Air pollution and birth weight: Evidence from extremely polluted places*

Maxim Massenkoff
Naval Postgraduate School

July 2023

Working Paper

Abstract

I document a surprising resilience to air pollution in the world’s most polluted places, at least as measured by birth outcomes. Average birth weight and the incidence of low birth weight in some of the most polluted cities in the world, and in highly polluted US counties in the 1970s, are essentially identical to the contemporary US. This is puzzling since quasi-experimental studies find large negative impacts of air pollution on fetal health. I discuss several possible explanations.

*Email: maxim.massenkoff@nps.edu. Jalaluddin Khan provided excellent research assistance. I thank to Gantuya Dorj for help understanding Ulaanbaatar birth data. For helpful feedback and comments, I thank Janet Currie, Stefano DellaVigna, Nathan Franz, Jacob Marcus, David Molitor, Joshua Rosenberg, Evan Rose, Aartik Sarma, Dean Spears, Alex Tabarrok, Christine Tsang, and Reed Walker. All errors are my own.
1 Introduction

The effects of air pollution are hard to measure because exposure is not random. For example, some research finds that poor people are more likely to live in polluted areas (e.g., Jbaily et al., 2022). Several studies on birth weight and pollution use quasi-experimental approaches to get around this issue. Alexander and Schwandt (2022) use purchases of high-pollution Volkswagens as an instrumental variable and find particulate matter has large negative impacts on birth weight. Currie and Walker (2011) find that the installation of automated toll booths increased birth weight in the surrounding areas, implying a large effect of nitrogen dioxide (NO$_2$) and other pollutants. These results carry significant implications for public health.

Consider the cross-sectional variation in air quality within the United States. Moving from a 10th to a 90th percentile county in terms of PM2.5 would more than double exposure. The results in Alexander and Schwandt (2022) suggest that this change would increase the chance of low birth weight by 40% and decrease birth weight by 142 grams—about half the 300-gram decrease observed during the Dutch famine (Stein and Susser, 1975). The effects in Currie and Walker (2011) are even larger.\textsuperscript{2}

Testing these predictions using birth outcomes in the contemporary US would not be helpful: the US has a relatively narrow range of pollution, and it’s easy to imagine that people in more polluted places are selected, so confounds could mask the impact. However, because of the fundamentally biological mechanism and large estimates, extremely polluted cities elsewhere in the world may complement the quasi-experimental approaches. Highly polluted cities like Beijing or Belgrade are poorer and have worse economic and health outcomes along many dimensions, so it seems reasonable to expect them to exhibit clear harm from pollution.

In this paper, I collect data on birth weights in the most polluted places in the world and on US counties that suffered high rates of pollution in the 1970s. In contrast to what might be expected from

\textsuperscript{1}Alexander and Schwandt (2022, Figure 6(a)) provide a useful summary of estimates of the effect on infant health. The automated toll booths decreased pollution through their impact on idling cars. Only NO$_2$ was consistently measured, but the change likely decreased other pollutants as well.

\textsuperscript{2}Estimates of the county-level distribution of PM2.5 exposure are population-weighted and use the county-level data on PM2.5 exposure for the years 2000-2016 from Wu et al. (2020), which is based on satellite estimates from Van Donkelaar et al. (2019) (as in Alexander and Schwandt (2022)). For calculating the percent change in low birth weight, I use the mean incidence of 6.7% calculated using 2019 US natality data (NCHS, 2023). The Currie and Walker (2011) estimates are summarized in Alexander and Schwandt (2022, Table A.9).
In the modern causal estimates, I uncover three surprises: (1) Birth weights, measured using average birth weight and low birth weight incidence, are normal in some of the world’s most polluted cities. For instance, babies born in Beijing, where particulate matter is 5-8 times as bad as in the United States, weigh about the same as American newborns. (2) Infants born in the most polluted US counties in the 1970s (e.g., Allegheny County, Pennsylvania), where pollution was twice as high as the most polluted counties today, also have normal birth weights. (3) Despite sizeable decreases in ambient PM2.5, NO₂, and other contaminants (EPA, 2022b,a), birth weights in these previously highly-polluted US counties have been stable over time.

I consider several explanations for why the birth weights in these cities are higher than expected given the causal estimates, and why birth weights haven’t improved in the formerly polluted parts of the US. The main question is whether, despite the observed normal birth weights, pollution is still quite harmful and something either masks its effects or makes these contexts unsuitable for the question.

One issue is the measurement of exposure in the international sample. The city PM2.5 estimates come from ground-level sensors (WHO, 2016b). These may not be informative of experienced pollution for the typical person: some cities in the sample cover a large area, and people spend most of their time indoors. But newer estimates integrating satellite measurements and population weights also document high levels of particulates in these places (WHO, 2023; Shaddick et al., 2018). If, despite this, most people in the highly polluted samples actually experience similar levels of exposure to the US, it would have significant implications for how we quantify the damages of air pollution. These measurements undergird estimates of the global burden of pollution (e.g., Landrigan et al., 2018; WHO, 2016a).

Further, where available, the data suggests that these birth samples are exposed to more pollution. The birth weight source from Ulaanbaatar documents high levels of indoor and outdoor exposure, measured for each subject individually (Barn et al., 2018a), and some of the birth samples are city-wide, so within-city selection cannot be biasing exposure. And studies employing wearable sensors also suggest that personal pollution exposure is higher in highly polluted cities. Even assuming that pollution is overestimated by 2-3x in these places, or assuming non-linear impacts (Miller et al., 2021), the collected birth weight statistics would exceed the causal predictions. It would be even less likely
that these issues apply in the historical US sample, where it seems widely acknowledged that the harms from pollution were far greater than today (e.g., Chay and Greenstone, 2003).

Another potential explanation is selection. Owing to the lower levels of economic development and worse health outcomes in the international sample, selection effects would seem to predict lower birth weights. And within the US data, maternal demographics do not predict a decrease in birth weight over time. Changing obstetric practices may have worked to decrease birth weight in the US (Tilstra and Masters, 2020; VanderWeele et al., 2012). Also, the US and other developed countries use a lower threshold for the viability of very premature infants, which will lower its average birth weight relative to low- and middle-income countries. But in either case, adjusting for this using US data still leaves a large gap between observed and predicted birth outcomes.

A related explanation is that extreme pollution induces a culling effect, disproportionately removing infants with worse birth outcomes from the population. Culling effects are difficult to assess but seem unlikely. First, culling does not seem to be happening during gestation or conception: The available data suggests that miscarriage rates are not much higher in polluted places and that in vitro fertilisation success rates (a more controlled setting for studying the difficulty of conception) in Beijing resemble those in Europe. Second, such culling effects would contradict a vast literature which has never documented that pollution increases aggregate birth weight.

A new and separate strain of experimental evidence also sheds light on this issue: randomized controlled trials of air purifiers and cooking technologies. Such interventions drastically reduce pollution. These trials measure pollution exposure more precisely with sensors placed on participants or in their homes, registering large effects of the interventions on personal PM 2.5 exposure. While the quasi-experimental evidence predicts impacts on birth weight in response to these declines in pollution, such effects are not found in these trials, although estimates are imprecise (Barn et al., 2018b; Jack et al., 2021).

These results, constructed from diverse sources, reveal a surprising resilience to pollution in almost every city studied. Ignoring birth weight, the bulk of the evidence still suggests that particulate matter affects other important aspects of health (Landrigan et al., 2018), so addressing air pollution likely remains an urgent priority. However, potential fetal impacts are especially concerning given the long-run effects (Almond and Currie, 2011; Currie et al., 2014; Black et al., 2007). It is possible
that birth weight is a complicated proxy for measuring the damage wrought by pollution and other health insults. Indeed, Goldin and Margo (1989) unearthed birth records from “poor houses” in 1800s Philadelphia. Despite clear disadvantages in other outcomes, the average birth weights are similar to birth weights today. In light of the uncertainties outlined here, larger-scale randomized trials of air purification could provide key evidence on this important issue.

This paper contributes to a literature on the harms of pollution. Birth measures such as average birth weight and the incidence of low birth weight are widely used health measures (Almond et al., 2005) and constitute a significant share of the outcomes studied in causal evidence on pollution. This might be because birth weight is more consistently measured (Currie, 2013). Another reason to be especially concerned about these outcomes is the potential for long-run impacts on adults (Almond and Currie, 2011). But research on pollution has also found effects on outcomes not studied here, such as infant mortality (e.g., Chay and Greenstone, 2003; Heft-Neal et al., 2020), old-age mortality (e.g., Deryugina et al., 2019), cardiovascular disease (e.g., Liang et al., 2020), and asthma (e.g., Alexander and Schwandt, 2022), to give just a few examples.

This paper also contributes to a literature on the external validity of empirical findings (e.g., Rosenzweig and Udry, 2020; List, 2020; Vivalt, 2020; DellaVigna and Linos, 2022) and combining evidence across studies (e.g., Meager, 2022). While the evidence assembled here is different from other observational and quasi-experimental approaches, policymakers must aggregate from disparate sources to evaluate policies (e.g., US EPA, 2021). And reasoning from extremes can be useful. For example, scientists suspecting a link between smoking and cancer applied tobacco tar to lab animals and documented the resulting tumors (Proctor, 2012). Both cases revolve around biological determinants of health—the causal effects of education, for example, might vary more widely across contexts. But for biological factors like pollution, tests like this may provide key evidence on the plausibility of the mechanism and effect sizes.

---

3For example, in its background research for numerous grants addressing air quality, Open Philanthropy (2022) highlights birth-related outcomes in particular:

The lack of reliable household PM2.5 concentration data makes it difficult to confidently discern health effects. The available evidence indicates that negative health outcomes of household air pollution in South Asia may include low birth weight, preterm birth, and other conditions that are correlated with an increased risk of infant death.

See Almond et al. (2005) for a careful examination of the health consequences of low birth weight.
2  Background: Quasi-experimental evidence on pollution and infant health

Alexander and Schwandt (2022) use the rollout of “cheating” Volkswagen cars to study the impact of particulate matter on birth outcomes. The authors show evidence that, conditional on a set of controls, sales of the problematic cars are as-good-as-random, and these cars have a large impact on pollution as measured by EPA sensors. This makes the car sales attractive as an instrumental variable. They report several key findings, but I focus on their two-stage least-squares (2SLS) estimates. This is for ease of interpretation; the main results are presented in terms of the effect of an additional car, but the 2SLS estimates are in terms of PM2.5 particulate matter. They find that a 1 point (or ten percent) increase in PM2.5 decreases birth weight by 23 grams and increases the probability low birth weight (< 2,500 grams) by 0.44 percentage points (Table A.8).

As the authors point out, these impacts are not clear from simple cross-sectional comparisons. Based only on their levels of pollution, San Diego and Fresno County are both predicted to have low birth weight rates around 10 percent, but in reality they closer to 7 percent. More systematic exercises show a similarly muted association. Alexander and Schwandt (2022) compare their instrumental variable estimates to between- and within-county OLS regressions of birth outcomes on pollution (Table A.8). The instrumental variable estimates of the effects on low birth weight of a one-unit increase of PM2.5 (-23.22 grams) are 6 times higher than the between-county estimates (-3.81) and 47 times higher than the within-county estimates (-.0494).

