9-2011

Three Essays on Racial Disparities in Infant Health and Air Pollution Exposure

Helen Scharber
University of Massachusetts Amherst, helenscharber@gmail.com

THREE ESSAYS ON RACIAL DISPARITIES IN INFANT HEALTH AND AIR POLLUTION EXPOSURE

A Dissertation Presented

by

HELEN SCHARBER

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

September 2011

Economics

THREE ESSAYS ON RACIAL DISPARITIES IN INFANT HEALTH AND AIR POLLUTION EXPOSURE

A Dissertation Presented

by

HELEN SCHARBER

Approved as to style and content by:

Michael Ash, Chair

James K. Boyce, Member

Arindrajit Dube, Member

Sylvia Brandt, Member

Michael Ash, Department Chair
Economics

ACKNOWLEDGMENTS

For their enthusiastic and patient support over the last several months, I would like to thank my dissertation committee. I am deeply grateful to my chair, Michael Ash, for his seemingly unlimited patience with questions related to the RSEI dataset, to Arindrajit Dube for his many valuable econometrics contributions, to Sylvia Brandt, for offering valuable suggestions as I was conceptualizing the research and to Jim Boyce for helping me see both the forest and the trees. I would also like to thank John Klinger, Julie Galernault, Larry Breitborde, Steve Cohn, Brian Doherty, Robert Pollin and the many other teachers who have guided me over the years.

For their collective wisdom, passion, compassion and camaraderie, I would like to thank my fellow graduate students. In particular, I would have been lost if not for the intellectual and emotional support of Jessica Carrick-Hagenbarth, Grace Chang, Jen Cohen, Hasan Comert, Nina Eichacker, Heidi Garrett-Peltier, Joo-Yeoun Suh and Ben Zipperer. I am also indebted to the staffs of Rao's Coffee and the Insight Meditation Society, Eric Burri, Don Blair, Judy Fogg, Robert Glassman, Tom Johnson, Carol Kaminsky, Alex Phillips and Carl South for their many and various contributions to the fulfillment of my material and spiritual needs.

Finally, to my parents, Mary and Gary Scharber. I couldn't have done it without you. Thank you so much for everything.

ABSTRACT

THREE ESSAYS ON RACIAL DISPARITIES IN INFANT HEALTH AND AIR POLLUTION EXPOSURE

SEPTEMBER 2011

HELEN SCHARBER

B.A., KNOX COLLEGE

M.A., KEELE UNIVERSITY

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Michael Ash

This three-essay dissertation examines racial disparities in infant health outcomes and exposure to air pollution in Texas. It also asks whether the EPA's Risk-Screening Environmental Indicators Geographic Microdata (RSEI-GM) might be used to assess the effects of little-studied toxic air pollutants on infant health outcomes. Chapter 1 contributes to the "weathering" literature, which has shown that disparities in infant health outcomes between non-Hispanic black and non-Hispanic white women tend to widen with age. In this study, we ask whether the same patterns are observed in Texas and among Hispanic women, since other studies have focused on black and white women from other regions. We find that black and Hispanic women in Texas do "weather" earlier than white mothers with respect to rates of low birthweight and preterm birth. This differential weathering appears to be mediated by racial disparities in the distribution and response to socioeconomic risk factors, though a large

gap between black and white mothers across all ages remains unexplained. Chapter 2 extends the statistical environmental justice literature by examining the distribution of toxic air pollution across infants in Texas. We find that, within Texas cities, being black or Hispanic is a significant predictor of how much pollution one is exposed to at birth. We further find that, among mothers who move between births, white mothers tend to move to significantly cleaner areas than black or Hispanic mothers. In Chapter 3, we use geocoded birth records matched to square-kilometer pollution concentration estimates from the RSEI-GM to ask whether the pollution-outcome relationships that emerge through regression analysis are similar to the effects found in previous research. If so, the RSEI-GM might be used to study the health effects of nearly 600 chemicals tracked in that dataset. We conclude, based on instability of results across various specifications and lack of correspondence to previous results, that this data is not appropriate for statistical epidemiology research and better exposure data for toxic air pollutants are needed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure **Page**

INTRODUCTION

Overview

In the United States, the idea that everyone should be guaranteed equal opportunities at birth is a nearly universal ethical precept. Yet as recently as 2008, the rates of preterm birth and low birthweight for Non-Hispanic Black infants in the U.S., at 17.5 and 13.7, were closer to those found in many African countries than the corresponding rates for Non-Hispanic White newborns, 11.1 and 7.2 [108]. Fetal growth restriction and low birthweight are troubling because they are associated with a higher risk of neonatal mortality, higher adult incidences of heart disease and diabetes, and diminished adult socioeconomic outcomes [101, 36]. Moreover, racial disparities in infant health outcomes have been found to widen with maternal age. Geronimus and others have attributed this early minority "weathering" to the cumulative and interacting effects of low socioeconomic status among minority women, but definitive explanations for the gap remain elusive [60, 61]. If we want to level the playing field, however, documenting and understanding these disparities are important first steps.

The right of all people to a clean and safe environment is another commonly held value, predicating other fundamental rights to life, liberty and the pursuit of happiness. Case studies and statistical analyses in the environmental justice literature, however, find that minority and low-income groups in the U.S. are disproportionately exposed to polluted air and water [74, 97]. That these groups bear the pollution costs of industrial production is particularly disturbing, given that they are unlikely enjoy the associated benefits, in terms of consumer surplus or profits, in equal proportion. Even more alarming, studies have found that minority race is a significant predictor

of pollution exposure, independent of income. Some have theorized that this environmental racism may contribute to the racial gap in health outcomes, though relatively little is known about many of the pollutants in question and little research has been done to explicitly connect these issues. Here, again, documenting the disparities and identifying the health effects of pollution are necessary steps toward erasing inequalities and guaranteeing fundamental rights.

In this dissertation, we use geocoded birth data from the state of Texas matched to industrial toxic air pollution data from the EPA's Risk-Screening Environmental Indicators Geographic Microdata (RSEI-GM) to document and explain racial disparities in infant health outcomes and infant pollution exposure. We further we ask whether it is possible to use our data to connect these two issues. In Chapter 1, we document racial disparities in the relationship between age and birth outcomes, finding that Black and Hispanic women in Texas "weather" earlier than White mothers. This differential weathering appears to be mediated by racial disparities in the distribution and response to socioeconomic risk factors, though a large gap between Black and White mothers across all ages remains unexplained. In Chapter 2, we find that, within Texas cities, being Black or Hispanic is a significant predictor of how much pollution one is exposed to at birth. We further find that, among mothers who move between births, White mothers move to significantly cleaner areas than Black or Hispanic mothers. In Chapter 3, we ask whether we can use matched birth and pollution data to better understand the effects of toxic air pollution on health. We conclude, based on instability of results across various specifications and lack of correspondence to previous results, that this data is not appropriate for statistical epidemiology research and better exposure data for toxic air pollutants are needed.

Chapter 1 Summary

Recently, a number of studies have found that well-established disparities in White and Black infant health outcomes tend to increase with maternal age. Since Geronimus's seminal article on infant mortality in 1992 [60], research has shown racial disparities in the relationship between age and low birthweight [61, 92, 34, 95], preterm birth [46, 4, 98, 87, 66] and other infant health outcomes [114, 21, 73]. To explain these differences, Geronimus proposed a "weathering" or "accelerated aging" hypothesis, which connects poorer reproductive outcomes as women age to declining overall health and further asserts that social inequalities lead to an earlier and greater decline in the health status of Black women [60, 61].

In Chapter 1, we ask whether Black and White mothers in Texas exhibit racial disparities in birth outcomes and age-outcome relationships similar to those found in other parts of the country, and we ask how U.S.- and foreign-born Hispanic mothers compare. We find large overall low birthweight (LBW) and preterm birth (PTB) disparities between Black and White mothers in Texas similar to those found in virtually every study examining Black-White differences in birth outcomes. We also find evidence of accelerated weathering among Black mothers, corresponding to the findings of most LBW weathering studies [61, 92, 95, 73] and some PTB studies [87, 66]. Our results also support the "Hispanic paradox" in that we observe relatively small overall disparities between White and Hispanic mothers. Significantly, however, we observe early weathering among both U.S.- and foreign-born Hispanic mothers compared to Whites. While one study examined age-LBW relationships for U.S.- and foreign-born Mexican women and found them to be similar, with U.S.-born women showing higher overall rates of LBW [114], no previous study had compared these relationships in Hispanic and non-Hispanic populations.

We also ask to what extent the weathering hypothesis, which asserts that racial disparities in weathering are caused by differential prevalence of socioeconomic status

(SES) related risk factors, can explain the observed disparities. We address this by including several such risk factors as independent variables in multivariate age-outcome regressions. We find that their inclusion does reduce observed disparities in both LBW and PTB, a finding that supports the weathering hypothesis and is consistent with previous research. That our group of confounding variables is far from complete provides support for the idea that exposure to low SES conditions may be more responsible for early weathering than we could observe. We find that the effect of most risk factor variables varies significantly by race, with Blacks generally experiencing more negative effects than Whites and foreign-born Hispanic mothers experiencing less negative effects. We further find that the effects of many risk factors, including those most associated with low SES—low education, unmarried status, smoking, and late prenatal care initation—increase with age.

Significantly, even after controlling for SES risk factors and including age and race interactions, we found an independent "effect" of being Black. In other words, our data is unable to explain the higher incidence of poor birth outcomes among Black women across ages. Indeed, we found that the coefficient on Black race either remained the same (PTB) or increased (LBW) after adding risk factor controls to the bivariate model; controlling for the average effects of education, marital status, smoking behavior, and other variables could not explain Black-White differences. Our control variables are certainly far from comprehensive, but previous research has also failed to fully explain these disparities. Researchers in the field of health inequalities have theorized about a range of material and psychosocial explanations for the gap, including disparities in access to nutritious food or healthcare, different behaviors and habits related to health, and different stress levels related to socioeconomic status [104]. Differential exposure to pollution has also been suggested as a causal link, and we explore this in subsequent chapters.

Chapter 2 Summary

Since the publication of the report *Toxic Wastes and Race in the United States* in 1987 [109], dozens of studies have documented the existence of systematic disparities in exposure to environmental hazards along race and class lines in the U.S., with minority race often predictive of disproportionate exposure even when controlling for socioeconomic status [74, 97]. In Chapter 2, we contribute to this literature, but taking a cue from a recent World Bank report, we focus on documenting inequities among infants in particular, on the basis that such inequities are particularly disturbing and unjustifiable [42].

We find, using summary distributional measures and multivariate regressions, that infants in Texas are not born into a land of equal opportunity, at least where industrial toxic air pollution is concerned. Rather, minority children are disproportionately exposed even before birth. Our results reflect what Ash and Fetter (2004) found for all residents at the national level; Black infants are more likely than others to be born into high-pollution cities, and within cities *and even within neighborhoods*, both Black and Hispanic infants are more likely to live on the wrong side of the environmental tracks [9].

Also in Chapter 2, we exploit our ability to track mothers across births in our dataset to ask how race is associated with the change in pollution exposure between sibling births. There has been much speculation but relatively little research on the dynamics of pollution distribution, and unfortunately, our dataset, covering 1995-2003, is not long enough to address questions about discriminatory facility siting. Still, trends in more recent years may be helpful in forming policies to ameliorate inequities in exposure. Our analysis finds different patterns for mothers who moved between births and mothers who did not. Relative to their original location, White mothers who move between births relocate to less polluted places than do Black and

Hispanic mothers, but this racial differentiation does not seem to exist among mothers who did not move between births.

Chapter 3 Summary

Unfortunately, research documenting racial disparities in air pollution exposure is more plentiful than research documenting the health effects of suspected toxicants. According to Suh et al. (2000), "relatively little has been done to characterize the concentrations, exposures, and health risks for most of the hazardous air pollutants (HAPs). Still less is known about the human health effects of HAP exposures at concentrations found in the ambient environment, as most of what is known has been obtained from occupational and animal studies" [105, p. 629]. In response to a 1997 Environmental Defense Fund report criticizing the lack of information about health effects of the chemicals most prevalent in U.S. production, the EPA published a report titled *What Do We Really Know About the Safety of High Production Volume Chemicals?*[47, 54]. In the 1998 report, the EPA revealed that no basic toxicity information was available for 43 percent of the 2,800 chemicals that are produced or imported in quantities of over 1 million pounds per year, and a full set of basic toxicity information was available for only 7 percent. Information on developmental or pediatric toxicity was available for only one-fifth of the chemicals at the time of publication, and more recent literature indicates that little progress has been made in the intervening decade [54, 63].

In Chapter 3, we ask whether the rich Risk-Screening Environmental Indicators Geographic Microdata (RSEI-GM), which is based on data from the EPA's Toxics Release Inventory (TRI), might be used to explore the effects of some of these chemicals on infant health outcomes. A few existing studies have used TRI data to explore effects on birth outcomes, but these statistical studies use county-level toxic exposure estimates that do not take into account within-county heterogeneity in their

analyses [3, 39]. In addition, these studies do not account for emissions that travel outside a county's boundaries, and they tend to lump different chemicals with presumably different physiological effects and levels of toxicity into the same measure. We are fortunate, at the University of Massachusetts Amherst, to have access to the EPA's RSEI microdata, which contains modeled concentration estimates of several hundred TRI chemicals for each square kilometer in the U.S. We match this data to geocoded birth and infant death records to assess the relationship between birth outcomes and the effects of several known developmental toxicants, including cadmium, epichlorohydrin, lead and toluene. Finding robust, negative relationships between these pollutants and birth outcomes would support the use of RSEI-GM in studies of the health effects of other TRI chemicals.

We use a variety of parametric specifications and one semiparametric specification to examine the relationships between air toxics and infant health outcomes. In addition to a lack of correspondence to previous results, we find a lack of consistency across different specifications, samples and assumed relationships between birth outcomes and pollution variables. In our parametric specifications, most estimated coefficients are not statistically significant, and those that achieve statistical significance in one specification often lose it or change sign or magnitude in a slightly different specification. Further, removing outliers drastically changes results, and results in the samples with outliers removed change depending on what share of observations are removed. In our semiparametric specifications, which impose no constraints on the unknown pollution-outcome relationship, we do observe some patterns, but even these are generally difficult to explain. We conclude that our merged birth-pollution data, while appropriate for studying geographical distribution of air pollution more broadly, are not precise enough for health research. Given that the RSEI data provides the best available information on the concentrations of many TRI pollutants, better data on actual ambient concentrations is clearly needed.

CHAPTER 1

ACCELERATED WEATHERING AMONG MINORITY MOTHERS IN TEXAS: EXPLORING THE EVIDENCE

1.1 Introduction

Statistics on racial and ethnic disparities in infant health in the U.S. are startling and sometimes confusing. In 2008, the preterm birth (PTB) and low birthweight (LBW) rates for Non-Hispanic White infants were 11.1 and 7.2, respectively. The corresponding rates for Non-Hispanic Black infants, at 17.5 and 13.7, rival those found in Eastern and Southern African countries [108]. Hispanic PTB and LBW rates, at 12.1 and 7.0, are much closer to White rates. Because poor health outcomes among Blacks have often been attributed to low socioeconomic status (SES), the combination of similarly low SES profiles among Hispanics with much better average health outcomes has become known as the "Hispanic paradox" [59]. While health-related behaviors and access to resources help explain some of the variation in health outcomes across race and ethnicity lines, researchers at the National Center for Health Statistics note that these disparities mostly "remain unexplained" [75].

In addition to well-established overall disparities in White and Black[1] infant health outcomes, research has also shown that the disparities tend to increase with age. Since Geronimus's seminal article on infant mortality in 1992 [60], research has also observed racial disparities in the relationship between age and low birthweight [61, 92, 34, 95], preterm birth [46, 4, 98, 87, 66] and other infant health outcomes [114, 21, 73].

[1]Henceforth, Non-Hispanic White and Non-Hispanic Black will be referred to as White and Black for conciseness.

To explain these differences, Geronimus proposed a "weathering" or "accelerated aging" hypothesis, which connects poorer reproductive outcomes as women age to declining overall health and further asserts that social inequalities lead to an earlier and greater decline in the health status of Black women [60, 61]. Perhaps due to the "Hispanic paradox" or the demographic characteristics of the particular cities that have been studied in previous weathering research, scant attention has been paid to the relationship between age and birth outcomes among Hispanics.[2]

Researchers have found that low birthweight and preterm delivery, two indicators of poor infant health that are sometimes but not always coincident, can have negative consequences ranging from unfavorable to dire. Low birthweight (LBW) and preterm birth (PTB) are both associated with higher rates of infant mortality. A recent National Center for Health Statistics analysis of 2006 data identified LBW as one of the three leading causes of infant death and found that 36 percent of infant deaths were "preterm-related" [78]. Other researchers have found associations between LBW and lower scores on tests of intellectual and social development in childhood [18, 20] as well as lower schooling attainments, earnings, and employment probabilities in adulthood [37]. Additionally, PTB has been associated with long-term disability [72]. Racial and ethnic disparities in infant health outcomes systematically prevent some citizens from enjoying the equal opportunities that the U.S. is supposed to provide.

Our first goal in this paper is to ask whether differential weathering is observed among Black and White mothers in Texas, since other studies in the literature have focused on other states and cities. We also extend previous research by expanding the analysis to include Hispanic mothers, a group that has been left out of previous analyses. After examining the simple age-outcome relationships, we will test Geronimus's hypothesis by controlling for various socioeconomic and medical risk factors that are

[2]Wildsmith does examine weathering among US-born and foreign-born Mexican-origin women, but comparisons with other races are not made.

differentally distributed across ages by race. We are interested in assessing the extent to which observed racial differences in outcomes or age-outcome relationships exist even after taking into account these risk factors. The results can help us understand the channels through which disadvantage is translated into poor birth outcomes, or at least alert us that these channels are not measured in our data and require further research. Finally, we will attempt to provide some limited insight into short- and long-term effects of various risk factors by examining how an increase or decrease in risk factors between births to the same mother affects her probability of LBW or PTB.

1.2 Background

In statistical analyses, the relationship between age and preterm birth (PTB) or low birthweight (LBW) has been observed to be U-shaped, regardless of race, with higher risk in adolescents, lower risk in early adulthood, and increasing risk at later ages [66, 98]. To the extent that this relationship holds within individual women independent of other conditions, we could say that most women "weather" with age, and we look for racial and ethnic differences in the average timing and extent of this weathering. Several researchers have noted that the elevated risk in adolescence can be largely explained by socioeconomic characteristics [98], whereas increased risk at older ages can be more attributed to biological factors related to aging [98, 4]. However, it is also theorized that still-growing adolescents may compete with the fetus for nutrients and/or that biological immaturity contributes to worse health outcomes for other reasons [46]. In an analysis that examined preterm births from 1975 to 1990, to mothers born between 1926 and 1980, Ananth et al. found that the age-outcome relationship was quite consistent across periods and cohorts, suggesting that the U-shaped relatioship is at least partly due to biological factors, since the

meaning/associated SES characteristics of younger and older pregnancy have changed over time [4].

Accelerated weathering among Blacks, defined here as either a relatively early turning point in the "U" and/or increasing disparities with Whites with age, has been observed in the bulk of the research examining racial differences in age-birth outcome relationships [61, 92, 34, 95, 46, 98, 87, 66]. Ananth et al. (2001) did not find evidence of early Black weathering, noting that while there were large overall Black-White disparities in PTB, relative risks remained constant across ages and both White and Black mothers experienced their lowest risk in the age category 25-29 [4]. Using age categories could have obscured turning points that actually do vary by race, but the finding that relative risks did not increase with age is inconsistent with most of the research.

Since the idea of biological differences contributing to racial health disparities has been largely discredited, understanding disparities in LBW and PTB rests on understanding the determinants of LBW and PTB in general, and these factors are not yet well understood. The most prevalent basic theory, illustrated in Geronimus's weathering hypothesis, is that low socioeconomic status (SES) contributes to poorer health in general correspondingly poorer reproductive outcomes. Several hypothesized mechanisms through which low SES acts to affect maternal health include difficulties accessing health care, adverse employment-related health effects, more obstacles to maintaining a healthy lifestyle and exposure to air pollutants [66]. Stressors including discrimination, domestic and neighborhood violence, financial troubles, housing insecurity and lack of social support are also more plentiful in low-SES neighborhoods [15, 65]. Recent research has identified several pathways that may link psychosocial stress to poor infant health outcomes, including health-depressing psychosocial factors such as anxiety and depression, stress hormones that may affect fetal development or initiate labor, risky coping behaviors such as smoking, alochol

use or poor nutrition, and depressed immune functioning that causes increased susceptibility to infection [87]. In addition, both fetal programming—a mother's own health as a newborn—and cumulative effects of low SES have been linked to negative infant health outcomes, highlighting the long-term nature of SES-related exposures [73].

Some research has simply documented age-outcome relationships by race [4, 21], but most has also tested hypotheses about the causes of observed weathering, operationalized by including control variables in regressions of negative health outcomes against age or examining the age-outcome relationship within different SES groups. These studies have generally focused on a city (e.g., Chicago [95, 34, 46, 73] and New York City [92]) or a state (e.g., Michigan [61] and North Carolina [21]), though some studies have analyzed birth outcomes nationally or in a geographically diverse group of cities [4, 87, 114, 66]. As mentioned previously, with two exceptions, this research excludes mothers who are not White or Black from the analysis. Most of these studies analyze either LBW or PTB, though subclassifications of these categories (e.g., moderately/very low birthweight or moderately/very/extremely preterm birth) are sometimes considered.

Research on the weathering hypothesis and LBW has typically shown that indicators of individual and neighborhood SES explain much of the observed disparities in the age-LBW relationship. In a study of Michigan births, Geronimus (1996) finds that disparities in prenatal care utilization, smoking, hypertension and a variety of other medical risk factors explain a large portion of the the widening disparities in LBW with age. The article further asserts that White women and Black women in high SES neighborhoods do not "weather" at all, since the probability of LBW, adjusted for maternal health characteristics, declines monotonically from age 15 to 34 for women in these groups (though relatively wide confidence intervals indicate that a constant or increasing relationship cannot be ruled out) [61]. Rauh et al. (2001) does

observe weathering with respect to moderately LBW among both Black and White mothers in New York City, but finds that controlling for Medicaid receipt as an indicator of individual poverty eliminates disparities in the age-outcome relationship [92]. Unlike Geronimus (1996), this paper observes weathering in Black mothers at all community poverty levels after individual social and health factors are adjusted for; part but not all of this difference may be due to a higher age cutoff in this study of 39. Rich-Edwards et al. (2003) finds that controlling for education, marital status, prenatal care utilization, smoking and neighborhood poverty accounted for much of the Black-White weathering disparities observed in Chicago [95]. Love et al. (2010), in a study of births in Cook County, IL, finds that the socioeconomic positions of a mother's current *and childhood* neighborhoods both influence infant health outcomes, with Black mothers who lived in poor areas as children and adults showing significantly accelerated weathering compared to mothers who did not, providing support for the cumulative effects of low SES [73].

Compared to research on LBW, evidence of racial disparities in PTB rates that widen with age has been less consistent, or perhaps just less comparable. As noted above, Ananth et al. (2001) shows weathering among all mothers and overall PTB disparities between White and Black mothers but not widening disparities with age, even with no SES controls beyond those for year of birth of the mother and child [4]. Ekwo and Moawad (2000) asks whether Black teens in Chicago are at greater risk for PTB than other age groups and find that they are not. The paper does find elevated risks after age 25, but as the sample is limited to Black mothers, racial disparities could not be assessed [46]. In a study of all White and Black births in U.S. metropolitan areas in 2000, Osypuk and Acevedo-Garcia (2008) finds evidence of accelerated weathering among Black mothers, even after controlling for a variety of SES and health behavior factors. The article also shows that the age-PTB disparities in hypersegregated areas are worse than in non-hypersegregated areas, supporting

the connection between area-level inequality and poor health documented in other research [87]. Holzman et al.(2009), using a rich sample births in several cities and counties across the country, finds no evidence of weathering between ages 20 and 39 among White, non-smoking mothers who lived in relatively high SES neighborhoods. The study did find PTB rate increases with age for Black women in general and for White smokers, but the age-PTB relationship was modified by neighborhood deprivation, with women living in high deprivation neighborhoods exhibiting the most weathering [66].

Geronimus (1996) challenges the assumption implicit in most research on racial health disparities that risk factors operate the same way across populations [61]. For example, a suggested explanation for the finding that low-income White mothers appear not to weather while low-income Black mothers do is that Black women may have lower purchasing power at a given income level, perhaps due to a race-biased geographical distribution of resources such as medical facilities and grocery stores. In a statistical analysis, this assumption can be challenged by estimating race-stratified models or including race interactions with independent variables. In this way, Rich-Edwards et al. (2003) finds that marriage is more strongly protective against LBW for Whites than for Blacks [95]. Unfortunately, such analyses cannot generally explain why effects vary across races, but acknowledging that they may is a useful first step.

Other research has challenged the implicit assumption that risk factors operate the same way across ages by including age-interacted independent variables in regressions. Rauh et al. (2001) includes age interactions with all control variables—smoking, Medicaid receipt, substance use, marital status and education—and finds that only the effect of Medicaid receipt on LBW increases with age, though a statistical significance cutoff of 0.99 may have hidden other probable age interactions [92]. Rich-Edwards et al. (2003) finds that the negative effects of most of the control variables included—education, marital status, prenatal care utilization, smoking and

neighborhood poverty—increase with age, though these age-risk factor relationships do not appear to vary by race [95]. Three non-mutually-exclusive explanations are put forth for these findings: 1) recent exposures to risk factors could have a greater impact among older women, 2) effects of exposure to risks could accumulate over time, causing worse outcomes in older women, and 3) women who are less healthy to begin with drift downwards into more risky environments as they age, a reverse causality argument that seems relatively unlikely.

While Geronimus's weathering hypothesis is the preferred explanation for racial disparities in age-birth outcome relationship in the literature reviewed here, another possibility arises from the observation that the ages at which mothers give birth are not randomly distributed and vary systematically by race. Several authors have noted that differences in childbearing patterns, including fertility-timing norms and interpregnancy intervals, could theoretically explain the apparent early weathering among Black mothers [61, 102, 87]. We note that these alternative theories cannot be separated from the weathering hypothesis on the basis of our analysis; they provide other possible explanations for the same observed relationships. It has also been hypothesized that women who have first births at later ages are distinct from women who become mothers earlier in life, in that they are more likely to have higher SES and less likely to have experienced a lifetime of disadvantage [114]. It is very likely that the "age effect" that remains after controlling for other variables is picking up some of these differences in SES, though including controls for parity and age-parity interaction terms can help take into account these differences [98].

Our paper expands on previous weathering research in a number of ways. As mentioned above, our focus on Texas not only allows us to assess whether previous regional results are generalizable to other parts of the country but also to explore age-birth outcome relationships in U.S.- and foreign-born Hispanic women, since Texas has large populations of both groups. Based on existing evidence, weathering seems

more apparent with respect to LBW than PTB, but since both outcomes have not been included in the same study, it is difficult to discern the extent to which different findings are an artifact of different data and model specifications. To explore this question, we include both LBW and PTB in our analysis. Further, other studies have generally used five-year age categories as the main explanatory variables, a specification that has the advantage of allowing flexibility in the nature of the age-birth outcome relationship. We choose instead to use a quadratic age specification, since evidence indicates that the age-outcome relationship is generally U-shaped and since using age categories could obscure the location of the "turning point"—the age at which birth outcomes begin to worsen—which may provide an important clue as to whether differential weathering occurs. Unlike most previous studies, we also interact all risk factor variables with both race and age, to assess whether racial differences in age-outcome relationships are mediated by racial-differences in the effects of risk factors across ages. Finally, we exploit our ability to track mothers across births in our dataset to look at "within-mother" effects of changes in risk factors, as an initial exploration into the short-term effectiveness of policy changes that could impact these factors.

1.3 Data and Methods

1.3.1 Methods

To assess whether disparities between White and minority mothers observed in other studies are upheld in Texas, we first look at simple mean LBW and PTB outcomes by age for the four race-ethnicity-natality groups in our study: U.S.-born Non-Hispanic White mothers (White), U.S.-born Non-Hispanic Black mothers (Black),

U.S.-born Hispanic mothers and foreign-born Hispanic mothers.[3] Most weathering research has described the relationship between age and birth outcomes using five-year age categories (e.g., 15-19, 20-24, etc.), potentially obscuring features of the relationship (e.g., the location of the turning point, if one exists) that we are able to see more clearly by examining the rates at each age. Here and in the multivariate models, we approximate the relationship between age and probability of negative birth outcome with a quadratic line, a decision informed by theory but ultimately supported by observation of the data, in which the age-outcome relationships turn out to be nearly perfectly "U"-shaped. We believe that specifying the age-PTB relationship in this way is an improvement on past research, in which age generally enters multivariate regression models linearly or in large categories, since it more closely captures the observed relationship between outcomes and age. We use the linear probability model[4] to estimate the bivariate relationship between age and infant health outcome:

$$y_{ijt} = R_j + \beta_{1r}(age_{ijt} * R_j) + \beta_{2r}(age^2_{ijt} * R_j) + Y_t + \varepsilon_{ijt} \tag{1.1}$$

where y_{ijt} takes on the value 1 if child i to mother j in year t is LBW (PTB) and 0 otherwise, and R_j is a categorical variable for the race of mother j. (Conventionally, this race is also assigned to child i). The variable age_{ijt} represents the age of mother j at the birth of child i, and the interaction terms $age_{ijt} * R_j$ and $age^2_{ijt} * R_j$ allow

[3]In addition to omitting the natality and ethnicity descriptors for White and Black mothers, we henceforth refer to these race-ethnicity-natality groups as race groups, though we acknowledge that we trade a great deal of accuracy for this more convenient shorthand.

[4]Previous research on the determinants of LBW and PTB has generally employed logit models to estimate the effects of a change in an independent variable on the probability of these binary outcomes. The two main arguments against using the linear probability model are 1) error terms automatically exhibit heteroskedasticity and 2) predicted values may fall outside the 0-1 interval. Heteroskedasticity can be remedied by calculating heteroskedasticity-robust standard errors, which we do here. Further, since our goal is not to predict individual values but to assess the extent to which risk factor variables mediate the relationship between age and birth outcome by race, we believe the linear probability model is an appropriate, convenient and easy-to-interpret means of carrying out this analysis. See Wooldridge 2002, p. 454-9.

the relationship between age and outcome to vary by race. Since we are using panel data, we also include a vector of year dummies, Y_t, to capture secular changes in the incidence of LBW and PTB over time. We hypothesize that we will observe accelerated weathering in Black mothers only, which would be evidenced in our analysis by a relatively early turning point and steeper increases in LBW or PTB rates with age.

If we observe accelerated weathering among any of the race groups in the bivariate analysis, we can explore the extent to which they can be explained by disparities in the distribution of risk factors, as suggested by Geronimus [61]. To assess how variation in risk factors affect the age-outcome relationships by race, we estimate the following model, similar to Equation (1.1) but with controls for various risk factors:

$$y_{ijt} = R_j + \beta_{1r}(age_{ijt} * R_j) + \beta_{2r}(age^2_{ijt} * R_j) + \theta risk_{ijt} + Y_t + \varepsilon_{ijt} \qquad (1.2)$$

where $risk_{ijt}$ is a vector of variables representing behavioral, socioeconomic and medical risk factors that are hypothesized to affect the probability of negative birth outcomes and are differentially distributed by age across races. We hypothesize that controlling for these factors will reduce weathering disparities (i.e., we expect to see more similar turning points and narrower differentials with age) as well as overall disparities (i.e., more similar constant terms).

Disparities that remain after controlling for the average effects of risk factors could exist partly because the effect of the risk factors vary by race and/or age, as other authors have suggested. To allow for the connection between risk factors and birth outcomes to vary systematically by race and by age, we also estimate models that allow for race-risk factor interactions (1.3) and age-race-risk factor interactions (1.4):

$$y_{ijt} = R_j + \beta_{1r}(age_{ijt} * R_j) + \beta_{2r}(age^2_{ijt} * R_j) + \theta_r(risk_{ijt} * R_j) + Y_t + \varepsilon_{ijt} \qquad (1.3)$$

$$y_{ijt} = R_j + \beta_{1r}(age_{ijt} * R_j) + \beta_{2r}(age^2_{ijt} * R_j) \\ +\omega_{r1}(risk_{ijt} * R_j * age_{ijt}) + \omega_{r2}(risk_{ijt} * R_j * age^2_{ijt}) + Y_t + \varepsilon_{ijt} \tag{1.4}$$

Model (1.3) implicitly assumes that the effect of any given risk factor is constant across ages but allows these age-averaged effects to vary by race. Model (1.4) allows risk factor effects to vary by age and allows the age-risk factor relationship to also vary by race. By including these race and age interactions, we ask how much of the age-outcome relationship is left unexplained when all women experience the same risks in the same way across ages as an average White women. We hypothesize that any discrepancies will be further narrowed in these estimations.

Because some of the risk factors in our analysis (described below) are both damaging to reproductive health in their own right and are correlated with low SES, which can harm health through various pathways described earlier, we are interested in understanding the extent to which altering a mother's risk profile could lead to relatively short-term changes in health. If these risk factors are mostly proxies for SES, which Love et al. (2010) shows has very long-term effects, the answer could be "not much." In order to gain some initial insight into this question, we estimate a model that utilizes only within-mother variation:

$$y_{ijt} = \beta_{1r}(age_{ijt} * R_j) + \beta_{2r}(age^2_{ijt} * R_j) + \theta_r(risk_{ijt} * R_j) + Y_t + \zeta_j + \varepsilon_{ijt} \tag{1.5}$$

In this model, the variable ζ_j is a mother fixed effect; it absorbs mother-specific characteristics influencing LBW and PTB that change slowly or not at all over time. Thus, if we observe that a within-mother, between-birth change in a risk factor is systematically tied to change in LBW or PTB, we can infer that there would be some relatively near-term effects of altering this risk factor. If we find small or nonexistent within-mother effects, it would follow that the OLS models are picking up mostly between-mother variation and that altering the risk factors (or the correlates they may

represent) would have little short-term impact. Here, the sample must be restricted to mothers who have had more than one birth in the sample period.

1.3.2 Data

The data for this study, all Texas birth records for the years 1994-2003 (N=3,532,103), were obtained from the Texas Department of State Health Services Center for Health Statistics. From this universe, we kept singleton births with probable gestational ages and birthweights[5] and non-missing values for birthweight, gestation and age. We further required that the mother's age be between 15-40 years and that the mother's race-ethnicity-natality status be either U.S.-born Non-Hispanic White (White), U.S.-born Non-Hispanic Black (Black), U.S.-born Hispanic or foreign-born Hispanic. Mothers under 15 and over 40, as well as foreign-born White and foreign-born Black mothers, were omitted due to relatively small numbers of such observations. The resulting dataset contains 3,159,044 observations and was used for the majority of our analyses. We also conduct one set of analyses using only records that could be matched to 2000 U.S. Census tract-level data, which reduces the number of usable observations to 2,727,139, and another set of analyses using only records with at least one sibling in the sample, which leaves 1,055,117 observations.

Our dependent variables, low birthweight (LBW) and preterm birth (PTB), are defined in the usual ways. The LBW variable was coded as 1 if birthweight was less than 2500 grams and 0 otherwise. The PTB variable was coded as 1 if gestational age was less than 37 weeks and 0 otherwise. Gestational age was calculated based on date of last menstrual period (LMP) and the clinical estimate of gestation was used in the small share of cases in which LMP was not reported.

---

[5]Following previous research, we excluded births with recorded birthweights of less than 500 grams or recorded gestational ages of less than 22 weeks, since these are often the product of coding errors [46, 4, 114].

As mentioned above, LBW and PTB are not mutually exclusive categories. Infants may be low birthweight because they were born preterm or because they experienced intrauterine growth restriction. Some researchers have used the category small for gestational age (SGA) to identify those births that are low birthweight due to growth restriction rather than early delivery. We did not employ this category here for two main reasons: 1) SGA traditionally adjusts for race, which does not make sense given the nature of our analysis and 2) other weathering research has used PTB rather than SGA. Of the 281,978 PTB cases and 184,394 LBW cases in our sample, 104,923 are both PTB and LBW. Although there is much overlap, differences between the LBW and PTB regressions may help us understand which factors are related to early delivery as opposed to growth restriction.

Our main right-hand-side variable, age, is not a typical explanatory variable in that we hypothesize that other variables can explain variation in LBW and PTB rates sometimes attributed to variation in age. Age is treated as a continuous variable in our regressions and is centered at 15, to allow for more straightforward interpretation of the constant terms. The other right-hand-side variables, which we call "risk factors", were chosen to represent—more or less directly depending on the variable—biological, behavioral and/or socioeconomic risk factors thought to influence infant health outcomes. The twelve risk factors, which enter as dummies, are 1) *low maternal education*, 2) *mother unmarried*, 3) *mother smoking*, 4) *first birth*, 5) *third or higher birth*, 6) *late prenatal care initiation*, 7) *few prenatal care visits*, 8) *excessive prenatal care visits*, 9) *diabetes*, 10) *chronic hypertension*, 11) *pregnancy-related hypertension*, and 12) *mother not living inside city limits*. These variables, described in more detail below, are set to 0 if the risk factor is not present, 1 if it is and 9 if the information is missing. In some specifications, we also include either *neighborhood median household income* or a *neighborhood deprivation index*, which entered as continuous variables. Our choice of risk factors was constrained by the data available on Texas birth records

and, since we were specifically looking for explanations of disparities in age-outcome relationships, a further consideration was the amount of between-race variation in the distribution of these risk factors across ages.

Several risk factor variables listed above require more explanation. *Low maternal education* is primarily intended to be a marker of socioeconomic status (SES), though it is possible that formal education, through nutrition or health classes, for example, might exert some influence birth outcomes. A maternal education variable has been used in some multivariate weathering analyses [92, 66] and not in others [61]. Geronimus (1996) specifically argues against including it, on the basis that it is not reliably recorded and is correlated with age and with age at first birth [61]. We include it despite these potential problems because, lacking information on a mother's income or poverty status, it is perhaps the best indicator on the birth record of SES. Further, we are less concerned with identifying the precise effects of a high school education on birth outcomes than in observing how the inclusion of this variable changes the age-outcome relationship. Finally, in some specifications, we also include a neighborhood SES variable, and we are interested in understanding the extent to which these variables pick up the same underlying characteristics. In our study, the low education variable takes on value 1 if a mother is at least 19 and has not completed high school or less than 19 and has not completed elementary school, and 0 otherwise.

Late prenatal care (PNC) initiation, *few PNC visits* and *many (excessive) PNC visits* are all defined as in Kotelchuck's Adequacy of Prenatal Care Utilization Index [69]. Here, Kotelchuck proposes a single index that accounts for both timing of PNC initiation and whether the number of visits is as expected given the month of initation. We separate these factors in our analysis, on the basis that late PNC initation and a lower-than-expected number of visits both signal lack of access to care, while a higher-than-expected number of visits signals pregnancy complications. *Late PNC initiation* takes on the value 1 if PNC began in month 5 or later, corresponding to the

"inadequate" and "intermediate" categories in Kotelchuck's Adequacy of Initiation of Prenatal Care Index, and 0 otherwise. The variable *few PNC visits* is coded as 1 if a mother received less than 80 percent of the expected visits for a mother who began PNC when she did, corresponding to the "inadequate" and "intermediate" categories in Kotelchuck's Adequacy of Received Prenatal Care Services Index. Many PNC visits takes on the value 1 if a mother received 110 percent or more of the expected number of visits, corresponding to the "adequate plus" category in Kotelchuck's Adequacy of Received Prenatal Care Services Index.

The remaining risk factor variables are defined in the expected ways. Unmarried status is a proxy for lack of social and financial support and may also capture more diffuse elements of SES. Smoking *per se* is known to be associated with growth restriction, but it, too, may be related to unmeasured lifestyle and socioeconomic characteristics. For example, smoking may be a response to stressful conditions that have independent effects on birth outcomes. First births and higher parity births are both associated with less favorable health outcomes than second births. When interacted with age, parity also becomes a signal of SES, since women with relatively high SES are more likely to delay childbearing. Not living within city limits is also included, as it may signal relatively low access to healthcare and other resources.

In some specifications, we also include an indicator of neighborhood SES as a risk factor variable. In one specification, we use 1999 median household income at the U.S. Census tract level, coded as a continuous variable and centered at the sample median value of $19,626. Our alternative measure for neighborhood SES is the deprivation index, defined as in Messer et al. (2006) [79]. This multi-dimensional index is intended to gauge relative deprivation more completely than a single measure, such as income. Tract-level variables included in the deprivation index are percent of males in management, percent of households with > 1 person per room, percent of individuals with 1999 income below federal poverty level, percent of families with

female headed household with dependent children, percent of households with income less than $30,000, percent of households with public assistance income, percent unemployed, and percent of adults with no high school education. To create the index, principal components analysis is used to infer the contribution of each of these variables to an underlying deprivation variable, with the factor loadings used to weight each variable's contribution to the tract-level neighborhood deprivation score. The deprivation index is then standardized to have a mean of 0 and standard deviation of 1. The deprivation index would seem to be a more complete indicator of neighborhood SES than median household income, but we find that the birth-weighted correlation between tract median household income and the neighborhood deprivation index is -0.85, so we expect similar results using either measure and do not include both in the same model.

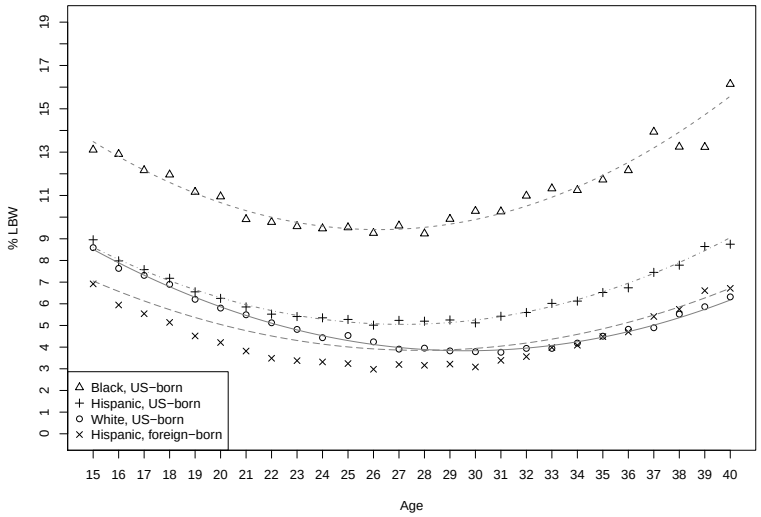
1.4 Results

1.4.1 Bivariate Relationship

Figure 1.1 shows the rate of low birthweight (LBW) births by maternal race and age. These rates are adjusted for year of birth but not for any other confounding factors. The lines in this figure represent a quadratic line fit to the data. The most striking result is the large disparity between White and Black rates of LBW across ages. Additionally, we see evidence of accelerated Black weathering; the Black-White disparity is smallest at age 15 and widens with age. Though the overall disparities between White and Hispanic mothers are comparatively small, we also see relatively early weathering in both Hispanic groups. The disparity with Whites is smallest at age 15 (negative, in the case of foriegn-born Hispanic mothers) and increases with age.

Figure 1.2 illustrates that racial differences in the relationship between preterm birth (PTB) and age follow a similar pattern, with minority mothers experiencing

Figure 1.1. Low birthweight rates by maternal race and age, Texas, 1994-2003

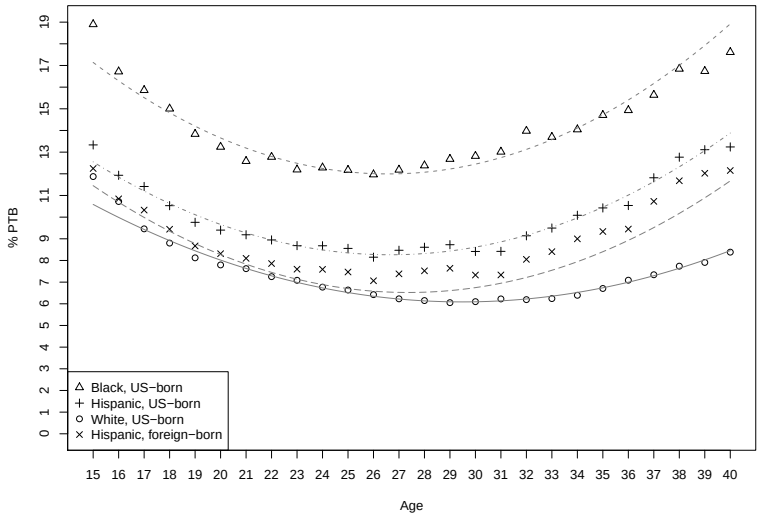


Source: Author's calculations based on data from the Texas Center for Health Statistics.
Notes: Rates are unadjusted except for birth year; plotted values represent 1994 rates.

an increase in PTB at earlier ages than White mothers. In the LBW relationship shown in Figure 1.1, we see that U.S.-born Hispanic mothers had similar rates of LBW to Whites at young ages and foreign-born Hispanic mothers had lower rates. With respect to PTB, we see that both U.S.- and foreign-born Hispanic mothers have slightly higher initial rates of PTB relative to White mothers, and the disparity increases with age.

The first column group in Table 1.3 lists the coefficients for the linear and squared terms of the quadratic relationship between age and LBW, adjusted only for year of birth, along with the implied turning point, which shows the approximate age at which outcomes begin to worsen for each race. As illustrated in Figure 1.1, this age is highest for Whites (29.6) and lowest for Blacks (26.2). Table 1.4 shows that the respective "turning point ages" are very similar with respect to PTB. We hypothesized that only Black women would exhibit accelerated weathering, but we find that all three minority groups do, though overall disparities are much lower for Hispanic women.

Figure 1.2. Preterm birth rates by maternal race and age, Texas, 1994-2003



Source: Author's calculations based on data from the Texas Center for Health Statistics.
Notes: Rates are unadjusted except for birth year; plotted values represent 1994 rates.

1.4.2 Sample risk factor characteristics

The main explanation for accelerated aging in the weathering hypothesis is the existence of disparities in exposure to risk factors that widen with age. Table 1.1 shows rates of those variables we call risk factors by race and age category, and Table 1.2 shows the ratio of these rates to the corresponding White rates. Looking first at the Black-White ratios, we see clear evidence of disparities widening across age with respect to several risk factors, including low education, unmarried status, late prenatal care initation, diabetes and both chronic and pregnancy-related hypertension. Black-White differences in smoking behavior are similar to those observed in other parts of the country; younger Black mothers are much less likely to smoke than White mothers of similar ages, but smoking rates increase with age for Black mothers and decrease with age for Whites.

With regard to U.S.-born Hispanic mothers, we also see relative "exposure" to risk factors increasing with age, though the magnitude of the disparities for several

Table 1.1. LBW, PTB and risk factor rates by maternal race and age group, Texas, 1994-2003.

	White, US-born (N=1270843)					Black, US-born (N=350761)				
Age Category	15-19	20-24	25-29	30-34	35-40	15-19	20-24	25-29	30-34	35-40
Share of births	11.2	25.2	28.5	23.8	11.3	22.4	35.4	22.1	13.6	6.4
Low birthweight (LBW)	7.0	5.3	4.3	4.1	5.2	12.1	10.2	9.7	10.9	13.1
Preterm birth (PTB)	9.4	7.8	6.8	6.7	7.8	15.7	13.1	12.7	13.9	16.1
Low education	57.5	21.2	7.6	3.8	3.2	61.1	21.7	11.2	7.7	7.5
Unmarried	63.4	31.0	11.4	6.6	7.3	93.0	74.4	49.6	36.0	33.8
Mother smoking	22.4	19.5	11.1	8.4	9.4	3.9	6.1	6.7	8.2	11.3
First birth	78.9	47.9	40.3	29.2	23.3	71.2	36.9	26.3	23.0	20.2
Third or higher birth	2.3	15.2	21.8	29.6	40.0	6.0	27.6	38.3	41.3	46.6
Late PNC initation	18.3	11.8	5.9	4.6	5.2	26.3	18.7	13.1	11.3	12.2
Few PNC visits	23.3	19.8	16.4	15.6	15.6	33.8	30.0	25.6	24.0	24.2
Many PNC visits	35.0	35.7	35.7	35.5	37.0	30.3	31.8	34.4	36.0	37.5
Diabetes	0.9	1.6	2.4	2.9	3.9	0.8	1.5	2.9	4.2	5.7
Chronic hypertension	0.3	0.5	0.8	0.9	1.4	0.4	0.7	1.6	2.7	4.4
Preg-related hypertension	5.4	4.7	4.5	3.9	4.2	6.1	4.6	4.9	5.4	6.2
Not in city limits	22.5	22.2	21.9	21.0	20.8	6.0	6.2	7.4	8.9	9.7

	Hispanic, US-born (N=764888)					Hispanic, foreign-born (N=772552)				
Age Category	15-19	20-24	25-29	30-34	35-40	15-19	20-24	25-29	30-34	35-40
Share of births	25.1	33.9	22.8	12.9	5.4	12.4	30.1	29.9	19.2	8.4
Low birthweight (LBW)	7.5	5.9	5.4	5.8	7.5	6.2	4.7	4.1	4.5	5.6
Preterm birth (PTB)	11.4	9.5	9.0	9.5	12.0	10.0	8.0	7.1	7.8	10.0
Low education	71.4	37.0	24.6	18.3	18.9	83.7	67.5	62.0	59.8	63.7
Unmarried	67.4	42.4	24.9	17.9	18.4	51.2	33.1	22.3	18.0	19.1
Mother smoking	3.2	3.7	3.2	3.1	4.1	0.7	0.7	0.8	1.1	1.4
First birth	69.5	35.2	23.6	18.6	14.8	75.2	45.3	25.3	16.0	12.0
Third or higher birth	5.0	25.8	39.4	46.5	54.3	3.5	17.9	37.9	54.5	64.6
Late PNC initation	23.5	17.6	11.4	9.1	9.9	33.3	26.7	22.4	20.7	21.7
Few PNC visits	29.0	26.2	22.6	20.8	20.0	45.3	41.9	38.8	36.3	35.1
Many PNC visits	32.1	33.6	35.6	37.0	40.1	22.0	22.4	23.8	25.9	28.4
Diabetes	0.9	1.9	3.4	5.3	7.9	0.7	1.5	2.9	5.0	7.7
Chronic hypertension	0.2	0.3	0.5	0.9	1.6	0.1	0.2	0.3	0.4	0.9
Preg-related hypertension	4.3	3.5	3.5	3.7	4.8	3.6	2.9	2.6	2.8	3.7
Not in city limits	9.7	10.0	10.7	11.5	11.9	8.5	8.3	9.0	9.7	9.9

Source: Author's calculations based on data from the Texas Center for Health Statistics.
Variable definitions: Low birthweight=birthweight less than 2500 grams. Preterm birth=gestation less than 37 weeks. Low education=less than high school for women ages 19 and up; less than sixth grade for women under 19. Late PNC initation=Inadequate or Intermediate ranking on Kotelchuck's prenatal care initiation index. Few PNC visits=Inadequate or Intermediate ranking on Kotelchuck's expected visits index. Many PNC visits=Adequate Plus ranking on Kotelchuck's expected visits index.

