Sensitivity analysis for live birth bias in the Ulaanbaatar Gestation and Air Pollution Research study

Lawrence C. McCandless, Enkhjargal Gombojav, Ryan W. Allen

Background: The Ulaanbaatar Gestation and Air Pollution Research (UGAAR) study is a randomized controlled trial of the effects of portable high efficiency particulate air (HEPA) filter air cleaner use during pregnancy on fetal growth and child health outcomes. A challenge with the study is that the outcome variables cannot be measured in the absence of a live birth. Thus pregnancy loss is a competing event for the outcome variables that can obscure the intervention-outcome relationship. This phenomenon is called “live birth bias” in the epidemiological literature, and it is an example of selection bias due to adjustment for variables affected by previous exposure.

Methods: In this investigation, we reanalyzed data from the UGAAR study and examined the impacts of the intervention on three health outcomes: preterm birth (PTB), birth weight, and full-scale IQ (FSIQ) measured with the Wechsler Preschool and Primary Scale of Intelligence III when children were four years old, while accounting for live birth bias. Specifically, we used a novel multiple imputation approach to simulate scenarios in which pregnancy losses had instead been born alive and experienced a range of health outcomes.

Results: After accounting for live birth bias, the association between the intervention and PTB diminished. Additionally, the magnitude of intervention effect on birth weight and FSIQ increased. FSIQ was less sensitive to live birth bias than birth weight.

Conclusion: We introduced a novel analysis approach to examine the role of live birth bias, and the findings will be useful in environmental epidemiology studies of birth cohorts.

Keywords: Birth cohort; Live birth bias; Neurodevelopment; Pregnancy outcomes; Sensitivity analysis

Introduction

The Ulaanbaatar Gestation and Air Pollution Research (UGAAR) study is a randomized controlled trial of the effects of portable HEPA filter air cleaner use during pregnancy on fetal growth, and child health outcomes in Mongolia’s highly polluted capital city.1–4 We found that the HEPA cleaner intervention had little effect on birth weight among all live births (23 g difference in mean birth weight; 95% CI: −79, 125 g).1 However, after adjusting for preterm birth (PTB), defined as birth at less than 37 weeks gestation, we found that the mean birth weight in the intervention group was 88 g (95% CI: 3, 173 g) greater than in the control group. This difference between the unadjusted and adjusted effect estimates was driven by an unexpected result; the HEPA cleaner intervention was associated with higher frequency of PTB (odds ratio = 2.3; 95% CI: 1.1, 5.0).

In a subsequent analysis of children’s cognitive performance at 4 years of age, we found that the HEPA cleaner intervention was associated with a 2.8-point increase (95% CI: 0.0, 5.7) in mean full-scale IQ (FSIQ).5 The effect estimate increased (3.2 points; 95% CI: 0.4, 6.0) after adjusting for PTB.

As is typical of birth cohort studies, our analyses of child health outcomes were restricted to live births (Figure 1).7,8 Women who experienced a pregnancy loss from spontaneous abortion or stillbirth were excluded from the analysis, because in these cases, we did not have data on fetal weight or FSIQ. As illustrated in Figure 1, the intervention was associated with a lower risk of pregnancy loss (OR = 0.6; 95% CI: 0.3, 1.1). Moreover, pregnancy loss shares common causes with PTB, such as birth defects and maternal stress.9 Thus restricting the analysis to live births may have biased the results. We
Hypothesized that infants selected into the intervention group had higher exposure to unmeasured risk factors for PTB. These factors would act like unmeasured confounders and distort the association between the intervention and PTB. This phenomenon is called live birth bias (LBB) in the epidemiology literature. It is an example of selection bias due to stratification on a variable that is affected by the exposure and outcome, or upstream causes of the exposure and outcome. The bias has been observed in other cohort studies of pregnant women leading to paradoxical effect estimates. For example, Raz et al. describe a study in which higher concentrations of air pollution were associated with lower rates of autism spectrum disorder. In another example, Liew et al. described an unexpected protective association between perfluoroalkyl substance exposure during pregnancy and child behavior problems or ADHD diagnosis. Other examples of LBB appear in the birth defects literature.

