Fetal Health and the Environment

HARVARD UNIVERSITY

Graduate School of Arts and Sciences

DISSERTATION ACCEPTANCE CERTIFICATE

The undersigned, appointed by the
Committee on Higher Degrees in Population Health Sciences,
have examined a dissertation entitled

"Fetal Health and the Environment"

presented by

MICHAEL LEUNG

candidate for the degree of Doctor of Philosophy
and hereby certify that it is worthy of acceptance.

Marc Weisskopf

Dr. Marc Weisskopf, Ph.D., Sc.D., Committee Chair, Harvard T.H. Chan School of Public Health

Brent A Coull

Dr. Brent Coull, Ph.D., Harvard T.H. Chan School of Public Health

Francine Laden

Dr. Francine Laden, Sc.D., Harvard T.H. Chan School of Public Health

Stefania Papatheodorou
Stefania Papatheodorou (Sep 8, 2021 11:51 EDT)

Dr. Stefania I. Papatheodorou, M.D., Ph.D., Harvard T.H. Chan School of Public Health

Date: 16 August 2021

Fetal Health and the Environment

Michael Leung

A dissertation presented to

The Department of Population Health Sciences

and

The Department of Epidemiology

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Population Health Sciences (Epidemiology)

Harvard University

Cambridge, Massachusetts

August 2021

Dissertation Advisor: Marc G. Weisskopf Michael Leung

Fetal Health and the Environment

Abstract

Whether fetuses are vulnerable to the effects of environmental exposures is a difficult question to answer due to the internal, and therefore, hidden nature of conceptions and early embryonic development. Studies of prenatal environmental exposures often assess fetal health using newborn anthropometry, such as birth weight, as these data are easily accessible from medical records and birth certificates. Although newborn size is a key development indicator for perinatal morbidity and mortality under the Barker hypothesis (also known as the Developmental Origins of Health and Disease hypothesis), it is measured at the end of pregnancy, and thus provides limited insights on the timing of when the growth-restricting effects manifest. That is, birth weight is insensitive to early to mid-pregnancy effects, as a fetus that experiences early fetal growth restriction can still catch up to achieve population growth standards by birth. Furthermore, birth weight is mainly driven by fat accumulation throughout pregnancy, and so may be a poor proxy for the development of other organs relevant for health (e.g., it may not reflect a reduction in head size, which has implications for later brain development). Finally, assessing newborn health necessitates the production of live births. Most epidemiologic analyses have focused on birth outcomes, and so ignore the effects on pregnancy loss, which not only is a relevant health outcome, but also a potential source of bias, as it is a competing event that prevents the birth outcome of interest from occurring. Thus, this dissertation comprises of three studies that aim to examine how environmental exposures affect *in utero* fetal developmental processes during pregnancy, rather than using proxies at birth.

In Chapter 1, we assessed through simulations the potential bias induced by restricting epidemiologic analyses to live births when pregnancy loss is influenced by the environmental exposure of interest *A* and any unmeasured factors *U* that also affect the child outcome. Few simulation studies have explored this topic, but they presuppose that there is no interaction between the exposure *A* and unmeasured factor *U*, which may not be a realistic assumption. In this chapter, we relax these assumptions and consider three fetal survival (or selection) mechanisms: 1) collider-stratification without interaction, where *A* and *U* independently affect selection, 2) depletion of susceptibles, where selection is dependent on the presence of both *A* and *U* (i.e., the interaction-only effect), and 3) when both mechanisms operate simultaneously. We show that ignoring pregnancy loss when estimating the effects of prenatal exposures on outcomes in live-born children lead to associations that are biased downwards, where the magnitude of the bias is determined by the selection mechanism, strength of selection, and prevalence of *U*. In Chapters 2 and 3, we examine the association between gestational exposure to $PM_{2.5}$ and ambient temperature, respectively, and fetal growth outcomes in a pregnancy cohort using spatiotemporally resolved data on exposures in Massachusetts, USA. Unlike prior studies that have mostly examined fetal growth using newborn size, we leveraged data from routine ultrasound measurements which allowed us to observe the developmental processes of distinct organs during pregnancy. These last two chapters demonstrate that gestational exposure to $PM_{2.5}$ and ambient temperature were associated with impaired fetal growth, where early to mid-pregnancy appears to be a critical window of exposure. Overall, the findings from this collection show the importance of studying *in utero* fetal health during pregnancy, rather than at its conclusion.

Table of Contents

List of Figures and Tables

Acknowledgements

The work presented here is foremost indebted to the study participants at Beth Israel Deaconess Medical Center, without whom this research would not have been possible. I also would like to express deep gratitude to my advisor, Dr. Marc G. Weisskopf, who was incredibly instructive in all my work. His flexibility with his busy schedule, and availability to always give constructive feedback were particularly appreciated. The members of my dissertation committee – Dr. Stefania Papatheodorou, Dr. Francine Laden, and Dr. Brent Coull – were also instrumental in helping me improve and complete this dissertation. I would like to thank my other co-authors, including Dr. Michele R. Hacker and Dr. Anna M. Modest for helping me navigate through the Beth Israel medical records and ultrasound data; Dr. Joel Schwartz and Dr. Yaguang Wei for their help with the large exposure datasets generated from their state-of-the-art spatiotemporal models; and Dr. Marianthi-Anna Kioumourtzoglou, Dr. Raanan Raz, and Dr. Blair J. Wylie for their substantive expertise.

I would like to thank Eric DiGiovanni, Ellen Furxhi, Warisha Amin, and Caroline Huntington from the Department of Epidemiology for their support and friendship throughout this whole process. Epi Tea Time with you all will always be a fond memory. Many thanks also to Bruce Villineau and Matthew Boccuzzi from the Department of Population Health Sciences for their administrative support despite the challenges we faced as a new PhD program at Harvard. I would also like to thank Dr. Miguel Hernan, Dr. Sonia Hernandez-Diaz, Dr. Murray Mittleman, and Dr. Barbra Dickerman, all of whom I had the pleasure of working with and learning from when helping them teach their epidemiologic methods and causal inference courses. I would also like to extend special thanks to Hanseul Kim, Alejandro Szmulewicz, Sarah Kalia, Lidia Moura,

and Nazleen Khan for their help in preparing for our written exam. They were a constant source of inspiration and motivation, and I could not have done it without them. Last but not least, I would like to thank my family for their unconditional support.

Chapter 1: Bias due to selection on live births in studies of environmental exposures during pregnancy: A simulation study

Michael Leung[1], Marianthi-Anna Kioumourtzoglou[2], Raanan Raz[3], Marc Weisskopf[1,4]

[1] Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts

[2] Department of Environmental Health Sciences, Columbia University Mailman School of Public Health, New York City, New York

[3] Braun School of Public Health and Community Medicine, the Hebrew University of Jerusalem - Hadassah, Jerusalem, Israel

[4] Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, Massachusetts

ABSTRACT

Background: Studies of the effects of prenatal environmental exposures on postnatal outcomes are particularly vulnerable to live birth bias; that is, the bias that arises from the necessary restriction of the analysis to live births when that is influenced by both the exposure under study A and unmeasured factors U that also affect the outcome.

Objectives: In the context of a recent publication of nitrogen dioxide (NO_2) and autism spectrum disorder (ASD) that found an odds ratio (OR) of 0.77 per 5.85 ppb NO_2 during pregnancy, we aimed to examine what parameters would be needed to account for this protective association through live birth bias.

Methods: We simulated the magnitude of bias under two selection mechanisms and when both mechanisms co-occur, assuming a true null effect. Simulation input parameters were based on characteristics of the original study and a range of plausible values for the prevalence of unmeasured factor U and the ORs for the selection effects (i.e., the effects of NO_2 and U on loss, and U on ASD). Each scenario was simulated 1000 times.

Results: We found that the magnitude of bias was small when NO_2 and U independently influenced pregnancy loss (collider-stratification without interaction), was stronger when NO_2-induced loss preferentially occurred in U=1 (depletion of susceptibles), and strongest when both mechanisms worked together. For example, ORs of 3.0 for NO_2-loss, U-loss, U-ASD and U prevalence=0.75 yielded NO_2-ASD ORs per 5.85 ppb NO_2 of 0.95, 0.89, and 0.75 for the three scenarios, respectively. The bias is amplified with multiple Us, yielding ORs as low as 0.51.

Discussion: Our simulations illustrate that live birth bias may lead to exposure-outcome associations that are biased downwards, where the extent of the bias depends on the fetal selection mechanism, the strength of that selection, and the prevalence of U.

INTRODUCTION

Understanding the health effects of environmental exposures is critical for identifying and developing preventive interventions for high-risk populations. However, these effects may not be identifiable if exposure prevents selection into the study and thus, observation of the outcome of interest. This process is a form of 'left truncation' and can lead to estimates of exposure-outcome associations that are biased[1,2]. Epidemiologic studies of environmental exposures are particularly vulnerable to left truncation as exposures are typically ubiquitous in time (e.g., participants are already exposed prior to study initiation), such that exposure-induced selection processes likely inform the formation of observational cohorts from which exposure-health effects are estimated.

In studies investigating the effects of prenatal exposures on outcomes in live-born children, left-truncation can induce a specific type of selection bias called live-birth bias[3,4]. These studies are typically based on cohorts formed by only live births, where selective survival between conception and birth can skew the distribution of prenatal exposures in the subset available for analysis (i.e., those conceptions that resulted in a live birth) from the exposure distribution among all conceptions, such that the estimated parameter in the analyzed subset differs from the parameter in the total population (i.e., all conceptions).

An example of possible live-birth bias is a recent analysis of traffic-related nitrogen dioxide (NO_2) and autism spectrum disorder (ASD), where the odds ratio was 0.77 per 5.85 parts per billion (ppb) increase in NO_2 during pregnancy when mutually adjusted for postnatal exposure to NO_2[5]; that is, prenatal exposure to NO_2 appeared to be protective against ASD. This paradoxical finding is unlikely to be causal as we are not aware of a possible biological mechanism for which

NO_2 may confer beneficial effects on the risk of ASD, or for human health in general for that matter. It is more likely that this strong protective association could be attributed to live-birth bias.

It has been suggested that there are two selection mechanisms that can lead to live-birth bias[3,4]. Although parameterized differently, both mechanisms can be envisioned as forms of collider-stratification bias[6], and can be represented by the directed acyclic graph (DAG) in Figure 1 – which is the same structure as the birth weight paradox[7]. We will refer to these here as "collider-stratification without interaction" and "depletion of susceptibles". In "collider-stratification without interaction", exposure A and some unmeasured factor U – for example, exposure to endocrine disrupting chemicals which have been associated with pregnancy loss and autism[8–11] - are independent and each affect selection (S). In "depletion of susceptibles", which is separate but related to the first mechanism, exposure A and unmeasured variable U do not have independent causal effects on fetal loss, but rather loss is dependent on the joint effects of A and U. A potential example of this mechanism is a gene-environment interaction whereby, exposure-induced loss preferentially occurs in those who have the genetic factor U (i.e., the subset of fetuses susceptible to ASD). Lastly, both mechanisms described above can also work in tandem as they operate through distinct mechanistic pathways, in that, A and U not only causally interact to affect fetal loss, but also, have independent causal effects on fetal loss. It is important to note that although the three mechanisms described above are parameterized differently, they are indistinguishable on a DAG since DAGs are nonparametric and thus cannot encode biases that depend on the specific parameterization of the effect. That is, they all represent the same causal structure (but are parameterized differently) where restricting the analysis to live births (i.e.,

conditioning on collider $S = 1$) induces a spurious association between A and U, which results in a biased $A - Y$ association (Figure 1).

Unlike other examples of selection bias, such as the birthweight paradox[7], obesity paradox[12–14] or loss to follow-up in cohort studies[15], live-birth bias is less amenable to addressing analytically as we cannot adjust for selection processes that we cannot observe (i.e., the necessary data to mitigate this bias are often not available). Thus, simulations are an invaluable tool for exploring the influence of live-birth bias on the estimation of the effects of exposure during pregnancy on outcomes in live-born children. Motivated by the findings of Raz et al.[5], we examine through simulations the magnitude of bias that would result from analyses under the two hypothetical selection mechanisms as well as when they operate simultaneously.

METHODS

Data-generating process

To examine bias from selection on live births under a true null effect of NO_2 on ASD, we simulated a pregnancy cohort of 100,000 conceptions, which we will refer to as the "total population", with data on entire-pregnancy NO_2 exposure A, an unmeasured factor U, the ASD outcome Y and selection indicator S (Figure 1). Entire-pregnancy NO_2 was normally distributed with mean 16.7 and standard deviation of 4.3 to reflect the distribution of NO_2 found in the original study[5]. For simplicity, we ignored the seasonal nature of the NO_2 exposure, and though we treated the exposure as Gaussian, the same principles would apply for a binary exposure. Unmeasured variable U and outcome Y were binary variables. The prevalence of U (π_U) was set to be either 0.25, 0.50 or 0.75. The baseline odds of Y were set to be 0.015 to reflect the low

incidence of ASD in the original analysis[5], and the baseline odds of fetal loss were set to be 0.05, such that the causal effects that lead to selection bias (i.e., $A \rightarrow S$, $U \rightarrow S$, $\{AU\} \rightarrow S$ [i.e., the effect of the $A - U$ interaction], and $U \rightarrow Y$), which we will henceforth refer to as "selection effects", lead to an overall loss in line with observed estimates[16]. All selection effects were modeled in terms of odds ratios, so that simulated probabilities were correctly bounded between 0 and 1; and for the $A \rightarrow S$, $U \rightarrow S$, $\{AU\} \rightarrow S$, and $U \rightarrow Y$ associations (OR_{AS}, OR_{US}, $OR_{\{AU\}S}$, OR_{UY}, respectively) that were not 1.0 as determined by the selection mechanism (see section *Selection Mechanisms* below) were set to all be the same and equal to 1.5, 2.0, 2.5 or 3.0 (for OR_{AS}, this is per 5.85 ppb increase in NO_2, the interquartile range in the original study). Here, we only considered effects of the same sign because exposures that are harmful for pregnancy loss are most likely also harmful for ASD (beneficial exposures would function in the same manner, in that, what is beneficial for loss is also beneficial for ASD, while those of opposite signs which we considered less plausible would lead to upwardly biased $A - U$ and $A - Y$ associations among live births). For simplicity of displaying, we will refer to the selection effects odds ratios as OR_S henceforth. The probability of loss and the ASD outcome Y for each fetus i was estimated using the following equations below. Equation 1 represents the probability that the pregnancy will result in a fetal loss given A and U. Equation 2 represents the probability of the outcome Y given U, where A is omitted since our simulations were conducted under the null; that is, there is no causal effect of A on Y.

$$P(loss_i) = \frac{\exp(\beta_0 + \beta_1 A_i + \beta_2 U_i + \beta_3 A_i * U_i)}{1 + \exp(\beta_0 + \beta_1 A_i + \beta_2 U_i + \beta_3 A_i * U_i)} \quad (1)$$

$$P(Y_i) = \frac{\exp(\gamma_0 + \gamma_1 U_i)}{1 + \exp(\gamma_0 + \gamma_1 U_i)} \quad (2)$$

Selection mechanisms

To examine bias from collider-stratification with no interaction (Mechanism 1 [M1]), where both A and U have independent causal effects on fetal loss, selection effects were set to the OR_S specified above, except that exp(β_3)=$OR_{\{AU\}S}$ was set to 1. For depletion of susceptibles (Mechanism 2 [M2]), exp(β_1)=OR_{AS} and exp(β_2)=OR_{US} were set to be 1, whereas exp(β_3)=$OR_{\{AU\}S}$ was set equal to the pre-specified OR_S; that is, A and U do not have independent causal effects on fetal loss and loss due to NO_2 could only occur in the subset of fetuses who were exposed to U. Finally, to examine bias from both mechanisms operating simultaneously (Both Mechanisms [M1+2]), where both A and U have independent causal effects on fetal loss and they causally interact on selection, exp(β_1)=OR_{AS}, exp(β_2)=OR_{US} and exp(β_3)=$OR_{\{AU\}S}$ were set to the specified OR_S. For all mechanisms, OR_{UY}=exp(γ_1) was set to the prespecified selection effect OR_S, such that the extent of the bias is driven by the differing parameterizations of the relations between A, U, and S across the three selection mechanisms (and not the $U - Y$ relationship, which is fixed to be constant for each scenario). In order to focus only on the bias induced by the selection effects, all simulations assumed that there was no confounding for the effect of NO_2, loss to follow-up among live-born children, outcome misclassification or exposure misclassification, such that observed associations can only be explained by live-birth bias.

Analysis

Each scenario was simulated 1,000 times. For each simulated dataset, we first restricted our analytic sample to live births (i.e., $S = 1$), and then performed a logistic regression of ASD status in children with NO_2 exposure to obtain the observed odds ratio $OR_{AY|S=1}$ (per 5.85 ppb), which approximates the risk ratio as the outcome is rare. With the distribution of point estimates generated over the 1000 iterations for each scenario, we computed the mean $OR_{AY|S=1}$ and percentile-based 95% simulation intervals (SIs), which are the 2.5th and 97.5th percentiles of the distribution. As the simulated truth is that there is no causal effect, the value of $OR_{AY|S=1}$ demonstrates the bias ratio, where greater departures from 1 indicate larger magnitudes of bias. Furthermore, simulation intervals demonstrate the range of $OR_{AY|S=1}$ estimates that are consistent with the data generating mechanism for the specified sample size; for example, if the 95% SIs generated by a given selection mechanism included the odds ratio of 0.77 found in the original study [5], it would suggest that this observed protective association would be consistent with live birth bias induced by that mechanism.

To better understand the drivers of bias from the different selection mechanisms, we also estimated the odds ratio for the association between A (NO_2 exposure, per 5.85 ppb) and U in the selected population ($OR_{AU|S=1}$) using a logistic regression, the prevalence of U in the selected population ($\pi_{U|S=1}$), and their respective 95% SIs. Since both parameters determine the strength of live-birth bias and are driven by the simulation inputs OR_S and π_U, we will henceforth refer to both parameters as "bias parameters". The simulation input OR_{AU} is expected to be 1 in the total population of all conceptions, but the parameter $OR_{AU|S=1}$ is expected to be below 1 in the selected population (i.e., fetuses that survived) since those exposed to both high air pollution A and U are strongly selected against because both factors increase the likelihood of loss. Thus,

those exposed to high air pollution in the selected population are less likely to be exposed to U (and vice versa) setting up an inverse association between A and U. Furthermore, the difference between $\pi_{U|S=1}$ and π_U indicates the extent to which the $U \rightarrow S$ and $A \rightarrow S$ determine the prevalence of U in those selected; that is, the expected value of $\pi_{U|S=1}$ is π_U in the absence of bias.

Finally, to examine the extent of the bias that would occur if there were multiple Us involved in the fetal selection process, we also estimated the value of $OR_{AY|S=1}$ and its 95% SI for each scenario in the presence of two, and then three Us, where all Us were set to have the same prevalence and effect on selection. All simulations and analyses were performed in R (R Foundation for Statistical Computing, Vienna, Austria)[17]. The simulation code and documentation are available at https://github.com/mleung-harvard/live-birth-bias-simulation and in the Appendix so that readers can explore the extent of potential biases with any parameters they wish.

RESULTS

The results of this simulation study on the bias in average $OR_{AY|S=1}$ are shown in Figure 2 and in Table S1. In the presence of collider-stratification with no interaction (M1), where both NO_2 exposure A and unmeasured variable U have independent causal effects on fetal loss and therefore selection S (i.e., OR_{AS} and OR_{US} were set to the prespecified selection effect, but $OR_{\{AU\}S}$ was set to 1), the bias was generally weak (Figure 2; Table S1). Selection effects of magnitudes 1.5 and 2.0 generated little to no bias on average across the three values of π_U with $OR_{AY|S=1}$ ranging from 0.99 to 1. Only when the selection effects reached 3.0 did we see larger

departures from the null (e.g., $OR_{AY|S=1}$ of 0.94 for π_U of 0.5), but these were still relatively weak such that the 95% SI (i.e., the distribution of point estimates consistent with this mechanism) still included the null (Table S1). Examining the bias parameters $OR_{AU|S=1}$ and $\pi_{U|S=1}$, we observed that stronger selection effects in the total population yielded a lower $OR_{AU|S=1}$ (i.e., a stronger inverse association between A and U), but a smaller $\pi_{U|S=1}$ in the selected population (Figure 3; Table S2); that is, with stronger selection effects, both parameters deviate further from the underlying population parameter, where $OR_{AU|S=1}$ would be 1 and $\pi_{U|S=1}$ would be equal to π_U in the absence of bias.

For depletion of susceptibles (M2), where fetal loss is solely dependent on the interaction between NO_2 exposure A and unmeasured variable U (i.e., OR_{AS} and OR_{US} were set to 1 but $OR_{\{AU\}S}$ was set to the prespecified selection effect), the magnitude of bias was slightly stronger compared with those generated by M1 (Figure 2; Table S1). Unlike with M1, with M2, $OR_{AY|S=1}$ was consistently low such that several 95% SI did not include the null (Table S1); for example, if the selection effects were 3.0 and 25% of the total population were exposed to U, then the observed OR for the NO_2-ASD association would be 0.91 (95% SI: 0.85, 0.97). When selection parameters were relatively weak (i.e., OR_S of 1.5 and 2), corresponding values of $OR_{AU|S=1}$ (for the same π_U) deviated further from the null under M2 compared to M1 (Figure 3; Table S2). For example, when the selection effect $OR_{\{AU\}S}$ was set to 1.5 (and both OR_{AS}=1 and OR_{US}=1) and $\pi_U = 0.75$, $OR_{AU|S=1}$ was 0.94 for depletion of susceptibles compared to 0.98 for M1.

When both mechanisms work together (M1+2), where both A and U have independent causal effects and causally interact on fetal loss (i.e., OR_{AS}, OR_{US}, and $OR_{\{AU\}S}$ were set to the prespecified selection effects), the magnitude of bias was usually strongest (Figure 2; Table S1). For example, if the selection effects were 3.0 and 50% of the total population were exposed to U (π_U=0.50), then the average OR for ASD by NO_2 among live births ($OR_{AU|S=1}$) would be 0.85 (95% SI: 0.74, 0.97) (Table S1). Even if we only change the selection effects to 2.0, the observed OR was 0.90 (95% SI: 0.82, 0.99) (Table S1). Examining the bias parameters, both $OR_{AU|S=1}$ and $\pi_{U|S=1}$ deviated further from their corresponding population parameters under M1+2, compared to both M1 and M2 (Figure 3; Table S2). For example, when the prevalence of U was 0.75 and the selection effect were set to 3 (OR_{AS}=3, OR_{US}=3, $OR_{\{AU\}S}$=3), $OR_{AU|S=1}$=0.21 and $\pi_{U|S=1}$=0.19 for M1+2 compared to $OR_{AU|S=1}$=0.77 and $\pi_{U|S=1}$=0.62 for M1 (OR_{AS}=3, OR_{US}=3, $OR_{\{AU\}S}$=1) and $OR_{AU|S=1}$=0.58 and $\pi_{U|S=1}$=0.60 for M2 (OR_{AS}=1, OR_{US}=1, $OR_{\{AU\}S}$=3).

In the presence of multiple Us, the bias is amplified with increasing number of Us, but the extent of the amplification differs by selection mechanism as shown in Figure 4 and in Table S3. For M1, the increase in bias is overall small, yielding small to moderate associations even in the presence of three Us. For example, when both OR_{AS} and OR_{US} for three Us were set to 1.5, the resulting $OR_{AY|S=1}$ were 0.99 for all values of π_U, compared with $OR_{AY|S=1}$ of 1 for all values of π_U when only one U was simulated. When both OR_{AS} and OR_{US} for three Us were set to 3.0, the resulting $OR_{AY|S=1}$ ranged from 0.87-0.90 across values of π_U, compared with $OR_{AY|S=1}$ of 0.94-0.95, when only one U was simulated. The amplification of bias with additional U parameters was stronger for M2 and strongest when both mechanisms co-occurred (M1+2). For example, when $OR_{\{AU\}S}$ for three $A-U$ interactions were set to 1.5 under M2 (OR_{AS} and OR_{US} set to 1),

$OR_{AY|S=1}$ ranged from 0.94-0.97 compared with 1.0 for all values of π_U when only one U was simulated. When OR_{AS}, OR_{US}, and $OR_{\{AU\}S}$ were set to 1.5 under mechanism M1+2 with three Us, $OR_{AY|S=1}$ ranged from 0.88-0.95, compared with 0.98 for all values of π_U when only one U was simulated. Bias increased under both mechanisms as the selection effects increased in magnitude and the prevalence of U was high. The most extreme bias occurred when there were three Us under M1+2, OR_S was set to 3.0, and π_U was set to 0.75, resulting in an average $OR_{AY|S=1}$ of 0.51 (95% SI: 0.34, 0.73) for the NO_2-ASD association when the population was restricted to live births.

