environmental hazard had a constant effect across individuals and that there was no confounding by other factors.

We then assigned each individual an underlying risk of spontaneous abortion, drawn from the following normal distribution:  $r \sim N(\mu = 0.2, \sigma = 0.05)$ . The time to spontaneous abortion in weeks was drawn from the following distribution:  $5 + \text{Weibull}(\lambda = 1.5, \gamma = 0.5) \times 10$ . This parameterization was chosen as it yielded a distribution that approximated the cumulative risk of spontaneous abortion by gestational age observed in California women (eTable 1; <http://links.lww.com/EE/A121>).<sup>24</sup> We then simulated spontaneous abortions, which we refer to as *C*, from a *Bernoulli* ( $P = r + z \times A$ ) model, in which *r* is the underlying risk drawn from the normal distribution described above, *z* is the absolute effect of exposure on risk among those who are exposed, and *A* is the binary indicator of exposure. We chose values of  $z \in [0, 0.2]$  by increments of 0.01. We selected 0.2, or 20 per 100 women, as the upper bound for the excess



**Figure 1.** Directed acyclic graph depicting selection with no live-birth bias for a target population of live births: the exposure of interest *A* is a cause of preterm birth *Y* and fetal loss *C*. Conditioning on *C* does not affect the estimation of causal effect of *A* on *Y* because *C* is not a collider.

risk of spontaneous abortion due to the environmental hazard because it allowed us to estimate the impact of exposure on live-birth bias under an extreme, but still potentially plausible, scenario.

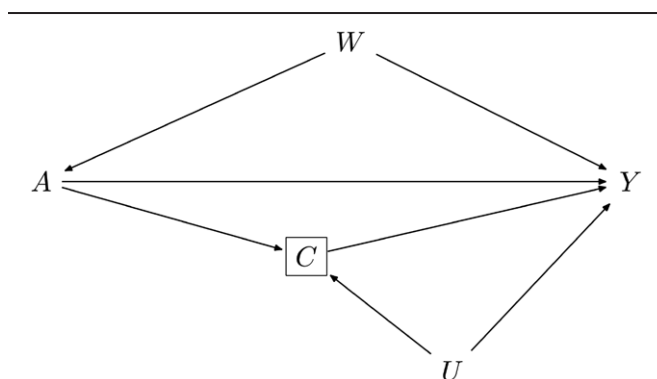
The underlying time to birth was simulated from the following model:  $43 - \text{Weibull}(\lambda = 1.75, \gamma = 1.2) \times 3$ . This parameterization was chosen as it approximated the distribution of live births by gestational age and the prevalence of preterm birth in California (eTables 2 and 3; <http://links.lww.com/EE/A121>).<sup>38,39</sup> Among exposed women, an amount  $s$  was subtracted from the underlying time to birth, to represent the effect of the environmental hazard on gestational length in weeks, where  $s \in [-0.05, 0, 0.05]$ . These values were chosen as they are similar to observed associations between in utero exposure to environmental hazards and gestational age at birth.<sup>40</sup> The outcome was preterm birth, which we call  $Y$ , and was classified as live births with a time to birth of less than 37 weeks completed gestation.

### Live-birth bias due to collider stratification

Next, we created a simulation that induced live-birth bias by including a confounder of fetal loss and preterm birth, which we call  $U$ . When conditioning on live births, the presence of  $U$  creates collider stratification bias, because both  $A$  and  $U$  point into  $C$ , creating a backdoor path between  $A$  and  $Y$  (Figure 2). We also add the social vulnerability factor  $W$ , which was drawn from a Bernoulli ( $P = 0.5$ ) model, so that approximately half the population is assumed to experience this social vulnerability. The unmeasured confounder was drawn from a Bernoulli ( $P = 0.2$ ) model, so approximately 20% of the population was assumed to have this factor that increased risk of both spontaneous abortion and preterm birth. We generated the exposure as a function of the social vulnerability factor, to reflect the fact that political, economic, and discriminatory forces expose lower income women and especially women of color to higher burdens of environmental hazards. The exposure was drawn from a Bernoulli ( $P = 0.1 + 0.25 \times W$ ) model, in which  $W$  represents the social vulnerability factor.

Spontaneous abortions were drawn from the following model: Bernoulli ( $P = r + z \times A + 0.2 \times U$ ), in which similar to above,  $r$  is the underlying risk of spontaneous abortion and  $z$  is the effect of the environmental hazard on fetal loss.

For those without the social vulnerability factor ( $W = 0$ ), the underlying time to birth was simulated from the same model described above:  $43 - \text{Weibull}(\lambda = 1.75, \gamma = 1.2) \times 3$ . For those with the social vulnerability factor ( $W = 1$ ), the underlying time to



**Figure 2.** Directed acyclic graph depicting live birth bias: the exposure of interest  $A$  is a cause of preterm birth  $Y$  and fetal loss  $C$ . Fetal loss and preterm birth also share an unmeasured common cause  $U$ . A social vulnerability factor  $W$  is a common cause of  $A$  and  $Y$ . By conditioning on  $C$ , as we do when using only live births in analyses, we create a backdoor path from the exposure to preterm birth.

birth was simulated from  $42.5 - \text{Weibull}(\lambda = 1.75, \gamma = 1.2) \times 3$ . This has the consequence of increasing the baseline prevalence of preterm birth among those with the social vulnerability factor. The final live birth time was calculated as follows:  $t - s \times A - 0.5 \times U$ , where  $t$  is the underlying live birth time,  $s \in [-0.05, 0, 0.05]$  is the amount that exposure affects time to birth, and those with  $U = 1$  have an additional 0.5 gestational weeks subtracted from their underlying live birth time.