Here, we sought to reanalyze data from the UGAAR study and examine the impacts of the intervention on three health outcomes: PTB, birth weight, and FSIQ, while accounting for LBB. Specifically, we used a novel multiple imputation (MI) approach to simulate scenarios in which pregnancy losses had instead been born alive and experienced a range of health outcomes. To our knowledge, MI has not been previously used to examine the influence of LBB.

Methods

The UGAAR study design and methods were published previously. We conducted the study in Ulaanbaatar, Mongolia’s largest city. Ulaanbaatar has relatively high concentrations of fine particulate matter (PM2.5) air pollution, particularly during colder months, due primarily to emissions from coal combustion in home heating stoves. In summary, we recruited 540 nonsmoking pregnant women from January 2014 to May 2015. Two hundred seventy-two (272) participants were randomly assigned to the control group and 268 participants were randomly assigned to the intervention group (Figure 1). Participants were enrolled in the first 18 weeks of pregnancy. The intervention group received one or two HEPA cleaners depending on home size. The control group received no HEPA cleaners. We collected the HEPA cleaners shortly after pregnancy ended. We did not use sham filtration, so participants were not blinded to intervention status, but those who measured childhood health outcomes were blinded. The HEPA cleaners reduced mean indoor PM2.5 concentrations by 29% (95% CI: 21%, 37%) during pregnancy, from a geometric mean (GM) of 24.3 μg/m3 in control homes to 17.3 μg/m3 in intervention homes.

We collected data on participants’ health, behavior, and lifestyle via questionnaires administered shortly after enrollment and later in pregnancy. After birth, we obtained birth weight, gestational age, sex, and other data from clinic records. Gestational age was determined from a combination of first trimester ultrasound, last menstrual period, and clinical assessment (symphysis-fundal height measurements and Dubowitz or Ballard score). We also collected information from clinic records on stillbirths, pregnancy complications, and comorbidities. Participants self-reported the occurrence and timing of spontaneous abortions.

The details of cognitive testing in the UGAAR study were published previously. In summary, we measured FSIQ when the children were a median of 48 months old (range: 48–51 months) using the Wechsler Preschool and Primary Scale of Intelligence, Fourth Edition (WPPSI-IV). The WPPSI-IV has been validated for measurement of cognitive functioning in children ages 2 years and 6 months to 7 years and 7 months.

There were 47 known pregnancy losses (29 control and 18 intervention), and 468 live births (225 control and 243 intervention) (Figure 1). For the purpose of this analysis, we ignored the 25 individuals who were lost to follow-up or excluded before

Figure 1. Trial profile for the UGAAR Study.
the end of pregnancy (e.g., who withdrew or who moved out of the study area).

In this analysis, we evaluated the role of the LBB in the UGAAR study by comparing results from three data analysis approaches. First, we repeated previous analyses restricted to live births, as is typical in the birth cohort studies.20-22 We used linear regression or logistic regression analysis to examine the effect of the intervention on three health outcomes: (1) PTB, (2) birth weight measured in grams, and (3) FSIQ. The analyses of PTB and birth weight included 465 participants after excluding three live births (two with a chromosomal abnormality and one with unknown gestational age at birth and birth weight), while the analysis of FSIQ included 383 participants who completed the WPPSI-IV. Two participants who withdrew from the study during pregnancy re-enrolled during postnatal follow-up. As a result, the birth results shown here differ slightly from the complete case birth weight results presented in Barn et al.1

Next, following Barn et al.1 and Ulziiikhuu et al.,4 we crudely corrected the imbalance in PTBs between intervention arms by estimating the effect of the intervention on birth weight and FSIQ adjusted for PTB. We also calculated the effect of the intervention among term births (≥37 weeks gestation). However, the use of stratification or adjustment for PTB to limit LBB is problematic. As discussed by Hernan,23 this method can introduce bias because it entails adjustment for variables on the causal pathway between the intervention and the outcome.