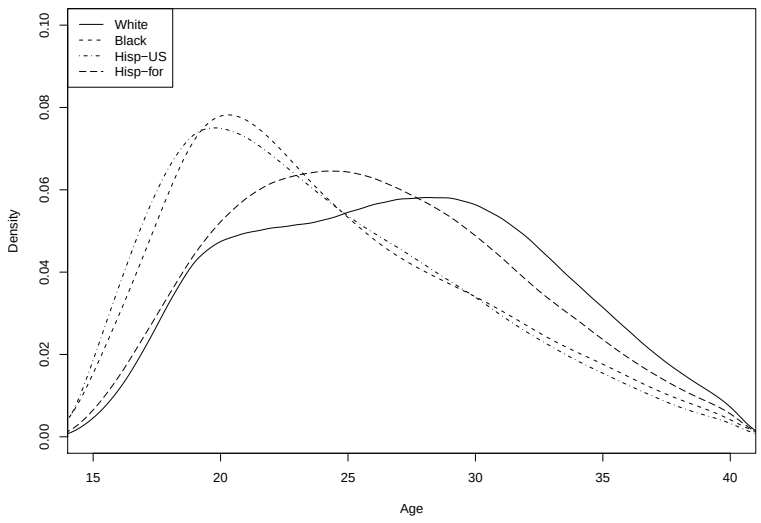
Table 1.2. LBW, PTB and risk factor rates–Minority to White ratios

	White, US-born (N=1270843)					Black, US-born (N=350761)				
Age Category	15-19	20-24	25-29	30-34	35-40	15-19	20-24	25-29	30-34	35-40
Share of births	1.0	1.0	1.0	1.0	1.0	2.0	1.4	0.8	0.6	0.6
Low birthweight (LBW)	1.0	1.0	1.0	1.0	1.0	1.7	1.9	2.3	2.7	2.5
Preterm birth (PTB)	1.0	1.0	1.0	1.0	1.0	1.7	1.7	1.9	2.1	2.1
Low education	1.0	1.0	1.0	1.0	1.0	1.1	1.0	1.5	2.0	2.3
Unmarried	1.0	1.0	1.0	1.0	1.0	1.5	2.4	4.3	5.5	4.6
Mother smoking	1.0	1.0	1.0	1.0	1.0	0.2	0.3	0.6	1.0	1.2
First birth	1.0	1.0	1.0	1.0	1.0	0.9	0.8	0.7	0.8	0.9
Third or higher birth	1.0	1.0	1.0	1.0	1.0	2.6	1.8	1.8	1.4	1.2
Late PNC initation	1.0	1.0	1.0	1.0	1.0	1.4	1.6	2.2	2.5	2.4
Few PNC visits	1.0	1.0	1.0	1.0	1.0	1.5	1.5	1.6	1.5	1.5
Many PNC visits	1.0	1.0	1.0	1.0	1.0	0.9	0.9	1.0	1.0	1.0
Diabetes	1.0	1.0	1.0	1.0	1.0	0.8	0.9	1.2	1.5	1.5
Chronic hypertension	1.0	1.0	1.0	1.0	1.0	1.4	1.4	2.1	2.9	3.0
Preg-related hypertension	1.0	1.0	1.0	1.0	1.0	1.1	1.0	1.1	1.4	1.5
Not in city limits	1.0	1.0	1.0	1.0	1.0	0.3	0.3	0.3	0.4	0.5

	Hispanic, US-born (N=764888)					Hispanic, foreign-born (N=772552)				
Age Category	15-19	20-24	25-29	30-34	35-40	15-19	20-24	25-29	30-34	35-40
Share of births	2.2	1.3	0.8	0.5	0.5	1.1	1.2	1.0	0.8	0.7
Low birthweight (LBW)	1.1	1.1	1.3	1.4	1.4	0.9	0.9	1.0	1.1	1.1
Preterm birth (PTB)	1.2	1.2	1.3	1.4	1.5	1.1	1.0	1.1	1.2	1.3
Low education	1.2	1.7	3.2	4.8	5.8	1.5	3.2	8.2	15.6	19.6
Unmarried	1.1	1.4	2.2	2.7	2.5	0.8	1.1	1.9	2.7	2.6
Mother smoking	0.1	0.2	0.3	0.4	0.4	0.0	0.0	0.1	0.1	0.2
First birth	0.9	0.7	0.6	0.6	0.6	1.0	0.9	0.6	0.5	0.5
Third or higher birth	2.2	1.7	1.8	1.6	1.4	1.5	1.2	1.7	1.8	1.6
Late PNC initation	1.3	1.5	1.9	2.0	1.9	1.3	1.4	1.7	1.8	1.8
Few PNC visits	1.2	1.3	1.4	1.3	1.3	1.9	2.1	2.4	2.3	2.2
Many PNC visits	0.9	0.9	1.0	1.0	1.1	0.6	0.6	0.7	0.7	0.8
Diabetes	1.0	1.2	1.5	1.8	2.0	0.8	0.9	1.2	1.7	2.0
Chronic hypertension	0.7	0.6	0.7	0.9	1.1	0.5	0.3	0.3	0.5	0.6
Preg-related hypertension	0.8	0.7	0.8	0.9	1.2	0.7	0.6	0.6	0.7	0.9
Not in city limits	0.4	0.5	0.5	0.5	0.6	0.4	0.4	0.4	0.5	0.5

Source: Author's calculations based on data from the Texas Center for Health Statistics.
Variable definitions: Low birthweight=birthweight less than 2500 grams. Preterm birth=gestation less than 37 weeks. Low education=less than high school for women ages 19 and up; less than sixth grade for women under 19. Late PNC initation=Inadequate or Intermediate ranking on Kotelchuck's prenatal care initiation index. Few PNC visits=Inadequate or Intermediate ranking on Kotelchuck's expected visits index. Many PNC visits=Adequate Plus ranking on Kotelchuck's expected visits index.

Figure 1.3. Distribution of births by maternal age and race, Texas, 1994-2003

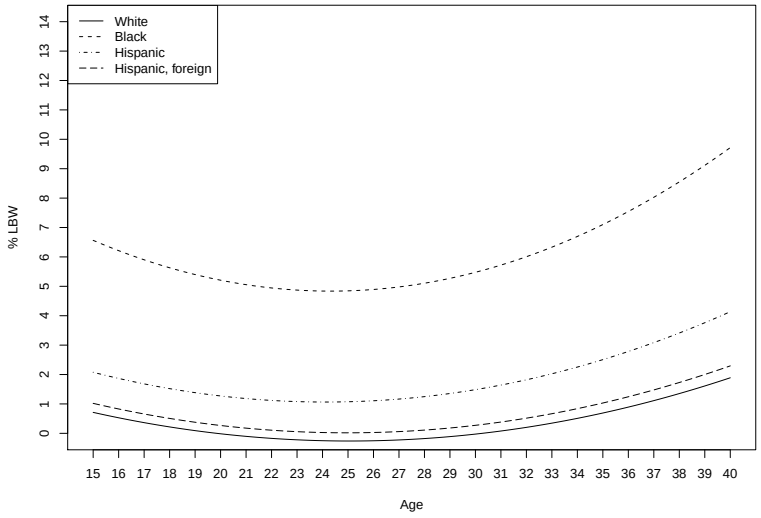


Source: Author's calculations based on data from the Texas Center for Health Statistics.
Notes: The density of births was calculated and plotted in R, using a Gaussian smoothing kernel estimated over 512 points of each age distribution with a smoothing bandwidth equal to the standard deviation of the smoothing kernel.

variables—including late prenatal care initation, diabetes, hypertension, and smoking status—is not as large as in the Black-White comparison. A similar pattern is seen in foreign-born Hispanic mothers, with even lower relative rates of smoking and medical risk factors. Both Hispanic groups are less likely to have a high school degree than Whites or Blacks. Surprisingly, perhaps, given possible language barriers, foreign-born Hispanic mothers are less likely to have late prenatal care initation than Black and U.S.-born Hispanic mothers.

Figure 1.3 shows the distribution of births across ages for each group of mothers. The distribution for Black and U.S.-born Hispanic mothers is similar, with a relatively large share of births occuring during the teens and early twenties. The age-birth distribution for foreign-born Hispanic mothers is remarkably similar to the White mother distribution, though Table 1.2 shows White mothers are much more likely to be having their first child at later ages, while foreign-born Hispanic mothers are more likely to have had one or more children already.

Figure 1.4. Age-LBW relationship, adjusted for average risk factor effects (M2)



Source: Author's calculations based on data from the Texas Center for Health Statistics.
Notes: The curves shown here are those implied by the race-specific constant terms and coefficients on age and age^2 in Table 1.3, Model 2.

1.4.3 Main effects models

If disparities in exposure to risk factors by race contributes to accelerated aging among minority women, then accounting for these risk factors should reduce observed weathering in the residual age-outcome relationship. The second column group in Table 1.3 shows the results of controlling for our twelve risk factor variables in the LBW model. In all regressions, the base value of each risk factor was set to 0, so the constant estimated in Model 2 (Equation (1.2)) reflects the unexplained LBW rate for a mother who experiences the mean effect of having a high school degree, is married, does not smoke, etc. We chose the protective counterparts to the risk factors as the base levels in our regressions to correspond to the experience of an average White woman in our sample.

Controlling for these risk factors (and the unobserved risks with which they may be correlated) does explain a large amount of previously unexplained variation in LBW, as can be seen in the much lower constant term in Model 2 compared to Model

Table 1.3. Low birthweight regressions: M1) Bivariate, M2) Multivariate, M3) Multivariate with race interactions

	Model 1: Bivariate				Model 2: Main effects				Model 3: Race interactions			
	White	B diff	H diff	HF diff	White	B diff	H diff	HF diff	White	B diff	H diff	HF diff
Constant	8.51^{**}	4.98^{**}	0.10	-1.46^{**}	0.71^{**}	5.85^{**}	1.36^{**}	0.31^{*}	0.51^{**}	2.27^{**}	1.43^{**}	2.12^{**}
Age	-0.639^{**}	-0.086^{*}	0.052^{*}	0.141^{**}	-0.194^{**}	-0.177^{**}	-0.028	-0.007	-0.128^{**}	$-0.076^{\#}$	-0.097^{**}	-0.132^{**}
Age^2	0.0218^{**}	0.0105^{**}	0.0024^{*}	-0.0024^{**}	0.0096^{**}	0.0102^{**}	0.0026^{*}	0.0004	0.0076^{**}	0.0067^{**}	0.0045^{**}	0.0043^{**}
Implied turning point	29.6	26.2	27.1	27.8	25.1	24.3	24.1	25.0	23.4	22.1	24.3	25.9
Low education						1.00^{**}			1.64^{**}	0.18	-0.68^{**}	-1.30^{**}
Unmarried						1.14^{**}			1.33^{**}	0.34^{*}	-0.37^{**}	-0.50^{**}
Mother smoking						4.71^{**}			4.15^{**}	3.13^{**}	0.47^{*}	-0.93^{**}
First birth						1.71^{**}			1.75^{**}	0.41^{**}	-0.06	-0.18^{*}
Third or higher birth						0.10^{**}			0.21^{**}	0.11	$-0.14^{\#}$	-0.39^{**}
Late PNC initation						0.33^{**}			0.53^{**}	-0.11	0.16	-0.64^{**}
Few PNC visits						1.94^{**}			2.09^{**}	1.81^{**}	0.00	-1.03^{**}
Many PNC visits						5.88^{**}			5.20^{**}	4.54^{**}	0.95^{**}	-0.09
Diabetes						-1.37^{**}			-1.07^{**}	-1.70^{**}	-0.66^{**}	0.09
Chronic hypertension						10.51^{**}			8.22^{**}	3.28^{**}	5.05^{**}	4.98^{**}
Preg hypertension						9.04^{**}			6.73^{**}	4.47^{**}	4.44^{**}	3.67^{**}
Not in city						-0.40^{**}			-0.36^{**}	-0.80^{**}	-0.04	0.27^{**}
Year controls		Y				Y				Y		
N		3159044				3159044				3159044		
Adjusted R^2		0.0079				0.0324				0.0339		

Source: Author's calculations based on data from the Texas Center for Health Statistics.
Notes: Results from OLS regressions with dependent variable low birthweight. Sample includes singleton Texas births to White, B(lack), H(ispanic, U.S. born) and H(ispanic), F(oreign-born) mothers in years 1994-2003. Standard errors (not shown) are adjusted for heterodkedasticity. ** indicates statistically significant at < 0.01; * at < 0.05; $^{\#}$ at < 0.10. Significance stars in White columns represent probability that estimated coefficient is different from zero; coefficients in B, H and HF columns represent difference from White coefficient and stars represent probability that the difference is greater than zero. All coefficients multiplied by 100 for legibility.

1. As hypothesized, these controls also reduce observed weathering, which can be seen both in the more similar "turning point" ages across races, as well as in Figure 1.4, where the gap between the minority and White curves does not widen with age as drastically as in the bivariate version (Figure 1.1). We also observe earlier turning points across races when controlling for risk factors. One possible explanation is that the negative biological effects of age on birthweight are counterbalanced by increasing rates of protective conditions and behaviors with age.

Including these controls does not, as we hypothesized, account for overall racial disparities. Referring again to Model 2 in Table 1.3, the unexplained disparity between Whites and the three minority groups, as represented by the race-specific constant terms, actually *increases* in the main effects regression. This finding is striking; while the risk factors we are able to control for are limited, the fact that disparities remain even after controlling for some proxies for SES indicate that belonging to a minority race in Texas may have implications for health beyond simple correlations with SES.

Because the risk factors are correlated both with each other and unobserved markers of SES, we should not place too much stock in the exact coefficient estimates of these variables *per se*. Broadly, though, we observe in Table 1.3 that all coefficient estimates are of the expected sign except for the indicator of whether the mother lives within city limits. (Diabetes has a well-known association with higher birthweights and was included in both LBW and PTB models for symmetry.) We hypothesized that living in a city would be protective, due to better access to resources and health care, but it seems that negative effects associated with urban living outweigh possible protective properties.

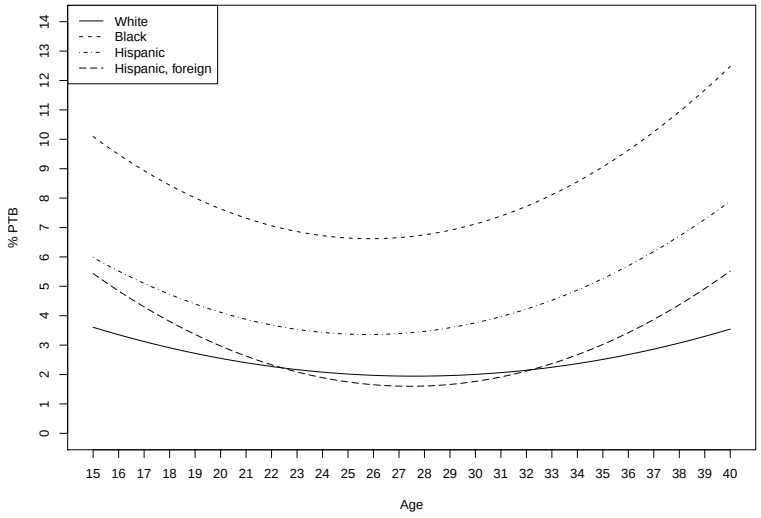
Turning now to the analagous model for PTB shown in Table 1.4, we see that controlling for risk factors again explains a high proportion of overall variation in preterm birth, as evidenced by a much lower constant term. As in the LBW model, we see that disparities by race are either maintained or increased with the addition

Table 1.4. Preterm birth regressions: M1) Bivariate, M2) Multivariate, M3) Multivariate with race interactions

	Model 1: Bivariate				Model 2: Main effects				Model 3: Race interactions			
	White	B diff	H diff	HF diff	White	B diff	H diff	HF diff	White	B diff	H diff	HF diff
Constant	10.59^{**}	6.55^{**}	1.97^{**}	0.87^{**}	3.61^{**}	6.49^{**}	2.37^{**}	1.83^{**}	3.43^{**}	3.83^{**}	2.57^{**}	3.58^{**}
Age	-0.621^{**}	-0.268^{**}	-0.115^{**}	-0.177^{**}	-0.263^{**}	-0.375^{**}	-0.222^{**}	-0.353^{**}	-0.232^{**}	-0.367^{**}	-0.295^{**}	-0.426^{**}
Age^2	0.0214^{**}	0.0170^{**}	0.0101^{**}	0.0108^{**}	0.0104^{**}	0.0189^{**}	0.0120^{**}	0.0144^{**}	0.0098^{**}	0.0183^{**}	0.0137^{**}	0.0166^{**}
Implied turning point	29.5	26.6	26.7	27.4	27.6	25.9	25.8	27.4	26.8	25.7	26.2	27.5
Low education						1.20^{**}			1.33^{**}	0.28	$-0.25^{\star}$	-0.31^{**}
Unmarried						1.54^{**}			1.40^{**}	0.71^{**}	0.07	0.07
Mother smoking						1.68^{**}			1.35^{**}	1.72^{**}	0.62^{**}	0.61
First birth						$0.09^{\star}$			0.57^{**}	-0.96^{**}	-0.99^{**}	-0.60^{**}
Third or higher birth						0.87^{**}			0.62^{**}	1.08^{**}	0.53^{**}	-0.17
Late PNC initation						1.49^{**}			1.70^{**}	0.93^{**}	1.08^{**}	-1.69^{**}
Few PNC visits						2.90^{**}			3.06^{**}	2.07^{**}	0.08	-1.32^{**}
Many PNC visits						7.64^{**}			7.14^{**}	3.95^{**}	0.75^{**}	-0.47^{**}
Diabetes						1.06^{**}			1.52^{**}	-1.59^{**}	-0.06	-1.08^{**}
Chronic hypertension						9.82^{**}			8.06^{**}	2.29^{**}	4.48^{**}	3.32^{**}
Preg hypertension						7.01^{**}			5.87^{**}	1.84^{**}	2.78^{**}	1.28^{**}
Not in city						-0.41^{**}			-0.37^{**}	-0.32	0.05	0.15
Year controls		Y				Y				Y		
N		3159044				3159044				3159044		
Adjusted R^2		0.0069				0.0264				0.0274		

Source: Author's calculations based on data from the Texas Center for Health Statistics.
Notes: Results from OLS regressions with dependent variable preterm birth. Sample includes singleton Texas births to White, B(lack), H(ispanic, U.S. born) and H(ispanic), F(oreign-born) mothers in years 1994-2003. Standard errors (not shown) are adjusted for heterodkedasticity. ** indicates statistically significant at < 0.01; * at < 0.05; $^{\#}$ at < 0.10. Significance stars in White columns represent probability that estimated coefficient is different from zero; coefficients in B, H and HF columns represent difference from White coefficient and stars represent probability that the difference is greater than zero. All coefficients multiplied by 100 for legibility.

Figure 1.5. Age-PTB relationship, adjusted for average risk factor effects (M2)



Source: Author's calculations based on data from the Texas Center for Health Statistics.
Notes: The curves shown here are those implied by the race-specific constant terms and coefficients on age and age^2 in Table 1.4, Model 2.

of controls. It appears, based on the impllied turning points and the age-PTB relationship shown in Figure 1.5, that controlling for these risk factors does not do as much to homogenize the age-PTB relationship as in the LBW model. Blacks and U.S.-born Hispanic mothers have turning points at relatively lower ages compared to White mothers in this model, signalling early weathering, and the size of the residual age disparity widens more among older women. At the same time, we see disparities decreasing for several years at younger ages before increasing again, which is a departure from the unadjusted age-PTB model, where racial disparities mostly increased monotonically with age.

We expect the coefficients to be slightly larger in the PTB model compared to the LBW model, since overall PTB rates are higher, but even taking this into account, there are some differences in risk factor effects. Smoking clearly has a greater impact on birthweight than gestational age; this observation helps explain why adding risk factor controls explains more of the variation in LBW than in PTB. It may also help

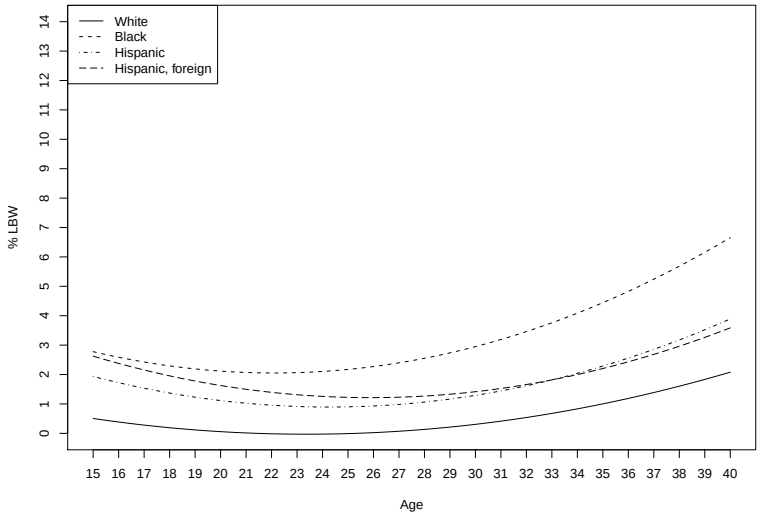
explain why minority weathering is less apparent after the addition of controls in the LBW model, since minority smoking rates increase relative to White rates with age. The results indicate that first births are relatively more likely to be LBW while third or higher births are more likely to be preterm. Diabetes, as expected, increases the risk of PTB but not LBW.

1.4.4 Race interaction models

The coefficients on risk factor variables in Model 2 represented population-averaged effects; in Model 3 (corresponding to Equation (1.3)), we allow the effects of all independent variables to vary by race. As shown in the third column group in Table 1.3, we find significant differences by race in the effects of these variables. In general, the risk factors included here are more risky for Blacks and U.S.-born Hispanics, relative to Whites, and less risky for foreign-born Hispanics. (We could also highlight the converse: the corresponding protective factors are more protective for Blacks and U.S.-born Hispanics and less protective for foreign-born Hispanics.) Having a high school degree, for example, is less protective for both Hispanic groups than it is for Whites or Blacks. Smoking and hypertension are more damaging for all minority mothers than for Whites.

There are several possible explanations for the differences in effects. An upward bias in the effect of hypertension on minority mothers, for example, could occur if there are racial differences in reporting and only more severe cases are recorded for minority mothers. Even if there were no disparities in reporting, the differences in effect could arise from different responses to hypertension. Better prenatal care among White mothers could help counter its effects. To give another example, differences in the effect of smoking could be due to the interaction of risk factors, with smoking more damaging to Black and U.S.-born Hispanic mothers because they also experience other stressors at higher levels. Differences in this effect might also reflect the fact

Figure 1.6. Age-LBW relationship, adjusted for race-specific risk factor effects (M3)



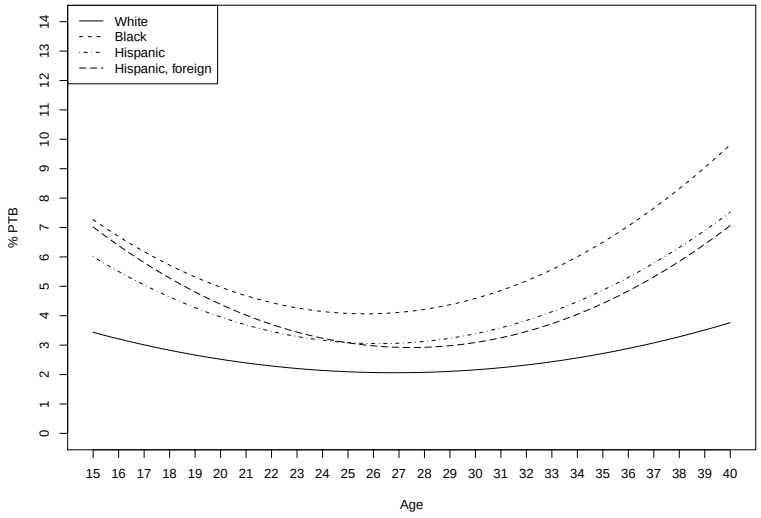
Source: Author's calculations based on data from the Texas Center for Health Statistics.
Notes: The curves shown here are those implied by the race-specific constant terms and coefficients on age and age^2 in Table 1.3, Model 3.

that the coefficient on this variable picks up not only the effects of smoking, *per se*, but also unobserved SES markers that are correlated with smoking.

The race-specific constant terms in Model 3 (Table 1.3) reflect unexplained LBW by race when 1) mothers experience the protective correlates of the risk factors (high school degree, married, etc.) and 2) respond to these protective factors in the way that White mothers do. Because, as noted above, these protective factors are experienced as relatively more protective among Black mothers, allowing for race interactions significantly reduces the overall disparity with Whites. On the other hand, because the same factors are relatively less protective for foreign-born Hispanic mothers, constraining these mothers to experience them in the same way as White mothers increases the overall disparity.

The results of allowing for race-specific effects on both overall and age-related disparities are illustrated in Figure 1.6. Compared to the relative positions of the curves in Figure 1.4, the Black curve is much lower and the foreign-born Hispanic curve is

Figure 1.7. Age-PTB relationship, adjusted for race-specific risk factor effects (M3)

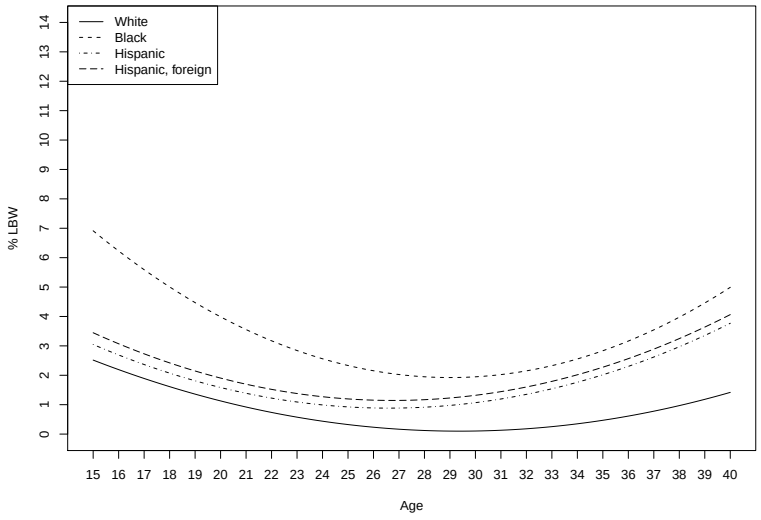


Source: Author's calculations based on data from the Texas Center for Health Statistics.
Notes: The curves shown here are those implied by the race-specific constant terms and coefficients on age and age^2 in Table 1.4, Model 3.

slightly higher. The early weathering effect basically disappears for both Hispanic groups in this version, with unexplained Hispanic-White disparities shrinking with age. The same is not true for Blacks; differences in the LBW rate still increase with age, even with race-specific controls.

Turning to the PTB race-interaction model estimation (Table 1.4), we see similar patterns in the effects of control variables by race, with risk factor coefficients generally larger for Blacks and smaller for foreign-born Hispanics relative to Whites. Because of these different responses, we see the same pattern in race-specific constants, with the overall Black disparity decreasing and the foreign-born Hispanic disparity increasing when protective factors are experienced as White mothers experience them. As in the average effects model, PTB shows greater declines among minority mothers than Whites at younger ages but bigger increases at later ages.

Figure 1.8. Age-LBW relationship, adjusted for race- and age-specific risk factor effects (M4)



Source: Author's calculations based on data from the Texas Center for Health Statistics.
Notes: The curves shown here are those implied by the race-specific constant terms and coefficients on age and age^2 in Table 1.5.

1.4.5 Race and age interaction models

Racial disparities in the age-outcome relationships that remain after controlling for race-specific risk factors could be caused by racial differences in the age distribution of births and/or by disparities in the ways risks interact with age. Geronimus (1996) attributes early weathering among Black mothers to greater stresses associated with low SES, the effects of which accumulate over time. If this is the case, we might expect to see the effect of risk factors increase more sharply with age for the minority groups in our sample. Even if the risk-age interaction does not vary by race, controlling for it may absorb disparities arising from the differences in birth timing patterns illustrated in Figure 1.3.

Table 1.5 shows the results of estimating a model in which the effects of all risk factors may vary by race and age and Figure 1.8 illustrates the residual age-LBW relationships. Compared to Model 3 with only race interactions, we notice that all curves, especially White and Black, have a much higher intercept and a steeper downward

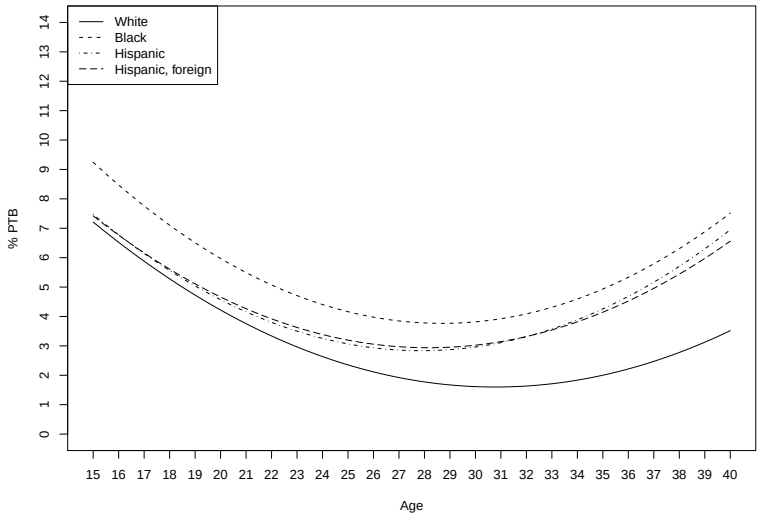
Table 1.5. Low birthweight regression: M4) Multivariate with age and race interactions

Model 4: Age and race interactions

	White	B diff	H diff	HF diff								
Age	−0.336**	−0.376**	−0.037	−0.055								
Age^2	0.0117**	0.0137**	0.0044#	0.0050*								
Implied turning point	29.4	29.0	26.6	26.8								
	Intercept (Effect at age 15)				Age Interaction				Age^2 Interaction			
	White	B diff	H diff	HF diff	White	B diff	H diff	HF diff	White	B diff	H diff	HF diff
Constant	2.52**	4.40**	0.53	0.93*								
Low education	1.15**	−0.53	−0.65*	−1.28**	−0.036	0.184#	0.038	0.096	0.0078**	−0.0056	−0.0040	−0.0092**
Unmarried	0.29	1.01#	0.63*	0.41	0.062	−0.221*	−0.095	−0.039	0.0043*	0.0083#	−0.0010	−0.0051#
Mother smoking	3.28**	−1.22	−0.02	1.94	−0.089#	0.640**	0.210	−0.483#	0.0114**	−0.0190*	−0.0097	0.0155
First birth	0.21	−0.88#	−0.09	0.26	0.137**	0.326**	0.119#	0.055	−0.0004	−0.0117**	−0.0053#	−0.0057*
Third or higher birth	1.48**	−0.06	−1.52**	−1.43**	−0.073	0.022	0.224**	0.145#	−0.0009	−0.0022	−0.0089**	−0.0049#
Late PNC initation	−1.62**	−0.18	1.31**	1.27**	0.268**	−0.021	−0.143#	−0.234**	−0.0045#	0.0055	0.0040	0.0035
Few PNC visits	2.96**	0.82#	−0.13	−1.12**	−0.139**	0.036	−0.054	−0.027	0.0042**	0.0042	0.0046#	0.0027
Many PNC visits	7.76**	3.37**	0.72*	−0.99**	−0.440**	0.124	−0.084	0.089	0.0152**	−0.0022	0.0060*	−0.0008
Diabetes	−3.04**	−3.40#	−0.40	0.70	0.353**	0.321	−0.036	−0.118	−0.0134**	−0.0116	0.0015	0.0047
Chronic hypertension	10.59**	−1.79	5.33	1.24	−0.451	0.874#	−0.141	0.613	0.0169	−0.0311#	0.0071	−0.0213
Preg hypertension	6.18**	−0.08	2.33**	−1.19	−0.047	1.040**	0.637**	0.965**	0.0064	−0.0397**	−0.0287**	−0.0362**
Not in city	−0.76**	−0.02	0.48	0.50	0.048	−0.020	−0.013	0.011	−0.0009	−0.0029	−0.0027	−0.0021
Year controls	Y											
N	3159044											
Adjusted R^2	0.0352											

Source: Author's calculations based on data from the Texas Center for Health Statistics.
Notes: Results from OLS regression with dependent variable low birthweight. Sample includes singleton Texas births to White, B(lack), H(ispanic, U.S. born) and H(ispanic), F(oreign-born) mothers in years 1994-2003. Standard errors (not shown) are adjusted for heterodkedasticity. ** indicates statistically significant at < 0.01; * at < 0.05; # at < 0.10. Significance stars in White columns represent probability that estimated coefficient is different from zero; coefficients in B, H and HF columns represent difference from White coefficient and stars represent probability that the difference is greater than zero. All coefficients multiplied by 100 for legibility.

Figure 1.9. Age-PTB relationship, adjusted for race- and age-specific risk factor effects (M4)



Source: Author's calculations based on data from the Texas Center for Health Statistics.
Notes: The curves shown here are those implied by the race-specific constant terms and coefficients on age and age^2 in Table 1.6.

slope at younger ages. In previous models, we assumed that risk/protective factors had the same effect across ages, and controlling for these average effects flattened the early-age parts of the curves. The more bowed age-LBW curves in this model intimate what we in fact find in this estimation, that the effect of these factors do vary across age.

Comparing the age-LBW relationships across races in Figure 1.8 indicate that disparities between Black and White mothers decline until around age 30 and then widen somewhat again. Increasing disparities with age are visible between Whites and both Hispanic groups, but the statistical significance of the difference is reduced in this model.

Table 1.6 and Figure 1.9 show corresponding results for PTB. Here, the White age-PTB curve becomes even more dramatically bowed. We can see from the constant terms in Table 1.6 that there is no statistically significant difference between White and either Hispanic group at age 15 and that the Black-White disparity is diminished.

Table 1.6. Preterm birth regression: M4) Multivariate with age and race interactions

Model 4: Age and race interactions

	White	B diff	H diff	HF diff								
Age	$-0.711^{\star\star}$	-0.090	-0.009	0.030								
Age^2	$0.0225^{\star\star}$	0.0067	$0.0054^{\#}$	0.0033								
Implied turning point	30.8	28.7	27.9	28.2								
	Intercept (Effect at age 15)				Age Interaction				Age^2 Interaction			
	White	B diff	H diff	HF diff	White	B diff	H diff	HF diff	White	B diff	H diff	HF diff
Constant	$7.21^{\star\star}$	$2.03^{\star\star}$	0.26	0.21								
Low education	$1.58^{\star\star}$	0.50	-0.41	-0.07	$-0.238^{\star\star}$	-0.006	0.057	0.096	$0.0158^{\star\star}$	-0.0005	-0.0020	$-0.0086^{\star}$
Unmarried	$1.32^{\star\star}$	$2.10^{\star\star}$	0.21	0.38	$-0.113^{\star}$	$-0.233^{\star}$	0.070	0.078	$0.0105^{\star\star}$	0.0059	$-0.0070^{\star}$	$-0.0096^{\star\star}$
Mother smoking	$-0.86^{\star\star}$	-1.04	0.86	$3.75^{\star}$	$0.254^{\star\star}$	$0.396^{\star}$	0.057	$-0.558^{\#}$	$-0.0040^{\#}$	-0.0102	-0.0038	$0.0191^{\#}$
First birth	$-3.11^{\star\star}$	-0.50	0.14	0.67	$0.464^{\star\star}$	0.030	-0.006	-0.052	$-0.0104^{\star\star}$	-0.0006	-0.0022	-0.0023
Third or higher birth	$2.48^{\star\star}$	0.45	-0.34	-1.13	-0.068	0.064	0.119	0.091	-0.0032	-0.0029	-0.0054	-0.0021
Late PNC initation	$0.70^{\star}$	$2.11^{\star\star}$	$2.75^{\star\star}$	0.27	0.105	$-0.274^{\star}$	$-0.360^{\star\star}$	$-0.241^{\star\star}$	-0.0013	$0.0121^{\star}$	$0.0146^{\star\star}$	0.0044
Few PNC visits	$3.70^{\star\star}$	$0.84^{\#}$	0.39	$-0.65^{\#}$	$-0.113^{\star}$	0.081	$-0.153^{\star}$	$-0.160^{\star}$	$0.0035^{\#}$	0.0036	$0.0091^{\star\star}$	$0.0075^{\star\star}$
Many PNC visits	$7.67^{\star\star}$	$3.45^{\star\star}$	$1.49^{\star\star}$	0.38	$-0.180^{\star\star}$	0.053	$-0.192^{\star\star}$	$-0.187^{\star}$	$0.0090^{\star\star}$	0.0003	$0.0093^{\star\star}$	$0.0083^{\star\star}$
Diabetes	$2.74^{\star\star}$	-3.25	-1.76	$-3.30^{\star}$	-0.088	0.286	0.254	0.295	-0.0004	-0.0105	-0.0085	-0.0084
Chronic hypertension	$10.12^{\star\star}$	-4.77	-0.26	-0.57	-0.481	$1.401^{\star}$	0.792	0.611	$0.0198^{\#}$	$-0.0543^{\star\star}$	-0.0269	-0.0208
Preg hypertension	$3.84^{\star\star}$	-0.83	$1.81^{\star}$	$-2.22^{\star}$	0.149	$0.602^{\star\star}$	$0.363^{\star}$	$0.633^{\star\star}$	0.0019	$-0.0203^{\star}$	$-0.0153^{\star}$	$-0.0214^{\star\star}$
Not in city	$-0.56^{\star}$	1.18	0.70	-0.32	0.012	-0.122	-0.054	0.110	0.0005	-0.0012	-0.0007	-0.0050
Year controls							Y					
N							3159044					
Adjusted R^2							0.0284					

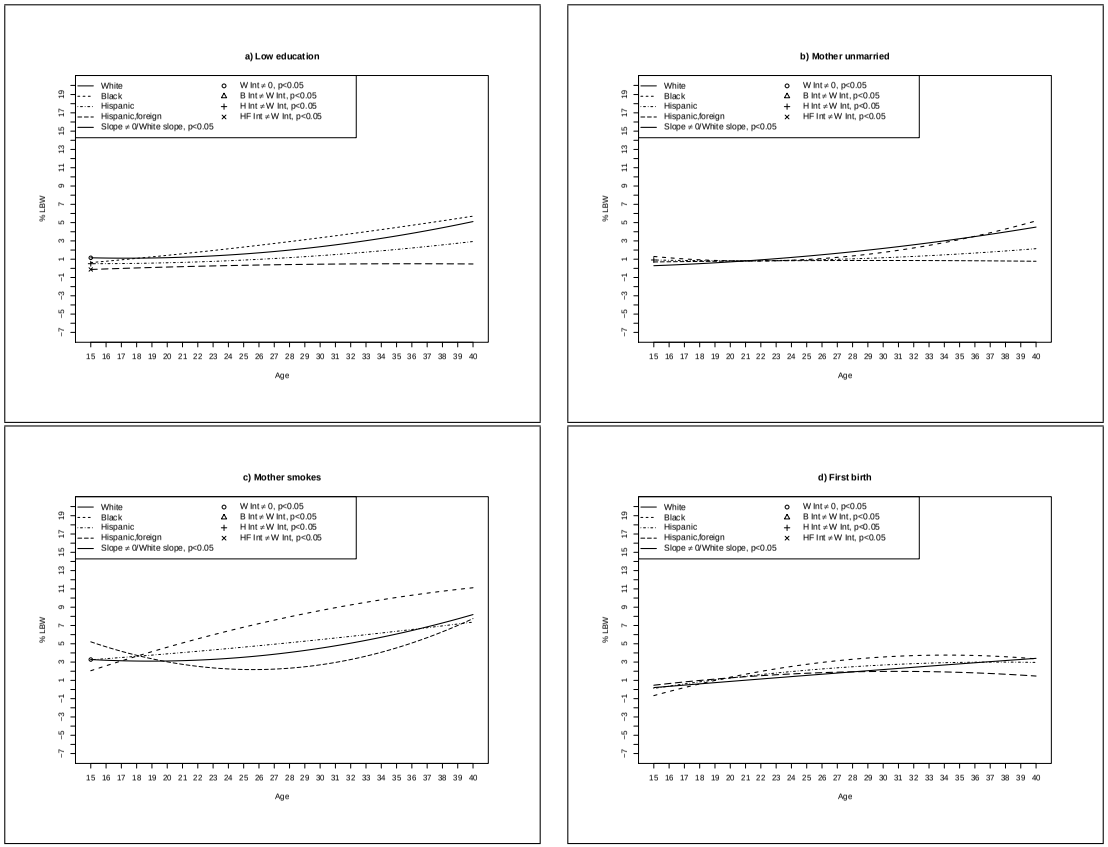
Source: Author's calculations based on data from the Texas Center for Health Statistics.
Notes: Results from OLS regression with dependent variable preterm birth. Sample includes singleton Texas births to White, B(lack), H(ispanic, U.S. born) and H(ispanic), F(oreign-born) mothers in years 1994-2003. Standard errors (not shown) are adjusted for heterodkedasticity. ** indicates statistically significant at < 0.01; * at < 0.05; $^{\#}$ at < 0.10. Significance stars in White columns represent probability that estimated coefficient is different from zero; coefficients in B, H and HF columns represent difference from White coefficient and stars represent probability that the difference is greater than zero. All coefficients multiplied by 100 for legibility.

Further, the age and age-squared coefficients are not statistically significant, though Figure 1.9 indicates that average disparities still increase with age. Still, it is notable that the age-LBW disparities shrink in this model because the left side of the White curve is "pulled upwards" rather than minority curves flattening. This would seem to contradict assertions that increased risk of negative health outcomes at young ages is due primarily to differences in SES, since imposing protective factors with age-specific effects on everyone results in a higher residual PTB rate among teens.

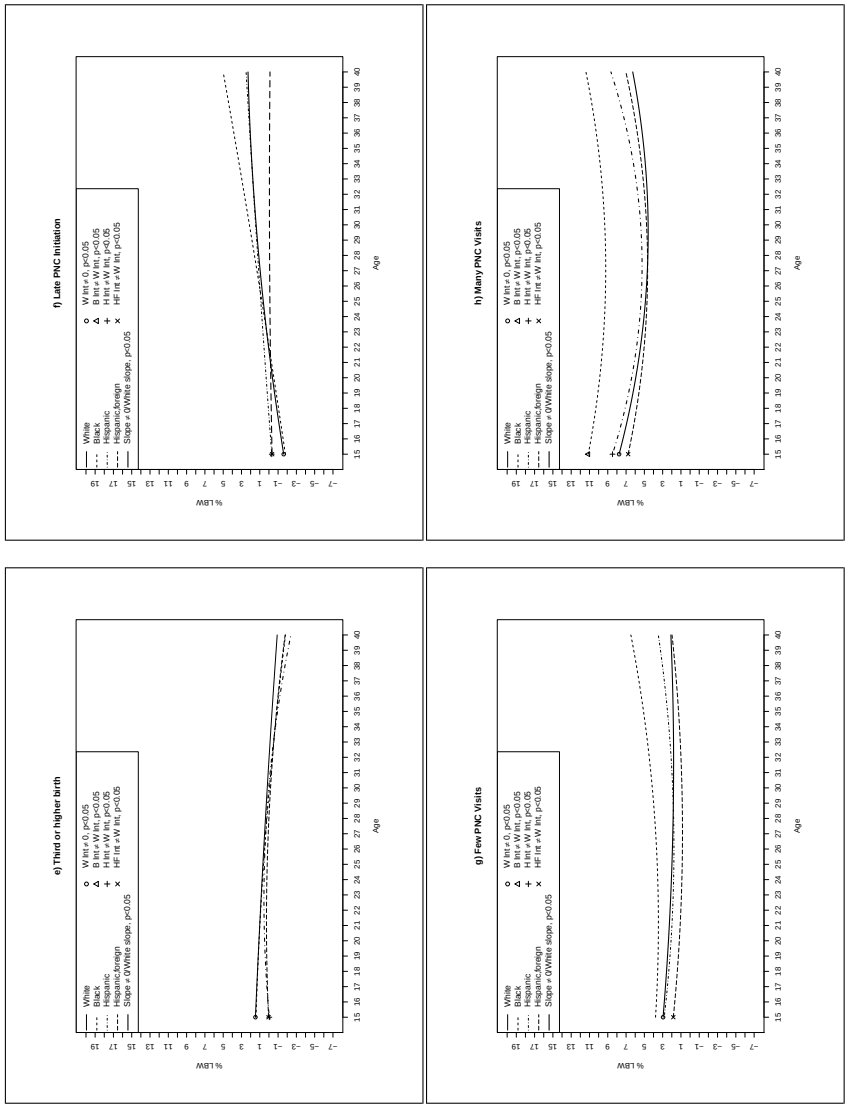
The many coefficients representing the age-outcome relationships for risk factors in Tables 1.5 and 1.6 are difficult to interpret, so we include Figures 1.10 (a-l) and 1.11 (a-l) to illustrate the implied relationships for LBW and PTB, respectively. Solid, bold lines in these figures indicate that the effect of a given risk factor changes with age for White mothers ($p - value$¡0.05), and we see that this is the case in the LBW regression for all risk variables except third or higher birth, the hypertension variables, and the indicator for living outside the city. In the PTB regression, the same three risk factors have age-invariant effects, along with late PNC initation and diabetes. (Variables with effects that do not vary by age are represented by lighter weight lines.)

A notable example of a risk factor that has a changing effect on LBW with age is low education, shown in Figure 1.10 (a). Though far from perfect, low education is perhaps the best indicator we have of low SES. One possible explanation for the increasing "effect" of low education with age is a close cousin of the weathering hypothesis; cumulative effects of stressors related to low SES may lead to declining health with age. In general, it is difficult to separate cumulative effects from more instantaneous effects of risk factors using this data, but because a high school degree cannot be taken away, we know that women without a high school degree at a given age did not have one at any previous age. According to these results, White, Black and U.S.-born Hispanic mothers all experience increasingly negative effects of low education with age, but foreign-born Hispanic mothers do not. This may have to

Figure 1.10. Effect of risk factors on LBW by race and age



Source: Author's calculations based on data from the Texas Center for Health Statistics.
Notes: The curves shown here are those implied by the race-specific coefficients for each risk factor and its interaction with age and age^2 in Table 1.5. A circle indicates that effect of a given risk factor on White mothers at age 15 is different from 0 with p¡0.05. A bold, solid line indicates that the effect of a given risk factor on White mothers varies with age. Other characters at the intercept and bold, dashed lines (details in legend) indicate different risk factor effects for minority mothers.


f) Late PNC Initiation
White
Black
Hispanic
Hispanic,foreign
Slope≠ 0/White slope, p<0.05
o W Int ≠ 0, p<0.05
Δ B Int ≠ W Int, p<0.05
+ H Int ≠ W Int, p<0.05
× HF Int ≠ W Int, p<0.05
% LBW
Age
h) Many PNC Visits
White
Black
Hispanic
Hispanic,foreign
Slope≠ 0/White slope, p<0.05
o W Int ≠ 0, p<0.05
Δ B Int ≠ W Int, p<0.05
+ H Int ≠ W Int, p<0.05
× HF Int ≠ W Int, p<0.05
% LBW
Age
e) Third or higher birth
White
Black
Hispanic
Hispanic,foreign
Slope≠ 0/White slope, p<0.05
o W Int ≠ 0, p<0.05
Δ B Int ≠ W Int, p<0.05
+ H Int ≠ W Int, p<0.05
× HF Int ≠ W Int, p<0.05
% LBW
Age
g) Few PNC Visits
White
Black
Hispanic
Hispanic,foreign
Slope≠ 0/White slope, p<0.05
o W Int ≠ 0, p<0.05
Δ B Int ≠ W Int, p<0.05
+ H Int ≠ W Int, p<0.05
× HF Int ≠ W Int, p<0.05
% LBW
Age

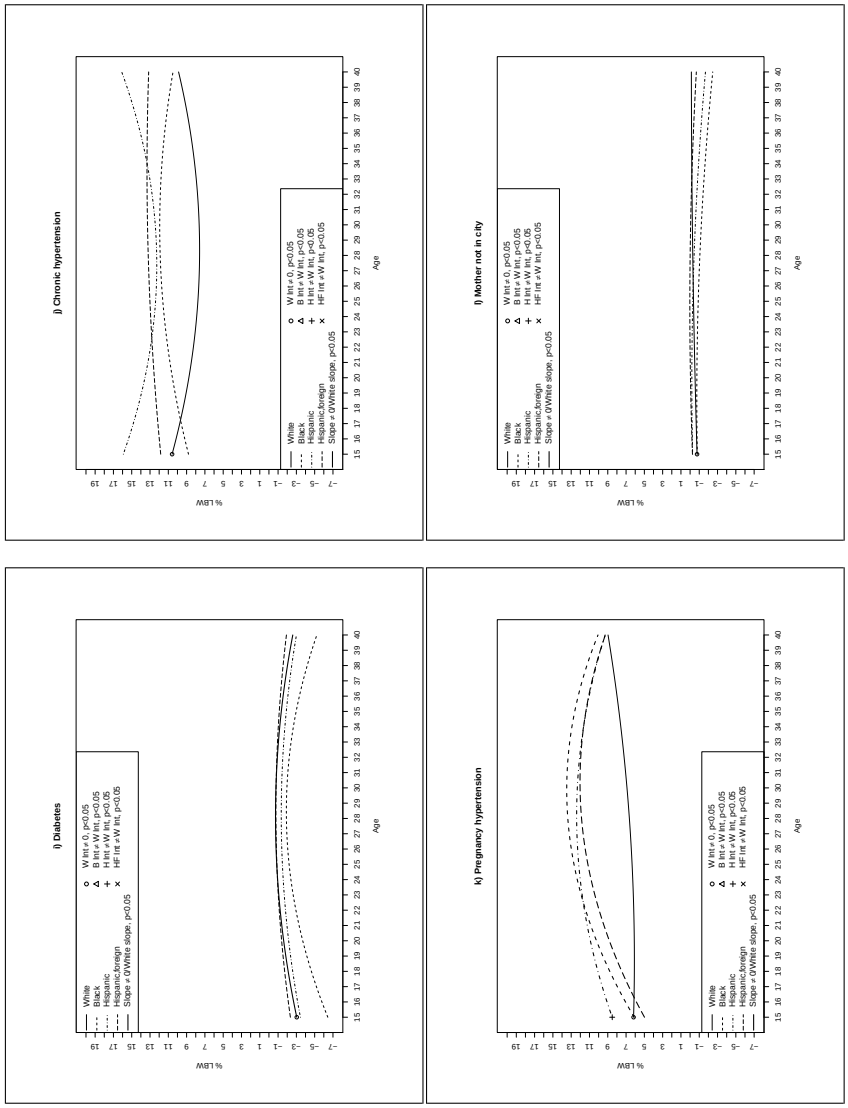
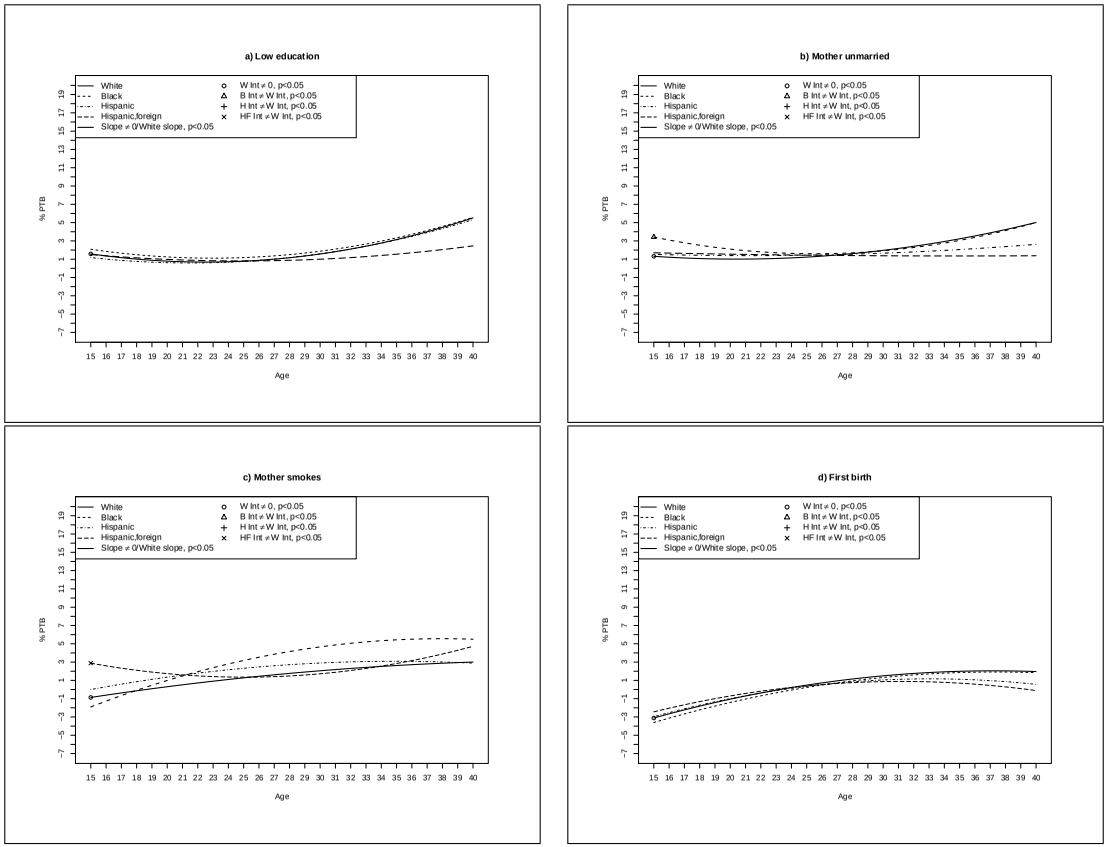
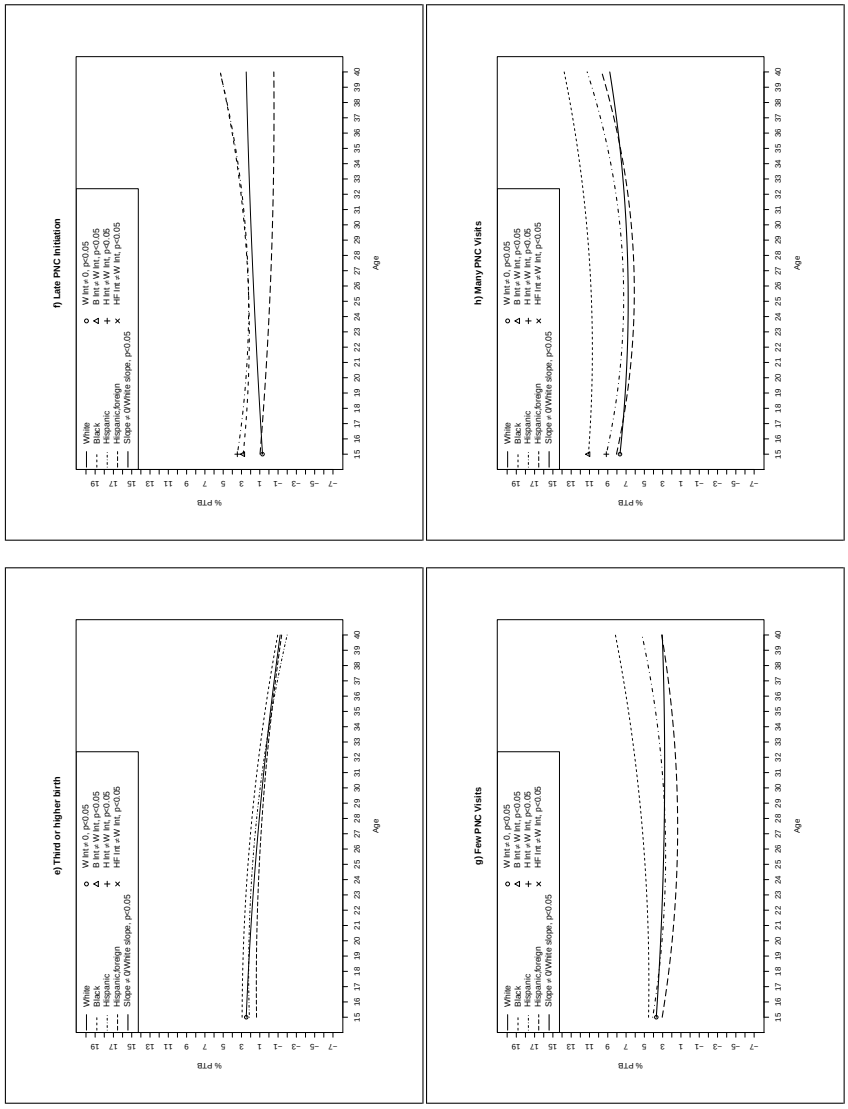
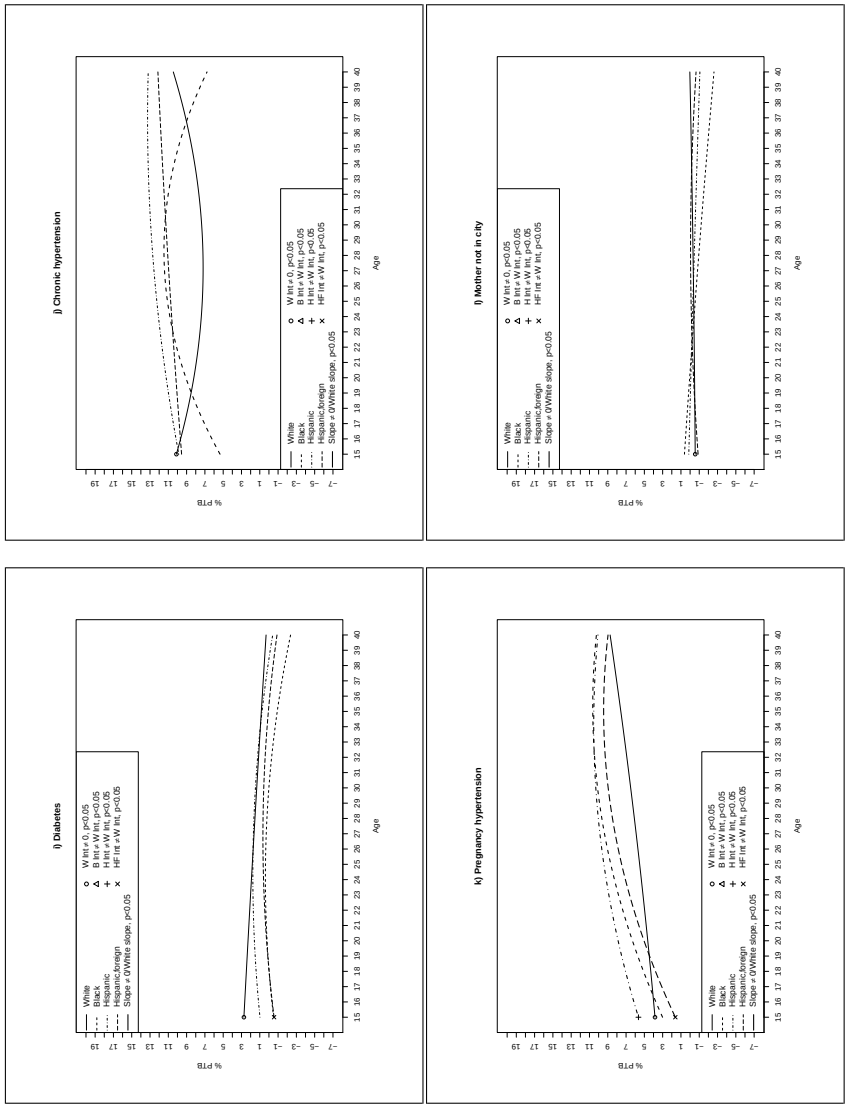

i) Diabetes
White
Black
Hispanic
Hispanic,foreign
Slope ≠ 0/White slope, p<0.05
o W Int ≠ 0, p<0.05
Δ B Int ≠ W Int, p<0.05
+ H Int ≠ W Int, p<0.05
x HF Int ≠ W Int, p<0.05
% LBW
Age
j) Chronic hypertension
k) Pregnancy hypertension
l) Mother not in city