We created a superpopulation of  $N = 5,000,000$  individuals, simulated as described above. We used this superpopulation to calculate the true effect of the environmental hazard on risk of preterm delivery if there was no fetal loss, overall and within strata of the social vulnerability factor. We then randomly sampled 625,000 individuals without replacement from the superpopulation for our analyses. This allowed us to calculate confidence interval (CI) coverage of our estimates. This procedure is described in more detail in the statistical analysis section.

### Live-birth bias due to depletion of susceptibles

Depletion of susceptibles refers to the phenomenon in which it is only those who are both susceptible and exposed who experience an effect of the environmental hazard on fetal loss. Therefore, those who are susceptible to the hazard's effect are differentially selected out of the live-birth cohort. Stated another way, susceptibility acts as an effect modifier of the environmental hazard on the risk of spontaneous abortion. This scenario is the same as the collider stratification simulation except we generated a susceptibility variable that operates similarly to  $U$ , the unmeasured confounder described earlier, which interacts with the exposure to affect fetal loss. This data-generating mechanism looks the same as that represented in Figure 2, because effect modification cannot be clearly represented on a directed acyclic graph.

### Statistical analysis

We used linear regression to estimate absolute differences in risk of preterm birth due to the environmental hazard. In the scenario in which there was no selection bias, models were run on the full set of conceptions (including spontaneous abortions as an adverse outcome in combination with preterm birth), the truncated set (only including live births), and the counterfactual set (the full set of conceptions including all live births had spontaneous abortions not occurred). The full and truncated data sets correspond to having an entire conception cohort versus only those that survived to be included in birth records, respectively. The counterfactual set allows us to compare the effects of exposure on risk of preterm birth had spontaneous abortion not occurred, that is, to include those who would have been at risk for preterm birth had they survived.

For the scenarios in which there is live-birth bias, either due to collider stratification or depletion of susceptibles, we estimated the effects of the environmental hazard using linear regression in the counterfactual and the truncated settings only, since the target population was live births. We controlled for social vulnerability as a confounder in the overall analyses, and also stratified by the social vulnerability factor.

To minimize the influence of sample size variation on the results, we used Monte Carlo simulation with 1000 replications. To assess whether the bias was strong enough to result in a qualitatively different inferential conclusion from the truth, we evaluated the CI coverage across the simulations by assessing whether the 95% CI of the truncated data set in the  $N = 625,000$  population included the point estimate of the counterfactual data set in the  $N = 5,000,000$  superpopulation.

We used R version 3.5.0 for this simulation; annotated code is available in the eAppendix; <http://links.lww.com/EE/A121>.

## Results

In the first simulation setting, the spontaneous abortion prevalence was 25.1%, the preterm birth prevalence was 10.4%, and the environmental hazard exposure prevalence was 50%. This is the scenario when, for a target population of live births, there is selection but no selection bias, because there is no collider stratification or other mechanism that induces bias. The effect estimates in the full and truncated data sets, which correspond to all conceptions and live births, respectively, demonstrate how critical it is to define the target population before determining whether restriction to live births induces bias (Figure 3). As the association between exposure and spontaneous abortion increased in magnitude (from a risk difference of 0 to 0.20, or 0 to 20 per 100 women), the effect of exposure on the risk of a pregnancy not resulting in a term birth (i.e., ending in spontaneous abortion or preterm delivery) among all conceptions increased linearly. Meanwhile, the effect in the truncated data set, which represents just live births, remained unchanged, with a modest risk difference effect size of 0.003, or 0.3 per 100 women. The estimates from the counterfactual and truncated data sets are the same on average, indicating that as long as the target population is live births (rather than all conceptions) restricting to live births in this scenario does not create bias.

In the presence of collider stratification, however, conditioning on live births opens the backdoor path between the environmental hazard and preterm birth, thus inducing bias even if the target population is live births. In the collider stratification scenario, the prevalence of the environmental hazard was 22.5%, the prevalence of spontaneous abortion was 26.3%, and the prevalence of preterm birth was 11.1% on average. As the relationship between the environmental hazard and spontaneous abortion strengthened, the live-birth bias increased (Figure 4). This bias was modest in magnitude, and although it was larger

for those in the socially vulnerable group (Figure 5), the coverage probabilities were not adversely affected (Figure 6).

The bias due to depletion of susceptibles was more pronounced, although the pattern was similar to that observed in the collider-bias scenario (Figure 7). In this scenario, the prevalence of the environmental hazard was 22.5%, the prevalence of spontaneous abortion was 20.5%, and the prevalence of preterm birth was 11.3%. Furthermore, the difference in bias between those with and without social vulnerability was larger (Figure 8). The CI coverage also varied more substantially by social vulnerability status. The coverage of the truncated estimate remained above 90% for those who were not in the socially vulnerable group for all but the most extreme simulation scenario. However, the coverage for those in the socially vulnerable group dropped precipitously, especially once the effect of exposure on spontaneous abortion exceeded a risk difference of 0.1 (Figure 9).