It is common for causal estimates to differ from the simple correlational or “OLS” analysis, so this is not necessarily a sign that the findings are inaccurate. For instance, the authors argue that this difference could be due to measurement error in pollution, which would bias the simpler OLS relationship toward zero. There could also be omitted variables causing positively-selected families to live in higher-pollution counties. In what follows, I perform a similar test of whether more polluted areas have worse birth outcomes. However, because I select from a sample of extremely polluted cities, the degree of confounding or measurement error would need to be much greater to explain a muted effect.

---

4PM2.5 refers to particulate matter that is 2.5 micrometers in diameter or smaller. It is typically measured in micrograms per cubic meter ($\mu g/m^3$). For brevity, I leave out $\mu g/m^3$ and refer to it as PM2.5.
3 Evidence from extremely polluted cities

In this section, I collect data on pollution and birth weight from the most polluted cities in the world. The relationship between pollution and birth outcomes within this collection of cities does not capture the causal effect of pollution. Rather, each individual city is seen as a test of the predictions from the causal estimates. Residents of these cities likely inhale much more particulate matter on average. I find that given the amount of pollution, birth weights are surprisingly normal in places like Beijing.

3.1 Extrapolation

The average PM2.5 was 85 in Beijing in 2013 (WHO, 2016b) compared to around 10 in the US (see Appendix C). How much lower should we expect their average birth weight to be? In what follows I use the relatively conservative estimates from Alexander and Schwandt (2022) to extrapolate. The effects measured in Currie and Walker (2011) are 80% larger (see Alexander and Schwandt, 2022, Figure 6(A)), so this exercise essentially subsumes their estimates.

Birth weight Based on the crudest linear interpretation of the 2SLS estimates in Alexander and Schwandt (2022)\(^5\), birth weight should be \((85 - 10) \times 23 = 1,725\) grams or 52% lower in Beijing compared to the US. This is an unrealistically large effect, so I explore alternative methods of extrapolation. An obvious option is to recast the estimates in log terms, noting however that Alexander and Schwandt (2022) explicitly test for non-linear effects in the US context and find no evidence of this (Table 5).

Abstracting away from all factors besides pollution, the authors’ estimates can be written as

\[
BirthWeight_c = \alpha - 23 * PM25_c + e_c.
\]

where \(BirthWeight_c\) is average birth weight in city \(c\), \(\alpha\) is a constant, \(PM25_c\) is the pollution in city \(c\) (or the US), and \(e_c\) is the error term. Since average PM 2.5 in the US is about 10, a 1-unit increase in PM 2.5 is a 10 percent increase. So birth weight goes down by 23 grams for every 10 percent increase in PM 2.5, or 2.3 grams for every 1 percent increase. This relationship could be captured in a linear-log

---

\(^5\)These come from the IV estimates in Table A.8 Panel C column (1).
specification as

\[ BirthWeight_c = \alpha - 230 \ast \ln(PM25_c) + e_c. \]

The predicted birth weight in Beijing, relative to the US, is thus:

\[ \Delta_{US-Beijing} = 230 \ast \ln(85/10) = 492 \text{ grams.} \] (1)

This is a more plausible estimate, but is still large, corresponding to a roughly one standard deviation decrease. While there may not be a gold standard study of causal effects of fetal insults, Goldin and Margo (1989) note that a common point of comparison is the effects of the Dutch famine on birth weights, a decrease of about 300 grams (Stein et al., 1975; Stein and Susser, 1975). In between-sibling designs, heavy smoking during pregnancy is associated with a 226 gram decrease in birth weight (Juárez and Merlo, 2013). So this conservative log extrapolation of the effects in Alexander and Schwandt (2022) implies that living in Beijing is 60% worse than enduring a famine and twice as bad as smoking, at least based on the available estimates.

**Low birth weight** The authors also study incidence of low birth weight, defined as birth weight under 2,500 grams. They estimate that a 1-unit increase in PM2.5 causes a 0.44 percentage point increase in the probability of low birth weight. This means the difference in low birth weight between Beijing and the US should be (85-10)*0.44 = 33 percentage points. This linear extrapolation is similarly extreme. We can cast the pollution in log terms such that a 10% increase in PM2.5 (off a base of 10) leads to a 0.44 percentage point increase in low birth weight:

\[ LBW_c = \delta + 0.044 \ast \ln(PM25_c) + e_c. \]

So the rate of low birth weight in Beijing should be

\[ \Delta_{US-Beijing} = 0.044 \ast \ln(85/10) = 0.094 \]

\[ ^6\text{Currie et al. (2009) perform the same design using data from New Jersey and get smaller estimates of the impact of smoking, a reduction of 61 grams for women smoking 10 cigarettes a day. Almond et al. (2005) estimate a 200 gram effect using propensity score matching.}\]
or 9 percentage points higher compared to the US. I use these linear and log extrapolations in the analysis that follows.

### 3.2 Data

I next present birth weight data on some of the most polluted cities in the world, collated from different medical sources that sample from the city itself. I choose cities rather than larger geographic areas as the level of analysis in order to have the sharpest possible measure of exposure to pollution. (Alexander and Schwandt (2022) use the US EPA’s county-level sensors.) I restricted to cities in countries with a GDP per capita of at least $2,000. I use this income threshold to make the groups slightly more comparable. The poorest countries in the world have more than ten times the rate of infant mortality compared to Western countries; birth outcomes are worse across all categories in these places, likely due to factors beyond just pollution.

I provide details on every source in Appendix C. Sources were found by searching Google and Google Scholar for the city name and the words “average birth weight” or “birth weight.” These statistics are not available for most cities. I first searched for the most polluted cities from each country, provided the city had a PM 2.5 of 30 or more, about two times higher than the 99th percentile county-year level in the United States (see Appendix C). If I could not find data on that city, I searched for the next most-polluted city in the country and so on. The data needed to be representative of the population sampled and include at least 100 births. A major difficulty is that medical researchers often make restrictions that render the sample unusable (e.g., pre-term or full-term infants only, or only infants without major birth defects). Several of the present studies exclude multiple births, which tend to have lower birth weight. For simplicity, I exclude them from the US to not bias the comparison in favor of non-US places.

In some cases, the statistics used for the highly polluted cities are less reliable and less representative compared to US birth records. For example, the reference for Beijing (Su et al., 2016) excludes mothers over 43 (this is above the 99th percentile of US mother ages), and some data is based on household surveys rather than administrative hospital data (e.g., the source for Tunis, Sassi et al., 2019). The Ulaanbaatar study, Barn et al. (2018b), excluded mothers who smoke. However, many of the sources listed in Appendix C are from hospitals or hospital networks, and restricting to such studies would not change the results.
Table 1: Average birth weight in extremely polluted cities

<table>
<thead>
<tr>
<th>City</th>
<th>Country</th>
<th>PM 2.5</th>
<th>Birth weight (grams)</th>
<th>Linear prediction</th>
<th>Log prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beijing</td>
<td>China</td>
<td>85</td>
<td>3,343</td>
<td>1,555</td>
<td>2,793</td>
</tr>
<tr>
<td>Ulaanbaatar</td>
<td>Mongolia</td>
<td>74</td>
<td>3,490</td>
<td>1,814</td>
<td>2,826</td>
</tr>
<tr>
<td>Tuzla</td>
<td>Bosnia and Herzegovina</td>
<td>65</td>
<td>3,387</td>
<td>2,026</td>
<td>2,856</td>
</tr>
<tr>
<td>Belgrade</td>
<td>Serbia</td>
<td>61</td>
<td>3,259</td>
<td>2,113</td>
<td>2,870</td>
</tr>
<tr>
<td>Abu Dhabi</td>
<td>United Arab Emirates</td>
<td>56</td>
<td>3,080</td>
<td>2,233</td>
<td>2,891</td>
</tr>
<tr>
<td>Hanoi</td>
<td>Vietnam</td>
<td>52</td>
<td>3,251</td>
<td>2,320</td>
<td>2,907</td>
</tr>
<tr>
<td>Skopje</td>
<td>North Macedonia</td>
<td>45</td>
<td>3,325</td>
<td>2,492</td>
<td>2,942</td>
</tr>
<tr>
<td>Tunis</td>
<td>Tunisia</td>
<td>38</td>
<td>3,281</td>
<td>2,641</td>
<td>2,979</td>
</tr>
<tr>
<td>Colombo</td>
<td>Sri Lanka</td>
<td>36</td>
<td>2,920</td>
<td>2,697</td>
<td>2,994</td>
</tr>
<tr>
<td>Temuco</td>
<td>Chile</td>
<td>31</td>
<td>3,386</td>
<td>2,803</td>
<td>3,026</td>
</tr>
<tr>
<td>Monza</td>
<td>Italy</td>
<td>31</td>
<td>3,171</td>
<td>2,803</td>
<td>3,026</td>
</tr>
<tr>
<td>Extreme city average</td>
<td></td>
<td>52</td>
<td>3,263</td>
<td>2,310</td>
<td>2,921</td>
</tr>
<tr>
<td>Nationwide</td>
<td>USA</td>
<td>10</td>
<td>3,286</td>
<td>3,286</td>
<td>3,286</td>
</tr>
</tbody>
</table>

Notes: PM 2.5 data is from the World Health Organization (WHO, 2016b). Birth weight sources are described in Appendix C. The math underlying the predictions is given in Section 3.1. The Extreme city average weights each city equally.

Pollution data comes from the World Health Organization (WHO, 2016b), and GDP data from the World Bank (World Bank, 2022a). The pollution data from WHO excludes some large countries, so in a few cases I used academic papers for PM 2.5 estimates (see Appendix C). If the birth data was collected more than five years from the pollution data, I searched for a more proximate pollution estimate. I left out samples where I did not have a pollution estimate within five years.

3.3 Birth outcomes in highly polluted cities

3.3.1 Average birth weight

I show the results for average birth weight in Figure 1(A) and Table 1. To illustrate the procedure for one city: Ulaanbaatar has a reported annual mean PM2.5 of 74 (WHO, 2016b). This is almost five times higher than the 99th percentile county-year in the US. The linear prediction of its average birth weight, using the estimates in Alexander and Schwandt (2022), is 1,814 grams. Using the log prediction derived in Equation 1, its birth weight should be 2,826 grams. Its average birth weight is 3,343 grams, an estimate from a recent randomized trial of air purifiers and infant health (Barn et al., 2018b). This far exceeds either of the causal predictions.

This is true of most of the considered cities: the average birth weight in the city substantially outstrips the predictions from the model, shown in the two columns on the right in Table 1. To take
Figure 1: Birth weights and pollution in the most polluted cities in the world

Notes: This figure shows average birth weights for cities with extreme levels of pollution. In panel (A), the y-axis gives the average birth weight (in grams) in that city and the x-axis gives its level of pollution as measured by PM2.5. The solid blue line gives the log prediction derived from the causal estimates in Alexander and Schwandt (2022); the dashed gold line is the linear formulation (see Section 3.1). Panel (B) is analogous but with the share of low birth weight (< 2,500 grams) in the y-axis. Samples differ across the two panels due to limitations in which outcomes were available.
another example, Beijing is predicted to have a birth weight of 2,793 grams. However, its average birth weight actually exceeds that of the US, at 3,343 grams. The equally-weighted average birth weight across all of these cities, shown at the bottom of Table 1, is just slightly lower than in the United States.