Therefore, in our third analysis approach we used MI to simulate scenarios in which pregnancy losses had instead been born alive and experienced a range of health outcomes. The approach is an example of sensitivity analysis for missing not at random (MNAR) data that examines how different assumptions influence the analysis results.24 Specifically, we assigned a value q to the probability of PTB among the pregnancy losses. The quantity q is a bias parameter that varies between 0% and 100%. For a given value of q (e.g., q = 50%), we then randomly assigned the binary PTB outcome to the 47 pregnancy losses. Conditional on the assigned values of PTB among the pregnancy losses, the missing values for birth weight and FSIQ were assumed to be MAR. Next, we applied MI to the dataset using the R package MICE to impute the missing values for birth weight and FSIQ, conditional on the assigned values for PTB.25 MICE imputes missing values using an iterative series of prediction models for the missing data. We generated 100 imputations using default settings for modeling missing variables, which used predictive mean matching for continuous variables and logistic models for categorical variables. We used Rubin’s rules for combining the multiply imputed estimates. We also imputed gestational age at birth and birth weight for one participant and FSIQ for the live births that did not complete the WPPSI-IV assessment. Furthermore, Table 2 illustrates that compared with term births, preterm infants had lower mean (±SD) birth weight (2,356 g ± 780 g vs. 3,545 g ± 432 g) and FSIQ (91.5 ± 12.6 vs. 99.4 ± 14.2 g). Upon adjustment for PTB, the magnitude of the intervention effect estimates increased to 88 g (95% CI: 3, 173) for mean birth weight and 3.2 (95% CI: 0.4, 6.0) for mean FSIQ.

Table 2 shows the effect of the HEPA air cleaner intervention on PTB and infant birth weight in an analysis restricted to 465 live births, and FSIQ in an analysis restricted to 383 children who completed the WPPSI-IV. Consistent with the results reported by Barn et al.1, the intervention was associated with a 2.3-fold (95% CI: 1.1, 5.0) higher odds of PTB and a 23 gram (95% CI: −79, 125 g) difference in mean birth weight. Among 383 children who remained in the study until age 4 and completed the WPPSI-IV cognitive assessment, the intervention was associated with a 2.8-point (95% CI: 0.0, 5.7) increase in mean FSIQ.

Table 1. Summary statistics by treatment group for 513 UGAAR study participants (465 live births and 47 pregnancy losses) included in a multiple imputation sensitivity analysis of live birth bias

| Enrollment season | Control (n = 253) | Intervention (n = 260) |
|-------------------|------------------|-----------------------|
| No                | 233 (92.1)       | 239 (91.9)            |
| Yes               | 17 (6.7)         | 18 (6.9)              |
| Not reported, N (%) | 3 (1.2)        | 3 (1.2)               |

| Maternal smoking at enrollment | Control (n = 253) | Intervention (n = 260) |
|-------------------------------|------------------|-----------------------|
| No                            | 69 (27.3)        | 58 (22.3)             |
| Yes                           | 173 (68.4)       | 194 (74.6)            |
| Not reported, N (%)           | 11 (4.3)         | 8 (3.1)               |

| Monthly household income at enrollment (Tugrik) |
|-----------------------------------------------|
| <800,000                                      |
| 800,000 to 1,199,999                           |
| ≥1,200,000                                    |
| Not reported, N (%)                           |
| 10 (4.0)                                     |
| 10 (3.8)                                     |

| Parity | Control (n = 253) | Intervention (n = 260) |
|--------|------------------|-----------------------|
| 0      | 73 (28.9)        | 82 (31.5)             |
| 1      | 99 (39.1)        | 93 (35.8)             |
| ≥2     | 60 (23.7)        | 66 (25.4)             |
| Not reported, N (%) | 21 (8.5) | 19 (7.3)               |

| Pregnancy outcome | Control (n = 253) | Intervention (n = 260) |
|-------------------|------------------|-----------------------|
| Live birth        | 224 (88.5)       | 242 (95.7)            |
| Spontaneous abortion | 24 (9.5)   | 10 (4.0)              |
| Stillbirth        | 5 (2.0)          | 8 (3.2)               |

| Child’s sex (live births) | Female | Male | Unknown, N (%) |
|---------------------------|--------|------|---------------|
|                           | 111 (49.6) | 111 (45.9) | 0 (0.0) |

Results

Among the 513 participants included in this sensitivity analysis, control and intervention participants had similar characteristics (Table 1). However, as we have described in previous publications, there were fewer spontaneous abortions in the intervention group.1-4