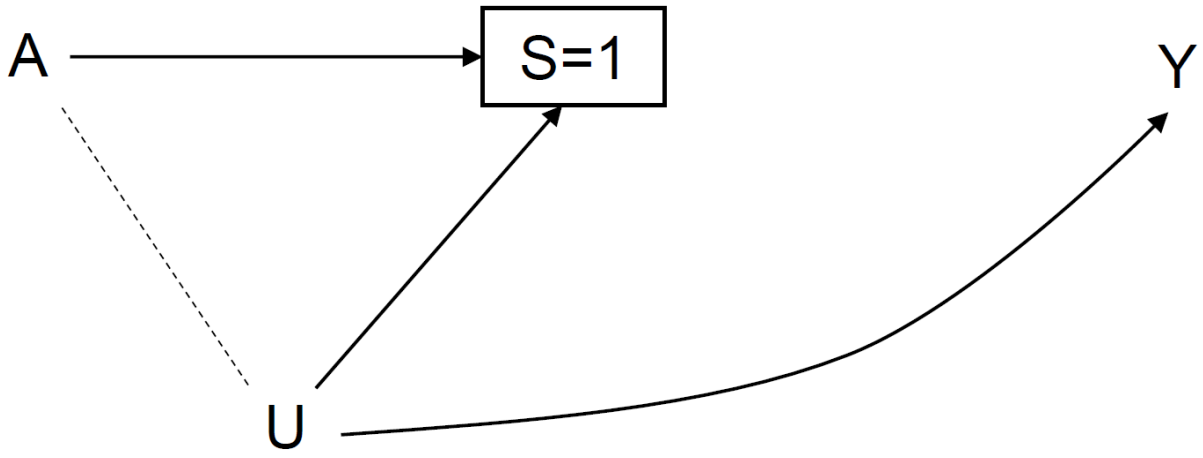

Figure 1.1. Directed Acyclic Graph (DAG) of the structure of live birth bias. Nitrogen dioxide (NO2) exposure A affects live births S that is also affected by an independent unmeasured risk factor U for autism spectrum disorder (ASD) outcome Y. Arrows are direct causal effects, and the dashed line is a spurious association induced between A and U after selection on live births (i.e., conditioning on $S = 1$). This DAG has the same structure as the birthweight paradox[7].

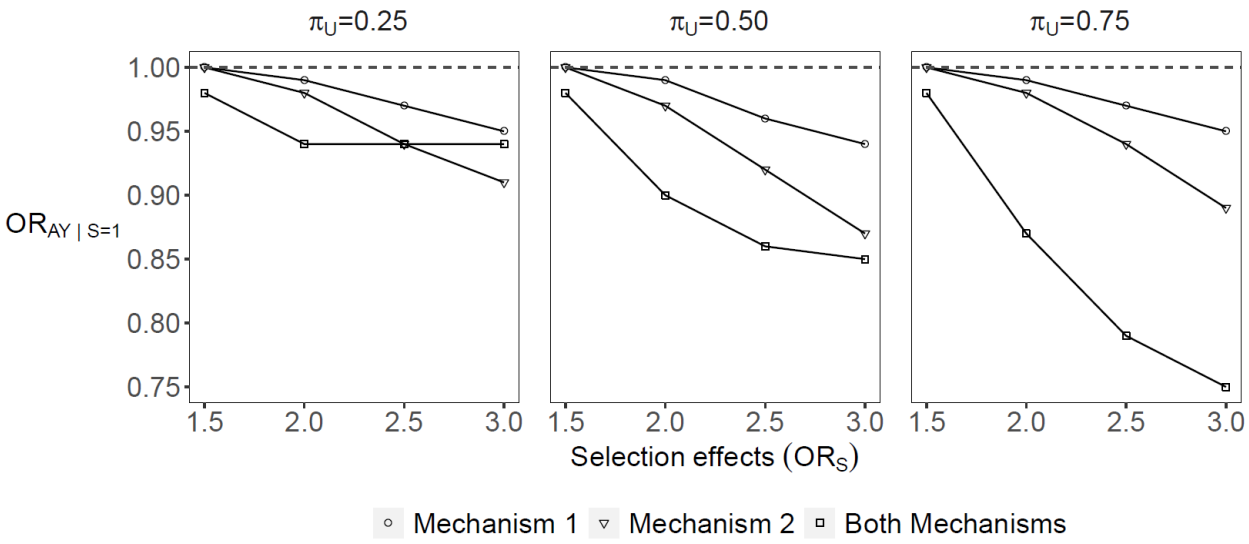

Figure 1.2. Live birth bias of OR_{AY} under different selection effects. Average odds ratios for the association between nitrogen dioxide (NO_2; exposure A) and autism spectrum disorder (ASD, outcome Y) among live births $S = 1$ ($OR_{AY|S=1}$) with varying simulation inputs for the prevalence of the unmeasured risk factor U (π_U) and the magnitude of selection effects (OR_S) under two selection mechanisms (collider-stratification without interaction, and depletion of susceptibles) and when they both co-occur with a single U, assuming a true null effect of NO_2 on ASD. Collider-stratification without interaction (Mechanism 1) occurs when A and U have independent causal effects on fetal loss, but with no interaction on the multiplicative scale ($OR_{\{AU\}S}$=1, and OR_{AS}=OR_{US}=OR_{UY}=OR_S); Depletion of susceptibles (Mechanism 2) occurs when A has a causal effect on fetal loss only in the subset of susceptible fetuses ($U = 1$), but neither A or U have independent causal effects on fetal loss (OR_{AS}=OR_{US}=1, and $OR_{\{AU\}S}$=OR_{UY}=OR_S); Both mechanisms occur when A and U have independent causal effects on fetal loss, and with interaction on the multiplicative scale (OR_{AS}=OR_{US}=$OR_{\{AU\}S}$=OR_{UY}=OR_S). Each scenario was simulated 1000 times. Points represent the mean $OR_{AY|S=1}$ in each scenario. Dashed lines indicate the true null effect of NO_2 on ASD (OR_{AY}=1) in the absence of live birth bias, where deviations from 1.0 quantify the magnitude of live birth bias. See Table S1 for corresponding numeric data, including 95% simulation intervals (SI).

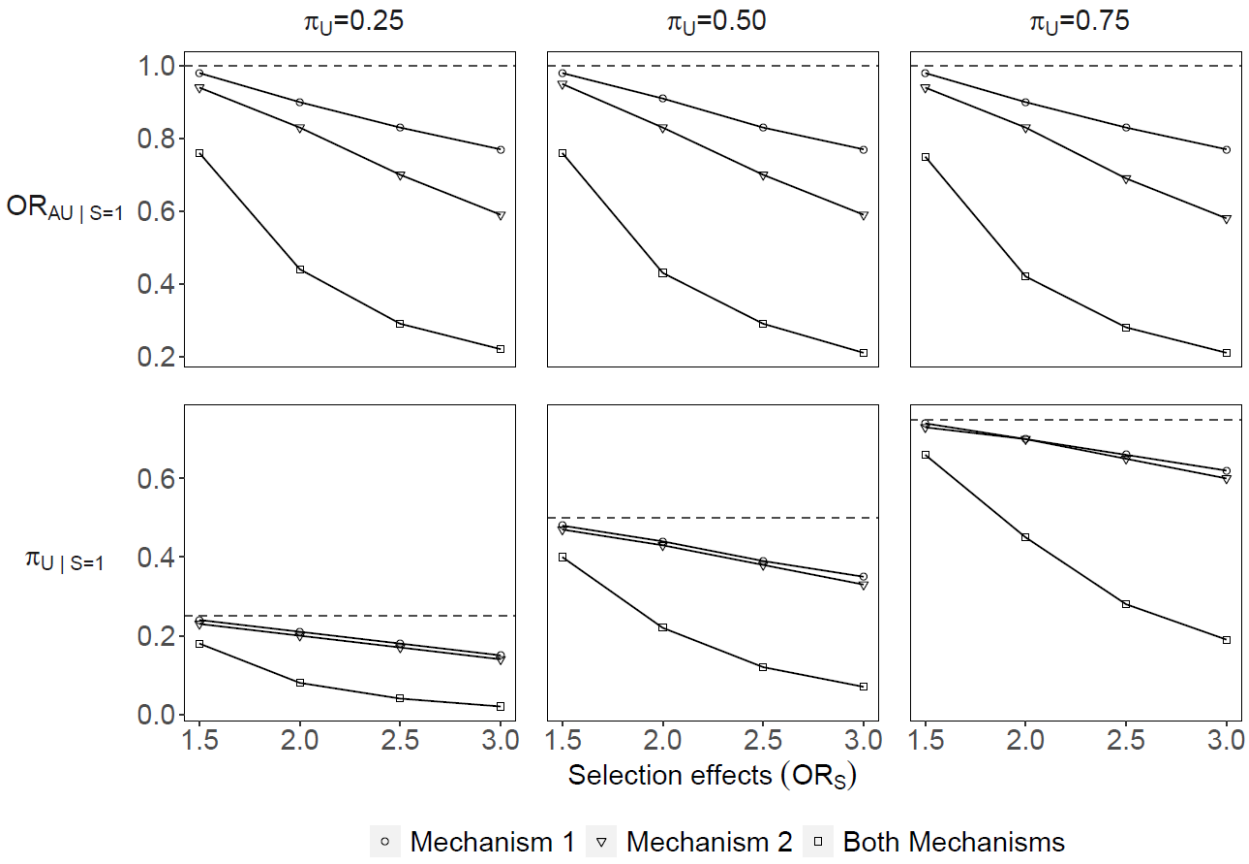

Figure 1.3. Bias parameters that drive live birth bias of OR_{AY} under different selection effects. Average bias parameters in the selected population with varying simulation inputs for the prevalence of the unmeasured risk factor U (π_U) and the magnitude of selection effects (OR_S) under two selection mechanisms and when they both co-occur with a single U, assuming a true null effect of nitrogen dioxide (NO_2; exposure A) on autism spectrum disorder (ASD, outcome Y). In the selected population (live births), $OR_{AU|S=1}$ is the association between A and U, and $\pi_{U|S=1}$ is the prevalence of U. Collider-stratification without interaction (Mechanism 1) occurs when A and U have independent causal effects on fetal loss, but with no interaction on the multiplicative scale ($OR_{\{AU\}S}$=1, and OR_{AS}=OR_{US}=OR_{UY}=OR_S); Depletion of susceptibles (Mechanism 2) occurs when A has a causal effect on fetal loss only in the subset of susceptible fetuses ($U = 1$), but neither A or U have independent causal effects on fetal loss (OR_{AS}=OR_{US}=1, and $OR_{\{AU\}S}$=OR_{UY}=OR_S); Both mechanisms occur when A and U have independent causal effects on fetal loss, and with interaction on the multiplicative scale (OR_{AS}=OR_{US}=$OR_{\{AU\}S}$=OR_{UY}=OR_S). Each scenario was simulated 1000 times. Points represent the mean value of the bias parameter in each scenario. Dashed lines indicate the expected values (in the absence of live birth bias) for $\pi_{U|S=1}$ ($\pi_{U|S=1}$=π_U), and $OR_{AU|S=1}$($OR_{AU|S=1}$=OR_{AU}=1), which are the parameters in the selected population that drive the strength of live birth bias of OR_{AY}. See Table S2 for corresponding numeric data, including 95% simulation intervals (SI).

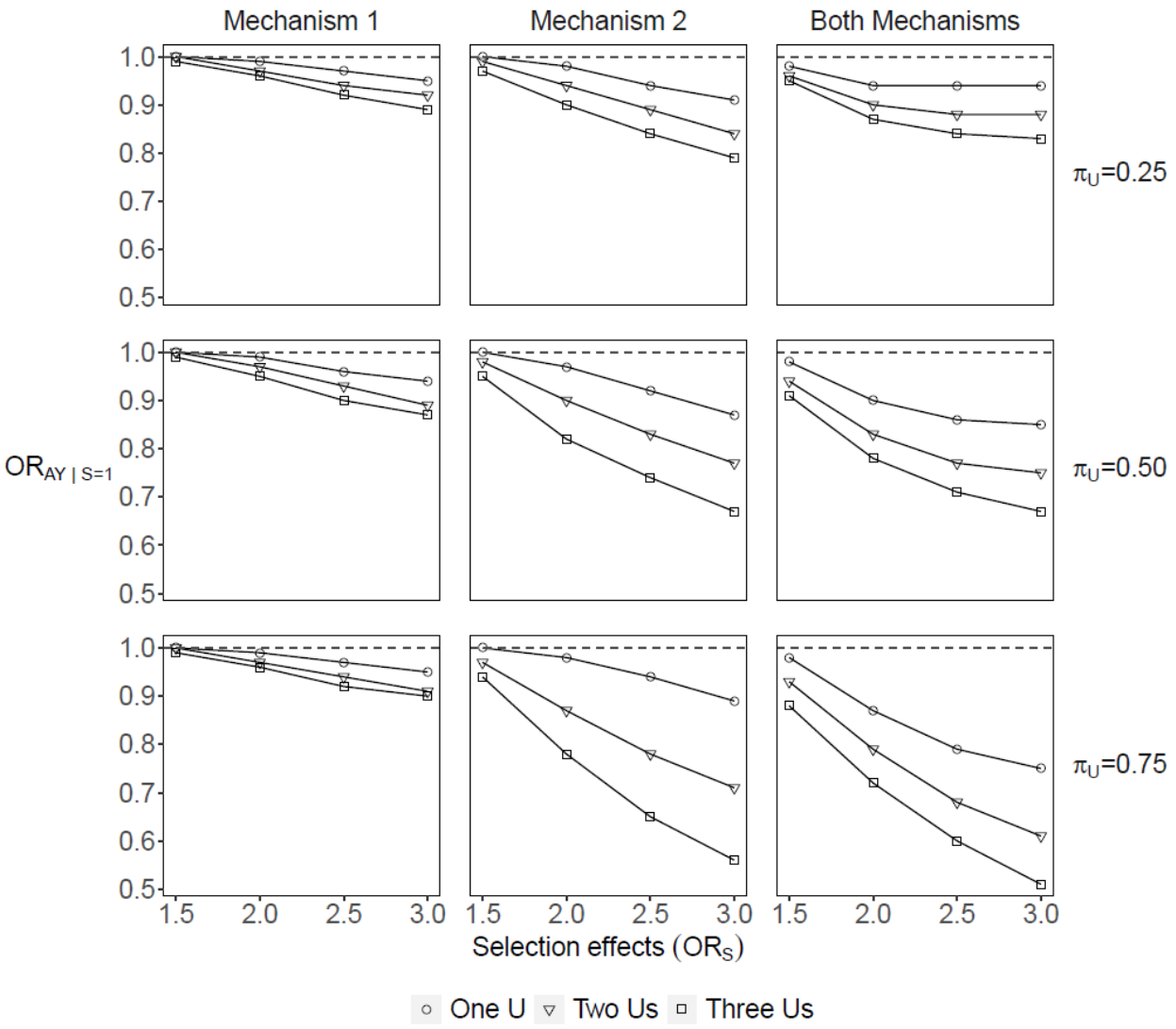

Figure 1.4. Live birth bias of OR_{AY} under different selection effects and different numbers of unmeasured risk factors for selection and the outcome. Average odds ratios for the association between nitrogen dioxide (NO_2; exposure A) and autism spectrum disorder (ASD, outcome Y) among live births $S = 1$ ($OR_{AY|S=1}$) with varying simulation inputs for the prevalence of the unmeasured risk factor U (π_U) and the magnitude of selection effects (OR_S) under two selection mechanisms and when they both co-occur with one, two or three Us, assuming a true null effect of NO_2 on ASD. U is a vector that consists of ≤ 3 unmeasured factors (U_1, U_2, U_3), where input parameters were applied equally for each unmeasured factor; thus, all references to U henceforth applies to each of the unmeasured factors U_1, U_2, U_3. Collider-stratification without interaction (Mechanism 1) occurs when A and U have independent causal effects on fetal loss, but with no interaction on the multiplicative scale ($OR_{\{AU\}S}$=1, and OR_{AS}=OR_{US}=OR_{UY}=OR_S); Depletion of susceptibles (Mechanism 2) occurs when A has a causal effect on fetal loss only in the subset of susceptible fetuses ($U = 1$), but neither A or U have independent causal effects on fetal loss (OR_{AS}=OR_{US}=1, and $OR_{\{AU\}S}$=OR_{UY}=OR_S); Both mechanisms occur when A and U have independent causal effects on fetal loss, and with

interaction on the multiplicative scale (OR_{AS}=OR_{US}=$OR_{\{AU\}S}$=OR_{UY}=OR_{S}). Each scenario was simulated 1000 times. Points represent the mean $OR_{AY|S=1}$ in each scenario. Dashed lines indicate the true null effect of NO_2 on ASD (OR_{AY}=1) in the absence of live birth bias, where deviations from 1.0 quantify the magnitude of live birth bias. See Table S3 for corresponding numeric data, including 95% simulation intervals (SI).

DISCUSSION

In our simulations, we found that the magnitude of bias was generally weak for collider-stratification without interaction (M1), which is consistent not only with previous research on live-birth bias[4], but also with existing literature that has focused on collider-stratification in other observational settings, such as the birthweight paradox[7], obesity paradox[12,13] and selection into genetic studies[18], where much stronger and perhaps implausible effects are required to induce substantial bias[19,20]. Thus, it is unlikely that collider-stratification without interaction alone can account for the observed association between pregnancy wide NO_2 exposure and ASD in the original study[5], as few simulated scenarios yielded odds ratios that were close to 0.77, the OR for NO_2 during pregnancy when mutually adjusted for postnatal exposure to NO_2[5]. Although other factors could have been at play as well, depletion of susceptibles (M2) or both mechanisms operating simultaneously (M1+2) would be more likely to be able to account for the observed association in the original study all on their own, as the simulations under these scenarios consistently generated moderate to strong protective associations due to more extreme selection bias parameters among live births—particularly when multiple Us were present.

To better understand the differences between the two selection mechanisms with just a single unmeasured variable U, we also estimated the $A \rightarrow U$ association ($OR_{AU|S=1}$) and prevalence of U ($\pi_{U|S=1}$) in the selected populations, as the magnitude of bias ($OR_{AY|S=1}$) is constrained by these two parameters, and the effect of U on ASD (OR_{UY}). However, as OR_{UY} was the same across mechanisms in our simulations, any discrepancies in $A - Y$ bias are driven by both $OR_{AU|S=1}$ and $\pi_{U|S=1}$. Here, M1 generally yielded weaker $OR_{AU|S=1}$ but similar $\pi_{U|S=1}$ compared to M2, which explains why M1 produces a weaker bias compared to M2. When both

mechanisms co-occur (M1+2), $OR_{AU|S=1}$ is lower (i.e., A and U are more strongly negatively associated) than either M1 or M2; this is unsurprising given that the effects of A and U on selection are "super-additive" under M1+2, in that the contributions of A and U together exceed the sum of their contributions when A and U are considered separately, as shown in Equation 1. Furthermore, $\pi_{U|S=1}$ is also lower under M1+2, however this does not necessarily generate more bias since similar to the magnitude of bias due to confounding[21], bias is maximized when $\pi_{U|S=1}$=0.50 ($\pi_{U|S=1}$ that is close to 0 or 1 actually reduces bias, as this is akin to conditioning on or stratifying by variable U). Thus, when $\pi_U \leq 0.5$, there exists a tension between the strength of $OR_{AU|S=1}$ and the distance of $\pi_{U|S=1}$ from 0.50, since stronger selection effects reduced $\pi_{U|S=1}$ (i.e., it moves further away from 0.50) such that it could offset the bias generated by the stronger $OR_{AU|S=1}$. For example, when π_U=0.25 and OR_S=3.0, the bias under M1+2 was weaker than under M2 ($OR_{AY|S=1}$=0.94 under M1+2 and $OR_{AY|S=1}$=0.91 under M2) because despite the stronger $OR_{AU|S=1}$ ($OR_{AU|S=1}$=0.22 under M1+2 and $OR_{AU|S=1}$=0.59 under M2) $\pi_{U|S=1}$ was further away from 0.50 ($\pi_{U|S=1}$=0.02 under M1+2 and $\pi_{U|S=1}$=0.14 for M2). On the other hand, when π_U=0.75, stronger selection effects strengthened $OR_{AU|S=1}$, and lowered $\pi_{U|S=1}$ closer to 0.50 (except for when OR_S=3.0 under M1+2, where $\pi_{U|S=1}$=0.19), such that both bias parameters worked in concert to increase the bias in the overall $OR_{AY|S=1}$ association.

Evaluating the range of plausible parameters for simulation inputs is important for bias analyses. With a single unmeasured variable U, neither M1 nor M2 produced associations near the observed OR of 0.77 from the original study[5]. On the other hand, when both mechanisms operate together (M1+2) and there is just a single unmeasured variable U, not only do the selection effects need to be quite strong, but also U needs to be relatively common in the population to

yield this observed OR. For example, an OR_S of 3.0 and π_U of 0.75 yielded an observed association of 0.75 (95% CI: 0.64, 0.89) in our simulations. Thus, with a single U, not only are there very few scenarios that could produce the observed estimate, but the magnitude of the input parameters that could potentially generate the bias strains credibility. That is, although it is perhaps plausible that one of the selection effects is that large, it seems unlikely that all three are. Therefore, strong bias is more likely under either M2 or M1+2 when U is a composite of uncontrolled variables (as well as those that have been controlled for, just imperfectly), as the OR_S would only need to be between 1.5 and 2.0 to generate a bias of similar magnitude. With selection effects of 2.0, and π_U=0.75 for each U, the $OR_{AY|S=1}$ was 0.87 and 0.78 under M2 when there were two and three Us, respectively; whereas, with the same input parameters, the $OR_{AY|S=1}$ was 0.79 and 0.72 under M1+2 when there were two and three Us, respectively. Even if these unmeasured factors were just as strongly associated with S and Y, but were less prevalent in the population, the mean $OR_{AY|S=1}$ is only slightly attenuated. For example, under the same selection effects, but now π_U=0.50 for each U, the $OR_{AY|S=1}$ was 0.90 and 0.82 under M2 when there were two and three Us, respectively, and the $OR_{AY|S=1}$ was 0.83 and 0.78 under M1+2 when there were two and three Us, respectively.

Potential candidates for U include prenatal stress[22–38], maternal smoking[39–42], genetic factors[43–45] and environmental stressors, such as endocrine-disrupting chemicals[8–11]. Many of these associations have been reported to be in line with or stronger than an OR_S of 1.5 – 2.0 and the collective exposure to these factors (or just a subset) in the population is likely not uncommon. For example, maternal smoking during pregnancy (any versus none) has been associated with an OR of 1.47 for stillbirth[42], and an OR of 1.56 for autism[41]. Furthermore, endocrine-disrupting

chemicals such as polychlorinated biphenyls (PCBs) have been associated with pregnancy loss with ORs ranging from 1.6-2.52 depending on the type of PCB when comparing those accidentally exposed versus those unexposed to accidental contamination of rice oil during the Yusho incident in Japan in 1968[10,46]. PCBs have also been associated with increased odds of autism, where a prior study reported ORs ranging from 1.20-1.97 depending on the type of PCB when comparing the highest to the lowest quartile of exposure[47]. Although we identified these Us for our illustrative example of NO_2 and ASD, they are also relevant for the estimation of the effects of any exposure during pregnancy that has the potential to cause loss and ASD. Furthermore, an OR_{AS} of 1.5 – 2.0 is also plausible for the effects of NO_2 on loss. Although past studies have reported ORs ranging from 1.04 – 1.27 for the association between NO_2 (typically per 10 ppb) and pregnancy loss[48], these are likely biased downwards since only a subset of pregnancy losses come to medical attention and can be studied. That is, early pregnancy loss, which has been estimated to be around 20-30%[16], are typically not observed, such that NO_2-induced loss early in pregnancy would go undetected and the resultant association would underestimate the true harmful effect of NO_2 on pregnancy loss. As our simulation code is available online, we encourage other investigators to evaluate the potential bias arising from live-birth bias with input parameter values that are relevant to their own research.