In Figure 1(A), I give a visual representation of these findings. In most cases, the birth weights lie substantially above even the log-predicted birth weight. All cities except Colombo have higher average birth weights than the log prediction. The birth weights are also surprisingly large if we impose a floor on the predicted birth weight after PM2.5 passes 20. In other words, while there is surely measurement error in the pollution numbers, these birth weights would still exceed the causal predictions if these cities’ true PM2.5 were only 20. This could also address the possibility that the effects of pollution are concave: that is, the effects of pollution decrease as pollution increases as suggested by Miller et al. (2021).

The data on Ulaanbaatar from Barn et al. (2018b,a) are especially useful here, because the researchers specifically selected pregnant women for the study, recorded all pregnancy outcomes including miscarriage, and measured pollution exposure indoors using sensors. Indoor PM2.5 was 24.5 (Barn et al., 2018a, Table 2) in the control group, 25% higher than ambient outdoor air quality in the worst counties in the US. The average birth weight was 3,490 grams in the control group, higher than in the US.

3.3.2 Low birth weight

Average birth weight may not be the appropriate measure if pollution makes extreme outcomes more likely while leaving most of the birth weight distribution unchanged. I collected data on the incidence of low birth weight, defined as weighing under 2,500 grams, using the same approach. These estimates were easier to find because it’s a more commonly measured outcome. The results are shown in Table 2 and Figure 1(B). They echo the findings with average birth weight. Low birth weight in most of the highly polluted cities is below the rate for the US, including in Ulaanbaatar and Beijing.

To take one example, WHO reports that PM 2.5 in Brescia, Italy is 31. The log formulation of estimates from Alexander and Schwandt (2022) predicts a low birth weight incidence of 13% based on this level of pollution. My source for low birth weight incidence is a vital statistics report from Italy. It reports a low birth weight incidence of 7% among parents living in Brescia (ATS Brescia, 2020), which
### Table 2: Low birth weight incidence in extremely polluted cities

<table>
<thead>
<tr>
<th>City</th>
<th>Country</th>
<th>PM 2.5</th>
<th>LBW</th>
<th>Linear prediction</th>
<th>Log prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doha</td>
<td>Qatar</td>
<td>93</td>
<td>0.12</td>
<td>0.44</td>
<td>0.16</td>
</tr>
<tr>
<td>Beijing</td>
<td>China</td>
<td>85</td>
<td>0.04</td>
<td>0.40</td>
<td>0.16</td>
</tr>
<tr>
<td>Makkah</td>
<td>Saudi Arabia</td>
<td>74</td>
<td>0.09</td>
<td>0.35</td>
<td>0.15</td>
</tr>
<tr>
<td>Ulaanbaatar</td>
<td>Mongolia</td>
<td>74</td>
<td>0.04</td>
<td>0.35</td>
<td>0.15</td>
</tr>
<tr>
<td>Tuzla</td>
<td>Bosnia and Herzegovina</td>
<td>65</td>
<td>0.06</td>
<td>0.31</td>
<td>0.15</td>
</tr>
<tr>
<td>Astana</td>
<td>Kazakhstan</td>
<td>53</td>
<td>0.05</td>
<td>0.26</td>
<td>0.14</td>
</tr>
<tr>
<td>Hanoi</td>
<td>Vietnam</td>
<td>52</td>
<td>0.03</td>
<td>0.25</td>
<td>0.14</td>
</tr>
<tr>
<td>Skopje</td>
<td>North Macedonia</td>
<td>45</td>
<td>0.05</td>
<td>0.22</td>
<td>0.13</td>
</tr>
<tr>
<td>Colombo</td>
<td>Sri Lanka</td>
<td>36</td>
<td>0.16</td>
<td>0.18</td>
<td>0.12</td>
</tr>
<tr>
<td>Muscat</td>
<td>Oman</td>
<td>35</td>
<td>0.10</td>
<td>0.18</td>
<td>0.12</td>
</tr>
<tr>
<td>Beirut</td>
<td>Lebanon</td>
<td>32</td>
<td>0.09</td>
<td>0.16</td>
<td>0.12</td>
</tr>
<tr>
<td>Brescia</td>
<td>Italy</td>
<td>31</td>
<td>0.07</td>
<td>0.16</td>
<td>0.12</td>
</tr>
<tr>
<td>Temuco</td>
<td>Chile</td>
<td>31</td>
<td>0.03</td>
<td>0.16</td>
<td>0.12</td>
</tr>
<tr>
<td><strong>Extreme city average</strong></td>
<td></td>
<td>52</td>
<td>0.07</td>
<td>0.25</td>
<td>0.14</td>
</tr>
<tr>
<td><strong>Nationwide</strong></td>
<td>USA</td>
<td>10</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
</tr>
</tbody>
</table>

**Notes:** PM 2.5 data is from the World Health Organization (WHO, 2016b) unless noted in Appendix C. LBW is the fraction of births with low birth weight (< 2,500 grams). Birth outcome sources are described in Appendix C. The math underlying the predictions is given in Section 3.1. The Extreme city average weights each city equally.

is about the same as the US.

Note that in this analysis I make the conservative adjustment of removing all multiple births from the US sample. In the US data, this decreases the low birth weight incidence from 8.3% to 6.7% (Appendix C). This is to deal with the fact that some of the international samples omit multiple births, and for others it is unclear (see Appendix C). Overall, this comparison is probably biased slightly toward finding higher rates of low birth weight in the international sample if most studies actually included multiple births.

### 3.4 Other international sources

In Appendix D, I discuss other noteworthy sources which did not meet the inclusion criteria. In particular, the city of Teplice in the Czech Republic was once one of the most polluted cities in the world. In the early 1990s, the Czech public health department entered into a special collaboration with the US Environmental Protection Agency to study the health consequences of pollution there. These studies yielded detailed estimates of birth outcomes and exposure, and similarly show no clear harms as far as the incidence of low birth weight and pre-term birth (Sram et al., 1996). I also discuss
Rich et al. (2015), a study of births and pollution in Beijing around the 2008 Summer Olympics. This has better pollution and birth data than the Beijing source employed in the main analysis, and has similarly normal birth outcomes, but the authors excluded preterm births.

3.5 Preterm birth

Alexander and Schwandt (2022) and Currie and Walker (2011) also study preterm birth as an outcome, defined as a birth before 37 weeks of gestation. Some of the studies in the international sample also recorded this and I note the results in Appendix C. In this case, there are fewer estimates, but the numbers give a similar impression: Beijing records a preterm birth rate of 5.0%, in Belgrade it is 8.7%, in Brescia it is 6.8%, and in Doha it is 14%. Except for Doha, these rates are lower than in the US. The preterm birth rate in 2021 was 10.5% (CDC, 2023) and about 8% for singletons (Martin and Osterman, 2021).

Together, these findings suggest that infants in highly-polluted areas around the world have surprising resilience to pollution—at least based on the most commonly used outcomes of birth weight and low birth weight. However, one potential risk is that these samples are somehow selected. For example, most studies are not population-level surveys and may consist of research hospitals serving richer communities. Next, I use US natality files to ask whether birth weights in previously highly polluted counties seem obviously impacted. This helps address selection concerns with the sparse international cities here and allows us to study other aspects of the birth such as gestation.

3.6 Historical birth weights in highly polluted US counties

The United States used to be much more polluted. Chay and Greenstone (2003, Figure 1) report that the national mean of total suspended particulates (TSP) was 93 micrograms per cubic meter in 1970 and just under 60 in 1990. Unfortunately, there’s not a clean way to convert TSP to PM2.5. PM2.5 is restricted to smaller particles and the EPA only began tracking it in 1998 (Voorheis et al., 2017). Lall et al. (2004) use data from sensors with both measurements and find that the average ratio between PM2.5 and TSP particulates is 0.30. Using this crude conversion factor (cf. Voorheis et al., 2017), PM2.5 pollution in the US was around 28 in 1970 and 17 in 1990. I use this factor going forward noting that it is an approximation. Other pollutants such as sulfur dioxide, nitrogen oxide, and lead have all seen similarly dramatic decreases.
Geocoded birth weights from the 1970s to 2005 are available from the NBER’s natality files (NCHS, 2023). These complement the data from non-US cities because there are minimal concerns about sample selection and it is easy to use the exact birth weights to measure the prevalence of low birth weight (< 2,500 grams) and very low birth weight (< 1,500 grams). Finally, information on the mothers allows us to probe the role of shifting maternal characteristics. Throughout I include twins from both the 1972 and contemporary samples, so these estimates are different from those in the international comparison.

In Table 3, I show the county-level TSP along with the imputed PM2.5, the average birth weight, the percent of low birth weight and very low birth weight infants, and the number of births in the most 15 polluted counties from 1972. The results resemble those from the non-US cities: The birth weights are surprisingly similar to the 2019 US sample, despite the vast difference in exposure to particulate pollution. For example, Allegheny, PA recorded the highest pollution that year, with an imputed PM2.5 of 43. But its mean birth weight and low birth weight incidence, taken based on 9,378 infants born there in 1972, are all no worse than in the country-level aggregates for the US in 2019. Notably, the nitrogen dioxide exposure in Allegheny is also extreme: 64 parts per billion (ppb) compared to a median of under 10 ppb in the contemporary US (US EPA, 2017, Figure 2.4).

The second to last row shows that the average birth weight in the polluted county sample is 3,286 grams, almost identical to the 2019 US average. We see a similar pattern in the two indicators for very small infants. The infants in 1972 are less likely to be low birth weight (7.1% in the 1972 sample vs. 8.3% in 2019) or very low birth weight (1.0% vs. 1.4%).