Figure 1.11. Effect of risk factors on PTB by race and age



Source: Author's calculations based on data from the Texas Center for Health Statistics.
Notes: The curves shown here are those implied by the race-specific coefficients for each risk factor and its interaction with age and age^2 in Table 1.6. A circle indicates that effect of a given risk factor on White mothers at age 15 is different from 0 with $p < 0.05$. A bold, solid line indicates that the effect of a given risk factor on White mothers varies with age. Other characters at the intercept and bold, dashed lines (details in legend) indicate different risk factor effects for minority mothers.


f) Late PNC Initiation
White
Black
Hispanic
Hispanic,foreign
Slope≠ 0/White slope, p<0.05
o W Int ≠ 0, p<0.05
Δ B Int ≠ W Int, p<0.05
+ H Int ≠ W Int, p<0.05
× HF Int ≠ W Int, p<0.05
% PTB
Age
h) Many PNC Visits
White
Black
Hispanic
Hispanic,foreign
Slope≠ 0/White slope, p<0.05
o W Int ≠ 0, p<0.05
Δ B Int ≠ W Int, p<0.05
+ H Int ≠ W Int, p<0.05
× HF Int ≠ W Int, p<0.05
% PTB
Age
e) Third or higher birth
White
Black
Hispanic
Hispanic,foreign
Slope≠ 0/White slope, p<0.05
o W Int ≠ 0, p<0.05
Δ B Int ≠ W Int, p<0.05
+ H Int ≠ W Int, p<0.05
× HF Int ≠ W Int, p<0.05
% PTB
Age
g) Few PNC Visits
White
Black
Hispanic
Hispanic,foreign
Slope≠ 0/White slope, p<0.05
o W Int ≠ 0, p<0.05
Δ B Int ≠ W Int, p<0.05
+ H Int ≠ W Int, p<0.05
× HF Int ≠ W Int, p<0.05
% PTB
Age


i) Diabetes
j) Chronic hypertension
k) Pregnancy hypertension
l) Mother not in city
% PTB
Age
White
Black
Hispanic
Hispanic,foreign
Slope ≠ 0/White slope, p<0.05
W Int ≠ 0, p<0.05
B Int ≠ W Int, p<0.05
H Int ≠ W Int, p<0.05
HF Int ≠ W Int, p<0.05

do with different correlations between low education and dietary or lifestyle habits between U.S. and foreign-born mothers or different ways in which low SES is translated to internal stress between these groups, but it is not possible to understand the mechanisms using this data.

Interestingly, the relationship between low education and PTB across ages appears somewhat different. Rather than increasing monotonically with age, as we see in the LBW case, the effect of low education appears to have a constant relationship with age until around age 26, at which point it increases. Again, the foreign-born Hispanic curve diverges from the other groups, with a barely increasing effect of low education with age.

A dashed, bold line in Figures 1.10 and 1.11 indicates that the effect of a risk factor for the corresponding minority group interacts with age differently from the White relationship. We do see some significant racial variation in the effects of risk factors across age, as we noted in the different effect of low education across ages for foreign-born Hispanic mothers, but the magnitude of the differences are not drastic in most cases. One notable exception is the effect of smoking on LBW. We see from the solid line in Figure 1.10 (c) that the effect of smoking increases with age for White mothers. Since we do not have information on lifetime smoking habits, we cannot infer whether this trend reflects cumulative effects of smoking, cumulative effects of other conditions that are correlated with smoking, more time-localized effects of smoking that interact with biology or associated SES factors that vary with age, or something else. Whatever the explanation, we see that the effect of smoking for Black mothers increases much more with age than for the other three groups. Since Black mothers are much less likely to smoke at younger ages than White mothers, this reduces the plausibility of the explanation that the relationship with age simply reflects the cumulative effects of smoking *per se*.

1.4.6 Models with neighborhood income or deprivation measures

Previous research showed that neighborhood SES exerts a large effect on birth outcomes, so we also estimate models that include such measures. Tables 1.7 and 1.8 show the effect of alternately adding median household income and the deprivation index (see p. 23 for a description) to the bivariate and main effects models. The first column in each column group reprints the estimates from models using the full sample and no neighborhood SES measure. The second column shows the estimation of the same model using the smaller sample that includes only geocoded births that could be attached to Census tracts. The foreign-born Hispanic constants and age coefficients change somewhat dramatically between these two versions; other coefficients do not. One possible explanation is that foreign-born Hispanic mothers who could not be geocoded, and are therefore omitted from the Census tract sample, are more likely to be migrant farmworkers who exhibit a systematically different relationship between age and negative birth outcomes. The representation of each race group is about the same in each sample.

The third column in Table 1.7 adds median household income to the bivariate LBW model. The income coefficient implies that a one standard deviation increase in median neighborhood income ($19,627, 2000 USD) decreases the probability of LBW by 0.65 percentage points. Since income and the deprivation index are highly correlated, it is not suprising that this estimate is nearly the same as the effect of a one standard deviation decrease in the neighborhood deprivation index, which decreases the probability of LBW by 0.64 points. The observed effect of neighborhood income in the bivariate PTB model is quite similar: the effect of a one standard deviation increase in income is a reduction in probability of PTB of 0.75 points and the effect of a one standard deviation decrease in the deprivation index is a reduction of 0.76 points.

Table 1.7. Low birthweight regressions: Controlling for neighborhood SES

	Model 1: Bivariate				Model 2: Main Effects			
	a) Full	b) CT	c) CT + Inc	d) CT + Dep	a) Full	b) CT	c) CT + Inc	d) CT + Dep
White Constant	8.51**	8.42**	8.31**	8.24**	0.71**	0.63**	0.66**	0.63**
B diff	4.98**	5.08**	4.93**	4.71**	5.85**	5.92**	5.80**	5.64**
H diff	0.10	0.12	0.00	−0.26#	1.36**	1.31**	1.20**	1.01**
FH diff	−1.46**	−1.25**	−1.37**	−1.72**	0.31*	0.42**	0.29*	0.04
Age	−0.639**	−0.631**	−0.581**	−0.585**	−0.194**	−0.192**	−0.167**	−0.172**
Age^2	0.0218**	0.0213**	0.0211**	0.0206**	0.0096**	0.0094**	0.0097**	0.0094**
Implied turning point	29.6	29.8	28.8	29.2	25.1	25.2	23.6	24.1
B*Age	−0.086*	−0.103**	−0.136**	−0.120**	−0.177**	−0.187**	−0.206**	−0.195**
B*Age^2	0.0105**	0.0113**	0.0117**	0.0117**	0.0102**	0.0107**	0.0108**	0.0108**
Implied turning point	26.2	26.3	25.9	25.9	24.3	24.4	24.1	24.1
H*Age	0.052*	0.052*	0.013	0.024	−0.028	−0.022	−0.045#	−0.038
H*Age^2	0.0024*	0.0026*	0.0031**	0.0033**	0.0026*	0.0025*	0.0026*	0.0029**
Implied turning point	27.1	27.2	26.8	26.7	24.1	24.0	23.6	23.6
FH*Age	0.141**	0.115**	0.077**	0.088**	−0.007	−0.021	−0.038	−0.028
FH*Age^2	−0.0024**	−0.0011	−0.0011	−0.0009	0.0004	0.0012	0.0009	0.0010
Implied turning point	27.8	27.8	27.6	27.6	25.0	25.0	24.7	24.6
Low education					1.00**	1.01**	0.93**	0.88**
Unmarried					1.14**	1.14**	1.08**	1.07**
Mother smoking					4.71**	4.69**	4.57**	4.62**
First birth					1.71**	1.70**	1.74**	1.75**
Third or higher birth					0.10**	0.11**	0.04	0.03
Late PNC initation					0.33**	0.42**	0.39**	0.37**
Few PNC visits					1.94**	1.95**	1.94**	1.92**
Many PNC visits					5.88**	5.92**	5.93**	5.93**
Diabetes					−1.37**	−1.38**	−1.44**	−1.42**
Chronic hypertension					10.51**	10.36**	10.29**	10.32**
Preg hypertension					9.04**	9.08**	9.05**	9.06**
Not in city					−0.40**	−0.56**	−0.46**	−0.47**
Tract med HH inc ($000)			−0.033**				−0.025**	
Tract deprivation index				0.64**				0.52**
Year controls	Y	Y	Y	Y	Y	Y	Y	Y
N	3159044	2727139	2727139	2727139	3159044	2727139	2727139	2727139
Adjusted R^2	0.0079	0.0082	0.0088	0.0087	0.0324	0.0329	0.0332	0.0332

Source: Author's calculations based on data from the Texas Center for Health Statistics merged with data from the 2000 U.S. Census.
Notes: Results from OLS regressions with dependent variable low birthweight. Full Sample includes singleton Texas births to White, B(lack), H(ispanic, U.S. born) and H(ispanic), F(oreign-born) mothers in years 1994-2003. C(ensus) T(ract) sample includes births from the full sample that could be matched to tract-level Census data on household income and variables included in the deprivation index, constructed using the method described in Messer et al. (2006). Deprivation index variables are % males in management, % households with > 1 person per room, % individuals with 1999 income below federal poverty level, % families with female headed household with dependent children, % households with income less than $30,000, % households with public assistance income, % unemployed, and % adults with no high school education. Standard errors (not shown) are adjusted for heterodkedasticity. ** indicates statistically significant at < 0.01; * at < 0.05; # at < 0.10. Significance stars in White columns represent probability that estimated coefficient is different from zero; coefficients in B, H and HF columns represent difference from White coefficient and stars represent probability that the difference is greater than zero. All coefficients multiplied by 100 for legibility.

Table 1.8. Preterm birth regressions: Controlling for neighborhood SES

	Model 1: Bivariate				Model 2: Main Effects			
	a) Full	b) CT	c) CT + Inc	d) CT + Dep	a) Full	b) CT	c) CT + Inc	d) CT + Dep
White Constant	10.59**	10.47**	10.35**	10.26**	3.61**	3.46**	3.48**	3.46**
B diff	6.55**	6.51**	6.34**	6.08**	6.49**	6.47**	6.35**	6.21**
H diff	1.97**	2.01**	1.86**	1.55**	2.37**	2.38**	2.27**	2.10**
FH diff	0.87**	1.20**	1.06**	0.64**	1.83**	2.15**	2.02**	1.79**
Age	−0.621**	−0.608**	−0.551**	−0.554**	−0.263**	−0.253**	−0.227**	−0.234**
Age^2	0.0214**	0.0207**	0.0205**	0.0199**	0.0104**	0.0099**	0.0102**	0.0099**
Implied turning point	29.5	29.7	28.4	28.9	27.6	27.7	26.1	26.8
B*Age	−0.268**	−0.271**	−0.309**	−0.291**	−0.375**	−0.376**	−0.395**	−0.384**
B*Age^2	0.0170**	0.0175**	0.0179**	0.0179**	0.0189**	0.0192**	0.0193**	0.0193**
Implied turning point	26.6	26.5	26.2	26.2	25.9	25.8	25.6	25.6
H*Age	−0.115**	−0.118**	−0.163**	−0.151**	−0.222**	−0.222**	−0.245**	−0.237**
H*Age^2	0.0101**	0.0105**	0.0110**	0.0114**	0.0120**	0.0123**	0.0124**	0.0127**
Implied turning point	26.7	26.6	26.3	26.3	25.8	25.7	25.4	25.4
FH*Age	−0.177**	−0.201**	−0.245**	−0.233**	−0.353**	−0.368**	−0.385**	−0.375**
FH*Age^2	0.0108**	0.0119**	0.0119**	0.0122**	0.0144**	0.0149**	0.0146**	0.0147**
Implied turning point	27.4	27.4	27.3	27.3	27.4	27.5	27.4	27.3
Low education					1.20**	1.16**	1.09**	1.04**
Unmarried					1.54**	1.51**	1.45**	1.44**
Mother smoking					1.68**	1.67**	1.55**	1.60**
First birth					0.09*	0.12**	0.16**	0.17**
Third or higher birth					0.87**	0.89**	0.82**	0.82**
Late PNC initation					1.49**	1.70**	1.67**	1.66**
Few PNC visits					2.90**	2.92**	2.90**	2.89**
Many PNC visits					7.64**	7.73**	7.73**	7.73**
Diabetes					1.06**	0.99**	0.93**	0.95**
Chronic hypertension					9.82**	9.81**	9.75**	9.78**
Preg hypertension					7.01**	6.97**	6.94**	6.95**
Not in city					−0.41**	−0.51**	−0.41**	−0.42**
Tract med HH inc ($000)			−0.038**				−0.026**	
Tract deprivation index				0.76**				0.49**
Year controls	Y	Y	Y	Y	Y	Y	Y	Y
N	3159044	2727139	2727139	2727139	3159044	2727139	2727139	2727139
Adjusted R^2	0.0069	0.0070	0.0075	0.0075	0.0264	0.0269	0.0271	0.0271

Source: Author's calculations based on data from the Texas Center for Health Statistics merged with data from the 2000 U.S. Census.
Notes: Results from OLS regressions with dependent variable low birthweight. Full Sample includes singleton Texas births to White, B(lack), H(ispanic, U.S. born) and H(ispanic), F(oreign-born) mothers in years 1994-2003. C(ensus) T(ract) sample includes births from the full sample that could be matched to tract-level Census data on household income and variables included in the deprivation index, constructed using the method described in Messer et al. (2006). Deprivation index variables are % males in management, % households with > 1 person per room, % individuals with 1999 income below federal poverty level, % families with female headed household with dependent children, % households with income less than $30,000, % households with public assistance income, % unemployed, and % adults with no high school education. Standard errors (not shown) are adjusted for heterodkedasticity. ** indicates statistically significant at < 0.01; * at < 0.05; # at < 0.10. Significance stars in White columns represent probability that estimated coefficient is different from zero; coefficients in B, H and HF columns represent difference from White coefficient and stars represent probability that the difference is greater than zero. All coefficients multiplied by 100 for legibility.

The last two columns in Tables 1.7 and 1.8 show the effects of alternately adding income and the deprivation index to the LBW and PTD main effects models, respectively. Since the coefficients on these neighborhood variables remain statistically significant in this version, we infer that neighborhood characteristics and/or unobserved individual characteristics for which these measures proxy have influence beyond the variables that are accounted for in the model. Also notable is the fact that coefficients on the individual-level variables do not change significantly, though we do see a slight reduction in magnitude of the estimated coefficients on the variables most associated with lower SES, including low education, unmarried status, smoking, three or more births, late PNC and few PNC visits. Though not presented here, the residual age-outcome curves look very similar with or without neighborhood SES controls.

Table 1.9. Low birthweight regressions: Effect of neighborhood income control on other coefficients

Model 2: Main effects	Model 3: Race interaction	Model 4: Age and race interaction
Coefficient increased more than 20% when income added to model and p-value< 0.10 in either version.		
None	B*Mother smoking *0.52 (0.020) to 0.66 (0.003)*	Unmarried *0.43 (0.071) to 0.61 (0.010)*
	B*Not in city *-0.90 (.000) to -0.56 (0.018)*	
	H*Not in city *-0.23 (0.081) to -0.05 (0.711)*	
Coefficient decreased more than 20% when income added to model and p-value< 0.10 in either version.		
FH diff *0.42 (0.005) to 0.29 (0.048)*	B*age *-0.081 (0.081) to -0.100 (0.032)*	H diff *0.67 (0.057) to 0.49 (0.181)*
H*age *-0.022 (0.377) to -0.045 (0.072)*	H*age *-0.095 (0.001) to -0.115 (0.000)*	Not in city *-0.69 (0.009) to -0.51 (0.052)*
Third or higher birth *0.11 (0.003) to 0.04 (0.279)*	Third or higher birth *0.24 (0.000) to 0.15 (0.005)*	B*Unmarried*Age^2 *0.0082 (0.079) to 0.0060 (0.21)*
	B*Low education *0.31 (0.079) to 0.25 (0.162)*	
	B*Unmarried *0.29 (0.046) to 0.12 (0.407)*	
	H*First birth *-0.22 (0.011) to -0.26 (0.002)*	

Source: Author's calculations based on data from the Texas Center for Health Statistics merged with data from the 2000 U.S. Census.
Notes: Results from OLS regressions with dependent variable low birthweight. Sample is Census Tract sample described in the notes to Table 1.7; it includes 2727139 births to mothers who could be located within a Census tract with median household income information. Numbers below variable names represent *coefficient in model without income control* (p-value) to *coefficient in model with income control* (p-value).

Table 1.10. Preterm birth regressions: Effect of neighborhood income control on other coefficients

Model 2: Main effects	Model 3: Race interaction	Model 4: Age and race interaction
Coefficient increased more than 20% when income added to model and p-value< 0.10 in either version.		
First birth	H*Mother smoking	B*age^2
0.12 (0.005) to 0.16 (0.000)	*0.69 (0.004) to 0.83 (0.001)*	*0.0069 (0.144) to 0.0096 (0.052)*
	H*Not in city	
	0.19 (0.257) to 0.39 (0.020)	
	FH*Third or higher birth	
	-0.22 (0.055) to -0.15 (0.190)	
Coefficient decreased more than 20% when income added to model and p-value< 0.10 in either version.		
None	None	Unmarried*age
		-0.079 (0.149) to -0.110 (0.046)

Source: Author's calculations based on data from the Texas Center for Health Statistics merged with data from the 2000 U.S. Census. *Notes:* Results from OLS regressions with dependent variable low birthweight. Sample is Census Tract sample described in the notes to Table 1.7; it includes 2727139 births to mothers who could be located within a Census tract with median household income information. Numbers below variable names represent *coefficient in model without income control* (p-value) to *coefficient in model with income control* (p-value).

We also investigated the effect of including median neighborhood income in the interaction models. Tables 1.9 and 1.10 list those variables for which estimated coefficients increased or decreased more than 20 percent in the version with income controlled, for the three models that include control variables. To be listed, we also required that the coefficient was statistically significant at 0.10 in one of the estimations. Relative to the number of estimated coefficients in the interaction models, the number that changed in this manner is very small and the direction of change is generally unsurprising. The similarity between models with and without neighborhood SES measures predicated our use of the full sample for the main analyses.

In the LBW model only, the main age-income interaction (not shown here) is statistically significant at $p < 0.01$, indicating the age-LBW relationship is mediated by neighborhood income, which other research has also shown. Upon closer analysis of the coefficients, however, we find that while neighborhood income does affect LBW in the expected direction, it does little to mediate the age-LBW relationship, at least for White mothers (i.e., though the relevant coefficients are statistically significant, they are small in magnitude). It is probable that here, too, controlling for all the individual-

level risk factors and their age interactions absorbed explanatory power that might have been attributed to neighborhood income in a model with fewer controls.

We do find that the age-income interaction coefficient for foreign-born Hispanic mothers is different from the White coefficient at at $p < 0.01$, and the relationship here is what we expect: living in a poorer area is associated with increasingly high LBW rates with age. The results imply, for example, that a foreign-born Hispanic mother living in a tract with 1999 median household income of $11,300, the 25th birth-weighted percentile, can expect a 0.3 percentage point greater increase in LBW probability between ages 15 and 40 than a mother living in a tract with income at the sample median of $19,600, all else equal. The age-income interaction coefficients for U.S.-born Hispanics and Blacks are not statistically significantly different from the White estimates, though the magnitude of the Black effect is similar to the foreign-born Hispanic effect.

1.4.7 Models with mother fixed effects

Though our main objective was not to estimate the effect of the independent variables used as controls, examining the short-term effect of changing these variables may have some policy relevance. We are able to do this with regressions utilizing only within-mother variation in the risk factors and dependent variables. Since there are many mothers with more than one child in our sample, OLS estimates using the full sample represent a weighted average of the estimates that would be obtained using only between-mother variation and estimates using only within-mother variation [11]. It is not possible to know, looking only at the OLS estimates, how much of each type of variation they represent. It could be, to give a somewhat dramatic example, that the observed effects of smoking are mostly attributable not to inhaling smoke *per se* but to characteristics associated with the type of woman who smokes. Within-mother regressions can help us address questions like these. If the OLS estimates

are much larger than the within-mother estimates, we can infer that the independent variables are markers of cumulative effects of either the risk factors as such and/or the unobserved variables they represent. If, on the other hand, the within estimates are similar in size to the OLS estimates, we can infer that changing the variable (or its close correlates) can have a more near-term effect on infant health.

Two qualifications must be noted before discussing the results. First, foreign-born Hispanic mothers were excluded from this analysis, since their representation in the sample that contained mother identifiers was proportionally very low. This under-representation is not surprising, since the mother identifying variable was derived from social security numbers. Second, the age variables were included in the within regression solely to separate the effects of advancing age from the effects of variation in other variables, which are naturally correlated with the passage of time. The coefficients on age in these models cannot be interpreted in the same way as in OLS since the within estimator does not allow the slope of the age-outcome relationship to be dependent on the level. By de-meaning all variables within mothers, the model does not distinguish, for example, between a mother who had two children at ages 15 and 20 and another who had two children at 35 and 40. Therefore, we do not compare the age effects in these regressions to those obtained using OLS.

The first column group in Table 1.11 (Table 1.12) shows the results from the LBW (PTB) regression with race interactions on all variables using the full Census tract sample with neighborhood income. The second column group shows the results from estimating the same model using only births with at least one sibling in the sample, since this is the sample that is used in the within-mother regression. The effects of risk factors in this reduced sample are attenuated somewhat but are quite similar to those from the larger sample.

The third column group in Table 1.11 (Table 1.12) shows the results of the within-mother regression with LBW (PTB) as the dependent variable. We see that, in

Table 1.11. Low birthweight regressions: Race interactions with mother fixed effects

Race interaction models with household income

	a) CT sample, OLS			b) CT sibling sample, OLS			c) CT sibling sample, Mother FE		
	White	B diff	H diff	White	B diff	H diff	White	B diff	H diff
Constant	0.45**	2.27**	1.26**	1.35**	2.70**	0.87**		2.71**	
Age	−0.101**	−0.100*	−0.115**	−0.147**	−0.162*	−0.099*	0.055	−0.24*	−0.113
Age^2	0.0077**	0.0074**	0.0047**	0.0065**	0.0093**	0.0051**	0.0001	0.0046	0.0067*
Low education	1.56**	0.25	−0.67**	1.40**	−0.10	−0.51**	−0.71**	1.14**	0.59#
Unmarried	1.29**	0.12	−0.37**	1.18**	−0.02	−0.40**	0.58**	−0.21	−0.12
Mother smoking	3.97**	3.11**	0.66**	3.87**	2.94**	0.78*	0.82**	2.44**	1.70**
First birth	1.81**	0.42**	−0.13	1.43**	0.53*	−0.04	1.72**	0.24	0.19
Third or higher birth	0.15**	0.06	−0.14	0.42**	0.26	−0.05	0.24*	0.24	−0.35*
Late PNC initation	0.56**	−0.19	0.12	0.33*	−0.17	0.27	0.33*	0.28	0.52*
Few PNC visits	2.12**	1.75**	−0.01	1.89**	1.93**	0.06	1.39**	1.12**	0.10
Many PNC visits	5.13**	4.62**	1.03**	4.63**	4.69**	1.11**	2.85**	3.32**	1.18**
Diabetes	−1.08**	−1.76**	−0.73**	−1.10**	−1.25*	−0.88**	−0.72**	−0.41	−0.57
Chronic hypertension	7.88**	3.62**	4.81**	5.48**	3.47**	6.92**	1.69*	0.87	5.08**
Preg hypertension	6.64**	4.53**	4.55**	5.72**	4.75**	5.08**	3.81**	4.17**	3.69**
Not in city	−0.37**	−0.56*	−0.05	−0.24**	−0.86*	−0.05	−0.02	−0.38	−0.14
Tract med HH inc ($000)	−0.0246**	−0.0205**	−0.0066**	−0.0190**	−0.0275**	−0.0113**	−0.0002	−0.0023	−0.0017
Year controls		Y			Y			N	
N		2079005			1055117			1055117	
Adjusted R^2		0.0372			0.0343				
Number mothers								467916	
Within R^2								0.0120	
Rho								0.4067	

Source: Author's calculations based on data from the Texas Center for Health Statistics merged with data from the 2000 U.S. Census.
Notes: Results from OLS (Columns a and b) and within mother (Column c) regressions with dependent variable low birthweight. C(ensus) T(ract) sample includes White, Non-Hispanic Black and Hispanic singleton Texas births from 1994-2003 that could be matched to Census tracts. Sibling sample includes births from the CT sample to mothers whose race did not change between birth records and who had at least two children in the sample. The mother F(ixed) E(ffect) regression in Column c utilizes only within-mother variation in the dependent and independent variables. *Rho* is the fraction of variance explained by mother fixed effects. Births to foreign-born Hispanic mothers were excluded from these samples due to low reporting of the mother identification variable. Tract-level median household income is from the 2000 U.S. Census. Standard errors (not shown) are adjusted for heterodkedasticity in (a) and (b) and clustered at mother level in (c). ** indicates statistically significant at < 0.01; * at < 0.05; # at < 0.10. Significance stars in White columns represent probability that estimated coefficient is different from zero; coefficients in B(lack) and H(ispanic, U.S. born) columns represent difference from White coefficient and stars represent probability that the difference is greater than zero. All coefficients multiplied by 100 for legibility.

Table 1.12. Preterm birth regressions: Race interactions with mother fixed effects

Race interaction models with household income

	a) CT sample, OLS			b) CT sibling sample, OLS			c) CT sibling sample, Mother FE		
	White	B diff	H diff	White	B diff	H diff	White	B diff	H diff
Constant	3.36**	3.77**	2.37**	4.42**	3.62**	1.82**		5.66**	
Age	−0.196**	−0.395**	−0.315**	−0.286**	−0.377**	−0.27**	0.156*	−0.802**	−0.46**
Age^2	0.0096**	0.0189**	0.0141**	0.0096**	0.0186**	0.0142**	−0.0006	0.0260**	0.0208**
Low education	1.22**	0.31	−0.21	1.08**	0.10	−0.23	−0.62*	1.14*	0.79*
Unmarried	1.33**	0.54**	0.03	1.35**	0.53*	0.00	0.23	0.40	0.09
Mother smoking	1.18**	1.61**	0.83**	1.12**	1.78**	0.91**	−0.44#	2.11**	1.48**
First birth	0.68**	−0.96**	−1.00**	0.34**	−1.04**	−0.88**	0.73**	−1.64**	−0.96**
Third or higher birth	0.52**	1.14**	0.59**	0.92**	1.28**	0.53**	0.84**	0.84*	−0.05
Late PNC initation	1.79**	0.72**	1.02**	1.52**	0.66*	1.31**	0.99**	0.55	1.26**
Few PNC visits	3.08**	2.05**	0.08	2.93**	2.02**	−0.03	2.21**	0.84**	−0.21
Many PNC visits	7.09**	4.15**	0.78**	6.83**	4.20**	0.77**	4.44**	3.31**	0.91**
Diabetes	1.49**	−1.61**	−0.12	1.28**	−1.79*	−0.34	−0.76#	−2.05#	0.29
Chronic hypertension	7.90**	2.48**	4.58**	6.08**	3.08**	6.50**	1.82*	3.61*	6.29**
Preg hypertension	5.78**	1.84**	2.85**	5.10**	2.12**	2.67**	2.90**	1.80*	1.48**
Not in city	−0.45**	0.01	0.39*	−0.38**	−0.29	0.31	−0.16	0.47	0.38
Tract med HH inc ($000)	−0.0258**	−0.0094*	−0.0088**	−0.0215**	−0.0085	−0.012**	0.0029	−0.0009	−0.0055
Year controls		Y			Y			N	
N		2079005			1055117			1055117	
Adjusted R^2		0.0314893			0.0306				
Number mothers								467916	
Within R^2								0.0092	
Rho								0.3753	

Source: Author's calculations based on data from the Texas Center for Health Statistics merged with data from the 2000 U.S. Census.
Notes: Results from OLS (Columns a and b) and within mother (Column c) regressions with dependent variable preterm birth. C(ensus) T(ract) sample includes White, Non-Hispanic Black and Hispanic singleton Texas births from 1994-2003 that could be matched to Census tracts. Sibling sample includes births from the CT sample to mothers whose race did not change between birth records and who had at least two children in the sample. The mother F(ixed) E(ffect) regression in Column c utilizes only within-mother variation in the dependent and independent variables. *Rho* is the fraction of variance explained by mother fixed effects. Births to foreign-born Hispanic mothers were excluded from these samples due to low reporting of the mother identification variable. Tract-level median household income is from the 2000 U.S. Census. Standard errors (not shown) are adjusted for heterodkedasticity in (a) and (b) and clustered at mother level in (c). ** indicates statistically significant at < 0.01; * at < 0.05; # at < 0.10. Significance stars in White columns represent probability that estimated coefficient is different from zero; coefficients in B(lack) and H(ispanic, U.S. born) columns represent difference from White coefficient and stars represent probability that the difference is greater than zero. All coefficients multiplied by 100 for legibility.

general, the effects of within-mother variation in these risk factors are smaller but not negligible, indicating that changing these variables may have short-term effects. The only variables that do not appear to have short-run effects are the neighborhood variables—moving in or out of a city and moving to a tract with higher or lower median neighborhood income.

While the OLS regressions show that mothers who have a high school degree are less likely to experience LBW and PTB, the within regressions indicate that for White mothers, obtaining a high school degree between births is actually deleterious to birth outcomes. Conjuring a reasonable explanation for this negative effect of education on White mothers is difficult. Receiving a high school degree appears to be protective for Black mothers and somewhat neutral for Hispanic mothers.

In the LBW regression, the coefficient on smoking is positive but smaller than in the OLS regression, indicating both that there are immediate effects and that cumulative effects of smoking or its correlates are quite important. The racial disparities in the effects of smoking are upheld in the within regression, with both Black and Hispanic mothers experiencing worse birth outcomes with smoking. With respect to PTB, which we saw earlier is less correlated with smoking, becoming a smoker actually appears to protect White mothers against PTB (the same is not true for LBW), while it has the expected effect for Black and Hispanic mothers. These racial differences could reflect smoking as a proxy for other SES factors, the interaction of smoking with other stressors, differences in the number of cigarettes smoked, differences in smoking prevlance across ages (which we did not control for here), or other explanations.

In both the LBW and PTB regressions, the coefficients on late PNC initation and few PNC visits are only slightly larger in the OLS regression compared to the within-mother regression, implying that access to PNC is not simply a proxy for the long-run effects of SES. Whether the within-mother effects of these variables

reflect the effect of better PNC itself and/or the effect of increased access to resources more broadly, it is interesting to note that there are relatively short-term effects of such changes. We might expect the coefficient on many PNC visits, which reflects pregnancy complications rather than access to care, to be more similar in magnitude between the OLS and within regressions. Rather, we observe that the effect is one-third to one-half the size in the within-mother LBW and PTB models. We can infer that there may be genetic or slow-changing characteristics that differ between women who do and do not many prenatal care visits.

Not surprisingly, chronic hypertension has a much larger between effect than within effect for both LBW and PTD, reflecting the cumulative nature of the effects of high blood pressure and its correlates. Racial disparities in the effect of hypertension are upheld for both Black and Hispanic mothers in the PTB regression but only for Hispanic mothers in the LBW regression. The within-mother effect of pregnancy-related hypertension is relatively closer in size to the OLS effect, reflecting its more limited duration.

1.5 Summary and Discussion

Our first goal was to assess whether Texas mothers exhibited racial disparities in birth outcomes and age-outcome relationships similar to those found in other parts of the country. In general, we found this to be the case. Large overall LBW and PTB disparities between Black and White mothers in Texas are similar to those found in virtually every study examining Black-White differences in birth outcomes. We also find evidence of accelerated weathering among Black mothers, corresponding to the findings of most LBW weathering studies [61, 92, 95, 73] and some PTB studies [87, 66]. We did not find, as Geronimus (1996) did, that Black LBW rates increase monotonically over time, though the use of five-year age categories in that paper may obscure the precise age-LBW relationship [61].

Our results also support the "Hispanic paradox" in that we observed relatively small overall disparities between White and Hispanic mothers. Significantly, however, we observed early weathering among both U.S.- and foreign-born Hispanic mothers compared to Whites. While one study examined age-LBW relationships for U.S.- and foreign-born Mexican women and found them to be similar, with U.S.-born women showing higher overall rates of LBW [114], no previous study had compared these relationships in Hispanic and non-Hispanic populations.

Our second goal was to test the weathering hypothesis, which asserts that racial disparities in weathering are caused by differential prevalence of SES-related risk factors. We did this by including several such risk factors in our age-outcome regressions, which we found were differentially distributed across ages by race, and we found that their inclusion did reduce observed disparities in both LBW and PTB. This finding supports the weathering hypothesis and is consistent with previous research. That our group of confounding variables was far from complete provides support for the idea that exposure to low SES conditions may be more responsible for early weathering than we could observe.

Like other studies, we found that low neighborhood SES was correlated with negative birth outcomes, but we did not find that the age-outcome relationships varied much by neighborhood income level, except for foreign-born Hispanic women. It is perhaps misleading to talk about a neighborhood exerting an "independent" effect on health, since these neighborhood effects must get into the body somehow, and individual SES variables may be thought of as channels through which poor environments lead to bad health outcomes. With that disclaimer, we did find a small, independent effect of neighborhood SES, even after individual-level variables are controlled for. Still, this additional effect is small, and including neighborhood SES induced only small reductions in the observed effects of individual-level SES-related variables. We therefore relied on the full, non-geocoded sample of births for

our main analyses, using several individual-level variables as proxies for SES. This finding may be helpful to future researchers, since non-geocoded birth data are much easier to obtain and prevents the need to select on a variable that may bias results.

We found that the effect of most risk factor variables varied significantly by race, with Blacks generally experiencing more negative effects than Whites and foreign-born Hispanic mothers experiencing less negative effects. As discussed above, it is impossible to disentangle the various possible explanations for race-specific effects from these data, and it is likely that several are operating at the same time. Race-varying risk factor effects could be observed if the variables actually measure different conditions depending on race (e.g., an unexpectedly high number of prenatal care visits may signify more extreme pregnancy complications in minorities than in Whites). They could also reflect different meanings attached to the same conditions (e.g., not having a high school education may translate to a lower relative social position among White women than among foreign-born Hispanic women). The differences could reflect different medical responses to the same conditions (e.g., White women with hypertension may receive better medical care than Black women with hypertension) or interactions with measured or unmeasured risk factors (e.g., poor nutrition amplifies risks.) Finally, the effect of risk factors could appear to vary by race if the effects varied by age, which we find that they do, since the age-birth distribution varies across races. If this were primarily driving the results, however, we would expect Black and U.S.-born Hispanic effects to be more similar.

We found that the effects of many risk factors, including those most associated with low SES—low education, mother unmarried, mother smoking, late PNC initation—increase with age. This is consistent with what Rich-Edwards et al. found in Chicago [95] and with the weathering hypothesis in general; we expect the effects of deprivation to accumulate and increase in severity with age. We also find some evidence that risk-outcome relationships interact with age differently across races,

but these differential age interactions seem to drive weathering disparities less than the combination of disparities in exposure to risk with age and differences in birth timing.

Even after controlling for SES risk factors and including age and race interactions, we found an independent "effect" of being Black. In other words, our data are unable to explain the higher incidence of poor birth outcomes among Black women across ages. Indeed, we found that the coefficient on Black race either remained the same (PTB) or increased (LBW) after adding risk factor controls to the bivariate model; controlling for the average effects of education, marital status, smoking behavior, etc. could not explain Black-White differences. Our control variables are certainly far from comprehensive, but previous research has also failed to fully explain these disparities. Researchers in the field of health inequalities have theorized about a range of material and psychosocial explanations for the gap, including disparities in access to nutritious food, healthcare, or pollution exposure, different behaviors and habits related to health, and different stress levels related to socioeconomic status [104]. Much work remains to be done on the role of these various and interacting factors on health disparities.

Because the relationship between race and SES cannot be changed quickly, conclusions drawn from previous weathering research have been necessarily broad and sometimes deflating. Love et al. (2010), which finds that the SES of a mother's childhood neighborhood affects birth outcomes, concludes that we must change economic and social environments for Black women over their life course, perhaps over multiple generations [73]. Geronimus (1996) somewhat more narrowly suggests targeting clinical interventions to the needs of socioeconomically disadvantaged Black women in their 20s and early 30s, and writes that anti-tobacco interventions for White teen and older Black mothers in particular could greatly reduce differential maternal age patterns of LBW [61]. Using a within-mother regression, we find support for this

suggestion, since the effects of many of the risk factors indeed appear to have both cumulative and more short-term effects. This is good news, in that interventions in areas such as smoking, hypertension and prenatal care access could help diminish disparities in the short run as we work toward changing economic and social environments in the long run.

CHAPTER 2

ARE WE POISONING THE MOST VULNERABLE?: THE DISTRIBUTION OF TOXIC AIR POLLUTION ACROSS NEWBORNS IN TEXAS

2.1 Introduction

This essay examines the racial distribution of industrial toxic air pollution across infants in Texas. Previous environmental justice literature has generally examined the distribution of environmental hazards across race groups in the population at large, but we focus on infants for several reasons. First, infants are perhaps more vulnerable to the effects of air pollution than any other age group. Second, data on infants and their mothers come from birth records, which are available on an annual basis and allow for analysis of time series properties of pollution distribution. Finally, since infants cannot be held responsible for choosing to live near air pollution, we believe this focus may underline the injustice of inequitable pollution exposure. The availability of geocoded birth records from Texas predicated our focus on that state, but with large numbers of Non-Hispanic White, Non-Hispanic Black and Hispanic residents, Texas provides a good context in which to study patterns of environmental exposure within these groups.

To carry out the analysis, we spatially match birth records, using the mother's home address, to geographic microdata from the U.S. Environmental Protection Agency's Risk-Screening Environmental Indicators (RSEI) project and to U.S. census data. For each metropolitan statistical area (MSA) in Texas, we construct two descriptive measures of pollution distribution—one that compares a race group's share of toxic burden to their share in the population and another that measures the share

of the group exposed to relatively high levels of pollution—and compare the infant measures to those for the population at large. We find, not very surprisingly, that the infant and all-population measures are similar, implying that geocoded birth records could be used as a substitute for census data in subsequent environmental justice studies. We find a great deal of variation among cities in Texas, but in almost every one, Black and Hispanic infants bear more than their fair share of pollution.

After presenting MSA-level distribution statistics, we turn to an analysis of predictors of infant toxic exposure. Specifically, we ask whether race/ethnicity is an important predictor of exposure even when controlling for other mother and area characteristics. In line with previous research, we find that within Core Based Statistical Areas (CBSAs), both Black and Hispanic infants are disproportionately exposed to pollution, and the correlation is not eliminated when controlling for other socioeconomic variables. We also find that more segregated cities tend to have higher overall levels of toxic air pollution.

Finally, we exploit our ability to track mothers across births in our dataset to ask how race is associated with the change in pollution exposure between sibling births. There has been much speculation but relatively little research on the dynamics of pollution distribution, and unfortunately, our dataset, covering 1995-2003, is not long enough to address questions about discriminatory facility siting. Still, trends in more recent years may be helpful in forming policies to ameliorate inequities in exposure. Our analysis finds different patterns for mothers who moved between births and mothers who did not. Relative to their original location, White mothers who moved between births move to less polluted places than do Black and Hispanic mothers, but this racial differentiation does not seem to exist among mothers who did not move between births.

2.2 Background and Motivation

Since the publication of the seminal report *Toxic Wastes and Race in the United States* in 1987 [109], dozens of studies have documented the existence of systematic disparities in exposure to environmental hazards along race and/or class lines in the U.S., with minority race often predictive of disproportionate exposure even when controlling for socioeconomic status [74, 97]. Findings of environmental injustice are powerful because they appeal, explicitly or implicitly, to the ethical precept that everyone is entitled to a healthy environment. Indeed, several countries have established this right in their constitutions, and others have explicitly interpreted a constitutionally-given right to life as inclusive of the right to pollution-free air and water [17]. In the U.S., President Clinton signed an Executive Order in 1994 requiring each federal agency to "make achieving environmental justice a part of its mission" and current EPA administrator, Lisa Jackson, has explicitly prioritized working toward environmental justice [32, 55]. Yet despite these commitments, the right to a clean and healthy environment, in the U.S. and elsewhere, is frequently not upheld.

Boyce (2000) contrasts this "rights-based approach" to the provision of environmental goods to the "wealth-based approach," in which access to clean air and water is based on willingness and ability to pay, and concludes that the latter—while perhaps lacking "prescriptive appeal"—more closely approximates what often happens in practice [17]. Though not referred to as such in economics textbooks, Boyce's "wealth-based approach" is a close cousin of the neoclassical theory of distribution, which is presented as having both descriptive and prescriptive relevance. The theory is based on the key economic insight that tradeoffs are necessary in a world where resources are scarce, and perfectly-informed economic actors may trade clean air and water for other goods, depending on their preferences and their budget constraint. One's budget constraint, in turn, is determined by his or her marginal economic product, which itself is determined through a combination of genes, work ethic and life

choices. Living near environmental hazards, according to this model, is a tradeoff that some people make, given their incomes. Further, the fact the some people have much larger incomes than others is considered fair and even desirable, since it is thought to provide incentives for others to work harder.

To the extent that the neoclassical theory of distribution has "prescriptive appeal" as a way to sort people into more and less polluted areas, it would seem to break down when the actors in question are infants, who cannot be expected to make rational tradeoffs but still suffer the health consequences of environmental hazards. Indeed, economic theory does not expect children to make such decisions. Rather, it accounts for a child's health and welfare by including it in the parents' utility function, which is maximized in the usual way, subject to the parents' budget constraint. Thus, an infant's "right" to a clean and health environment is conditioned on her parents' ability to pay for it. Even those who generally agree with the prescriptions of economic theory, however, are likely to agree that infants should be born with equal opportunities, regardless of their parents' situations. This commitment to equal opportunities at birth is the premise of a recent report from the World Bank, an institution that has relied heavily on the prescriptions of standard economic theory in recent decades [42]. The report acknowledges that, though inequality among adults may be tolerable because it is perceived as fair, people view inequality arising from differences in opportunities as fundamentally unfair [42]. According to the report, "equality of opportunity seeks to level the playing field so that circumstances such as gender, ethnicity, birthplace, or family background, which are beyond the control of an individual, do not influence a person's life chances" [42, p.1]. After establishing this shared value, the authors use similar methods to the ones we use here to document the vast inequality of opportunity among children in Latin America, in order to argue for policies that would help level the playing field.

Documenting environmental injustice among infants may provide a more convincing and urgent call for change than documenting environmental inequities in the population at large. Disproportionate exposure to pollution among infants is disturbing not only because infants cannot choose to escape it, but because they are more susceptible than adults to its negative health effects. As Landrigan et al. write in an article on health effects of pesticide exposure, infants and children are uniquely vulnerable to environmental toxicants [70]. First, because the central nervous system is not fully developed until at least six months after birth, toxic chemicals may disrupt its development in fetuses and infants in a way not likely to affect older children or adults. Similarly, there may be critical windows of development during which exposure to toxics have particularly deleterious effects. Other reasons offered for the susceptibility of fetuses and infants have to do with their small size and young age. A given dose of pollution could be expected to have much larger effects in a seven-pound infant than in a 150-pound adult, for example, and because of their longer lifespans, infants simply have more years during which to develop chronic disease from early exposures.

Unfortunately, relatively little is known about how many types of environmental pollutants affect infants, due to data shortcomings, difficulty in separating effects of different exposures, and a range of other hindrances. The effects of maternal smoking on infant health provide uncontroversial evidence that environmental contaminants have the potential to impair health at birth. Results from a number of studies indicate that early exposures to other types of pollution can significantly damage the life prospects of developing fetuses and infants. Chay and Greenstone (2003) used recession-induced variation in total suspended particulates (TSPs) to show that a one unit decline in TSPs led to reductions in the incidence of low birthweight and between four and seven fewer infant deaths per 100,000 live births [30]. Rauh et al. (2006) found that children who were prenatally exposed to high levels of cholor-

pyrifos, a pesticide, were five times more likely to be developmentally delayed than those in a lower exposure group [93]. Reyes (2005) found that the phaseout of leaded gasoline in the U.S. led to a three to four percent decrease in low birthweight and infant mortality, not to mention the developmental delays with which lead is more famously associated. Currie and Schmeider (2009) found that a number of common industrial chemicals have deleterious effects for newborns, with toluene and cadmium, in particular, significantly raising the likelihood of low birthweight and infant mortality [39]. In another paper, Currie (2011) estimates that differences in exposure to releases of industrial toxic chemicals—the environmental hazard we focus on in this study—may explain six percent of the low birthweight gap between infants of White college educated mothers and infants of Black high school dropout mothers [36].

These findings show that pollution exposure can compromise the health of infants, but they do not explain the mechanism. A growing body of research in epigenetics may help fill that gap. Epigenetics is the study of heritable changes in appearance or gene expression caused by mechanisms other than changes in the underlying DNA sequence [90]. This research suggests that, rather than a relatively small number of genes determining our heritable physical and personality traits, the expression of certain traits results from an interplay between genes and the environment. The epigenome, literally "above the genome," can be thought of as a series of switches that turn parts of the genome on and off. Environmental factors can trigger these switches, and perhaps most impressively, changes in the arrangement of these switches can be passed from parents to children.[1]

Thus, environmental conditions, including pollution exposure, may play a larger role than previously thought in determining a child's health at birth. Health at birth, in turn, has been found to affect well-being later in life. Though it does not capture

[1]The "series of switches" analogy was borrowed from Currie (2011).

all aspects of fetal health, low birthweight has long been used as a summary measure, thanks to its wide availability and relatively accurate measurement. Low birthweight is associated with a higher risk of neonatal mortality, as well as higher adult incidences of heart disease and diabetes [101]. Presumably through its effects on health, low birthweight is also associated with negative socioeconomic outcomes. Even within sibling or twin pairs, children of lower birth weight have been shown to have worse outcomes in schooling attainment, test scores, use of disability programs, propensity to live in high income areas, and wages [36]. It may be, as Currie writes, that "poor health in childhood is an important mechanism for intergenerational transmission of education and economic status" [35, p. 88].

Given the potential for long-term negative effects of early exposure to air pollution, further research into these effects certainly seems warranted. Recent studies have rightfully focused on identifying the effect of individual pollutants on health, so as not to misattribute the effect of the pollutant to the effect of a correlate. Yet people are generally exposed to multiple types of air pollutants, and the cumulative effects of these mixtures may be multiplicative rather than additive [99]. Cumulative exposures to environmental conditions, defined more broadly, are also widespread. As Evans et al. (2002) note, "the poor are most likely to be exposed not only to the worst air quality, the most noise, the lowest-quality housing and schools, etc., but of particular consequence, also to lower-quality environments on a wide array of multiple dimensions" [58, p. 304]. Already high levels of biophysical and psychosocial stresses may result in marginal effects of pollution exposure that are worse in low SES populations. Consideration of cumulative effects magnifies the injustice of disproportionate toxic exposure among infants.