Although our simulations could generate the magnitude of the protective effect reported from the original nested case-control study using a range of plausible input parameter values, they are simplified depictions of potential causal structures and, therefore, should not be directly compared with estimates from analyses using real data. For example, for simplicity, we assumed no confounding of the exposure effects, no loss to follow-up among live-born children, no

measurement error, and no seasonal or time trends in the exposure or outcome. It is unlikely that all these assumptions would hold in a real analysis. The potential bias in the original nested case-control study may actually be a net downward bias (assuming that exposure to NO_2 during pregnancy is not neuroprotective) arising from a combination of residual and/or unmeasured confounding, selection bias due to non-random attrition between birth and ASD assessment (although such selection would be subject to the same issues we describe here for live-birth bias), exposure measurement error (which typically biases the estimate towards the null), outcome misclassification, and model misspecification (e.g., imperfect control for seasonal trends) in addition to live-birth bias. However, the simplicity of our current simulation study is also its strength, though, in that in our simulations we can isolate and fully identify the bias due to the specific fetal selection mechanism. Furthermore, as prior knowledge of the magnitude and sign of the selection effects (OR_{AS}, OR_{US}, $OR_{\{AU\}S}$, OR_{UY}) is limited since pregnancy loss is a challenging outcome to study[16,49], we set these ORs to be equal in our simulations for simplicity, but presumably similar associations could be seen with some OR lower and others higher. Along similar lines, we ran our simulations under the null for simplicity, which is sufficient to evaluate the magnitude of bias as it does not depend on the effect of NO_2 on ASD. For example, if we observe an $OR_{AY|S=1}$ of 0.75 for a given selection mechanism under the null, then a true effect of OR_{AY}=1.33 (i.e., the inverse of 0.75) would be rendered null by this selection mechanism. Finally, we also assumed that exposure groups are exchangeable in the total population of all conceptions (e.g., conceptions are not affected by selection processes induced by pre-conception exposures). However, this may not be the case as there are likely selection processes that influence fertility (i.e., the likelihood of conception). Excluding women of reproductive age who are trying, but are unable, to conceive (as pregnancy is a requirement to study exposures during

pregnancy) may lead to biased exposure-health effects in the set of actual conceptions, which differs from the total population of intended conceptions. This bias would act through similar mechanisms that we address in this paper's simulations, and therefore could amplify live-birth bias in an analysis with real data (Figure S1).

Here, we show that live-birth bias under plausible simulation parameters can lead to associations of NO_2 and ASD that are biased downwards, where the largest bias occurs when both M1 and M2 both operate simultaneously (i.e., M1+2). This bias may explain the inconsistent body of literature[5,50–53], where truly adverse effects may appear not as harmful, null, or even protective. Although we used NO_2 and ASD for our illustrative example, this bias can extend to other studies relevant to fetal programming[54], which can limit the identification of harmful prenatal exposure effects and prevent the development of interventions during pregnancy aimed at promoting better health. For example, it is possible that live-birth bias can also explain the unexpected negative associations between prenatal exposure to perfluoroalkyl substances (PFAS) with ASD[55] and attention-deficit/hyperactivity disorder[56–59].

In order to rule out live-birth bias as a threat to internal validity, we would need to show that the exposure in question does not affect selection (i.e., fetal loss); that is, if we find that exposure does not affect selection, either independently or in conjunction with another risk factor for fetal loss, then the association with the outcome cannot be biased through this mechanism. If, on the other hand, exposure is associated with selection, then to potentially mitigate or eliminate this bias, we would need to collect information on U (something we would need to plan for in the study design phase) and adjust for it in the analysis. Lastly, if there is reason to believe that there

are no common causes of selection and the outcome, then there would be no live-birth bias even if the exposure affects selection (although this is impossible to verify in practice). It would still be worthwhile to quantify the effect of exposure on selection, as it gives us insight into the change in the potential number of losses and the child outcome (when considered jointly with its effect on the outcome), if we were to intervene to set exposure to another level. All of these analyses require estimating the effects on selection which is no simple feat, but new approaches for studying pregnancy loss without needing to enumerate the population at risk (i.e., all conceptions)[60], makes such an undertaking less daunting. Thus, our study findings highlight the need for cautious interpretations of studies of the effects of prenatal exposures on postnatal outcomes, and for more investment into research on the determinants of pregnancy loss.

Chapter 2: Exposure to $PM_{2.5}$ during pregnancy and fetal growth in Massachusetts, USA

Michael Leung[1], Marc G. Weisskopf[1,2], Francine Laden[1,2,3], Brent Coull[4], Anna M. Modest[5,6], Michele R. Hacker[1,5,6], Blair J. Wylie[5,6], Yaguang Wei[2], Joel Schwartz[1,2], Stefania Papatheodorou[1]

[1]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA

[2]Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA

[3]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA

[4]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA

[5]Department of Obstetrics and Gynecology, Beth Israel Deaconess Medical Center, Boston, MA

[6]Department of Obstetrics, Gynecology and Reproductive Biology, Harvard Medical School, Boston, MA

ABSTRACT

Background: Prior studies have examined the association between particulate matter less than 2.5 μm ($PM_{2.5}$) and fetal growth with either limited spatial or temporal resolution.

Objectives: In this study, we examined the association between $PM_{2.5}$ exposure during pregnancy and fetal growth outcomes (ultrasound biometric parameters and birth weight) in a pregnancy cohort using spatiotemporally resolved $PM_{2.5}$ in Massachusetts, USA.

Methods: We used prenatal and obstetric data from 10,008 pregnancies that delivered at Beth Israel Deaconess Medical Center from 2011-2016. There were 26,151, 26,038, 26,035, and 25,978 ultrasound measurements for biparietal diameter (BPD), head circumference (HC), femur length (FL), and abdominal circumference (AC), respectively, and 9,991 measurements for birth weight. We used linear mixed effects models to examine the associations of $PM_{2.5}$ in the first 16 weeks of pregnancy with anatomic ultrasound scans (<24 weeks), growth ultrasound scans (≥24 weeks), and birth weight. All models were adjusted for sociodemographic characteristics, maternal comorbidities, long-term trends, and temperature.

Results: Higher $PM_{2.5}$ exposure in the first 16 weeks was associated with reductions in all fetal growth outcomes, where associations were particularly strong for BPD, AC, and birth weight. For example, a 5 $\mu g/m^3$ increase in $PM_{2.5}$ was associated with a BPD z-score reduction of -0.18 before 24 weeks, an AC z-score reduction of -0.15 after 24 weeks, and a birth weight z-score reduction of -0.12. Our secondary analyses examining the associations with cumulative $PM_{2.5}$ exposure up until the assessment of fetal growth produced attenuated associations.

Conclusions: Gestational exposure to $PM_{2.5}$ was associated with impaired fetal growth at levels below the current national standards, where early to mid-pregnancy appears to be a critical window of exposure.

INTRODUCTION

Particulate matter less than 2.5 microns ($PM_{2.5}$) is a considerable threat to health worldwide[61–63]. Pregnant individuals and their fetuses are likely vulnerable to the effects of $PM_{2.5}$ due to changes in maternal physiology, and the rapid speed of fetal organ formation and development[64]. A recent systematic review showed compelling evidence for an association between maternal exposure to $PM_{2.5}$ and impaired fetal growth as characterized by birth weight[65], a key developmental indicator for perinatal morbidity and mortality as well as later life cardiometabolic outcomes[54]. Several biological mechanisms have been proposed to explain these associations, including inflammation, oxidative stress, endocrine disruption, coagulation changes, and placental dysfunction[66–69].

As the sequence of events during fetal development is very specific (e.g., cell differentiation, organ development, changes in fetal metabolism, etc.), the timing of exposure to $PM_{2.5}$ during pregnancy is likely to manifest in distinct effects on fetal growth parameters. However, the use of newborn anthropometry does not allow for the identification of these developmental windows. Routine ultrasound measurements would make these internal (and therefore hidden) processes observable. However, only a few studies have used fetal ultrasound parameters to examine the timing of when the growth-restricting effects of $PM_{2.5}$ manifest[70–72]. Although all three studies found that increased prenatal $PM_{2.5}$ exposure was associated with reduced fetal growth from mid-gestation onwards, they suffer from several limitations with regards to exposure assessment: 1) limited spatial resolution, in that two of the three used the nearest land-based monitors for exposure assessment[71,73], and 2) assessment of long-term cumulative $PM_{2.5}$ exposure without considering other potentially relevant developmental windows (two of the three assessed $PM_{2.5}$

concentrations from conception to the date of ultrasound[71,73], while the third assessed annual average $PM_{2.5}$ concentrations[70]). Thus, we aimed to overcome these limitations by examining the association between $PM_{2.5}$ exposure during several exposure windows and fetal growth in a pregnancy cohort with routine ultrasound and spatiotemporally resolved $PM_{2.5}$ from Massachusetts, USA.

METHODS

Study population

We used prenatal and obstetric data from Beth Israel Deaconess Medical Center (BIDMC), a large tertiary-care hospital which covers a wide geographic area in Massachusetts with no specific pattern of referral, therefore representing the general population residing in the study area. In this study, we included all pregnancies delivered by the practices in which all obstetric ultrasounds are performed through BIDMC. We restricted the analyses to live births ≥20 weeks of gestation from 2011 through 2016, which is the period for which both ultrasound and $PM_{2.5}$ data were available. Of these, we excluded individuals with multifetal gestations due to different growth trajectories, and residential addresses outside of Massachusetts. Full addresses were available for each delivery and were geocoded to latitude and longitude using the Google Maps Application Programing Interface. This study was approved by the institutional review boards of Harvard T.H. Chan School of Public Health and Beth Israel Deaconess Medical Center.

$PM_{2.5}$ exposure

We assigned $PM_{2.5}$ exposure based on where the pregnant individual resided at birth in Massachusetts from a state-of-the-art spatiotemporal model that estimates daily $PM_{2.5}$

concentration for each 1x1 km grid across the Continental United States[74]. Briefly, the model uses an ensemble of three machine learning algorithms (artificial neural network, random forest, and gradient boosting) that incorporates satellite-based aerosol optical depth, simulation outputs from three chemical transport models, land-use predictors, and meteorological predictors to estimate daily concentrations of $PM_{2.5}$ at each grid. The predictive model is calibrated using data from 1,928 monitoring stations that belong to the Environmental Protection Agency Air Quality System plus additional monitoring from the National Park Service IMPROVE network, the SEARCH network in the Southeastern U.S. and the MATES III and MATES IV networks in California. Ten-fold cross-validation revealed good model performance with a total R^2 of 0.86 for the entire United States. These predicted $PM_{2.5}$ data have been used in a large body of work that have examined associations with birth outcomes in Massachusetts[75–77].

Our primary window of exposure, over which daily $PM_{2.5}$ levels were averaged, was the first 16 weeks of pregnancy, as this is the period where organ formation takes place and most major functional defects occur in the fetal anatomy[78]. Furthermore, as a secondary analysis, we also considered cumulative $PM_{2.5}$ exposure up until the assessment of fetal growth; that is, from conception to the ultrasound for fetal ultrasound parameters, and from conception to birth for birth weight.

Fetal ultrasounds and birth weight

We used repeated ultrasound biometric measurements of biparietal diameter (BPD), head circumference (HC), femur length (FL) and abdominal circumference (AC), all of which were reported in millimeters (mm). Ultrasounds were interpreted by maternal-fetal medicine

specialists or radiologists. Gestational age at the time of ultrasound was based on the best obstetric estimate combining information from the last menstrual period and the earliest ultrasound performed in pregnancy[79].

Ultrasounds were classified *a priori* into two groups based on the type of scan. The standard of care is one scan at about 18-20 weeks to evaluate fetal anatomy but can sometimes be delayed to 21-23 weeks if the participant is late for prenatal care. Scans conducted later in pregnancy are typically ordered for pregnancies at risk for impaired fetal growth or other complications. Thus, we will refer to scans before 24 weeks as "anatomic scans", and those at 24 weeks and beyond as "growth scans" henceforth. We excluded measurements that were considered implausible, defined as values below or above 4 standard deviations (SD) from the mean of the cohort at that gestational age. Furthermore, to enable comparisons across gestational weeks, we generated z-scores for each of the fetal growth measures by applying the INTERGROWTH-21st standards for fetal growth[80]. Since these standards are only available up to 40 weeks of gestation, ultrasound scans conducted after 40 weeks were excluded. Finally, we also abstracted birth weight from the medical records, which was reported in grams, and generated z-scores using the INTERGROWTH-21st standards for newborn size[80].

Covariates

Temperature data were obtained from the Land Data Assimilation Systems at the NASA Goddard Earth Sciences Data and Information Services Center with 12x12 km spatial resolution [81]. Individual-level covariates were abstracted from the medical records and included maternal age (continuous), race (White, Black, Asian, Hispanic, or Other), education (high school or

higher or less than high school), insurance type (private or public/uninsured), parity (nulliparous or parous), and fetal sex (male or female). Finally, for area-level socioeconomic status, we used the national percentile rankings of the Area Deprivation Index (ADI)[82], which is a neighborhood disadvantage metric derived from 17 census variables on income, education, employment, and housing quality from the American Community Survey (ACS). The ADI from the 2009 – 2013 ACS was linked to each pregnancy at the census block group level.

Epidemiologic air pollution studies also typically control for smoking status[83], and population density[84]; however, neither was included as a covariate in our analyses, as they are already controlled for by design with a form of restriction. That is, of those included in our study, few women reported to have smoked during pregnancy, and only 1 participant lived in a small town/rural area according to the Rural-Urban Commuting Area (RUCA) Codes.

Statistical Analysis

We fitted linear mixed effects models to assess the associations of $PM_{2.5}$ in the first 16 weeks of pregnancy with anatomic ultrasound scans (<24 weeks), growth ultrasound scans (≥24 weeks), and birth weight. To adjust for confounding by long-term and seasonal trends, we used natural splines with 4 degrees of freedom per year (i.e., 24 degrees of freedom for the entire 6-year study period). All covariates previously described also were included in the model, where linear and quadratic terms were used for continuous variables. Furthermore, we included a random intercept for each pregnancy because ultrasounds within each pregnancy were likely to be correlated.

We conducted several additional analyses. First, we examined the associations with cumulative $PM_{2.5}$ exposure, which is the average $PM_{2.5}$ from conception to ultrasound for fetal ultrasound parameters, and whole pregnancy average $PM_{2.5}$ for birth weight. Furthermore, prior analyses have identified several subgroups that are at an increased risk of growth impairment, such as male fetuses[85], and those of Black women and individuals with lower SES (e.g., less education, lower median household income, etc.)[65]. Thus, for each developmental window, we assessed whether the association was separately modified by fetal sex, maternal race, maternal education, and ADI. To do so, we included a product term between $PM_{2.5}$ and the modifier. For each potential modifier, we tested whether the association differed from that of the reference level, which was "female" for fetal sex, "White" for maternal race, "less than college" for maternal education, and the first quartile for ADI. Finally, since all $PM_{2.5}$ exposures in this cohort were below the national ambient standard, it was not necessary to conduct a low dose analysis[86]. All analyses were performed in R (version 3.6.1; R Foundation for Statistical Computing, Vienna, Austria)[17].

RESULTS

Maternal and fetal characteristics for the 10,008 pregnancies included in our study are shown in Table 1, where this sample of eligible participants was formed as in Figure S1. Participants were, on average, 31 years of age at the time of conception, majority White (51%), and had private insurance (78%); about half had a college education or higher (48%). Furthermore, the mean ADI was 22, suggesting that participants in this cohort lived in more advantaged neighborhoods compared to the rest of the US given that an ADI of 50 is the median ADI nationwide. In our sample, the average $PM_{2.5}$ exposure during the first 16 weeks of pregnancy was 7.4 $\mu g/m^3$, where

exposure levels decreased from 9.3 μg/m^3 in the beginning of 2011 to 5.9 μg/m^3 at the end of 2016 (Figure S2). Furthermore, average temperature during the first 16 weeks of pregnancy was 10º Celsius and remained relatively stable over the years of the study period (Figure S2).

Most participants had 2 or more ultrasound measurements (73%) during pregnancy (Table 1), where summary statistics for the fetal ultrasound parameters are displayed in Table 2. The mean z-scores for all ultrasound parameters were similar for anatomic and growth scans, where fetuses had, on average, smaller BPD, but larger HC, FL and AC during pregnancy compared to the international standard [80]. At delivery, the mean birth weight z-score was 0.33, which suggests that newborns in our sample were slightly heavier than the international norm.

The associations between $PM_{2.5}$ exposure in the first 16 weeks of pregnancy and fetal growth outcomes (ultrasound biometric parameters and birth weight) are shown in Table 3. For both anatomic and growth scans, we observed that increased $PM_{2.5}$ was linearly associated with reduced ultrasound parameters, where associations were particularly strong for BPD and AC. For example, a 5 μg/m^3 increase in $PM_{2.5}$ was associated with z-score reductions among anatomic scans (<24 weeks) of -0.18 (95% CI: -0.29, -0.07) for BPD, -0.09 (95% CI: -0.20, -0.02) for HC, -0.07 (95% CI: -0.19, -0.05) for FL, and -0.11 (95% CI: -0.23, 0.01) for AC, and z-score reductions among growth scans (≥24 weeks) of -0.14 (95% CI: -0.23, -0.05) for BPD, -0.03 (95% CI: -0.12, 0.06) for HC, -0.05 (95% CI: -0.14, 0.04) for FL, and -0.15 (95% CI: -0.24, -0.06) for AC. Furthermore, $PM_{2.5}$ exposure in the first 16 weeks was also associated with reduced birth weight, where a 5 μg/m^3 increase was associated with a z-score reduction of -0.12 (95% CI: -0.22, -0.01). Analyses with cumulative $PM_{2.5}$ exposure produced attenuated

associations for the ultrasound parameters, where most 95% CIs include the null. However, the association with birth weight was similar to our primary analysis (Table S1).

Results from our effect modification analyses can be found in Tables S2-5. We did not find any evidence for effect modification by fetal sex (Table S2). However, we found that maternal race, education, and ADI potentially modified the association during specific windows. For maternal race, we found that the negative associations for all ultrasound parameters ≥24 weeks were stronger for Hispanic women compared to White women (Table S3). On the other hand, the growth-restricting effects of $PM_{2.5}$ were not observed after 24 weeks (i.e., among growth scans and birth weight) in Asian women (Table S3). For maternal education, we found that the negative associations appeared to be stronger for ultrasound parameters <24 weeks for those with college or higher education (Table S4). For ADI, we found that the fourth quartile association with FL <24 weeks, fourth quartile association with BPD ≥24 weeks, and second quartile association with HC ≥24 appeared to be attenuated compared the first quartile (Table S5).

Table 2.1. Maternal and fetal characteristics of deliveries at Beth Israel Deaconess Medical Center, Boston, Massachusetts from 2011-2016 (N=10,008)

Characteristics	**N (%)**
Age (years)	
Mean (SD)	31 (5.4)
Education	
College or higher	4769 (48)
Lower than college	3416 (34)
Not specified	1823 (18)
Race	
White	5117 (51)
Black	1750 (17)
Asian	940 (9)
Hispanic	931 (9)
Not specified	1270 (13)
Parity	
Nulliparous	4941 (49)
Parous	5067 (51)
Child sex	
Female	4927 (49)
Male	5081 (51)
Insurance	
Private	7797 (78)
Public or uninsured	2211 (22)
Area Deprivation Index (percentile)	
Mean (SD)	22 (20)
Average $PM_{2.5}$ in first 16 weeks of pregnancy ($\mu g/m^3$)	
Mean (SD)	7.4 (1.5)
Average temperature in first 16 weeks of pregnancy (^{o} C)	
Mean (SD)	10 (7.5)
Number of ultrasounds	
2+ ultrasounds	7320 (73)
1 ultrasound	2688 (27)

Table 2.2. Summary statistics for fetal ultrasound parameters and birth weight from deliveries at Beth Israel Deaconess Medical Center, Boston, Massachusetts in 2011-2016 (N=10,008)

Growth outcome	N	Mean (SD)
Anatomic scans (<24 weeks)		
BPD z-score	10,207	-0.67 (1.10)
HC z-score	10,164	0.37 (1.10)
FL z-score	10,143	0.82 (1.14)
AC z-score	10,112	0.52 (1.17)
Growth scans (≥24 weeks)		
BPD z-score	15,944	-0.86 (1.12)
HC z-score	15,874	0.33 (1.15)
FL z-score	15,892	1.11 (1.10)
AC z-score	15,866	0.56 (1.10)
Birth weight z-score	9,991	0.33 (1.00)

Table 2.3. Linear mixed effects model estimates and 95% CIs for the association between $PM_{2.5}$ during the first 16 weeks of pregnancy and fetal growth outcomes (ultrasound parameters and birth weight) from deliveries at Beth Israel Deaconess Medical Center, Boston, Massachusetts in 2011-2016 (N=10,008)

Growth outcome	Estimate	95% CI	*p*-Value
Anatomic scans (<24 weeks)			
BPD z-score	-0.18	(-0.29, -0.07)	0.002
HC z-score	-0.09	(-0.20, 0.02)	0.09
FL z-score	-0.07	(-0.19, 0.05)	0.23
AC z-score	-0.11	(-0.23, 0.01)	0.08
Growth scans (≥24 weeks)			
BPD z-score	-0.14	(-0.23, -0.05)	0.0003
HC z-score	-0.03	(-0.12, 0.06)	0.56
FL z-score	-0.05	(-0.14, 0.04)	0.30
AC z-score	-0.15	(-0.24, -0.06)	0.001
Birth weight			
Birth weight z-score	-0.12	(-0.22, -0.01)	0.03

Note: Estimates represent the difference in mean z-score with a 5 $\mu g/m^3$ increase in $PM_{2.5}$ during the first 16 weeks of pregnancy after adjusting for maternal age, race, education, insurance type, parity, fetal sex, conception year, day of the year of conception, temperature, and Area Deprivation Index.

DISCUSSION

This large retrospective cohort study showed that higher $PM_{2.5}$ exposure was associated with reductions in all four ultrasound parameters as well as birth weight at levels below the current annual ambient standard of 12 μg/m^3 for $PM_{2.5}$[86]. Associations were particularly strong for BPD, AC, and birth weight; for example, a 5 μg/m^3 increase in $PM_{2.5}$ was associated with a BPD z-score reduction of -0.18 before 24 weeks. These findings have implications for later health and childhood development. For example, head size, as measured by BPD and HC, has been associated with brain development and cognitive achievement in childhood[87–89]; height, as approximated *in utero* by FL, has been associated with educational attainment and economic productivity in adulthood[90,91]; abdominal circumference, which is an indicator for the size of the fetal liver and the amount of subcutaneous fat deposition, has been associated with later cardiometabolic conditions[92]; and birth weight, a summary measure of *in utero* growth, is a key indicator for later life morbidity and mortality[54].

Our findings are consistent with prior studies that have found that increased prenatal $PM_{2.5}$ exposure was associated with reduced fetal ultrasound parameters, even though our estimates are not directly comparable[70–72]. One study conducted in Beijing did not assess the same fetal growth parameters, where they found that a 10 μg/m^3 increase in $PM_{2.5}$ from conception to ultrasound was associated with a 0.3 z-score reduction in estimated fetal weight; however, they did not examine associations with either HC, FL or AC, all of which were used to compute estimated fetal weight[93]. Furthermore, another study conducted in Shanghai used raw ultrasound measurements and did not standardize their parameters by gestational age, but still aggregated all ultrasound measurements together as a single outcome[71]. They found that a 10 μg/m^3 increase in

$PM_{2.5}$ from conception to ultrasound was associated with reduced BPD, FL, and AC by about 5.5 mm each. Finally, the last study conducted in Scotland internally standardized their fetal ultrasound parameters and thus their z-scores are not directly comparable to ours, but they found that $PM_{2.5}$ exposure was associated with reduced BPD[70]. Despite the heterogeneity in analytic treatments, the current body of evidence (including our contribution) suggests a robust signal for the association between prenatal $PM_{2.5}$ exposure and impairment of ultrasound parameters of fetal growth. This is further corroborated by the negative associations we found between $PM_{2.5}$ and birth weight, which is concordant with the findings from 25 out of the 29 studies included in a recent systematic review examining this association[65].