These estimates are at the county level, but the US natality data includes the city of residence for cities with populations exceeding 250,000 (NCHS, 1972). In theory, city-level pollution should be more representative of exposure because it is measured in a smaller area. Seven cities register as extremely polluted (TSP over 100) and are large enough to be indicated in the natality files: Birmingham, Denver, Detroit, El Paso, Newark, Phoenix, and Wichita. I show results for them in Figure A.1 with the estimates reported in Table B.1. These cities similarly surpass the predictions of the causal evidence although their birth outcomes are slightly worse, driven by Newark and Detroit. In particular, the low birth weight incidence in the combined group is 9.9% compared to 8.3% in the contemporary US. But the more conservative log formulation predicts that low birth weight incidence should be 14%.
Table 3: Pollution and birth outcomes in the most polluted US counties in 1972

<table>
<thead>
<tr>
<th>Place</th>
<th>TSP</th>
<th>Est. PM 2.5</th>
<th>Birth weight</th>
<th>% LBW</th>
<th>% VLBW</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allegheny, PA</td>
<td>144</td>
<td>43</td>
<td>3,271</td>
<td>6.9</td>
<td>1.1</td>
<td>9,378</td>
</tr>
<tr>
<td>Riverside, CA</td>
<td>140</td>
<td>42</td>
<td>3,333</td>
<td>6.5</td>
<td>0.9</td>
<td>3,599</td>
</tr>
<tr>
<td>El Paso, TX</td>
<td>140</td>
<td>42</td>
<td>3,268</td>
<td>6.8</td>
<td>0.7</td>
<td>4,509</td>
</tr>
<tr>
<td>Webb, TX</td>
<td>137</td>
<td>41</td>
<td>3,327</td>
<td>5.8</td>
<td>1.1</td>
<td>1,140</td>
</tr>
<tr>
<td>Madison, IL</td>
<td>135</td>
<td>41</td>
<td>3,333</td>
<td>6.2</td>
<td>1.2</td>
<td>2,164</td>
</tr>
<tr>
<td>Tulare, CA</td>
<td>132</td>
<td>40</td>
<td>3,377</td>
<td>5.5</td>
<td>0.5</td>
<td>1,767</td>
</tr>
<tr>
<td>Kern, CA</td>
<td>133</td>
<td>40</td>
<td>3,302</td>
<td>6.9</td>
<td>0.9</td>
<td>3,006</td>
</tr>
<tr>
<td>Mahoning, OH</td>
<td>130</td>
<td>39</td>
<td>3,287</td>
<td>7.7</td>
<td>1.1</td>
<td>2,286</td>
</tr>
<tr>
<td>Washington, PA</td>
<td>129</td>
<td>39</td>
<td>3,290</td>
<td>7.6</td>
<td>1.2</td>
<td>1,456</td>
</tr>
<tr>
<td>San Bernardino, CA</td>
<td>132</td>
<td>39</td>
<td>3,316</td>
<td>7.0</td>
<td>1.1</td>
<td>5,580</td>
</tr>
<tr>
<td>Denver, CO</td>
<td>125</td>
<td>38</td>
<td>3,134</td>
<td>9.4</td>
<td>1.1</td>
<td>4,262</td>
</tr>
<tr>
<td>Beaver, PA</td>
<td>127</td>
<td>38</td>
<td>3,257</td>
<td>8.5</td>
<td>1.4</td>
<td>1,399</td>
</tr>
<tr>
<td>Pinal, AZ</td>
<td>127</td>
<td>38</td>
<td>3,299</td>
<td>7.4</td>
<td>1.0</td>
<td>793</td>
</tr>
<tr>
<td>Scott, IA</td>
<td>128</td>
<td>38</td>
<td>3,427</td>
<td>4.1</td>
<td>1.2</td>
<td>1,170</td>
</tr>
<tr>
<td>Black Hawk, IA</td>
<td>125</td>
<td>37</td>
<td>3,328</td>
<td>7.9</td>
<td>1.5</td>
<td>957</td>
</tr>
<tr>
<td>1972 Combined</td>
<td>135</td>
<td>40</td>
<td>3,286</td>
<td>7.1</td>
<td>1.0</td>
<td>43,466</td>
</tr>
<tr>
<td>2019 USA</td>
<td>33*</td>
<td>10</td>
<td>3,254</td>
<td>8.3</td>
<td>1.4</td>
<td>3,753,815</td>
</tr>
</tbody>
</table>

Notes: This table shows pollution and birth weights for extremely polluted counties in the US in 1972. TSP is total suspended particles as reported in the EPA Air Quality Trends report for 1972. The PM2.5 column is estimated by multiplying TSP by 0.3, except for the USA row. Birth weight is from the NHCS natality file for 1972 and (for the USA row) 2019. LBW is defined as birth weight under 2,500 grams, VLBW is birth weight under 1,500 grams. N is the sample size in that place.

I next use the county-level data to study how the change in pollution from 1972 to 2002 might have improved birth outcomes.7 Figure 2 shows two observations for each county in the high-pollution sample: one from 1972 and one from 2002. The maroon circles show data points from 1972, the same as in Table 3. The green diamonds show the same data from 2002. On average, these counties saw a 64% decrease in PM2.5, from an (imputed) average of 39 to a (measured) average of 14.

The decrease in PM2.5 should have led to increases in birth weight. Instead, most move laterally in the plot: PM2.5 decreased without any meaningful improvement in birth outcomes. The average percent change in birth weights is about zero.

Older US sources Data on birth weights before the 1960s is scarce, but there are a few notable examples. Goldin and Margo (1989) study birth records from the Almshouse Hospital in Philadelphia in 1848–1837. The overall mean birth weight is 3,377 grams, although it’s unclear how polluted Philadel-

---

7I stop at 2002 because the natality files switch their birth weight coding in 2004 (and from 2005 onward the geocodes are only available in the limited access data).
Figure 2: Birth weights and pollution predictions in the most polluted US counties in 1972-2002

Notes: This figure shows county-level PM 2.5 measurements and average birth weight for 1972 and 2002 for the 15 most polluted counties from 1972. The math underlying the predictions is given in Section 3.1.

phia was at this time. It’s striking, however, how these mothers were selected: “The majority of the women were destitute, abandoned, and without other housing; they were, in other words, the poorest of the poor” (Goldin and Margo, 1989, p. 370). Costa (1998) uses records from the New York Lying-In hospital for 1910-1931. Birth weights average 3,463 grams (Table 1). It is harder to know the health conditions of these mothers, but the birth weights are also similar to today.

3.7 Summing up
These estimates show that extrapolations from Alexander and Schwandt (2022) and Currie and Walker (2011) do not match birth outcomes in highly-polluted cities. These observations are not meant to trace out the relationship between birth weight and pollution. Instead, each observation shows a
distinct prediction error compared to the simplified calculations based on the causal estimates. But my extrapolation was maximally simple: I only considered pollution as a predictor.

4 Potential explanations

In this section, I consider factors that could explain why infants in these places do not exhibit lower birth weights. I first discuss potential issues in using air quality sensors to measure exposure to pollution. Next, I discuss the broad range of selection issues that could affect the comparison, including obstetric practices and sample selection. I give special consideration to culling effects, and that hospitals in the US may be better equipped to care for very premature infants.

4.1 How bad is exposure in the international cities?

One potential explanation is that experienced pollution is actually not so extreme in these cities. If these comparisons are a useful check on existing results, it must be that sampled mothers from the highly-polluted areas are exposed to much more pollution compared to contemporary US mothers. For the highly-polluted US sample, this assumption seems uncontroversial given the widely-documented decline in pollution in the US since the 1970s (e.g., Chay and Greenstone, 2003; US EPA, 2023).

For the international cities, it is less clear. Pollution sensors might be placed at the urban center, far from the typical resident. Beijing, for example, has an area of around 6,000 square miles, although this is an outlier. Brescia, another one of the cities in the international sample, is 35 square miles, and the data was restricted to parents who lived in Brescia (ATS Brescia, 2020). Currie et al. (2023, Figure B.1) use data from Di et al. (2016) to form a more granular measure of pollution exposure in the US, using 1 kilometer × 1 kilometer cells rather than the typical county-level estimates. They find that up to half of the White-Black pollution gap in PM 2.5 exposure is within commuting zones (which contain around four counties each on average). Cities in the present sample might similarly exhibit variation in exposure.

Newer studies integrate satellite measurements, ground-level sensors, and information about the spatial distribution of population to estimate exposure (WHO, 2023; Shaddick et al., 2018). In Table B.2, I show population-weighted estimates of country-level urban PM2.5 exposure from WHO (2023) based on Shaddick et al. (2018). These are not provided at the city level. But the exposure in ur-
ban areas in these countries generally confirm that pollution is far worse in these places. For example, WHO (2023) estimates that the PM2.5 exposure in urban places is 49.8 in China and 47.1 in Mongolia.

Another core issue with many studies of pollution exposure is that pollution data comes from outdoor sensors while people spend most of their time indoors. This makes the connection between indoor and outdoor pollution important to understand. In a large meta-analysis, Chen and Zhao (2011) find that the approximate ratio of indoor to outdoor PM2.5 is about 1 (IQR: 0.83-1.58), pooling 87 studies mostly in the US (Table S1). But these do not include the highly polluted places in my sample.

One study from the birth weight statistics collected here shows clearly that experienced pollution exposure is high among the mothers in the sample. Barn et al. (2018a) used sensors to measure indoor and outdoor exposure in their trial in Ulaanbaatar. They found a geometric mean indoor PM2.5 of 25, and outdoor PM 2.5 was over 50 for almost all households (Table 1). This is similar to the apartment dwellers tracked in another study of Ulaanbaatar, where researchers registered a personal PM2.5 exposure of 27 (Kim et al., 2021). But none of the other birth samples are linked to person-level pollution exposure.

I searched for studies of indoor pollution or personal exposure from the list of cities included in the international sample and from the historical US. The evidence here is limited because, unlike ambient pollution, there are not systematic studies meant to generate estimates of personal exposure at the city or county level, even within the US. In general, however, studies of indoor environments or personal exposure in these places find similarly extreme levels of pollution.

China has some of the best evidence. Liang et al. (2019) had 50 Beijing residents wear portable environmental sensors as they went about their usual routines, recording an average PM2.5 exposure of 38, more than double the worst outdoor measures in the US. A similar study found an average PM2.5 exposure of 127 across 60 office workers in Beijing (Baccarelli et al., 2014). Rich et al. (2015), a study of air pollution and birth weight outcomes around the Olympics in Beijing, find mean PM 2.5 exposure of 61 when restricting to districts where the mothers in the sample reside.8 Zhang et al. (2021) reviews studies of China from 1980-2019 and finds an average indoor PM2.5 of 95 in urban residences. This suggests that high levels of pollution in China result in high levels of exposure.

---

8This study also finds normal birth weights in Beijing. However, they restricted to term births so this was not used for the main analysis.
Apart from China, researchers have studied personal exposure in a handful of other cities from the international sample. Jorquera et al. (2018) find that median indoor PM2.5 across 63 households in urban Temuco, Chile was 44.4. Chamseddine et al. (2019) report high PM2.5 levels across several rooms in a hospital in urban Beirut (Table 3). Tran et al. (2021) record an average indoor PM2.5 of 52.1 and outdoor PM2.5 of 54.4 across 32 urban residential homes in Hanoi. Borgini et al. (2015) had 90 Milan students ages 12 to 18 wear a portable sensor to track their pollution exposure (Milan is not in the sample but is in the same Lombardy region as Monza and Brescia). PM2.5 exposure was 35 indoors and 46 outside (Table 1). There is less evidence from the 1970s United States as the study of indoor pollution was in its infancy.\(^9\)

These two strands of evidence—the satellite-based measures and more limited studies using personal sensors—suggest that higher city-level outdoor air pollution, as documented in WHO (2016b), is associated with higher levels of exposure. And in general, it would certainly be surprising if almost all of the international samples collected here managed to avoid the pollution ascribed to their cities by WHO (2016b), as these same numbers underlie estimates of the global health consequences of pollution (Landrigan et al., 2018).

4.2 Selection effects

Maternal characteristics in US sample

Many important maternal characteristics in the US have shifted over this time. This could mask improvements in infant outcomes caused by the drop in pollution from 1972 to 2002. In Figure A.3, I show how trends in birth weight and low birth weight change with the inclusion of controls for mother age, race, live birth order, and state of residence. The trends are unchanged, suggesting that maternal characteristics are not a major confound for the comparisons across time periods. (These results are identical when I restrict to the polluted county sample.)

Maternal characteristics in international sample

Unsurprisingly, the highly-polluted international cities are in countries that are poorer than the United States, so selection effects would generally predict worse outcomes. I show country-level life expectancy and GDP for these places in Table B.2. The

---

97% of households had indoor PM2.5 above 25, and 88% had outdoor PM2.5 above 25 (Fig. 4).