We have attempted to show why disproportionate pollution exposure among infants is a cause for concern. Next, we take a brief look at the industrial composition of toxic emissions in Texas, to help contextualize our analyses. Table 2.1 lists the top

Table 2.1. TRI toxicity-weighted pounds emitted by industries in Texas, Top two chemicals, 1995-2003.

Industry	Tox-wt lbs 1995-2003	% of total	Chemical 1	Tox-wt lbs 1995-2003	% of all-indust total	Chemical 2	Tox-wt lbs 1995-2003	% of all-indust total
Industrial Organic Chemicals	1456.21	66.98%	Diisocyanates	1004.20	46.19%	Diaminotoluene (mixed isomers)	325.02	14.95%
Synthetic Plastics Materials And Synthetic Resins	472.25	21.72%	1,2,3-Trichloropropane	429.01	19.73%	Titanium tetrachloride	6.69	0.31%
Petroleum Refining	47.35	2.18%	Sulfuric acid	16.73	0.77%	Chlorine	11.49	0.53%
Industrial Inorganic Chemicals	36.77	1.69%	Chlorine	13.89	0.64%	1,2,3-Trichloropropane	6.92	0.32%
Miscellaneous Wood Products	17.37	0.80%	Acrolein	13.90	0.64%	Diisocyanates	1.36	0.06%
Primary Smelting And Refining Of Nonferrous Metals	17.33	0.80%	Arsenic	4.63	0.21%	Lead	3.94	0.18%
Motor Vehicles And Motor Vehicle Equipment	12.97	0.60%	Diisocyanates	10.96	0.50%	Asbestos (friable)	1.57	0.07%
Miscellaneous Plastics Products	12.60	0.58%	Diisocyanates	9.53	0.44%	Toluenediisocyanate	2.07	0.10%
Hydraulic Cement	10.84	0.50%	Sulfuric acid	7.98	0.37%	Manganese	2.03	0.09%
Iron And Steel Foundries	10.60	0.49%	Diisocyanates	7.29	0.34%	Chromium	1.57	0.07%

Source: Author's calculations based on TRI public release data for years 1995-2003 and toxicity weights from RSEI version 2.15.
Notes: Figures represent billions of toxicity-weighted pounds released. Only emissions to the air from facilities located in Texas are included. Values exclude emissions from facilities belonging to four-digit SIC codes that were not required to report to the TRI during the whole sample period (1995-2003). Percentages represent percent of total toxicity-weighted pounds released by all industries in the sample.

ten industries by aggregate toxicity-weighted TRI emissions over the years 1995-2003, the sample period used in our analyses below, and the top two air toxics emitted by each of these industries.[2] From this table, we see that a relatively small group of industries are responsible for the lion's share of toxicity-weighted emissions, with the ten industries on this list emitting 96 percent of all toxicity-weighted pounds released during these years. Even more strikingly, the top two industries—industrial organic chemicals and synthetic plastics materials—are responsible for almost 89 percent. We should note that some facilities that currently report to the TRI that were not required to report before 1998, most prominently electric utilities, but also including some facilities in the petroleum and chemical industries, were not included in this analysis.[3]

Table 2.1 also shows the top two chemicals, by toxicity-weighted pounds, emitted by each industry over the 1995-2003 period. We see that diisocyanates, a chemical with a very high toxicity weight, occupies five of the 20 slots in the table, and releases of diisocyanates from the organic chemicals industry is responsible for almost half the total toxicity-weighted releases over the sample period. 1,2,3-Trichloropropane from the synthetics plastics industry and diaminotoluene from the organic chemicals industry account for about 20 and 15 percent of total toxicity-weighted releases, respectively. Also making appearances on this list, though with lower contributions, are a handful of more common substances, including sulfuric acid, arsenic, asbestos, chlorine, and acrolein, as well as heavy metals lead, manganese and chromium.

---

[2]Here, industries are based on three-digit SIC codes.

[3]Specifically, there are 7,982 Texas TRI-reporting facilities in the sample used to produce these tables; another 1,836 were required to report at some point between 1998-2003 but were not included here, because they belong to four-digit SIC classifications that did not report to the TRI before 1998. These 1,836 facilities include 569 that belong to the four-digit SIC code for "refuse systems," which may include incinerators, hazardous waste facilities, landfills, etc.; 558 in the "petroleum bulk stations and terminals" category (petroleum refining is a separate category that is included in our sample); 355 in the "chemicals and allied products, not elsewhere classified" category; and 322 electric utilities.

Table 2.2. TRI toxicity-weighted pounds emitted by industries in Texas, Top three cities, 1995-2003.

Industry	Tox-wt lbs 1995-2003	City 1	Tox-wt lbs 1995-2003	% of all-indust total	City 2	Tox-wt lbs 1995-2003	% of all-indust total	City 3	Tox-wt lbs 1995-2003	% of all-indust total
Industrial Organic Chemicals	1456.21	Baytown	644.08	29.63%	La Porte	433.70	19.95%	Pasadena	273.34	12.57%
Synthetic Plastics Materials And Synthetic Resins	472.25	Deer Park	440.58	20.26%	Houston	8.02	0.37%	Wadsworth	6.65	0.31%
Petroleum Refining	47.35	Pasadena	7.74	0.36%	Corpus Christi	6.73	0.31%	Texas City	6.70	0.31%
Industrial Inorganic Chemicals	36.77	Freeport	28.39	1.31%	Corpus Christi	4.32	0.20%	La Porte	1.31	0.06%
Miscellaneous Wood Products	17.37	Diboll	13.95	0.64%	Jasper	1.48	0.07%	Nacogdoches	0.67	0.03%
Primary Smelting And Refining Of Nonferrous Metals	17.33	Amarillo	8.59	0.40%	Rockdale	7.15	0.33%	El Paso	1.59	0.07%
Motor Vehicles And Motor Vehicle Equipment	12.97	Sealy	10.61	0.49%	Dallas	1.52	0.07%	Denton	0.47	0.02%
Miscellaneous Plastics Products	12.60	New Braunfels	7.49	0.34%	Terrell	1.59	0.07%	Dallas	1.06	0.05%
Hydraulic Cement	10.84	Midlothian	8.46	0.39%	New Braunfels	1.32	0.06%	San Antonio	0.44	0.02%
Iron And Steel Foundries	10.60	Tyler	5.11	0.23%	Houston	3.01	0.14%	Temple	0.80	0.04%

Source: Author's calculations based on TRI public release data for years 1995-2003 and toxicity weights from RSEI version 2.15.
Notes: Figures represent billions of toxicity-weighted pounds released. Only emissions to the air from facilities located in Texas are included. Values exclude emissions from facilities belonging to four-digit SIC codes that were not required to report to the TRI during the whole sample period (1995-2003). Percentages represent percent of total toxicity-weighted pounds released by all industries in the sample.

Table 2.2 shows, for each of the top ten industries, the three cities with the highest total toxicity-weighted releases between 1995-2003. Chemical facilities in Baytown, LaPorte and Pasadena together account for 62 percent of total toxicity-weighted air emissions during the period, and emissions from the plastics industry in Deer Park account for another 20 percent. These four cities lie just to the east of Houston, which itself appears twice in the table, near the Gulf of Mexico. Other Gulf cities in Table 2.2 are Freeport, Wadsworth and Corpus Christi. With the exception of Amarillo, in the northern part of the state, and El Paso, just across the border from Juarez, all the cities on this list lie on or east of interstate I-35, which runs through San Antonio and Dallas.

Texas clearly has some very polluted industrial areas, as well as rural areas with little industry, but it is not clear from these tables whether Black and Hispanic residents in Texas are disproportionately exposed. Previous environmental justice research suggests that they are. In a book of case studies, journalist Steve Lerner documents the environmental justice struggles of residents living near oil refineries in two Texas cities, Port Arthur and Corpus Christi. These residents, disproportionately minority and low-income, suffered from a range of poor health conditions but often could not afford to move [71]. In a statistical study of industrial polluters in Texas, Wolverton (2009) found that the non-White share of a census tract was positively correlated with the presence of a polluting facility in 1990 [117]. Currie (2011), in a study of infants in Texas and four other states, finds that, within zip codes, Black and Hispanic infants are more likely to be born near polluting factories and Black infants are more likely to be born near Superfund sites [36].

Our study, focused on industrial toxic air pollution exposure of infants in Texas, extends previous environmental justice research in a number of ways. No previous study has compared the race distribution of pollution across infants to its distribution across races in the population at large. Doing so is interesting in its own right, since

infants are uniquely susceptible to pollution exposure, but if the distribution is similar across the two populations, it may support further environmental justice research using geocoded birth records rather than census data. Birth records are potentially useful to environmental justice studies for at least two reasons. First, they allow for precise location of a person in what Morello-Frosch and Lopez call the environmental "riskscape" [83]. This method, known as "point interpolation," precludes the need for "areal interpolation" of demographic characteristics from census areas onto the areas potentially exposed to a hazard, since these two geographic units generally have different boundaries [29]. Second, unlike census data, birth records are available on an annual basis and can therefore be used to study time series properties of pollution distribution.

Currie (2011) also uses geocoded birth records and measures the distance between infants and sources of industrial air pollution. Living near a factory is a reasonable but rough indication of pollution exposure. Using the Risk-Screening Environmental Indicators (RSEI) geographic microdata allows us to get a more accurate idea of actual exposure (see full RSEI description on p. 85). The RSEI model takes into account the fate and transport of chemicals in order to model their concentrations in the area surrounding each polluting facility. In addition, because chemicals vary in their degree of toxicity, the model weights more dangerous chemicals more heavily in producing a summary pollution measure. In an assessment of various methods of spatially matching people to pollution, Chakraborty and Maantay (2011) call the RSEI data "particularly suitable for environmental justice research" [29, p. 124].

In our regression analyses, where we ask whether minority race/ethnicity predisposes newborns to being born in a more polluted environment, we also include an indicator of city-level segregation in some specifications. Segregation is closely related to environmental racism since, if there were not disproportionate numbers of minorities in some places compared to others within a given geograhic unit of analy-

sis, disproportionate exposures would not be possible. Even so, explicit connections between degree of segregation and pollution exposure in the environmental justice literature are rare. Morello-Frosch and Jesdale (2006), in one of the few studies to make this link, finds that estimated cancer risks associated with ambient air toxics are highest in neighborhoods located in metropolitan areas that are highly segregated, as measured by the multigroup dissimilarity index [82]. The authors further find that these higher exposures affect populations across racial lines, not only the segregated minority. While not studying segregation specifically, Ash et al. (2010) find that cities with more disproportionate minority pollution exposure also have higher levels of pollution overall, indicating that "environmental justice is good for White folks" [8].

As segregation deserves more attention in the environmental justice literature, so too do questions about the dynamics of pollution distribution. In particular, the issue of whether environmental injustices arose because pollution followed minorities or because minorities followed pollution (sometimes called the "move-in versus siting" question) has been the subject of much speculation but relatively little research (see [86, 12, 88, 117, 10]). Unfortunately, the nature and short time period of our data do not allow us to determine whether facilities in Texas were sited in a discriminatory manner.[4] What we can do, thanks to our ability to track multiple births to the same mother, is ask whether race has been a significant predictor of the change in

[4]Wolverton (2009), studying manufacturing plants sited in Texas after 1975, concludes that though minorities are currently disproportionately represented in tracts containing factories, race is no longer significant when plant location is matched to socioeconomic characteristics at the time of siting. It is true that the coefficient on "percent minority" in the regression that controls only for area socioeconomic characteristics and voting behavior (Table 3, Column 4) is small and insignificant. In the two models preferred by the author, however, in which input costs are also controlled for (Table 3, Columns 5-6), the coefficients on percent minority are as large as in the contemporary regression and the standard errors indicate that these coefficients are approaching statistical significance ($\beta = 0.52$ and $SE = 0.33$ in Model 5; $\beta = 0.0.44$ and $SE = 0.28$ in Model 6) [117, p. 22]. Also, like other studies that have addressed the move-in versus siting question ([86, 12]), this study looks only at the demographic characteristics of tracts containing factories, which are often located near a tract border, and ignores the characteristics of adjacent neighborhoods.

pollution exposure between births over the 1995-2003 period. Further, we can look separately at mothers who moved between births and mothers who did not, to see whether different patterns exist for these two groups.

As Pastor et al. (2001) note, appropriate policy responses to environmental injustice can be informed by whether "move-in" or "siting" is more to blame, and both these dynamics may operate to different extents in different areas. If disproportionate exposure is due primarily to discriminatory siting, zoning and permitting processes should be revised to prevent such discrimination. If minorities are moving to more polluted areas, the authors suggest that appropriate policy would seek to provide greater access to data on the health risks of such pollution and enforce existing statutes that prohibit the steering of minority house-seekers to particular neighborhoods [88]. Unfortunately, the move-in versus siting debate has seemed to be as much about supporting ideological commitments as about finding appropriate ways to reduce environmental injustice. A finding of poor or minority "move-in" has confirmed, for some, that environmental injustice has not been imposed on its victims but rather arises from households maximizing their utility subject to budget constraints. Setting aside the flaws in the line of reasoning that equates this sort of "choice" with fairness, it is worth reiterating that a growing fetus assaulted with toxic chemicals is beginning life with a deficit, however those conditions arose. More research about pollution and population dynamics should be conducted with an eye toward finding appropriate, place-specific solutions.

2.3 Data and Methods

2.3.1 Methods

To assess the distribution of industrial toxic air pollution across all people and newborns in Texas, we use two different measures that capture somewhat different aspects of environmental justice. The first is what we call the discrepancy score,

which measures the difference between the share of the total burden of industrial air toxic releases borne by a minority group and the share of that minority group in the area. Because the two largest minority groups in Texas are Black and Hispanic, and because we expect these groups may have different exposure patterns, we calculate Black and Hispanic discrepancy scores separately, both for the state as a whole and for each Metropolitan Statistical Area (MSA). A positive Hispanic discrepancy score, for example, would indicate that the Hispanic population in the specified area is exposed to pollution disproportionate to their share in the population. It is worth noting that the discrepancy score is a purely distributional measure; that is, it takes the size of the "pollution pie" in an area as given and asks whether the pie is divided equitably. A city with very low overall pollution could still have a large discrepancy if that pollution were divided very inequitably.

We first find the "all resident" Black and Hispanic discrepancy scores, defined and calculated analogously to the *Minority Discrepancy* measure in Ash et al. (2009) and Ash et al. (2010):

$$AllDiscrep_{\overline{m}j} = \frac{\sum_{i=1}^{n} \gamma_{\overline{m}ij} \times Tox_{ij}}{\sum_{i=1}^{n} \delta_{ij} \times Tox_{ij}} - \frac{Pop_{\overline{m}j}}{Pop_j} \tag{2.1}$$

where $\overline{m}$ indexes the minority group for which the discrepancy measure is calculated, j indexes the MSA (or the state of Texas) and i indexes the Census tract within the MSA (state). Tox_{ij} is the estimate of the toxicity-weighted concentration of chemicals in tract i (from RSEI data, described below), $Pop_{\overline{m}j}$ is the population of minority group $\overline{m}$ in MSA j, and Pop_j is total MSA population. Instead of simply weighting the Tox variables in the first term by $Pop_{\overline{m}j}$ in the numerator and Pop_j in the denominator, we first weight age and sex groups within the population by inhalation exposure factors (IEFs) used in the RSEI model to reflect biological differences in in-

halation uptake of pollutants.[5] Thus, $\gamma_{\overline{m}ij}$ and δ_{ij} reflect the IEF-weighted population in tract i of the minority group and of all residents, respectively.

We are interested in comparing the distribution of pollution across newborns in Texas to the distribution across all residents. To do so, we calculate discrepancy measures analogous to the one described above, but including only children born during the year of analysis (here, 1999):

$$NewbornDiscrep_{\overline{m}j} = \frac{\sum_{i=1}^{n} Births_{\overline{m}ij} \times Tox_{ij}}{\sum_{i=1}^{n} Births_{ij} \times Tox_{ij}} - \frac{Births_{\overline{m}j}}{Births_j} \tag{2.2}$$

where most variables and indices are defined as in (2.1). $Births_{\overline{m}ij}$ represents births to minority group $\overline{m}$ in tract i (in year y, which is not indexed here) and $Births_{ij}$ represents all births in tract i. We do not account for IEFs in this equation, since the IEF recommended for male and female infants is the same [51].

The first two discrepancy measures assign tract-level pollution concentration measures to 1) all residents and 2) newborns. We are also able to match birth records to the pollution concentration estimate from the square kilometer grid cell in which the mother lived at the time of birth. Since this is the pollution measure we use in subsequent analyses, we are also interested to see how using this more geographically-precise exposure measure affects the discrepancy score. Equation (2.2) can also be used to describe the cell-level discrepancy measure, but in this case, i indexes the grid cell instead of the tract, and because some cells do not lie fully within MSAs (as tracts do), births outside the MSA are not included in the weights. Since we have grid cell pollution measures for multiple years, we also calculate and graphically display the state-level Black and Hispanic newborn discrepancy measures for years 1995-2003.

---

[5]The inhalation exposure factors used are as follows: male age 0-17 (0.341), male age 18-44 (0.209), male age 45-64 (0.194), male age 65 and up (0.174), female age 0-17 (0.310), female age 18-44 (0.186), female age 45-64 (0.165), female age 65 and up (0.153). For more information of the derivation of the IEFs, see Technical Appendix C for RSEI Version 2.1.5 [51].

Since our first measure is purely distributional, we also employ a measure that accounts for absolute exposure, which may be more relevant to health concerns. Specifically, we look at the share of people (newborns) of each race group that were exposed to industrial toxic air pollution above a given threshold. Since our RSEI exposure measure represents the aggregation of chemicals with different biological effects, many of which are not well understood, it is difficult to define a threshhold that we can say with certainty is associated with negative health effects for infants. We arbitrarily choose the 1999 U.S. 75 percentile level to represent "high" exposure, though we also display the threshold measures using the 50 percentile and 90 percentile pollution levels as cutoffs in appendix Tables A.4-A.6. As with the discrepancy measures, we calculate this measure for 1) all residents matched to tract-level exposure estimates, 2) newborns matched to tract-level exposure estimates, and 3) newborns matched to square kilometer grid cell exposure estimates. Using the third sample, we also plot this distributional measure for years 1995-2003 to show changes over time.

After documenting the distribution of pollution across infants in Texas, we turn to an analysis of the predictors of newborn pollution exposure. We start from the premise that no fetus or infant should be exposed to disproportionately high levels of pollution. We ask whether characteristics of a newborn's mother, especially her race, and characteristics of her neighborhood and city are significantly correlated with pollution exposure. Our basic model is thus:

$$Tox_{ijct} = \rho_j + \beta X_{ijt} + \gamma Z_k + \omega seg_l + Y_t + \epsilon_{ijcklt} \tag{2.3}$$

where i indexes the infant, j indexes the mother and t indexes the year of birth. The indices c, k and l index the square kilometer grid cell, census tract and core based statistical area (CBSA, see p. 87 for description) corresponding to the mother's home address. Tox_{ijct} is the toxicity-weighted RSEI pollution concentration estimate for grid cell c in year t, and this represents the pollution exposure of infant

i. The vector ρ_j includes race/ethnicity dummy variables, which signify whether the mother identifies as Non-Hispanic White, Non-Hispanic Black or Hispanic. The vector X_{ijt} contains variables representing the mother's socioeconomic status, here age, educational attainment and marital status. The vector Z_k contains tract-level socioeconomic characteristics thought to be correlated with pollution levels, including median household income and its square, population density, share of adults without a high school degree, share of houses that are vacant, share of houses that are owner occupied and share of residents employed in manufacturing. The vector seg_l contains Black and Hispanic segregation measures for CBSA l. Finally, the vector Y_t includes year dummy variables to capture secular variation in pollution levels.

Ash and Fetter (2004) suggest that "environmental justice studies should account for variations in base levels of pollution in order to avoid collapsing variation within cities and variation among cities into a single coefficient" [9, p. 460]. While we do pool all Texas births in the estimation of Model (2.3), in our second specification, we include fixed effects for the 68 CBSAs in Texas in order to look exclusively at variation within cities:

$$Tox_{ijct} = \rho_j + \beta X_{ijt} + \gamma Z_k + Y_t + \phi_l + \epsilon_{ijcklt} \tag{2.4}$$

where ϕ_l is a CBSA fixed effect. Since segregation is a CBSA-level variable, it is not included in this model.

Because there is some variation in pollution exposure estimates within tracts, we can also ask whether mother characteristics, including race/ethnicity, are predictive of exposure within neighborhods:

$$Tox_{ijct} = \rho_j + \beta X_{ijt} + Y_t + \tau_k + \epsilon_{ijckt} \tag{2.5}$$

where τ_k is a tract fixed effect. It should be noted here that, while the RSEI data does provide fine geographic detail, the pollution concentrations in our dataset are based

on self-reported firm-level data that has been run through a dispersion model. While very useful for assessing broad pollution distribution patterns, the accuracy as we "zoom in" may not be as good, due to possible inaccuracies in the reported data and modeling assumptions (described below). With these qualifiers, we propose the tract fixed effect model as an initial exploration into the within-neighborhood patterns of pollution exposure.

Though the models presented above are similar to many in the environmental justice literature, with a measure of pollution exposure regressed on demographic characteristics, there are some small differences. Where other studies typically use a geographic area (e.g., Census tract or circular buffer around a pollution source) as the unit of analysis, we use individual infants, so neighborhoods are represented in proportion to how many births occured there over the sample period. Further, as we do not have appropriate data to do so, we do not theorize about whether the independent variables, including characteristics of individual mothers, caused the contemporaneous levels of pollution in their neighborhood. Rather, we frame our analysis as an investigation of the predictors of newborn toxic exposure.

In our second set of regression analyses, we attempt to address a variation of the move-in versus siting question using a novel method. In our version, we ask whether the difference in birth-year pollution exposure between siblings depends on race/ethnicity, and we stratify the sample into births to mothers who moved between births and mothers who did not move. With these two samples, we ask both "Do minorities follow the pollution?" and "Does the pollution follow minorities?" As alluded to above, our sample does not cover a long enough period to shed a great deal of light on this issue, but it does allow us to look for evidence of trends in recent years.

To address these questions, we use the following model:

$$DiffTox_{jt} = \rho_j + \beta X_{jt} + \gamma Z_k + \delta Diff Z_{jt} + Y_t + Diff Y_{jt} + \phi_l + \epsilon_{jklt} \qquad (2.6)$$

where j indexes the mother of the pair of sibling births that constitute each observation, t indexes the year of birth of the older sibling, and k and l are defined as above. $DiffTox_{jt}$ represents the difference in birth-year pollution exposure between the older sibling and the younger sibling. The variables ρ_j, Z_k and Y_t are defined as in Equation (2.3). The vector X_{jt} still represents mother characteristics, though variables indicating the change (or lack thereof) in education and marital status between the two births are also included. In the analysis using the sample of mothers who moved between births, the vector $DiffZ_{jt}$ is included to capture the difference in tract characteristics between the old and new neighborhoods. Vector $DiffY_{jt}$ captures the number of years between births, since we would expect the magnitude of the change in pollution to be correlated with the amount of time that passed between births. Finally, we include CBSA fixed effects, ϕ_l, to focus on patterns within CBSAs.[6]

We estimate separate samples for mothers who moved between births and mothers who did not, on the basis that such mobility may be correlated with increasing access to resources, though we acknowledge that reasons for moving may also vary systematically by race. It is worth noting here that the sample of non-moving mothers utilizes only temporal variation in the pollution measures, and in a separate analysis of year-to-year changes in TRI reporting across the U.S., we found that around 20 percent of facilities reporting to the TRI in two subsequent years report no change in release. Since it is unlikely that these facilities literally had no change in releases between reporting years, this finding calls into question the accuracy of the reported

[6]In order to include CBSA fixed effects, we exclude a relatively small number of mothers who moved to a new CBSA between births.

changes, though we cannot infer whether those reporting no change are systematically under- or overreporting.

We cluster standard errors at the CBSA level in all regressions. Doing so helps account for correlation between observations in regressions with CBSA-level segregation measures or fixed effects, and it should correct for some spatial autocorrelation in the other regressions as well. Spatial autocorrelation describes the correlation between observations that are geographically close to one another, since observations from nearby locations are often more similar than what could be expected on a random basis, and it violates the uncorrelated errors assumption of OLS [28].

2.3.2 Data

Our birth data come from Texas birth records for the years 1995-2003 and were provided by the Texas Center for Health Statistics. From these records, we use information on the child's year of birth and the mother's race, age, educational attainment, marital status and residence at the time of birth.

Using maternal residence information, we spatially match birth records to toxic exposure data from the geographic microdata of the U.S. EPA's Risk-Screening Environmental Indicators (RSEI) project. The RSEI data contain estimates of location-specific exposure to toxic air pollutants emitted by industrial plants across the United States [50]. RSEI uses information on annual releases of more than 600 chemicals from more than 20,000 facilities, reported in the Toxics Release Inventory (TRI). The TRI was created at the direction of the Congress under the Emergency Planning and Community Right-to-Know Act (EPCRA), passed in 1986 in response to the disastrous Bhopal chemical plant explosion. EPCRA requires industrial facilities to submit annual data to EPA on deliberate and accidental releases of toxic chemicals into air, surface water, and the ground and on transfers to offsite facilities.

To make the TRI data more meaningful, the RSEI model estimates local concentrations by incorporating information on the fate and transport of releases. Fate and transport information comes from a plume model that accounts for chemical decay rates, stack heights, exit-gas velocities, average temperature and prevailing winds. For each air release (each facility x chemical combination), RSEI estimates the concentration in each square kilometer of a 101-km by 101-km grid centered on the releasing facility.

Although all TRI chemicals are toxic, their human health hazards vary widely. RSEI incorporates data on relative toxicity to construct a measure of exposure that is additive across chemicals. Toxicity here refers to chronic human health effects from long-term exposure, including cancer and non-cancer effects such as developmental toxicity, reproductive toxicity, and neurotoxicity. The toxicity weights are based on a peer-reviewed methodology, taking into account the single most sensitive chronic human health endpoint (cancer or non-cancer).

To construct our toxicity measure, the RSEI toxicity-weighted concentrations are added across chemicals and facilities to characterize the total exposure to industrial air toxics in each square kilometer grid cell in which a pregnant mother resided. RSEI data are superior to other exposure measures due to some of the characteristics just described: they capture a wide range of chemicals from the largest industrial emitters, they account for differences in toxicity among chemicals, and they circumvent the "how near is near?" question by modeling concentrations in each grid cell. Still, it should be noted that the data do not capture emissions from mobile or small point sources, and the concentration estimates are only as good as the underlying self-reported emissions data and the fate and transport model.

The birth record and toxic exposure data are also spatially matched to tract-level characteristics data and CBSA-level segregation data derived from the 2000 U.S. census. In our analysis we occasionally refer somewhat imprecisely to tracts as

neighborhoods and CBSAs as cities. Census tracts generally contain between 2,500 and 8,000 residents and are designed, at least initially, to be roughly homogeneous with respect to population characteristics, economic status, and living conditions [111]. The term Core Based Statistical Area (CBSA) came into use in the 2000 census and refers collectively to Metropolitan Statistical Areas, which must have at least one urbanized area of 50,000 or more residents, and Micropolitan Statistical Areas, which must have at least one urban cluster of at least 10,000 but less than 50,000 residents. CBSAs are comprised of whole counties, and tracts do not cross county boundaries. The temporal variation in our birth records and RSEI measures cannot be matched in the census data, but since our data period (1995-2003) is never more than five years from the census year, and since neighborhood characteristics have been known to change slowly, we feel confident that the census data closely approximates neighborhood conditions in non-2000 birth years.

Our main sample includes Texas births from 1995-2003, limited to these years because the list of chemicals tracked in RSEI doubled in 1994, making the composite measure more comprehensive after that year, and birth records were available only through 2003. From this original dataset of 3,205,504 observations, we drop 416,332 that are not geocoded, in order to merge them with pollution and census data. We further exclude 85,303 observations outside of CBSAs, both because we might expect different relationships between socioeconomic characteristics and pollution in more rural areas and to facilitate the analyses with segregation and CBSA fixed effects. Finally, we include only non-Hispanic White, non-Hispanic Black and Hispanic births, because other race-ethnicity groups in Texas have relatively small numbers of births. Thus, the final sample used to estimate Models (2.3)-(2.5) includes 2,596,691 observations.

We refine the sample further to estimate Model (2.6). Because our unit of observation in this analysis represents a change in mother and neighborhood characteristics

between two sibling births, we take from the initial sample the set of births with at least one sibling in the sample. Though we had no reason to exclude multiple births in the initial sample, we do so here, since we are interested in observing changes over time between siblings who are born in different time periods. Also, because race is the independent variable we are most interested in, we exclude births from mothers whose reported race changed between births. Finally, in order to focus on variation within cities, we exclude siblings who were born in different CBSAs. We sort these 1,031,026 births by mother and by year of birth and then first difference the RSEI pollution variable, so that our final dataset contains 528,356 observations representing changes between two births to the same mother.

The distribution of our RSEI toxic concentration variable has a long right tail, so we use a logged version as the dependent variable when estimating Models (2.3)-(2.5) to better approximate the OLS normality assumptions. Since some births are matched to an initial toxic exposure measure of 0, we add 1 before logging so that we can include these observations. For Model (2.6), the dependent variable is the difference in this logged exposure measure between two sequential (in the dataset) siblings.

Our mother socioeconomic status variables, obtained from birth record data, include variables for race/ethnicity, age, educational attainment and marital status. The race variables are indicators for whether the mother is non-Hispanic White, non-Hispanic Black or Hispanic. For Models (2.3)-(2.5), the age variable indicates whether the mother is in her teens, twenties (omitted), thirties or forties or above at the time of birth. The education variable indicates whether the mother did not complete high school (no HS), completed high school (HS, omitted), or completed four-year college (BA). The marital status variable is a dummy that is set equal to 1 if the mother is unmarried. Model (2.6) also contains age, education and marital status variables, though the definitions change somewhat. The age variables indicate age category at

the time of the older sibling's birth. The education variable in this model takes on one of six levels, depending on the mother's educational attainment at the births of the younger and older siblings. The levels are 1) no HS $\rightarrow$ no HS, 2) no HS $\rightarrow$ HS, 3) no HS $\rightarrow$ BA, 4) HS $\rightarrow$ HS (omitted), 5) HS $\rightarrow$ BA, and 6) BA $\rightarrow$ BA. A variable accounting for possible changes in marital status can take on one of four possible levels: 1) unmarried $\rightarrow$ unmarried, 2) unmarried $\rightarrow$ married, 3) married $\rightarrow$ unmarried, and 4) married $\rightarrow$ married (omitted). From the birth records, we also use controls for year of birth in Models (2.3)-(2.5), and year of later birth and number of years since earlier birth in Model (2.6).

The neighborhood characteristics data, taken from the 2000 census, include variables commonly used in environmental justice studies to proxy neighborhood socioeconomic position, political power and other factors likely to be correlated with pollution levels. Our main socioeconomic status variable is tract median household income. We also include its square, since other authors have found an inverse-U relationship between income and pollution [12, 19]. According to the theory, in poorer areas, pollution reflects higher levels of economic activity and increases with income, but after a point, residents have the economic and political resources to fight or move away from the pollution. Following previous studies, we also include population density, share of adults age 25 and up without a high school degree (% no HS), share of vacant housing (% vacant), share of owner-occupied housing (% owner occ), and share of people employed in manufacturing (% mfg empl). For the sample of non-moving mothers in Model (2.6), we include the same set of neighborhood variables. When using the sample of mothers who moved, these variables indicate the neighborhood characteristics at the time of the older sibling's birth, and we also include variables representing the change in these values between the two births.

Finally, we derive three segregation measures using census data. As Massey and Denton (1988) show, cities can be segregated in different ways, and we are interested in

exploring how various dimensions of segregation may be correlated with pollution [77]. The most widely-used measure of segregation is the dissimilarity index, a measure of evenness, or the degree to which the proportion of a particular racial or ethnic group living in neighborhoods (here, census tracts) approximates that group's relative share of an entire city (here, a CBSA). The dissimilarity index ranges from 0 to 1, and conceptually, it represents the proportion of Black (Hispanic) residents that would have to change location to achieve an even distribution. We also use a measure of exposure, the isolation index, which measures the extent to which a particular minority group is likely to have contact with members of the same group. This index also varies from 0 to 1 and can be interpreted as the probability that a randomly drawn member of the Black (Hispanic) population shares a census tract with another Black (Hispanic) person. While evenness does not depend on the relative sizes of race groups in the city—it is a purely distributive measure—isolation does, since minorities have a greater chance of living in a high-minority neighborhood in a city where there are a relatively high number of minorities overall. We also use a measure of residential clustering, the index of spatial proximity, which assesses the extent to which Black (Hispanic) census tracts are continguous, forming "a single large ethnic or racial enclave," versus being more evenly spread throughout a CBSA [77, p. 293]. The index equals 1 if there is no differential clustering between Blacks (Hispanics) and Whites and is greater than 1 when members of each group live nearer one another than each other. We show formulas for calculating these segregation measures in an appendix on p. 188.

Table 2.3. RSEI scores and demographic summary statistics for Texas MSAs, 1999-2000.

Metropolitan Area	RSEI Score Med Resident	All Residents (Apr 2000) Pop	White Share	Black Share	Hispanic Share	Births (1999) Pop	White Share	Black Share	Hispanic Share
Texas (all)	46.40	20851820	52.4	11.3	32.0	310638	39.6	11.9	44.7
Abilene	23.64	160245	73.4	6.5	17.1	2269	63.4	7.4	26.3
Amarillo	80.68	226522	71.9	5.3	19.1	3322	62.6	5.2	29.7
Austin-Round Rock	30.71	1249763	60.7	7.6	26.2	18745	47.9	8.4	39.4
Beaumont-Port Arthur	917.56	385090	64.0	24.6	8.0	4845	54.4	30.3	11.5
Brownsville-Harlingen	6.48	335227	14.5	0.3	84.5	6518	5.5	0.2	93.6
College Station-Bryan	6.26	184885	65.8	12.1	17.3	2211	50.9	14.4	29.6
Corpus Christi	109.37	403280	41.0	3.7	52.7	6089	29.0	3.3	65.7
Dallas-Fort Worth-Arlington	34.49	5161544	59.0	13.6	21.6	84946	47.6	14.3	33.0
El Paso	579.11	679622	17.0	2.7	78.3	12846	12.0	2.3	84.7
Houston-Sugar Land-Baytown	273.38	4715407	48.2	16.6	28.7	76878	36.5	16.8	41.6
Killeen-Temple-Fort Hood	9.58	330714	59.2	18.9	15.7	5555	54.9	22.6	18.4
Laredo	23.32	193117	4.8	0.2	94.4	5036	1.3	0.1	98.3
Longview	263.73	194042	72.4	18.0	7.8	2024	60.8	22.6	15.1
Lubbock	3.84	249700	62.0	7.5	28.0	3507	48.5	9.2	40.4
McAllen-Edinburg-Mission	0.95	569463	10.4	0.4	88.4	7134	7.1	0.2	91.6
Midland	2.70	116009	61.9	7.0	28.9	1821	47.4	7.8	43.1
Odessa	38.66	121123	51.1	4.3	42.4	2193	34.6	4.2	60.1
San Angelo	0.30	105781	63.2	4.2	30.5	1518	50.6	4.2	43.9
San Antonio	36.79	1711703	40.6	5.9	50.4	26403	29.8	5.3	62.3
Sherman-Denison	3.08	110595	84.3	5.8	6.4	1217	75.2	9.6	13.1
Tyler	60.71	174706	68.0	19.0	11.1	2471	56.7	20.6	21.0
Victoria	120.57	111663	53.2	5.3	39.2	1555	38.3	4.3	56.6
Waco	23.20	213517	64.7	15.0	17.9	2945	50.5	19.2	29.2
Wichita Falls	42.56	151524	75.9	8.6	11.0	1899	70.3	10.3	16.9

Source: Author's calculations based on data from the 2000 U.S. Census and 1999 birth data from the Texas Center for Health Statistics merged with tract-level RSEI exposure measures for 1999.

2.4 Results

2.4.1 Distribution of pollution across newborns in Texas

Table 2.3 presents demographic and toxic pollution summary statistics for the state of Texas and its 24 MSAs at the time of the 2000 U.S. Census.[7] The first column to the right of the MSA names gives the RSEI concentration for the median resident for the area indicated. The population-weighted median score for Texas in 1999, at 46.4, was somewhat lower than the U.S. median score of 77.7. The MSA scores show that there is much variation within the state, however, with very high median scores in some highly industrial MSAs, including the areas containing Beaumont, El Paso, Houston and Longview.

The next column group presents population counts for MSAs in Texas, along with shares of White, Black and Hispanic residents, documenting the wide variation in racial composition of Texas cities. In the four MSAs with median RSEI scores above 200, three have a higher share of Blacks than in the state as a whole: Beaumont, Houston and Longview. The fourth, El Paso, has a much higher proportion of Hispanic residents than the state as a whole. We do not, however, see a clear correlation between high Hispanic share and high RSEI score. Border cities with large Hispanic populations, including Brownsville, Laredo and McAllen, have relatively low toxic scores associated with less heavy industry.

Since newborns are the main subject of later analyses, the third column group shows birth counts for the year 1999. These counts do not include 43,387 recorded 1999 births that could not be geocoded (about 12 percent of all 1999 births), since these births could not be matched to pollution scores. We see that White mothers have the lowest birth rate of the three groups and Hispanic mothers have the highest.

[7]Texarkana, TX-Texarkana, AR was not included in this analysis since we do not have all birth records from this MSA.

Table 2.4. Black and Hispanic discrepancies in distribution of air toxics, all residents and newborns, Texas MSAs, 1999.

Metropolitan Area	Pop (Apr 2000)	RSEI Score Med Resident	Black Discrepancy All Tract	Black Discrepancy Rank in US	Black Discrepancy Newborn Tract	Black Discrepancy Newborn km^2	Hispanic Discrepancy All Tract	Hispanic Discrepancy Rank in US	Hispanic Discrepancy Newborn Tract	Hispanic Discrepancy Newborn km^2
Texas (all)	20851820	46.40	0.8	33	-0.9	2.2	12.1	6	11.8	5.6
Abilene	160245	23.64	2.3	148	2.3	6.1	10.3	27	11.6	3.9
Amarillo	226522	80.68	0.7	214	0.7	0.5	0.7	177	-1.8	-1.2
Austin-Round Rock	1249763	30.71	1.8	165	1.5	1.5	3.2	81	3.5	2.9
Beaumont-Port Arthur	385090	917.56	-2.2	348	-3.7	-0.1	1.4	132	0.6	1.3
Brownsville-Harlingen	335227	6.48	0.1	283	0.0	-0.1	5.2	60	1.4	2.5
College Station-Bryan	184885	6.26	7.9	63	8.2	6.5	7.6	39	16.9	16.2
Corpus Christi	403280	109.37	0.2	268	0.1	-0.3	0.9	161	-0.8	0.9
Dallas-Fort Worth-Arlington	5161544	34.49	4.3	102	2.0	8.4	10.9	25	11.6	10.3
El Paso	679622	579.11	-0.9	335	-0.5	-0.7	-8.1	360	-4.9	4.5
Houston-Sugar Land-Baytown	4715407	273.38	-3.1	354	-4.9	-4.8	14.6	18	14.5	13.5
Killeen-Temple-Fort Hood	330714	9.58	-4.4	356	-6.4	-5.8	8.2	36	10.2	15.7
Laredo	193117	23.32	0.2	257	0.1	0.1	-2.2	352	-1.0	-0.9
Longview	194042	263.73	10.4	43	8.8	6.2	5.0	63	3.2	4.5
Lubbock	249700	3.84	9.7	52	7.6	4.4	17.6	11	15.0	19.0
McAllen-Edinburg-Mission	569463	0.95	0.0	284	-0.1	-0.1	3.3	79	1.9	2.4
Midland	116009	2.70	-1.3	343	-1.9	-1.4	3.9	71	0.6	3.7
Odessa	121123	38.66	9.9	48	4.3	4.5	-0.5	333	0.4	4.1
San Angelo	105781	0.30	2.2	151	0.4	5.6	15.9	15	11.3	-5.6
San Antonio	1711703	36.79	0.9	203	0.4	-0.5	-2.2	353	-3.7	-3.1
Sherman-Denison	110595	3.08	2.9	132	3.3	3.1	3.1	85	4.1	3.9
Tyler	174706	60.71	5.7	89	3.3	2.0	6.8	46	7.7	4.5
Victoria	111663	120.57	1.0	194	0.1	0.5	6.7	48	2.4	3.1
Waco	213517	23.20	0.3	242	-1.3	0.0	5.2	61	2.9	-0.6
Wichita Falls	151524	42.56	6.0	83	4.7	6.7	4.9	65	0.5	2.3

Source: Author's calculations based on 1) data from the 2000 U.S. Census merged with tract-level RSEI scores for 1999 and 2) 1999 birth data from the Texas Center for Health Statistics merged with tract-level and km^2-grid-cell-level RSEI scores for 1999.
Notes: The discrepancy measures are calculated as described on p. 79 and represent the difference between share of toxic exposure and share in the population. The *All Tract* discrepancy measure is calculated using a sample of all residents matched to tract-level RSEI scores. This measure adjusts for age- and sex-specific Inhalation Exposure Factors (see p. 79). *Rank in US* ranks Texas MSAs by minority discrepancy (descending) out of 363 MSAs in the U.S. in 2000. The *Newborn Tract* measure uses a sample of babies born in 1999 matched to tract-average RSEI scores. The *Newborn* km^2 measure uses the sample of babies born in 1999 but matches them to to km^2-grid-cell-level RSEI scores.

Table 2.4 shows discrepancy scores for Black and Hispanic populations, which represent the difference between the share of pollution borne by these groups and their share in the population.[8] The columns labeled "All, Tract" show the scores for the sample that includes all residents, in which each person is assigned the RSEI concentration of their census tract.[9] At the state level, we find Black residents breathed slightly more than their "fair" share of pollution in 1999, with a discrepancy score of 0.8. Compared to other states, Texas has the 33rd highest Black discrepancy. The state-level Hispanic discrepancy is much higher, at 12.1, placing Texas higher than all but five states on this measure. Another way of comparing discrepancy scores in Texas to those in other states is to look at average MSA discrepancy rankings. If the 24 MSAs in Texas had the 24 highest discrepancy scores, the average rank would be 12.5. If these MSAs had the lowest rankings of the 363 MSAs, the average rank would be 351.5. In Texas, the observed average rank for Black discrepancy is 198.2 and the average rank for Hispanic discrepancy is 112.4. These measures indicate that Blacks in Texas experience approximately average levels of environmental injustice and Hispanics in Texas experience worse than average injustice.

Looking at the distribution of pollution in individual MSAs, we see that the Black discrepancy is positive in all but five cities and the Hispanic discrepancy is positive in all but four. El Paso is the only MSA in which both the Black and Hispanic discrepancy is negative. Within cities, pollution seems to be borne more heavily by one group than another in general, though Black and Hispanic discrepancies are relatively similar in a handful of MSAs, including Wichita Falls, College Station, Tyler, Austin, Sherman and Amarillo; most of these cities have Hispanic populations

---

[8]Previous analyses have used the same measure with all minorities [7, 8]. We choose to analyze Black and Hispanic discrepancies separately here to explore differences in distributional patterns across cities, but we provide measures of overall minority discrepancy in appendix tables on pages 182-184.

[9]As described on p. 79, this measure assigns different Inhalation Exposure Factor (IEF) weights to age-sex subgroups, to better account for probable differences in exposure.

that are proportionally low relative to the state share. In Longview and Odessa, the Black discrepancy is significantly higher than the Hispanic discrepancy.

The "Newborn, Tract" columns in Table 2.4 show discrepancy measures for the sample that includes babies born in 1999, matched to tract-level RSEI scores. At the state level, we see that the newborn Black discrepancy is -0.9, indicating that Black infants are not exposed to inequitable levels of pollution by this measure. The newborn Black discrepancy is lower than the all-resident measure in all but four MSAs. The newborn measure could be lower if the newborn toxic share were relatively low, compared to the all-resident values, and/or the newborn population share were relatively high. Appendix Table A.2, which shows the discrepancy scores along with the toxic and population shares that comprise them, indicates that both these factors are partially responsible.

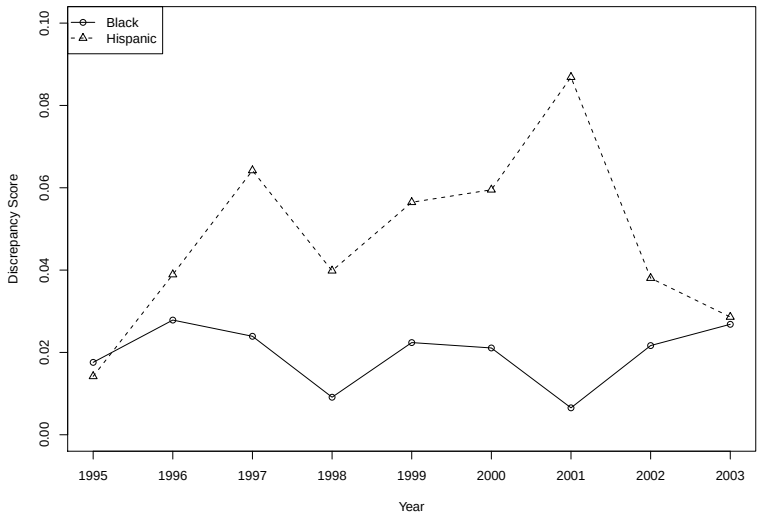
At the state level, the newborn Hispanic discrepancy, at 11.8, is also reduced slightly compared to the all-resident measure, though it is still high. Across MSAs, we see that Hispanic newborns experience greater discrepancies than the general Hispanic population in 10 of 24 cities. Based on disaggregated toxic and population share values in an appendix table on p. 183, we see that Hispanic newborns have higher toxic shares in each city than the Hispanic population generally—not surprising given relatively high Hispanic birth rates—but the overall discrepancy score is sometimes lower due to an even higher share of Hispanic newborns in the population.

Finally, the "Newborn, km^2" columns use the same sample, but infants are matched to the pollution estimate from the square kilometer grid cell their mother lived in at the time of birth, rather than the census tract average. Using this somewhat more precise measure allows us to check whether the tract-average estimate is a good approximation. Further, for other birth years in our sample (1995-1998 and 2000-2003) we only have cell-level pollution estimates, so it allows us to connect these statistics to analyses using other years.

At the state level, we see that the newborn Black discrepancy appears to rise when cell-level pollution measures are used (from -0.9 to 2.2). Assuming that the cell-level estimates more closely approximate actual exposure, this indicates that tract estimates slightly underestimate exposure. Comparing the tract and cell estimate columns, we see that the discrepancy score stayed the same in three of 24 cities and increased in 12. We see relatively large increases in newborn Black discrepancy from one column to the next in Abilene, Beaumont, Dallas and San Angelo.

For Hispanic infants, on the other hand, using grid cell pollution estimates halves the apparent state-level discrepancy compared to using tract estimates (11.8 to 5.6). Across MSAs, the grid cell discrepancy measure increases relative to the tract measure for Hispanic newborns in as many cases as it did for Black newborns. Abiline and San Angelo have significantly lower newborn Hispanic discrepancy scores using the cell-level pollution data, but Dallas, Kileen, and Lubbock all have significantly higher scores.

Figure 2.1. Black and Hispanic Newborn Discrepancy, Texas, 1995-2003



Source: Author's calculations based on 1999 birth data from the Texas Center for Health Statistics merged with km^2-grid-cell-level RSEI scores for 1999.
Notes: The Discrepancy measures are calculated as described on p. 79 and represent the difference between share of toxic exposure and share in the population.

Because we have both birth records and square kilometer grid cell pollution estimates for the years 1995 through 2003, we can also examine how the newborn discrepancy measure changed over that period. Figure 2.1 shows the Black and Hispanic newborn discrepancy measures corresponding to the state-level values shown for 1999 in the "Newborn, km^2" columns of Table 2.4. From this figure, we see that while the 1999 discrepancy measure for Black newborns is a good indicator of its value across the 1995-2003 time period, the Hispanic discrepancy measure is much more variable. Had our "snapshot" year been 1995 or 2003, we would not have concluded that pollution was more inequitably distributed across Hispanic newborns than Black newborns. It is likely that this volatility is a function of relatively more dynamic Hispanic community in Texas, both in terms of growth and distribution across the state, though it is difficult to gain deeper understanding with our data.

To address the concern that very large toxicity-weighted concentration values—some of which may be reporting errors—are skewing our results, we also conducted the discrepancy analysis on a sample with large outliers removed. We chose the cutoff value to be the 97.5 percentile value from the entire U.S. in 1999, a toxicity-weighted concentration of 2,991. Results of this exercise are reported in Table 2.5, the no-outliers analog of Table 2.4. We learn that removing outliers does have a somewhat significant impact on our discrepancy measures at the state level and in high-exposure cities. Using the sample with outliers removed increases the tract-level Texas black discrepancy from 0.8 to 4.4 and changes its rank in the U.S. from 33 to 18. It also increases the all-Texas newborn discrepancy measures. Conversely, removing outliers decreases the tract-level Hispanic discrepancy from 12.1 to 6.9 and reduces its rank within the U.S. from 6 to 8. Likewise, the Texas newborn Hispanic discrepancy measures are decreased.

The change in Hispanic measures at the state level seems to be a function of more Hispanic residents living outside cities, since a look at the MSA-level discrepancy

Table 2.5. Black and Hispanic discrepancies in distribution of air toxics, all residents and newborns only, Texas MSAs, 1999. Observations with exposure above the U.S. 97.5 percentile removed.