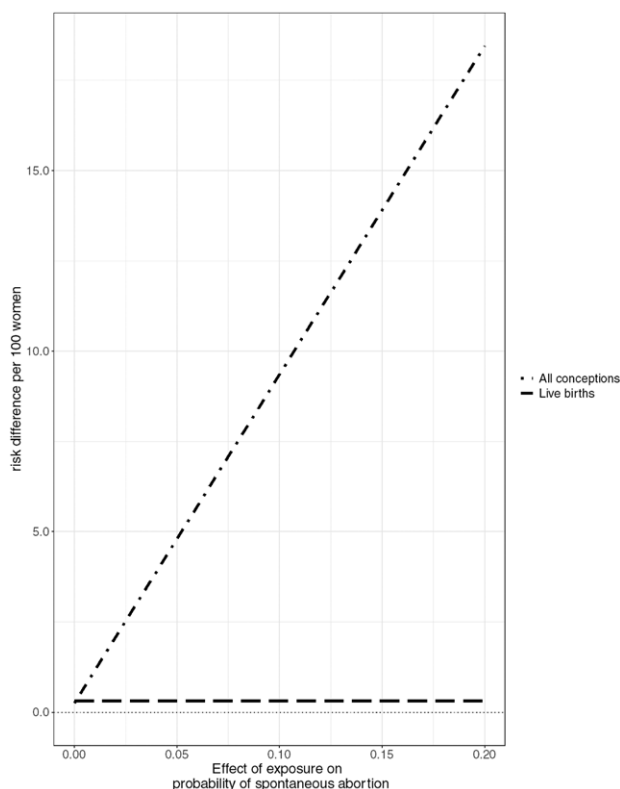
We repeated these analyses for alternate scenarios in which the exposure had no effect or had a protective effect on risk of preterm birth. The patterns were identical to those presented here (eFigures 1–6; <http://links.lww.com/EE/A121>).

## Discussion

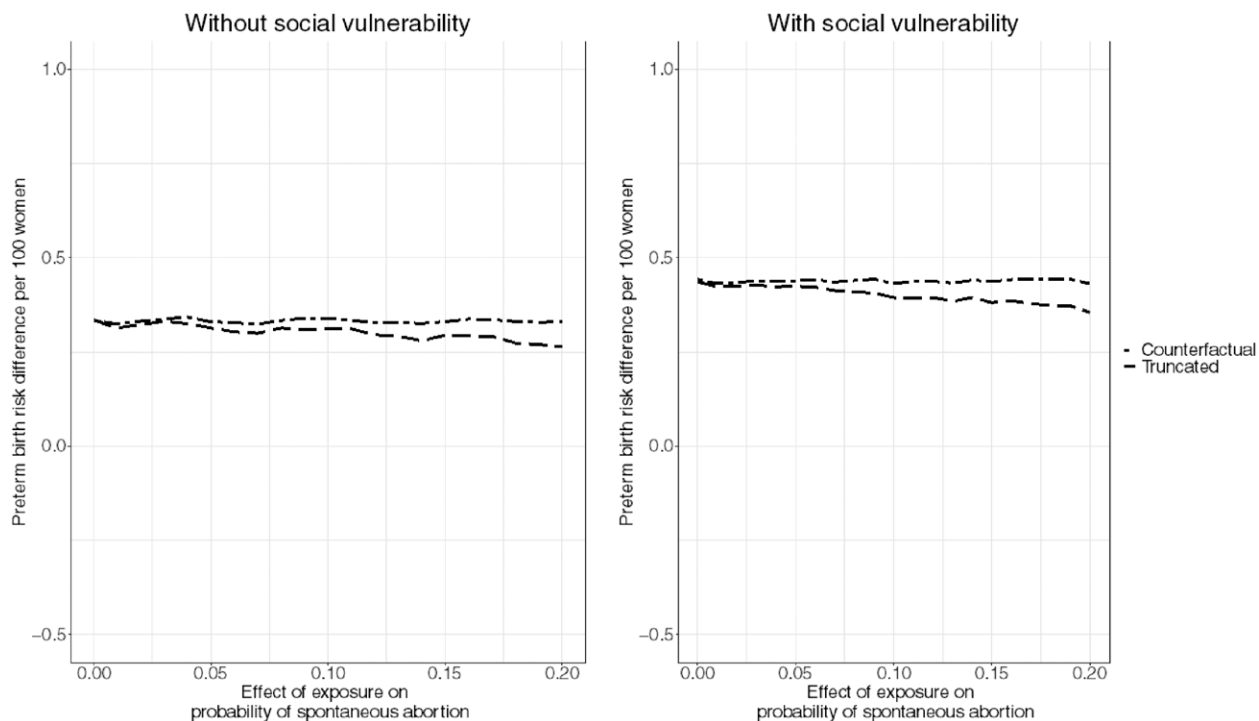
This simulation study demonstrates that live-birth bias can result in underestimation of effects of hazardous environmental exposures on preterm birth, especially among socially vulnerable groups who are more likely to experience the exposure and have greater risk for preterm delivery. Our findings suggest that exclusion of fetal loss from studies of preterm birth likely biases effect estimates downward and, therefore, may hinder identification of harmful exposures regardless of whether the desired level of inference is all conceptions or all live births. Depletion of susceptible fetuses, in particular, resulted in larger live-birth bias, whereas the bias due to collider stratification was smaller in magnitude. Our finding that the extent of downward bias was more pronounced for women with the social vulnerability factor suggests current models may underestimate disparities in maternal and infant health as a result of harmful environmental exposures.

This finding is especially relevant in the context of the high prevalence of preterm birth among Black women in California and the United States.<sup>38</sup> While not much evidence exists regarding determinants of early fetal loss by race in the United States, one study found higher risk of self-reported miscarriage among Black compared with white women in the Southeastern states.<sup>41</sup> Additionally, the male-to-female live birth ratio (i.e., the sex ratio) has been used to assess fetal loss as male fetuses are presumed to be more vulnerable to in utero stressors. For example, evidence shows the sex ratio decreased after the 11 September 2001 terrorist attacks<sup>42</sup> and after the collapse of the East German economy,<sup>43</sup> and increased after the Clean Air Act Amendments of 1970.<sup>44</sup> Analysis of California births from 1960 to 1996 shows that Black women have consistently lower sex ratios compared with white women in California.<sup>45</sup> Therefore, evidence suggests Black women may experience greater early fetal loss, potentially related to their higher burden of exposure to environmental hazards and other forms of structural racism. In combination with our simulation results, this suggests that current estimates of disparities in adverse birth outcomes by race are likely underestimated.