Among studies of fetal growth and particulate matter exposure (both $PM_{2.5}$ and PM_{10}), there is still inconsistency with regards to the critical window of exposure. The three studies focusing on $PM_{2.5}$ only examined long-term exposure, either entire pregnancy average[70] or the cumulative exposure from the date of ultrasound to the date of ultrasound[71,72], and did not, or were not able to, consider other exposure windows. Studies examining PM_{10} have also identified that PM_{10} exposure in the first 16 weeks of pregnancy[94], and also, in the third trimester[95,96], were associated with reduced ultrasound parameters. Here, we found that both exposure windows we considered produced negative associations, but those using cumulative $PM_{2.5}$ produced weaker associations for all outcomes (i.e., fetal ultrasound parameters, and birth weight). This pattern indicates that the effects of $PM_{2.5}$ potentially interfere with fetal tissue development in early pregnancy, rather than with the period of rapid growth in late pregnancy. Thus, it is imperative that future studies also assess exposures early in pregnancy in addition to cumulative exposures otherwise harmful exposures may be overlooked. That is, if the potential critical window is indeed in early

pregnancy, then only assessing cumulative exposures by averaging values in the critical window with other periods later in pregnancy would add measurement error, such that the estimate is biased towards the null.

In our effect modification analyses, we found that the associations did not differ by fetal sex. Although this is contrary to previous literature suggesting that stressful exposures during pregnancy may be more harmful to male births, those investigations mostly focused on acute exposures[85,97–99]. We, however, found suggestive evidence that race, education, and ADI could modify the association during specific developmental windows. Associations with fetal ultrasound parameters were stronger for individuals who were Hispanic and were null for those who were Asian. These findings however should be interpreted with caution given that the 95% CIs were wide due to few Asian and Hispanic women in our cohort (about 900 deliveries each over the 6-year study period), where future studies should examine these racial disparities in other settings with different demographic distributions. Furthermore, we also found stronger associations for individuals who were more educated and lived in areas with the least disadvantage. This finding, which was conditional on the other covariates, such as maternal age, was not expected. One possible explanation is that although less socially vulnerable individuals (e.g., those with lower ADI) are presumably, on average, healthier, they may be less resilient to the effects of $PM_{2.5}$. This explanation is consistent with findings from other health disparities work that show that minority groups in the US tend to have a lower risk of experiencing certain health outcomes, such as depression, anxiety and suicide[100]. Potential effect heterogeneity among fetal ultrasound measurements were also mostly observed in birthweight; however, these two sets of measurements are not directly comparable for several reasons. First, birth weight poorly

reflects fetal growth during early pregnancy; that is, it is hard to distinguish a healthy newborn from one that experiences fetal growth restriction early in pregnancy but later catches up to achieve population growth standards by birth. Furthermore, birth weight is a summary measure of *in utero* growth and may be more sensitive to symmetric growth restriction. Thus, its use in this setting may not be appropriate given that we show that the effect of $PM_{2.5}$ manifests in distinct effects on different organs (i.e., the effect on the fetal anatomy is asymmetric)

This study has several strengths. We assessed $PM_{2.5}$ exposure at home addresses at birth with high spatial and temporal resolution using a state-of-the-art model, where several exposure windows were considered. We used repeated ultrasound measures to longitudinally assess growth trajectories, which enabled the assessment and identification of critical windows of development. We controlled for an extensive panel of confounders, the most important being socioeconomic status, for which we had both individual and neighborhood-level variables. Furthermore, we also controlled for prenatal smoking and population density by design with restriction as our study population had few individuals who smoked during pregnancy or lived in non-urban areas. These variables could therefore not confound the exposure-outcome association.

Some limitations also should be acknowledged. Our $PM_{2.5}$ exposure assessment was based on modelled outdoor values at the home address, which may not reflect personal $PM_{2.5}$ exposure. This measurement error will bias the estimate towards the null, but this could be offset by better control for confounding as more proxy measures of exposure are less vulnerable to individual-level confounding[101]. Furthermore, we used maternal residence at delivery to assign exposures,

which could result in further misclassification due to residential mobility during pregnancy. Yet, past simulations have shown that ignoring residential mobility had only a minor impact on point estimates, and that the identification of critical windows was robust to this type of exposure misclassification[102]. We also ignored the time-activity pattern of participants during pregnancy, as we used exposure grids where the participant was reported to have resided at delivery. Yet participants are less likely to spend time at home during early pregnancy[103]. However, time-activity patterns are unlikely to be related to fetal growth conditional on the covariates included, and thus the bias is expected to be nondifferential and towards the null. Furthermore, our analyses were restricted to live-born children and so our estimates may be biased upwards[3,4,104,105]. That is, if pregnancy loss is driven by both the exposure and other unmeasured factors that also affect the outcome, then the $PM_{2.5}$-growth associations estimated in the subpopulation of "healthier" live births are likely biased. These considerations suggests that perhaps the associations are even stronger than they are shown here, which further supports the need to reduce $PM_{2.5}$ exposure in early to mid-pregnancy. Finally, this study was retrospective, and so analyses of growth scans should be interpreted with caution since high-risk pregnancies could be overrepresented (i.e., low-risk pregnancies with normal fetal growth are unlikely to have third-trimester pregnancy scans).

In conclusion, we show that gestational exposure to $PM_{2.5}$ was associated with fetal ultrasound parameters and birth weight at levels below the current national standards[86], where early to mid-pregnancy appears to be a critical window of exposure. These findings add to the growing body of literature documenting the harmful effects of $PM_{2.5}$ not only during pregnancy, but also for overall health. Future work should explore this topic further in other settings and different

populations, at exposure levels higher than this study, while also considering early pregnancy as a critical window of exposure. Clinicians should keep up-to-date with emerging research to inform pregnant women about the potential adverse effects of $PM_{2.5}$ exposure. In terms of policy implications, our findings suggest the need to focus efforts on reducing exposures even at "safe" concentrations.

Chapter 3: Temperature during pregnancy and fetal growth in Massachusetts, USA

Michael Leung[1], Francine Laden[1,2,3], Brent Coull[4], Anna M. Modest[5,6], Michele R. Hacker[1,5,6], Blair J. Wylie[5,6], Yaguang Wei[2], Joel Schwartz[1,2], Stefania Papatheodorou[1], Marc G. Weisskopf[1,2]

[1]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA

[2]Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA

[3]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA

[4]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA

[5]Department of Obstetrics and Gynecology, Beth Israel Deaconess Medical Center, Boston, MA

[6]Department of Obstetrics, Gynecology and Reproductive Biology, Harvard Medical School, Boston, MA

ABSTRACT

Background: No prior study has examined the association between exposure to ambient temperature during pregnancy and fetal ultrasound parameters.

Objectives: In this study, we examined the association between gestational exposure to ambient temperature and fetal growth outcomes (ultrasound biometric parameters and birth weight) in a pregnancy cohort using spatiotemporally resolved temperature in Massachusetts, USA.

Methods: We used ultrasound measurements of biparietal diameter (BPD), head circumference (HC), femur length (FL), and abdominal circumference (AC), in addition to birth weight from 10,008 pregnancies that delivered at Beth Israel Deaconess Medical Center from 2011-2016. We fitted linear mixed effects models to examine the associations of temperature in the first 16 weeks of pregnancy with anatomic ultrasound scans (<24 weeks), growth ultrasound scans (≥24 weeks), and birth weight. All models were adjusted for sociodemographic characteristics, maternal comorbidities, long-term trends, $PM_{2.5}$, and humidity.

Results: Higher temperatures in the first 16 weeks were associated with reductions in all fetal growth outcomes, where associations were particularly strong for head size parameters. For example, a 5 ^{o}C in temperature was associated with a BPD and HC z-score reductions of -0.20 and -0.22, respectively, before 24 weeks, and z-score reductions of -0.20 and -0.14, respectively, after 24 weeks. Our secondary analyses examining the associations with long-term average temperature, and average temperature 1 month prior to growth assessment produced attenuated associations, except for birth weight which was more strongly associated with long-term average temperature.

Conclusions: Higher temperatures were associated with impaired fetal growth, with head size parameters being particularly sensitive

INTRODUCTION

The climate crisis has led to warmer global temperatures, where eight of the ten hottest years on record have occurred in the past decade[106–108]. Apart from having been recognized as an environmental emergency, several organizations including the World Health Organization (WHO) and the Lancet Countdown have identified human health as one of its major consequences[109,110]. Pregnant individuals and their fetuses have been recognized as one group that is particularly vulnerable to heat stress[111,112]. The physiological and anatomical changes that occur during pregnancy (e.g., increased internal heat production with fetal and placental metabolism) present challenges to thermoregulation. The impaired ability to respond to high temperatures can result in cell death, disturbance of cell migration, disruption of gene expression, and damage to blood vessels and the placenta[113–117].

Two recent systematic reviews have shown that heat exposure during pregnancy is associated with lower birth weight[65,118], which has implications for subsequent health and development[54,119,120]. However, the use of newborn anthropometry to assess fetal growth provides limited insights on the timing of when the growth-restricting effects of heat manifest during pregnancy. Routine ultrasound measurements would make the ontogenetic processes of fetal development observable, and thus would be an invaluable tool for identifying developmental windows during which a fetus is susceptible to heat stress. Yet, no study, to date, has examined the associations between heat exposure and ultrasound parameters of fetal growth. Thus, in a pregnancy cohort with routine ultrasound and spatiotemporally resolved temperature from Massachusetts, USA, we aimed to examine the association between ambient temperature during pregnancy and fetal growth.

METHODS

Study population

This study population has been described previously[121]. Briefly, we used prenatal and obstetric data from Beth Israel Deaconess Medical Center (BIDMC). Pregnant individuals who delivered at ≥20 weeks of gestation from 2011 to 2016 by the practices in which all obstetric ultrasounds are performed through BIDMIC were eligible (n=12,967). Of these, we excluded pregnant individuals with multifetal gestations because the growth trajectory is different than singleton pregnancies (n=844) and residential addresses outside of Massachusetts (n=408). We further excluded those without any ultrasound measurements (n=1,707), leading to a final sample size of 10,008 participants (Figure S1). Full addresses were available for each delivery and were geocoded to latitude and longitude using the Google Maps Application Programing Interface. This study was approved by the institutional review boards of Harvard T.H. Chan School of Public Health and Beth Israel Deaconess Medical Center.

Temperature

To assess temperature, we downloaded daily temperature data on a 12 km grid from Phase 2 of the North American Land Data Assimilation Systems (NLDAS-2) at the NASA Earth Sciences Data and Information Services Center[122]. These data were derived from the 32 km grids of the National Centers for Environmental Prediction (NCEP) North American Regional Reanalysis (NARR). The 32 km NARR grids were spatially interpolated to the finer resolution of the NLDAS-2 1/8-degree grid (approximately 12 km). This spatial interpolation process has been described in more detail elsewhere[123].

For each delivery, we assigned the 12 km grid in which the pregnant individual reported to have resided at the time of birth. After grid assignment, we estimated the average temperature in the first 16 weeks of pregnancy, as this has been previously reported to be a critical window of development[78,121]. As a secondary analysis, we also considered two other exposure windows: 1) long-term average temperature exposure from conception up until the assessment of fetal growth; that is, from conception to the ultrasound for fetal ultrasound parameters, and from conception to birth for birth weight, and 2) average temperature exposure 1 month prior to growth assessment.

Fetal ultrasound

Ultrasound scans at about 18-23 weeks are typically conducted to evaluate fetal anatomy, while scans conducted later are typically done to screen for pregnancy complications (e.g., impaired growth, obstetric indications such as breech presentation etc.). Thus, the ultrasound data were categorized into two types of scans for the analyses: the "anatomic scan" prior to 24 weeks, and the "growth scan" at 24 weeks and beyond. To assess impairment of fetal growth in each of these windows, we used four ultrasound biometric parameters – biparietal diameter (BPD), head circumference (HC), femur length (FL) and abdominal circumference (AC) – all of which were recorded in millimeters (mm). The gestational age at the ultrasound examination was based on the best obstetric estimate combining information from the last menstrual period and the earliest ultrasound performed in pregnancy[79].

For a given gestational week, we considered ultrasound measurements 4 standard deviations away from the mean of the cohort implausible, and thus excluded these measurements. Furthermore, to enable comparisons across gestational weeks, we generated z-scores for each of

the fetal ultrasound parameters by applying the INTERGROWTH-21st standards for fetal growth[80]. Since these standards are only available up to 40 weeks of gestation, ultrasound scans conducted after 40 weeks were excluded. Finally, we also abstracted birth weight from the medical records, which was reported in grams, and generated age- and sex-specific z-scores using the INTERGROWTH-21st standards for newborn size[80].

Covariates

Data on specific humidity, defined as the mass of water vapor in a unit mass of moist air, were obtained from NASA NLDAS-2 with 12 km spatial resolution[122]. From the medical records, we abstracted data on the following maternal and fetal characteristics: maternal age (continuous), race (White, Black, Asian, Hispanic, or Other), educational attainment (college or higher or less than college), insurance type (private or public/uninsured), parity (nulliparous or parous), and fetal sex (male or female). Furthermore, for area-level socioeconomic status, we used the national percentile rankings of the Area Deprivation Index (ADI), which ranges from 1 to 100, with 1 being the least disadvantaged and 100 being the most[82]. The ADI was calculated at the census block group level and represents a composite measure of neighborhood socioeconomic disadvantage derived from 17 census variables on income, employment, and housing from the American Community Survey[82]. Finally, for comparability with prior studies, we also included exposure to particulate matter less than 2.5 microns ($PM_{2.5}$) in our models. The $PM_{2.5}$ data were assigned based on where the pregnant individual resided at birth in Massachusetts from a state-of-the-art spatiotemporal model that predicts daily $PM_{2.5}$ concentration for each 1 km grid across the Continental United States[74].

Statistical Analysis

We fitted generalized additive mixed models (GAMMs) to examine the relationship of temperature in the first 16 weeks with anatomic ultrasound scans (<24 weeks), growth ultrasound scans (≥24 weeks), and birth weight. We used a penalized spline for temperature to allow for a nonlinear exposure-response function in each window[124]. To adjust for confounding by long-term and seasonal trends, we included conception year and day of the year, where the latter was modeled using natural splines with 4 degrees of freedom per year. All other covariates previously described were included in the model, with linear and quadratic terms used for continuous variables. Furthermore, we included a random intercept for each pregnancy because ultrasounds within each pregnancy were likely to be correlated. Since our GAMMs showed all associations between temperature and fetal ultrasound parameters to be approximately linear (Figures S2-3), we present estimates from linear mixed effects models.

We conducted several additional analyses. First, we examined the associations between each outcome and the other two exposure windows we considered: 1) long-term temperature exposure, defined as the average temperature exposure from conception to ultrasound for fetal ultrasound parameters, and whole pregnancy average temperature exposure for birth weight, and 2) average temperature exposure 1 month prior to growth assessment. Furthermore, we also conducted several subgroup analyses to assess for potential effect modification, focusing on effect modifiers reported in prior studies. Specifically, prior work has shown that the growth-restricting effect of heat exposure is larger in pregnant individuals at the age extremes (<22 or >40 years)[125–127]; those who were Black or Hispanic[126,127]; and those of low socioeconomic status[125–127]. We also considered potential effect modification by fetal sex as other birth

outcomes research has shown that prenatal exposures may be more harmful to male fetuses[85,97–99]. Thus, for each type of scan, we assessed whether the association of temperature in the first 16 weeks and fetal growth was separately modified by maternal age (<22, 22-40, and >40 years), race, education, ADI (divided into quartiles), and fetal sex. To do so, we included a product term between temperature and the modifier. For each potential modifier, we tested whether the association differed from that of the reference level, which was "22-40 years" for maternal age, "White" for race, "less than college" for education, the first quartile for ADI, and "female" for fetal sex. We used a Bonferroni correction to account for multiple testing when examining modifiers with more than two levels. All analyses were performed in R (version 3.6.1; R Foundation for Statistical Computing, Vienna, Austria)[17].

RESULTS

Characteristics of the 10,008 pregnancies that constituted our study population are shown in Table 1. On average, participants were 31 years of age at conception, with the majority being white (52%) and having private insurance (78%). About half had completed college or higher (52%), and about half were nulliparous (49%). Furthermore, the mean ADI percentile was 22 (median nationwide is 50), which indicates that our cohort was comprised of individuals that lived in neighborhoods with less disadvantage relative to the rest of the US. In our sample, average temperature in the first 16 weeks of pregnancy was 10º C (Table 1 and Figure S2). Peak temperature exposures in the first 16 weeks were approximately 20 º C, which was experienced by pregnancies that were conceived in the summer of each year. On the other hand, the coldest temperatures varied throughout the study period, where conceptions beginning in the winters of 2010, 2013, and 2015 experienced average temperatures below 0 º C during the first 16 weeks of

pregnancy, and those that occurred in the winters of other years experienced milder conditions (Figure S4). Average specific humidity in the first 16 weeks of pregnancy was 0.0068 g/kg, where the pattern over the study period closely resembled that of temperature (Figure S4). Most pregnancies had at least 2 ultrasound scans (73%) (Table 1). Summary statistics for the four ultrasound parameters are displayed in Table 2. On average, fetuses in our sample had smaller BPD, but larger HC, FL and AC compared to the international standard[80]. Furthermore, newborns in our sample had an average birth weight z-score of 0.33, which is slightly heavier than the international norm (Table 2).

The associations between temperature in the first 16 weeks of pregnancy and fetal growth outcomes (ultrasound parameters and birth weight) are shown in Table 3. For both anatomic and growth scans (i.e., scans performed <24 weeks and ≥24 weeks, respectively), we observed that increased temperature was linearly associated with reduced ultrasound parameters, where associations were particularly strong for head size parameters (BPD and HC). For example, a 5 ºC increase in temperature in the first 16 weeks was associated with z-score reductions among anatomic scans of -0.20 (95% CI: -0.34, -0.06) for BPD, -0.22 (95% CI: -0.36, -0.09) for HC, -0.15 (95% CI: -0.30, 0.00) for FL, and -0.15 (95% CI: -0.30, 0.00) for AC, and z-score reductions among growth scans of -0.20 (95% CI: -0.32, -0.09) for BPD, -0.14 (95% CI: -0.26, -0.02) for HC, -0.01 (95% CI: -0.13, 0.10) for FL, and -0.17 (95% CI: -0.29, -0.06) for AC. Furthermore, temperature exposure in the first 16 weeks was also associated with reduced birth weight, with a 5 ºC increase associated with a z-score reduction of -0.11 (95% CI: -0.19, -0.02). Analyses with the two other exposure windows produced associations that were attenuated

relative to the effect estimates for the 16-week exposures, with most 95% CIs including null. However, the estimates for birth weight were comparable to our primary analysis (Tables S1-2).

Results from our effect modification analyses can be found in Tables S3-7. We found that maternal age and race potentially modified the associations during specific windows. For maternal age, pregnant individuals <22 years had stronger negative associations with AC among growth scans compared to those 22-40 years, while associations with FL among growth scans appear slightly positive for individuals >40 years (Table S3). For maternal race, Black individuals had stronger negative associations with HC among growth scans compared to White individuals (Table S4). Furthermore, Asian individuals had stronger negative associations with HC among growth scans, AC among growth scans, and birth weight, compared to their White counterparts (Table S4). Finally, Hispanic individuals had stronger negative associations with BPD among growth scans compared to White individuals (Table S5). We did not find any evidence for effect modification by education, ADI, or fetal sex (Tables S6-7).

Table 3.1. Maternal and fetal characteristics of deliveries at Beth Israel Deaconess Medical Center, Boston, Massachusetts from 2011-2016 (N=10,008)

Characteristics	**N (%)**
Age (years)	
Mean (SD)	31 (5.4)
Education	
College or higher	4769 (48)
Lower than college	3416 (34)
Not specified	1823 (18)
Race	
White	5117 (51)
Black	1750 (17)
Asian	940 (9)
Hispanic	931 (9)
Not specified	1270 (13)
Parity	
Nulliparous	4941 (49)
Parous	5067 (51)
Child sex	
Female	4927 (49)
Male	5081 (51)
Insurance	
Private	7797 (78)
Public or uninsured	2211 (22)
Area Deprivation Index (percentile)	
Mean (SD)	22 (20)
Average temperature in first 16 weeks of pregnancy (º C)	
Mean (SD)	10 (7.5)
Average specific humidity in first 16 weeks of pregnancy (g/kg)	
Mean (SD)	0.0068 (0.003)
Average $PM_{2.5}$ in first 16 weeks of pregnancy ($\mu g/m^3$)	
Mean (SD)	7.4 (1.5)
Number of ultrasounds	
2+ ultrasounds	7320 (73)
1 ultrasound	2688 (27)

Table 3.2. Summary statistics for fetal ultrasound parameters and birth weight from deliveries at Beth Israel Deaconess Medical Center, Boston, Massachusetts in 2011-2016 (N=10,008)

Growth outcome	N	Mean (SD)
Anatomic scans (<24 weeks)		
BPD z-score	10,207	-0.67 (1.10)
HC z-score	10,164	0.37 (1.10)
FL z-score	10,143	0.82 (1.14)
AC z-score	10,112	0.52 (1.17)
Growth scans (≥24 weeks)		
BPD z-score	15,944	-0.86 (1.12)
HC z-score	15,874	0.33 (1.15)
FL z-score	15,892	1.11 (1.10)
AC z-score	15,866	0.56 (1.10)
Birth weight z-score	9,991	0.33 (1.00)

Table 3.3. Linear mixed effects model estimates and 95% CIs for the association between temperature during the first 16 weeks of pregnancy and fetal growth outcomes (ultrasound parameters and birth weight) from deliveries at Beth Israel Deaconess Medical Center, Boston, Massachusetts in 2011-2016 (N=10,008)

Growth outcome	Estimate per 5 ºC	95% CI	*p*-Value
Anatomic scans (<24 weeks)			
BPD z-score	-0.20	(-0.34, -0.06)	0.006
HC z-score	-0.22	(-0.36, -0.09)	0.001
FL z-score	-0.15	(-0.30, 0.00)	0.05
AC z-score	-0.15	(-0.30, 0.00)	0.05
Growth scans (≥24 weeks)			
BPD z-score	-0.20	(-0.32, -0.09)	0.001
HC z-score	-0.14	(-0.26, -0.02)	0.02
FL z-score	-0.01	(-0.13, 0.10)	0.82
AC z-score	-0.17	(-0.29, -0.06)	0.004
Birth weight			
Birth weight z-score	-0.11	(-0.19, -0.02)	0.01

Note: Adjusted for maternal age, race, education, insurance type, parity, fetal sex, conception year, day of the year of conception, Area Deprivation Index, humidity, and $PM_{2.5}$.

DISCUSSION

In this large retrospective pregnancy cohort from Massachusetts, USA, we found that increased temperature was linearly associated with reductions in all four ultrasound parameters - BPD, HC, FL and AC - which has implications for subsequent health and development, as each of these ultrasound parameters has been associated with later health outcomes[87–92]. Negative associations with birth weight were also observed, which is concordant with the published literature[65,118]. Although head size parameters BPD and HC appear particularly sensitive to heat exposure, the associations with other ultrasound parameters and birth weight are still important for public health given that exposure to high temperatures is common and escalating[106–108].

Several studies have explored the question of critical exposure windows in fetal growth using birth weight as an outcome, but this body of work has produced results that are inconclusive[118]. In this study, we found that the associations with ultrasound parameters were stronger for temperature exposure in the first 16 weeks of pregnancy compared to long-term temperature exposure (i.e., from conception to the time of fetal growth assessment) and exposure 1 month prior to growth assessment, which indicates that exposure to heat potentially interferes with early fetal development rather than growth later in pregnancy. One possible mechanism is that increased heat exposure leads to the production of heat-shock proteins, which can disrupt normal protein synthesis in early pregnancy, leading to altered fetal organ development[128]. Given the strong associations we see for head size parameters, the ontogenetic processes of early brain development (e.g., neurogenesis, axonal and dendritic growth, synaptogenesis etc.) appear to be particularly vulnerable, where any perturbation in these processes (which occur in a specific order) could potentially have long-term effects on brain development[129].

We also found that birth weight was associated with temperature in the first 16 weeks and long-term average temperature, but not exposure 1-month prior to growth assessment, where associations were strongest when using long-term average temperature. However, our results with birth weight are not directly comparable to our ultrasound results for several reasons. First, we expect the critical window for birth weight to be different from fetal ultrasound parameters, as the biological mechanism underpinning the association is different. With birth weight, the effects of heat mainly relate to fat accumulation which occurs throughout pregnancy (as opposed to organ development which occurs early to mid-pregnancy). Suboptimal thermoregulation in response to rising temperatures could divert too much blood away from the developing fetus, which could lead to lower birth weight by depriving the fetus of adequate nutrition[130]. Second, birth weight is a summary measure of *in utero* growth and may be more sensitive to either symmetric growth restriction or growth restriction related to larger organs (e.g., a reduction in just head size does not necessarily translate to a reduction in birth weight).