Metropolitan Area	Pop (Apr 2000)	RSEI Score Med Resident	Black Discrepancy All Tract	Black Discrepancy Rank in US	Black Discrepancy Newborn Tract	Black Discrepancy Newborn km^2	Hispanic Discrepancy All Tract	Hispanic Discrepancy Rank in US	Hispanic Discrepancy Newborn Tract	Hispanic Discrepancy Newborn km^2
Texas (all)	20851820	46.40	4.4	18	2.5	4.0	6.9	8	6.0	2.1
Abilene	160245	23.64	2.3	148	2.3	4.4	10.3	20	11.6	19.0
Amarillo	226522	80.68	0.7	225	0.7	6.5	0.7	175	-1.8	16.2
Austin-Round Rock	1249763	30.71	1.8	167	1.5	1.2	3.2	77	3.5	13.8
Beaumont-Port Arthur	385090	917.56	4.2	97	2.7	6.1	2.2	108	1.6	3.9
Brownsville-Harlingen	335227	6.48	0.1	298	0.0	6.3	5.2	52	1.4	2.9
College Station-Bryan	184885	6.26	7.9	56	8.2	6.7	7.6	35	16.9	2.3
Corpus Christi	403280	109.37	0.4	244	0.0	4.5	5.2	51	3.1	4.1
Dallas-Fort Worth-Arlington	5161544	34.49	4.3	96	2.0	-2.7	10.9	18	11.6	11.9
El Paso	679622	579.11	0.2	271	0.1	3.1	-1.7	352	-2.7	3.9
Houston-Sugar Land-Baytown	4715407	273.38	-0.2	328	-2.0	3.0	10.0	21	9.4	3.1
Killeen-Temple-Fort Hood	330714	9.58	-4.4	361	-6.4	1.5	8.2	31	10.2	5.7
Laredo	193117	23.32	0.2	268	0.1	1.5	-2.2	353	-1.0	2.9
Longview	194042	263.73	5.7	80	3.4	-1.8	3.7	64	2.6	8.7
Lubbock	249700	3.84	9.7	43	7.6	0.5	17.6	10	15.0	3.1
McAllen-Edinburg-Mission	569463	0.95	0.0	300	-0.1	0.1	3.3	75	1.9	3.2
Midland	116009	2.70	-1.3	353	-1.9	-1.4	3.9	60	0.6	3.7
Odessa	121123	38.66	9.9	39	4.3	-0.1	-0.5	336	0.4	2.5
San Angelo	105781	0.30	2.2	149	0.4	-0.1	15.9	12	11.3	2.4
San Antonio	1711703	36.79	0.9	210	0.4	-0.2	-2.2	354	-3.7	1.7
Sherman-Denison	110595	3.08	2.9	132	3.3	5.6	3.1	80	4.1	-5.6
Tyler	174706	60.71	5.7	81	3.3	0.2	6.8	38	7.7	-0.6
Victoria	111663	120.57	1.0	204	0.1	0.0	6.7	41	2.4	-0.6
Waco	213517	23.20	0.3	255	-1.3	0.5	5.2	53	2.9	-1.2
Wichita Falls	151524	42.56	6.0	76	4.7	-0.4	4.9	55	0.5	-2.0

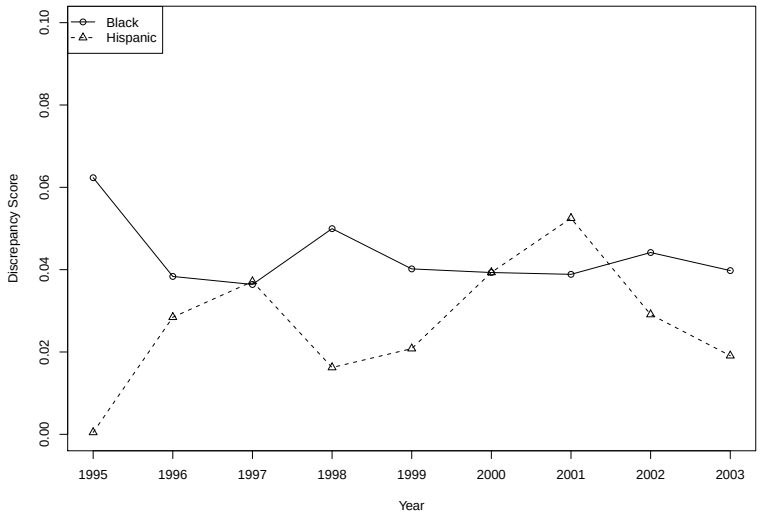
Source: Author's calculations based on 1) data from the 2000 U.S. Census merged with tract-level RSEI scores for 1999 and 2) 1999 birth data from the Texas Center for Health Statistics merged with tract-level and km^2-grid-cell-level RSEI scores for 1999.
Notes: Observations with a toxicity-weighted concentration above 2,991, the 97.5 percentile tract concentration in the U.S. in 1999, were excluded from this sample. The Discrepancy measures are calculated as described on p. 79 and represent the difference between share of toxic exposure and share in the population. The *All Tract* discrepancy measure is calculated using a sample of all residents matched to tract-level RSEI scores. This measure adjusts for age- and sex-specific Inhalation Exposure Factors (see p. 79). *Rank in US* ranks Texas MSAs by minority discrepancy (descending) out of 363 MSAs in the U.S. in 2000. The *Newborn Tract* measure uses a sample of babies born in 1999 matched to tract-average RSEI scores. The *Newborn* km^2 measure uses the sample of babies born in 1999 but matches them to to km^2-grid-cell-level RSEI scores.

measures shows that Hispanic tract-level scores mostly remained the same between the two samples, increasing in three cities (of 24) and decreasing in three. Black tract-level discrepancies increased in four cities and decreased in one. Not surprisingly, since our removal of outliers affected tracts with high concentrations, the five MSAs with changes in tract-level discrepancy were among the most polluted in the group: Beaumont, Corpus Christi, El Paso, Houston and Longview.

While the majority of tract-level discrepancy measures remained the same when high outliers were removed, grid-cell-level newborn discrepancy measures changed in each MSA except one (Midland). Comparing the Black newborn measures in Table 2.5 to those in Table 2.4, we see that discrepancy scores increased in the same four cities as in the tract-level analysis; overall, black scores increased in 13 MSAs and decreased in 10. In the full sample analysis, 15 cities had positive black newborn discrepancies, and in the sample with outliers removed, 18 cities appear to have positive discrepancies. Turning to the Hispanic newborn cell-level scores, we see 12 MSAs with increased scores relative to the full sample version and 11 MSAs with decreased scores. In both analyses, there are 19 MSAs with positive Hispanic discrepancy scores, but only one MSA (Waco) overlaps.

Figure 2.2 is the analog to Figure 2.1, excluding observations with cell-level exposures above a toxicity-weighted concentration of 2,991. While this cutoff excluded exactly 2.5 percent of U.S. tracts by exposure in 1999, it excludes about 1.6 percent of all 1995-2003 Texas birth records. Using this smaller sample preserves the trend of the Hispanic newborn discrepancy time series, but in each year, the Texas-level discrepancy appears about 2 percentage points lower. Conversely, using the sample with outliers removed shifts the Black newborn discrepancy scores up and does alter the trend somewhat, though as in Figure 2.1, the Black discrepancy trend line in Figure 2.2 is relatively less volatile compared to the Hispanic trend line.

Figure 2.2. Black and Hispanic Newborn Discrepancy, Texas, 1995-2003. Observations with cell-level exposure above the U.S. 97.5 percentile removed.



Source: Author's calculations based on 1999 birth data from the Texas Center for Health Statistics merged with km^2-grid-cell-level RSEI scores for 1999.
Notes: Observations with a grid-cell-level toxicity-weighted concentration above 2,991, the 97.5 percentile tract concentration in the U.S. in 1999, were excluded from this sample. The Discrepancy measures are calculated as described on p. 79 and represent the difference between share of toxic exposure and share in the population.

The broad picture that emerges from these analyses is that Black and Hispanic newborns are disproportionately exposed to toxic pollution in Texas, though the minority discrepancy measure is not extremely robust to the exclusion of outliers. To account for this issue and to compliment the purely distributional discrepancy measures, we present in Table 2.6 the percentage of people and infants, by race/ethnicity, exposed to "high" levels of toxic air pollution. As described above, high exposure is defined here as the exposure of the 75th-percentile resident of the U.S. as a whole in 1999, a RSEI score of 247.[10]

As we observed in Table 2.3, levels of pollution exposure vary widely across Texas cities, and the difference in exposure within race groups across cities is generally

[10]We also conducted these analyses with outliers removed, but differences between the two versions were very small and are not reported here. Results are available from the author.

Table 2.6. Share of population with high RSEI scores by race, all residents and newborns, Texas MSAs, 1999.

Metropolitan Area	Pop (Apr 2000)	RSEI Score Med Resident	White, % >US 75p (RSEI=247) All Tract	White, % >US 75p (RSEI=247) Newborn Tract	White, % >US 75p (RSEI=247) Newborn km^2	Black, % >US 75p (RSEI=247) All Tract	Black, % >US 75p (RSEI=247) Newborn Tract	Black, % >US 75p (RSEI=247) Newborn km^2	Hispanic, % >US 75p (RSEI=247) All Tract	Hispanic, % >US 75p (RSEI=247) Newborn Tract	Hispanic, % >US 75p (RSEI=247) Newborn km^2
Texas (all)	20851820	46.40	16.7	17.7	15.0	28.3	28.3	25.6	24.4	26.9	19.0
Abilene	160245	23.64	3.6	2.6	0.9	3.9	4.1	3.6	4.4	4.7	1.3
Amarillo	226522	80.68	2.2	1.5	0.2	0.1	1.2	0.0	0.5	0.2	0.1
Austin-Round Rock	1249763	30.71	3.2	3.4	2.7	4.9	5.5	3.7	4.6	6.0	3.0
Beaumont-Port Arthur	385090	917.56	87.2	88.8	88.2	98.9	99.5	99.3	97.2	98.9	98.6
Brownsville-Harlingen	335227	6.48	0.6	0.8	0.8	0.0	0.0	0.0	1.1	1.3	2.5
College Station-Bryan	184885	6.26	1.2	0.1	0.0	1.3	0.9	0.0	0.2	0.2	0.0
Corpus Christi	403280	109.37	19.7	19.3	16.9	22.1	21.9	15.4	20.5	20.6	16.8
Dallas-Fort Worth-Arlington	5161544	34.49	2.8	2.9	1.7	5.0	4.7	4.1	8.6	9.1	7.0
El Paso	679622	579.11	95.8	96.5	4.9	98.3	99.3	8.8	86.4	88.2	9.1
Houston-Sugar Land-Baytown	4715407	273.38	48.1	48.6	44.7	57.4	55.8	51.1	65.6	66.9	63.4
Killeen-Temple-Fort Hood	330714	9.58	4.9	4.7	6.6	5.2	4.4	6.4	11.1	10.6	15.9
Laredo	193117	23.32	19.2	20.3	9.4	30.3	33.3	33.3	11.2	9.8	6.9
Longview	194042	263.73	48.7	64.2	60.7	58.6	71.2	70.1	57.5	68.2	67.5
Lubbock	249700	3.84	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.6
McAllen-Edinburg-Mission	569463	0.95	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Midland	116009	2.70	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Odessa	121123	38.66	3.8	2.2	0.9	23.9	10.9	12.0	3.6	3.3	3.7
San Angelo	105781	0.30	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
San Antonio	1711703	36.79	2.0	2.2	1.0	0.6	0.4	0.1	1.5	1.5	0.9
Sherman-Denison	110595	3.08	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Tyler	174706	60.71	2.0	2.0	7.1	4.3	3.9	3.9	3.9	4.0	5.2
Victoria	111663	120.57	7.5	8.2	2.2	4.8	1.5	3.0	7.5	3.5	1.0
Waco	213517	23.20	2.2	2.5	0.3	2.6	3.7	0.2	7.8	7.4	0.0
Wichita Falls	151524	42.56	5.9	6.6	4.0	19.3	15.3	9.7	17.8	11.9	10.6

Source: Author's calculations based on 1) data from the 2000 U.S. Census merged with tract-level RSEI scores for 1999 and 2) 1999 birth data from the Texas Center for Health Statistics merged with tract-level and km^2-grid-cell-level RSEI scores for 1999.
Notes: % >US 75p refers to the share of the population (by race/ethnicity) with 1999 RSEI scores above 247.0, the U.S. population-weighted 75-percentile RSEI score. The *All Tract* measure is calculated using a sample of all residents matched to tract-level RSEI scores. The *Newborn Tract* measure uses a sample of babies born in 1999 matched to tract-average RSEI scores. The *Newborn km^2* measure uses the sample of babies born in 1999 but matches them to km^2-grid-cell-level RSEI scores.

greater than the difference within cities across races. Still, there is considerable evidence of environmental injustice using this measure. At the state level, just under 17 percent of all White Texas residents live in top-quartile RSEI score tracts, while 28 percent of Blacks and 24 percent of Hispanics do. The discrepancy measures indicated that the Hispanic population in Texas experienced higher levels of environmental injustice than the Black population, but this measure adds another dimension to the EJ picture, and we see that a greater proportion of Blacks than Hispanics are exposed to high levels of toxic air pollution. The share of Hispanic residents living in highly polluted areas is still almost 8 percentage points higher than the White share, however. Amarillo and San Antonio are the only two MSAs in Texas in which the White group have the highest share of residents above the 75 percentile score. Conversely, there are a relatively large number of cities in which high levels of pollution are disproportionately borne by Blacks (Laredo, Odessa), Hispanics (Dallas, Houston, Kileen, Waco) or both groups (Beaumont, Longview, Wichita Falls).

Not surprisingly, the shares of newborns exposed to high pollution are similar, in most places, to the shares of all residents exposed to high pollution. In Longview, we find that newborns are disproportionately exposed, across races, ostensibly due to the geographic distribution of childbearing-age versus older residents across the city. In Odessa, where Black environmental injustice is high, Black newborns are relatively less exposed to high RSEI scores than Black residents in general.

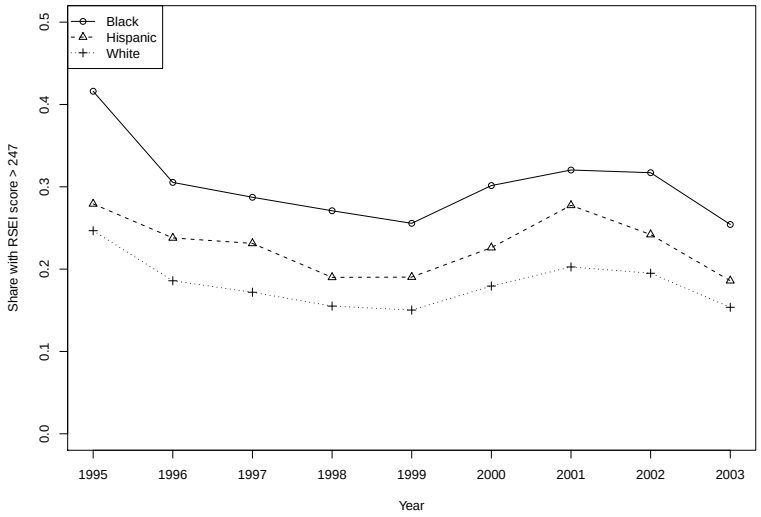
Comparing the estimates obtained matching newborns to 1) tract-level pollution estimates and 2) square kilometer grid cell estimates, we see that the share of highly exposed newborns is reduced somewhat when the latter is used. This finding indicates that newborns (and perhaps people in general) live in the less polluted portions of tracts. For the White and Black groups, the share of highly-exposed newborns falls by about 3 percentage points at the state level when the grid cell measure is used; for Hispanic newborns, the measure drops by almost 8 percentage points,

indicating that, within neighborhoods, Hispanics live farther from industrial activity than Blacks. In Laredo, using the grid cell measure halves the percentage of White newborns experiencing high levels of pollution from 20.3 to 9.4, and in Tyler, using this measure increases the White highly-exposed newborn share from 2.0 to 7.1.

Using the more refined measure in El Paso leads to remarkably different conclusions for all race groups: assigning newborns to tract-level estimates implies that 96.5, 99.3 and 88.2 percent of White, Black and Hispanic newborns, respectively, are exposed to RSEI scores above 247; using the grid-cell estimates changes those figures to 4.9, 8.8, and 9.1. A closer examination of the sources of pollution in El Paso helps explain why. Manganese and chromium emissions from one mining company accounted for over 90 percent of the aggregate toxicity-weighted RSEI concentration in 1999. It seems likely that the lack of concordance between tract- and cell-level pollution estimates can be traced to the fact that these metals, due to their weight, are assumed not to travel as far as other chemicals in the fate and transport model. Further, the tracts that receive this pollution, according to the RSEI model, are relatively large (e.g., the most affected tract is 8.7 square kilometers), so it is likely that residents in affected tracts are living farther away from the mines and the highest concentrations of manganese and chromium.

Again, access to newborn and grid-cell-level pollution data for the years 1995-2003 allows us to look at time trends in the share of newborns exposed to high levels of pollution. The 1999 values in Figure 2.3 are analagous to the Texas-level shares in the "Newborn, km^2" columns of Table 2.6. The patterns we see here are more consistent than in Figure 2.1, which showed the purely distributional discrepancy measures. Across all nine years, Black newborns are most likely to be exposed to high levels of pollution, followed by Hispanic then White newborns. The trend lines appear to rise and fall in line with economic activity. We also observe that the overall slope does

Figure 2.3. Share of newborns with high RSEI scores by race, Texas, 1995-2003



Source: Author's calculations based on 1999 birth data from the Texas Center for Health Statistics merged with km^2-grid-cell-level RSEI scores for 1999.
Notes: A "high" RSEI score is greater than or equal to 247.0, the U.S. population-weighted 75-percentile RSEI score in 1999 (see Table 2.6).

not appear to be strictly decreasing, in line with national TRI trends, though again, our period of analysis is quite short.

2.4.2 Predictors of newborn pollution exposure

Table 2.7 shows summary statistics for the births in our first regression sample. The toxicity-weighted concentration measures reinforce the existence of the racial inequities we observed in the last section: Black newborns experience the highest average toxic exposure and White newborns experience the lowest.

Below these rows, we show mean values of mother and area chacteristics, other than race, that may help predict a newborn's pollution exposure. On virtually every measure, average Black and Hispanic levels of these characteristics would seem to predict higher levels of exposure, based on our hypotheses of how these factors correlate with pollution. White mothers are more likely to be older, married and more educated when they have children; we assume these characteristics facilitate living

Table 2.7. Sample means, Texas newborns, 1995-2003.

	All	White	Black	Hispanic
N / Share of sample	2596691	0.405	0.125	0.470
Tox-wt concentration (RSEI)	328.1	280.6	378.3	355.7
	(1588.0)	(1672.7)	(1152.8)	(1611.5)
Log tox-wt concentration (log(RSEI+1))	4.137	3.994	4.695	4.113
	(1.835)	(1.712)	(1.601)	(1.963)
Age	26.16	27.67	24.87	25.19
	(6.11)	(6.01)	(6.08)	(5.93)
Education	12.08	13.66	12.53	10.59
	(3.07)	(2.41)	(2.13)	(3.06)
Married	0.686	0.814	0.391	0.654
Tract med HH inc ($000)	41.38	52.18	35.27	33.69
	(20.18)	(22.65)	(15.50)	(13.83)
Tract pop dens (000/km^2)	1.463	1.026	1.567	1.812
	(1.510)	(0.987)	(1.458)	(1.778)
Tract % no HS	28.35	16.57	27.93	38.63
	(19.50)	(12.86)	(15.50)	(19.42)
Tract % vacant	7.37	6.72	7.82	7.81
	(5.20)	(4.97)	(4.96)	(5.40)
Tract % owner occ	61.19	67.97	53.05	57.50
	(23.23)	(22.15)	(24.15)	(22.35)
Tract % mfg empl	12.04	12.45	11.40	11.85
	(5.38)	(5.31)	(4.98)	(5.51)
CBSA seg: Black dissim	0.516	0.509	0.522	0.521
	(0.081)	(0.077)	(0.073)	(0.085)
CBSA seg: Hisp dissim	0.429	0.427	0.431	0.430
	(0.067)	(0.073)	(0.073)	(0.059)
CBSA seg: Black isolation	0.309	0.339	0.385	0.263
	(0.150)	(0.126)	(0.102)	(0.165)
CBSA seg: Hisp isolation	0.488	0.422	0.412	0.566
	(0.182)	(0.137)	(0.122)	(0.197)
CBSA seg: Black clustering	1.079	1.087	1.102	1.066
	(0.048)	(0.045)	(0.044)	(0.047)
CBSA seg: Hisp clustering	1.104	1.104	1.103	1.104
	(0.051)	(0.047)	(0.043)	(0.055)

Source: Author's calculations based on 1) birth data from the Texas Center for Health Statistics (1995-2003) merged with 2) km^2-grid-cell-level RSEI scores corresponding to birth years and 3) data from the 2000 U.S. Census.

in better neighborhoods with lower pollution. Looking at tract characteristics, we observe that White mothers live in significantly higher-income neighborhoods that are characterized by lower population density, more educated adults, lower housing vacancy rates and more owner-occupied houses. White newborns are more likely to be born into neighborhoods with higher shares of manufacturing employment, which we expect to be a significant predictor of pollution, but this is the only neighborhood-level exposure variable that would predict higher levels of toxic exposure to White newborns.

In terms of exposure to segregation, Black and Hispanic newborns are significantly more likely to live in cities with high Black and Hispanic isolation, respectively. Exposure to the other segregation measures is distributed more equally across races, but the effects of living in a city with high Black clustering, for example, may be very different for Black and White residents.

To take an initial look at the relative importance of city location versus neighborhood (within city) location in terms of pollution exposure, we also calculated the between- and within-CBSA variance in our dependent variable, logged toxicity-weighted concentration. With a sample mean of 4.14 (shown above), the between-city standard deviation in our pollution variable, at 1.87, is slightly higher than the within-city standard deviation of 1.11. This indicates that city status is more important than within-city neighborhood location in predicting pollution exposure, though there is also a fair amount of within-city variance.

In Table 2.8, we present the results from OLS regressions of pollution exposure on mother and area characteristics for all births in Texas. In the bivariate pollution-race regression, shown in Column 1, the Black coefficient of 0.703 implies that the toxicity-weighted pollution exposure of Black newborns is about twice as high as the

Table 2.8. Predictors of toxic exposure among newborns in Texas, 1995-2003.

	1	2	3	4	5	6	7
Mother Black	0.703**	0.631**	0.734**	0.654**	0.536**	0.108#	0.207**
Mother Hispanic	0.133	0.017	0.125	−0.022	−0.055	0.254**	0.308*
Mother age: teens		−0.169**	−0.169**	−0.077*	−0.053**	0.0013	−0.016
Mother age: 30s		0.054	0.027	0.027	0.006	0.027*	0.030#
Mother age: 40s		0.121#	0.102*	0.111**	0.085**	0.103**	0.105**
Mother ed: no HS		0.264**	0.305*	0.127#	0.103	−0.037	−0.039
Mother ed: BA		−0.048	−0.103	−0.089	−0.090	0.021	0.016
Mother unmarried		0.137**	0.161*	0.132*	0.113**	0.105*	0.103*
Tract med HH inc ($000)			0.023	0.049	0.024	−0.030*	−0.004
Tract income squared			−0.000143	−0.000251	−0.000138	0.000113#	0.000000
Tract pop dens				0.098	0.060	0.049	0.057
Tract % no HS				0.015	0.005	−0.007	0.005
Tract % vacant				−0.036#	−0.030#	−0.022	−0.021
Tract % owner occ				−0.018*	−0.015*	−0.003	−0.006#
Tract % mfg empl				0.072*	0.071*	0.054*	0.043#
CBSA seg: Black dissim					0.218		
CBSA seg: Hisp dissim					0.517*		
CBSA seg: Black isolation						1.135**	
CBSA seg: Hisp isolation						0.589*	
CBSA seg: Black clustering							1.075**
CBSA seg: Hisp clustering							0.075
Constant	4.190**	4.135**	4.074**	3.950**	3.275**	2.661**	2.930**
Year controls	Y	Y	Y	Y	Y	Y	Y
N	2596691	2596691	2596691	2596691	2596691	2596691	2596691
Adjusted R^2	0.020	0.026	0.032	0.149	0.195	0.393	0.351

Source: Author's calculations based on 1) birth data from the Texas Center for Health Statistics (1995-2003) merged with 2) km^2-grid-cell-level RSEI scores corresponding to birth years and 3) data from the 2000 U.S. Census.
Notes: Results from OLS regressions with dependent variable $log(RSEI + 1)$. Sample includes Texas births to White, Black and Hispanic mothers whose birth records could be geocoded and located within a Core Based Statistical Area (CBSA). Standard errors (not shown) are clustered at the CBSA level and adjusted for heteroskedasticity. ** indicates statistically significant at < 0.01; * at < 0.05; # at < 0.10. Variables are defined as described on p. 88. Tract and segregation variables were centered to facilitate interpretation of the constant term. Segregation variables were normalized, so associated coefficients represent the change in the dependent variable with a 1 standard deviation increase in the segregation measure.

exposure of White newborns.[11] The Hispanic coefficient implies about a 14 percent higher exposure relative to Whites, but this estimate is not statistically significant.

In Column 2 of Table 2.8, we add mother characteristics to the right-hand side of the model equation, on the basis that differences in age, education and marital status between races—and the differences in access to resources they imply—may explain some of the differences in exposure. We find that the Black coefficient is attenuated but still very high, implying that at the same maternal age, education and marital status, we still expect an infant born to a Black mother in Texas to be exposed to pollution about 88 percent higher than an infant born to a White mother with similar characteristics. We also find that births to mothers who did not graduate high school are exposed to about 30 percent higher pollution than mothers who did; births to unmarried mothers are exposed to about 15 percent higher RSEI scores; and exposure tends to increase with age, when education and marital status are controlled for.

Adding controls for neighborhood income in Column 3 is associated, counter-intuitively, with an increase in the Black coefficient, indicating that greater Black exposure to pollution is not a simple function of access to resources. The coefficients on neighborhood income and its square imply a positive relationship between income and pollution at the state level through a neighborhood income of about $80,000, though lack of statistical significance at the 10 percent level implies that income is not a particularly consistent determinant of pollution exposure when the entire state is the area of analysis. Adding income also does not affect the coefficients on the mother characteristic variables to a great extent.

Including other neighborhood-level variables, as shown in Column 4, significantly increases the prediction power of the model, with the adjusted R^2 value increasing from 0.03 to 0.15. These variables all have the expected signs, with population den-

[11] We include year controls in what we call the bivariate regression.

sity, share of adults who did not graduate high school and share of adults employed in manufacturing all predictors of greater pollution exposure and vacancy rate and share of owner-occupied houses associated with lower exposure, though only the vacancy, owner-occupied and manufacturing employment variables are statistically significant. We find that controlling for neighborhood characteristics decreases the apparent protectivity of teen motherhood and reduces the risk associated with low education, indicating, as expected, that mother characteristics help predict pollution exposure via their ability to predict the type of area a woman might live in. This regression shows that, at the state level, neighborhood income is not as good an indicator of pollution exposure as these other variables. Notably, the Black coefficient in this model remains very high, indicating that differential exposure to these neighborhood characteristics can explain only a small part of the discrepancy.

Columns 5-7 in Table 2.8 show the results from adding three segregation measures, one at a time, to the model with all mother and area characteristics. From these results, we see that knowing the racial composition and distribution of the city a child is born into can tell us a great deal about that child's early pollution exposure. In Column 5, we observe that being born in a city with Hispanic dissimilarity one standard deviation above the mean is associated with 68 percent higher exposure overall, other characteristics equal. The coefficient on the Black dissimilarity measure is also positive but smaller and not statistically significant. In Column 6, we see that the coefficients on both isolation measures are large and statistically significant, with a one standard deviation increase from the mean in Black isolation being associated three times higher exposure. In Column 7, we see a coefficient on Black clustering that is similar in magnitude to the Black isolation coefficient, though the coefficient on Hispanic clustering is small and not statistically significant. Unlike the dissimilarity index, the Black isolation and clustering measures are highly correlated with the share of Black residents in the city (and, by extension, with one another). These results

reinforce what we noticed earlier—cities with a greater proportion of Black residents tend to be more polluted. While the same is not necessarily true in the Hispanic case, we do see that cities with a more uneven distribution of Hispanics tend to be more highly polluted.

Controlling for segregation decreases the Black coefficient and increases the Hispanic coefficient, indicating that Black newborns are exposed to more pollution on average because they live in cities with proportionally more Black residents, and Hispanics are exposed to less pollution on average because they live in cities with less pollution and proportionally fewer Black residents. At any given level of segregation, Black and Hispanic newborns are still more likely than White newborns to experience higher levels of pollution, though living in a more segregated city appears to make all newborns worse off in terms of pollution exposure.

The models estimated in Table 2.9 are similar to those in 2.8, Columns 5-7, but instead of controlling for segregation alone, we control for all CBSA-specific characteristics by including CBSA fixed effects. We see that, though Hispanics tend to live in less polluted places across the state, within cities, Hispanic newborns are even more disproportionately exposed to pollution than are Black newborns. Controlling for mother characteristics (Column 2) attenuates the apparent disparities, but only to a small extent, and the effects of maternal age, education and marital status are of the expected signs and statistically significant.

Controlling for income and its square (Column 3) reduces the Black and Hispanic coefficients by roughly the same amount and renders the Black coefficient statistically insignificant, implying that some but not all of the racial disparities in exposure can be explained by neighborhood socioeconomic status. While income had a positive relationship with pollution at the state level, indicating that more industrial regions within the state are more prosperous, the income coefficients in this model imply a negative relationship between income and industrial air pollution within cities. The

Table 2.9. Predictors of toxic exposure among newborns within Texas CBSAs (CBSA fixed effects), 1995-2003.

	1	2	3	4	5		
					White	Black diff	Hisp diff
Mother Black	0.280**	0.220**	0.074	0.140*		0.261*	
Mother Hispanic	0.419**	0.342**	0.225**	0.122*			0.154
Mother age: teens		−0.034*	−0.021	−0.007	−0.010#	−0.015#	0.002
Mother age: 30s		−0.044**	−0.003	−0.001	0.004	−0.011	−0.008
Mother age: 40s		−0.026#	0.017*	0.026**	0.041*	−0.047	−0.038#
Mother ed: no HS		0.090**	0.031#	0.000	0.019	−0.016	−0.028
Mother ed: BA		−0.073#	0.022	0.020	0.019	−0.028	0.002
Mother unmarried		0.074**	0.044**	0.045**	0.059**	0.001	−0.040
Tract med HH inc ($000)			−0.019**	0.007	0.003	−0.009	0.024*
Tract income squared			0.000076**	−0.000035	−0.000009	0.000037	−0.000164*
Tract pop dens				0.026	0.138**	−0.184**	−0.124**
Tract % no HS				0.0097**	0.007#	0.002	0.007
Tract % vacant				−0.018#	−0.025*	0.011*	0.016*
Tract % owner occ				−0.0092**	−0.010**	0.007*	0.000
Tract % mfg empl				0.038#	0.044	−0.018	−0.005
Constant	4.101**	4.120**	4.183**	4.101**	4.069**		
Year controls	Y	Y	Y	Y		Y	
N	2596691	2596691	2596691	2596691		2596691	
Adjusted R^2	0.043	0.047	0.066	0.116		0.129	

Source: Author's calculations based on 1) birth data from the Texas Center for Health Statistics (1995-2003) merged with 2) km^2-grid-cell-level RSEI scores corresponding to birth years and 3) data from the 2000 U.S. Census.
Notes: Results from within-CBSA (CBSA fixed effect) regressions with dependent variable $log(RSEI + 1)$. Sample includes Texas births to White, Black and Hispanic mothers whose birth records could be geocoded and located within a Core Based Statistical Area (CBSA). Standard errors (not shown) are clustered at the CBSA level and adjusted for heteroskedasticity. ** indicates statistically significant at < 0.01; * at < 0.05; # at < 0.10. In Model 5, significance stars in White column represents probability that estimated coefficient is different from zero; coefficients in Black Diff and Hisp Diff columns represent difference from White coefficient and stars represent probability that the difference is greater than zero. Variables are defined as described on p. 88. Tract variables were centered to facilitate interpretation of the constant term.

coefficient on income squared indicates the relationship is convex, but the implied turning point occurs at a neighborhood income of $120,000, well beyond most observed values. In Column 4, as in the analogous column in Table 2.8, the explanatory power of the model nearly doubles when other neighborhood-level controls are added. The income coefficients lose statistical significance, but the coefficients on all other tract-level variables are statistically significant in the expected directions.

The last three columns of Table 2.9 allow the effects of all mother and area characteristics to vary by race, and we find evidence that several do. While the relationship between income and pollution, once other neighborhood variables are controlled for, is not distinguishable from zero in the case of White and Black births, neighborhood income appears to be positively correlated with pollution exposure for Hispanic newborns. This may have to do with the average income levels between races; a relatively high-income Hispanic resident may be *more* likely to live near industrial activity while a relatively high-income White resident may be *less* likely. The effects of changes in population density and share of vacant houses also vary across race groups. For White newborns, being born into a more densely populated neighborhood is associated with higher pollution exposure, while for Blacks, it is associated with slightly lower exposure and for Hispanics, population density does not appear to matter much. Again, this could have to do with different average population density levels between races. For Whites, higher population density could be associated with the difference between less-polluted suburban and more-polluted urban neighborhoods, while for Blacks, it may represent the difference between less- and more-dense urban areas, if less-dense areas are more likely to be near polluting factories. Finally, we see that the share of owner-occupied housing, a variable included to indicate political power to resist pollution, is more protective for Whites than for Blacks. This could be the case if establishment in one's neighborhood does not translate to political power for Black residents as readily as it does for Whites.

Table 2.10. Predictors of toxic exposure among newborns within Texas neighborhoods (tract fixed effects), 1995-2003.

	1	2
Mother Black	0.023**	0.021**
Mother Hispanic	0.025*	0.022*
Mother age: teens		−0.0047
Mother age: 30s		−0.0032*
Mother age: 40s		0.0007
Mother ed: no HS		0.0084*
Mother ed: BA		0.0005
Mother unmarried		0.0079
Constant	4.257**	4.255**
Year controls	Y	Y
N	2596691	2596691
Adjusted R^2	0.039	0.039

Source: Author's calculations based on 1) birth data from the Texas Center for Health Statistics (1995-2003) merged with 2) km^2-grid-cell-level RSEI scores corresponding to birth years and 3) data from the 2000 U.S. Census.
Notes: Results from within-neighborhood (tract fixed effect) regressions with dependent variable $log(RSEI + 1)$. Sample includes Texas births to White, Black and Hispanic mothers whose birth records could be geocoded and located within a Core Based Statistical Area (CBSA). Standard errors (not shown) are clustered at the CBSA level and adjusted for heteroskedasticity. ** indicates statistically significant at < 0.01; * at < 0.05; # at < 0.10. Variables are defined as described on p. 88.

Recall that newborns were assigned to the RSEI exposure measure for the square kilometer grid cell in which their mothers resided at the time of birth, as well as the census tract corresponding to the mother's address. A single grid cell, moreover, may span more than one tract and vice versa. This allows us to take an initial, rough look at whether minorities are likely to be exposed to more pollution, even within neighborhoods. We present the results of this analysis in Table 2.10, where we do find evidence that Black and Hispanic newborns are systematically born into the more polluted parts of census tracts. Being either Black or Hispanic is associated with about two percent higher pollution exposure, all other tract variables held constant. This finding holds even when mother characteristics are controlled for.

2.4.3 Predictors of the change in pollution exposure between sibling births

In the first set of regression analyses, we explored the predictors of pollution exposure among newborns in Texas. In the second set, we look at the predictors of the *change* in pollution exposure experienced by a single mother between the

Table 2.11. Sample means, change-between-sibling-birth sample, Texas, 1995-2003.

	Did not move between births				Moved between births (same CBSA)			
	All	White	Black	Hispanic	All	White	Black	Hispanic
N / Share of sample	269270	0.469	0.124	0.408	259086	0.393	0.187	0.420
Δ Log RSEI	-0.029	-0.008	-0.015	-0.057	-0.120	-0.162	-0.054	-0.109
	(0.848)	(0.785)	(0.712)	(0.949)	(1.248)	(1.252)	(1.093)	(1.306)
Years between births	2.461	2.471	2.307	2.497	2.914	2.964	2.794	2.922
	(1.323)	(1.211)	(1.357)	(1.428)	(1.570)	(1.514)	(1.576)	(1.614)
Age	27.95	30.00	25.68	26.28	25.86	27.34	24.78	24.96
	(5.92)	(5.44)	(5.81)	(5.71)	(5.32)	(5.39)	(5.02)	(5.05)
Δ Ed: No HS → No HS	0.208	0.076	0.202	0.363	0.275	0.150	0.231	0.411
Δ Ed: No HS → HS	0.055	0.028	0.099	0.072	0.086	0.062	0.119	0.094
Δ Ed: No HS → BA	0.0008	0.0006	0.0010	0.0008	0.0013	0.0014	0.0015	0.0011
Δ Ed: HS → HS	0.429	0.396	0.513	0.442	0.457	0.477	0.521	0.410
Δ Ed: HS → BA	0.025	0.032	0.026	0.018	0.026	0.035	0.026	0.017
Δ Ed: BA → BA	0.263	0.453	0.130	0.084	0.134	0.257	0.073	0.046
Δ Marital: No → No	0.203	0.073	0.555	0.245	0.279	0.139	0.565	0.282
Δ Marital: No → Yes	0.084	0.053	0.091	0.118	0.164	0.145	0.159	0.184
Δ Marital: Yes → No	0.028	0.014	0.029	0.044	0.052	0.040	0.043	0.068
Δ Marital: Yes → Yes	0.684	0.860	0.324	0.592	0.504	0.675	0.232	0.465
Tract med HH inc ($000)	44.38	56.50	33.97	33.61	41.81	52.81	34.48	34.80
	(22.75)	(24.12)	(15.82)	(14.37)	(20.53)	(23.27)	(15.30)	(14.38)
Tract pop dens	1.235	0.962	1.404	1.497	1.383	0.987	1.634	1.642
	(1.157)	(0.905)	(1.287)	(1.296)	(1.278)	(0.967)	(1.431)	(1.365)
Tract % no HS	26.95	15.24	30.52	39.33	26.91	17.12	28.50	35.36
	(19.36)	(12.29)	(15.37)	(18.95)	(18.07)	(13.08)	(15.40)	(18.70)
Tract % vacant	7.22	6.31	8.16	7.99	7.17	6.52	7.81	7.50
	(5.12)	(4.72)	(5.06)	(5.42)	(4.84)	(4.53)	(4.91)	(5.02)
Tract % owner occ	66.85	72.69	56.73	63.20	61.71	69.19	50.96	59.50
	(20.21)	(19.24)	(22.48)	(18.43)	(22.96)	(21.80)	(24.11)	(21.06)
Tract % mfg empl	12.16	12.66	11.76	11.69	12.02	12.86	11.33	11.54
	(5.37)	(5.21)	(5.14)	(5.57)	(5.24)	(5.21)	(4.83)	(5.33)
Δ Tract med HH inc					3.45	5.10	2.22	2.46
					(18.02)	(21.68)	(16.22)	(14.56)
Δ Tract pop dens					-0.212	-0.263	-0.121	-0.206
					(1.591)	(1.238)	(1.881)	(1.735)
Δ Tract % no HS					-2.24	-1.31	-2.18	-3.14
					(17.89)	(14.35)	(18.07)	(20.55)
Δ Tract % vacant					-0.218	-0.184	-0.391	-0.172
					(5.763)	(5.455)	(6.254)	(5.813)
Δ Tract % owner occ					5.06	7.81	1.80	3.95
					(29.05)	(29.29)	(32.21)	(27.05)
Δ Tract % mfg empl					0.241	0.498	0.118	0.055
					(5.078)	(5.060)	(5.222)	(5.021)

Source: Author's calculations based on 1) birth data from the Texas Center for Health Statistics (1995-2003) merged with 2) km^2-grid-cell-level RSEI scores corresponding to birth years and 3) data from the 2000 U.S. Census.
Notes: Each observation in this sample represents the change between siblings to a single mother. The level variables represent the level at the time of the older sibling's birth.

births of two siblings. Table 2.11 shows means for this sample, which we limited to mothers who did not change CBSAs and stratified by whether or not the mother moved between births. Not surprisingly, given secular declines in TRI emissions, we see that pollution exposure decreased between births on average for each group in the sample. Because the dependent variable is $log(tox_{b2}) - log(tox_{b1}) = log(tox_{b2}/tox_{b1})$, where $b1$ and $b2$ represent sequential births and tox is the logged RSEI score (plus 1), we can interpret it as roughly the percent difference in pollution exposure between the earlier and later births. In the sample overall, non-moving mothers experienced approximately three percent lower pollution exposure at the time of the later birth and moving mothers experienced a 12 percent lower pollution on average. Among non-movers, the average reduction for Black and Hispanic mothers was greater than for White mothers, but among women who moved, Black and Hispanic mothers saw smaller reductions than did White mothers.

We assumed that levels of mother and area characteristics, as well as changes in these levels, might help predict the change in pollution exposure from one birth to another. For non-movers, we see quite large differences across the race groups, with White mothers being older, better educated and more likely to be married at the time of the later birth, as well as less likely to have a marital status change between births. As in the first sample, White mothers also live in considerably "nicer" neighborhoods than minority non-moving mothers, in terms of income, education level of adults, and share of owner-occupied houses; White mothers also live in neighborhoods with more manufacturing employment. Among mothers who moved, Whites are still older, more educated, more likely to be married, etc., but the differentials between races in the mover subsample are smaller.

Table 2.12 shows the results from regressing the change in pollution exposure between births on mother and area characteristics, with fixed effects included to control for characteristics that do not vary within CBSAs. In each model, we controlled

Table 2.12. Predictors of change in toxic exposure between sibling births in Texas (CBSA fixed effects), 1995-2003.

	Did not move between births				Moved between births (same CBSA)			
	1	2	3	4	1	2	3	4
Mother Black	$-0.015^{\star}$	$-0.020^{\star}$	$-0.023^{\#}$	$-0.022^{\star}$	$0.087^{\star\star}$	$0.046^{\star\star}$	$0.029^{\star}$	$0.032^{\star}$
Mother Hispanic	-0.003	-0.007	-0.009	-0.010	$0.059^{\star\star}$	$0.025^{\star}$	0.017	$0.026^{\#}$
Age: teens		$0.027^{\star\star}$	$0.027^{\star\star}$	$0.026^{\star\star}$		$0.020^{\#}$	$0.018^{\#}$	0.013
Age: 30s		-0.002	-0.001	0.002		$-0.024^{\star\star}$	-0.013	-0.010
Age: 40s		$0.018^{\star}$	$0.020^{\star}$	$0.024^{\star\star}$		-0.014	-0.003	-0.010
Δ Ed: No HS $\rightarrow$ No HS		-0.009	-0.010	-0.012		$0.019^{\star}$	0.001	0.007
Δ Ed: No HS $\rightarrow$ HS		-0.016	-0.017	$-0.018^{\#}$		0.007	-0.005	-0.004
Δ Ed: No HS $\rightarrow$ BA		0.044	0.044	0.051		-0.013	-0.010	0.003
Δ Ed: HS $\rightarrow$ BA		-0.022	-0.021	-0.017		$-0.033^{\#}$	-0.014	-0.015
Δ Ed: BA $\rightarrow$ BA		-0.010	-0.008	-0.001		$-0.039^{\#}$	-0.011	-0.016
Δ Marital: No $\rightarrow$ No		0.004	0.003	0.002		$0.062^{\star\star}$	$0.036^{\star\star}$	$0.029^{\star\star}$
Δ Marital: No $\rightarrow$ Yes		-0.003	-0.003	-0.004		$0.030^{\star}$	0.016	0.010
Δ Marital: Yes $\rightarrow$ No		-0.001	-0.002	-0.003		$0.074^{\star\star}$	$0.05^{\star\star}$	$0.041^{\star}$
Tract med HH inc ($000)			-0.0003	-0.0014			$-0.0069^{\star\star}$	0.0006
Tract income squared			0.000000	0.000005			$0.000058^{\star\star}$	-0.000012
Tract pop dens				-0.0037				0.0145
Tract % no HS				0.0004				0.0005
Tract % vacant				-0.0005				-0.0041
Tract % owner occ				$0.0012^{\#}$				0.0003
Tract % mfg empl				-0.0011				-0.0013
Δ Tract med HH inc							$-0.0070^{\star\star}$	0.0021
Δ Tract income squared							-0.000008	0.000007
Δ Tract pop dens								0.013
Δ Tract % no HS								$0.007^{\star\star}$
Δ Tract % vacant								$-0.012^{\star}$
Δ Tract % owner occ								$-0.006^{\star\star}$
Δ Tract % mfg empl								0.013
Constant	0.030	0.030	0.031	0.029	-0.029	-0.045	-0.047	-0.034
Year controls	Y	Y	Y	Y	Y	Y	Y	Y
Δ Year controls	Y	Y	Y	Y	Y	Y	Y	Y
N	269270	269270	269270	269270	259086	259086	259086	259086
Adjusted R^2	0.044	0.044	0.044	0.045	0.022	0.023	0.035	0.058

Source: Author's calculations based on 1) birth data from the Texas Center for Health Statistics (1995-2003) merged with 2) km^2-grid-cell-level RSEI scores corresponding to birth years and 3) data from the 2000 U.S. Census.
Notes: Results from within-city (CBSA fixed effect) regressions where the dependent variable is the difference in log toxic exposure to mother m between sibling births. Sample includes Texas births to White, Black and Hispanic mothers who had at least two children in the sample and whose birth records could be geocoded and located within a Core Based Statistical Area (CBSA). Mothers who changed CBSAs between births were excluded. Standard errors (not shown) are clustered at the CBSA level and adjusted for heteroskedasticity. Regressions were performed on samples stratified by whether or not mother m moved between births. ** indicates statistically significant at < 0.01; * at < 0.05; $^{\#}$ at < 0.10. Variables are defined as described on p. 88.

for secular trends in emissions by including indicators for number of years between births as well as the year of the later birth. Looking at the first column group, we see that Black non-moving mothers experience a change in pollution roughly 2 percent lower than White mothers, a finding robust to the inclusion of controls. We included controls for mother and area characteristics here on the basis that places with higher socioeconomics status (where mothers with higher SES live) might have bigger decreases or smaller increases in pollution exposure, though we see that these control variables are generally small and statistically insignificant.

The second column group in Table 2.12 shows the results of an analogous model, using the sample of mothers who changed neighborhoods between births. The first subcolumn shows the results from a regression with race variables only, and we see that Black and Hispanic mothers who move between births experience smaller decreases or larger increases, on average, than White mothers who move. As we move rightward in the table, we see that including other SES variables attenuates the race effect, implying that residents with greater upward mobility, who are disproportionately White, are moving to relatively cleaner areas. The signs of the coefficients on the mother characteristic variables are just as we would expect. Increases in age at the later birth, higher education, and being or getting married are all correlated with greater reductions in pollution exposure than the alternatives. In subcolumn 3, we add variables for the second tract's income as well as the change in income from the first tract. We find both sets of variables to be statistically significant in the expected directions. Adding neighborhood income diminishes the size and significance of mother variables, showing, as we would expect, that the effect of mother SES on change in pollution is mediated through her ability to move to a nicer and less polluted neighborhood. Finally, in subcolumn 4, we add variables for both tract characteristics in the new neighborhood as well as the change between new and old neighborhoods. We find that the level values are not important predictors of the change in pollution,

but as expected, the change values are. As we would expect, they are similar in size and statistical significance to the level values in Column 4 of Table 2.9.

2.5 Discussion and Conclusions

The distributional measures and regression results reported here all point to the same basic conclusion: infants in Texas are not born into a land of equal opportunity, at least where industrial toxic air pollution is concerned. Rather, minority children are disproportionately exposed even before birth. Our results reflect what Ash and Fetter (2004) found for all residents at the national level; Black infants are more likely than others to be born into high-pollution cities, and within cities *and even within neighborhoods*, both Black and Hispanic infants are more likely to live on the wrong side of the environmental tracks [9].

Though our analysis includes only 68 CBSAs in Texas, we also find evidence, following Morello-Frosch and Jesdale (2006), that segregation is correlated with higher toxic exposure [82]. The segregation measures that are most predictive of higher exposure, however, are those highly correlated with percentage of Black residents in the city's population, making it difficult to separate the "effects" of segregation versus Black share. Either way, the higher pollution exposure in these cities, though still disproportionately borne by minority residents, affects everyone, providing support for the conclusion in Ash et al. (2010) that environmental injustice is bad for White folks, too [8].

In our analysis of changes in maternal pollution exposure between sibling births, we found that, relative to their initial location, White mothers who moved between births tended to move to relatively cleaner places than Black or Hispanic mothers who moved between births. We did not find evidence that minority mothers who did not move were subjected to bigger increases (or smaller reductions) than White mothers who did not move, at least over the 1995-2003 period. Thus, we do find some

evidence for the "move in" hypothesis, but our data covers too few years and is not appropriate for making conclusions about "siting."

These results do not imply an abstention of responsibility for disproportionate exposure among newborns; lack of information about the extent and effects of pollution exposure undermines the argument that parents who move to a relatively polluted place are behaving in an economically rational manner. Rather, these results provide useful information about what research and policy approaches might be most appropriate. If we want to truly understand how pollution came to be distributed the way it is, case studies, perhaps informed by the MSA-specific environmental justice statistics presented here, will be necessary. Wolverton (2009) concludes, using regression analysis, that the presence of other manufacturing firms and land/labor costs, rather than demographic characteristics like race, are the real drivers of location decision. But statistical correlation cannot show intent, and in the words of Pastor et al., "the real rationales for [factory] location will need to be uncovered by specific case studies" [88, p. 5].

We also need better data and more research on the effects of the industrial chemicals included in this analysis. Science has yet to determine exactly how many of these chemicals affect developing fetuses, but lack of proof should not be confused with proof of lack of effects. Suggestive evidence from a handful of chemicals, historical failure of the "innocent until proven guilty" approach, and the implication from epigenetic research that toxicants may affect gene expression all point to the importance of conducting more research and taking a precautionary approach in the meantime. Such research would be greatly facilitated by better data on ambient concentrations of air toxics. Though RSEI data is produced using a highly sophisticated model, models can only produce estimates. Air monitoring can supplement and help verify existing modeled concentration data.

Since industrial air pollution is only one of many environmental and social stressors more common in minority and low-income populations, Morello-Frosch et al. (2011) recommend that governmental bodies "use cumulative impact screening to map, characterize, and target vulnerable communities for interventions that improve existing conditions and prevent future harm" [84, p. 883]. Rather than focusing narrowly on pollutants, as encouraged by much existing environmental policy, cumulative impact assessments would acknowledge and account for the multiple, interactive stressors faced disproportionately by groups with low socioeconomic status. Such studies can inform policies to ameliorate existing inequities, or at least prevent their worsening. An example of such a policy, cited by Morello-Frosch et al., is the 2009 Environmental Justice Ordinance in Cincinnati, Ohio, which requires new or expanding industrial facilities to demonstrate that they will not cause a cumulative adverse impact to the health and environment of the community in order to receive a permit [84].

Finally, an obvious—albeit politically difficult—set of policy strategies include enforcing and strengthening existing pollution laws, regulating known toxics that are currently "regulated" only on right-to-know basis, and requiring companies (rather than the government or citizens) to prove that chemicals are safe before using them in production. Making polluting more costly in these ways can also spur advances in cleaner production, which would ultimately help ensure a clean and healthy environment for all people.

CHAPTER 3

USING RSEI DATA TO ASSESS THE EFFECTS OF TOXIC AIR POLLUTION ON INFANT HEALTH: AN EXPLORATION

3.1 Introduction

In the U.S., thousands of different chemicals are released into the air as byproducts of industrial production, though the health effects of breathing these substances are not well understood. According to the U.S. Environmental Protection Agency (EPA), no basic toxicity information is available for nearly half of the chemicals used in quantities of over 1 million pounds per year [54]. Further, information on the toxicity of these chemicals to developing fetuses and/or children—perhaps the most vulnerable groups in society—is available for only one-fifth of the chemicals. Some common pollutants are subject to legal limits, but others are regulated only on a "right-to-know" basis, in which information on emissions is made available to the public, so that citizens might use the information to pressure companies to improve their environmental performance or inform housing decisions. Such information may not be useful, however, when the effects of these pollutants are unknown. Thus, better research on the health effects of industrial pollutants is clearly warranted.

In this paper, we ask whether the Risk-Screening Environmental Indicators Geographic Microdata (RSEI-GM), which is based on data from the Toxics Release Inventory (TRI), itself a product of right-to-know legislation, might be used to explore the effects of some of these chemicals on infant health outcomes including birthweight, gestational age and infant death. A few existing studies have used TRI data to explore the effects of air toxics on birth outcomes, but these statistical studies use county-level

toxic exposure estimates that do not take into account within-county heterogeneity in their analyses [3, 39]. In addition, these studies do not account for emissions that travel outside of a county's boundaries, and they tend to lump different chemicals with presumably different physiological effects and levels of toxicity into the same measure. We are fortunate, at the University of Massachusetts Amherst, to have access to the EPA's RSEI microdata, which contains modeled concentration estimates of several hundred TRI chemicals for each square kilometer in the U.S. We match this data to geocoded birth and infant death records to assess the relationship between birth outcomes and the effects of several known developmental toxicants, including cadmium, epichlorohydrin, lead and toluene. Finding robust, negative relationships between these pollutants and birth outcomes would support the use of RSEI-GM in studies of the health effects of other TRI chemicals.

Like other studies in the statistical epidemiology literature, ours trades off precise exposure data in an experiment with few subjects for less-precise exposure data and a large sample. Ideally, research on infant health effects of pollution would assess exposure using biomarkers or perhaps ambient exposure data in the areas where pregnant mothers live. Our exposure data is based on annual, self-reported firm data that is modeled to approximate the average concentration level in a given area, so it is far from ideal. On the other hand, it is the best available data for many of these pollutants, and we have access to around three million geocoded birth records from the state of Texas between 1994-2003. We use a number of regression specifications, both parametric and semi-parametric, and we ultimately find that our results are not only unexpected but very sensitive to model specification. We conclude that the RSEI microdata, while appropriate for studying geographical distribution of air pollution more broadly, are probably not precise enough for infant health research.

3.2 Background

Before discussing what is known about the health effects of toxic air pollution, it is necessary to consider which of the thousands of substances present in the ambient environment deserve that designation. According to the U.S. Environmental Protection Agency (EPA), "toxic air pollutants, also known as hazardous air pollutants, are those pollutants that are known or suspected to cause cancer or other serious health effects, such as reproductive effects or birth defects, or adverse environmental effects"[52]. In the EPA's lexicon, toxic air pollutants are distinguished from what are termed "criteria" or "common" air pollutants—carbon monoxide, lead, nitrogen dioxide, ozone, particulate matter, and sulfur dioxide—more on the basis of their ubiquity in the ambient environment than on their hazardousness to health.[1] Of the 84,000 chemicals currently registered for commercial use with the EPA, 188 have been designated hazardous air pollutants [53, 52], though it is likely that many more are toxic to human health.