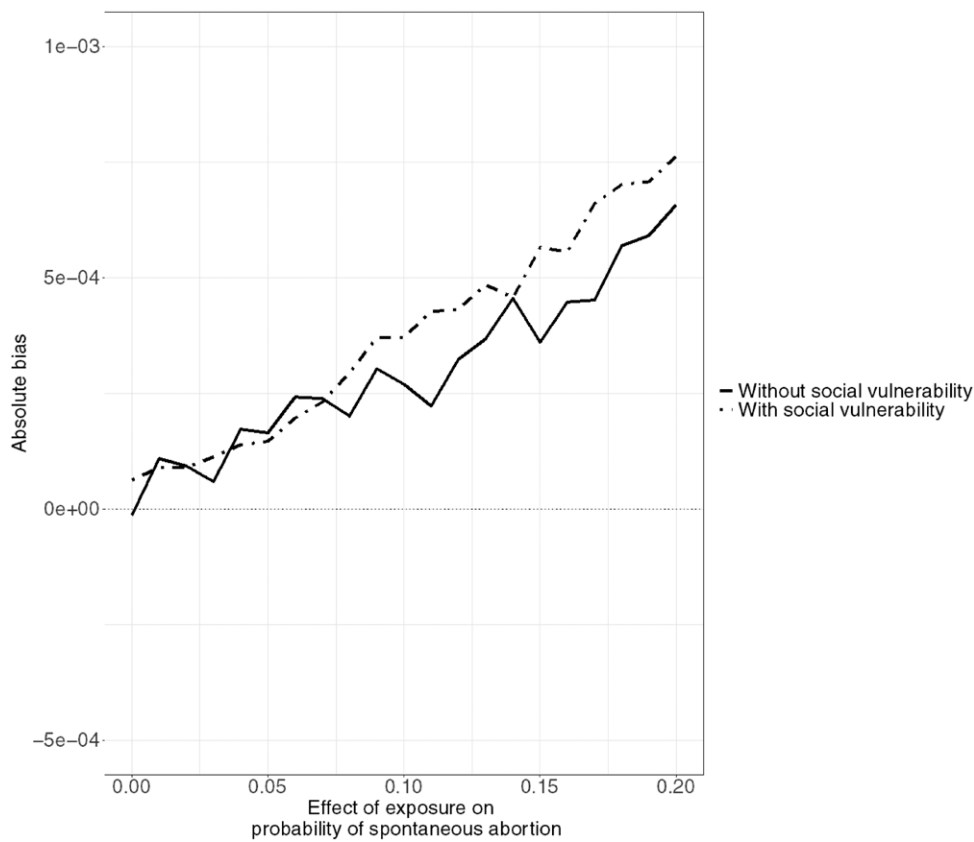
Previous studies have examined magnitudes of collider stratification bias and have found them in general to be small, unless the relationships between  $U$  and either the  $A$  and  $Y$  variables are very strong. For example, one prior study examined the potential magnitudes of collider stratification bias in a general setting,<sup>46</sup> and found having odds ratios of 4, 8, and 16 between  $U$  and  $A$  or  $U$  and  $Y$  would result in biases on the odds ratio scale of 1.14, 1.43, and 1.93, respectively.<sup>46</sup> Additionally, a study that



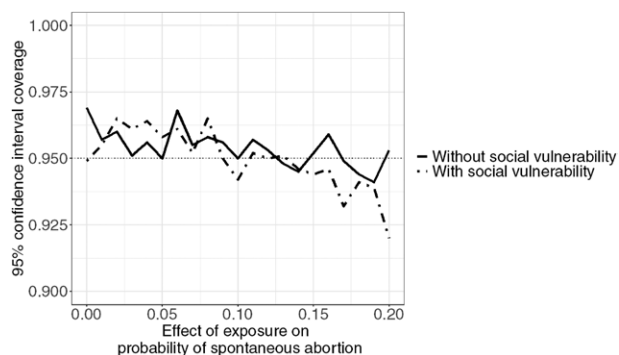
**Figure 3.** Effect of exposure on the risk of a pregnancy not resulting in term birth among all conceptions or all live births when there is no unmeasured common cause of fetal loss and preterm birth.



**Figure 4.** Effect of exposure on the risk of preterm birth among people with and without social vulnerability in the counterfactual and truncated data when there is collider stratification bias. The counterfactual data captures the risk of preterm birth if there were no spontaneous abortion, while the truncated data limits to live births.



**Figure 5.** Bias in effect estimates among people with and without social vulnerability due to collider stratification.