In effect, this approach used relationships between PTB, birth weight, and FSIQ among live births to impute values of birth weight and FSIQ among pregnancy losses based on PTB, which is governed by q. In order to reliably impute birth weight and FSIQ, we included additional variables in the imputation model that were associated with birth weight or FSIQ including: the intervention, season of enrolment, maternal age at birth, pre-pregnancy BMI, monthly household income, maternal smoking at baseline, vitamin intake at baseline, parity, and sex of the baby. We used a total of 100 MIs.
These effect estimates were similar to those obtained from an analysis restricted to term births (Table 2). This demonstrates the influence of LBB on the results: PTB acts like a confounder for the effect of the intervention on birth weight and FSIQ. However stratification and adjustment for PTB does not eliminate LBB because it conditions on variables affected by previous exposure.23 Figure 2 presents a directed acyclic graph (DAG) for the relationship between variables in the UGAAR study, which is similar to the figures given in Raz et al10 and Leung et al.12 The DAG is simplified to make it easier to grasp and includes unmeasured common causes of pregnancy loss and PTB, such as maternal stress or infection. Restricting the analysis to live births opens a backdoor path from the intervention to PTB, birth weight, and FSIQ. For the analysis of birth weight and FSIQ, conditioning on PTB blocks the backdoor path. However, it also blocks the causal effect of the intervention on birth weight or FSIQ that is mediated by PTB.

Figure 3 presents the results of the sensitivity analysis illustrating the impact of LBB as a function of q, which varies from 0 to 1. After accounting for LBB, the log odds ratio for the effect of the intervention on PTB shifted toward zero. For example, when q = 0.25, which implies that 25% of pregnancy losses would have resulted in PTBs, the 95% confidence interval for PTB covered zero. Figure 3 illustrates that as q increases from 0 to 1 the effect of the intervention on birth weight and FSIQ also increases. Furthermore, birth weight was more sensitive to LBB than FSIQ. For example, the point estimate of the intervention

### Table 2

|                   | Control (n = 223) | Intervention (n = 242) |
|-------------------|------------------|------------------------|
|                   | n, mean %, SD    | n, mean %, SD          | Effect (odds ratio or mean difference) |
| Preterm births    | 10 4.5%          | 24 9.9%                | 2.3 (1.1, 5.0) |
| Birthweight       |                  |                        |                                    |
| All live births   | 3446 529         | 3469 585               | 23 (-79, 125)                      |
| Term births       | 3500 453         | 3589 407               | 89 (8, 171)                        |
| Preterm births    | 2305 756         | 2376 805               |                                               |
| Effect of intervention on birthweight regression adjusted for PTB | | | |
| Control (n=182)   |                  | Intervention (n=201)   |
| FSIQ              |                  |                        | 88 (3, 173) |
| All live births   | 97.3 14.9        | 100.2 13.4             | 2.8 (0.0, 5.7)                      |
| Term births       | 97.7 14.9        | 101.0 13.2             | 3.3 (0.4, 6.3)                      |
| Preterm births    | 90.8 14.6        | 91.8 12.0              | -                                    |
| Effect of intervention on IQ regression adjusted for PTB | | | 3.2 (0.3, 6.0) |

We obtained gestational age at birth and birthweight for two participants who withdrew from the study during pregnancy but re-enrolled during postnatal follow-up. As a result, the birth results shown here for 465 participants differ slightly from the complete case birthweight results for 463 participants presented in Barn et al.1
effect on birth weight changed by 219% from 26g when \( q = 0 \) to 83g when \( q = 1 \), whereas for FSIQ the point estimates changed by only 19% from 2.7 when \( q = 0 \) to 3.2 when \( q = 1 \). For birth weight and FSIQ, there was wide uncertainty in the 95% confidence intervals.