In this essay, we use a more expansive, but still not exhaustive, definition of toxic air pollutants: those substances tracked in the EPA's Toxics Release Inventory (TRI) and released into air. The TRI was created by the Emergency Planning and Community Right-to-Know Act (EPCRA) of 1986, passed in response to the deadly Union Carbide plant explosion in Bhopal, India and subsequent chemical release from a sister plant in West Virginia. EPCRA required manufacturing plants with more than ten full-time employees that either used or produced more than threshold amounts of some 300 toxic chemicals to report their annual emissions of these chemicals to the EPA for inclusion in the TRI. Since the first TRI was released in 1988, there have been several changes to reporting requirements. In 1995, the list of chemicals doubled; in 1998 the type of facilities required to report was expanded to include,

[1]Lead is also considered a toxic air pollutant and is tracked in the Toxic Release Inventory.

most notably, some electric utilities; in 2000, the thresholds for reporting of chemicals that persist in the environment were lowered and several chemicals were delisted. The current TRI toxic chemical list contains 581 individually listed chemicals and 30 chemical categories [54].[2]

Little is known about either the health effects of most toxic air pollutants or the biological mechanisms by which they might affect health, a gap frequently lamented in the epidemiological, medical and public health literatures. According to Suh et al. (2000), "relatively little has been done to characterize the concentrations, exposures, and health risks for most of the hazardous air pollutants (HAPs). Still less is known about the human health effects of HAP exposures at concentrations found in the ambient environment, as most of what is known has been obtained from occupational and animal studies" [105, p. 629]. In response to a 1997 Environmental Defense Fund report criticizing the lack of information about health effects of the chemicals most prevalent in U.S. production, the EPA published a report titled *What Do We Really Know About the Safety of High Production Volume Chemicals?*[47, 54]. In the 1998 report, the EPA revealed that no basic toxicity information was available for 43 percent of the 2,800 chemicals that are produced or imported in quantities of over 1 million pounds per year, and a full set of basic toxicity information was available for only 7 percent. Information on developmental or pediatric toxicity was available for only one-fifth of the chemicals at the time of publication, and more recent literature indicates that little progress has been made in the intervening decade [54, 63]. While the industrial air toxics listed in the TRI are not exactly synonymous with the 188 HAPs referred to in Suh et al. (2000) or the nearly 3,000 high production volume

[2]As described below, our analysis includes only those chemicals that are a) tracked during the entirety of our 1994-2003 sample period, b) emitted by industries required to report during the entirety of our sample period, and c) found in the vicinity of mothers in our sample, according to the RSEI model.

chemicals in the EPA report, the lack of toxicity information is common to all three overlapping groups.

Further research into the human health effects of industrial air toxics is clearly warranted, and infants are a good group to study for several reasons. In the absence of well-understood biological mechanisms connecting toxic pollution exposure to health outcomes, statistical studies showing consistent associations between exposure and response provide the next best evidence [76]. Unfortunately, identifying consistent connections between pollution exposure and adult health is often confounded by the cumulative effects of everything adults have breathed, eaten, and experienced throughout their lifetimes, many of which cannot, in statistical terms, be "controlled for." While fetal growth and infant health are influenced by cumulative maternal exposures, the potential for confounding effects is greatly reduced.

In addition to this practical concern, there are also compelling public health reasons for focusing on infants. As Landrigan et al. write in an article on health effects of pesticide exposure, "Infants and children are not little adults. They are uniquely vulnerable to environmental toxicants" [70, p. 257]. Fetuses and infants are uniquely vulnerable, according to the authors, for several reasons. First, because the central nervous system is not fully developed until at least six months after birth, toxic chemicals may disrupt nervous system development in fetuses and infants in a way not likely to affect older children or adults. Similarly, there may be critical windows of development during which exposure to toxics have particularly deleterious effects. Other reasons offered for the susceptibility of fetuses and infants have to do with their small size and young age. A given dose of pollution could be expected to have much larger effects in a seven-pound infant than in a 150-pound adult, for example, and because of their longer lifespans, infants simply have more years during which to develop chronic disease from early exposures.

A final reason we offer for studying infants is in response to the justification given in standard economic theory for disparities in the distribution of pollution and attendant health problems. According to this theory, rationally acting adults maximize utility subject to budget constraints, and one's budget, according to the neoclassical theory of distribution, is determined meritocratically according to his or her marginal contribution to production. If a person lives near large quantities of toxic air pollution, it is assumed that it was a utility-maximizing choice on her part, given her budget constraint, and made with perfect information about the costs of doing so. If that person develops cancer years later due to the toxic air pollution, it is further assumed that the present discounted value of those health effects were all included in her initial cost-benefit analysis. Importantly, it was her utility-maximizing *choice* to live there. The same argument, however, cannot be applied to fetuses and infants. Fetuses cannot choose to move away from pollution exposure, and even neoclassical economic theory can not justify inequities in health borne by infants born to mothers with low socioeconomic status.

Thus, more research on the effects of toxic air pollution on infants seems justified. With only a few exceptions, research on the effects of pollution on infant health outcomes has focused on criteria pollutants, for at least two reasons. Criteria pollutants are, by definition, more widespread in the ambient environment, and monitoring stations installed to comply with Clean Air Act regulations make estimating exposure to these pollutants relatively easy and accurate.

Four recent econometric analyses are the only studies, to our knowledge, to have used toxic release inventory (TRI) data to study infant health effects of toxic air pollution. Comprising the four are Currie and Schmeider (2009), which was published in the *American Economic Review*; Agarwal, Banternghansa and Bui (2010), published in the *Journal of Health Economics*; Carman (2009), a doctoral dissertation;

and Somov (2004), a master's thesis. Hereafter, I'll refer to the studies as CS, ABB, Carman and Somov, respectively.

Although these studies all employ regression analysis to study the effect of TRI releases on infant health outcomes, the studies vary enough in both method and data used to make results difficult to compare. (See Table 3.1.) First, both the infant health outcomes used as dependent variables and the TRI data used as explanatory variables vary across the studies. In terms of infant health outcomes, ABB, CS and Somov all study infant mortality. ABB also includes fetal mortality, and CS includes gestational age, birth weight, low birth weight (LBW, <2,500 grams) and very low birth weight (VLBW, <1,500 grams). Carman studies preterm birth (PTB, <37 weeks gestation) and LBW.

ABB and Somov use only aggregate toxic pollution measures. Somov uses all TRI releases (to any media), and ABB runs separate regressions for all TRI releases combined, releases by media (air, water or land), and releases by media and presumed health effect (carcinogen, developmental toxin or neither). CS focuses on TRI fugitive air releases, since "emissions that go up a smoke stack are more likely to be treated in some fashion (e.g., with scrubbers) and travel farther than those that do not" and therefore "should be less likely to affect those in the immediate vicinity of the plant"[3][39, p. 177]. CS also estimates separate models using progressively smaller groups of chemicals as explanatory variables: all fugitive releases of air toxics; a subset of 80 chemicals designated as developmental toxicants; subsets of heavy metals and volatile organic compounds (VOCs) within the developmental toxicants group; and toluene, lead, epichlorohydrin and cadmium separately. Carman studies fugitive and stack air releases of lead and toluene.

[3]For the definition of fugitive and stack releases from the RSEI methodology document, see p.144.

Table 3.1. Epidemiological studies of effects of toxic pollution on infant health

Study	Years	Location	Unit	Outcomes	Toxics	Exposure measure	Effects	Notes
Agarwal et al. (ABB, 2010)	1989-2002	U.S.	County-year, weighted by # births	Fetal mortality, infant mortality	All TRI releases; TRI grouped by media (air, water, land); TRI grouped by media and effect (carcinogens, reproductive, residual)	County releases(lbs)/cty land area(sq. mi.)	Carcinogenic air, residual air, and residual water had stat sig effects on infant mortality. No effects on fetal death.	Controls for emissions from mobile sources and non-TRI-reporting facilities.
Carman (2009)	2003	Allegheny County, PA	Individual births	PTB, LBW	Fugitive and stack air releases of lead and toluene	Releases weighted by distance from mother's home	No effects of lead or toluene exposure on PTB or LBW.	Categorical exposure variable (four quartiles). Fewer than 25 (85) LBW (PTB) observations in each quartile of lead and toluene releases.
Currie and Schmeider (CS, 2009)	1988-1999	U.S.	County-year, weighted by mean # births in county 1988-1999	Gestational age, birth weight, LBW, VLBW, infant mortality	Fugitive TRI air releases and subsets: Developmental releases; VOCs; Heavy metals; Toluene, epichlorohydrin, lead, cadmium	County releases(lbs)/cty land area(sq. mi.)	Large and stat sig effects of toxics (esp. toluene, cadmium and lead) on LBW, VLBW and infant mortality. Small effects on gestational age and continuous birth weight.	Specification check using stack instead of fugitive releases did not find stat sig effects on infant health outcomes.
Somov (2004)	1997	U.S.	County	Infant mortality	All on- and off-site TRI releases	County releases(lbs)/cty land area(sq. mi.)	TRI releases had stat sig effects on infant mortality.	Separate regressions with criteria pollutants, drinking water violations as pollution variable—not stat sig.

ABB, CS and Somov all use the county as the geographic unit of aggregation, with infant health outcomes aggregated to annual county means (for gestational age and birth weight) and rates (for fetal and infant mortality, LBW, VLBW and PTB). ABB and CS both use panels that cover more than 10 years of TRI and health outcome data, making their unit of analysis the county-year, while Somov conducts a cross-sectional analysis using data from TRI reporting year 1997. ABB and CS also weight their observations, by number of live births in the county-year, in the first case, and by the mean number of births in the county over the sample period, in the second. Since the infant health outcomes are measured in terms of county means and rates, the TRI releases chosen for analysis are adjusted to represent a measure of average exposure. These three studies all use the same exposure measure: TRI releases (measured in pounds) divided by county land area (measured in miles). This exposure measure (matched to county mean health outcomes) not only misses a great deal of variation within the county, it also pools together chemicals thought to vary in toxicity by seven orders of magnitude, except in the cases where CS analyzes individual chemicals separately. These analyses also implicitly assume that chemicals released by a facility in County X affect only County X, an assumption CS attempts to approximate by focusing on fugitive emissions, which are thought to stay closer to the emitting facility than stack emissions.

Carman differs from the other studies in geographic scope, unit of observation, and exposure measure employed. While the other studies include all counties in the U.S., Carman studies individual birth outcomes in Allegheny County, Pennsylvania. Carman estimates maternal exposure to toluene and lead by weighting pounds of releases by the inverse of the distance between the mother's house and the emitting factory. Also unlike the other studies, Carman uses a categorical exposure measure (exposure quartiles) instead of a continuous one (pounds per square mile).

Carman more precisely matches birth outcomes to toxic exposure than the other studies; surprisingly, perhaps, it is also the only study to find no negative effects of toxics on infant health outcomes. Controlling for race, marital status, education and smoking, Carman found that mothers in the second, third and fourth quartiles of toluene and lead exposure were not more likely than mothers in the first quartile to experience preterm birth or have a baby of low birth weight [26].

CS, on the other hand, finds that fugitive air releases of TRI chemicals have small, negative effects on birth weight and gestational age and relatively large effects on the probability of LBW, VLBW and infant mortality. Interestingly, CS finds that the very chemicals that Carman found to have no effect on LBW have a relatively large effect on LBW, VLBW and infant mortality. As one example, the results in CS indicate that a two-standard-deviation increase in toluene releases would increase the incidence of low and very low birth weight by 1.9 and 2.7 percent, respectively. Regarding the effect of lead, toluene, and cadmium on infant mortality, CS holds that "reductions in these three chemicals alone can account for about 3.9 percent of the reduction in infant mortality during the late 1980s and 1990s from 9.2 to 6.9 deaths per 1,000 live births" [39, p. 181].

What accounts for the differences between CS and Carman, and for Carman's surprising finding that lead and toluene exposure do not negatively affect infant health? This finding may be done to Carman's relatively small sample size (n=2,789 births), with a correspondingly low number of adverse health outcomes, or a possible lack of variation in pollution exposure between mothers in Allegheny County in 2003. Another explanation involves the exposure measures used. On one hand, the distance-weighted emissions measure used by Carman would seem to offer more precise exposure estimates than total county releases divided by land area. On the other, if Carman's estimates were off, the impact of the estimation error could have a greater effect than error in CS's ecological study. It is also possible that CS's county-

year-level pollution variables are correlated with other unobserved factors that affect infant health and the effect attributed to pollution is in fact the effect of one of these unobservables. CS includes fixed effects for county and year, though, so this explanation seems unlikely unless county characteristics changed significantly over the sample period.

Somov and ABB both found negative effects of air toxics on infant health outcomes. According to Somov's regression results, the elasticity of infant mortality with respect to total TRI emissions (measured in lbs/mi^2) is about 0.01. ABB separates TRI releases by media and health effect and finds that the categories of releases with statistically significant effects in the expected direction are carcinogenic air releases, residual (non-carcinogenic, non-developmental) air releases and residual water releases. According to the authors' estimates, the respective elasticities with respect to infant mortality are 0.0027, 0.0288 and 0.0043. Notably, the "developmental toxicant" release category does not appear to increase the likelihood of infant mortality, whether the medium is air, water or land. Like CS, these papers indicate that toxic air pollutants lead to adverse pregnancy outcomes, but they also pool chemicals with potentially very different physiological impacts and very different levels of toxicity. Still, ABB's findings do not inspire confidence in using toxicological theory to define appropriate groups of pollution variables.

Why did CS find that developmental toxicants negatively affect infant health while ABB did not? Both define developmental toxicants as the chemicals so designated by the California Office of Environmental Health Hazard Assessment, so there is no disagreement there [23]. Differences in health outcomes studied and the variable used to measure toxic exposure make the results difficult to compare directly. Only one result from a specification check in CS seems at all comparable to any result in ABB. In this specification check, CS estimates the effect of an additional pound per square mile of stack (versus fugitive) releases of developmental toxicants on infant

mortality and finds the effect to be small and statistically insignificant (coef=0.0248 and SE=0.0674). Likewise, ABB find the effect of all (stack+fugitive) air releases of developmental toxicants to have a small and insignificant effect on infant mortality (in their richest specification, coef=0.0010 and SE=0.0488). Thus, it seems that CS's focus on fugitive emissions, which ostensibly allows for better spatial matching between pollution and infant health outcomes, can explain the different findings. Specification differences may also have played a role: unlike CS, ABB includes a pollution squared term to allow for non-linearities in the dose-response function, as well as proxies for toxic releases from non-reporting facilities and from mobile sources.

The different findings among these papers are difficult to reconcile, and we will not attempt to do so here. We are skeptical, at any rate, of the usefulness of looking for infant health effects of groups of chemicals (e.g., all TRI releases) that may operate on different physiological systems and produce different effects. Still, we are interested in using previous research to inform our own study. The RSEI microdata could potentially be used to study the effect of hundreds of little-studied chemicals on infant health, but we first want to check whether estimated effects of more commonly studied chemicals are similar to those found in previous research. As a point of comparison, we prefer CS. Unlike Carman and Somov, it was published in a peer-reviewed journal, and unlike ABB, it reports the effects of several individual chemicals. Carman also reports on two individual chemicals, but we are more inclined to trust the results reported in CS, since they are based on more geographic and temporal variation. Thus, we choose to focus on the same four suspected developmental toxicants (cadmium, epichlorohydrin, lead and toluene) and three chemical groups (developmental, and subsets of this group that include volatile organic compounds and heavy metals) for which results are reported in CS. If this initial exploration using the RSEI-GM yields similar results, we will feel more confident in studying other chemicals and health outcomes using this data.

CS is not, of course, the first paper to present research on the effects of cadmium, epichlorohydrin, lead or toluene on developing fetuses. To identify other studies, we first turned to the EPA's Integrated Risk Information System (IRIS) database, which contains information on health effects of over 500 chemicals, synthesized from scientific papers by EPA staff, and divided into cancer and noncancer effects from oral and inhalation pathways [57]. Information from IRIS is the preferred source for the toxicity weights used in the RSEI model. Since we are interested in the developmental effects of air pollution, these—to the extent that they are known—should be listed in the inhalation, non-cancer effects section of each chemical's page in IRIS.

Of the four chemicals on our list, toluene has the lowest "uncertainty factor" in IRIS, due to a relatively large number of studies with human subjects, but the reviewed studies are mostly on neurological effects of occupationally-exposed workers. One sentence in the toluene section is devoted to developmental effects: "Animal studies have demonstrated reproductive and developmental effects of toluene at exposure levels higher than those used for the determination of the point of departure," the level of exposure where negative health effects are thought to begin [57]. Other research indicates that toluene exposure is linked to preterm birth, perinatal death, and growth retardation, though these studies have small sample sizes and are on women who inhaled large amounts of toluene resulting from sniffing glue or other inhalants [115, 6]. It seems that little is known about the developmental effects of toluene at levels present in the ambient environment.

We might expect the effects of lead on developing fetuses to be relatively well-studied, based on the widespread acknowledgment of lead's negative effects on cognitive development, but non-cancer effects of lead inhalation are "not assessed under the IRIS program" [57]. Further, providing an upper threshold for oral lead ingestion has proved politically difficult, with the EPA deferring this decision to the Center for Disease Control (CDC), and the CDC refusing to lower the blood level of concern

from 10 g/dL on the basis that "there is no evidence of a threshold below which adverse effects are not experienced, [so] any decision to establish a new level of concern would be arbitrary and provide uncertain benefits" [27]. Looking outside the body of research reviewed by governmental bodies turned up a relatively recent review of the effects of lead on reproductive and developmental outcomes, which concludes that maternal lead exposure during pregnancy is inversely related to fetal growth and high paternal exposures appear to increase an infants risk of preterm birth and low birth weight [13].

Like lead, non-cancer effects of cadmium inhalation are not assessed in IRIS. The toxicity weights for cadmium used in the RSEI model are taken from the California EPA's Consolidated Table of Approved Risk Assessment Health Values [22], but while categories for "developmental" and "reproductive" toxicants are listed in this table, cadmium is listed only as a kidney and respiratory toxicant, despite its inclusion on California's list of known developmental toxicants [23]. Several recent studies, however, have found evidence that cadmium exposure increases the risk of preterm birth and, as a result, low birthweight [85, 64, 107].

Non-cancer effects of epichlorohydrin inhalation are examined in the IRIS database, but the one "critical effect" listed is described as "changes in the nasal turbinates," a respiratory problem, and this finding is based on a 1979 study of rats [57]. Three human studies from 1980 and 1981 are cited that find no evidence of reduced sperm count due to epichlorohydrin exposure in occupationally-exposed males, but there seems to be little conclusive evidence of other developmental effects of exposure. A review of recent literature also turns up little research of human developmental effects, with one recent rat study finding a decrease in male fertility [100].

We take two main conclusions from this brief review. First, evidence from both CS and other literature indicates that lead, cadmium, epichlorohydrin and toluene inhibit fetal development, and we hypothesize that our analyses will support these findings.

Second, this evidence is far from conclusive and can mostly be blamed on lack of appropriate research. The four chemicals we focus on here are all considered high production volume chemicals, used or produced in quantities of over 1 million pounds per year. If even these chemicals are lacking solid information on developmental toxicity, the effects of less common chemicals are even more uncertain. More research is clearly needed.

3.3 Data and Methods

3.3.1 Methods

To assess the nature of the relationship between pollution and birth outcomes, we estimate several equations, beginning with a bivariate model:

$$y_{ijct} = tox_{ct} + Y_t + \varepsilon_{ijct} \tag{3.1}$$

where y_{ijct} is the birth outcome of infant i to mother j living in square-kilometer grid cell c at time t and the variable tox_{ct} is a measure of the pollution in grid cell c at time t. Our bivariate model also contains year dummies, Y_t.

Because pollution exposure is likely to be correlated with other area and mother characteristics that are related to birth outcomes, we include variables to help identify the effect of pollution:

$$y_{ijct} = tox_{ct} + \beta X_{ijt} + \gamma Z_k + Y_t + \varepsilon_{ijckt} \tag{3.2}$$

where X_{ijt} is a vector of mother and child characteristics, described on p.138, and Z_k includes Census tract median income and a neighborhood deprivation index measure.

To further control for area effects that may be correlated with both pollution exposure and birth outcomes, we also include a model that uses only variation within core-based statistical areas (CBSAs, described on p.141):

$$y_{ijct} = tox_{ct} + \beta X_{ijt} + \gamma Z_k + \phi_l + Y_t + \varepsilon_{ijcklt} \tag{3.3}$$

where ϕ_l is a CBSA fixed effect. We cluster standard errors at the CBSA level in this model, for obvious reasons, and also in Models (3.1) and (3.2), to make confidence intervals more comparable and in attempt to absorb some spatial autocorrelation.

Models (3.4) and (3.5) employ an identification strategy that attempts to isolate the effect of pollution on birth outcomes using only within-mother variation:

$$y_{ijct} = tox_{ct} + \beta X_{ijt} + \gamma Z_k + \zeta_j + Y_t + \varepsilon_{ijckt} \tag{3.4}$$

where ζ_j is a mother fixed effect. This model holds constant features of a mother that do not change or change slowly over time, such as race, genetic endowment, and dietary and lifestyle habits. Here, standard errors are clustered at the mother level.

Our final model also includes mother fixed effects but only includes mothers who did not move between births:

$$y_{ijct} = tox_{ct} + \beta X_{ijt} + \zeta_j + Y_t + \varepsilon_{ijct} \tag{3.5}$$

In addition to using a different sample, Model (3.5) differs from Model (3.4) in not including tract-level variables, since, as described below, our measures of these do not vary over the years in our sample.

In our initial estimations, the variable tox_{ct} is the estimated concentration of pollution in cell c at time t, and we implicitly assume a linear relationship between concentration and birth outcome. In subsequent estimations, we include squared and cubed pollution terms, to allow the relationship to be more flexible, as well as categorical pollution terms.

To impose even fewer restrictions on the nature of the pollution-outcome relationship, we also employ a semiparametric regression method. Because the size of our dataset prevented us from using standard procedures for semiparametric analysis,

such as local linear regression, we devised our own approach. First, we order our observations by pollution exposure, making the difference in exposure between subsequent observations approximately zero, and we first difference all other variables. We then regress the first-differenced outcome variables on the first-differenced control variables to "clean" the outcome variables of the effects of the controls and save the residuals. Finally, we want to visually examine the relationship between pollution exposure and outcome residuals, but because there are so many observations, we reduce the number of points by plotting the mean residual outcome in each pollution percentile against the right endpoint of the associated percentile.[4]

The method just described is analagous, in terms of control variables employed, to parametric model (3.2) above. We also employ the semiparametric analog to the non-moving mother fixed effects model, represented in equation (3.5) by subtracting the "mother mean" value from all variables before following the steps outlined above. In the plots showing results (pp.167-176), the mother-mean-differenced pollution exposure values have been retained on the x-axis, as a reminder of the mean differencing and to show the extent of variation in the pollution variable, but the mean-differenced outcome residuals have been shifted up by a constant to make the y-axis range correspond to the range in the non-mean-differenced plots.

3.3.2 Data

Our birth data come from a dataset of all 1994-2003 births in Texas that are matched, if applicable, to infant death records. This data was generously provided by the Texas Center for Health Statistics. Three of our dependent variables—gestation in weeks, birthweight in grams and internal infant death (death at less than one year of age from health-related causes)—come directly from this data. Gestational

[4]Unfortunately, time limits and technical difficulties prevented us from carrying out the bootstrapping method we devised to obtain confidence intervals, so the plots resulting from this method provide only suggestive evidence of pollution-outcome relationships.

age was calculated based on date of last menstrual period (LMP) and the clinical estimate of gestation was used in the small share of cases in which LMP was not reported. Our other dependent variables—preterm birth (PTB), low birthweight (LBW) and very low birthweight (VLBW)—are calculated based on gestational age and birthweight. These variables are defined in the usual ways: PTB is birth at less than 37 weeks of gestation, LBW is birthweight below 2,500 grams and VLBW is birthweight below 1,500 grams. With the exception of PTB, this is the same group of dependent variables used in Currie and Schmedier (CS, 2009).

The categorical control variables in our analyses are also based on birth record data, including maternal age, educational attainment, marital status and smoking behavior, as well as child sex, parity and information on prenatal care (PNC) initiation timing and frequency. The definition of most control variable categories can be inferred from Table 3.3, but the PNC variables require some explanation. *Late PNC initiation*, *few PNC visits* and *many (excessive) PNC visits* are all defined as in Kotelchuck's Adequacy of Prenatal Care Utilization Index [69]. *Late PNC initiation* takes on the value 1 if PNC began in month 5 or later, corresponding to the "inadequate" and "intermediate" categories in Kotelchuck's Adequacy of Initiation of Prenatal Care Index, and 0 otherwise. The variable *few PNC visits* is coded as 1 if a mother received less than 80 percent of the expected visits for a mother who began PNC when she did, corresponding to the "inadequate" and "intermediate" categories in Kotelchuck's Adequacy of Received Prenatal Care Services Index. *Many PNC visits* takes on the value 1 if a mother received 110 percent or more of the expected number of visits, corresponding to the "adequate plus" category in Kotelchuck's Adequacy of Received Prenatal Care Services Index. For all control variables, we include a "missing" category to retain maximum statistical power.

The birth records also provide several other useful pieces of information. If a pregnancy spans more than one calendar year, we weight pollution concentration

values from the two years by the share of the pregnancy that happened in each year before adding them, to better match women to annual pollution data. To produce these "year weights", we use information on the date of birth as well as gestational length. We also use birth year as a control variable in our regressions. In some specifications, we use mother identification numbers, which were generated at the Texas Center for Health Statistics based on social security numbers, to track births to the same mother. Finally, we use latitude and longitude coordinates, representing the mother's home address location, to merge birth data with pollution data from the Risk-Screening Environmental Indicators Geographic Microdata (RSEI-GM) and tract-level data from the 2000 U.S. Census.

The RSEI microdata contain estimates of location-specific exposure to toxic air pollutants emitted by industrial plants across the United States [50]. RSEI uses information on releases of several hundred chemicals from thousands of facilities, reported in the Toxics Release Inventory (TRI). Each year, facilities are required to report to the EPA on deliberate and accidental releases of toxic chemicals into air, surface water, and the ground and on transfers to offsite facilities. To make the TRI data more meaningful, the RSEI model estimates local concentrations by incorporating information on the fate and transport of releases. Fate and transport information comes from a plume model that accounts for chemical decay rates, stack heights, exit-gas velocities, average temperature and prevailing winds. For each air release (each facility x chemical combination), RSEI estimates the concentration in each square kilometer of a 101-km by 101-km grid centered on the releasing facility.

To construct our pollution measures, the RSEI concentrations for individual chemicals are added across facilities to characterize the total exposure to a given chemical in the square kilometer grid cell in which the pregnant mother resided during the year(s) of her pregnancy. We also use some aggregate measures of chemical groups, and in these cases, concentrations are added across chemicals to produce the pollution

measure. In some specifications, we apply RSEI toxicity weights before aggregating chemicals, on the basis that not all chemicals are equally toxic. Toxicity here refers to chronic human health effects from long-term exposure, including—in addition to reproductive and developmental toxicity—potential to cause cancer and neurotoxicity. The RSEI toxicity weights are based on a peer-reviewed methodology, taking into account the single most sensitive chronic human health endpoint (cancer or non-cancer). As noted above, the EPA's IRIS database is the preferred source of toxicity information, and developmental toxicity was not the most sensitive chronic human health endpoint for any of the four chemicals we focus on.

As mentioned above, our focus chemicals are toluene, epichlorohydrin, lead and cadmium, since effects of these four chemicals on birth outcomes are reported in CS. Finding similar effects will build confidence in using RSEI microdata to study the relationship between other chemicals and birth outcomes. The chemical groups we use are also those reported in CS, including the group of developmental toxicants and subsets of the developmental list that include VOCs, in one case, and heavy metals in another. The VOCs group includes benzene, carbon disulphide, dibromoethane, epichlorohydrin, ethylene oxide and toluene. The heavy metals are arsenic, cadmium, lead and mercury. These chemicals were chosen in CS "on the basis of frequency of releases and known toxicity" [39, p.178]. The developmental chemicals come from a list of developmental and reproductive toxicants published by the California Office of Environmental Health Hazard Assessment (OEHHA), which is produced in accordance with Proposition 65 and requires the governor to publish a list of chemicals known to the state to cause cancer or reproductive toxicity at least once per year [23]. While 82 chemicals reported to the TRI are currently identified as developmental or reproductive toxicants on the OEHHA list, only 22 of these were both in the TRI list during our sample period and found in the areas where our mothers lived; these 22

chemicals comprise our developmental toxicants group. (See Appendix Table C.13 for a list of these chemicals.)

The birth record and pollution dataset is then merged with tract-level data from the 2000 U.S. Census, again using mother home address information. Census tracts generally contain between 2,500 and 8,000 residents and are designed, at least initially, to be roughly homogeneous with respect to population characteristics, economic status, and living conditions [111]. The term Core Based Statistical Area (CBSA) came into use in the 2000 census and refers collectively to Metropolitan Statistical Areas, which must have at least one urbanized area of 50,000 or more residents, and Micropolitan Statistical Areas, which must have at least one urban cluster of at least 10,000 but less than 50,000 residents. CBSAs are comprised of whole counties, and tracts do not cross county boundaries. There are 68 CBSAs in Texas and therefore 68 CBSA fixed effects included in our estimation of Model (3.3).

From the census data, we take tract-level median household income and variables for our tract-level deprivation index. The deprivation index, defined as in Messer et al. (2006), is intended to gauge relative deprivation more completely than a single measure such as income [79]. Tract-level variables included in the deprivation index are percent of males in management, percent of households with more than one person per room, percent of individuals with 1999 income below the federal poverty line, percent of families with a female headed household and dependent children, percent of households with income less than $30,000, percent of households with public assistance income, percent unemployed, and percent of adults with no high school education. To create the index, principle components analysis was used to infer the contribution of each of these variables to an underlying deprivation variable, with the factor loadings used to weight each variable's contribution to the tract-level neighborhood deprivation score. The deprivation index was then standardized across tracts to have a mean of 0 and standard deviation of 1. Though they are highly correlated

($\rho = -0.85$), including both median income and the deprivation index in our regressions is part of our strategy to identify the effects of pollution, since the correlation between pollution and unobserved socioeconomic status variables can confound correct identification. Unfortunately, the temporal variation in our birth records and RSEI measures is not matched in the 2000 census data, but since neighborhood characteristics have been known to change slowly, we feel confident that the census data provide a reasonable approximation of neighborhood conditions in other birth years.

The full sample of Texas births from 1994-2003 includes 3,532,103 observations. From this universe, we kept singleton births with non-missing values for birthweight, gestation and age and probable gestational ages and birthweights.[5] These exclusions left 3,413,392 records. From these records, we also excluded approximately 13 percent that were not geocoded, since they could not be matched to pollution data, and another 3 percent that were not located in CBSAs, since living in very rural areas might imply different access to resources and health services. The full sample used in our analyses, then, has 2,868,306 observations.

For reasons mentioned above, our dataset is unique and potentially valuable for pollution-health research for a number of reasons. The geographical specificity of both the birth records and the RSEI data allows us to match health outcomes to toxic pollution concentrations in a much more precise way than has been possible in previous research. Though we begin with an initial exploration of a handful of chemicals, the RSEI database contains information on more than 500 little-researched toxic substances. Further, we have the ability to track mothers throughout the dataset and therefore use mother fixed effects, which should allow us to more precisely identify the effect of pollution.

---

[5]Following previous research, we excluded births with recorded birthweights of less than 500 grams or recorded gestational ages of less than 22 weeks.

While seemingly the best available data for our purposes, our dataset also has a number of shortcomings. First, birthweight, gestational age and infant death incidence are hardly the only three health outcomes that may be affected by pollution exposure, but they are readily available and reliably measured in general. If this method of analysis seems promising, matching RSEI pollution data to data from birth defects registries could help expand our knowledge of pollution effects on infant health.

A second category of potential problems are those related to exposure misclassification, which can happen for a variety of reasons related to both the birth data and the RSEI data. First, it has been estimated that a fairly large share of women (20-30 percent) move during pregnancy [25], and even among those who do not, we lack information on how much time pregnant women spend near their home, how much time is spent outdoors, and how protective the homes are against the invasion of ambient air pollution. Unfortunately, there is little we can do to correct for these data shortcomings.

In addition, the pollution concentration data is modeled based on annual, self-reported data and does not account for mobile or small point sources. Further, the chemicals and industries required to report to the TRI changed over the years in our sample. Again, there is little we can do to address most of these shortcomings, though we do attempt to address some of them. In some specifications, we exclude very high concentrations that may be an artifact of misreporting, though we have no way of distinguishing these "outliers" from truly high concentrations. For babies whose time in utero spanned two calendar years, we weight the annual estimated RSEI concentrations from these two years by the share of each year the mother was pregnant. While we cannot account for pollution that is not reported to the TRI, we do report the correlations between 2002 RSEI and National Air Toxics Assessment (NATA) tract-level concentrations for those pollutants that are available in both

databases in Appendix Table C.13.[6] To maintain consistency in our pollution data panel, we exclude chemicals and contributions from industries that were not reported to the TRI during all years in our sample (1993-2003). Finally, because the quality of the dispersion model will affect the quality of the concentration estimates, we use only fugitive releases in one specification, since this type of release is subject to many fewer modeling assumptions [50, pp.39-47].[7]

The third set of problems are related to identification. Separating the effect of pollution from the effect of unobserved mother and area characteristics that may be correlated with birth outcomes is a major challenge for this type of research. Unfortunately, birth data lacks information about a mother's income; this shortcoming is why we include neighborhood socioeconomic status measures. Using mother fixed effects to control for slow-changing mother characteristics would seem to attenuate the identification issue, but doing so presents another potential problem. It is often the case that second births have higher birthweight and longer gestation than first births. If mothers are systematically moving to cleaner areas between births, the improvement in birth outcomes may be mistakenly attributed to pollution reduction, though we do control for parity.

To check whether variation in pollution is correlated with mobility in our sample, we first estimated actual within-family variation in the four pollutants in our study and then compared it to what the within-family variation would have been if the

---

[6]NATA concentration data is also modeled, but the underlying data comes from different sources and is considered to be more comprehensive. Unfortunately, we cannot use it for our research because it is available for only a few years and different versions are not comparable.

[7]According to the RSEI methodology manual, "Stack (or point) air releases include releases to air through stacks, confined vents, ducts, pipes, or other confined air streams, and represent the majority of air releases (87% of on-site air releases). Fugitive releases to air include all other on-site air releases, including leaks, evaporation from surface impoundments, and releases for building ventilation systems."

mother had stayed in the location where she was first observed.[8] While the actual and non-moving variation was quite similar for lead (0.0025 vs. 0.0021), it was quite different for toluene (0.636 vs 0.271), epichlorohydrin (0.043 vs 0.034) and cadmium (0.00025 vs. 0.00016), which suggests that results from the full mother fixed effects sample may not be trustworthy.

Thus, our preferred specification, *a priori*, uses non-moving mother fixed effects as represented in Model (3.5). Because the pollution distributions have long right tails, we also prefer the sample with the top one percent of observations removed. This specification would seem to best address potential correlation between pollution and unobserved influences on infant health, as well as problems with outliers. Where space allows, we also present estimates from Model (3.5) without outliers removed, lest these are providing important information, as well as estimates from Model (3.3), in case the exclusive use of intertemporal variation in the non-moving mother fixed effects version is too limiting.

Finally, a note on statistical power is useful at this point. In econometric analyses, avoiding Type I error, or incorrectly rejecting the null hypothesis when it is true, is generally given precedence over avoiding Type II error, or incorrectly failing to reject the null when it is false. Depending on the subject matter, however, this tradeoff may be more or less appropriate. As Moore argues in *Children and Pollution*, having a very low tolerance for Type I error is inappropriate when it comes to interpreting statistical studies of the effects of pollution on health [81]. A precautionary principle would imply allowing relatively more Type I error and relatively less Type II error; that is, it would imply being more worried about inaccurately calling a dangerous chemical safe than about incorrectly identifying a safe chemical as dangerous. It

[8]We borrow this technique from Currie et al. (2009), which also included mother fixed effects in its regression analyses.

seems, however, that both convention and corporate power result in a general failure to err on the side of precaution and the insistence on very low Type I error probability.

If our basic null hypothesis is that a given chemical does not affect infant health outcomes, but that chemical *does* have an effect in reality, what is the probability that we will correctly reject the null? This probability, also known as the statistical power of our test, depends not only on the level of Type I error we are willing to allow, but also the sample size and the true effect size. For example, based on a mean birthweight of 3,330 grams, we find that we would need a sample of around 75,000 observations to detect the effect of a chemical that reduced birthweight by 5 grams with 0.80 statistical power, allowing a relatively large 10 percent probability of Type I error. To detect the effect of a chemical that increased the probability of LBW by 1 in 1,000 from a baseline rate of 6 percent with 0.80 power, we would need a sample of around 350,500. As shown in Table 3.2, our smallest sample contains 621,559 observations, which exceeds the bar set in these hypothetical examples. Our effective sample is smaller, however, because multiple mothers are assigned to the same grid-cell-year pollution concentrations and because we use only within-CBSA and within-mother variation in some cases. As an example, the 621,559 mothers in our non-moving mother sample represented 164,269 unique cell-year pollution observations. Thus, effects that we may consider relatively large by public health standards (a 1 in 1,000 increase in LBW is nearly a two percent increase) may be too small to be picked up in our statistical tests. Since we are unable to obtain a larger sample, we proceed with our tests, keeping in mind that we may want to be lenient with respect to Type I error and that our statistical power may not be large enough to pick up small effects.

3.4 Results

Table 3.2 presents summary statistics for the samples used in our analyses. The first column displays means from the sample containing all singleton births in Texas

Table 3.2. Sample Means

	All Singleton	Geocoded and in CBSA	> 1 sibling	> 1 sibling and no move
N	3,413,392	2,868,306	1,269,917	621,559
Outcomes				
Gestation (wks)	39.14	39.14	39.09	39.10
	(2.35)	(2.34)	(2.33)	(2.28)
Birthweight (g)	3325.2	3326.7	3330.4	3352.9
	(552.3)	(552.7)	(547.1)	(544.8)
Preterm Birth (PTB)	0.089	0.089	0.090	0.086
Low Birth Weight (LBW)	0.059	0.059	0.057	0.052
Very Low Birth Weight (VLBW)	0.0092	0.0092	0.0086	0.0081
Infant Death	0.0040	0.0040	0.0049	0.0051
RSEI concentration measures				
All releases ($\mu g/m^3$)		1.117	1.124	1.070
		(4.825)	(4.963)	(4.746)
Developmental ($\mu g/m^3$)		0.197	0.196	0.183
		(1.229)	(1.246)	(1.133)
VOCs ($\mu g/m^3$)		0.154	0.156	0.146
		(1.105)	(1.137)	(1.036)
Heavy metals ($\mu g/m^3$)		0.000888	0.000784	0.000756
		(0.00550)	(0.00485)	(0.00480)
Toluene ($\mu g/m^3$)		0.109	0.111	0.103
		(1.005)	(1.039)	(0.934)
Epichlorohydrin ($\mu g/m^3$)		0.00166	0.00164	0.00174
		(0.0777)	(0.0753)	(0.0888)
Lead ($\mu g/m^3$)		0.000722	0.000650	0.000632
		(0.00426)	(0.00388)	(0.00390)
Cadmium ($\mu g/m^3$)		0.0000319	0.0000258	0.0000259
		(0.000525)	(0.000469)	(0.000488)
Demographic characteristics				
Mother age	26.03	26.19	25.79	26.87
	(6.07)	(6.10)	(5.84)	(5.98)
Mother yrs of ed	12.02	12.12	12.61	12.93
	(3.06)	(3.08)	(2.66)	(2.76)
Mother married	0.69	0.70	0.70	0.76
Mother smoking	0.070	0.122	0.132	0.122
Child female	0.489	0.489	0.489	0.488
Birth order	2.045	2.034	2.175	2.225
	(1.211)	(1.203)	(1.200)	(1.226)
Late PNC initation	0.410	0.400	0.377	0.348
Few PNC visits	0.260	0.257	0.229	0.219
Many PNC visits	0.323	0.323	0.337	0.337
Tract med HH inc ($000)		41.529	42.783	44.843
		(20.28)	(21.20)	(22.85)
Tract deprivation index		0.0968	0.0257	-0.0273
		(1.013)	(1.003)	(1.038)

Source: Author's calculations based on 1) birth data from the Texas Center for Health Statistics (1994-2003) merged with 2) km^2-grid-cell-level RSEI concentrations and 3) tract-level data from the 2000 U.S. Census.
Notes: PTB is birth at < 37 weeks gestation; LBW is birthweight < 2500 g; VLBW is birthweight < 1500 g; Infant death is internal death within first year. The RSEI concentration measures shown here are weighted averages of the annual measures for the year(s) in which the child was in utero. Following Currie and Schmeider (2009), the VOCs we examine are benzene, carbon disulphide, dibromoethane, epichlorohydrin, ethylene oxide and toluene. The heavy metals are arsenic, cadmium, lead and mercury. Late PNC initation is Inadequate or Intermediate ranking on Kotelchuck's prenatal care initiation index; Few PNC visits is Inadequate or Intermediate ranking on Kotelchuck's expected visits index; Many PNC visits is Adequate Plus ranking on Kotelchuck's expected visits index [69]. Deprivation index variables are described on p.141.

in the years 1994-2003. Because some of these records are not geocoded and therefore cannot be matched to pollution measures, we do not use this sample in our analyses, but we include summary statistics here to check the generalizability of the CBSA sample. We see that the birth outcome means from the CBSA sample are nearly identical to those from the full sample. Mothers in these two samples are demographically similar as well, though we observe that mothers in the CBSA sample are considerably more likely to smoke during pregnancy. We also observe that the rates seem somewhat low; the 2000 rate of LBW in Texas was reported as 7.4 by the U.S. Bureau of Vital Statistics and the infant death rate was 5.8 deaths per 1,000 live births [68]. However, apparent underreporting does not differentially affect the geocoded sample, and we would expect this to bias our results downward.

Compared to the CBSA sample, the group that includes births with at least one sibling in the sample (used in the mother fixed effects specification) has a slightly lower mean probability of LBW and higher mean probability of infant death. The mothers in this sample are a bit younger on average, though more educated and living in tracts with a lower deprivation index score.

The final sample, which includes births with at least one sibling to mothers who did not move between births, is somewhat less similar to the CBSA sample. Mean birthweight is more than 25 grams higher and probability of LBW and VLBW are correspondingly lower. Surprisingly, the rate of infant death is markedly higher, at 5.1 in 1,000 compared to 4.0 in the CBSA sample. Mothers in this sample appear to have higher socioeconomic status than those in the previous two samples; they are older, more educated, more likely to be married and living in tracts with higher median income and lower deprivation index scores.

Table 3.3 is included to show how the regressions are specified, since coefficients on control variables are suppressed in subsequent tables. We show results for the linear probability model regression of LBW on epichlorohydrin. The five columns correspond

Table 3.3. Low birthweight - epichlorohydrin regressions

	(3.1) Bivar	p-val	(3.2) Multivar	p-val	(3.3) CBSA FE	p-val	(3.4) Mom FE	p-val	(3.5) No Move FE	p-val
Epichlorohydrin conc	0.0167	(0.4104)	-0.0518	(0.2406)	-0.0524	(0.0002)	0.2860	(0.4237)	0.1867	(0.7243)
Child female			0.7792	(0.0000)	0.7789	(0.0000)	0.9984	(0.0000)	0.8878	(0.0000)
Birth order 1			1.9148	(0.0000)	1.9222	(0.0000)	1.9485	(0.0000)	1.8890	(0.0000)
Birth order 3			-0.1019	(0.0080)	-0.0998	(0.0097)	0.1137	(0.1293)	0.2595	(0.0145)
Birth order ≥ 4			0.4523	(0.0000)	0.4603	(0.0000)	0.4515	(0.0006)	0.7584	(0.0001)
Mother age < 19			0.8538	(0.0000)	0.8585	(0.0000)	0.3647	(0.0026)	0.0968	(0.6284)
Mother age 25-34			0.3247	(0.0002)	0.3448	(0.0000)	0.0907	(0.3482)	0.0989	(0.4989)
Mother age ≥ 35			2.0482	(0.0000)	2.0852	(0.0000)	0.1704	(0.3343)	-0.0711	(0.7634)
Mother ed $<$ HS			-0.0657	(0.6166)	-0.0731	(0.5748)	-0.0397	(0.7584)	0.1002	(0.6308)
Mother ed $>$HS, $<$ BA			-0.3745	(0.0000)	-0.3685	(0.0000)	-0.1412	(0.2147)	-0.1277	(0.4471)
Mother ed $\geq$ BA			-0.9581	(0.0000)	-0.9554	(0.0000)	0.0313	(0.8524)	0.2381	(0.3132)
Mother unmarried			2.0379	(0.0000)	1.9787	(0.0000)	0.5162	(0.0000)	0.1807	(0.2719)
Mother smoking			4.1636	(0.0000)	4.0682	(0.0000)	1.7595	(0.0000)	1.3034	(0.0000)
Late PNC initiation			-0.1239	(0.5288)	-0.1480	(0.4507)	0.1946	(0.0014)	0.2421	(0.0062)
Few PNC visits			2.0708	(0.0000)	2.0477	(0.0000)	1.5715	(0.0000)	1.5707	(0.0000)
Many PNC visits			6.2957	(0.0000)	6.3786	(0.0000)	3.7831	(0.0000)	3.6398	(0.0000)
Tract med HH inc ($000)			-0.0188	(0.0018)	-0.0182	(0.0001)	0.0040	(0.2484)		
Tract deprivation index			0.3186	(0.0748)	0.4562	(0.0016)	0.1517	(0.0563)		
1995	0.0478	(0.5943)	0.0087	(0.9109)	0.0114	(0.8842)	-0.0430	(0.7204)	0.1331	(0.4394)
1996	0.0251	(0.8385)	0.0282	(0.8054)	0.0290	(0.8010)	-0.0164	(0.8859)	0.0501	(0.7555)
1997	0.0630	(0.6364)	0.0498	(0.7039)	0.0502	(0.7033)	-0.0240	(0.8419)	-0.1082	(0.5302)
1998	0.1697	(0.0923)	0.1889	(0.0951)	0.1911	(0.0986)	0.0717	(0.5774)	-0.0682	(0.7156)
1999	0.1192	(0.2450)	0.0504	(0.4957)	0.0496	(0.5152)	0.1723	(0.2162)	-0.0459	(0.8222)
2000	0.0866	(0.4674)	0.0460	(0.5931)	0.0499	(0.5831)	-0.1077	(0.4782)	-0.3943	(0.0772)
2001	0.2366	(0.0645)	0.2642	(0.0078)	0.2711	(0.0111)	0.0913	(0.5834)	-0.1509	(0.5387)
2002	0.3163	(0.0215)	0.3223	(0.0279)	0.3381	(0.0282)	0.1153	(0.5303)	-0.2769	(0.3100)
2003	0.4926	(0.0131)	0.4363	(0.0407)	0.4461	(0.0411)	0.2615	(0.1920)	-0.1769	(0.5557)
Constant	5.7182	(0.0000)	0.8761	(0.0004)	0.8538	(0.0000)	2.3856	(0.0000)	2.2791	(0.0000)
N	2,868,306		2,868,306		2,868,306		1,269,917		621,559	
Moms w/ var							26,280		11,503	

Source: Author's calculations based on 1) birth data from the Texas Center for Health Statistics (1994-2003) merged with 2) km^2-grid-cell-level RSEI epichlorohydrin concentrations and 3) tract-level data from the 2000 U.S. Census.
Notes: Dependent variable is low birthweight. Linear probability model regressions: OLS for models (3.1) and (3.2); within -CBSA, -mother and -non-moving mother regressions used in models (3.3), (3.4) and (3.5), respectively. Standard errors adjusted for heteroskedasticity and clustered at CBSA level in models (3.1)-(3.3) and at mother level in models (3.4) and (3.5). All RSEI epichlorohydrin concentration measures used here are weighted averages of annual measures for the year(s) in which the child was in utero. Control variables are defined as described on p.138. For models (3.4) and (3.5), "moms w/ var" indicates the number of mothers in the sample who had variation in both epichlorohydrin exposure and infant death.

to the models as described in Section 3.3.1. All coefficients are multiplied by 100; thus, the coefficient on epichlorohydrin in the bivariate model indicates that a 1 μg/m^3 increase in epichlorohydrin corresponds to a 0.0167 percentage point increase in the probability of being born LBW. As is true for other birth outcomes and chemicals, as we will discuss below, the estimated coefficient and its statistical significance is not stable across models. Including control variables (Column 2) and CBSA fixed effects (Column 3) both produce a coefficient of approximately -0.052, which is statistically significant at conventional levels in the CBSA fixed effect model. Using only within mother variation (Column 4) and within non-moving mother variation (Column 5) changes the sign of the coefficient back to positive, though these estimates are not statistically significant.

Coefficients on the control variables are generally of the expected signs and statistically significant. Females and firstborns are more likely to be LBW than males and second births, as are births to smokers and mothers who had relatively few or many prenatal care visits. In those models that utilize between-mother variation in control variables (Columns 2 and 3), age, education and marital status are also reliable predictors of LBW probability, with teen, older and unmarried mothers more predisposed to have a LBW child and mothers with above a high school education less so. Neighborhood median income and deprivation index scores are also related to LBW in the expected ways, with higher income and lower deprivation being associated with lower likelihood of LBW.[9] Finally, coefficients on the year dummies show a general (though non-monotonic) trend toward higher incidences of LBW over the sample period, especially in the final three years.

At the bottom of Table 3.3, below the sample size figure, is a row titled "moms w/ var." While the mother and non-moving mother fixed effect samples include 1,269,917

[9]Because income and deprivation index variables come from the 2000 U.S. census and do not vary across the years in our sample, they are not included in the non-moving mother regressions.

and 621,559 observations, respectively, it is often the case that a much smaller number of mothers have between-birth variation in both the birth outcome and the chemical in question. This is especially true in the case of relatively low-probability outcomes such as VLBW and infant death. Here, we see that the mother fixed effects regression included 26,280 mothers who had between-birth variation in both LBW incidence and epichlorohydrin exposure and the non-moving mother sample had 11,503 such mothers. We include an indicator of this type of variation with most regression coefficients from the mother fixed effect regressions.

Because we chose the chemicals and chemical groups we did based on results from Currie and Schmeider (CS, 2009), we present in Table 3.4 a comparison between results from that paper and our own. Reprinted in the upper left quadrant of the table are results presented in CS, though coefficients have been adjusted to show the effect of a one standard deviation increase in the chemical or chemical group listed, and coefficients on LBW, VLBW and infant death have been multiplied by 100.[10] Every coefficient except one (the coefficient on lead in the infant death regression) is of the expected sign, with chemical exposure associated with shorter gestation, lower birthweight, and higher probabilities of LBW, VLBW and infant death. Most of these estimates are also statistically significant, and as noted in CS, many represent sizable effects.

As described above, our data and methods are intentionally different in a number of ways from those used in CS. Ideally, however, we would present the effect of the same size increase used in CS, based on our estimates. The nature of the data makes this difficult, since our unit of exposure is a concentration (μg/m^3) and the unit used in CS is total pounds released normalized by two-dimensional land area (lbs/mi^2).

[10]Coefficients in the original table show the effect of a one pound/mi^2 increase in the chemical.