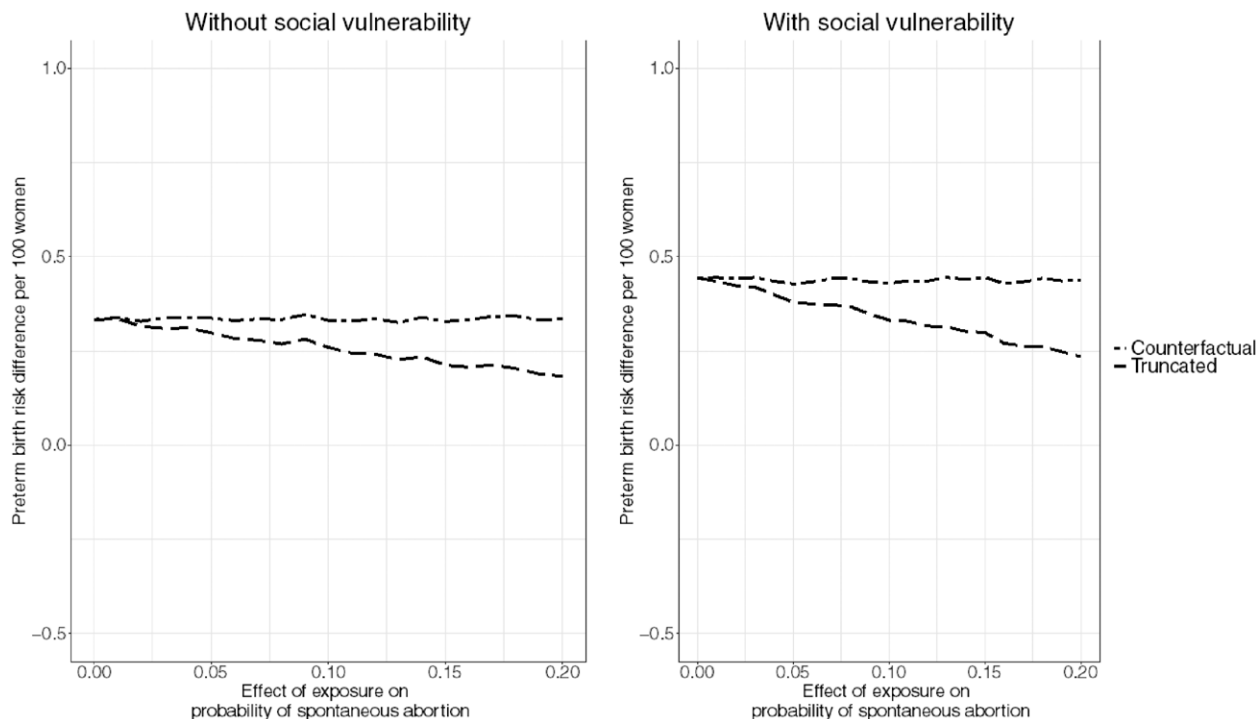


**Figure 6.** Coverage probabilities among people with and without social vulnerability when there is collider stratification bias. The horizontal dotted line is at the 95% coverage level.

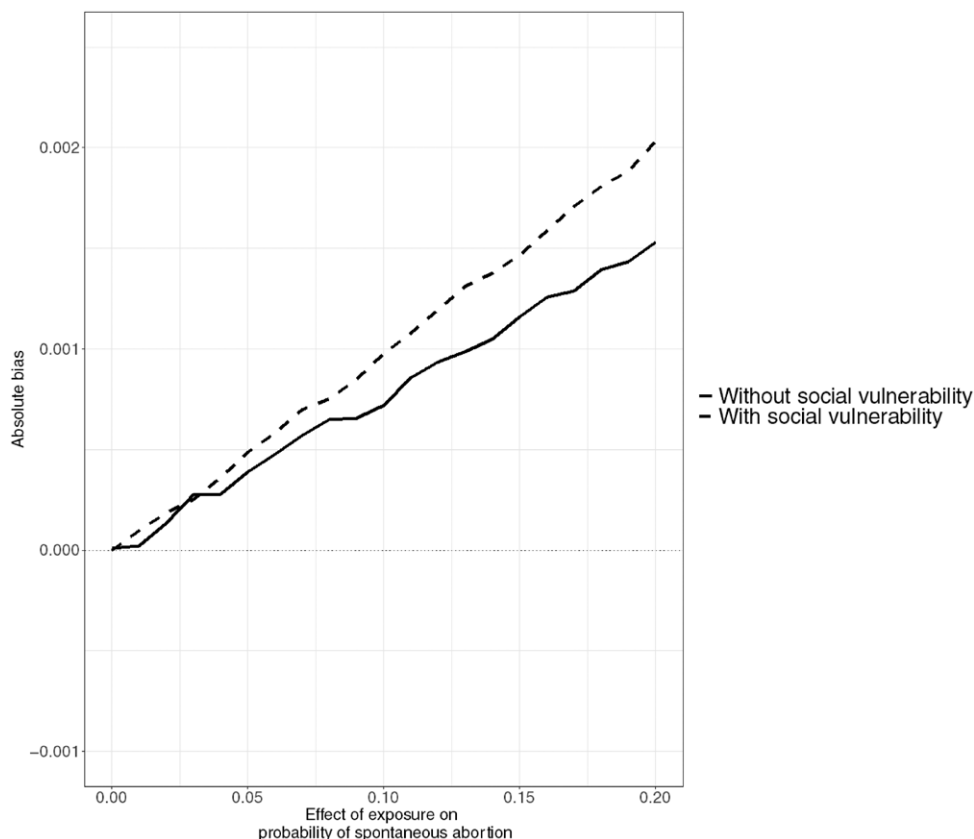
examined collider-stratification bias as a potential explanation for the birthweight paradox found that collider stratification bias was sufficient to explain the paradox when the relationships between exposure, birthweight, and outcome were strong,<sup>30</sup> that is, when *A* reduced birthweight *C* by 400g, *U* reduced birthweight *C* by 500g, and *U* increased risk of neonatal mortality *Y* by 20%. However, others have shown that when interaction exists between variables involved in collider stratification, the biases can be much larger.<sup>47</sup> For example, a simulation study examining collider-stratification bias due to mortality in studies of cognitive decline found that biases were substantially larger when interaction existed between the exposure *A* and unmeasured confounder *U* to affect the censoring variable *C*.<sup>48</sup> Given that depletion of susceptibles is a form of collider stratification with interaction between the susceptibility factor and the exposure, our findings that depletion of susceptibles corresponded to larger estimates of live-birth bias compared with collider stratification are consistent with those found in prior work.