One unexpected finding was that cold temperatures did not seem to have an impact on fetal growth in our cohort, where temperature effects on health have often been shown to be either U- or J-shaped[118,131]. This may be because our cohort is comprised of highly educated individuals who live in urban neighborhoods with little disadvantage, such that they may be better at mitigating the effects of extreme cold (e.g., heating, adequate clothing, staying indoors etc.). Furthermore, colder winters in New England are shorter and are becoming less common[132], where only 9% of the study participants over the six-year study period experienced averaged temperature below freezing. Thus, although there may be a true effect of cold temperatures on

fetal growth, the relatively mild climatic conditions of Massachusetts in recent years (i.e., the limited sample size at temperature extremes) may not be adequate to detect any signal that would exist in the data. These considerations suggest that perhaps the linear dose-repose relationship we observed may not be generalizable to other settings with different temperature distributions and seasonality patterns.

In our effect modification analyses, we found that the associations did not differ by educational attainment or ADI. One possible explanation, akin to the rationale for the lack of effect for cold temperatures, is that the socioeconomic distribution of our cohort falls within a narrow range, such that between-group comparisons did not yield any differences. Past studies that found differences were able to do so because, perhaps, they included individuals from a broader range of socioeconomic contexts. For example, a study using birth certificates in California in which individuals from lower socioeconomic groups were better represented in the analytic sample found differences by educational attainment[127]. Furthermore, we did not find differences by fetal sex which is contrary to some of the previous literature[85,97–99], but coincides with our previous report on gestational exposures to $PM_{2.5}$ which showed that male fetuses were equally vulnerable to environmental stressors compared to their female counterparts[121]. We did, however, find that maternal age could potentially modify the association between heat stress and impaired fetal growth. Negative associations were stronger for pregnant individuals who were <22 years, which coincides with previous literature suggesting that younger mothers are at an increased risk of adverse birth outcomes due to a constellation of potential reasons, including biological immaturity, socioeconomic disadvantages, behavioral factors, and lack of access to high quality antenatal care[133]. Furthermore, it is notable that we found differences by race (despite controlling

for socioeconomic status), where Black, Asian, and Hispanic individuals appeared to be more vulnerable to the effects of heat than White individuals. One possible reason according to recent research is that non-White individuals are more likely to experience more intense urban heat islands[134], a phenomenon which our 12 km exposure grids may not be able to capture. Further research is warranted using temperature data with finer spatial resolution in other settings with different demographic and socioeconomic compositions to better understand the intersectionality between race/ethnicity and socioeconomic status on the impact of rising temperatures during pregnancy.

To our knowledge, this is the first study to use fetal ultrasound parameters to investigate the growth-restricting effect of heat, which is a key strength of our analyses. Other strengths include the use of temperature data with high spatial and temporal resolution, where three exposure windows were considered. Furthermore, our adjustment for confounding variables was comprehensive, where we controlled for individual, area-level socioeconomic, and meteorological variables in addition to long-term time trends. However, we also acknowledge several limitations. Although the resolution of the temperature data is reported to be 12 km, it is, in reality, coarser because the spatial interpolation process of the NARR 32 km grid may not capture urban-scale features (e.g., urban heat islands), which may be relevant for human health[135], especially in the context of racial disparities. Furthermore, there were additional sources of measurement error with regards to exposure assignment: 1) heat exposure was based on outdoor values, and so may not reflect personal exposure, and 2) assignment was based on residential address at delivery, and so is agnostic towards both residential mobility and time-activity patterns during pregnancy. However, these errors are likely nondifferential with respect

to fetal growth, and so the expected direction of the bias is towards the null. Our estimates are also vulnerable to live birth bias since our analyses were restricted to live-born children[3,4,104,105]. That is, if heat exposure also causes pregnancy loss, then our associations are likely biased upwards since fetuses who are more susceptible to the effects of heat do not survive to birth. This suggests that the estimates we present here are conservative and that the growth-restricting effects of heat could possibly be even stronger. Finally, our study was based on retrospective ultrasound data, and so estimates from the analyses of growth scans (i.e., scans ≥24 weeks) should be interpreted with caution as they might have been conducted among individuals at higher risk for pregnancy complications.

In conclusion, we show that heat exposure during pregnancy was associated with reductions in all fetal ultrasound parameters and birth weight. Head size parameters were particularly sensitive, where early to mid-pregnancy appears to be a critical exposure window. Future research should explore this topic further in other settings with different temperature distributions/seasonality, different demographic/socioeconomic compositions, and using temperature data with finer spatial resolution. Long-term follow-up studies of neurodevelopment in children are warranted to examine the practical implications of our findings. These novel findings contribute to the growing body of evidence documenting the overall health impact of rising temperatures, which further highlights the importance of investment into preventive measures for pregnant individuals, heat warning systems, and more broadly, advocacy for regulations to mitigate the climate crisis.

Chapter 4: Conclusion

This dissertation expands our knowledge about pregnancy as a vulnerable period for health. In Chapter 1, we show that estimating the effects of prenatal exposures on health outcomes in live-born children is a challenge when the exposure under study affects pregnancy loss, and if fetal survival and the outcome of interest share common causes. Analyses restricted to live births can lead to associations that are biased downwards, which can produce a body of evidence that is inconsistent – associations with harmful exposures may appear less harmful, null, or even paradoxically protective. This finding implies that not accounting for live-birth bias can limit the identification of harmful prenatal exposures and prevent the development of interventions aimed at promoting better health for pregnant individuals and their fetuses. Despite important theoretical and methodological advances, further investment into collecting the necessary data to mitigate this bias is required if we are to make advances in this field. Specifically, we need better study designs that can capture conceptions or viable fetuses at the relevant exposure window and/or measurements that can capture information on possible selection-outcome confounders.

In Chapters 2 and 3, we found that increased $PM_{2.5}$ and ambient temperature during pregnancy was associated with reduced fetal ultrasound parameters and birth weight in a pregnancy cohort from Massachusetts, USA. For both environmental exposures, we found that early to mid-pregnancy appears to be a critical exposure window which indicates that these exposures primarily interfere with early tissue development rather than with the period of rapid growth late in pregnancy. Notably, Massachusetts is a low pollution environment where $PM_{2.5}$ concentrations were below the current national standards, and so in terms of policy implications, these findings suggest the need to focus efforts on reducing $PM_{2.5}$ exposures at even

concentrations considered to be safe for human health. With ambient temperature, we found that fetal head size was particularly sensitive to the effects of heat, and so it would be interesting to explore whether this affects later neurodevelopment in future long-term follow-up studies of these children. Given that exposure to high temperatures is common and escalating, the importance of advocacy for regulations to mitigate the climate crisis cannot be understated. Overall, this dissertation aims to unbox the black box of fetal health during pregnancy and highlights that studies which exclusively focus on health outcomes among live births may miss *in utero* events that are relevant for health.

References

1. Howards P, Hertz-Picciotto I, Poole C. Conditions for bias from differential left truncation. *Am J Epidemiol*. 2006;165(4):444-452.
2. Lisonkova S, Joseph KS. Left truncation bias as a potential explanation for the protective effect of smoking on preeclampsia. *Epidemiology*. 2015;26(3):436-440.
3. Raz R, Kioumourtzoglou MA, Weisskopf MG. Live-Birth Bias and Observed Associations between Air Pollution and Autism. *Am J Epidemiol*. 2018;187(11):2292-2296.
4. 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(1):345-354.
5. Raz R, Levine H, Pinto O, Broday DM, Yuval, Weisskopf MG. Traffic-Related Air Pollution and Autism Spectrum Disorder: A Population-Based Nested Case-Control Study in Israel. *Am J Epidemiol*. 2018;187(4):717-725.
6. Hernan M, Robins J. *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC.; 2020.
7. Hernández-Díaz S, Schisterman EF, Hernán MA. The birth weight "paradox" uncovered? *Am J Epidemiol*. 2006;164(11):1115-1120.
8. Pelch KE, Bolden AL, Kwiatkowski CF. Environmental chemicals and autism: A scoping review of the human and animal research. *Environ Health Perspect*. 2019;127(4):046001-1-046001-046012.
9. Jensen TK, Andersen LB, Kyhl HB, Nielsen F, Christesen HT, Grandjean P. Association between perfluorinated compound exposure and miscarriage in Danish pregnant women.

PLoS One. 2015;10(4).

10. Krieg SA, Shahine LK, Lathi RB. Environmental exposure to endocrine-disrupting chemicals and miscarriage. *Fertil Steril*. 2016;106(4):941-947.
11. Kalkbrenner AE, Schmidt RJ, Penlesky AC. Environmental chemical exposures and autism spectrum disorders: a review of the epidemiological evidence. *Curr Probl Pediatr Adolesc Health Care*. 2014;44(10):277-318.
12. Sperrin M, Candlish J, Badrick E, Renehan A, Buchan I. Collider bias is only a partial explanation for the obesity paradox. *Epidemiology*. 2016;27(4):525-530.
13. Glymour MM, Vittinghoff E. Selection bias as an explanation for the obesity paradox: Just because it's possible doesn't mean it's plausible. *Epidemiology*. 2014;25(1):4-6.
14. Lajous M, Banack HR, Kaufman JS, Hernán MA. Should patients with chronic disease be told to gain weight? the obesity paradox and selection bias. *Am J Med*. 2015;128(4):334-336.
15. Howe CJ, Cole SR, Lau B, Napravnik S, Eron JJ. Selection Bias Due to Loss to Follow Up in Cohort Studies. *Epidemiology*. 2016;27(1):91-97.
16. Wilcox AJ, Weinberg CR, O'connor JF, et al. Incidence of Early Loss of Pregnancy. *N Engl J Med*. 1988;319(4):189-194.
17. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austra. Published 2014. http://www.r-project.org/
18. Munafò MR, Tilling K, Taylor AE, Evans DM, Davey Smith G. Collider scope: when selection bias can substantially influence observed associations. *Int J Epidemiol*. 2018;47(1):226-235.
19. Greenland S. Quantifying biases in causal models: Classical confounding vs collider-

stratification bias. *Epidemiology*. 2003;14(3):300-306.

20. Smith LH, Vanderweele TJ. Bounding Bias Due to Selection. *Epidemiology*. 2019;30(4):509-516.
21. Walker A. *Observation and Inference*. Epidemiology Resources Incorporated; 1991.
22. Brunton PJ. Effects of maternal exposure to social stress during pregnancy: Consequences for mother and offspring. *Reproduction*. 2013;146(5):R175-89.
23. Plana-Ripoll O, Parner E, Olsen J, Li J. Severe stress following bereavement during pregnancy and risk of pregnancy loss: Results from a population-based cohort study. *J Epidemiol Community Health*. 2015;70(5):424-429.
24. Bercum FM, Rodgers KM, Benison AM, et al. Maternal stress combined with terbutaline leads to comorbid autistic-like behavior and epilepsy in a rat model. *J Neurosci*. 2015;35(48):15894-15902.
25. Walder DJ, Laplante DP, Sousa-Pires A, Veru F, Brunet A, King S. Prenatal maternal stress predicts autism traits in 61/2 year-old children: Project Ice Storm. *Psychiatry Res*. 2014;219(2):353-360.
26. Roberts AL, Lyall K, Rich-Edwards JW, Ascherio A, Weisskopf MG. Association of maternal exposure to childhood abuse with elevated risk for autism in offspring. *JAMA Psychiatry*. 2013;70(5):508-515.
27. Roberts AL, Koenen KC, Lyall K, Ascherio A, Weisskopf MG. Women's posttraumatic stress symptoms and autism spectrum disorder in their children. *Res Autism Spectr Disord*. 2014;8(6):608-616.
28. Beversdorf DQ, Manning SE, Hillier A, et al. Timing of prenatal stressors and autism. *J Autism Dev Disord*. 2005;35(4):471-478.

29. Kinney DK, Munir KM, Crowley DJ, Miller AM. Prenatal stress and risk for autism. *Neurosci Biobehav Rev*. 2008;32(8):1519-1532.

30. Kinney DK, Miller AM, Crowley DJ, Huang E, Gerber E. Autism prevalence following prenatal exposure to hurricanes and tropical storms in Louisiana. *J Autism Dev Disord*. 2008;38(3):481-488.

31. Coughlan C, Walters S, Ledger W, Li TC. A comparison of psychological stress among women with and without reproductive failure. *Int J Gynecol Obstet*. 2014;124(2):143-147.

32. Dean RG, Dean J, Heller GZ, Leader LR. A mass shooting at Port Arthur, Tasmania, Australia: a study of its impact on early pregnancy losses using a conception time-based methodology. *Hum Reprod*. 2015;30(11):2671-2676.

33. Nepomnaschy PA, Welch KB, McConnell DS, Low BS, Strassmann BI, England BG. Cortisol levels and very early pregnancy loss in humans. *Proc Natl Acad Sci U S A*. 2006;103(10):3938-3942.

34. Wainstock T, Lerner-Geva L, Glasser S, Shoham-Vardi I, Anteby EY. Prenatal stress and risk of spontaneous abortion. *Psychosom Med*. 2013;75(3):228-235.

35. Wisborg K, Barklin A, Hedegaard M, Henriksen TB. Psychological stress during pregnancy and stillbirth: Prospective study. *BJOG An Int J Obstet Gynaecol*. 2008;115(7):882-885.

36. Li W, Newell-Price J, Jones GL, Ledger WL, Li TC. Relationship between psychological stress and recurrent miscarriage. *Reprod Biomed Online*. 2012;25(2):180-189.

37. Bruckner TA, Mortensen LH, Catalano RA. Spontaneous Pregnancy Loss in Denmark Following Economic Downturns. *Am J Epidemiol*. 2016;183(8):701-708.

38. Class QA, Abel KM, Khashan AS, et al. Offspring psychopathology following

preconception, prenatal and postnatal maternal bereavement stress. *Psychol Med*. 2014;44(1):71-84.

39. Jung Y, Lee AM, McKee SA, Picciotto MR. Maternal smoking and autism spectrum disorder: Meta-analysis with population smoking metrics as moderators. *Sci Rep*. 2017;7(1).

40. Pineles BL, Park E, Samet JM. Systematic review and meta-analysis of miscarriage and maternal exposure to tobacco smoke during pregnancy. *Am J Epidemiol*. 2014;179(7):807-823.

41. Caramaschi D, Taylor AE, Richmond RC, et al. Maternal smoking during pregnancy and autism: using causal inference methods in a birth cohort study. *Transl Psychiatry*. 2018;8(1):262.

42. Marufu TC, Ahankari A, Coleman T, Lewis S. Maternal smoking and the risk of still birth: systematic review and meta-analysis. *BMC Public Health*. 2015;15(1):239.

43. Page JM, Silver RM. Genetic causes of recurrent pregnancy loss. *Clin Obstet Gynecol*. 2016;59(3):498-508.

44. Risch N, Spiker D, Lotspeich L, et al. A genomic screen of autism: Evidence for a multilocus etiology. *Am J Hum Genet*. 1999;65(2):493-507.

45. Grove J, Ripke S, Als TD, et al. Identification of common genetic risk variants for autism spectrum disorder. *Nat Genet*. 2019;51(3):431-444.

46. Tsukimori K, Tokunaga S, Shibata S, et al. Long-term effects of polychlorinated biphenyls and dioxins on pregnancy outcomes in women affected by the Yusho incident. *Environ Health Perspect*. 2008;116(5):626-630.

47. Bernardo BA, Lanphear BP, Venners SA, et al. Assessing the relation between plasma

PCB concentrations and elevated autistic behaviours using bayesian predictive odds ratios. *Int J Environ Res Public Health*. 2019;16(3).

48. Grippo A, Zhang J, Chu L, et al. Air pollution exposure during pregnancy and spontaneous abortion and stillbirth. *Rev Environ Health*. 2018;33(3):247-264.
49. Finer LB, Zolna MR. Declines in unintended pregnancy in the United States, 2008-2011. *N Engl J Med*. 2016;374(9):843-852.
50. Lam J, Sutton P, Kalkbrenner A, et al. A systematic review and meta-analysis of multiple airborne pollutants and autism spectrum disorder. *PLoS One*. 2016;11(9).
51. Yang C, Zhao W, Deng K, Zhou V, Zhou X, Hou Y. The association between air pollutants and autism spectrum disorders. *Environ Sci Pollut Res*. 2017;24(19):15949-15958.
52. Flores-Pajot MC, Ofner M, Do MT, Lavigne E, Villeneuve PJ. Childhood autism spectrum disorders and exposure to nitrogen dioxide, and particulate matter air pollution: A review and meta-analysis. *Environ Res*. 2016;151:763-776.
53. Weisskopf MG, Kioumourtzoglou MA, Roberts AL. Air Pollution and Autism Spectrum Disorders: Causal or Confounded? *Curr Environ Heal reports*. 2015;2(4):430-439.
54. Barker DJP. The developmental origins of adult disease. *J Am Coll Nutr*. 2004;23(6 Suppl):588S-595S.
55. Hertz-Picciotto I, Park HY, Dostal M, Kocan A, Trnovec T, Sram R. Prenatal exposures to persistent and non-persistent organic compounds and effects on immune system development. In: *Basic and Clinical Pharmacology and Toxicology*. Vol 102. ; 2008:146-154.
56. Fei C, Olsen J. Prenatal exposure to perfluorinated chemicals and behavioral or

coordination problems at age 7 years. *Environ Health Perspect*. 2011;119(4):573-578.

57. Liew Z, Ritz B, von Ehrenstein OS, et al. Attention deficit/hyperactivity disorder and childhood autism in association with prenatal exposure to perfluoroalkyl substances: A nested case–control study in the Danish National Birth Cohort. *Environ Health Perspect*. 2015;123(4):367-373.

58. Ode A, Källén K, Gustafsson P, et al. Fetal exposure to perfluorinated compounds and attention deficit hyperactivity disorder in childhood. *PLoS One*. 2014;9(4):e95891.

59. Stein CR, Savitz DA, Bellinger DC. Perfluorooctanoate and neuropsychological outcomes in children. *Epidemiology*. 2013;24(4):590-599.

60. Kioumourtzoglou MA, Raz R, Wilson A, et al. Traffic-related Air Pollution and Pregnancy Loss. *Epidemiology*. 2019;30(1):4-10.

61. Burnett R, Chen H, Szyszkowicz M, et al. Global estimates of mortality associated with longterm exposure to outdoor fine particulate matter. *Proc Natl Acad Sci U S A*. 2018;115(38):9592-9597.

62. Forouzanfar MH, Afshin A, Alexander LT, et al. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet*. 2016;388(10053):1659-1724.

63. Cohen AJ, Brauer M, Burnett R, et al. Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015. *Lancet*. 2017;389(10082):1907-1918.

64. Wilcox A. *Fertility and Pregnancy: An Epidemiologic Perspective*. Oxford University Press; 2011.

65. 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. *JAMA Netw Open*. 2020;3(6):e208243.

66. Glinianaia S V., Rankin J, Bell R, Pless-Mulloli T, Howel D. Does particulate air pollution contribute to infant death? A systematic review. *Environ Health Perspect*. 2004;112(14):1365-1370.

67. Liu Y, Wang L, Wang F, Li C. Effect of fine particulate matter (PM2.5) on rat placenta pathology and perinatal outcomes. *Med Sci Monit*. 2016;22:3274-3280.

68. Veras MM, Damaceno-Rodrigues NR, Caldini EG, et al. Particulate urban air pollution affects the functional morphology of mouse placenta. *Biol Reprod*. 2008;79(3):578-584.

69. Slama R, Darrow L, Parker J, et al. Meeting report: Atmospheric pollution and human reproduction. In: *Environmental Health Perspectives*. Vol 116. ; 2008:791-798.

70. Clemens T, Turner S, Dibben C. Maternal exposure to ambient air pollution and fetal growth in North-East Scotland: A population-based study using routine ultrasound scans. *Environ Int*. 2017;107:216-226.

71. Cao Z, Meng L, Zhao Y, et al. Maternal exposure to ambient fine particulate matter and fetal growth in Shanghai, China. *Environ Heal A Glob Access Sci Source*. 2019;18(1).

72. Lin L, Li Q, Yang J, et al. The associations of particulate matters with fetal growth in utero and birth weight: A birth cohort study in Beijing, China. *Sci Total Environ*. 2020;709.

73. Zhao N, Qiu J, Ma S, et al. Effects of prenatal exposure to ambient air pollutant PM10 on ultrasound-measured fetal growth. *Int J Epidemiol*. 2018;47(4):1072-1081.

74. Di Q, Amini H, Shi L, et al. An ensemble-based model of PM2.5 concentration across the

contiguous United States with high spatiotemporal resolution. *Environ Int*. 2019;130.

75. Qiu X, Fong KC, Shi L, et al. Prenatal exposure to particulate air pollution and gestational age at delivery in Massachusetts neonates 2001–2015. *Environ Epidemiol*. 2020;4(5):e113.

76. Fong KC, Kosheleva A, Kloog I, et al. Fine Particulate Air Pollution and Birthweight: Differences in Associations Along the Birthweight Distribution. *Epidemiology*. 2019;30(5):617-623.

77. Fong KC, Kloog I, Coull BA, et al. Residential greenness and birthweight in the state of Massachusetts, USA. *Int J Environ Res Public Health*. 2018;15(6).

78. CDC. Fetal Development Chart. https://www.cdc.gov/ncbddd/fasd/documents/fasdbrochure_final.pdf

79. American College of Obstetricians and Gynecologists and Committee on Obstetric Practice. Methods for estimating the due date. *Obs Gynecol*. 2017;(129):e150-4.

80. Papageorghiou AT, Ohuma EO, Altman DG, et al. International standards for fetal growth based on serial ultrasound measurements: The Fetal Growth Longitudinal Study of the INTERGROWTH-21st Project. *Lancet*. 2014;384(9946):869-879.

81. NASA. Land Data Assimilation Systems (LDAS). https://ldas.gsfc.nasa.gov/data

82. Kind AJH, Buckingham WR. Making Neighborhood-Disadvantage Metrics Accessible — The Neighborhood Atlas. *N Engl J Med*. 2018;378(26):2456-2458.

83. Smarr MM, Vadillo-Ortega F, Castillo-Castrejon M, O'Neill MS. The use of ultrasound measurements in environmental epidemiological studies of air pollution and fetal growth. *Curr Opin Pediatr*. 2013;25(2):240-246.

84. Cefalu M, Dominici F. Does exposure prediction bias health-effect estimation?: The

relationship between confounding adjustment and exposure prediction. *Epidemiology*. 2014;25(4):583-590.

85. Jedrychowski W, Perera F, Mrozek-Budzyn D, et al. Gender differences in fetal growth of newborns exposed prenatally to airborne fine particulate matter. *Environ Res*. 2009;109(4):447-456. doi:10.1016/j.envres.2009.01.009

86. United States Environmental Protection Agency. NAAQS Table. https://www.epa.gov/criteria-air-pollutants/naaqs-table

87. Jaekel J, Sorg C, Baeuml J, Bartmann P, Wolke D. Head growth and intelligence from birth to adulthood in very preterm and term born individuals. *J Int Neuropsychol Soc*. 2019;25(1):48-56.

88. Kirkegaard H, Moller S, Wu C, et al. Associations of birth size, infancy, and childhood growth with intelligence quotient at 5 years of age: a Danish cohort study. *Am J Clin Nutr*. Published online 2020.

89. Villar J, Gunier RB, Tshivuila-Matala COO, et al. Fetal cranial growth trajectories are associated with growth and neurodevelopment at 2 years of age: INTERBIO-21st Fetal Study. *Nat Med*. Published online 2021:1-6.

90. Magnusson PEK, Rasmussen F, Gyllensten UB. Height at age 18 years is a strong predictor of attained education later in life: Cohort study of over 950 000 Swedish men. *Int J Epidemiol*. 2006;35(3):658-663.

91. McGovern ME, Krishna A, Aguayo VM, Subramanian S V. A review of the evidence linking child stunting to economic outcomes. *Int J Epidemiol*. 2017;46(4):1171-1191.

92. Rückinger S, Beyerlein A, Jacobsen G, von Kries R, Vik T. Growth in utero and body mass index at age 5years in children of smoking and non-smoking mothers. *Early Hum*

Dev. 2010;86(12):773-777.

93. Hadlock FP, Harrist RB, Sharman RS, Deter RL, Park SK. Estimation of fetal weight with the use of head, body, and femur measurements-A prospective study. *Am J Obstet Gynecol*. 1985;151(3):333-337.

94. Hansen CA, Barnett AG, Pritchard G. The effect of ambient air pollution during early pregnancy on fetal ultrasonic measurements during mid-pregnancy. *Environ Health Perspect*. 2008;116(3):362-369.