One early example is Moschandreas et al. (1981), which studies the relationship between indoor and outdoor TSP at ten residences and two office buildings in the Boston area. They found that indoor TSP was consistently higher than outdoor TSP.
US is richer than all countries in the sample. And in 2019, US life expectancy was 79, exceeding most counties in the sample but not Italy, Qatar, Lebanon, and Chile. All of these countries rank lower on health and development (United Nations, 2020). With worse attributes along these dimensions, it would be surprising if some unconsidered factor were bolstering birth weight in these cities.

The mothers in the samples could be different from other mothers in the city in terms of their pollution exposure, health, and demographics. Currie and Schwandt (2016) argue that mothers exposed to the 9/11 dust cloud may have been positively selected along key dimensions, leading to a spurious cross-sectional finding that mothers close to the dust cloud saw better birth outcomes. I noted above that the Ulaanbaatar sample is selected: in particular, it is restricted to non-smokers. Two sources are city-wide: Brescia (ATS Brescia, 2020) and Makkah (General Authority for Statistics, 2016). This would seem to address issues around selection within the city. Descriptives are not available for most of the samples included. But the Hanoi sample (Tran et al., 2012, Table 3) shows similar incomes compared to the rest of Vietnam. And 13.6% of the Temuco sample reports smoking during pregnancy, compared to 7.2% in the US (Drake et al., 2018).

It’s possible that inherited factors not related to economic development tend to lead to consistent differences in birth outcomes along racial or ethnic lines. In practice this seems not to threaten the conclusions here as reference materials on birth weight tend to suggest larger weights among European populations (Janssen et al., 2007). Age is another potential confound. Birth weight increases with maternal age until around 30 (Wang et al., 2020). Except for China, where the average age at childbirth is similar to the US, this tends to bias comparisons in favor of the US.\(^\text{11}\) Taken together, it seems unlikely that background traits of the mothers in the sample predict better birth outcomes overall.

**Obstetric practices and gestation length** Some research suggests that changing obstetric practices have worked to decrease birth weights in the US: now, c-sections and inductions are more likely to happen before the due date, and pregnancies are not allowed to continue much beyond the 40th week (Tilstra and Masters, 2020; VanderWeele et al., 2012).

These changes could mask improvements in infant health brought by cleaner air in the analysis of US counties from 1972 to 2002. One way to address this is to calculate the birth weight per week and

---

\(^{11}\)The average age at first birth in China was 26.9 years in 2016 (He et al., 2019, Table 4). In the US it was 26.3 years in 2014 (Mathews and Hamilton, 2016).
then multiply by 40 to get a constant-gestation birth weight:

\[
\text{ConstantGestationBW}_i = 40 \times (\text{BirthWeight}_i / \text{Gestation}_i)
\]

I show the results in Figure A.2. In contrast to birth weights, constant gestation birth weights have increased—by about 2.6% from 1972 to 2002 in the polluted US county sample. This is small compared to the changes in pollution. The average decrease in pollution was around 60%, so this corresponds with an elasticity of 0.07. The implied elasticity from Alexander and Schwandt (2022) is 1.0 (Figure 6A). The other main outcome used in Figure 1 is low birth weight. This has increased over time in the US sample, and this remains so when I use the constant-gestation correction.

**Viability** Premature infants are smaller, and the US and other developed countries attain viability—typically defined as a 50% chance the infant survives—at lower lengths of gestation, about 23–24 weeks according to Glass et al. (2015). Some countries in my international sample impose higher thresholds for viability. For example, in China, official guidelines are to only provide full care for infants of at least 28 weeks of gestation (Han et al., 2022). Italy, however, which provides two of the polluted cities from the international sample, almost always provides care at 25 weeks, with debates around the proper practice at 22-25 weeks (Pignotti and Moratti, 2010).

Based on its 2019 data, if the US used China’s same threshold of 28 weeks, average birth weight would be 13 grams more on average, and the incidence of low birth weight would be 6.2% instead of 6.7% (all excluding multiple births). The difference would be even smaller using gestational cutoff of 25 weeks, a conservative guess at Italy’s policy. Data on the thresholds employed by the lower-income countries in the sample is difficult to find. Hayden et al. (2020) surveyed hospitals in the Philippines, which has a lower GDP per capita than all the countries from the international sample except Vietnam, Iran, and Nigeria. They found that most hospitals provide care at 27-28 weeks of gestation, similar to the official guidelines reported for China. The vast majority of hospitals reported resuscitating at 31-32 weeks.

Even imposing a minimum gestation of 32 weeks on the US sample would leave a large gap between the predicted birth weights and actual birth weights. Restricting to gestation lengths of 32 weeks or more, US birth weights would be about 30 grams higher and 1pp less likely to be low birth
weight. (The changes are similar with or without multiple births.) This extreme adjustment would account for just a small share of the gap between the observed birth outcomes and the log predictions discussed above.

4.3 Survivorship bias or culling effects

A related explanation is culling. Mothers in polluted and impoverished areas may have higher rates of miscarriage and stillbirth or more difficulty conceiving. If the fetuses that suffer a miscarriage or were not conceived would have had low birth weights, newborn infants in the polluted cities could be positively selected through a culling effect. This question—whether health insults have culling or scarring effects—is prominent in the fetal origins literature.\(^\text{12}\) Could selective survival lead to a healthier infant population, at least as measured by birth weight?

This would have the interesting implication that interventions reducing pollution exposure could decrease average birth weights through its effects on survival. I could find no studies arguing that fetal insults increase aggregate average birth outcomes through culling, so this explanation would contradict a vast literature (see the meta-analysis in, e.g., Bekkar et al., 2020; Li et al., 2017). In the developing world, studies of household fuels tend to find that indoor pollutants both increase stillbirths and decrease birth weight, consistent with a scarring rather than culling effect (Pope et al., 2010; Amegah et al., 2014). Further, while data on miscarriage is sparse and complicated to measure (for the US, see Mukherjee et al., 2013), the rates of stillbirth (Hug et al., 2021) and low birth weight (Blencowe et al., 2019) at a national level tend to move together, which also suggests that deprivation tends to lead to scarring. Still, given the unique sample of extremely polluted cities, the mechanism deserves consideration.\(^\text{13}\)

\(^\text{12}\)See Almond and Currie (2011), Almond (2006), and Chay and Greenstone (2003) for discussion and specific examples. Floris et al. (2021) and Brown and Thomas (2019) argue that the flu pandemic of 1918 induced negative selection; i.e., infants that survived to birth had lower socioeconomic status. Bruckner and Catalano (2007) find evidence for culling: males born in times of low male:female sex ratios (a proxy for fetal stress) have lower mortality, which suggests that infants who survive to birth in inhospitable conditions have better health outcomes. Almond and Currie (p. 157 2011) discuss this in the context of the null findings from the famines in Finland and Leningrad, where mortality might have been high enough to induce a culling effect. Bozzoli et al. (2007), looking at the relationship between neonatal mortality and adult height, find a negative relationship overall (consistent with scarring), but some evidence of a positive relationship in the poorest countries (consistent with culling).

\(^\text{13}\)Some research suggests that the share of male infants falls in response to stressors (e.g., Catalano et al., 2006). But birth weight is slightly higher for males (Van Vliet et al., 2009), so effects operating through the sex ratio could decrease, not increase, aggregate birth weight. And in a large study of Ulaanbaatar (Dorj et al., 2014), the authors record 78,076 female and 82,381 male births, a male-female sex ratio of 51.34%. This is about equal to the sex ratio at birth in the United States (Mathews et al., 2005), although Dorj et al. (2014) include only full-term births.
Culling through miscarriage  To affect birth weight, culling would need to happen either through miscarriage, stillbirth, or (a decrease in) conceptions. Observational studies tend to find that pollution increases spontaneous abortion (Zhang et al., 2019; Grippo et al., 2018). A simpler test is whether the miscarriage/stillbirth rate is in general higher in polluted areas.

The Barn et al. (2018b) study provides a useful estimate of this statistic, which is difficult to find for the countries in the sample. The mothers were recruited at a median gestation of 11 weeks and at or before 18 weeks of gestation. They report that 10.1% had either a miscarriage or stillbirth. I could find only one other detailed study of miscarriage from a sample with high pollution exposure. Dellicour et al. (2016) recruited 1,134 pregnant women in rural Siaya County, Kenya to estimate weekly miscarriage rates. Their sample included 508 mothers captured before the 12th week of gestation. While pollution exposure is not reported in Dellicour et al. (2016), 93% of rural Kenyans primarily use charcoal and firewood, so the exposure to indoor pollutants is likely high (Kenya Ministry of Energy (2019); also see Dida et al. (2022) for data on a neighboring county). Their survival rates are similar: after 11 weeks of gestation, the risk of miscarriage is 8.8% (Dellicour et al., 2016, Table 2).

There are few studies of US samples to compare to, but the existing estimates are comparable to Ulaanbaatar and Kenya. Mukherjee et al. (2013) calculates cumulative miscarriage risk for a sample of mothers in North Carolina. The risk conditional on making it through the 9th week of pregnancy is 8%. Ammon Avalos et al. (2012) collate data from four previous studies in the US and find that miscarriage rates as of the 9th week of gestation range from 6 to 10% (Figure 3), arguing that the highest estimate is most reliable because it recruited people earlier in pregnancy. The risk of miscarriage as of the 11th week of gestation is 7.7%.

These estimates suggest that miscarriage rates are not distinctly higher among populations with severe pollution exposure, although the comparison is not perfect because Barn et al. (2018b) do not report specific miscarriage rates or stillbirth rates. Instead, they report on the risk of pre-term birth. However, the risk of pre-term birth is significantly higher in the treatment group compared to the control group (OR=2.37; 95% CI: 1.11, 5.07). This could be a statistical fluke, as the two groups are identical in their risk of spontaneous abortion (24 in control vs. 10 in treatment). Barn et al. (2018b, Table 1)

Stillbirth is unfortunately not reported, but another recent study in Siaya, Eilerts-Spinelli et al. (2022), finds a rate of about 1%.

Specifically, “The cumulative risk of embryonic loss (gestational weeks 6–9) was 15% for whites and 14% for blacks. The cumulative risk for early fetal loss (weeks 10–15) was 6% and 8% and, for late fetal loss (weeks 16–19), was <1% and 2% for whites and blacks, respectively” (Mukherjee et al., 2013).
calculate the weekly risk of miscarriage. Assuming all mothers in Barn et al. (2018b) were at exactly
11 weeks of gestation, a bounding exercise shows it is unlikely that this could explain the normal birth
weights. Assume the miscarriage rate at 11 weeks is 8% in the US (Ammon Avalos et al., 2012) and—
roughly—10% in Ulaanbaatar (Barn et al., 2018b). Then there’s a missing 2% of infants in Ulaanbaatar
who presumably would have had lower than average birth weights. Assuming these babies are at
the 1st percentile of the US birth weight distribution, weighing 1,275g, adding them back into the
sample would decrease the Ulaanbaatar average by 43g, from 3,403g to 3,357g, which would still put
Ulaanbaatar above the US average. Similarly, the incidence of low birth weight would still be lower
in Ulaanbaatar if we counted these infants.