**Discussion**

In this paper, we demonstrate the influence of LBB on the results of the UGAAR study with respect to three health outcomes: PTB, birth weight, and FSIQ. After accounting for LBB, the association between the intervention and PTB diminished. Additionally, the magnitude of intervention effect on birth weight and FSIQ increased. We also found that FSIQ was less sensitive to LBB than birth weight. This finding is also illustrated in Table 2 by the fact that regression adjustment for PTB had a relatively smaller impact on the effect estimate for FSIQ than for birth weight. One explanation for this finding is because PTB is a stronger predictor for birth weight than FSIQ. Table 2 illustrates that the mean birth weight of preterm infants is roughly two standard deviations lower than that of term infants. Preterm infants also have lower mean FSIQ than term infants, but the difference is only roughly one standard deviation.

In contrast to our approach, most of the existing methods to account for LBB use statistical models for the probability of selection.2,3 For example, Leung et al12 and Liew et al11 simulated the magnitude of bias under various selection mechanisms that model the prevalence of hypothetical unmeasured factors and selection effects. This approach requires that the investigator specify the association between the unmeasured factor, the probability of selection, and the outcome variable. In practice, there may be multiple unmeasured factors, and these associations may be poorly understood. In contrast, our method treats LBB as missing data that is MNAR. Rather than using a selection model, we use a pattern-mixture model, which requires the user to specify the probability distribution of the outcome variable in the missing data.16,17 In UGAAR, this distribution is represented by the quantity \( q \), which is the probability of PTB in the pregnancy losses. Our method has the advantage that it is more straightforward to implement because it models the outcome variable directly without alluding to unmeasured variables. A similar analysis strategy has been used in bias analysis for unmeasured confounding where the investigator models the distribution of potential outcomes directly without modeling the distribution and behavior of the underlying unmeasured confounders.21

Our statistical method to account for LBB makes several simplifying assumptions. First, when simulating PTB in the pregnancy losses, we assumed that the probability of PTB does not depend on any measured covariates, such as maternal smoking. This assumption is probably incorrect. To account for uncertainty in specifying \( q \) in the sensitivity analysis, Figure 3 presented results over a range of different values for \( q \). A further assumption of our sensitivity analysis was that the impact of LBB on birth weight and FSIQ was mediated entirely by PTB. This assumption is reflected in Figure 2 with the absence of additional unmeasured causes of pregnancy loss that influenced birth weight and FSIQ directly. This assumption is likely to be false. For example, growth restriction during fetal development is associated with a higher risk of pregnancy loss, and is also an important predictor of birth weight. Finally, for simplicity, we assumed that PTB was binary, but our method could easily be extended to impute PTB as an ordinal variable in order to account for the fact that very preterm infants (e.g., <32 weeks gestation) do worse with regard to birth weight and FSIQ compared to late preterm infants (e.g., 35 to 37 weeks gestation). We can also adjust for gestational age as a continuous variable using splines.26

Our study has several strengths for illustrating LBB in birth cohort studies. The intervention was demonstrably associated with a higher probability of a live birth compared to control. Furthermore, there was an excess of PTBs in the intervention group. Because the UGAAR study uses a randomized design, this minimizes the influence of measured and unmeasured confounders when comparing outcomes in the intervention versus control. Thus the mostly likely explanation for this observed increase in PTB among the intervention group participants is LBB. However, caution is warranted in interpreting the results for birth weight and FSIQ due to the wide and overlapping confidence intervals for effect estimates. In contrast, in other published investigations of LBB the association between exposure and pregnancy loss has been more modest and more speculative.10,11 To summarize, we have illustrated a novel sensitivity analysis to examine the role of LBB, and the findings will be useful in future environmental epidemiology studies using pregnancy and birth cohorts.