95. Lamichhane DK, Ryu J, Leem J-H, et al. Air pollution exposure during pregnancy and ultrasound and birth measures of fetal growth: A prospective cohort study in Korea. *Sci Total Environ*. 2018;619-620:834-841.

96. van den Hooven EH, Pierik FH, de Kluizenaar Y, et al. Air pollution exposure during pregnancy, ultrasound measures of fetal growth, and adverse birth outcomes: A prospective cohort study. *Environ Health Perspect*. 2012;120(1):150-156.

97. Bruckner TA, Catalano R, Ahern J. Male fetal loss in the U.S. following the terrorist attacks of September 11, 2001. *BMC Public Health*. 2010;10(1):273.

98. Fukuda M, Fukuda K, Shimizu T, Moøer H. Decline in sex ratio at birth after Kobe earthquake. *Hum Reprod*. 1998;13(8):2321-2322.

99. Mocarelli P, Brambilla P, Gerthoux MP, Patterson DG, Needham LL. Change in sex ratio with exposure to dioxin [14]. *Lancet*. 1996;348(9024):409.

100. Breslau J, Aguilar-Gaxiola S, Kendler KS, Su M, Williams D, Kessler RC. Specifying race-ethnic differences in risk for psychiatric disorder in a USA national sample. *Psychol Med*. 2006;36(1):57-68.

101. Weisskopf MG, Webster TF. Trade-offs of Personal Versus More Proxy Exposure

Measures in Environmental Epidemiology. *Epidemiology*. 2017;28(5):635-643.

102. Warren JL, Son J-Y, Pereira G, Leaderer BP, Bell ML. Investigating the Impact of Maternal Residential Mobility on Identifying Critical Windows of Susceptibility to Ambient Air Pollution During Pregnancy. *Am J Epidemiol*. 2018;187(5):992-1000.
103. Nethery E, Brauer M, Janssen P. Time-activity patterns of pregnant women and changes during the course of pregnancy. *J Expo Sci Environ Epidemiol*. 2009;19(3):317-324.
104. Goin DE, Casey JA, Kioumourtzoglou M-A, Cushing LJ, Morello-Frosch R. Environmental hazards, social inequality, and fetal loss: Implications of live-birth bias for estimation of disparities in birth outcomes. *Environ Epidemiol*. 2021;5(2):e131.
105. Leung M, Kioumourtzoglou MA, Raz R, Weisskopf MG. Bias due to Selection on Live Births in Studies of Environmental Exposures during Pregnancy: A Simulation Study. *Environ Health Perspect*. 2021;129(4):47001.
106. IPCC. *Global Warming of 1.5 ºC.*; 2018.
107. Haustein K, Allen MR, Forster PM, et al. A real-time Global Warming Index. *Sci Rep*. 2017;7(1).
108. NASA, NOAA. 2018 fourth warmest year in continued warming trend, according to NASA, NOAA. Published 2019. https://www.nasa.gov/press-release/2018-fourth-warmest-year-in-continued-warming-trend-according-to-nasa-noaa
109. World Health Organization. Climate change and health. Published 2018. https://www.who.int/news-room/fact-sheets/detail/climate-change-and-health
110. Watts N, Amann M, Arnell N, et al. The 2018 report of the Lancet Countdown on health and climate change: shaping the health of nations for centuries to come. *Lancet*. 2018;392(10163):2479-2514.

111. The American College of Obstetricians and Gynecologists. *Climate Change and Women's Health: Position Statement*.; 2018.

112. USAID. *Heat Waves and Human Health: Emerging Evidence and Experience to Inform Risk Management in a Warming World*.; 2019. https://www.climatelinks.org/resources/heat-waves-and-human-health-emerging-evidence-and-experience-inform-risk-management

113. Laburn HP. How does the fetus cope with thermal challenges? *Physiology*. 1996;11(2):96-100.

114. Wells JCK. Thermal environment and human birth weight. *J Theor Biol*. 2002;214(3):413-425.

115. Di Renzo GC, Conry JA, Blake J, et al. International Federation of Gynecology and Obstetrics opinion on reproductive health impacts of exposure to toxic environmental chemicals. *Int J Gynecol Obstet*. 2015;131(3):219-225.

116. Edwards MJ, Saunders RD, Shiota K. Effects of heat on embryos and foetuses. In: *International Journal of Hyperthermia*. Vol 19. ; 2003:295-324.

117. Wang J, Liu X, Dong M, et al. Associations of maternal ambient temperature exposures during pregnancy with the placental weight, volume and PFR: A birth cohort study in Guangzhou, China. *Environ Int*. 2020;139.

118. Chersich MF, Pham MD, Areal A, et al. Associations between high temperatures in pregnancy and risk of preterm birth, low birth weight, and stillbirths: Systematic review and meta-analysis. *BMJ*. 2020;371.

119. Grantham-McGregor S, Cheung YB, Cueto S, et al. Developmental potential in the first 5 years for children in developing countries. *Lancet (London, England)*. 2007;369(9555):60-

70.

120. Black MM, Walker SP, Fernald LCH, et al. Early childhood development coming of age: science through the life course. *Lancet*. 2016;380:1-14.

121. Leung M, Weisskopf MG, Laden F, et al. Exposure to PM2.5 during pregnancy and ultrasound parameters of fetal growth in Massachusetts, USA. Submitted to: Environmental Health Perspectives. Published online 2021.

122. NASA. NLDAS-2 Model Data Description/Information. https://ldas.gsfc.nasa.gov/index.php/nldas/v2/models

123. Cosgrove BA, Lohmann D, Mitchell KE, et al. Real-time and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project. *J Geophys Res Atmos*. 2003;108(D22):2002JD003118.

124. Wood S. *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC Press; 2006.

125. Strand LB, Barnett AG, Tong S. The influence of season and ambient temperature on birth outcomes: A review of the epidemiological literature. *Environ Res*. 2011;111(3):451-462.

126. Ngo NS, Horton RM. Climate change and fetal health: The impacts of exposure to extreme temperatures in New York City. *Environ Res*. 2016;144(Pt A):158-164.

127. Basu R, Rau R, Pearson D, Malig B. Temperature and Term Low Birth Weight in California. *Am J Epidemiol*. 2018;187(11):2306-2314.

128. Bennett G. Hyperthermia: malformations to chaperones. *Birth Defects Res B Dev Reprod Toxicol*. 2010;89(4):279-288.

129. Thompson RA, Nelson CA. Developmental science and the media: Early brain development. *Am Psychol*. 2001;56(1):5-15.

130. Bouchama A, Knochel J. Heat stroke. *N Engl J Med*. 2002;346(25):1978-19788.

131. Gasparrini A, Guo Y, Hashizume M. Mortality risk attributable to high and low ambient temperature: a multicountry observational study. *Lancet*. 2015;386(9991):369-375.

132. Climate Central. *The Case of the Shifting Snow*.; 2020. https://www.climatecentral.org/news/report-the-case-of-the-shifting-snow

133. Londero AP, Rossetti E, Pittini C, Cagnacci A, Driul L. Maternal age and the risk of adverse pregnancy outcomes: a retrospective cohort study. *BMC Pregnancy Childbirth*. 2019;19(1):261.

134. Hsu A, Sheriff G, Chakraborty T, Manya D. Disproportionate exposure to urban heat island intensity across major US cities. *Nat Commun*. 2021;12(1):2721.

135. Crosson WL, Al-Hamdan MZ, Insaf TZ. Downscaling NLDAS-2 daily maximum air temperatures using MODIS land surface temperatures. *PLoS One*. 2020;15(1).

Appendix

Table S1.1. Average odds ratios for the association between nitrogen dioxide (NO_2; exposure A) and autism spectrum disorder (ASD, outcome Y) among live births $S = 1$ ($OR_{AY|S=1}$) and 95% simulation intervals (SI) over 1000 simulations with input parameters π_U=0.25, 0.50, 0.75 and OR_S=1.5, 2.0, 2.5, 3.0 under two selection mechanisms and when they both co-occur with a single unmeasured U, assuming a true null effect of NO_2 on ASD

Inputs		Average NO_2-ASD odds ratio among live births, $OR_{AY|S=1}$ (95% SI)		
π_U	OR_S	Mechanism 1: Collider-stratification without interaction[a]	Mechanism 2: Depletion of susceptibles[b]	Both Mechanisms: Mechanism 1 + Mechanism 2[c]
0.25	1.5	1.00 (0.93,1.07)	1.00 (0.93,1.07)	0.98 (0.91,1.06)
	2.0	0.99 (0.92,1.06)	0.98 (0.91,1.04)	0.94 (0.86,1.04)
	2.5	0.97 (0.89,1.06)	0.94 (0.88,1.01)	0.94 (0.84,1.04)
	3.0	0.95 (0.86,1.06)	0.91 (0.85,0.97)	0.94 (0.83,1.06)
0.50	1.5	1.00 (0.93,1.07)	1.00 (0.93,1.06)	0.98 (0.91,1.05)
	2.0	0.99 (0.91,1.06)	0.97 (0.91,1.03)	0.90 (0.82,0.99)
	2.5	0.96 (0.88,1.04)	0.92 (0.86,0.98)	0.86 (0.77,0.96)
	3.0	0.94 (0.86,1.03)	0.87 (0.82,0.94)	0.85 (0.74,0.97)
0.75	1.5	1.00 (0.93,1.06)	1.00 (0.94,1.06)	0.98 (0.91,1.05)
	2.0	0.99 (0.92,1.06)	0.98 (0.92,1.04)	0.87 (0.78,0.96)
	2.5	0.97 (0.89,1.05)	0.94 (0.89,1.00)	0.79 (0.69,0.90)
	3.0	0.95 (0.86,1.04)	0.89 (0.83,0.94)	0.75 (0.64,0.89)

Note: π_U is the prevalence of unmeasured U in the target population of all conceptions; OR_S is the magnitude of the select effects (OR_{AS}, OR_{US}, $OR_{\{AU\}S}$, OR_{UY}) in the target population of all conceptions. The expected value of $OR_{AY|S=1}$ is 1.0 in the absence of live birth bias, where deviations from 1.0 quantify the magnitude of live birth bias.

[a] Collider-stratification without interaction: NO_2 exposure A and unmeasured factor U have independent causal effects on fetal loss, but with no interaction on the multiplicative scale ($OR_{\{AU\}S}$=1, and OR_{AS}=OR_{US}=OR_{UY}=OR_S)

[b] Depletion of susceptibles: NO_2 exposure A has a causal effect on fetal loss only in the subset of susceptible fetuses ($U = 1$), but neither A or U have independent causal effects on fetal loss (OR_{AS}=OR_{US}=1, and $OR_{\{AU\}S}$=OR_{UY}=OR_S)

[c] Both Mechanisms: NO_2 exposure A and unmeasured factor U have independent causal effects on fetal loss, and with interaction on the multiplicative scale (OR_{AS}=OR_{US}=$OR_{\{AU\}S}$=OR_{UY}=OR_S)

Table S1.2. Average bias parameters and 95% simulation intervals (SI) over 1000 simulations with input parameters π_U=0.25, 0.50, 0.75 and OR_S=1.5, 2.0, 2.5, 3.0 under two selection mechanisms and when they both co-occur with a single unmeasured U, assuming a true null effect of nitrogen dioxide (NO_2; exposure A) on autism spectrum disorder (ASD, outcome Y)

Inputs		Average bias parameters among live births (95% SI)					
		Mechanism 1: Collider-stratification without interaction[a]		Mechanism 2: Depletion of susceptibles[b]		Both Mechanisms: Mechanism 1 + Mechanism 2[c]	
π_U	OR_S	$OR_{AU\|S=1}$	$\pi_{U\|S=1}$	$OR_{AU\|S=1}$	$\pi_{U\|S=1}$	$OR_{AU\|S=1}$	$\pi_{U\|S=1}$
0.25	1.5	0.98 (0.96,1.00)	0.24 (0.23,0.24)	0.94 (0.93,0.96)	0.23 (0.23,0.23)	0.76 (0.74,0.78)	0.18 (0.18,0.18)
	2.0	0.90 (0.88,0.93)	0.21 (0.21,0.21)	0.83 (0.81,0.85)	0.20 (0.20,0.20)	0.44 (0.42,0.46)	0.08 (0.08,0.09)
	2.5	0.83 (0.80,0.85)	0.18 (0.17,0.18)	0.70 (0.69,0.72)	0.17 (0.17,0.17)	0.29 (0.27,0.31)	0.04 (0.04,0.04)
	3.0	0.77 (0.74,0.80)	0.15 (0.15,0.15)	0.59 (0.58,0.61)	0.14 (0.14,0.14)	0.22 (0.20,0.24)	0.02 (0.02,0.03)
0.50	1.5	0.98 (0.96,1.00)	0.48 (0.48,0.49)	0.95 (0.93,0.96)	0.47 (0.47,0.48)	0.76 (0.74,0.77)	0.40 (0.39,0.40)
	2.0	0.91 (0.89,0.93)	0.44 (0.44,0.45)	0.83 (0.82,0.85)	0.43 (0.43,0.43)	0.43 (0.42,0.45)	0.22 (0.21,0.22)
	2.5	0.83 (0.81,0.85)	0.39 (0.39,0.40)	0.70 (0.69,0.71)	0.38 (0.38,0.38)	0.29 (0.27,0.30)	0.12 (0.11,0.12)
	3.0	0.77 (0.75,0.80)	0.35 (0.34,0.35)	0.59 (0.57,0.60)	0.33 (0.33,0.33)	0.21 (0.20,0.23)	0.07 (0.07,0.07)
0.75	1.5	0.98 (0.96,1.00)	0.74 (0.73,0.74)	0.94 (0.93,0.96)	0.73 (0.73,0.73)	0.75 (0.73,0.77)	0.66 (0.66,0.67)
	2.0	0.90 (0.88,0.93)	0.70 (0.70,0.71)	0.83 (0.81,0.85)	0.70 (0.69,0.70)	0.42 (0.41,0.44)	0.45 (0.45,0.46)
	2.5	0.83 (0.81,0.85)	0.66 (0.66,0.67)	0.69 (0.68,0.71)	0.65 (0.64,0.65)	0.28 (0.27,0.30)	0.28 (0.28,0.29)
	3.0	0.77 (0.75,0.80)	0.62 (0.61,0.62)	0.58 (0.56,0.59)	0.60 (0.59,0.60)	0.21 (0.19,0.22)	0.19 (0.18,0.19)

Note: π_U is the prevalence of unmeasured U in the total population of all conceptions; OR_S is the magnitude of the select effects (OR_{AS}, OR_{US}, $OR_{\{AU\}S}$, OR_{UY}) in the total population of all conceptions; $OR_{AU|S=1}$ is the association between A and U among live births; $\pi_{U|S=1}$ is the prevalence of U among live births. The expected values (in the absence of live birth bias) for $\pi_{U|S=1}$ is π_U, and for $OR_{AU|S=1}$ is OR_{AU} (which is 1.0 in the total population of all conceptions). Both $\pi_{U|S=1}$ and $OR_{AU|S=1}$ are parameters in the selected population (live births) that drive the strength of live birth bias, where deviations from their expected values as a consequence of the input parameters determine the strength of live birth bias under each mechanism.

[a] Collider-stratification without interaction: NO_2 exposure A and unmeasured factor U have independent causal effects on fetal loss, but with no interaction on the multiplicative scale ($OR_{\{AU\}S}$=1, and OR_{AS}=OR_{US}=OR_{UY}=OR_S)

[b] Depletion of susceptibles: NO_2 exposure A has a causal effect on fetal loss only in the subset of susceptible fetuses ($U = 1$), but neither A or U have independent causal effects on fetal loss (OR_{AS}=OR_{US}=1, and $OR_{\{AU\}S}$=OR_{UY}=OR_S)

[c] Both Mechanisms: NO_2 exposure A and unmeasured factor U have independent causal effects on fetal loss, and with interaction on the multiplicative scale (OR_{AS}=OR_{US}=$OR_{\{AU\}S}$=OR_{UY}=OR_S)

Table S1.3. Average odds ratios for the association between nitrogen dioxide (NO_2; exposure A) and autism spectrum disorder (ASD, outcome Y) among live births $S = 1$ ($OR_{AY|S=1}$) and 95% simulation intervals (SI) over 1000 simulations with input parameters π_U=0.25, 0.50, 0.75 and OR_S=1.5, 2.0, 2.5, 3.0 under two selection mechanisms and when they both co-occur with multiple unmeasured Us, assuming a true null effect of NO_2 on ASD

Inputs		Average NO_2-ASD odds ratio among live births, $OR_{AY\|S=1}$ (95% SI)								
		Mechanism 1: Collider-stratification without interaction[a]			Mechanism 2: Depletion of susceptibles[b]			Both Mechanisms: Mechanism 1 + Mechanism 2[c]		
π_U	OR_S	One U	Two Us	Three Us	One U	Two Us	Three Us	One U	Two Us	Three Us
0.25	1.5	1.00 (0.93,1.07)	1.00 (0.93,1.07)	0.99 (0.93,1.06)	1.00 (0.93,1.07)	0.99 (0.92,1.05)	0.97 (0.91,1.04)	0.98 (0.91,1.06)	0.96 (0.89,1.03)	0.95 (0.88,1.02)
	2.0	0.99 (0.92,1.06)	0.97 (0.90,1.05)	0.96 (0.90,1.03)	0.98 (0.91,1.04)	0.94 (0.89,1.01)	0.90 (0.85,0.96)	0.94 (0.86,1.04)	0.90 (0.82,1.00)	0.87 (0.78,0.96)
	2.5	0.97 (0.89,1.06)	0.94 (0.87,1.02)	0.92 (0.85,0.99)	0.94 (0.88,1.01)	0.89 (0.84,0.95)	0.84 (0.79,0.90)	0.94 (0.84,1.04)	0.88 (0.78,0.99)	0.84 (0.74,0.94)
	3.0	0.95 (0.86,1.06)	0.92 (0.82,1.01)	0.89 (0.80,0.98)	0.91 (0.85,0.97)	0.84 (0.79,0.90)	0.79 (0.74,0.85)	0.94 (0.83,1.06)	0.88 (0.77,1.00)	0.83 (0.70,0.96)
0.50	1.5	1.00 (0.93,1.07)	1.00 (0.93,1.06)	0.99 (0.93,1.05)	1.00 (0.93,1.06)	0.98 (0.92,1.04)	0.95 (0.89,1.01)	0.98 (0.91,1.05)	0.94 (0.87,1.01)	0.91 (0.83,0.99)
	2.0	0.99 (0.91,1.06)	0.97 (0.90,1.03)	0.95 (0.89,1.01)	0.97 (0.91,1.03)	0.90 (0.84,0.95)	0.82 (0.77,0.87)	0.90 (0.82,0.99)	0.83 (0.74,0.92)	0.78 (0.69,0.89)
	2.5	0.96 (0.88,1.04)	0.93 (0.86,1.00)	0.90 (0.84,0.97)	0.92 (0.86,0.98)	0.83 (0.77,0.88)	0.74 (0.68,0.79)	0.86 (0.77,0.96)	0.77 (0.67,0.89)	0.71 (0.59,0.84)
	3.0	0.94 (0.86,1.03)	0.89 (0.82,0.97)	0.87 (0.80,0.95)	0.87 (0.82,0.94)	0.77 (0.71,0.82)	0.67 (0.62,0.72)	0.85 (0.74,0.97)	0.75 (0.63,0.88)	0.67 (0.53,0.84)
0.75	1.5	1.00 (0.93,1.06)	1.00 (0.94,1.06)	0.99 (0.94,1.05)	1.00 (0.94,1.06)	0.97 (0.91,1.04)	0.94 (0.88,0.99)	0.98 (0.91,1.05)	0.93 (0.86,1.01)	0.88 (0.80,0.98)
	2.0	0.99 (0.92,1.06)	0.97 (0.91,1.03)	0.96 (0.91,1.02)	0.98 (0.92,1.04)	0.87 (0.82,0.93)	0.78 (0.73,0.84)	0.87 (0.78,0.96)	0.79 (0.70,0.90)	0.72 (0.59,0.86)
	2.5	0.97 (0.89,1.05)	0.94 (0.88,1.01)	0.92 (0.87,0.99)	0.94 (0.89,1.00)	0.78 (0.73,0.83)	0.65 (0.60,0.70)	0.79 (0.69,0.90)	0.68 (0.57,0.82)	0.60 (0.44,0.77)
	3.0	0.95 (0.86,1.04)	0.91 (0.83,0.99)	0.90 (0.83,0.98)	0.89 (0.83,0.94)	0.71 (0.66,0.76)	0.56 (0.51,0.61)	0.75 (0.64,0.89)	0.61 (0.48,0.77)	0.51 (0.34,0.73)

Note: U is a vector that consists of ≤ 3 unmeasured factors (U_1, U_2, U_3), where input parameters were applied equally for each unmeasured factor; thus, all references to U henceforth applies to each of the unmeasured factors U_1, U_2, U_3. π_U is the prevalence of unmeasured U in the target population of all conceptions; OR_S is the magnitude of the select effects (OR_{AS}, OR_{US}, $OR_{\{AU\}S}$, OR_{UY}) in the target population of all conceptions. The expected value of $OR_{AY|S=1}$ is 1.0 in the absence of live birth bias, where deviations from 1.0 quantify the magnitude of live birth bias. Estimates shown in the column "One U" are equivalent to the $OR_{AY|S=1}$ displayed in Table S1.

[a] Collider-stratification without interaction: NO_2 exposure A and unmeasured factor U have independent causal effects on fetal loss, but with no interaction on the multiplicative scale ($OR_{\{AU\}S}$=1, and OR_{AS}=OR_{US}=OR_{UY}=OR_S)

[b] Depletion of susceptibles: NO_2 exposure A has a causal effect on fetal loss only in the subset of susceptible fetuses ($U = 1$), but neither A or U have independent causal effects on fetal loss (OR_{AS}=OR_{US}=1, and $OR_{\{AU\}S}$=OR_{UY}=OR_S)

[c] Both Mechanisms: NO_2 exposure A and unmeasured factor U have independent causal effects on fetal loss, and with interaction on the multiplicative scale (OR_{AS}=OR_{US}=$OR_{\{AU\}S}$=OR_{UY}=OR_S)

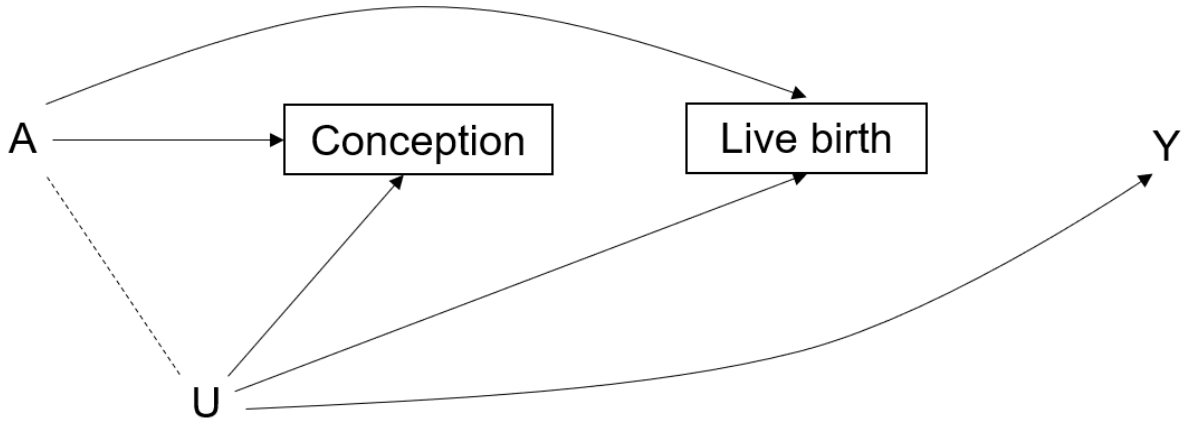

Figure S1.1. A directed acyclic graph (DAG) representing the additional bias that could arise from conditioning on the ability to conceive, where exposure A affects conceptions and live births, both of which are also affected by an independent unmeasured risk factor U for the outcome Y. Here, conditioning on both conceptions and live births induces a stronger A-U association in the selected population, which produces an A-Y association with a stronger downward bias compared to that generated from the causal structure in Figure 1.