Censoring the US birth distribution  These potential explanations around miscarriage and viability
can be viewed as a censoring of the birth weight distributions in the international samples relative to
the US. While we do not have data on the full distributions in the non-US samples, we can use the
contemporary US data to illustrate just how censored the samples would have to be for the predictions
to line up with the causal estimates. Here I show that such censoring would need to be extreme.

Figure A.4 studies how excluding the bottom 0 to 5 percent of births from the US birth weight
distribution would change birth outcomes in the US and the causal predictions in the international
sample. As before, multiple births are excluded. Panel (A) shows the results using birth weight. The
estimates of the US average birth weight mechanically increase as the amount of censoring increases
(solid black line). The red and blue dashed lines show that the average predictions also increase,
because they were derived using the US as a reference (see Section 3.1). Even censoring the bottom 5
percent of the US birth distribution, the polluted city average birth weight, shown in the flat maroon
line, exceeds the predictions.

Panel (B) shows the same exercise for low birth weight incidence. The US incidence of low birth
weight mechanically decreases with the level of censoring (solid black line). In this case, the average
log prediction comes close the observed polluted city average when removing the bottom 5 percent of
the US birth weight distribution.

This level of censoring seems implausible given the available evidence. I estimated above that the
rate of either stillbirth or miscarriage might be 2 percentage points higher in Ulaanbaatar. This is half
the needed censoring, and would not perfectly correlate with birth weight. Finally, several papers
from the international sample provide estimates of the rate of stillbirth: Monza (0.4%), Brescia (0.3%), Tuzla (0.7%) and Belgrade (1.3%) (see Appendix C). Except for Belgrade, these are lower than in the US (0.6%; Gregory et al. (2022)).

**Culling at conception** A culling effect could also happen earlier, before the infant is conceived. Pollution could inhibit conception for “weaker” embryos. There is some indirect evidence for this: Births in Ulaanbaatar exhibit seasonality, with 13% more births in the Fall compared to the Spring (Dorji, 2015, Table 1). If this drives culling, birth weights should be higher in the season with fewer births, but they are the same (Dorji, 2015, Table 1).

Apart from seasonality, another test is whether couples in highly polluted places have to spend longer trying to conceive. In vitro fertilization (IVF) procedures offer a more controlled environment for measuring this. Choe et al. (2018) find that IVF implantation is slightly less successful for women in more polluted areas. On the other hand, a careful examination by Zhou et al. (2018) finds that implantation success rates are about the same in Beijing as in Europe.

In sum, it’s hard to find evidence that pollution is actually quite harmful in these places through its effects on miscarriage and conception. Miscarriage rates by week of gestation appear similar to the US. And IVF implantation has similar success rates in Beijing and Europe. These findings are based on sparse data. But separate from these specific mechanisms around miscarriage or conception, such culling effects would be quite surprising given that no previous study has found positive effects of pollution on birth weight.

### 4.4 Timing of exposure

The cities and counties in the highly polluted places studied here have stably high levels of pollution. In contrast, much of the quasi-experimental literature is based on shorter-term changes in exposure (as is the case in Alexander and Schwandt (2022) and Currie and Walker (2011), see also the studies reviewed in Currie et al. (2014)). Could it be that short-term effects are larger than long-term effects? This would also be surprising if the primary mechanism is the accumulation of particulate matter in lungs (Lippmann et al., 2003), although this has not been clearly established.

---

18 Badarch et al. (2021) find greater seasonality, but it’s based on 10,000 births from a single hospital. Dorji (2015) samples 160,000 births from all private hospitals in Ulaanbaatar.
There are certainly examples from biology where early exposures make individuals more resilient. Human hands and fingers develop calluses in response to repeated friction, which makes them more robust to future exposures. And some research suggests that early exposure to allergens decreases allergic responses (Hesselmar et al., 1999). In the current context, this explanation could hold that, due to their early exposure, mothers in highly-polluted places are more resilient to pollution than contemporary US mothers. The harmful effects of pollution are thus less likely to impact fetuses in more polluted places.

There is little research available which can speak to this claim. One potential piece of evidence against this is that Currie et al. (2009) find that pollution effects on fetuses were more harmful for mothers who smoked. On the other hand, Checkley et al. (2021) find that wildfire smoke increases premature births much more in places with less smoke exposure on average (Figure 4). Overall, it would seem surprising if birth outcomes were worse among infants exposed to a temporary increase in pollution from 5 to 15 PM2.5 compared to infants born in an environment with a constant PM2.5 of 25. However, it is hard to definitively rule out such a mechanism. Either way, this is more of an explanation for the surprising resilience rather than an argument against it.

5 Experimental evidence

A sparse experimental literature tests the idea that short-term changes in air quality could impact fetal health. Based on the quasi-experimental estimates, air purifiers could be a highly cost-effective intervention. Low-cost air purifiers decrease indoor particulate matter by 30-60% (Barn et al., 2018a) and cost between $100 and $200 as of February 2022. People spend most of their time indoors, and outdoor pollution affects indoor pollution (Leung, 2015). Replacing biomass fuels with ethanol could have similar benefits.

To my knowledge, there is only one randomized controlled trial of air purifiers and infant health, performed in Ulaanbaatar, one of the most polluted cities in the world (Barn et al., 2018b). The researchers randomized 540 mothers at 11 weeks gestation. The purifiers caused a 7.2 ppm decrease in indoor PM2.5 (Barn et al., 2018a). However, the effects on birth weight were null with a 95% confidence interval of -84g to 120g. Although this confidence interval encompasses practically large effects, this rejects the linear predicted effects from Alexander and Schwandt (2022, Table A.8(c)), which imply a birth weight increase of 167 grams (23.2×7.2). But the wide confidence interval does not rule out my
log formulation (see Section 3.1) of the results.

Interestingly, other interventions using air purifiers and liquefied gas achieve large reductions in indoor PM2.5 without substantial effects on health outcomes (Shao et al. (2017); Checkley et al. (2021); Dong et al. (2019), also see the review in GiveWell (2022)). These studies use more accurate measures of exposure. In particular, the subjects in Checkley et al. (2021), a trial of liquefied petroleum gas, wear a sensor on their clothing that measures particulate matter in their immediate environment. The sensors record a PM 2.5 of 30 in the treatment group vs. 98 in the control group, but the paper finds no effects on lung function.

Another notable example is Jack et al. (2021), a pre-registered study in rural Ghana with 1,414 households that also provided liquefied petroleum gas to treated families. The gas decreased indoor PM 2.5 by 25 (Figure 3), measured using sensors that were physically placed on the mothers for 72 hours. The birth weights across arms were practically identical. Control infants were 29 grams heavier than treated infants, with the 95% confidence interval rejecting birth weight benefits above 56 grams. The log formulation (see Section 3.1) of the results in Alexander and Schwandt (2022) suggests that this should have increased birth weight by 90 grams. Two similar cooking fuel trials in rural Nepal also found null results on health outcomes (Katz et al., 2020). As Jack et al. (2021) note, studies of clean fuels have had surprisingly small effects on birth weight.\(^\text{19}\) A meta-analysis of the effects in Jack et al. (2021) and Barn et al. (2018b) using a random-effects framework yields a 95% confidence interval of (-0.13, 0.09) standard deviations. Based on birth weight standard deviation of 450 grams, this rejects benefits above 40 grams, while the pooled decrease in PM 2.5 exposure across the two trials was 18.

As Alexander and Schwandt (2022) note, effects from correlational studies of air pollution and health outcomes may be biased downward due to measurement error, but quasi-experimental approaches can correct for this bias. These randomized studies also address the measurement error issue. If pollution sensors are doing a poor job at measuring the biologically-relevant pollution exposure, the random assignment inherent in these trials can be viewed as an instrumental variable similar to the Volkswagen sales. These experimental studies, while unable to reject some large impacts on fetal health, present another puzzle compared to the quasi-experimental evidence.\(^\text{19}\) Thompson et al. (2011) estimate a positive impact on birth weight of chimneys but the confidence interval is wide and includes zero.
6 Discussion

This paper uses the causal effects of air pollution estimated in Alexander and Schwandt (2022) to predict birth outcomes in extremely polluted cities in middle-income countries and in US counties from the 1970s. I find that birth weights in highly polluted cities, which average five times the particulate matter of the contemporary United States, are surprisingly normal. So are birth outcomes in the most polluted US places in the 1970s, which, despite decreases in pollution, show no signs of improvement when I analyze outcomes in the same areas from 2002. The findings present a puzzle. Based on the causal estimates, infants in these extreme environments should have worse outcomes.

In addition to the birth weight estimates in this paper, it is striking that Goldin and Margo (1989) find normal birth weights by today’s standards in a 19th century poor house. Are birth weights unusually hard to change? While far from exhaustive, I searched all reviews of randomized trials in the Cochrane Library targeting either birth weight or low birth weight. According to meta-analyses, many treatments come up short, including: zinc (Carducci et al., 2021), calcium (Buppasiri et al., 2015), deworming (Salam et al., 2021), vitamin E (Rumbold et al., 2015), vitamin A (McCauley et al., 2015), vitamin C (Rumbold and Crowther, 2015), iodine (Harding and De-Regil, 2017), and magnesium (Makrides et al., 2014). On the other hand, the reviews find either increases in birth weight or decreases in low birth weight for: folic acid (Lassi et al., 2013), vitamin D (Palacios et al., 2019), omega-3 (Middleton et al., 2018), and anti-malarial bed nets (Gamble et al., 2006).\(^\text{20}\) Also, birth outcomes within the US still vary substantially across groups: There is a stark income gradient (Martinson and Reichman, 2016), with mothers in the bottom income quintile more than twice as likely to have a low birth weight infant compared to mothers in the top quintile of income (Table 2), suggesting that access to resources could drive poor birth outcomes.

Pollution has many potential health effects (Landrigan et al., 2018), but impacts on fetal health are especially important because of the potential long-run consequences on life outcomes (Almond and Currie, 2011). Indeed, investment in child health has some of the best benefit-cost ratios compared to other policies (Hendren and Sprung-Keyser, 2020). Given the uncertainties outline here, a randomized study of air purification could provide key evidence on the causal effects of air pollution.