Table 3.4. Implied change in birth outcome with 1 SD increase in pollutant—comparing our results to those obtained by Currie and Schmeider (2009)

Currie and Schmeider (2009, p.181)

	Unit: lbs/mi^2		Implied change with 1 SD increase (p-val)				
	Mean	SD	Gest	Bwt	LBW	VLBW	Death
Developmental	212.4	791.1	-0.0195	-2.25	0.0681	0.0181	0.0196
			(0.0000)	(0.0095)	(0.0000)	(0.0086)	(0.0000)
VOCs	200.8	775	-0.0190	-2.20	0.0677	0.0182	0.0191
			(0.0001)	(0.0125)	(0.0000)	(0.0077)	(0.0000)
Toluene	176.5	742.6	-0.0182	-2.37	0.0726	0.0186	0.0205
			(0.0005)	(0.0046)	(0.0000)	(0.0047)	(0.0000)
Epichlorohydrin	0.391	4.319	-0.0040	-0.745	0.0054	0.0085	0.0060
			(0.1703)	(0.0190)	(0.6500)	(0.0005)	(0.0553)
Heavy metals	1.691	12.33	-0.0111	-2.22	0.0694	0.0129	0.0108
			(0.0590)	(0.0664)	(0.1298)	(0.2874)	(0.4083)
Lead	1.518	11.82	-0.0043	-0.932	0.0189	0.0044	-0.0020
			(0.1785)	(0.1374)	(0.1210)	(0.6716)	(0.8155)
Cadmium	0.116	2.273	-0.0061	-1.21	0.0481	0.0093	0.0103
			(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)

Non-moving mother fixed effects, Top 1% obs removed

	Unit: μg/m^3		Implied change with 1 SD increase (p-val)				
	Mean	SD	Gest	Bwt	LBW	VLBW	Death
Developmental	0.197	1.229	0.0775	3.29	-0.737	-0.452	-0.124
			(0.0853)	(0.7117)	(0.1043)	(0.0169)	(0.4261)
VOCs	0.154	1.105	0.0924	4.93	-0.543	-0.267	-0.169
			(0.0799)	(0.6277)	(0.2982)	(0.2405)	(0.3924)
Toluene	0.109	1.005	0.102	16.68	-0.921	-0.551	-0.438
			(0.1084)	(0.1767)	(0.1497)	(0.0415)	(0.0641)
Epichlorohydrin	0.00166	0.0777	0.373	-134.6	-5.021	-0.134	1.148
			(0.3165)	(0.0630)	(0.1681)	(0.9331)	(0.4106)
Heavy metals	0.000888	0.0055	0.0499	17.50	-0.278	-0.140	0.0085
			(0.0788)	(0.0010)	(0.3249)	(0.2430)	(0.9272)
Lead	0.000722	0.00426	0.0533	19.33	-0.398	-0.152	-0.0043
			(0.0423)	(0.0001)	(0.1204)	(0.1687)	(0.9619)
Cadmium	0.0000319	0.000525	0.177	29.87	-0.572	-0.331	0.0787
			(0.1155)	(0.1695)	(0.6078)	(0.4759)	(0.8218)

Non-moving mother fixed effects

	Unit: μg/m^3		Implied change with 1 SD increase (p-val)				
	Mean	SD	Gest	Bwt	LBW	VLBW	Death
Developmental	0.197	1.229	0.0158	2.76	-0.107	-0.100	-0.0146
			(0.1164)	(0.1262)	(0.2173)	(0.0328)	(0.6278)
VOCs	0.154	1.105	0.0058	2.34	-0.104	-0.0651	-0.0049
			(0.5700)	(0.1999)	(0.2196)	(0.1311)	(0.8593)
Toluene	0.109	1.005	-0.0011	2.24	-0.0908	-0.0470	-0.0044
			(0.9156)	(0.2415)	(0.2844)	(0.3036)	(0.8772)
Epichlorohydrin	0.00166	0.0777	0.0081	0.157	0.0145	-0.0081	-0.0093
			(0.0423)	(0.7760)	(0.7243)	(0.1132)	(0.0839)
Heavy metals	0.000888	0.0055	0.0104	2.95	-0.0661	-0.0131	-0.0214
			(0.0423)	(0.7760)	(0.7243)	(0.1132)	(0.0839)
Lead	0.000722	0.00426	0.0129	3.45	-0.10	-0.0137	-0.0198
			(0.0219)	(0.0001)	(0.0459)	(0.4491)	(0.2412)
Cadmium	0.0000319	0.000525	-0.0138	-0.534	0.0089	0.0215	-0.0396
			(0.1821)	(0.6200)	(0.8234)	(0.2400)	(0.1367)

CBSA fixed effects, Top 1% obs removed

	Unit: μg/m^3		Implied change with 1 SD increase (p-val)				
	Mean	SD	Gest	Bwt	LBW	VLBW	Death
Developmental	0.197	1.229	0.0128	6.56	-0.126	-0.0365	-0.0075
			(0.5380)	(0.5089)	(0.6027)	(0.4691)	(0.6808)
VOCs	0.154	1.105	0.0216	7.95	-0.241	-0.0287	-0.0188
			(0.2692)	(0.4215)	(0.2333)	(0.5058)	(0.2092)
Toluene	0.109	1.005	0.0317	4.66	-0.216	-0.0227	-0.0423
			(0.3365)	(0.7087)	(0.4833)	(0.7717)	(0.0538)
Epichlorohydrin	0.00166	0.0777	-0.110	63.71	-2.621	-0.621	0.172
			(0.2209)	(0.0161)	(0.0018)	(0.0000)	(0.0966)
Heavy metals	0.000888	0.0055	0.0171	4.73	-0.123	-0.0102	0.0172
			(0.4212)	(0.4864)	(0.4589)	(0.7881)	(0.4813)
Lead	0.000722	0.00426	0.0110	1.62	-0.0451	-0.0059	0.0174
			(0.6412)	(0.8134)	(0.7955)	(0.8883)	(0.5260)
Cadmium	0.0000319	0.000525	0.111	21.73	-0.714	0.0418	-0.109
			(0.2351)	(0.3880)	(0.0864)	(0.5710)	(0.0154)

Source: Upper-left quadrant: author's calculations based on tables in Currie and Schmeider (2009). Other quadrants: author's calculations based on 1) birth data from the Texas Center for Health Statistics (1994-2003) merged with 2) km^2-grid-cell-level RSEI chemical concentrations and 3) tract-level data from the 2000 U.S. Census.
Notes: Samples marked "Top 1% obs removed" exclude top 1% of all infants by exposure from full CBSA sample. P-values given in parentheses. P-values for Currie and Schmeider results calculated based on reported coefficients, standard errors and estimated degrees of freedom. Coefficients on LBW, VLBW and infant death multiplied by 100. Results in upper right, lower left and lower right quadrants correspond to results presented in Tables C.2, C.1 and C.3, respectively.

Thus, we present the effect of a one standard deviation increase, using a standard deviation from our CBSA sample.

Results from estimates using our preferred non-moving mother fixed effects model with outliers removed are presented in the upper right quadrant of Table 3.4. In most cases, signs of the estimated coefficients are opposite what we would expect (and opposite those found in CS), with chemical exposure being associated with longer gestation, higher birthweight, and lower probabilities of LBW, VLBW and infant death. Only one coefficient, that on epichlorohydrin in the birthweight regression, is of the expected sign and statistically significant at less than 0.10. It is also noteworthy that many of the coefficients estimated in this regression, statistically significant or not, seem improbably large in absolute magnitude. The coefficient on epichlorohydrin in the birthweight regression implies that a one standard deviation increase is associated with close to a 135 gram reduction in birthweight; another statistically significant coefficient implies that a one standard deviation increase in toluene exposure would reduce the probability of infant death by 0.44 percentage points from a baseline sample infant death rate of 0.40 percent.

The lower left quadrant in Table 3.4 shows the same non-moving mother fixed effect regressions using the full sample, without removing pollution outliers. The signs of coefficients are still unexpected in general, though their magnitudes seem more reasonable. Four coefficients are statistically significant in both quadrants: those on heavy metals and lead in the gestation regression, on lead in the birthweight regression and on developmental toxicants in the VLBW regression. In each case, the sign is the same in both samples (positive for gestation and birthweight, negative for VLBW), but the absolute magnitude is three to five times higher in the sample with outliers removed.

In the lower right quadrant, we present results from CBSA fixed effect regressions of the same birth outcomes and pollutants, on the basis that the exclusive use of

possibly inaccurate temporal pollution variation in the mother fixed effect regressions may bias our results. We also remove outliers from this sample. Here, too, gestation and birthweight coefficients tend to be positive and LBW, VLBW and infant death coefficients tend to be negative. Two coefficients are statistically significant in both this quadrant and the one directly above it. The effect of a one standard deviation increase in toluene on the probability infant death is -0.04 percentage points here, compared to -0.44 in the first sample. The coefficient on epichlorohydrin implies a 63.7 gram increase in birthweight here, compared to the 134.6 gram decrease implied in the quadrant above. Similarly, the coefficient on epichlorohydrin implies a 0.17 percentage point increase in the probability of infant death in this sample, whereas the analagous coefficient in the lower left quadrant implies a 0.009 percentage point decrease.

In Appendix Tables C.1-C.3, we present results from the three specifications presented in Table 3.4, with additional estimates from a pollution variable that represents the group of all chemicals that were in the TRI between 1994-2003 and toxicity-weighted versions of the aggregate pollution variables (developmental, VOCs and heavy metals). We find that the toxicity-weighted aggregates are not more likely to be statistically significant or of the expected sign than non-toxicity-weighted versions. In general, these results show a major lack of robustness to changes in specification and sample.

In Table 3.5, we focus on the four individual chemicals presented in CS—toluene, epichlorohydrin, lead and cadmium—and show results from LBW regressions using several different specifications. (Analagous tables for other birth outcomes are presented in Appendix Tables C.4-C.8.) Columns represent the five models described in section 3.3.1. Pollutants are represented in superrows, and subrows indicate the form of pollution variable used and whether or not outliers were removed. Coefficients here show the effect of a one $\mu g/m^3$ increase in the chemical indicated. Since lead

Table 3.5. Low birthweight (LBW) regressions with alternate pollution variables

Pollution Variable	N	Bivar	p-val	Multivar	p-val	CBSA FE	p-val	Mom FE	p-val	moms w/ var	No Move Mom FE	p-val	moms w/ var
Toluene													
A) Fug+stack, yr wts	2,868,306	0.082	(0.0320)	0.032	(0.1814)	0.019	(0.3016)	0.015	(0.6146)	50,705	-0.090	(0.2844)	22,772
B) Fug+stack concep yr	2,868,306	0.078	(0.0300)	0.030	(0.1883)	0.018	(0.3116)	0.012	(0.7052)	50,327	-0.096	(0.2693)	22,523
C) Fug+stack, birth yr	2,868,306	0.085	(0.0312)	0.036	(0.1382)	0.023	(0.2087)	0.017	(0.5684)	50,089	-0.070	(0.3726)	22,418
Like (A), top 2.5 pct rm	2,796,587	1.328	(0.0371)	-0.403	(0.5378)	-0.635	(0.1541)	-0.958	(0.0416)	48,800	-1.585	(0.1037)	22,060
Like (A), top 1.0 pct rm	2,839,622	0.987	(0.0050)	-0.067	(0.8665)	-0.215	(0.4833)	-0.487	(0.0978)	49,878	-0.916	(0.1497)	22,464
Like (A), fugitive only	2,868,306	0.077	(0.0371)	0.036	(0.1283)	0.021	(0.2212)	0.020	(0.5208)	49,765	-0.094	(0.2588)	22,205
Epichlorohydrin													
A) Fug+stack, yr wts	2,868,306	0.017	(0.4104)	-0.052	(0.2406)	-0.052	(0.0002)	0.286	(0.4237)	26,280	0.187	(0.7243)	11,503
B) Fug+stack concep yr	2,868,306	0.027	(0.1404)	-0.042	(0.2969)	-0.043	(0.0011)	0.235	(0.5165)	22,765	0.326	(0.5480)	9,880
C) Fug+stack, birth yr	2,868,306	-0.024	(0.2619)	-0.097	(0.0401)	-0.098	(0.0000)	0.341	(0.3234)	24,191	0.147	(0.7674)	10,542
Like (A), top 2.5 pct rm	2,796,599	79.414	(0.0191)	-18.599	(0.7894)	-37.903	(0.0394)	-36.888	(0.4556)	24,634	-97.013	(0.2049)	10,844
Like (A), top 1.0 pct rm	2,839,622	40.532	(0.0230)	-23.891	(0.5411)	-33.726	(0.0018)	-62.122	(0.0333)	25,606	-64.619	(0.1681)	11,207
Like (A), fugitive only	2,868,306	0.047	(0.0083)	-0.020	(0.5961)	-0.023	(0.0666)	0.302	(0.4063)	17,068	0.179	(0.7306)	7,417
Lead													
A) Fug+stack, yr wts	2,868,306	11.03	(0.4216)	-0.24	(0.9773)	-5.71	(0.2734)	-0.24	(0.9735)	50,921	-24.29	(0.0459)	22,756
B) Fug+stack concep yr	2,868,306	8.86	(0.5013)	-0.69	(0.9287)	-5.56	(0.2551)	0.083	(0.9906)	49,885	-23.38	(0.0548)	22,047
C) Fug+stack, birth yr	2,868,306	14.25	(0.2913)	2.20	(0.7920)	-3.34	(0.5355)	-0.91	(0.8974)	50,288	-22.72	(0.0388)	22,280
Like (A), top 2.5 pct rm	2,796,591	199.27	(0.1001)	90.93	(0.3194)	-23.04	(0.7025)	41.50	(0.4878)	49,175	-13.42	(0.8978)	22,016
Like (A), top 1.0 pct rm	2,839,622	125.94	(0.1298)	47.27	(0.4617)	-10.58	(0.7955)	-36.38	(0.3066)	50,273	-93.34	(0.1204)	22,460
Like (A), fugitive only	2,868,306	25.81	(0.0144)	4.73	(0.5731)	-2.19	(0.7036)	-7.86	(0.4727)	47,909	-38.91	(0.0113)	21,133
Cadmium													
A) Fug+stack, yr wts	2,868,306	26.2	(0.5634)	-33.3	(0.2668)	-38.6	(0.0538)	158.9	(0.0504)	21,049	17.0	(0.8234)	8,751
B) Fug+stack concep yr	2,868,306	5.2	(0.9201)	-37.1	(0.1720)	-41.5	(0.0283)	134.0	(0.0592)	18,883	47.2	(0.5251)	7,056
C) Fug+stack, birth yr	2,868,306	48.1	(0.1677)	-26.0	(0.4158)	-31.9	(0.1556)	180.0	(0.0681)	18,050	-44.7	(0.7508)	6,733
Like (A), top 2.5 pct rm	2,796,598	6,063.0	(0.1261)	-737.4	(0.8627)	-1,171.4	(0.6150)	-3,494.3	(0.2123)	18,822	-4,788.0	(0.3406)	7,444
Like (A), top 1.0 pct rm	2,839,623	2,248.4	(0.3393)	-904.3	(0.5789)	-1,359.2	(0.0864)	-1,377.9	(0.2566)	19,766	-1,089.8	(0.6078)	7,815
Like (A), fugitive only	2,868,306	68.2	(0.0000)	-14.6	(0.5224)	-21.4	(0.0364)	207.1	(0.0487)	14,207	12.3	(0.9303)	5,007

Source: Author's calculations based on 1) birth data from the Texas Center for Health Statistics (1994-2003) merged with 2) km^2-grid-cell-level RSEI chemical concentrations and 3) tract-level data from the 2000 U.S. Census.
Notes: Dependent variable is occurance of low birthweight (< 2500 grams). Linear probability model OLS regressions used for models in columns (1) and (2); within -CBSA, -mother and -non-moving mother regressions used in columns (3), (4) and (5), respectively. Standard errors adjusted for heteroskedasticity and clustered at CBSA level in columns (1)-(3) and at mother level in columns (4) and (5). Numbers shown here are estimated coefficients and p-values from the pollution variable only. All coefficients multipled by 100. "Moms w/ var" indicates the number of mothers in the sample who had variation in both chemical exposure and child LBW.

and cadmium, especially, are found in concentrations much lower than one $\mu g/m^3$, estimated coefficients on these chemicals look large.

Because the time infants spend in utero often spans two TRI reporting years, our preferred pollution variable, as described above, is the average estimated concentrations from these years, weighted by the share of each year spent in utero. Results from regressions using these pollution variables are presented in the rows in Table 3.5 marked "A) Fug+stack, yr wts." Rows (B) and (C) show results from regressions that instead use estimated concentrations from the conception year and the birth year, respectively, to check whether results are robust to these alternate specifications. Comparing results in rows (A)-(C), we see that the sign and magnitude of coefficients are generally similar. Several exceptions exist (e.g., the coefficient on birth year epichlorohydrin in the bivariate regression, birth year lead in the multivariate regression, conception year lead in the mother fixed effects regression, and birth year cadmium in the non-moving mother regression), but the coefficients in these cases are not statistically significant. Thus, it seems that the year-weighted concentrations are reasonable to use in general.

In the fourth and fifth subrows under each pollutant in Table 3.5, we show the results of estimating the regressions after removing the top 2.5 percent or the top 1.0 of observations by chemical exposure from the CBSA sample, respectively.[11] Compared to the sample that uses all observations, removing outliers changes the sign in about half the coefficients shown here, though only the coefficient on toluene in the mother fixed effects model and the coefficients on epichlorohydrin in the bivariate and mother fixed effects models are statistically significant in the outliers-removed regressions and not the all-observations version. As we observed in Table 3.4, removing outliers

[11]Selection for the mother fixed effect samples took place after the removal of outliers, so it may not be the case that exactly 2.5 (1.0) percent of the observations of these smaller samples were removed.

significantly increases the absolute magnitude of the coefficients. For toluene in the bivariate model, and epichlorohydrin and cadmium in the CBSA fixed effects model, coefficients are statistically significant and of the same sign in the regressions with and without outliers removed, but in each case, the coefficient increases by more than an order of magnitude in the versions without outliers.

CS uses fugitive chemical releases only, on the basis that emissions that go up a smoke stack are more likely to be treated with scrubbers and travel farther than those that do not. Indeed, results from the paper show coefficients on fugitive emissions are statistically significant where the coefficients on analagous stack emissions are not [39, p.182]. In our preferred specification, we use concentration estimates based on both fugitive and stack emissions, since we expect the dispersion model employed by RSEI to estimate both well. However, a close look at the methodology for modeling air releases does show that many more assumptions, having to do with meteorological conditions and chemical dispersion behavior at the height of the stack, are employed when modeling stack releases [50, pp.39-47]. For this reason, we also estimate regressions using concentrations from fugitive emissions only. We find that in the bivariate model, the coefficients on epichlorohydrin, lead and cadmium more than double and become statistically significant when fugutive releases only are used instead of fugitive and stack. In other models, however, coefficients are generally similar, and dissimilarity in size or sign is accompanied by lack of statistical significance.

In general, the results presented in Table 3.5 indicate that using year-weighted exposure measures and both fugitive and stack emissions are reasonable, but the lack of robustness across models and samples with and without outliers presents questions about the validity of any particular results. Relatively few coefficients are statistically significant at conventional levels and those that achieve significance in one model often do not in a slightly different specification. Though only LBW regression results are

shown here, similar patterns are borne out in analagous appendix tables with different birth outcomes (see Tables C.4-C.8).

Table 3.6. Regressions with categorical pollution variables, non-moving mother fixed effects

	Gest (wks)	Bwt (g)	PTB	LBW	VLBW	Infant Death
Toluene, top 25 pct	0.0313	1.73	-0.483	-0.184	-0.018	-0.092
p-val	(0.0480)	(0.5749)	(0.0165)	(0.2314)	(0.7869)	(0.0943)
moms w/ var	*33,215*	*37,455*	*6,796*	*3,948*	*759*	*549*
Toluene, top 10 pct	0.0172	2.77	-0.591	-0.365	-0.087	-0.154
p-val	(0.4381)	(0.5199)	(0.0376)	(0.0957)	(0.3506)	(0.0455)
moms w/ var	*17,638*	*19,761*	*3,887*	*2,281*	*416*	*314*
Toluene, top 5 pct	-0.0134	-4.63	-0.385	0.073	-0.081	0.0026
p-val	(0.6460)	(0.4073)	(0.3071)	(0.8033)	(0.5348)	(0.9803)
moms w/ var	*9,797*	*11,013*	*2,147*	*1,265*	*249*	*183*
Epichlorohydrin, top 25 pct	0.0154	4.55	0.166	0.433	0.304	0.197
p-val	(0.7499)	(0.6295)	(0.7768)	(0.3269)	(0.1392)	(0.2228)
moms w/ var	*4,402*	*4,940*	*832*	*487*	*93*	*80*
Epichlorohydrin, top 10 pct	-0.0017	-8.82	-0.257	0.059	0.075	0.078
p-val	(0.9330)	(0.0221)	(0.3044)	(0.7576)	(0.3651)	(0.2684)
moms w/ var	*19,825*	*22,785*	*3,851*	*2,286*	*451*	*309*
Epichlorohydrin, top 5 pct	0.0275	-5.11	-0.445	-0.681	-0.064	0.052
p-val	(0.2751)	(0.2875)	(0.1604)	(0.0041)	(0.5396)	(0.5514)
moms w/ var	*12,849*	*14,698*	*2,597*	*1,489*	*294*	*219*
Lead, top 25 pct	0.00029	2.10	0.237	0.068	0.004	0.041
p-val	(0.9826)	(0.4253)	(0.1633)	(0.6049)	(0.9497)	(0.3660)
moms w/ var	*40,547*	*46,471*	*7,766*	*4,604*	*844*	*591*
Lead, top 10 pct	0.0318	12.17	0.160	-0.168	-0.147	0.0076
p-val	(0.1206)	(0.0021)	(0.5365)	(0.4013)	(0.0936)	(0.9140)
moms w/ var	*18,681*	*21,123*	*3,839*	*2,299*	*426*	*304*
Lead, top 5 pct	0.0446	13.31	0.132	-0.099	0.038	-0.022
p-val	(0.0899)	(0.0086)	(0.6921)	(0.7028)	(0.7444)	(0.8165)
moms w/ var	*11,037*	*12,523*	*2,254*	*1,391*	*264*	*188*
Cadmium, top 25 pct	0.0106	10.61	0.040	-0.147	-0.070	-0.082
p-val	(0.6168)	(0.0100)	(0.8830)	(0.4722)	(0.3937)	(0.2466)
moms w/ var	*16,784*	*19,133*	*3,176*	*1,839*	*327*	*244*
Cadmium, top 10 pct	0.0042	4.04	-0.023	-0.082	-0.175	0.050
p-val	(0.8480)	(0.3480)	(0.9359)	(0.7092)	(0.0521)	(0.4856)
moms w/ var	*15,304*	*17,353*	*3,098*	*1,871*	*328*	*212*
Cadmium, top 5 pct	0.0529	13.40	-0.488	-0.456	-0.186	0.014
p-val	(0.0870)	(0.0262)	(0.2203)	(0.1344)	(0.1601)	(0.8848)
moms w/ var	*8,255*	*9,314*	*1,719*	*1,045*	*202*	*119*

Source: Author's calculations based on 1) birth data from the Texas Center for Health Statistics (1994-2003) merged with 2) km^2-grid-cell-level RSEI chemical concentrations and 3) tract-level data from the 2000 U.S. Census.
Notes: Dependent variable listed in column headings. Regressions use variation within mothers who did not move between births. Standard errors adjusted for heteroskedasticity and clustered at mother level. Pollution variables indicate whether infant is in top X% of all infants by exposure in full CBSA sample. "Moms w/ var" indicates the number of mothers in the sample who had between-birth variation in both chemical exposure category and infant health outcome. Coefficients in PTB, LBW, VLBW and infant death regressions multiplied by 100.

One way of incorporating information from high exposure values without allowing them to exert undue influence is to use categorical pollution variables. Table 3.6 presents results from regressions using the non-moving mother fixed effect model with categorical variables for toluene, epichlorohydrin, lead and cadmium. Because

creating categorical variables requires an arbitrary cutoff, we present three versions here, making the "high pollution" category alternately include the top 25 percent, top 10 percent and top 5 percent of observations by exposure value from the full CBSA sample.[12]

Results obtained using the categorical pollution variables do not appear to be more consistent than those from the continuous pollution variable regressions. Whatever the estimated effect is when high exposure is defined as the top 25 percent of exposures, we might expect it to intensify in the top 10 and top 5 percent versions. We see this in the case of the birthweight-lead regression, in which the positive coefficient gets bigger and achieves statistical significance as the high pollution category narrows. We also see this pattern on the toluene coefficients in the PTB and infant death regressions, though it is not consistent across all three categorical variables. In the PTB case, the negative coefficient gets larger in absolute magnitude between the top 25 and top 10 percent categories (-0.48 to -0.59), but gets smaller (-0.39) and loses statistical significance when the top 5 percent category is used. In the infant death case, the negative coefficient also grows in absolute magnitude between the first two categories (-0.09 to -0.15) but changes sign (0.003) and loses statistical significance in the 5 percent category.

In general, however, we do not see expected trends in the magnitude of coefficients across categorical variable definitions, and statistical significance seems randomly scattered around the table. The coefficient on high lead exposure, for example, is -0.15 and statistically significant when the 10 percent category is used, but the coefficients are positive and not statistically significant when high exposure is defined as the top 25 or top 5 percent of exposures. It is also the case that using categories reduces the

[12]Selection for the mother fixed effect samples took place after the designation of pollution categories, so it may not be the case that 25 (10, 5) percent of the observations of these smaller samples are in the "high pollution" category.

number of mothers who exhibited variation in both exposure category and outcome across births, which would seem most problematic in the cases of VLBW and infant death, though coefficients from regressions with these outcomes are not more bizarre or less likely to be statistically significant than others.

Table 3.7. Regressions with categorical pollution variables, CBSA fixed effects

	Gest(wks)	Bwt (g)	PTB	LBW	VLBW	Infant Death
Toluene, top 25 pct	0.0154	3.74	-0.136	-0.136	-0.0329	-0.0187
p-val	(0.0681)	(0.0751)	(0.2145)	(0.1013)	(0.2360)	(0.0594)
Toluene, top 10 pct	0.0163	6.72	-0.158	-0.178	-0.0329	-0.00572
p-val	(0.0655)	(0.0942)	(0.2957)	(0.0776)	(0.1411)	(0.5223)
Toluene, top 5 pct	0.0188	3.54	-0.176	-0.0434	-0.0140	-0.00920
p-val	(0.3288)	(0.6437)	(0.4392)	(0.7990)	(0.7592)	(0.5341)
Epichlorohydrin, top 25 pct	-0.0915	-23.90	0.694	0.409	0.101	-0.0205
p-val	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0014)
Epichlorohydrin, top 10 pct	-0.00279	7.22	-0.126	-0.151	-0.0118	0.0066
p-val	(0.6147)	(0.0000)	(0.0120)	(0.0012)	(0.0300)	(0.2551)
Epichlorohydrin, top 5 pct	0.0115	14.22	-0.280	-0.380	-0.0905	-0.0177
p-val	(0.0183)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0309)
Lead, top 25 pct	-0.00379	-3.56	0.00306	-0.0345	-0.0164	-0.0107
p-val	(0.5278)	(0.0471)	(0.9640)	(0.4206)	(0.5073)	(0.4832)
Lead, top 10 pct	0.00947	3.43	-0.0604	-0.0816	-0.00867	0.0273
p-val	(0.6186)	(0.4704)	(0.5259)	(0.5847)	(0.8391)	(0.2710)
Lead, top 5 pct	0.0148	3.93	-0.0815	-0.0515	0.0168	0.0474
p-val	(0.5941)	(0.6407)	(0.5972)	(0.7967)	(0.6969)	(0.0436)
Cadmium, top 25 pct	-0.0142	-4.00	0.221	-0.0834	0.0156	-0.0328
p-val	(0.6610)	(0.6503)	(0.2180)	(0.5475)	(0.3183)	(0.0680)
Cadmium, top 10 pct	-0.00341	-2.89	0.0975	0.0337	0.0125	-0.0192
p-val	(0.8942)	(0.7212)	(0.3809)	(0.8234)	(0.3750)	(0.0999)
Cadmium, top 5 pct	0.0326	9.99	-0.169	-0.258	-0.0242	-0.0270
p-val	(0.2358)	(0.2115)	(0.2660)	(0.1000)	(0.4543)	(0.2761)

Source: Author's calculations based on 1) birth data from the Texas Center for Health Statistics (1994-2003) merged with 2) km^2-grid-cell-level RSEI chemical concentrations and 3) tract-level data from the 2000 U.S. Census.
Notes: Dependent variable listed in column headings. Regressions include CBSA fixed effects. Standard errors adjusted for heteroskedasticity and clustered at CBSA level. Pollution variables indicate whether infant is in top X% of all infants in CBSA sample by exposure. Coefficients in PTB, LBW, VLBW and infant death regressions multiplied by 100.

Table 3.7 shows results from analogous regressions using the CBSA fixed effects model and sample, in order to incorporate the spatial variation in pollution that is missing from the non-moving mother fixed effects version. Notably, the coefficients on almost all the epichlorohydrin variables in these regressions are highly statistically significant, and the trend for each outcome except infant death indicates that higher doses of the chemical lead to better birth outcomes. For these five outcomes, the coefficient on the top 25 percent epichlorohydrin category indicates that high exposure has a negative effect on birth outcome, but in each case, the sign changes when high

exposure is defined as membership in the top 5 percent. We see this pattern for some of the other chemicals and outcomes in the CBSA regressions, though without statistical significance.

In general, we see that the results from the categorical regressions are not robust to different category definitions and there is little correspondance between the CBSA fixed effects and the non-moving mother fixed effects regressions.

In Tables 3.8 and 3.9, we present results from another attempt to allow for a nonlinear relationship between pollutant and birth outcome, this time by adding squared and cubed pollution terms to the original linear model. Allowing for these polynomial relationships avoids the need to make arbitrary cutoff decisions, as in the categorical variable case, and includes more mothers with variation in both birth outcome and the pollution variable. The question of what to do with outliers reasserts itself here, however, and we choose to focus on the sample with the top 1 percent of observations by exposure removed, though we cannot know whether these values represent truly high exposures or errors in reporting.

Looking first at Table 3.8, which presents results from the gestation, birthweight and PTB regressions that allow the relationship with pollution to be alternately linear, quadratic and cubic, we see that the higher order polynomials do not seem to better describe the relationship between pollution and outcome in most cases. Eleven of the 36 regression results in this table have at least one statistically-significant pollution term; eight of these are the linear pollution term, two are a squared term in combination with a statistically-significant linear term, and one is a cubed term in isolation. In Table 3.9, which shows results from analogous LBW, VLBW and infant death regressions, it does appear that a quadratic or cubic specification better represents the relationship between outcome and pollution in some cases, based on observed statistical significance (e.g., the quadratic specification of the VLBW-

Table 3.8. Gestation, birthweight and PTB regressions with polynomial pollution variables, non-moving mother fixed effects, top 1% of obs by exposure removed

	Gest (wks)			Bwt (g)			PTB		
	$conc$	$conc^2$	$conc^3$	$conc$	$conc^2$	$conc^3$	$conc$	$conc^2$	$conc^3$
Toluene									
moms w/ var	*210,004*	.	.	*246,624*	.	.	*38,230*	.	.
Linear	0.101	.	.	16.6	.	.	-1.73		
p-val	(0.1084)	.	.	(0.1767)	.	.	(0.0376)	.	.
Quadratic	0.310	-0.243	.	5.83	12.5	.	-5.29	4.14	
p-val	(0.0258)	(0.0907)	.	(0.8289)	(0.6544)	.	(0.0034)	(0.0243)	.
Cubic	0.539	-0.946	0.0000012	69.4	-183.5	134.6	-5.76	5.58	-0.991
p-val	(0.0192)	(0.1092)	(0.2228)	(0.1190)	(0.1065)	(0.0777)	(0.0495)	(0.4509)	(0.8403)
Epichlorohydrin									
moms w/ var	*104,840*	.	.	*122,845*	.	.	*18,837*	.	.
Linear	4.80	.	.	-1,732	.	.	-36.5		
p-val	(0.3165)	.	.	(0.0630)	.	.	(0.5468)	.	.
Quadratic	-0.992	646.3	.	-3,918	243,845	.	-149.1	12,564	
p-val	(0.9309)	(0.5802)	.	(0.0750)	(0.2792)	.	(0.2993)	(0.3939)	.
Cubic	1.25	15.5	40,599	-5,362	650,649	-26,183,122	-171.2	18,791	-400,792
p-val	(0.9525)	(0.9976)	(0.9015)	(0.1906)	(0.5226)	(0.6853)	(0.5172)	(0.7745)	(0.9234)
Lead									
moms w/ var	*209,588*	.	.	*245,990*	.	.	*38,206*	.	.
Linear	12.5	.	.	4,537	.	.	6.8		
p-val	(0.0423)	.	.	(0.0001)	.	.	(0.9296)	.	.
Quadratic	10.7	247.0	.	5,112	-79,194	.	203.2	-27,055	
p-val	(0.4465)	(0.8893)	.	(0.0586)	(0.8114)	.	(0.2582)	(0.2269)	.
Cubic	-3.59	5441.4	-409,005	1,215	1,336,144	-111,442,945	420.4	-105,955	6,212,576
p-val	(0.8824)	(0.4626)	(0.4719)	(0.7951)	(0.3426)	(0.2973)	(0.1718)	(0.2552)	(0.3823)
Cadmium									
moms w/ var	*75,851*	.	.	*88,480*	.	.	*13,460*	.	.
Linear	337.9	.	.	56,898	.	.	-3,644		
p-val	(0.1155)	.	.	(0.1695)	.	.	(0.1740)	.	.
Quadratic	191.1	543.7	.	119,898	-233,309	.	-645	-11,107	
p-val	(0.7049)	(0.7458)	.	(0.2229)	(0.4748)	.	(0.9216)	(0.6070)	.
Cubic	-656.6	9,250	-18,494	283,460	-1,913,428	3,568,695	7,729	-97,128	182,716
p-val	(0.4684)	(0.2444)	(0.2575)	(0.1053)	(0.2094)	(0.2557)	(0.5075)	(0.3388)	(0.3830)

Source: Author's calculations based on 1) birth data from the Texas Center for Health Statistics (1994-2003) merged with 2) km^2-grid-cell-level RSEI chemical concentrations and 3) tract-level data from the 2000 U.S. Census.
Notes: Dependent variable listed in column headings. Sample excludes top 1% of all infants by exposure from full CBSA sample. Regressions use only variation within mothers who did not move between births (non-moving mother fixed effects). Standard errors adjusted for heteroskedasticity and clustered at mother level. "Linear", "quadratic" and "cubic" refer to models in which the relationship between birth outcome and pollution is allowed to have the shape indicated. "Moms w/ var" indicates the number of mothers in the sample who had between-birth variation in both chemical exposure and infant health outcome. Coefficients in PTB regressions multiplied by 100.

Table 3.9. LBW, VLBW and infant death regressions with polynomial pollution variables, non-moving mother fixed effects, top 1% of obs by exposure removed

	LBW			VLBW			Infant Death		
	$conc$	$conc^2$	$conc^3$	$conc$	$conc^2$	$conc^3$	$conc$	$conc^2$	$conc^3$
Toluene									
moms w/ var	*22,435*	.	.	*4,136*	.	.	*2,849*	.	.
Linear	-0.916			-0.548			-0.436	.	.
p-val	(0.1497)	.	.	(0.0415)	.	.	(0.0641)	.	.
Quadratic	-0.836	-0.094		0.059	-0.706		-0.545	0.126	.
p-val	(0.5404)	(0.9464)	.	(0.9219)	(0.2546)	.	(0.2632)	(0.8148)	.
Cubic	-5.34	13.8	-9.53	-1.34	3.59	-2.95	-1.35	2.60	-1.70
p-val	(0.0168)	(0.0156)	(0.0125)	(0.1541)	(0.1330)	(0.0663)	(0.1100)	(0.2656)	(0.3071)
Epichlorohydrin									
moms w/ var	*11,207*	.	.	*2,118*	.	.	*1,450*	.	.
Linear	-64.6			-1.72			14.8	.	.
p-val	(0.1681)	.	.	(0.9331)	.	.	(0.4106)	.	.
Quadratic	-106.0	4,616		14.1	-1,760		35.9	-2,361	.
p-val	(0.3350)	(0.6836)	.	(0.7665)	(0.7174)	.	(0.3983)	(0.6017)	.
Cubic	166.8	-72,230	4,946,015	19.6	-3,319	100,384	121.0	-26,311	1,541,518
p-val	(0.4122)	(0.1533)	(0.1244)	(0.8266)	(0.8817)	(0.9437)	(0.1268)	(0.1924)	(0.2412)
Lead									
moms w/ var	*22,425*	.	.	*4,149*	.	.	*2,858*	.	.
Linear	-93.3			-35.7			-1.01	.	.
p-val	(0.1204)	.	.	(0.1687)	.	.	(0.9619)	.	.
Quadratic	-100.5	981.4		-11.7	-3,299		69.2	-9,674	.
p-val	(0.4702)	(0.9554)	.	(0.8432)	(0.6380)	.	(0.1415)	(0.0932)	.
Cubic	274.2	-135,105	10,715,363	-85.3	23,422	-2,103,966	96.6	-19,619	783,078
p-val	(0.2569)	(0.0680)	(0.0601)	(0.4189)	(0.4719)	(0.3882)	(0.2470)	(0.4442)	(0.6857)
Cadmium									
moms w/ var	*7,813*	.	.	*1,402*	.	.	*979*	.	.
Linear	-1,090			-630.8			149.8	.	.
p-val	(0.6078)	.	.	(0.4759)	.	.	(0.8218)	.	.
Quadratic	-6,098	18,547		-4,756	15,275		921.3	-2,857	.
p-val	(0.2245)	(0.2781)	.	(0.0208)	(0.0223)	.	(0.5544)	(0.5598)	.
Cubic	-1,391	-29,801	102,696	-7,101	39,366	-51,169	1,319	-6,937	8,667
p-val	(0.8762)	(0.7030)	(0.5280)	(0.0468)	(0.2081)	(0.4197)	(0.6613)	(0.7902)	(0.8680)

Source: Author's calculations based on 1) birth data from the Texas Center for Health Statistics (1994-2003) merged with 2) km^2-grid-cell-level RSEI chemical concentrations and 3) tract-level data from the 2000 U.S. Census.
Notes: Dependent variable listed in column headings. Sample excludes top 1% of all infants by exposure from full CBSA sample. Linear probability model regressions using only variation within mothers who did not move between births (non-moving mother fixed effects). Standard errors adjusted for heteroskedasticity and clustered at mother level. "Linear", "quadratic" and "cubic" refer to models in which the relationship between birth outcome and pollution is allowed to have the shape indicated. "Moms w/ var" indicates the number of mothers in the sample who had between-birth variation in both chemical exposure and infant health outcome. Coefficients multiplied by 100.

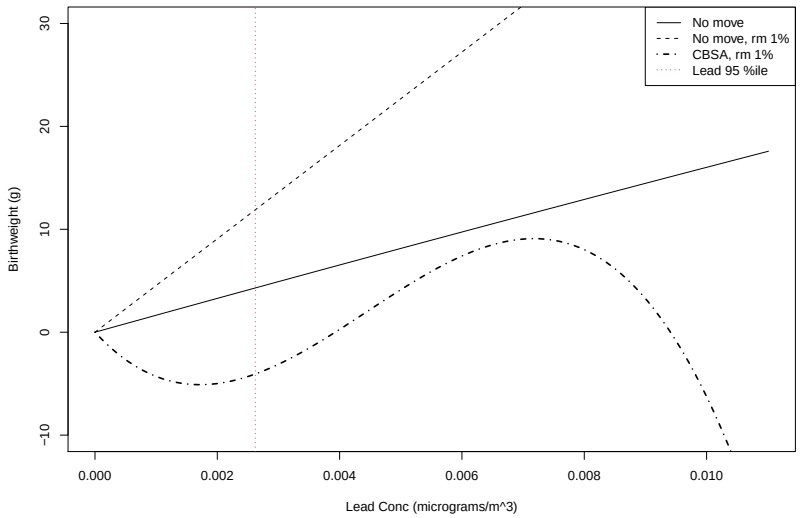
cadmium regression and the cubic specifications in the LBW-toluene and LBW-lead regressions.)

We would expect that a significant linear term in the linear-only regressions would predict a significant linear term in the quadratic and cubic regressions (likewise for a significant squared term in the quadratic regression implying a significant squared term in the cubic regression), even if the higher order terms are not significant. We do see this pattern in several cases in Tables 3.8 and 3.9, but the fact that there are seven outcome-pollution combinations for which this is not the case raises some doubt as to the trustworthiness of the statistically significant results.

Moreover, a comparison of these results to those obtained using the same model without outliers removed (Appendix Tables C.9 and C.10) and using the CBSA fixed effects model with outliers removed (Appendix Tables C.11 and C.12) further calls into question the credibility of the results. Each of the models has a handful of statistically significant pollution coefficients, but these are not generally robust across models.

In Figures 3.1 and 3.2, we present the two outcome-pollution combinations that did have statistically significant pollution terms across the three models. Figure 3.1 shows the implied relationship between birthweight and lead based on coefficient estimates from the three models. Because of the long right tail of the lead exposure distribution, we limited the x-axis range to the 0-99 percentiles. The coefficient estimates from non-moving mother fixed effect regressions with and without outliers removed imply a relationship between lead and birthweight that increases monotonically over the range. The sample that includes outliers actually produced statistically significant coefficients on the squared and cubed lead terms as well, but the inflection points occur after the 99th exposure percentile. The coefficients on the CBSA fixed effects regression with outliers removed imply a birthweight-lead relationship that decreases initially, then increases at some point just below the 95th exposure percentile, then decreases again at higher levels of exposure.

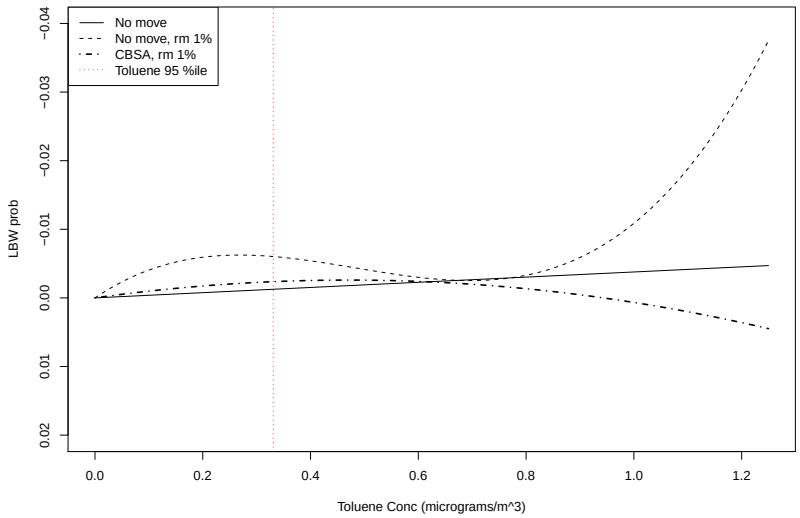
Figure 3.1. Polynomial lead-birthweight relationship; 0-99 percentile lead concentrations on x-axis



Source: Author's calculations based on 1) birth data from the Texas Center for Health Statistics (1994-2003) merged with 2) km^2-grid-cell-level RSEI chemical concentrations and 3) tract-level data from the 2000 U.S. Census.
Notes: "No move" line corresponds to estimated coefficients from the cubic lead-birthweight model in Table C.9. "No move, rm 1%" line corresponds to estimated coefficients from the cubic lead-birthweight model in Table 3.8. "CBSA, rm 1%" line corresponds to estimated coefficients from the cubic lead-birthweight model in Table C.11.

Figure 3.2 shows the implied relationship between LBW and toluene based on coefficient estimates from the same three models. The y-axis has been inverted here so that a "better" outcome (lower probability of LBW) is farther from the origin. Once again, the estimate from the non-moving mother fixed effects regression with outliers included implies a monotonically improving relationship between LBW and toluene over the 0-99 percentile exposure range. The same relationship implied by coefficients from the mother fixed effects regression with outliers removed implies that toluene decreases the probability of LBW until somewhere beyond the 95th exposure percentile, and then increases LBW probability. The CBSA fixed effects estimates, on the other hand, imply decreasing probability of LBW initially, which changes to increasing somewhere below the 95th percentile and eventually changes to decreasing again at higher levels.

Figure 3.2. Polynomial toluene-LBW relationship; 0-99 percentile toluene concentrations on x-axis

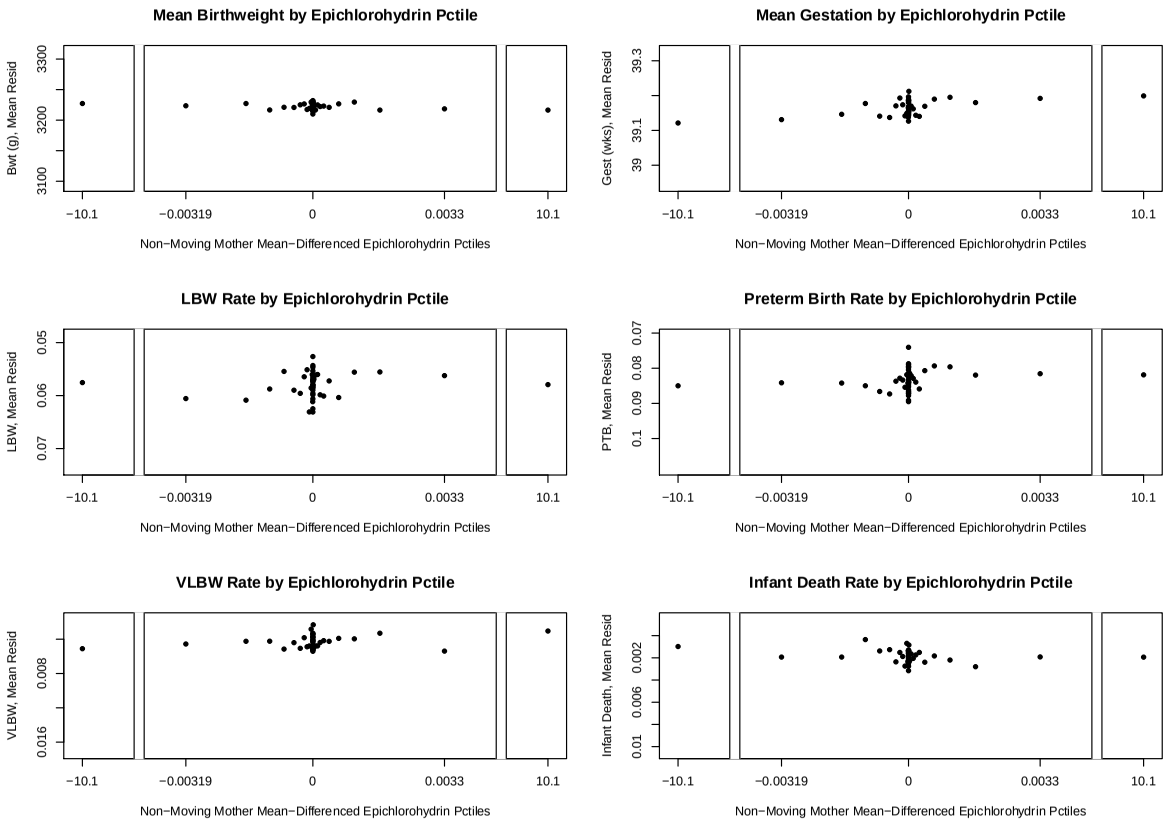


Source: Author's calculations based on 1) birth data from the Texas Center for Health Statistics (1994-2003) merged with 2) km^2-grid-cell-level RSEI chemical concentrations and 3) tract-level data from the 2000 U.S. Census.
Notes: "No move" line corresponds to estimated coefficients from the quadratic toluene-birthweight model in Table C.10. "No move, rm 1%" line corresponds to estimated coefficients from the cubic toluene-LBW model in Table 3.9. "CBSA rm 1%" line corresponds to estimated coefficients from the quadratic toluene-LBW model in Table C.12.

Lack of correspondence across models would not necessarily be a cause for concern if one model is clearly better than others on theoretical grounds, but both the CBSA and mother fixed effects models are problematic. The former suffers from difficulty in isolating the effect of pollution, the latter uses possibly inaccurate temporal variation, and both rely on self-reported and modeled pollution data. Thus, we interpret lack of robustness across various specifications of the polynomial models as further evidence that even those estimates that achieve statistical significance likely do not represent the true relationship between a given pollutant and its effect on infant health.

As a final exploration, we use the semiparametric method described beginning on p.136, which places no constraints on the relationship between birth outcomes and pollution. Figure 3.3 shows the relationship between mother-mean-differenced birth outcomes (y-axis) and mother-mean-differenced epichlorohydrin exposure (x-axis) from the sample of non-moving mothers, after the outcomes have been "cleaned"

Figure 3.3. Epichlorohydrin - birth outcome relationships; uses variation within non-moving mothers (semi-parametric method)



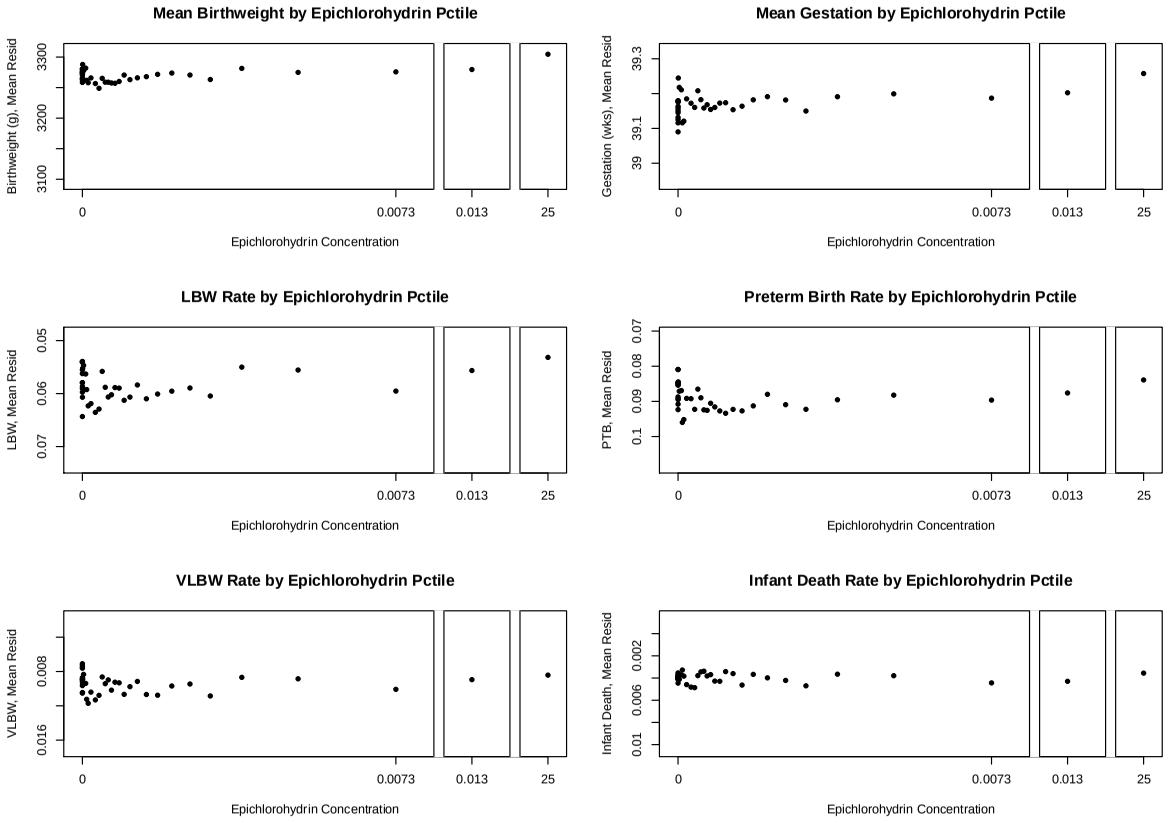
Source: Author's calculations based on 1) birth data from the Texas Center for Health Statistics (1994-2003) merged with 2) km^2-grid-cell-level RSEI chemical concentrations and 3) tract-level data from the 2000 U.S. Census.
Notes: See p.136 for description of method.

of the effect of other control variables. Each dot represents one percentile of mean-differenced exposure, with all zero observations collapsed to a single point. Percentiles corresponding to negative mean-differenced exposures have been placed on the x-axis at the lower endpoint of the given percentile; those corresponding to positive mean-differenced exposures were placed at the upper endpoint. Thus, the lowest number on the x-axis corresponds to the minimum mean-differenced exposure and the highest number corresponds to the maximum. Since the distribution of exposure typically has a long right tail, gaps are placed in plot where relatively big jumps occur. Mean-differenced birth outcome values have been shifted up by a constant, to allow the y-axis range to match that used in Figure 3.4, and the y-axis is inverted for PTB, LBW, VLBW and infant death, so that "better" outcomes are farther from the origin on each graph.

We expect the relationship between epichlorohydrin and any birth outcome to exhibit a downward slope. This relationship does not appear to be borne out in Figure 3.3, in which no clear relationships are visible. If we exclude the dots associated with top and bottom few percentiles in the infant death graph, we see something like a downward-sloping relationship, but as in the other graphs, there is also a good deal of variation in birth outcomes at mean-differenced concentrations that are very near zero.