Simulation Code

```
library(Rlab)
library(dplyr)

# set seed
set.seed()

lbb <- function(iter, n, p, mean, sd, con,
                bL, bY, or1, or2, or3, or4, or5, nU)

{
  ## iter   number of iterations
  ## n      sample size for each of the generated datasets
  ## p      prevalence of U
  ## mean   mean of normally distributed exposure
  ## sd     standard deviation of normally distributed exposure
  ## con    exposure contrast (i.e., increment of exposure for associations)
  ## bL     baseline odds of loss
  ## bY     baseline odds of outcome
  ## or1    odds ratio for exposure-loss effect
  ## or2    odds ratio for U-loss effect
  ## or3    odds ratio for exposure*U loss interaction effect
  ## or4    odds ratio for U-outcome effect
  ## or5    odds ratio for exposure-outcome effect

  # create matrix to store simulation results
  results.cs <- rep(NA, iter)
  results.dos <- rep(NA, iter)
  results.both <- rep(NA, iter)

  for (i in 1:iter){
    set.seed(i)

    # monitor simulation
    if((i %% 10) == 0) print(c(iter, i))

    # specify contrast
    mean_std <- mean/con
    sd_std <- sd/con

    ###########################################
    ### Mechanism 1: collider-stratification ###
    ###########################################

    # data generating mechanism
    cs <- data.frame("id" = 1:n) %>%
      mutate(no2 = rnorm(n, mean_std, sd_std), # exposure is gaussian
             U = rbern(n, p),
             prob_loss = plogis(log(bL) + log(or1)*no2 + log(or2)*U),
             loss = rbern(n, prob_loss),
             pY = plogis(log(bY) + log(or4)*U + log(or5)*no2),
             Y = rbern(n, pY))

    # fit a logistic model among live births
    model1 <- cs %>%
      glm(formula = Y ~ no2,
```

```
          family = binomial(link = "logit"),
          data = .,
          subset = loss==0)

    results.cs[i] <- exp(model1$coef[2])

    ##############################################
    ### Mechanism 2: depletion of susceptibles ###
    ##############################################

    # data generating mechanism
    dos <- data.frame("id" = 1:n) %>%
      mutate(no2 = rnorm(n, mean_std, sd_std),
             U = rbern(n, p),
             prob_loss = plogis(log(bL) + log(or3)*no2*U),
             loss = rbern(n, prob_loss),
             pY = plogis(log(bY) + log(or4)*U + log(or5)*no2),
             Y = rbern(n, pY))

    # fit a logistic model among live births
    model2 <- dos %>%
      glm(formula = Y ~ no2,
          family = binomial(link = "logit"),
          data = .,
          subset = loss==0)

    results.dos[i] <- exp(model2$coef[2])

    ########################
    ### Both Mechanisms ###
    ########################

    # data generating mechanism
    both <- data.frame("id" = 1:n) %>%
      mutate(no2 = rnorm(n, mean_std, sd_std),
             U = rbern(n, p),
             prob_loss = plogis(log(bL) + log(or1)*no2 + log(or2)*U +
log(or3)*no2*U),
             loss = rbern(n, prob_loss),
             pY = plogis(log(bY) + log(or4)*U + log(or5)*no2),
             Y = rbern(n, pY))

    # fit a logistic model among live births
    model3 <- both %>%
      glm(formula = Y ~ no2,
          family = binomial(link = "logit"),
          data = .,
          subset = loss==0)

    results.both[i] <- exp(model3$coef[2])

  }

  # turn results into data frame
  simResults <- data.frame(results.cs, results.dos, results.both)

  # get mean ORs for ASD-no2 association
```

```
  csOR <- mean(simResults$results.cs)
  dosOR <- mean(simResults$results.dos)
  bothOR <- mean(simResults$results.both)

  # get the 95% simulation intervals
  csOR.si <- quantile(simResults$results.cs, c(0.025,0.975))
  dosOR.si <- quantile(simResults$results.dos, c(0.025,0.975))
  bothOR.si <- quantile(simResults$results.both, c(0.025,0.975))

  # put results together
  lbb.cs <- paste(formatC(csOR,digits=2,format="f"),
                  " (",formatC(csOR.si[1],digits=2,format="f"),
                  ",",formatC(csOR.si[2],digits=2,format="f"),")",sep="")
  lbb.dos <- paste(formatC(dosOR,digits=2,format="f"),
                   " (",formatC(dosOR.si[1],digits=2,format="f"),
                   ",",formatC(dosOR.si[2],digits=2,format="f"),")",sep="")
  lbb.both <- paste(formatC(bothOR,digits=2,format="f"),
                    " (",formatC(bothOR.si[1],digits=2,format="f"),
                    ",",formatC(bothOR.si[2],digits=2,format="f"),")",sep="")
  lbb.results <- cbind(lbb.cs,lbb.dos,lbb.both)
  lbb.results

}

# example inputs
lbb(iter=1000, n=100000, p=0.75, mean=16.7, sd=4.3, con=5.85,
    bL=0.05, bY=0.015, or1=2, or2=2, or3=2, or4=2, or5=1)
```

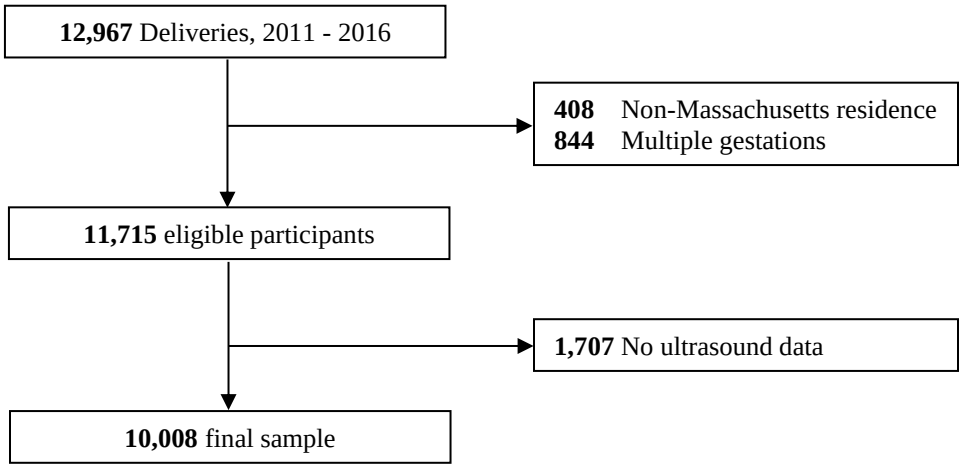

Figure S2.1. Flow chart of participant selection in deliveries at Beth Israel Deaconess Medical Center, Boston, Massachusetts from 2011-2016

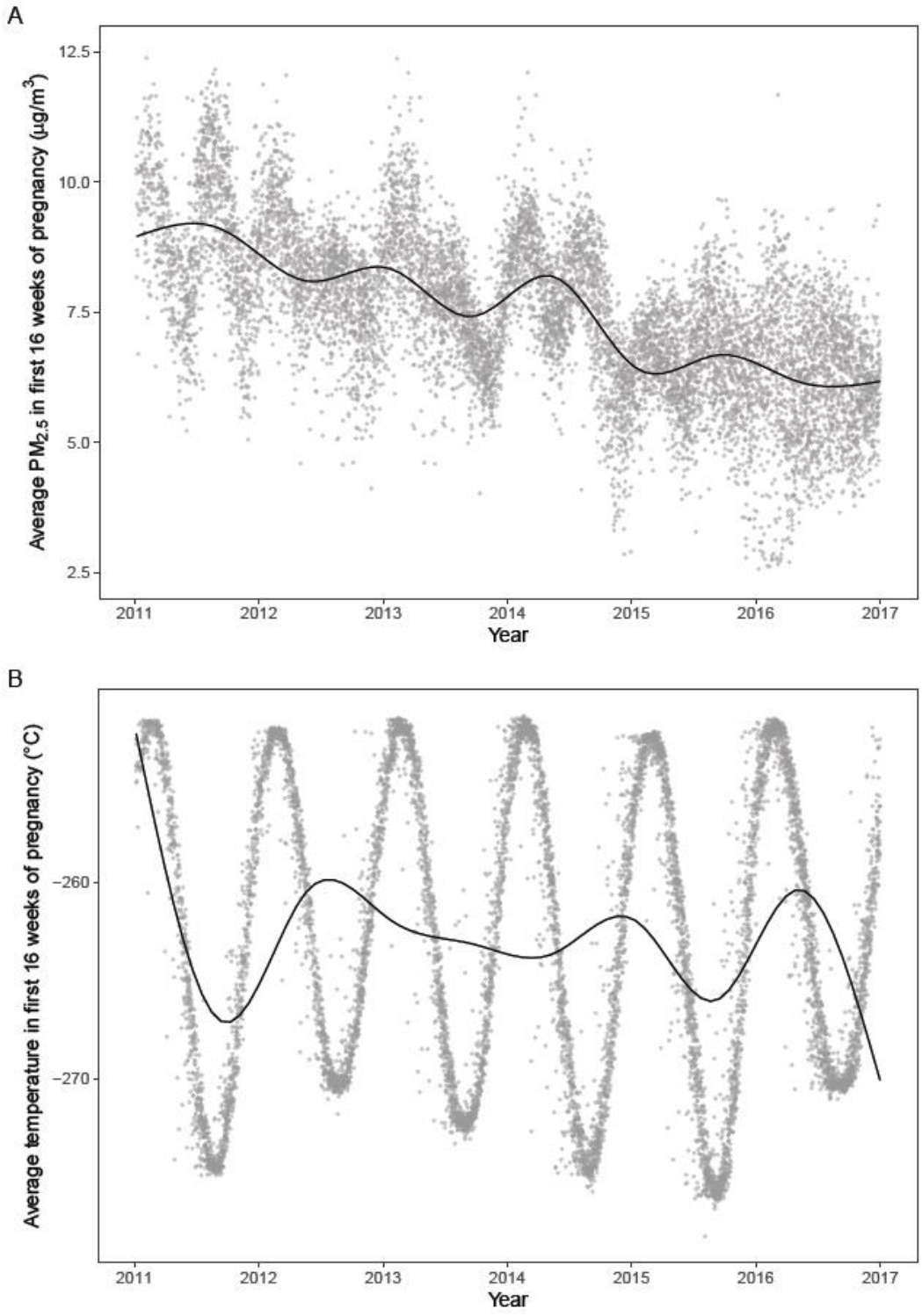

Figure S2.2. Distributions of A) average $PM_{2.5}$ in the first 16 weeks of pregnancy, and B) average temperature in the first 16 weeks of pregnancy over the study period (2011 – 2016). Grey points represent the mean value for each pregnancy. The solid black line represents the fitted mean values during pregnancy from a generalized additive model with a penalized spline for date of birth.

Table S2.1. Linear mixed effects model estimates and 95% CIs for the association between cumulative $PM_{2.5}$ and fetal growth outcomes (ultrasound parameters and birth weight) from deliveries at Beth Israel Deaconess Medical Center, Boston, Massachusetts in 2011-2016 (N=10,008)

Growth outcome	Estimate	95% CI	*p*-Value
Anatomic scans (<24 weeks)			
BPD z-score	-0.15	(-0.25, -0.05)	0.004
HC z-score	-0.08	(-0.18, 0.01)	0.09
FL z-score	-0.03	(-0.13, 0.08)	0.59
AC z-score	-0.08	(-0.18, 0.03)	0.17
Growth scans (≥24 weeks)			
BPD z-score	-0.07	(-0.17, 0.02)	0.15
HC z-score	0.04	(-0.06, 0.14)	0.43
FL z-score	-0.05	(-0.14, 0.05)	0.33
AC z-score	-0.05	(-0.14, 0.05)	0.31
Birth weight			
Birth weight z-score	-0.12	(-0.24, -0.01)	0.03

Note: Estimates represent the difference in mean z-score with a 5 $\mu g/m^3$ increase in cumulative $PM_{2.5}$ from conception to the assessment of fetal growth after adjusting for maternal age, race, education, insurance type, parity, fetal sex, conception year, day of the year of conception, temperature, and Area Deprivation Index.

Table S2.2. Linear mixed effects model estimates and 95% CIs for the association between $PM_{2.5}$ in the first 16 weeks and fetal growth outcomes (ultrasound parameters and birth weight) by fetal sex from deliveries at Beth Israel Deaconess Medical Center, Boston, Massachusetts in 2011-2016 (N=10,008)

Growth outcome	Female		Male	
	Estimate (95% CI)	*p*-Value	Estimate (95% CI)	*p*-Value
Anatomic scans (<24 weeks)				
BPD z-score	-0.18 (-0.33, -0.03)	Ref	-0.18 (-0.33, -0.03)	0.97
HC z-score	-0.11 (-0.25, 0.04)	Ref	-0.08 (-0.23, 0.06)	0.78
FL z-score	-0.07 (-0.22, 0.09)	Ref	-0.08 (-0.23, 0.08)	0.88
AC z-score	-0.10 (-0.26, 0.06)	Ref	-0.12 (-0.28, 0.05)	0.87
Growth scans (≥24 weeks)				
BPD z-score	-0.13 (-0.29, 0.03)	Ref	-0.15 (-0.30, 0.01)	0.84
HC z-score	-0.04 (-0.19, 0.12)	Ref	-0.02 (-0.17, 0.14)	0.82
FL z-score	-0.07 (-0.22, 0.09)	Ref	-0.03 (-0.19, 0.13)	0.70
AC z-score	-0.11 (-0.28, 0.06)	Ref	-0.19 (-0.35, -0.02)	0.37
Birth weight				
Birth weight z-score	-0.14 (-0.26, -0.02)	Ref	-0.10 (-0.21, 0.02)	0.50

Note: Estimates represent the difference in mean z-score with a 5 $\mu g/m^3$ increase in $PM_{2.5}$ during the first 16 weeks of pregnancy after adjusting for maternal age, race, education, insurance type, parity, conception year, day of the year of conception, temperature, and Area Deprivation Index; P-values were computed by testing whether the association for each level of the potential modifier differs from that of the reference level, which is "Female" for fetal sex.

Table S2.3. Linear mixed effects model estimates and 95% CIs for the association between $PM_{2.5}$ in the first 16 weeks and fetal growth outcomes (ultrasound parameters and birth weight) by maternal race from deliveries at Beth Israel Deaconess Medical Center, Boston, Massachusetts in 2011-2016 (N=10,008)

Growth outcome	White		Black		Asian		Hispanic	
	Estimate (95% CI)	*p*-Value	Estimate (95% CI)	*p*-Value	Estimate (95% CI)	*p*-Value	Estimate (95% CI)	*p*-Value
Anatomic scans (<24 weeks)								
BPD z-score	-0.20 (-0.34, -0.06)	Ref	-0.10 (-0.33, 0.12)	0.40	-0.24 (-0.51, 0.04)	0.78	-0.17 (-0.50, 0.15)	0.87
HC z-score	-0.13 (-0.27, 0.01)	Ref	0.05 (-0.18, 0.27)	0.12	-0.16 (-0.41, 0.10)	0.85	-0.02 (-0.36, 0.32)	0.51
FL z-score	-0.10 (-0.25, 0.05)	Ref	-0.05 (-0.29, 0.19)	0.70	-0.07 (-0.35, 0.21)	0.84	-0.01 (-0.35, 0.33)	0.62
AC z-score	-0.16 (-0.32, -0.01)	Ref	-0.04 (-0.29, 0.20)	0.32	-0.12 (-0.41, 0.17)	0.75	-0.00 (-0.37, 0.37)	0.38
Growth scans (≥24 weeks)								
BPD z-score	-0.20 (-0.35, -0.05)	Ref	-0.14 (-0.36, 0.08)	0.59	0.03 (-0.25, 0.32)	0.11	-0.44 (-0.73, -0.15)	0.11
HC z-score	-0.06 (-0.21, 0.09)	Ref	-0.08 (-0.30, 0.15)	0.91	0.08 (-0.19, 0.34)	0.31	-0.33 (-0.65, -0.02)	0.09
FL z-score	-0.10 (-0.25, 0.06)	Ref	-0.06 (-0.31, 0.18)	0.78	0.16 (-0.13, 0.45)	0.08	-0.31 (-0.59, -0.03)	0.13
AC z-score	-0.22 (-0.38, -0.06)	Ref	-0.13 (-0.39, 0.12)	0.50	0.08 (-0.22, 0.38)	0.05	-0.26 (-0.56, 0.05)	0.81
Birth weight								
Birth weight z-score	-0.15 (-0.27, -0.03)	Ref	-0.14 (-0.32, 0.04)	0.93	0.06 (-0.16, 0.28)	0.06	-0.15 (-0.38, 0.08)	0.99

Note: Estimates represent the difference in mean z-score with a 5 $\mu g/m^3$ increase in $PM_{2.5}$ during the first 16 weeks of pregnancy after adjusting for maternal age, education, insurance type, parity, fetal sex, conception year, day of the year of conception, temperature, and Area Deprivation Index; P-values were computed by testing whether the association for each level of the potential modifier differs from that of the reference level, which is "White" for maternal race.

Table S2.4. Linear mixed effects model estimates and 95% CIs for the association between $PM_{2.5}$ in the first 16 weeks and fetal growth outcomes (ultrasound parameters and birth weight) by maternal education from deliveries at Beth Israel Deaconess Medical Center, Boston, Massachusetts in 2011-2016 (N=10,008)

Growth outcome	Less than college		College or higher	
	Estimate (95% CI)	*p*-Value	Estimate (95% CI)	*p*-Value
Anatomic scans (<24 weeks)				
BPD z-score	-0.09 (-0.28, 0.09)	Ref	-0.23 (-0.37, -0.09)	0.14
HC z-score	-0.05 (-0.23, 0.13)	Ref	-0.11 (-0.25, 0.03)	0.51
FL z-score	0.07 (-0.11, 0.25)	Ref	-0.14 (-0.29, 0.01)	0.03
AC z-score	0.02 (-0.26, 0.06)	Ref	-0.18 (-0.34, -0.02)	0.04
Growth scans (≥24 weeks)				
BPD z-score	-0.17 (-0.35, 0.02)	Ref	-0.14 (-0.29, 0.02)	0.76
HC z-score	-0.08 (-0.27, 0.11)	Ref	0.01 (-0.15, 0.16)	0.37
FL z-score	-0.09 (-0.27, 0.09)	Ref	-0.03 (-0.19, 0.13)	0.52
AC z-score	-0.21 (-0.28, 0.06)	Ref	-0.10 (-0.27, 0.06)	0.28
Birth weight				
Birth weight z-score	-0.14 (-0.28, 0.00)	Ref	-0.10 (-0.22, 0.02)	0.64

Note: Estimates represent the difference in mean z-score with a 5 $\mu g/m^3$ increase in $PM_{2.5}$ during the first 16 weeks of pregnancy after adjusting for maternal age, race, insurance type, parity, fetal sex, conception year, day of the year of conception, temperature, and Area Deprivation Index; P-values were computed by testing whether the association for each level of the potential modifier differs from that of the reference level, which is "Less than college" for maternal education.

Table S2.5. Linear mixed effects model estimates and 95% CIs for the association between $PM_{2.5}$ in the first 16 weeks and fetal growth outcomes (ultrasound parameters and birth weight) by Area Deprivation Index (ADI) from deliveries at Beth Israel Deaconess Medical Center, Boston, Massachusetts in 2011-2016 (N=10,008)

Growth outcome	ADI Quartile 1		ADI Quartile 2		ADI Quartile 3		ADI Quartile 4	
	Estimate (95% CI)	*p*-Value	Estimate (95% CI)	*p*-Value	Estimate (95% CI)	*p*-Value	Estimate (95% CI)	*p*-Value
Anatomic scans (<24 weeks)								
BPD z-score	-0.22 (-0.38, -0.05)	Ref	-0.11 (-0.29, 0.08)	0.31	-0.17 (-0.36, 0.01)	0.68	-0.19 (-0.41, 0.02)	0.84
HC z-score	-0.14 (-0.31, 0.02)	Ref	-0.04 (-0.23, 0.14)	0.34	-0.12 (-0.31, 0.07)	0.83	-0.01 (-0.21, 0.20)	0.25
FL z-score	-0.15 (-0.32, 0.03)	Ref	-0.01 (-0.20, 0.19)	0.21	-0.13 (-0.33, 0.07)	0.87	0.06 (-0.16, 0.28)	0.09
AC z-score	-0.18 (-0.36, 0.01)	Ref	-0.12 (-0.32, 0.09)	0.62	-0.15 (-0.35, 0.06)	0.81	0.03 (-0.20, 0.27)	0.11
Growth scans (≥24 weeks)								
BPD z-score	-0.23 (-0.41, -0.04)	Ref	-0.09 (-0.29, 0.10)	0.25	-0.15 (-0.35, 0.05)	0.51	-0.02 (-0.24, 0.19)	0.09
HC z-score	-0.11 (-0.29, 0.07)	Ref	0.09 (-0.10, 0.27)	0.07	-0.09 (-0.29, 0.11)	0.86	0.05 (-0.20, 0.22)	0.17
FL z-score	-0.07 (-0.24, 0.11)	Ref	0.01 (-0.19, 0.22)	0.49	-0.12 (-0.32, 0.08)	0.64	0.00 (-0.21, 0.21)	0.55
AC z-score	-0.19 (-0.38, -0.00)	Ref	-0.12 (-0.32, 0.07)	0.56	-0.18 (-0.39, 0.03)	0.90	-0.06 (-0.29, 0.16)	0.30
Birth weight								
Birth weight z-score	-0.17 (-0.31, -0.04)	Ref	-0.11 (-0.26, 0.04)	0.46	-0.09 (-0.24, 0.07)	0.33	-0.10 (-0.26, 0.07)	0.41

Note: Estimates represent the difference in mean z-score with a 5 $\mu g/m^3$ increase in $PM_{2.5}$ during the first 16 weeks of pregnancy after adjusting for maternal age, race, education, insurance type, parity, fetal sex, conception year, day of the year of conception, and temperature; P-values were computed by testing whether the association for each level of the potential modifier differs from that of the reference level, which is the first quartile for Area Deprivation Index.

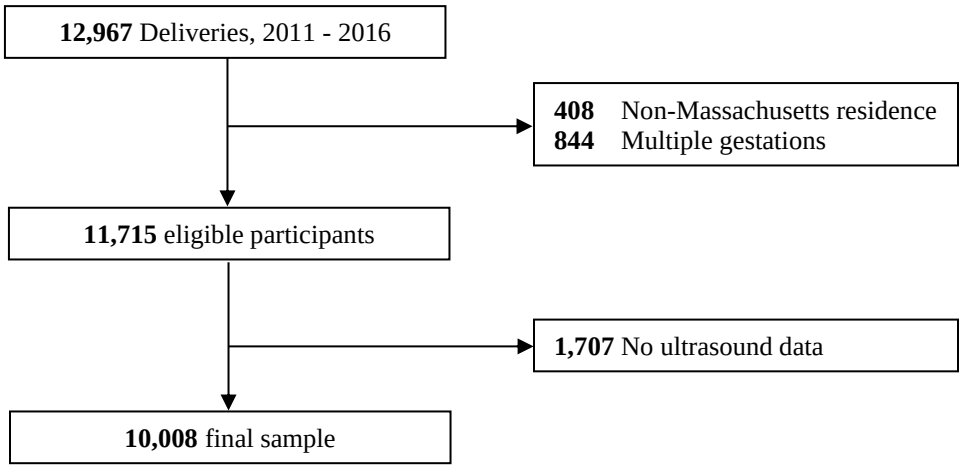

Figure S3.1. Flow chart of participant selection in deliveries at Beth Israel Deaconess Medical Center, Boston, Massachusetts from 2011-2016

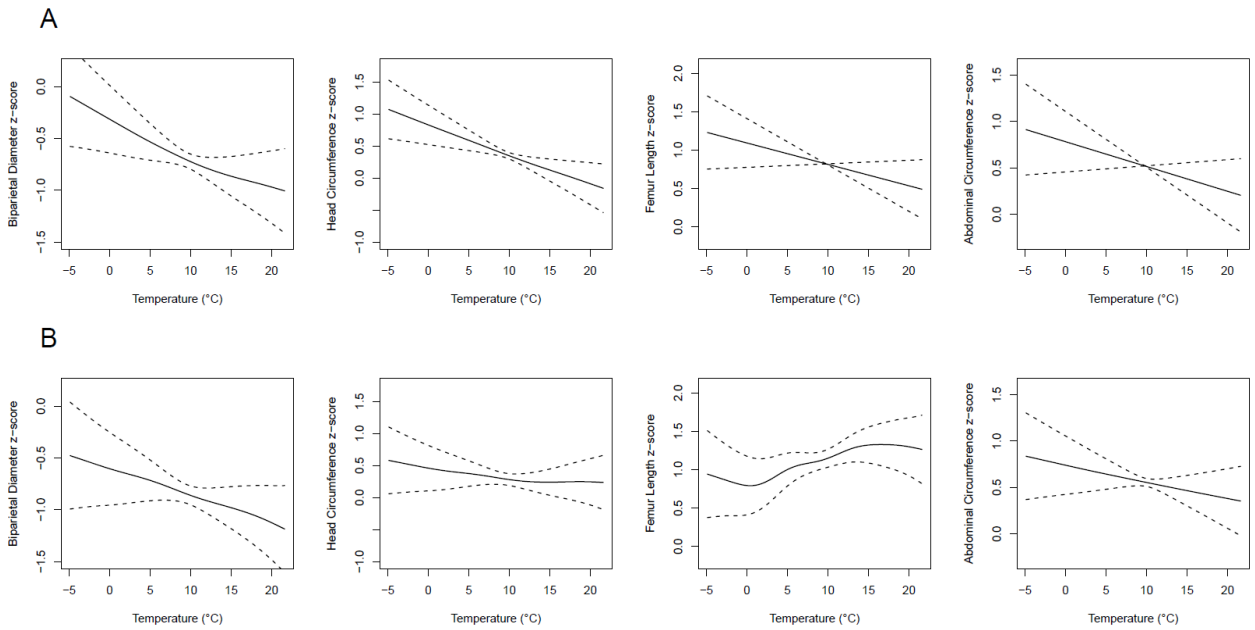

Figure S3.2. Relationship between temperature in the first 16 weeks of pregnancy and ultrasound biometric parameter z-scores estimated using a generalized additive mixed model among A) anatomic scans (<24 weeks), and B) growth scans (≥24 weeks) from deliveries at Beth Israel Deaconess Medical Center, Boston, Massachusetts in 2011-2016. The solid line represents the predicted z-score by temperature given that all other covariates are at their respective means. The dashed line represents the 95% confidence intervals. Analyses were adjusted for maternal age, race, education, insurance type, parity, fetal sex, conception year, day of the year of conception, Area Deprivation Index, humidity, and $PM_{2.5}$.