To make our simulation as realistic as possible, we selected parameters based on published estimates whenever possible. However, there is little evidence about drivers of fetal loss or potential magnitudes of effects of environmental exposures on fetal loss before 20 weeks. We attempted to account for a variety of possible scenarios by drawing an underlying risk of spontaneous abortion from a normal distribution, and varying the potential effects from 0 to 0.2, or 0 to 20 per 100 women, on the additive scale. Additionally, we simulated the socially vulnerable group to experience a greater burden of exposure and higher risk for preterm birth but not a larger effect of the exposure on fetal loss, because, to our knowledge, no studies have demonstrated the latter relationship. It is possible, however, that socially vulnerable women experience higher prevalence of other risk factors that have been linked to increased risk of fetal loss, including alcohol use, smoking, or other substance use,<sup>49–51</sup> or experience higher prevalence of complications that might increase risk of miscarriage such as infection and depression.<sup>52,53</sup> We would expect a higher prevalence of fetal loss or a stronger effect of environmental hazards on fetal loss to increase live-birth bias; therefore, our results for the socially vulnerable group may be conservative. We also selected multiple potential effects of exposure on preterm birth among those who did not experience fetal loss, where exposure either (1) reduced risk of preterm birth; (2) had no effect on preterm birth; or (3) increased risk of preterm birth. We found in each of these three cases that selection bias resulting from left-truncation of the data would cause absolute downward bias of the true effect (eFigures 1–6; <http://links.lww.com/EE/A121>). Therefore, researchers might find an effect estimate that is more protective than the truth, protective while no true effect exists, or no effect despite there being a harmful effect for scenarios 1, 2, and 3, respectively.

There is an ongoing debate in the perinatal epidemiology literature regarding the implications of live-birth bias when studying effects of in utero exposures on conditions that are diagnosed only among babies who survive until a certain



**Figure 7.** Effect of exposure on the risk of preterm birth among people with and without social vulnerability in the counterfactual and truncated data when there is depletion of susceptibles. The counterfactual data captures the risk of preterm birth if there were no spontaneous abortion, while the truncated data limits to live births.

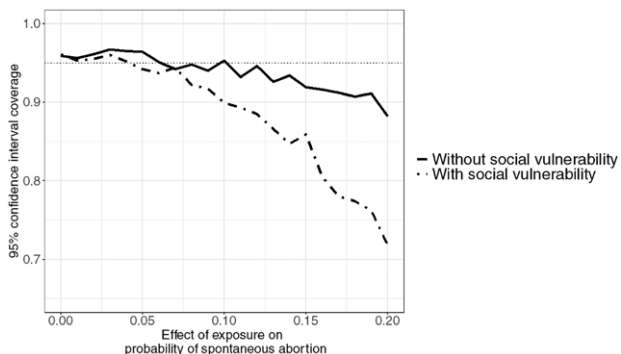


**Figure 8.** Bias in effect estimates among people with and without social vulnerability due to depletion of susceptibles.

age.<sup>10,28,54,55</sup> This debate tends to center on whether it is advisable to account for fetal deaths when the population at-risk of these conditions is children, not all conceptions. In this case, we agree that the relevant target population for estimating risk of childhood diseases is children who survive long enough to be diagnosed. It is possible the data-generating mechanism for such a question resembles Figure 1, in which there is selection but no selection bias for a target population of live births. In this scenario, we agree that restriction to live births, without additional analysis, could be a reasonable approach. However, researchers cannot be certain that there are no unmeasured common causes of fetal loss and the outcome and, as the results of this study show, failing to account for those factors is a threat to internal validity, even when the target population is live births or children who survive to a certain age. Adjusting

for common causes of fetal loss and adverse pregnancy outcomes will not change the target population nor the question of interest—but it will reduce the bias induced by collider stratification and depletion of susceptibles. Therefore, we must improve our understanding of the mechanisms and risk factors for fetal loss. Otherwise, our estimates of hazardous environmental exposures will continue to fall victim to the problem of unknown unknowns—we do not know what factors may be biasing our results or by how much.

Addressing this challenge is especially important for groups that face social or economic marginalization because, as our simulation results show, the magnitude of live-birth bias can be larger for these populations that may already experience a greater burden of exposure and have higher prevalence of preterm birth. Thus, without accounting for live-birth bias, we may underestimate true disparities in pregnancy outcomes across race, class, or other dimensions of social inequity. Therefore, excluding early pregnancy losses from studies of harmful exposures and preterm delivery not only likely biases downwards the true public health costs of these exposures; it likely underestimates the risk attributable to these exposures to an even greater degree for women of color and lower-income women. This may be especially true in countries like the United States, where there are large discrepancies in health by socioeconomic status and race/ethnicity, and may be less relevant in countries with relatively less social inequality. To better understand the magnitudes of these biases in real populations, it would be necessary to improve our efforts to identify or enroll pregnancy cohorts in which early pregnancy loss can be studied.<sup>20,56</sup> Capturing the relationship between hazardous social and environmental exposures and early fetal loss is important to elucidate potential interventions that reduce preventable pregnancy loss and to better characterize risk factors for both fetal loss and preterm birth that, if adjusted for, could reduce live-birth bias and improve



**Figure 9.** Coverage probabilities among people with and without social vulnerability when there is bias due to depletion of susceptibles. The horizontal dotted line is at the 95% coverage level.

our understanding of how to reduce disparities in adverse birth outcomes due to environmental hazards.