**Acknowledgments**

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

**References**

1. Barn P, Gombojav E, Ochir C, et al. The effect of portable HEPA filter air cleaner use during pregnancy on fetal growth: the UGAAR randomized controlled trial. Environ Int. 2018;121:981–989.
2. Barn P, Gombojav E, Ochir C, et al. Coal smoke, gestational cadmium exposure, and fetal growth. Environ Res. 2019;179:108830.
3. Barn P, Gombojav E, Ochir C, et al. The effect of portable HEPA filter air cleaners on indoor PM2.5 concentrations and second hand tobacco smoke exposure among pregnant women in Ulaanbaatar, Mongolia: the UGAAR randomized controlled trial. Sci Total Environ. 2018;615:1379–1389.
4. Tamana SK, Gombojav E, Kanlic A, et al. Portable HEPA filter air cleaner use during pregnancy and children’s body mass index at two years of age: the UGAAR randomized controlled trial. Environ Int. 2021;156:106728.
5. Ekhbat U, Gombojav E, Banzrai C, et al. Portable HEPA filter air cleaner use during pregnancy and children’s behavior problem scores: a secondary analysis of the UGAAR randomized controlled trial. Environ Health. 2021;20:78.
6. Ulzakhuu B, Gombojav E, Banzrai C, et al. Portable HEPA filter air cleaner use during pregnancy and children’s cognitive performance at four years of age: the UGAAR Randomized Controlled Trial. Environ Health Perspect. 2022;130:67006.
7. Chiu Y-H, Stensrud MJ, Dahabreh IJ, et al. The effect of prenatal treatments on offspring events in the presence of competing events: an application to a randomized trial of fertility therapies. Epidemiology. 2020;31:636–643.
8. Tchetgen Tchetgen EJ, Phiri K, Shapiro R. A simple regression-based approach to account for survival bias in birth outcomes research. Epidemiology. 2015;26:473–480.
9. Stein Z, Susser M, Warburton D, et al. Spontaneous abortion as a screening device. The effect of fetal survival on the incidence of birth defects. Am J Epidemiol. 1975;102:275–290.
10. Riaz R, Kioumourtzoglou M-A, Weisskopf MG. Live-birth bias and observed associations between air pollution and autism. Am J Epidemiol. 2018;12:2292–2296.
11. Liew Z, Olsen J, Cui X, et al. Bias from conditioning on live birth in pregnancy cohorts: an illustration based on neurodevelopment in children after prenatal exposure to organic pollutants. Int J Epidemiol. 2015;44:345–354.
12. Leung M, Kioumourtzoglou M-A, Raz R, et al. Bias due to selection on live births in studies of environmental exposures during pregnancy: a simulation study. Environ Health Perspect. 2021;129:47001.
13. Hernán MA, Hernández-Díaz S, Werler MM, et al. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. Am J Epidemiol. 2002;155:176–184.
14. Hook EB. Incidence and prevalence as measures of frequency of birth defects. Am J Epidemiol. 1982;116:743–747.
15. Mason CA, Kirby RS, Sever LE, et al. Prevalence is the preferred measure of frequency of birth defects. Birth Defects Res Clin Mol Teratol. 2005;73:690–692.
16. Carpenter J, Kenward M. Multiple Imputation and its Application. John Wiley and Sons; 2012.
17. Little RJA, Rubin DB. Statistical Analysis with Missing Data. John Wiley & Sons; 2014.
18. Allen RW, Gombojav E, Barkhasragchaa B, et al. An assessment of air pollution and its attributable mortality in Ulaanbaatar, Mongolia. Air Qual Atmos Health 2013;6:137–150.
19. Hill LD, Drew Hill L, Edwards R, et al. Health assessment of future PM2.5 exposures from indoor, outdoor, and secondhand tobacco smoke concentrations under alternative policy pathways in Ulaanbaatar, Mongolia. PLoS One. 2017;12:e0186834.
20. Hu JMY, Arndtke TE, Janssen P, et al. Prenatal exposure to endocrine disrupting chemical mixtures and infant birth weight: a Bayesian analysis using kernel machine regression. Environ Res. 2021;195:110749.
21. 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;16:457.
22. Woods MM, Lanphear BP, Braun JM, et al. Gestational exposure to endocrine disrupting chemicals in relation to infant birth weight: a Bayesian analysis of the HOME Study. Environ Health. 2017;16:115.
23. Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. Epidemiology. 2004;15:615–625.
24. van Buuren S, Groothuis-Oudshoorn K. mice: multivariate imputation by chained equations inR. J Stat Softw. 2011;45:1–67.
25. Vanderweele TJ, Arah OA. Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments, and confounders. Epidemiology 2011;22:42–52.
26. Harley KG, Chevrier J, Aguilar Schall R, et al. Association of prenatal exposure to polybrominated diphenyl ethers and infant birth weight. Am J Epidemiol. 2011;174:885–892.