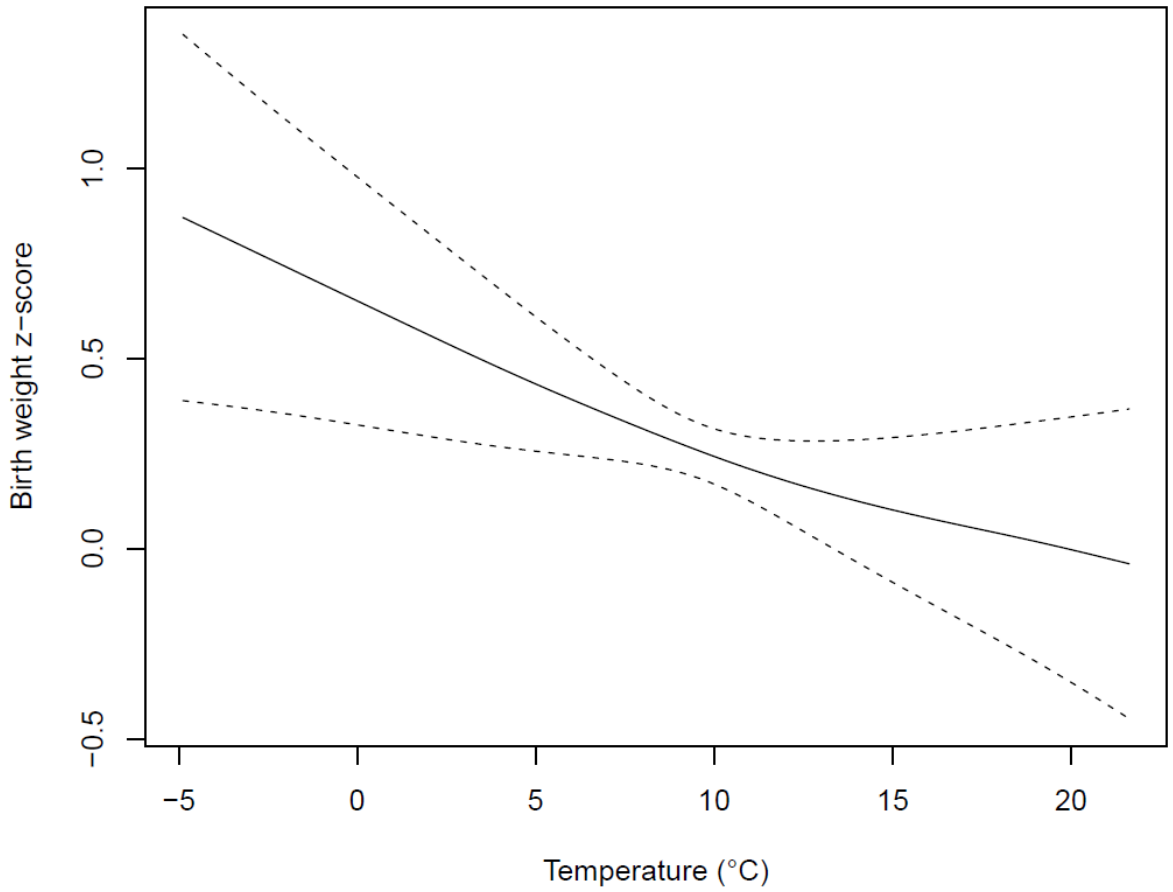

Figure S3.3. Relationship between temperature in the first 16 weeks of pregnancy and birth weight z-score estimated using a generalized additive mixed model from deliveries at Beth Israel Deaconess Medical Center, Boston, Massachusetts in 2011-2016. The solid line represents the predicted z-score by temperature given that all other covariates are at their respective means. The dashed line represents the 95% confidence intervals. Analyses were adjusted for maternal age, race, education, insurance type, parity, fetal sex, conception year, day of the year of conception, Area Deprivation Index, humidity, and $PM_{2.5}$.

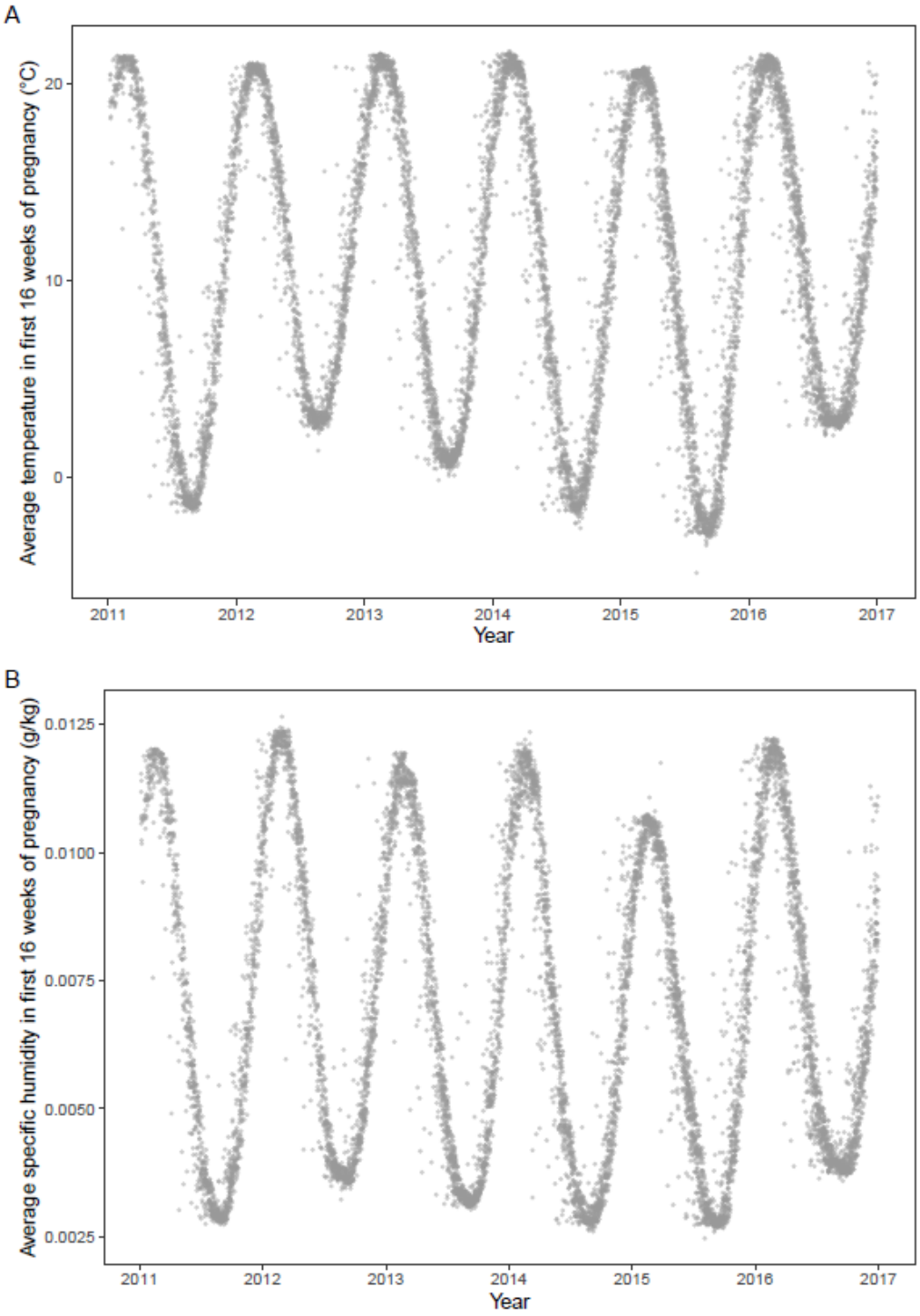

Figure S3.4. Distributions of A) average temperature in the first 16 weeks of pregnancy, and B) average specific humidity in the first 16 weeks of pregnancy over the study period (2011 – 2016). Grey points represent the mean value for each pregnancy.

Table S3.1. Linear mixed effects model estimates and 95% CIs for the association between cumulative temperature and fetal growth outcomes (ultrasound parameters and birth weight) from deliveries at Beth Israel Deaconess Medical Center, Boston, Massachusetts in 2011-2016 (N=10,008)

Growth outcome	Estimate per 5 ºC	95% CI	*p*-Value
Anatomic scans (<24 weeks)			
BPD z-score	-0.06	(-0.19, 0.06)	0.32
HC z-score	-0.11	(-0.24, 0.01)	0.07
FL z-score	-0.08	(-0.21, 0.05)	0.25
AC z-score	-0.10	(-0.23, 0.04)	0.15
Growth scans (≥24 weeks)			
BPD z-score	0.02	(-0.11, 0.14)	0.77
HC z-score	0.01	(-0.12, 0.14)	0.87
FL z-score	0.14	(0.02, 0.27)	0.03
AC z-score	-0.05	(-0.17, 0.08)	0.48
Birth weight			
Birth weight z-score	-0.16	(-0.26, -0.07)	0.001

Note: Adjusted for maternal age, race, education, insurance type, parity, fetal sex, conception year, day of the year of conception, Area Deprivation Index, humidity, and $PM_{2.5}$.

Table S3.2. Linear mixed effects model estimates and 95% CIs for the association between temperature 1 month prior to growth assessment and fetal growth outcomes (ultrasound parameters and birth weight) from deliveries at Beth Israel Deaconess Medical Center, Boston, Massachusetts in 2011-2016 (N=10,008)

Growth outcome	Estimate per 5 ºC	95% CI	*p*-Value
Anatomic scans (<24 weeks)			
BPD z-score	-0.00	(-0.09, 0.08)	0.95
HC z-score	-0.02	(-0.11, 0.06)	0.59
FL z-score	-0.03	(-0.12, 0.06)	0.54
AC z-score	-0.04	(-0.13, 0.06)	0.44
Growth scans (≥24 weeks)			
BPD z-score	-0.02	(-0.09, 0.04)	0.48
HC z-score	-0.07	(-0.14, 0.00)	0.06
FL z-score	0.04	(-0.03, 0.11)	0.24
AC z-score	-0.03	(-0.10, 0.04)	0.43
Birth weight			
Birth weight z-score	-0.03	(-0.09, 0.02)	0.31

Note: Adjusted for maternal age, race, education, insurance type, parity, fetal sex, conception year, day of the year of conception, Area Deprivation Index, humidity, and $PM_{2.5}$.

Table S3.3. Linear mixed effects model estimates and 95% CIs for the association between temperature in the first 16 weeks and fetal growth outcomes (ultrasound parameters and birth weight) by mother's age from deliveries at Beth Israel Deaconess Medical Center, Boston, Massachusetts in 2011-2016 (N=10,008)

Growth outcome	Age 22-40 years		Age <22 years		Age >40 years	
	Estimate[a] (95% CI)	*p*-Value	Estimate[a] (95% CI)	*p*-Value	Estimate[a] (95% CI)	*p*-Value
Anatomic scans (<24 weeks)						
BPD z-score	-0.18 (-0.34, -0.03)	Ref	-0.23 (-0.40, -0.05)	0.24	-0.20 (-0.37, -0.03)	0.74
HC z-score	-0.21 (-0.72, 0.29)	Ref	-0.27 (-0.79, 0.26)	0.18	-0.22 (-0.73, 0.28)	0.80
FL z-score	-0.15 (-0.31, 0.02)	Ref	-0.16 (-0.35, 0.03)	0.75	-0.14 (-0.31, 0.04)	0.80
AC z-score	-0.14 (-0.53, 0.26)	Ref	-0.16 (-0.57, 0.26)	0.63	-0.17 (-0.57, 0.23)	0.42
Growth scans (≥24 weeks)						
BPD z-score	-0.21 (-0.39, -0.02)	Ref	-0.25 (-0.45, -0.05)	0.19	-0.18 (-0.38, 0.01)	0.51
HC z-score	-0.15 (-0.33, 0.04)	Ref	-0.19 (-0.39, 0.00)	0.21	-0.11 (-0.31, 0.08)	0.31
FL z-score	-0.02 (-0.22, 0.18)	Ref	-0.03 (-0.24, 0.18)	0.81	0.07 (-0.14, 0.28)	0.01*
AC z-score	-0.17 (-0.35, 0.02)	Ref	-0.24 (-0.44, -0.04)	0.07	-0.21 (-0.41, -0.00)	0.25
Birth weight						
Birth weight z-score	-0.10 (-0.27, 0.07)	Ref	-0.15 (-0.33, 0.03)	0.13	-0.11 (-0.28, 0.06)	0.18

[a] Estimates per 5 ºC increase in temperature, adjusted for maternal race, education, insurance type, parity, fetal sex, conception year, day of the year of conception, Area Deprivation Index, humidity, and $PM_{2.5}$; P-values were computed by testing whether the association for each level of the potential modifier differs from that of the reference level, which is "22-40" for maternal age.

* Statistically significant after Bonferroni correction

Table S3.4. Linear mixed effects model estimates and 95% CIs for the association between temperature in the first 16 weeks and fetal growth outcomes (ultrasound parameters and birth weight) by maternal race from deliveries at Beth Israel Deaconess Medical Center, Boston, Massachusetts in 2011-2016 (N=10,008)

Growth outcome	White		Black		Asian		Hispanic	
	Estimate[a] (95% CI)	*p*-Value	Estimate[a] (95% CI)	*p*-Value	Estimate[a] (95% CI)	*p*-Value	Estimate[a] (95% CI)	*p*-Value
Anatomic scans (<24 weeks)								
BPD z-score	-0.18 (-0.34, -0.03)	Ref	-0.22 (-0.38, -0.06)	0.13	-0.21 (-0.38, -0.04)	0.40	-0.17 (-0.34, -0.00)	0.59
HC z-score	-0.21 (-0.36, -0.06)	Ref	-0.23 (-0.39, -0.07)	0.43	-0.24 (-0.40, -0.08)	0.30	-0.19 (-0.35, -0.02)	0.44
FL z-score	-0.15 (-0.33, 0.04)	Ref	-0.15 (-0.34, 0.04)	0.94	-0.18 (-0.37, 0.01)	0.26	-0.15 (-0.35, 0.04)	0.92
AC z-score	-0.14 (-0.31, 0.03)	Ref	-0.15 (-0.32, 0.01)	0.57	-0.18 (-0.35, -0.00)	0.24	-0.12 (-0.30, 0.06)	0.57
Growth scans (≥24 weeks)								
BPD z-score	-0.18 (-0.36, 0.00)	Ref	-0.22 (-0.40, -0.03)	0.14	-0.23 (-0.42, -0.04)	0.14	-0.24 (-0.43, -0.06)	0.03
HC z-score	-0.12 (-0.31, 0.06)	Ref	-0.17 (-0.35, 0.02)	0.08	-0.18 (-0.37, 0.01)	0.08	-0.17 (-0.36, 0.02)	0.12
FL z-score	-0.01 (-0.20, 0.17)	Ref	-0.01 (-0.20, 0.18)	0.99	-0.05 (-0.24, 0.14)	0.20	-0.01 (-0.21, 0.18)	0.97
AC z-score	-0.17 (-0.35, 0.01)	Ref	-0.17 (-0.36, 0.01)	0.87	-0.23 (-0.42, -0.04)	0.06	-0.16 (-0.35, 0.03)	0.77
Birth weight								
Birth weight z-score	-0.09 (-0.18, -0.01)	Ref	-0.10 (-0.18, -0.01)	0.80	-0.16 (-0.25, -0.07)	0.0001*	-0.10 (-0.18, -0.01)	0.86

[a] Estimates per 5 ºC, adjusted for maternal age, education, insurance type, parity, fetal sex, conception year, day of the year of conception, Area Deprivation Index, humidity, and $PM_{2.5}$; P-values were computed by testing whether the association for each level of the potential modifier differs from that of the reference level, which is "White" for maternal race.

* Statistically significant after Bonferroni correction

Table S3.5. Linear mixed effects model estimates and 95% CIs for the association between temperature in the first 16 weeks and fetal growth outcomes (ultrasound parameters and birth weight) by maternal education from deliveries at Beth Israel Deaconess Medical Center, Boston, Massachusetts in 2011-2016 (N=10,008)

Growth outcome	Less than college		College or higher	
	Estimate[a] (95% CI)	*p*-Value	Estimate[a] (95% CI)	*p*-Value
Anatomic scans (<24 weeks)				
BPD z-score	-0.20 (-0.36, -0.04)	Ref	-0.20 (-0.36, -0.04)	0.85
HC z-score	-0.23 (-0.38, -0.07)	Ref	-0.22 (-0.38, -0.07)	0.72
FL z-score	-0.16 (-0.33, 0.01)	Ref	-0.15 (-0.32, 0.02)	0.51
AC z-score	-0.16 (-0.33, 0.01)	Ref	-0.14 (-0.31, 0.03)	0.24
Growth scans (≥24 weeks)				
BPD z-score	-0.20 (-0.39, -0.02)	Ref	-0.21 (-0.39, -0.02)	0.92
HC z-score	-0.15 (-0.33, 0.03)	Ref	-0.14 (-0.33, 0.04)	0.80
FL z-score	-0.00 (-0.18, 0.18)	Ref	-0.03 (-0.22, 0.15)	0.09
AC z-score	-0.17 (-0.36, 0.02)	Ref	-0.18 (-0.36, 0.01)	0.73
Birth weight				
Birth weight z-score	-0.10 (-0.19, -0.02)	Ref	-0.11 (-0.20, -0.03)	0.38

[a] Estimates per 5 ºC, adjusted for maternal age, race, insurance type, parity, fetal sex, conception year, day of the year of conception, Area Deprivation Index, humidity, and $PM_{2.5}$; P-values were computed by testing whether the association for each level of the potential modifier differs from that of the reference level, which is "Less than college" for maternal education.

Table S3.6. Linear mixed effects model estimates and 95% CIs for the association between temperature in the first 16 weeks and fetal growth outcomes (ultrasound parameters and birth weight) by Area Deprivation Index (ADI) from deliveries at Beth Israel Deaconess Medical Center, Boston, Massachusetts in 2011-2016 (N=10,008)

Growth outcome	ADI Quartile 1		ADI Quartile 2		ADI Quartile 3		ADI Quartile 4	
	Estimate[a] (95% CI)	*p*-Value	Estimate[a] (95% CI)	*p*-Value	Estimate[a] (95% CI)	*p*-Value	Estimate[a] (95% CI)	*p*-Value
Anatomic scans (<24 weeks)								
BPD z-score	-0.19 (-0.35, -0.03)	Ref	-0.17 (-0.33, -0.02)	0.46	-0.18 (-0.34, -0.02)	0.58	-0.21 (-0.37, -0.06)	0.32
HC z-score	-0.22 (-0.38, -0.07)	Ref	-0.20 (-0.36, -0.04)	0.30	-0.22 (-0.37, -0.06)	0.74	-0.23 (-0.20, 0.11)	0.68
FL z-score	-0.15 (-0.31, 0.01)	Ref	-0.15 (-0.32, 0.02)	0.97	-0.14 (-0.31, 0.02)	0.79	-0.16 (-0.32, 0.00)	0.76
AC z-score	-0.15 (-0.32, 0.01)	Ref	-0.14 (-0.31, 0.03)	0.65	-0.17 (-0.34, 0.00)	0.58	-0.14 (-0.31, 0.03)	0.55
Growth scans (≥24 weeks)								
BPD z-score	-0.19 (-0.37, -0.00)	Ref	-0.20 (-0.39, -0.02)	0.61	-0.20 (-0.38, -0.01)	0.67	-0.21 (-0.40, -0.02)	0.40
HC z-score	-0.13 (-0.31, 0.06)	Ref	-0.14 (-0.33, 0.04)	0.52	-0.13 (-0.31, 0.06)	0.96	-0.16 (-0.22, 0.15)	0.28
FL z-score	-0.02 (-0.20, 0.16)	Ref	-0.01 (-0.20, 0.17)	0.88	0.00 (-0.18, 0.19)	0.34	-0.02 (-0.20, 0.16)	0.85
AC z-score	-0.17 (-0.36, 0.01)		-0.18 (-0.37, 0.00)	0.69	-0.18 (-0.36, 0.00)	0.73	-0.17 (-0.35, 0.02)	0.88
Birth weight								
Birth weight z-score	-0.10 (-0.30, 0.09)	Ref	-0.11 (-0.31, 0.09)	0.76	-0.12 (-0.32, 0.07)	0.39	-0.10 (-0.30, 0.09)	0.96

[a] Estimates per 5 ºC, adjusted for maternal age, race, education, insurance type, parity, fetal sex, conception year, day of the year of conception, humidity, and $PM_{2.5}$; P-values were computed by testing whether the association for each level of the potential modifier differs from that of the reference level, which is the first quartile for Area Deprivation Index.

Table S3.7. Linear mixed effects model estimates and 95% CIs for the association between temperature in the first 16 weeks and fetal growth outcomes (ultrasound parameters and birth weight) by fetal sex from deliveries at Beth Israel Deaconess Medical Center, Boston, Massachusetts in 2011-2016 (N=10,008)

Growth outcome	Female		Male	
	Estimate[a] (95% CI)	*p*-Value	Estimate[a] (95% CI)	*p*-Value
Anatomic scans (<24 weeks)				
BPD z-score	-0.20 (-0.38, -0.02)	Ref	-0.20 (-0.38, -0.01)	0.93
HC z-score	-0.23 (-0.38, -0.07)	Ref	-0.22 (-0.38, -0.07)	0.87
FL z-score	-0.15 (-0.32, 0.01)	Ref	-0.15 (-0.31, 0.01)	0.94
AC z-score	-0.16 (-0.33, 0.02)	Ref	-0.15 (-0.32, 0.03)	0.52
Growth scans (≥24 weeks)				
BPD z-score	-0.21 (-0.40, -0.01)	Ref	-0.20 (-0.40, -0.00)	0.74
HC z-score	-0.14 (-0.33, 0.04)	Ref	-0.14 (-0.32, 0.04)	0.91
FL z-score	-0.00 (-0.19, 0.18)	Ref	-0.03 (-0.21, 0.16)	0.16
AC z-score	-0.18 (-0.36, 0.00)	Ref	-0.16 (-0.35, 0.02)	0.39
Birth weight				
Birth weight z-score	-0.11 (-0.20, -0.03)	Ref	-0.10 (-0.19, -0.02)	0.25

[a] Estimates per 5 ºC, adjusted for maternal age, race, education, insurance type, parity, conception year, day of the year of conception, Area Deprivation Index, humidity, and $PM_{2.5}$; P-values were computed by testing whether the association for each level of the potential modifier differs from that of the reference level, which is "female" for fetal sex.