## Acknowledgments

The authors declare that they have no conflicts of interest with regard to the content of this report.

## References

- Ferguson KK, O'Neill MS, Meeker JD. Environmental contaminant exposures and preterm birth: a comprehensive review. *J Toxicol Environ Health B Crit Rev*. 2013;16:69–113.
- Bekkar B, Pacheco S, Basu R, DeNicola N. Association of air pollution and heat exposure with preterm birth, low birth weight, and stillbirth in the US: a systematic review. *JAMA Netw Open*. 2020;3:e208243.
- Ferguson KK, Chin HB. Environmental chemicals and preterm birth: biological mechanisms and the state of the science. *Curr Epidemiol Rep*. 2017;4:56–71.
- Schisterman EF, Cole SR, Ye A, Platt RW. Accuracy loss due to selection bias in cohort studies with left truncation. *Paediatr Perinat Epidemiol*. 2013;27:491–502.
- Parazzini F, Chatenoud L, Tozzi L, Benzi G, Dal Pino D, Fedele L. Determinants of risk of spontaneous abortions in the first trimester of pregnancy. *Epidemiology*. 1997;8:681–683.
- Coker AL, Sanderson M, Dong B. Partner violence during pregnancy and risk of adverse pregnancy outcomes. *Paediatr Perinat Epidemiol*. 2004;18:260–269.
- Gaskins AJ, Hart JE, Chavarro JE, et al. Air pollution exposure and risk of spontaneous abortion in the Nurses' Health Study II. *Hum Reprod*. 2019;34:1809–1817.
- Arbuckle TE, Savitz DA, Mery LS, Curtis KM. Exposure to phenoxy herbicides and the risk of spontaneous abortion. *Epidemiology*. 1999;10:752–760.
- Raz R, Kioumourtoglou MA, Weisskopf MG. Live-birth bias and observed associations between air pollution and autism. *Am J Epidemiol*. 2018;187:2292–2296.
- Snowden JM, Bovbjerg ML, Dissanayake M, Basso O. The curse of the perinatal epidemiologist: inferring causation amidst selection. *Curr Epidemiol Rep*. 2018;5:379–387.
- Bruckner TA, Catalano R. Selection in utero and population health: theory and typology of research. *SSM Popul Health*. 2018;5:101–113.
- Nobles J, Hamoudi A. Detecting the effects of early-life exposures: why fecundity matters. *Popul Res Policy Rev*. 2019;38:783–809.
- Macklon NS, Geraedts JP, Fauser BC. Conception to ongoing pregnancy: the 'black box' of early pregnancy loss. *Hum Reprod Update*. 2002;8:333–343.
- Wilcox AJ, Weinberg CR, O'Connor JF, et al. Incidence of early loss of pregnancy. *N Engl J Med*. 1988;319:189–194.
- Morello-Frosch R, Shenassa ED. The environmental "riskycape" and social inequality: implications for explaining maternal and child health disparities. *Environ Health Perspect*. 2006;114:1150–1153.
- Gronlund CJ. Racial and socioeconomic disparities in heat-related health effects and their mechanisms: a review. *Curr Epidemiol Rep*. 2014;1:165–173.
- Blumenshine P, Egerter S, Barclay CJ, Cubbin C, Braveman PA. Socioeconomic disparities in adverse birth outcomes: a systematic review. *Am J Prev Med*. 2010;39:263–272.
- Chambers BD, Baer RJ, McLemore MR, Jelliffe-Pawlowski LL. Using index of concentration at the extremes as indicators of structural racism to evaluate the association with preterm birth and infant mortality—California, 2011–2012. *J Urban Health*. 2019;96:159–170.
- Savitz DA, Hertz-Picciotto I, Poole C, Olshan AF. Epidemiologic measures of the course and outcome of pregnancy. *Epidemiol Rev*. 2002;24:91–101.
- Hogue CJ. Invited commentary: preventable pregnancy loss is a public health problem. *Am J Epidemiol*. 2016;183:709–712.
- Kinlaw AC, Buckley JP, Engel SM, Poole C, Brookhart MA, Keil AP. Left truncation bias to explain the protective effect of smoking on preeclampsia: potential, but how plausible? *Epidemiology*. 2017;28:428–434.
- Basso O. Implications of using a fetuses-at-risk approach when fetuses are not at risk. *Paediatr Perinat Epidemiol*. 2016;30:3–10.
- Hu L, Du J, Lv H, et al. Influencing factors of pregnancy loss and survival probability of clinical pregnancies conceived through assisted reproductive technology. *Reprod Biol Endocrinol*. 2018;16:74.
- Ammon Avalos L, Galindo C, Li DK. A systematic review to calculate background miscarriage rates using life table analysis. *Birth Defects Res A Clin Mol Teratol*. 2012;94:417–423.
- Howards PP, Hertz-Picciotto I, Poole C. Conditions for bias from differential left truncation. *Am J Epidemiol*. 2007;165:444–452.
- Lisonkova S, Joseph KS. Left truncation bias as a potential explanation for the protective effect of smoking on preeclampsia. *Epidemiology*. 2015;26:436–440.
- Suarez EA, Landi SN, Conover MM, Jonsson Funk M. Bias from restricting to live births when estimating effects of prescription drug use on pregnancy complications: a simulation. *Pharmacoepidemiol Drug Saf*. 2018;27:307–314.