\(^{20}\text{Cochrane reports “low-quality evidence” that iron (Peña-Rosas et al., 2015) and gum disease treatment (Iheozor-Ejiofor et al., 2017) can increase birth weight. They also find borderline insignificant effects of social and emotional support (East et al., 2019).}\)
A  Additional Figures

Figure A.1: City-level birth and pollution in the US, 1972-2002

Notes: Birth outcomes from NCHS Natality Files, pollution data from EPA Air Quality Trends report. Low birth weight is defined as birth weight under 2,500 grams.
Figure A.2: Low birth weight and constant gestation birth weight, US counties 1972-2002

Notes: Birth outcomes from NCHS Natality Files, pollution data from EPA sensors. Low birth weight is defined as birth weight under 2,500 grams. The constant gestation birth weight is calculated at the infant level as:

\[ 40 \times (\text{Birth Weight}) / (\text{Weeks of Gestation}) \]
**Figure A.3**: Trends in US birth weights with and without controlling for maternal characteristics

Notes: These panels show the estimated coefficients on year dummies in regressions using a 1 percent sample of the natality files from 1968 to today. The controlled regression includes fixed effects for mother age, race, live birth order, and state of residence. In Panel (A), the outcome is birth weight in grams. In Panel (B), the outcome is an indicator for being low birth weight (<2,500 grams). Source is the US Natality files (NCHS, 2023).
Figure A.4: Results censoring the US birth weight distribution

Notes: These panels show how the international sample’s birth weights would compare to their predicted values when censoring the bottom 0 to 5 percent of the US birth weight distribution. In Panel (A), the solid black line shows how average US birth weight would change removing the bottom X percent of its distribution, where X is the value in the x-axis. The flat red line shows the average birth weight across the sampled cities. The two prediction lines show how the predictions from the causal estimates would change when the censored US estimate is used as the reference. Panel (B) is analogous, but with percent low birth weight as the outcome.
### Additional Tables

**Table B.1: Birth weights and pollution in the most polluted US cities, 1972**

<table>
<thead>
<tr>
<th>Place</th>
<th>TSP</th>
<th>PM 2.5</th>
<th>Birth weight</th>
<th>% LBW</th>
<th>% VLBW</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phoenix, Arizona</td>
<td>159</td>
<td>48</td>
<td>3,298</td>
<td>7.0</td>
<td>0.9</td>
<td>5,583</td>
</tr>
<tr>
<td>El Paso, Texas</td>
<td>142</td>
<td>43</td>
<td>3,267</td>
<td>6.7</td>
<td>0.7</td>
<td>4,112</td>
</tr>
<tr>
<td>Wichita, Kansas</td>
<td>142</td>
<td>43</td>
<td>3,276</td>
<td>8.8</td>
<td>1.0</td>
<td>2,407</td>
</tr>
<tr>
<td>Newark, New Jersey</td>
<td>134</td>
<td>40</td>
<td>3,101</td>
<td>13.1</td>
<td>2.1</td>
<td>3,879</td>
</tr>
<tr>
<td>Birmingham, Alabama</td>
<td>131</td>
<td>39</td>
<td>3,217</td>
<td>8.9</td>
<td>1.0</td>
<td>3,094</td>
</tr>
<tr>
<td>Detroit, Michigan</td>
<td>102</td>
<td>31</td>
<td>3,151</td>
<td>11.8</td>
<td>2.2</td>
<td>13,511</td>
</tr>
<tr>
<td>Denver, Colorado</td>
<td>104</td>
<td>31</td>
<td>3,134</td>
<td>9.4</td>
<td>1.1</td>
<td>4,262</td>
</tr>
<tr>
<td>1972 Combined</td>
<td>124</td>
<td>37</td>
<td>3,193</td>
<td>9.9</td>
<td>1.5</td>
<td>36,848</td>
</tr>
<tr>
<td>USA, 2019</td>
<td>33*</td>
<td>10</td>
<td>3,254</td>
<td>8.3</td>
<td>1.4</td>
<td>3,753,815</td>
</tr>
</tbody>
</table>

*Notes: This table shows pollution and birth weights for extremely polluted counties in the US in 1972. TSP is total suspended particles as reported in the EPA Air Quality Trends report for 1972. The PM2.5 column is estimated by multiplying TSP by 0.3, except for the USA row. Birth weight is from the NHCS natality file for 1972 and (for the USA row) 2019. LBW is defined as birth weight under 2,500 grams, VLBW is birth weight under 1,500 grams. N is the sample size in that place.*
Table B.2: International sample with country-level attributes

<table>
<thead>
<tr>
<th>City</th>
<th>Country</th>
<th>Life expectancy</th>
<th>GDP per capita</th>
<th>Birth weight (g)</th>
<th>LBW</th>
<th>Urban PM2.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td>USA</td>
<td>78.8</td>
<td>65,280</td>
<td>3,286</td>
<td>0.07</td>
<td>8.4</td>
</tr>
<tr>
<td>Doha</td>
<td>Qatar</td>
<td>80.2</td>
<td>62,276</td>
<td>0.12</td>
<td></td>
<td>62.4</td>
</tr>
<tr>
<td>Abu Dhabi</td>
<td>United Arab Emirates</td>
<td>78.0</td>
<td>42,701</td>
<td>3,080</td>
<td>0.07</td>
<td>43.0</td>
</tr>
<tr>
<td>Brescia</td>
<td>Italy</td>
<td>83.2</td>
<td>33,642</td>
<td>0.07</td>
<td></td>
<td>17.3</td>
</tr>
<tr>
<td>Monza</td>
<td>Italy</td>
<td>83.2</td>
<td>33,642</td>
<td>3,171</td>
<td>0.09</td>
<td>17.3</td>
</tr>
<tr>
<td>Makkah</td>
<td>Saudi Arabia</td>
<td>75.1</td>
<td>23,140</td>
<td>0.10</td>
<td></td>
<td>57.0</td>
</tr>
<tr>
<td>Muscat</td>
<td>Oman</td>
<td>77.9</td>
<td>17,701</td>
<td>0.03</td>
<td></td>
<td>35.8</td>
</tr>
<tr>
<td>Temuco</td>
<td>Chile</td>
<td>80.2</td>
<td>14,742</td>
<td>3,386</td>
<td></td>
<td>22.6</td>
</tr>
<tr>
<td>Beijing</td>
<td>China</td>
<td>76.9</td>
<td>10,144</td>
<td>3,343</td>
<td>0.04</td>
<td>49.8</td>
</tr>
<tr>
<td>Astana</td>
<td>Kazakhstan</td>
<td>73.2</td>
<td>9,813</td>
<td>0.05</td>
<td></td>
<td>30.9</td>
</tr>
<tr>
<td>Belgrade</td>
<td>Serbia</td>
<td>75.7</td>
<td>7,417</td>
<td>3,259</td>
<td></td>
<td>26.3</td>
</tr>
<tr>
<td>Tuzla</td>
<td>Bosnia and Herzegovina</td>
<td>77.4</td>
<td>6,120</td>
<td>3,387</td>
<td>0.06</td>
<td>31.9</td>
</tr>
<tr>
<td>Skopje</td>
<td>North Macedonia</td>
<td>75.8</td>
<td>6,070</td>
<td>3,325</td>
<td>0.05</td>
<td>30.8</td>
</tr>
<tr>
<td>Ulaanbaatar</td>
<td>Mongolia</td>
<td>69.9</td>
<td>4,405</td>
<td>3,490</td>
<td>0.04</td>
<td>47.1</td>
</tr>
<tr>
<td>Colombo</td>
<td>Sri Lanka</td>
<td>77.0</td>
<td>3,852</td>
<td>2,920</td>
<td>0.16</td>
<td>23.9</td>
</tr>
<tr>
<td>Tunis</td>
<td>Tunisia</td>
<td>76.7</td>
<td>3,575</td>
<td>3,281</td>
<td></td>
<td>26.4</td>
</tr>
<tr>
<td>Hanoi</td>
<td>Vietnam</td>
<td>75.4</td>
<td>2,715</td>
<td>3,251</td>
<td>0.03</td>
<td>21.1</td>
</tr>
</tbody>
</table>

*Note:* Life expectancy and GDP per capita data are at the country level for 2019. Source for life expectancy is World Bank (2022b) and source for GDP is World Bank (2022a). The estimate of country-level urban PM2.5 is from WHO (2023). Sources for city-level birth outcomes are given in Appendix C.
References


Buppasiri, Pranom, Pisake Lumbiganon, Jadsada Thinkhamrop, Chetta Ngamjarus, Malinee Laopaiboon, and Nancy Medley, “Calcium supplementation (other than for preventing or treating hypertension) for improving pregnancy and infant outcomes,” *Cochrane Database of Systematic Reviews*, 2015, (2).

Carducci, Bianca, Emily C Keats, and Zulfiqar A Bhutta, “Zinc supplementation for improving pregnancy and infant outcome,” *Cochrane Database of Systematic Reviews*, 2021, (3).


Collivignarelli, Maria Cristina, Alessandro Abbà, Francesca Maria Caccamo, Giorgio Bertanza, Roberta Pedrazzani, Marco Baldi, Paola Ricciardi, and Marco Carnevale Miino, “Can particulate matter be identified as the primary cause of the rapid spread of CoViD-19 in some areas of Northern Italy?,” *Environmental Science and Pollution Research*, 2021, 28, 33120–33132.


Di, Qian, Itai Kloog, Petros Koutrakis, Alexei Lyapustin, Yujie Wang, and Joel Schwartz, “Assessing PM2. 5 exposures with high spatiotemporal resolution across the continental United States,” *Environmental science & technology*, 2016, 50 (9), 4712–4721.

Dong, Wei, Shan Liu, Mengtian Chu, Bin Zhao, Di Yang, Chen Chen, Mark R Miller, Miranda Loh, Junhui Xu, Rui Chi et al., “Different cardiorespiratory effects of indoor air pollution intervention with ionization air purifier: findings from a randomized, double-blind crossover study among school children in Beijing,” Environmental Pollution, 2019, 254, 113054.

Donkelaar, Aaron Van, Randall V Martin, Chi Li, and Richard T Burnett, “Regional estimates of chemical composition of fine particulate matter using a combined geoscience-statistical method with information from satellites, models, and monitors,” Environmental science & technology, 2019, 53 (5), 2595–2611.


East, Christine E, Mary A Biro, Suzanne Fredericks, and Rosalind Lau, “Support during pregnancy for women at increased risk of low birthweight babies,” Cochrane Database of Systematic Reviews, 2019, (4).


Han, Tao, Dan Wang, Wenyu Xie, Changgen Liu, Qian Zhang, Zhichun Feng, and Qiuping Li, “Obstetricians’ Attitudes Toward the Treatment of Extremely Preterm Infants in China,” *JAMA Network Open*, 2022, 5 (9), e2233511–e2233511.


Li, Xiangyu, Shuqiong Huang, Anqi Jiao, Xuhao Yang, Junfeng Yun, Yuxin Wang, Xiaowei Xue, Yuanyuan Chu, Feifei Liu, Yisi Liu et al., “Association between ambient fine particulate matter and preterm birth or term low birth weight: an updated systematic review and meta-analysis,” *Environmental Pollution*, 2017, 227, 596–605.

Liang, Lu, Peng Gong, Na Cong, Zhichao Li, Yu Zhao, and Ying Chen, “Assessment of personal exposure to particulate air pollution: the first result of City Health Outlook (CHO) project,” *BMC public health*, 2019, 19 (1), 1–12.


Lucila, Rey-Ares, Irazola Vilma Edit, Althabe Fernando, Sobrino Edgardo, Mazzoni Agustina, Serón Pamela, Lanas Fernando, Calandreli Matías, and Rubinstein Adolfo, “Lower respiratory tract infection in children younger than five years of age and adverse pregnancy outcomes related to household air pollution in Bariloche (Argentina) and Temuco (Chile),” *Indoor air*, 2016, 26 (6), 964.


Makrides, Maria, Danielle D Crosby, Emily Shepherd, and Caroline A Crowther, “Magnesium supplementation in pregnancy,” *Cochrane Database of Systematic Reviews*, 2014, (4).


Moschandreas, Demetrios J, DJ Pelton, and J Zabransky, Comparison of indoor and outdoor air quality, Electric Power Research Institute, 1981.