- Liew Z, Olsen J, Cui X, Ritz B, Arah OA. Bias from conditioning on live birth in pregnancy cohorts: an illustration based on neurodevelopment in children after prenatal exposure to organic pollutants. *Int J Epidemiol*. 2015;44:345–354.
- Snowden JM, Basso O. Causal inference in studies of preterm babies: a simulation study. *BJOG*. 2018;125:686–692.
- Whitcomb BW, Schisterman EF, Perkins NJ, Platt RW. Quantification of collider-stratification bias and the birthweight paradox. *Paediatr Perinat Epidemiol*. 2009;23:394–402.
- Culhane JF, Goldenberg RL. Racial disparities in preterm birth. *Semin Perinatol*. 2011;35:234–239.
- Burris HH, Collins JW Jr, Wright RO. Racial/ethnic disparities in preterm birth: clues from environmental exposures. *Curr Opin Pediatr*. 2011;23:227–232.
- Miranda ML, Maxson P, Edwards S. Environmental contributions to disparities in pregnancy outcomes. *Epidemiol Rev*. 2009;31:67–83.
- Martin JA, Hamilton BE, Osterman MJ, Driscoll AK. Births: final data for 2018. In: *National Vital Statistics Reports*. Vol. 68. National Center for Health Statistics: Centers for Disease Control and Prevention; 2019:1–47.
- Woodruff TJ, Parker JD, Kyle AD, Schoendorf KC. Disparities in exposure to air pollution during pregnancy. *Environ Health Perspect*. 2003;111:942–946.
- Cassidy-Bushrow AE, Sitarik AR, Havstad S, et al. Burden of higher lead exposure in African-Americans starts in utero and persists into childhood. *Environ Int*. 2017;108:221–227.
- Faber DR, Krieg EJ. Unequal exposure to ecological hazards: environmental injustices in the Commonwealth of Massachusetts. *Environ Health Perspect*. 2002;110(suppl 2):277–288.
- Ratnasiri AWG, Parry SS, Arief VN, et al. Temporal trends, patterns, and predictors of preterm birth in California from 2007 to 2016, based on the obstetric estimate of gestational age. *Matern Health Neonatol Perinatol*. 2018;4:25.
- Barradas DT, Dietz PM, Pearl M, England LJ, Callaghan WM, Kharrazi M. Validation of obstetric estimate using early ultrasound: 2007 California birth certificates. *Paediatr Perinat Epidemiol*. 2014;28:3–10.
- Rauch SA, Braun JM, Barr DB, et al. Associations of prenatal exposure to organophosphate pesticide metabolites with gestational age and birth weight. *Environ Health Perspect*. 2012;120:1055–1060.
- Mukherjee S, Velez Edwards DR, Baird DD, Savitz DA, Hartmann KE. Risk of miscarriage among black women and white women in a U.S. Prospective Cohort Study. *Am J Epidemiol*. 2013;177:1271–1278.
- Catalano R, Bruckner T, Marks AR, Eskenazi B. Exogenous shocks to the human sex ratio: the case of September 11, 2001 in New York City. *Hum Reprod*. 2006;21:3127–3131.
- Catalano RA. Sex ratios in the two Germanies: a test of the economic stress hypothesis. *Hum Reprod*. 2003;18:1972–1975.
- Sanders NJ, Stoecker C. Where have all the young men gone? Using sex ratios to measure fetal death rates. *J Health Econ*. 2015;41:30–45.
- Smith D, Von Behren J. Trends in the sex ratio of California births, 1960–1996. *J Epidemiol Community Health*. 2005;59:1047–1053.
- Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. *Epidemiology*. 2003;14:300–306.
- Whitcomb BW, McArdle PF. Collider-stratification bias due to censoring in prospective cohort studies. *Epidemiology*. 2016;27:e4–e5.
- Mayeda ER, Tchetgen Tchetgen EJ, Power MC, et al. A simulation platform for quantifying survival bias: an application to research on determinants of cognitive decline. *Am J Epidemiol*. 2016;184:378–387.
- Harlap S, Shiono PH. Alcohol, smoking, and incidence of spontaneous abortions in the first and second trimester. *Lancet*. 1980;2:173–176.
- Ness RB, Grisso JA, Hirschinger N, et al. Cocaine and tobacco use and the risk of spontaneous abortion. *N Engl J Med*. 1999;340:333–339.
- Henriksen TB, Hjollund NH, Jensen TK, et al. Alcohol consumption at the time of conception and spontaneous abortion. *Am J Epidemiol*. 2004;160:661–667.



52. Sugiura-Ogasawara M, Furukawa TA, Nakano Y, Hori S, Aoki K, Kitamura T. Depression as a potential causal factor in subsequent miscarriage in recurrent spontaneous aborters. *Hum Reprod.* 2002;17:2580–2584.
53. Giakoumelou S, Wheelhouse N, Cuschieri K, Entrican G, Howie SE, Horne AW. The role of infection in miscarriage. *Hum Reprod Update.* 2016;22:116–133.
54. Werler MM, Parker SE. Bias from conditioning on live-births in pregnancy cohorts: an illustration based on neurodevelopment in children after prenatal exposure to organic pollutants (Liew *et al.* 2015). *Int J Epidemiol.* 2015;44:1079–1080.
55. Liew Z, Olsen J, Cui X, Ritz B, Arah OA. Response to Werler and Parker letter: comment on live-birth bias in pregnancy cohorts. *Int J Epidemiol.* 2015;44:1080–1081.
56. Buck GM, Lynch CD, Stanford JB, et al. Prospective pregnancy study designs for assessing reproductive and developmental toxicants. *Environ Health Perspect.* 2004;112:79–86.