Shaddick, Gavin, Matthew L Thomas, Amelia Green, Michael Brauer, Aaron Donkelaar, Rick Burnett, Howard H Chang, Aaron Cohen, Rita Van Dingenen, Carlos Dora et al., “Data integration


Tran, Toan K, BO Eriksson, Chuc TK Nguyen, Peter Horby, Goran Bondjers, and Max Petzold, “DodaLab: an urban health and demographic surveillance site, the first three years in Hanoi, Vietnam,” Scandinavian journal of public health, 2012, 40 (8), 765–772.


Zhang, Ao, Yumeng Liu, Bin Zhao, Yinping Zhang, Haidong Kan, Zhuohui Zhao, Furong Deng, Chen Huang, Xiangang Zeng, Yuexia Sun et al., “Indoor PM2.5 concentrations in China: a concise review of the literature published in the past 40 years,” Building and Environment, 2021, 198, 107898.


C Data sources

All pollution data comes from WHO (2016b) unless noted below. Preterm means born at gestation < 37 weeks unless noted.

Abu Dhabi, United Arab Emirates

Taha et al. (2022). Mean birth weight: 3,080g (Table 1). N=1,610. Study inclusion: “Mothers with complete data on sociodemographic factors (such as age and education), health factors related to pregnancy and the mode of delivery, and breastfeeding practices (such as breastfeeding initiation) were considered as participants in this study” (p. 3). Their treatment of multiple births is not mentioned. Pollution year: 2013. An alternate pollution estimate closer in time is very similar: Abuelgasim and Farahat (2020).

Astana, Kasakhstan

Aimukhametova et al. (2012). Low birth weight incidence: 5.1% (Table 2). Preterm: 8.3% (Table 2). N = 157. Women with completed singleton pregnancies at a tertiary hospital in Astana who gave birth at 24 weeks or more of gestation (p. 84).

Pollution source Kenessariyev et al. (2013). Annual average PM2.5 estimate: 52.9 (Table 4). Used because Kazakhstan is not included in the WHO (2016b) database.

Beijing, China

Su et al. (2016). Mean birth weight: 3,343g. Low birth weight incidence: 3.6%. (In both cases these are weighted averages taken using outcomes and group sizes in Table 3.) Preterm: 5.0% (Table 2). N=5,479. Sampling: ‘Medical and obstetrical data for each participant was collected from 15 hospitals in Beijing by a systemic cluster sampling survey conducted from 20 June 2013 to 30 November 2013” (p. 1062). Eligibility was defined as all women “who delivered a live born singleton infant between 20 June 2013 and 30 November 2013 and that were born at 1970 or later” (p. 1062), so mothers could not be older than 43. Pollution year: 2013.
Beirut, Lebanon

Tamim et al. (2007). Low birth weight incidence: 9.18% This comes from counting the births in the low birth weight categories broken out in Table 2. N=18,727. Sampling: All births from nine major hospitals in Beirut and its suburbs in 2001-2002. Pollution year: 2014. An alternate pollution estimate closer in time to the births is very similar: Saliba et al. (2004).

They report a preterm rate of 15.9%, but defined as gestation ≤ 37 weeks instead of < 37 weeks. The US rate would be higher than 15.9% if defined this way.

Belgrade, Serbia

Maric et al. (2010). Mean birth weight: 3,259; N=2,581 (Table 2). Preterm: 8.7% (Table 2; this is defined as ≤ 37 weeks, p. 83). Sample: Births at the Institute of Gynecology and Obstetrics in Belgrade, 1996 to 2003. Only the control group, which was not exposed to the bombings in March-June 1999, is used. Multiple births included (Table 2). Stillbirth rate: 1.3%.


Brescia, Italy


Colombo, Sri Lanka


Doha, Qatar

Abdulkader et al. (2013). Low birth weight incidence: 12.36% (Table 1). Preterm: 14% (Table 3). N = 890. Sampling: births at Women’s Hospital HMC from Mach to April 2011. “All babies born live during the study period, including multiples, were included in the study” (p. 33). Pollution year: 2012.
**Hanoi, Vietnam**

Nguyen et al. (2012). Mean birth weight: 3,251 grams; N=537. Estimate is the weighted average of “urban boy” and “urban girl” birth weight means in Table 1. Survey data on children born at the Dodalab (Tran et al., 2012) in Dongda, Hanoi, March, 2009 to June 2010. The authors excluded 20 children who were either twins or born with a congenital disease.

**Pollution source** Luong et al. (2017). Sample period: September 2010 to September 2011. PM 2.5 Estimate: 67. Used because Vietnam is missing from the WHO (2016b) list.

**Makkah, Saudi Arabia**

General Authority for Statistics (2016). Low birth weight incidence: 9.28%. Calculated using the counts reported in Table 24-1 “Number of Saudi live births during the 5 years preceding the survey, by sex and weight of the child at birth and administrative Area” on page 78. N = 443,128. No sample restrictions reported. Pollution year: 2014.

**Monza, Italy**

Ornaghi et al. (2022). Average birth weight: 3,170.5 grams. Reported in the text on page 468. Preterm: 6.7% (Table 2). N = 1,882. Hospital data from women giving both at a hospital in Monza. (The 2020 sample excludes women with a symptomatic COVID infection; I use the 2019 estimate.) Stillbirth rate: 0.4%. Pollution year: 2013. Alternative PM2.5 estimate: 31 (Collivignarelli et al., 2021, Figure 4).

**Muscat, Oman**

Ministry of Health Sultanate of Oman (2012). Low birth weight incidence: 9.95% (Table 8-5: Births in MOH Hospitals and Health Centres according to Governorates during 2012). N=13,576. All births in Royal Hospital, Khawlah Hospital, and Qurayyat Hospital in 2012. Pollution year: 2009.

**Skopje, North Macedonia**


**Temuco, Chile**

Lucila et al. (2016). Average birth weight: 3,386 grams; N=339 (Table 1). Low birth weight incidence:
3.3% (Table 1). Preterm: 6.5% (Table 1). Sample: consenting households with “at least one child under the age of five had resided in the home in the past five years and/or a child had been born alive or stillborn in the household in the last 3 years” (p. 4). Birth weight was collected from health records (p. 4). Pollution year: 2014.

**Tunis, Tunisia**


**Tuzla, Bosnia and Herzegovina**


**Ulaanbaatar, Mongolia**

Barn et al. (2018b, Table 3). Mean birth weight: 3,490 grams, N=223. (This mean is reported in Section 2.7. Table 3 reports the main outcomes but using median rather than mean birth weight). Sampling: “We recruited women who met the following criteria: ≥ 18 years of age, ≤ 18 weeks of a single-gestation pregnancy, non-smoker, living in an apartment, planning to give birth in a medical facility in the city, and not using a residential portable air cleaner at enrollment. We excluded women who lived in ger households because electricity is unreliable in ger neighborhoods and gers may have higher indoor-outdoor air exchange rates, which reduces HEPA cleaner effectiveness” (p. 982).

Since Ulaanbaatar features prominently in this analysis, I note another pollution source (Ganbat et al., 2020). Averaging across eight sensors placed throughout the city (Table 1), the authors report a yearly mean PM2.5 of between 50 and 86 in 2014-2019 (Table 3).

**United States**

I use US natality data from 2019 to calculate birth weight outcomes in the US (NCHS, 2023). This allows me to exclude non-singleton births.

- **Singletons only**: Mean birth weight: 3,286 grams. Low birth weight incidence: 6.67%. N=3,630,218.
• **All births**: Mean birth weight: 3,254 grams. Low birth weight incidence: 8.31%. N=3,753,815.

**Pollution** I use a PM2.5 of 10 for the US. This is from the outcome mean in Table 2B column (1) of *(Alexander and Schwandt, 2022)*, which uses EPA data from 2007-2015. This is slightly higher than other estimates. *WHO (2023)* report an average of 8.4 for the US in 2010-2019. The population-weighted average across counties in *Wu et al. (2020)* for 2000-2016 is 9.1. Using these lower estimates of PM2.5 would heighten the difference between the predicted and actual birth weights in the polluted places. The US 99th percentile of 15.35 comes from taking the population-weighted 99th percentile of county-year PM2.5 estimates from *Wu et al. (2020)*, weighted by population.
D Additional sources

Here I detail some sources that did not satisfy the inclusion criteria.

Teplice, 1994-1997

Teplice was one of the most polluted cities in the world in the 1990s, but not currently, so it was not included in WHO (2016b). The Czech government began a collaboration with the US EPA in the early 1990s to study the health consequences of pollution. Sram et al. (1996) (the “Teplice Program”) is a detailed report on pollution exposure and its consequences in Teplice. The city had a mean PM2.5 of 122 in January-March 1993 and 28.7 in May-August 1993, and Dejmek et al. (1999) report a mean of 47.6 from May 1993 to March 1996.

At the time of writing, Sram et al. (1996) were in the middle of a study of birth outcomes in Teplice from 1994-1997. They report data for the first 15 months of the study, covering all hospitalized births in Teplice. Out of the 1,626 births in Teplice, the incidence of low birth weight was 8.8% and the incidence of premature birth was 6.2%. The same figures for the 1,380 births to European parents were 6.4% and 5.0% (Table 8). In the US in 2021, the preterm birth rate was 10.5% (CDC, 2023) and about 8% for singletons (Martin and Osterman, 2021). These findings were expanded on in Dejmek et al. (1999), but for that study the authors dropped preterm births. They report a similar fraction of preterm births. Mean birth weight in the sample of term births was 3,354 grams.

Their research also confirmed, using personal censors from the US EPA, that personal exposure was indeed high in Teplice. In a sample of women with outdoor jobs, PM2.5 exposure as measured by the personal sensors ranged from 33 to 106 depending on the month (Table 2 in Sram, 2020).

Beijing term births, 2007-2009

Rich et al. (2015) study the effect of a large pollution reduction in Beijing due to the Olympics. They analyze birth outcomes in four districts where they deployed specific pollution sensors. Their birth data comes from Beijing hospitals in those districts in 2007-2009 and they begin with 140,298 births. Unfortunately, their study is restricted only to term births (37-41 weeks of gestation) so it cannot be

\[^{21}\text{It’s possible Sram et al. (1996) restricted to singletons although it’s not reported. Sram et al. (1996) also note that, at the time, 42% of women from Teplice smoked in early pregnancy (p. 709). This compares to 7.2% in the US in 2016 (Drake et al., 2018).}\]

\[^{22}\text{Dejmek et al. (1999, p. 477) dropped 92 preterm births out of 2,045 births or 4.5% after dropping non-European births, twins, duplicate mothers. Mean birth weight is calculated using the two groups in Table 1.}\]
used as population estimates. But in their methods (p. 881), they describe that this entails dropping about 5% of the sample. This suggests that the rate of premature birth is much lower in Beijing, since recent estimates put the US around 10% (CDC, 2023). Their sensors measured PM2.5 above 60 for all three years of the sample (Figure 1).

**Thailand cities, 2006**

Kosirinond and Son (2008) is a United Nations report on Thailand birth weights in 2006 in multiple places, including two of the highly-polluted cities in WHO (2016b): Saraburi: low birth weight incidence: 9.12%, N=8,741. Nakhon Sawan: 9.10%, N=10,548 (“Low birth weight by sex, and province” p. 221). The calculation includes all live births (see the data description on page 218). However, the WHO (2016b) pollution estimates are from 2014, more than five years from these birth weight outcomes. I could not find pollution estimates taken closer to 2006.