In case using only within-mother variation is too restrictive, we also present results from the same semiparametric method applied to non-mean-differenced variables in Figure 3.4. In addition to utilizing between-person and geographic variation, we are able to take advantage of a larger sample in these regressions. There do appear to be some trends in the six graphs in Figure 3.4, but the relationships are not easy to explain. For example, in the gestation and PTB graphs, we see something like a U-shaped relationship, if we ignore the observations that are very near zero and the top seven percentiles. But as in the within-mother versions, we see as much

Figure 3.4. Epichlorohydrin - birth outcome relationships, pooled sample (semi-parametric method)



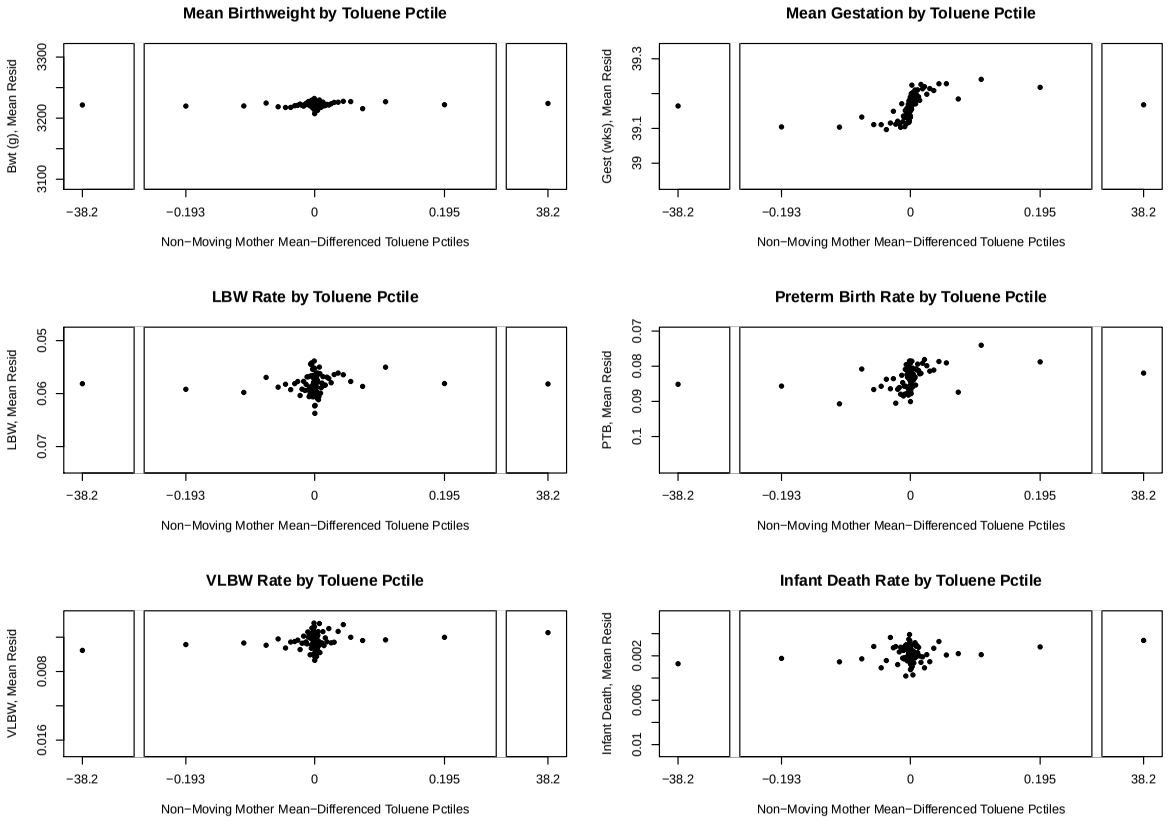
Source: Author's calculations based on 1) birth data from the Texas Center for Health Statistics (1994-2003) merged with 2) km^2-grid-cell-level RSEI chemical concentrations and 3) tract-level data from the 2000 U.S. Census.
Notes: See p.136 for description of method.

variation in birth outcomes at exposures very near to zero as at higher exposures. Further, as in the analagous multivariate specifications in the parametric regressions, we are hesitant to place too much faith in the relationships using this method, since the effects of pollution exposure may be conflated with unobserved maternal or area characteristics that are also correlated with birth outcomes.

We show analagous graphs for toluene in Figures 3.5 and 3.6, for lead in Figures 3.7 and 3.8, and for cadmium in Figures 3.9 and 3.10. Figures 3.5 and 3.6 suggest a positive relationship between toluene and gestation, and Figure 3.6 indicates that high levels of toluene exposure are related to increased probability of LBW, though this relationship is not obvious in the within-mother version. In Figure 3.7, we do not see obvious relationships between lead and birth outcomes using only within-mother variation. We do observe some difficult-to-explain trends in Figure 3.8, including roughly U-shaped relationships between lead and birthweight, in one case, and LBW, in another, when the last 10 percentiles and some near-zero percentiles are ignored. Finally, while we do not see clear relationships between cadmium and birth outcomes in Figure 3.9, we do see a very clear relationship between cadmium and PTB in Figure 3.10, if the last few percentiles are ignored. If we ignore near-zero percentile means, we also see some rough evidence for relationships between cadmium and LBW, VLBW and reduced gestation.

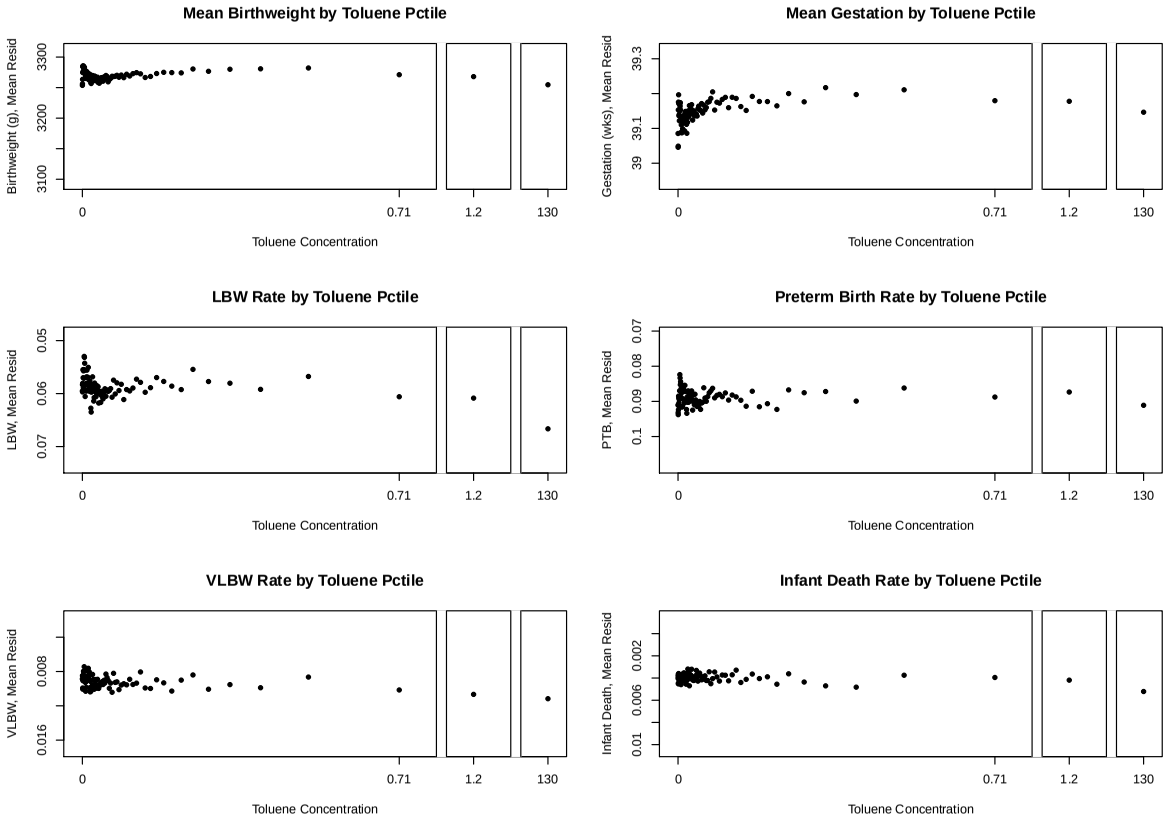
Of all our specifications, we are most inclined to look to the semiparametric graphs for suggestive evidence of the pollution-birth outcome relationships. Graphs analogous to those shown in Figures 3.3-3.10 were produced for the other 170 chemicals that were emitted in Texas and tracked during our sample period; these are available from the author. As in the figures presented here, patterns are also discernable in some of these graphs, though they are often difficult to explain. We hesitate to place too much faith in those that show the expected negative relationship between birth outcome and pollutant, since this would require us to take the confusing patterns—

Figure 3.5. Toluene - birth outcome relationships; uses variation within non-moving mothers (semi-parametric method)



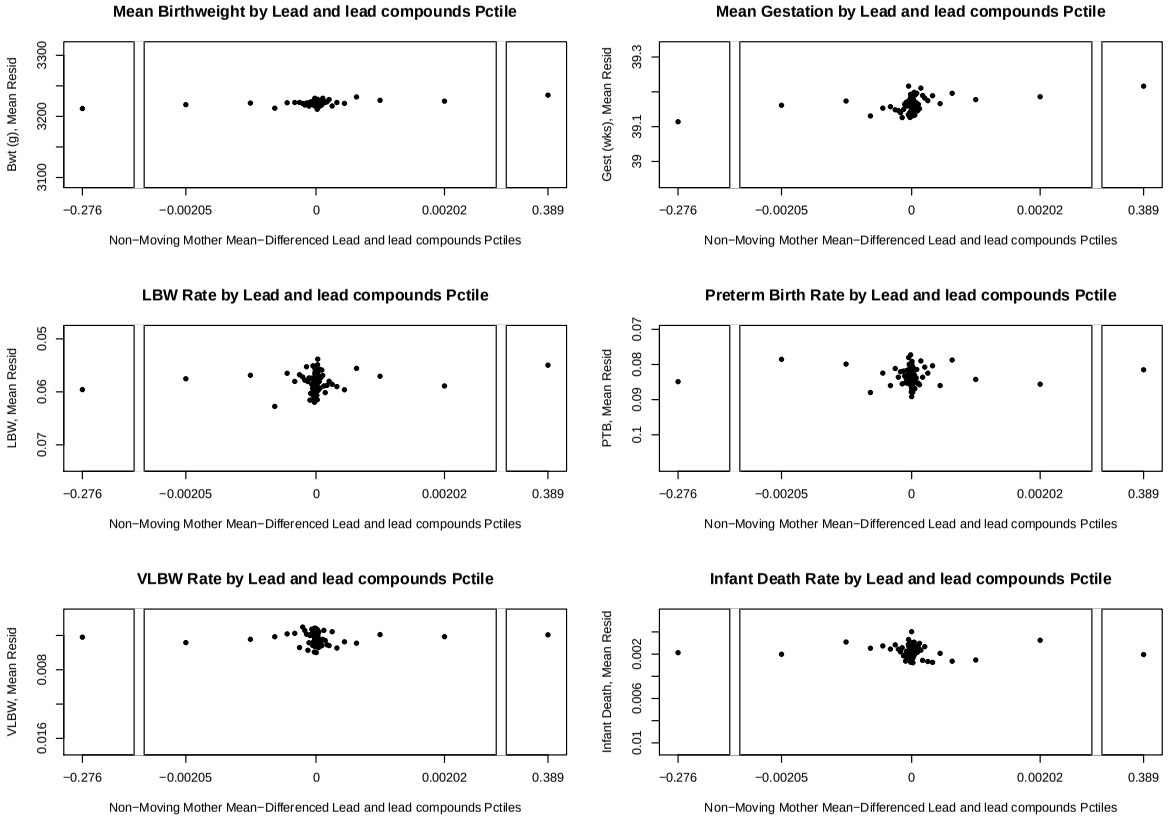
Source: Author's calculations based on 1) birth data from the Texas Center for Health Statistics (1994-2003) merged with 2) km^2-grid-cell-level RSEI chemical concentrations and 3) tract-level data from the 2000 U.S. Census.
Notes: See p.136 for description of method.

Figure 3.6. Toluene - birth outcome relationships, pooled sample (semi-parametric method)



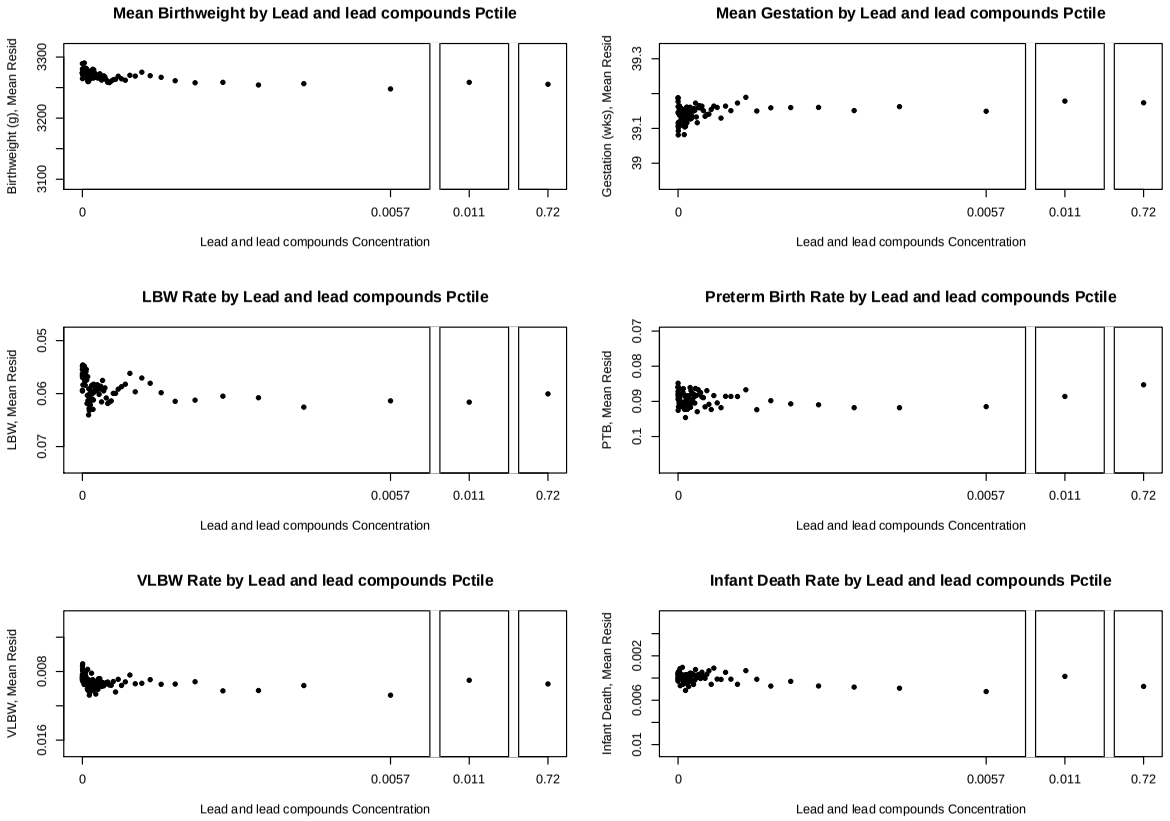
Source: Author's calculations based on 1) birth data from the Texas Center for Health Statistics (1994-2003) merged with 2) km^2-grid-cell-level RSEI chemical concentrations and 3) tract-level data from the 2000 U.S. Census.
Notes: See p.136 for description of method.

Figure 3.7. Lead - birth outcome relationships w/ controls; uses variation within non-moving mothers (semi-parametric method)



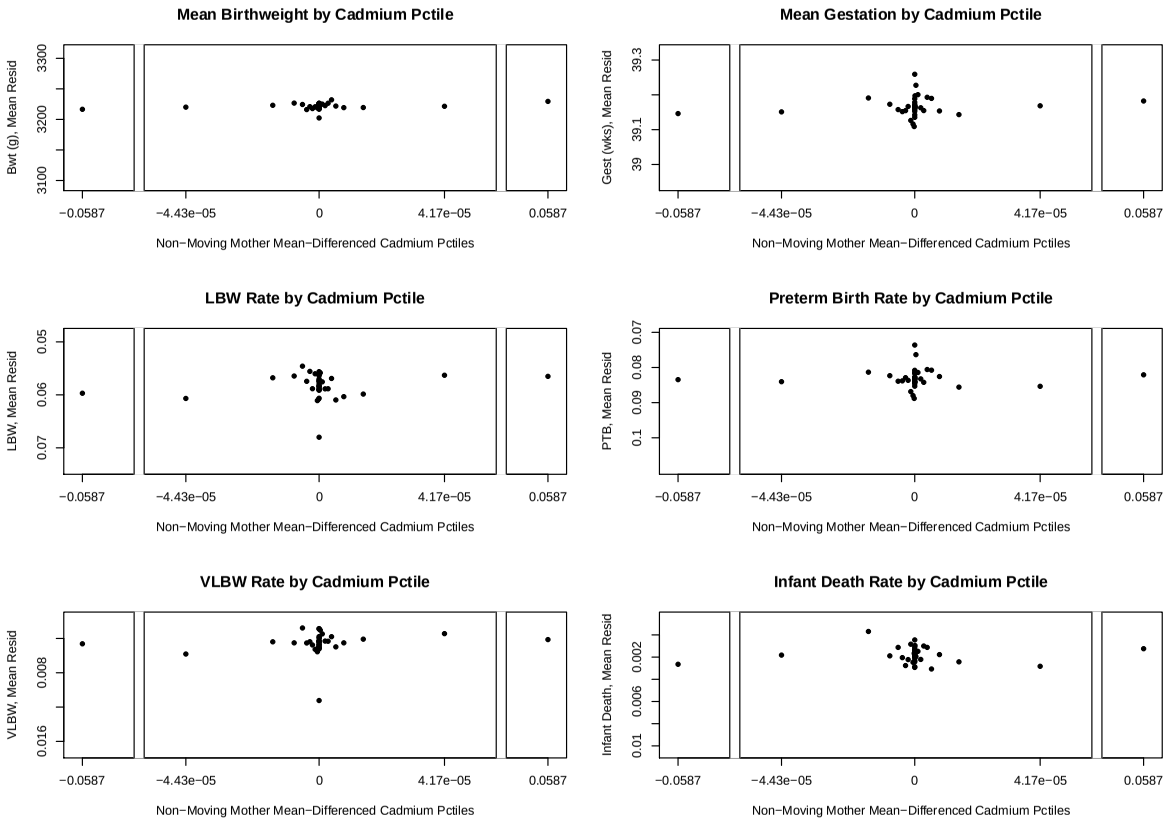
Source: Author's calculations based on 1) birth data from the Texas Center for Health Statistics (1994-2003) merged with 2) km^2-grid-cell-level RSEI chemical concentrations and 3) tract-level data from the 2000 U.S. Census.
Notes: See p.136 for description of method.

Figure 3.8. Lead - birth outcome relationships, pooled sample (semi-parametric method)



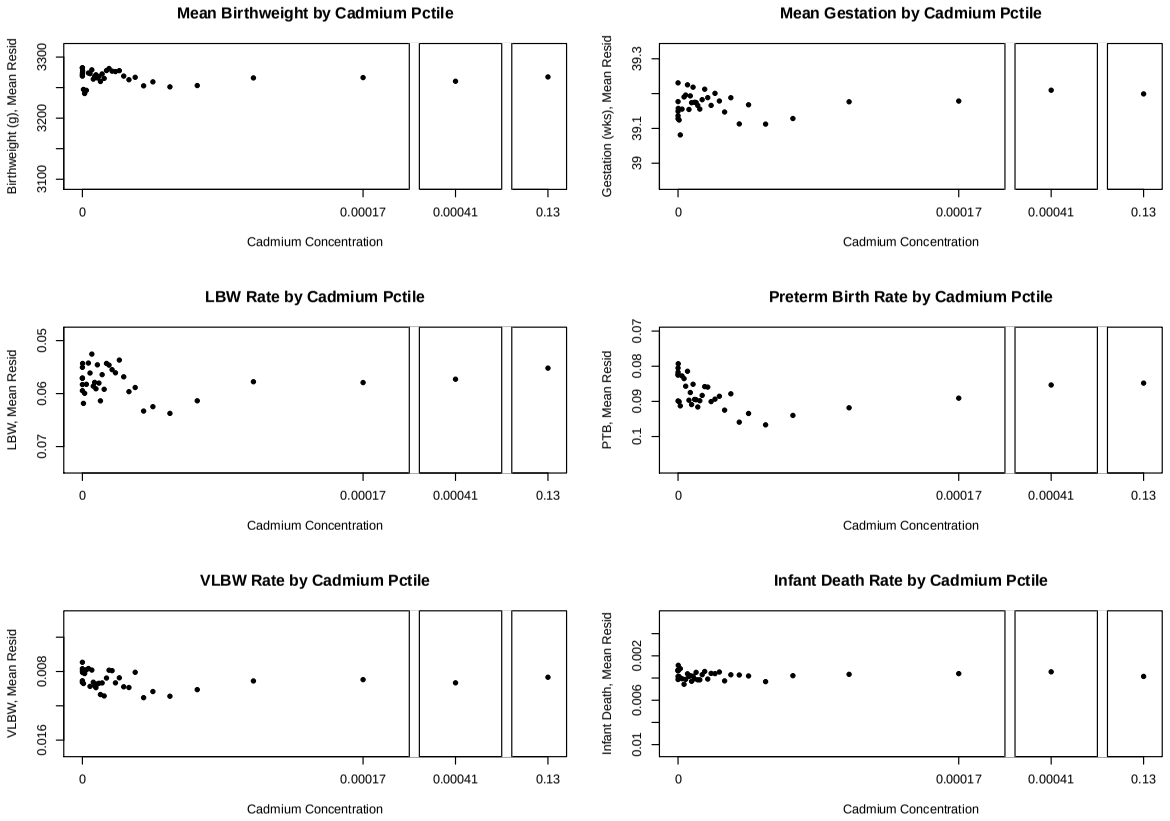
Source: Author's calculations based on 1) birth data from the Texas Center for Health Statistics (1994-2003) merged with 2) km^2-grid-cell-level RSEI chemical concentrations and 3) tract-level data from the 2000 U.S. Census.
Notes: See p.136 for description of method.

Figure 3.9. Cadmium - birth outcome relationships w/ controls; uses variation within non-moving mothers (semi-parametric method)



Source: Author's calculations based on 1) birth data from the Texas Center for Health Statistics (1994-2003) merged with 2) km^2-grid-cell-level RSEI chemical concentrations and 3) tract-level data from the 2000 U.S. Census.
Notes: See p.136 for description of method.

Figure 3.10. Cadmium - birth outcome relationships, pooled sample (semi-parametric method)



Source: Author's calculations based on 1) birth data from the Texas Center for Health Statistics (1994-2003) merged with 2) km^2-grid-cell-level RSEI chemical concentrations and 3) tract-level data from the 2000 U.S. Census.
Notes: See p.136 for description of method.

some that look like sine curves—equally seriously. In the end, we conclude that our findings do not support using this data for health outcome research.

3.5 Discussion and Conclusions

In our analysis, we set out to determine whether the rich RSEI microdata might be suitable for initial research into the health effects of little-studied toxic air pollutants. We conclude that the modeled concentrations available in this dataset, while appropriate for illustrating general patterns in the distribution of pollution, are probably not accurate enough to proxy for exposure in health studies. We chose to focus on four relatively common air toxics—cadmium, epichlorohydrin, lead and toluene—which had been found to negatively affect birth outcomes in previous research. That our results did not align with relationships found in other studies between these chemicals and birthweight, gestation and infant death was an initial clue that the data may not be appropriate. We determined even before conducting the analyses that we may not have had enough statistical power to pick up small effects, given our sample. More convincing, however, is the lack of consistency in results between different specifications, samples and assumed relationships between birth outcomes and pollution variables. In our parametric specifications, most estimated coefficients were not statistically significant, and those that achieved statistical significance in one specification often lost it or changed sign or magnitude in a slightly different specification. Further, removing outliers drastically changed results, and results in the samples with outliers removed changed depending on what share of observations were removed. In our semiparametric specifications, which imposes no constraints on the unknown pollution-outcome relationship, we did observe some patterns, but even these were generally difficult to explain.

Our theoretically preferred specification included mother fixed effects, to control for the effects of those mother characteristics that change slowly or not at all over

time and to avoid bias resulting from correlation between pollution variables and unobserved explanatory variables. We further preferred limiting our sample to infants of mothers who did not move between births, since we found evidence that mothers were systematically moving to cleaner areas between births, and removing those observations that were associated with the top one percent of exposure measures in the full sample, since these might be associated with misreporting. Had our pollution exposure data come from biomarkers or even ambient monitoring, we might have had more trust in the results obtained from estimating this model. However, the apparent advantages of this model are accompanied by a number of downsides. First and foremost, including non-moving mother fixed effects results utilizes only intertemporal variation. This is not a problem for birth outcomes, since these are measured with relative accuracy, but using only year-to-year variation in TRI data raises some doubt. We found, in an analysis not reported here, that around 20 percent of facilities reporting to the TRI in two subsequent years report no change in release. For those that do report a change, it is not known how closely the reported change corresponds to the actual change. Further, since there is no intertemporal variation in our Census income and deprivation measures, which proxied for socioeconomic status, it is possible that changes in reported emissions—to the extent that they correspond to actual changes in emissions—were picking up effects of changing economic patterns. Finally, using fixed effects in general is known to exacerbate bias from measurement error, especially when measurement error in the explanatory variable is correlated with the explanatory variable itself, which could very well be the case in the RSEI data [118].

We also employed specifications that incorporated both geographic and temporal variation, an advantage over the mother fixed effects models, but these models have the distinct disadvantage of probable correlation between pollution variables and the error term. Thus, given the nature of our data, no one model was clearly preferable *a*

priori. Even *post hoc*, none of the models consistently produced results that confirmed our hypotheses, obviating the need to explain why one model "outperformed" another.

As described in some detail above, problems with our data, especially related to exposure misclassification, are likely most responsible for the confusing nature of our results. This misclassification could result from women moving during pregnancy; lack of information on activity patterns and housing quality; misreported emissions data; lack of information on variability in emissions/concentrations throughout the year; lack of data on mobile, small point sources and sources that did not report to the TRI in the early years of our sample; and/or the inaccuracies in the dispersion model's predictions. We undertook this analysis despite these issues because the TRI/RSEI database does provide the best available information on emissions of many of the chemicals included. It is not entirely surprising, however, that these data were not good enough indicators of actual exposure to study health outcomes at the individual level.

The basic conclusion we draw from this exercise is that more research is needed on the developmental effects of the thousands of chemicals in use in production today. While statistical studies cannot explain the biological mechanisms connecting pollutant and health outcome, they are important first steps in flagging public health risks to researchers and policymakers. Such research, however, can only be undertaken with reasonably good exposure data. The EPA currently conducts ambient monitoring of carbon monoxide, nitrogen oxides, sulfer dioxide, VOCs, particulate matter and ammonia, to comply with criteria pollutant standards in the Clean Air Act. Also, thanks in part to a *USA Today* report series using RSEI data and produced in cooperation with researchers at the University of Massachusetts, EPA administrator Lisa Jackson ordered increased monitoring of toxic air pollution outside schools in 2009, initially focusing on 63 schools in 22 states [112, 56]. Expanding the extent of monitoring, both geographically and in terms of chemicals monitored, could increase

the usefulness of studies like this one by making better data available. Given the country's "innocent until proven guilty" attitude toward industrial chemicals and the preference for right-to-know regulation, more useful information on the pollution we are breathing would seem to be a national imperative.

APPENDIX A

ADDITIONAL DISTRIBUTION MEASURES FOR CHAPTER 2

Table A.1. Minority Discrepancy: Race distribution of air toxics in Texas MSAs, all residents, 1999.

Metropolitan Area	Pop (Apr 2000)	Min Tox Shr	Min Pop Shr	Min Discrep	MD Rank	Black Tox Shr	Black Pop Shr	Black Discrep	BD Rank	Hisp Tox Shr	Hisp Pop Shr	Hisp Discrep	HD Rank
Texas (all)	20851820	59.5	47.6	11.9	17	12.1	11.3	0.8	33	44.1	32.0	12.1	6
Lubbock	249700	64.4	38.0	26.5	13	17.2	7.5	9.7	52	45.7	28.0	17.6	11
San Angelo	105781	54.5	36.8	17.7	42	6.4	4.2	2.2	151	46.5	30.5	15.9	15
Longview	194042	43.1	27.6	15.5	50	28.4	18.0	10.4	43	12.8	7.8	5.0	63
Dallas-Fort Worth-Arlington	5161544	55.3	41.0	14.3	60	17.9	13.6	4.3	102	32.6	21.6	10.9	25
Wichita Falls	151524	36.8	24.1	12.7	71	14.6	8.6	6.0	83	15.9	11.0	4.9	65
Abilene	160245	39.1	26.6	12.5	72	8.8	6.5	2.3	148	27.4	17.1	10.3	27
College Station-Bryan	184885	46.5	34.2	12.4	75	20.0	12.1	7.9	63	24.9	17.3	7.6	39
Tyler	174706	44.1	32.0	12.1	77	24.6	19.0	5.7	89	17.9	11.1	6.8	46
Houston-Sugar Land-Baytown	4715407	60.6	51.8	8.8	110	13.5	16.6	-3.1	354	43.3	28.7	14.6	18
Odessa	121123	57.5	48.9	8.6	112	14.3	4.3	9.9	48	41.9	42.4	-0.5	333
Victoria	111663	54.1	46.8	7.2	132	6.4	5.3	1.0	194	45.9	39.2	6.7	48
Austin-Round Rock	1249763	45.4	39.3	6.0	151	9.4	7.6	1.8	165	29.4	26.2	3.2	81
Sherman-Denison	110595	21.6	15.7	5.9	153	8.7	5.8	2.9	132	9.5	6.4	3.1	85
Waco	213517	40.7	35.3	5.4	159	15.3	15.0	0.3	242	23.1	17.9	5.2	61
Brownsville-Harlingen	335227	90.5	85.5	4.9	169	0.4	0.3	0.1	283	89.7	84.5	5.2	60
McAllen-Edinburg-Mission	569463	92.6	89.6	2.9	218	0.4	0.4	0.0	284	91.7	88.4	3.3	79
Midland	116009	40.6	38.1	2.5	228	5.7	7.0	-1.3	343	32.9	28.9	3.9	71
Amarillo	226522	29.4	28.1	1.2	271	5.9	5.3	0.7	214	19.7	19.1	0.7	177
Killeen-Temple-Fort Hood	330714	41.7	40.8	0.8	285	14.5	18.9	-4.4	356	23.8	15.7	8.2	36
Corpus Christi	403280	59.5	59.0	0.5	297	3.8	3.7	0.2	268	53.6	52.7	0.9	161
Beaumont-Port Arthur	385090	35.1	36.0	-0.9	330	22.4	24.6	-2.2	348	9.4	8.0	1.4	132
San Antonio	1711703	57.8	59.4	-1.6	339	6.8	5.9	0.9	203	48.2	50.4	-2.2	353
Laredo	193117	93.4	95.2	-1.8	341	0.4	0.2	0.2	257	92.2	94.4	-2.2	352
El Paso	679622	74.5	83.0	-8.5	359	1.9	2.7	-0.9	335	70.2	78.3	-8.1	360

Source: Author's calculations based on data from the 2000 U.S. Census merged with tract-level RSEI exposure measures for 1999.
Notes: The minority share of toxic score (*Min Tox Shr*) and minority discrepancy (*Min Discrep*) are calculated as described on p. 79. *Min Discrep* and *Min Pop Shr* may not sum to *Min Tox Shr* due to rounding. *MD Rank* ranks Texas MSAs by minority discrepancy (descending) out of 363 MSAs in the U.S. in 2000.

Table A.2. Newborn Minority Discrepancy: Race distribution of air toxics across newborns in Texas MSAs, 1999. (Tract-level pollution measure).

Metropolitan Area	Births in 1999	Min Tox Shr	Min Bir Shr	Min Discrep	Black Tox Shr	Black Bir Shr	Black Discrep	Hisp Tox Shr	Hisp Bir Shr	Hisp Discrep
Texas (all)	310638	70.2	60.4	9.8	11.0	11.9	-0.9	56.5	44.7	11.8
Lubbock	3507	73.6	51.5	22.0	16.7	9.2	7.6	55.5	40.4	15.0
College Station-Bryan	2211	70.3	49.1	21.2	22.6	14.4	8.2	46.5	29.6	16.9
Abilene	2269	49.2	36.6	12.7	9.7	7.4	2.3	37.8	26.3	11.6
Dallas-Fort Worth-Arlington	84946	64.9	52.4	12.5	16.3	14.3	2.0	44.6	33.0	11.6
Longview	2024	51.2	39.2	12.0	31.4	22.6	8.8	18.3	15.1	3.2
San Angelo	1518	61.2	49.4	11.8	4.5	4.2	0.4	55.1	43.9	11.3
Tyler	2471	54.0	43.3	10.6	23.9	20.6	3.3	28.8	21.0	7.7
Houston-Sugar Land-Baytown	76878	70.7	63.5	7.1	11.9	16.8	-4.9	56.1	41.6	14.5
Sherman-Denison	1217	31.8	24.8	7.0	12.9	9.6	3.3	17.2	13.1	4.1
Austin-Round Rock	18745	58.3	52.1	6.1	9.9	8.4	1.5	42.9	39.4	3.5
Odessa	2193	71.1	65.4	5.8	8.5	4.2	4.3	60.5	60.1	0.4
Wichita Falls	1899	35.3	29.7	5.6	15.1	10.3	4.7	17.3	16.9	0.5
Victoria	1555	64.1	61.7	2.4	4.4	4.3	0.1	59.0	56.6	2.4
Waco	2945	51.1	49.5	1.7	17.9	19.2	-1.3	32.1	29.2	2.9
McAllen-Edinburg-Mission	7134	94.5	92.9	1.6	0.1	0.2	-0.1	93.6	91.6	1.9
Killeen-Temple-Fort Hood	5555	46.6	45.1	1.5	16.1	22.6	-6.4	28.6	18.4	10.2
Brownsville-Harlingen	6518	95.6	94.5	1.1	0.2	0.2	0.0	95.1	93.6	1.4
Laredo	5036	98.0	98.7	-0.7	0.2	0.1	0.1	97.3	98.3	-1.0
Corpus Christi	6089	70.3	71.0	-0.8	3.4	3.3	0.1	64.9	65.7	-0.8
Midland	1821	51.3	52.6	-1.3	5.9	7.8	-1.9	43.7	43.1	0.6
Amarillo	3322	36.1	37.4	-1.3	5.9	5.2	0.7	27.9	29.7	-1.8
San Antonio	26403	67.1	70.2	-3.1	5.7	5.3	0.4	58.5	62.3	-3.7
Beaumont-Port Arthur	4845	42.4	45.6	-3.2	26.5	30.3	-3.7	12.1	11.5	0.6
El Paso	12846	83.3	88.0	-4.7	1.8	2.3	-0.5	79.8	84.7	-4.9

Source: Author's calculations based on 1999 birth data from the Texas Center for Health Statistics merged with tract-level RSEI exposure measures for 1999.
Notes: Sample includes all births in Texas in 1999, except for 43,387 births that could not be matched to Census tracts. The minority share of toxic score (*Min Tox Shr*) and minority discrepancy (*Min Discrep*) are calculated as described on p. 79. *Min Discrep* and *Min Bir Shr* may not sum to *Min Tox Shr* due to rounding.

Table A.3. Newborn Minority Discrepancy: Race distribution of air toxics across newborns in Texas MSAs, 1999. (Square kilometer grid cell-level pollution measure).

Metropolitan Area	Births in 1999	Min Tox Shr	Min Bir Shr	Min Discrep	Black Tox Shr	Black Bir Shr	Black Discrep	Hisp Tox Shr	Hisp Bir Shr	Hisp Discrep
Texas (all)	310638	67.6	60.4	7.3	14.2	11.9	2.2	50.3	44.7	5.6
Lubbock	3507	73.9	51.5	22.4	13.5	9.2	4.4	59.4	40.4	19.0
College Station-Bryan	2211	68.0	49.1	18.9	20.9	14.4	6.5	45.8	29.6	16.2
Dallas-Fort Worth-Arlington	84946	69.7	52.4	17.2	22.7	14.3	8.4	43.3	33.0	10.3
Longview	2024	49.8	39.2	10.6	28.8	22.6	6.2	19.5	15.1	4.5
Abilene	2269	46.0	36.6	9.4	13.5	7.4	6.1	30.2	26.3	3.9
Wichita Falls	1899	38.8	29.7	9.1	17.0	10.3	6.7	19.2	16.9	2.3
Odessa	2193	74.2	65.4	8.8	8.7	4.2	4.5	64.2	60.1	4.1
Killeen-Temple-Fort Hood	5555	52.5	45.1	7.4	16.8	22.6	-5.8	34.0	18.4	15.7
Sherman-Denison	1217	31.7	24.8	6.9	12.7	9.6	3.1	17.1	13.1	3.9
Houston-Sugar Land-Baytown	76878	70.0	63.5	6.5	12.0	16.8	-4.8	55.1	41.6	13.5
Austin-Round Rock	18745	58.4	52.1	6.3	9.9	8.4	1.5	42.3	39.4	2.9
Tyler	2471	49.3	43.3	5.9	22.6	20.6	2.0	25.6	21.0	4.5
Victoria	1555	65.2	61.7	3.5	4.8	4.3	0.5	59.7	56.6	3.1
El Paso	12846	91.5	88.0	3.4	1.6	2.3	-0.7	89.1	84.7	4.5
Midland	1821	54.6	52.6	2.1	6.4	7.8	-1.4	46.7	43.1	3.7
Brownsville-Harlingen	6518	96.5	94.5	2.0	0.1	0.2	-0.1	96.1	93.6	2.5
McAllen-Edinburg-Mission	7134	94.9	92.9	2.0	0.1	0.2	-0.1	94.1	91.6	2.4
Beaumont-Port Arthur	4845	46.9	45.6	1.3	30.2	30.3	-0.1	12.9	11.5	1.3
Corpus Christi	6089	71.3	71.0	0.3	3.0	3.3	-0.3	66.7	65.7	0.9
San Angelo	1518	49.5	49.4	0.1	9.7	4.2	5.6	38.3	43.9	-5.6
Waco	2945	48.9	49.5	-0.6	19.1	19.2	0.0	28.6	29.2	-0.6
Laredo	5036	97.8	98.7	-0.9	0.2	0.1	0.1	97.4	98.3	-0.9
Amarillo	3322	36.4	37.4	-1.0	5.7	5.2	0.5	28.4	29.7	-1.2
San Antonio	26403	66.5	70.2	-3.8	4.9	5.3	-0.5	59.1	62.3	-3.1

Source: Author's calculations based on 1999 birth data from the Texas Center for Health Statistics merged with square kilometer grid cell-level RSEI exposure measures for 1999.
Notes: Sample includes all births in Texas in 1999, except for 43,387 births that could not be matched to Census tracts. The minority share of toxic score (*Min Tox Shr*) and minority discrepancy (*Min Discrep*) are calculated as described on p. 79. *Min Discrep* and *Min Bir Shr* may not sum to *Min Tox Shr* due to rounding.

Table A.4. Share of population with high RSEI scores by race, Texas MSAs, 1999.

Metropolitan Area	Pop (Apr 2000)	White				Black				Hispanic			
		Race shr in pop	% > US 50p	% > US 75p	% > US 90p	Race shr in pop	% > US 50p	% > US 75p	% > US 90p	Race shr in pop	% > US 50p	% > US 75p	% > US 90p
Texas (all)	20851820	52.4	33.5	16.7	5.9	11.3	50.7	28.3	11.2	32.0	40.6	24.4	10.7
Abilene	160245	73.4	13.5	3.6	1.2	6.5	27.1	3.9	1.3	17.1	30.2	4.4	2.5
Amarillo	226522	71.9	57.5	2.2	0.0	5.3	64.3	0.1	0.0	19.1	46.7	0.5	0.0
Austin-Round Rock	1249763	60.7	13.5	3.2	0.4	7.6	21.4	4.9	0.3	26.2	18.0	4.6	0.2
Beaumont-Port Arthur	385090	64.0	98.7	87.2	51.1	24.6	100.0	98.9	74.7	8.0	99.9	97.2	73.4
Brownsville-Harlingen	335227	14.5	1.9	0.6	0.0	0.3	1.9	0.0	0.0	84.5	4.3	1.1	0.0
College Station-Bryan	184885	65.8	1.2	1.2	0.0	12.1	1.3	1.3	0.0	17.3	0.2	0.2	0.0
Corpus Christi	403280	41.0	50.9	19.7	10.9	3.7	67.6	22.1	9.4	52.7	74.6	20.5	9.0
Dallas-Fort Worth-Arlington	5161544	59.0	11.9	2.8	0.3	13.6	18.6	5.0	1.7	21.6	31.1	8.6	0.6
El Paso	679622	17.0	99.7	95.8	52.8	2.7	100.0	98.3	36.9	78.3	99.1	86.4	39.6
Houston-Sugar Land-Baytown	4715407	48.2	83.6	48.1	16.1	16.6	94.7	57.4	18.5	28.7	93.6	65.6	31.3
Killeen-Temple-Fort Hood	330714	59.2	19.8	4.9	0.0	18.9	11.8	5.2	0.0	15.7	24.5	11.1	0.0
Laredo	193117	4.8	34.9	19.2	5.4	0.2	38.3	30.3	10.0	94.4	16.9	11.2	5.5
Longview	194042	72.4	96.5	48.7	13.5	18.0	93.9	58.6	37.6	7.8	95.6	57.5	35.5
Lubbock	249700	62.0	0.4	0.0	0.0	7.5	0.3	0.0	0.0	28.0	2.0	0.0	0.0
McAllen-Edinburg-Mission	569463	10.4	0.0	0.0	0.0	0.4	0.0	0.0	0.0	88.4	0.0	0.0	0.0
Midland	116009	61.9	0.0	0.0	0.0	7.0	0.0	0.0	0.0	28.9	0.0	0.0	0.0
Odessa	121123	51.1	22.3	3.8	0.1	4.3	42.8	23.9	22.6	42.4	18.6	3.6	2.1
San Angelo	105781	63.2	0.0	0.0	0.0	4.2	0.0	0.0	0.0	30.5	0.0	0.0	0.0
San Antonio	1711703	40.6	14.0	2.0	0.0	5.9	12.3	0.6	0.0	50.4	7.9	1.5	0.0
Sherman-Denison	110595	84.3	0.0	0.0	0.0	5.8	0.0	0.0	0.0	6.4	0.0	0.0	0.0
Tyler	174706	68.0	16.3	2.0	2.0	19.0	48.0	4.3	4.3	11.1	59.9	3.9	3.9
Victoria	111663	53.2	70.9	7.5	0.0	5.3	85.0	4.8	0.0	39.2	77.2	7.5	0.0
Waco	213517	64.7	20.6	2.2	0.0	15.0	13.9	2.6	0.0	17.9	17.9	7.8	0.0
Wichita Falls	151524	75.9	19.7	5.9	1.5	8.6	39.2	19.3	3.6	11.0	32.6	17.8	2.7

Source: Author's calculations based on data from the 2000 U.S. Census merged with tract-level RSEI exposure measures for 1999.
Notes: Columns with headings beginning *% > US...* represent the share of the population (by race/ethnicity) with 1999 tract-level RSEI scores above the U.S. population-weighted 50, 75 and 90 percentile RSEI scores. In 1999, the population-weighted median U.S. RSEI score was 77.7, the 75 percentile score was 247.0, and the 90 percentile score was 725.5.

Table A.5. Share of newborns with high RSEI scores by race, Texas MSAs, 1999. (Tract-level pollution measure).

Metropolitan Area	Births 1999	White Race shr in pop	White % > US 50p	White % > US 75p	White % > US 90p	Black Race shr in pop	Black % > US 50p	Black % > US 75p	Black % > US 90p	Hispanic Race shr in pop	Hispanic % > US 50p	Hispanic % > US 75p	Hispanic % > US 90p
Texas (all)	310638	39.6	35.2	17.7	6.4	11.9	52.3	28.3	11.2	44.7	14.0	7.5	3.0
Abilene	2269	63.4	19.1	2.6	0.8	7.4	39.6	4.1	1.8	26.3	34.6	4.7	3.4
Amarillo	3322	62.6	62.6	1.5	0.0	5.2	79.2	1.2	0.0	29.7	47.9	0.2	0.0
Austin-Round Rock	18745	47.9	13.7	3.4	0.2	8.4	24.8	5.5	0.1	39.4	21.0	6.0	0.1
Beaumont-Port Arthur	4845	54.4	99.8	88.8	53.6	30.3	100.0	99.5	74.4	11.5	100.0	98.9	83.2
Brownsville-Harlingen	6518	5.5	2.5	0.8	0.0	0.2	8.3	0.0	0.0	93.6	5.7	1.3	0.0
College Station-Bryan	2211	50.9	0.1	0.1	0.0	14.4	0.9	0.9	0.0	29.6	0.2	0.2	0.0
Corpus Christi	6089	29.0	50.6	19.3	11.7	3.3	66.7	21.9	9.5	65.7	73.5	20.6	9.2
Dallas-Fort Worth-Arlington	84946	47.6	11.7	2.9	0.3	14.3	18.2	4.7	1.8	33.0	32.7	9.1	0.6
El Paso	12846	12.0	99.9	96.5	57.4	2.3	100.0	99.3	51.2	84.7	99.9	88.2	41.9
Houston-Sugar Land-Baytown	76878	36.5	84.5	48.6	16.4	16.8	96.7	55.8	17.4	41.6	94.7	66.9	32.7
Killeen-Temple-Fort Hood	5555	54.9	15.9	4.7	0.0	22.6	11.5	4.4	0.0	18.4	26.5	10.6	0.0
Laredo	5036	1.3	34.4	20.3	3.1	0.1	33.3	33.3	33.3	98.3	13.2	9.8	5.7
Longview	2024	60.8	98.7	64.2	24.4	22.6	99.1	71.2	50.4	15.1	100.0	68.2	46.6
Lubbock	3507	48.5	0.4	0.0	0.0	9.2	1.2	0.0	0.0	40.4	2.3	0.0	0.0
McAllen-Edinburg-Mission	7134	7.1	0.0	0.0	0.0	0.2	0.0	0.0	0.0	91.6	0.0	0.0	0.0
Midland	1821	47.4	0.0	0.0	0.0	7.8	0.0	0.0	0.0	43.1	0.0	0.0	0.0
Odessa	2193	34.6	19.6	2.2	0.3	4.2	25.0	10.9	9.8	60.1	18.4	3.3	2.0
San Angelo	1518	50.6	0.0	0.0	0.0	4.2	0.0	0.0	0.0	43.9	0.0	0.0	0.0
San Antonio	26403	29.8	14.6	2.2	0.0	5.3	11.4	0.4	0.0	62.3	7.7	1.5	0.0
Sherman-Denison	1217	75.2	0.0	0.0	0.0	9.6	0.0	0.0	0.0	13.1	0.0	0.0	0.0
Tyler	2471	56.7	16.1	2.0	2.0	20.6	47.3	3.9	3.9	21.0	65.0	4.0	4.0
Victoria	1555	38.3	73.3	8.2	0.0	4.3	83.6	1.5	0.0	56.6	80.8	3.5	0.0
Waco	2945	50.5	24.5	2.5	0.0	19.2	14.0	3.7	0.0	29.2	17.8	7.4	0.0
Wichita Falls	1899	70.3	24.0	6.6	2.0	10.3	35.2	15.3	6.1	16.9	25.6	11.9	1.9

Source: Author's calculations based on 1999 birth data from the Texas Center for Health Statistics merged with tract-level RSEI exposure measures for 1999.
Notes: Sample includes all births in Texas in 1999, except for 43,387 births that could not be matched to Census tracts. Columns with headings beginning *% > US...* represent the share of births (by race/ethnicity) with 1999 tract-level RSEI scores above the U.S. population-weighted 50, 75 and 90 percentile RSEI scores. In 1999, the population-weighted median U.S. RSEI score was 77.7, the 75 percentile score was 247.0, and the 90 percentile score was 725.5.

Table A.6. Share of newborns with high RSEI scores by race, Texas MSAs, 1999. (Square kilometer grid cell-level pollution measure).

Metropolitan Area	Births 1999	White Race shr in pop	White % > US 50p	White % > US 75p	White % > US 90p	Black Race shr in pop	Black % > US 50p	Black % > US 75p	Black % > US 90p	Hispanic Race shr in pop	Hispanic % > US 50p	Hispanic % > US 75p	Hispanic % > US 90p
Texas (all)	310638	39.6	30.8	15.0	5.3	11.9	49.1	25.6	9.0	44.7	13.1	6.8	2.4
Abilene	2269	63.4	4.7	0.9	0.8	7.4	8.9	3.6	3.6	26.3	6.9	1.3	1.3
Amarillo	3322	62.6	4.6	0.2	0.0	5.2	0.0	0.0	0.0	29.7	1.0	0.1	0.0
Austin-Round Rock	18745	47.9	12.2	2.7	0.4	8.4	18.5	3.7	0.2	39.4	16.5	3.0	0.7
Beaumont-Port Arthur	4845	54.4	99.3	88.2	48.2	30.3	100.0	99.3	74.0	11.5	100.0	98.6	81.2
Brownsville-Harlingen	6518	5.5	2.2	0.8	0.0	0.2	0.0	0.0	0.0	93.6	4.7	2.5	0.0
College Station-Bryan	2211	50.9	1.9	0.0	0.0	14.4	4.7	0.0	0.0	29.6	4.7	0.0	0.0
Corpus Christi	6089	29.0	49.9	16.9	2.8	3.3	68.7	15.4	2.0	65.7	74.0	16.8	3.9
Dallas-Fort Worth-Arlington	84946	47.6	8.7	1.7	0.4	14.3	15.3	4.1	1.4	33.0	27.3	7.0	1.6
El Paso	12846	12.0	37.0	4.9	1.0	2.3	41.0	8.8	0.3	84.7	42.1	9.1	3.9
Houston-Sugar Land-Baytown	76878	36.5	81.6	44.7	15.1	16.8	95.2	51.1	12.6	41.6	93.9	63.4	30.2
Killeen-Temple-Fort Hood	5555	54.9	15.8	6.6	1.4	22.6	14.3	6.4	1.6	18.4	27.5	15.9	3.6
Laredo	5036	1.3	29.7	9.4	1.6	0.1	33.3	33.3	0.0	98.3	12.9	6.9	0.7
Longview	2024	60.8	90.7	60.7	24.6	22.6	95.4	70.1	52.4	15.1	89.5	67.5	48.2
Lubbock	3507	48.5	0.3	0.2	0.0	9.2	1.9	0.0	0.0	40.4	2.9	0.6	0.0
McAllen-Edinburg-Mission	7134	7.1	0.0	0.0	0.0	0.2	0.0	0.0	0.0	91.6	0.0	0.0	0.0
Midland	1821	47.4	0.0	0.0	0.0	7.8	0.0	0.0	0.0	43.1	0.0	0.0	0.0
Odessa	2193	34.6	18.3	0.9	0.3	4.2	28.3	12.0	7.6	60.1	20.2	3.7	2.1
San Angelo	1518	50.6	0.0	0.0	0.0	4.2	0.0	0.0	0.0	43.9	0.0	0.0	0.0
San Antonio	26403	29.8	9.8	1.0	0.1	5.3	2.3	0.1	0.0	62.3	4.9	0.9	0.0
Sherman-Denison	1217	75.2	0.0	0.0	0.0	9.6	0.0	0.0	0.0	13.1	0.0	0.0	0.0
Tyler	2471	56.7	20.2	7.1	1.2	20.6	54.1	3.9	1.4	21.0	69.0	5.2	1.0
Victoria	1555	38.3	75.8	2.2	0.5	4.3	86.6	3.0	0.0	56.6	79.2	1.0	0.2
Waco	2945	50.5	10.6	0.3	0.2	19.2	9.2	0.2	0.2	29.2	9.4	0.0	0.0
Wichita Falls	1899	70.3	27.2	4.0	1.1	10.3	39.3	9.7	7.1	16.9	25.9	10.6	1.9

Source: Author's calculations based on 1999 birth data from the Texas Center for Health Statistics merged with square kilometer grid cell-level RSEI exposure measures for 1999.
Notes: Sample includes all births in Texas in 1999, except for 43,387 births that could not be matched to Census tracts. Columns with headings beginning *% > US...* represent the share of births (by race/ethnicity) with 1999 grid cell-level RSEI scores above the U.S. population-weighted 50, 75 and 90 percentile RSEI scores. In 1999, the population-weighted median U.S. RSEI score was 77.7, the 75 percentile score was 247.0, and the 90 percentile score was 725.5.

APPENDIX B

SEGREGATION MEASURES FOR CHAPTER 2

Formulas for the dissimilarity index (D), the isolation index ($_mP^*_m$) and the index of spatial proximity (S_P):

$$D = \sum_{i=1}^{k} [t_i|p_{im} - P_m|]/(2TP_m(1 - P_m)) \tag{B.1}$$

$$_mP^*_m = \sum_{i=1}^{k} [(t_i p_{im}/TP_m)(p_{im})] \tag{B.2}$$

$$S_P = (TP_m P_{mm} + TP_n P_{nn})/NP_{tt} \tag{B.3}$$

where

T = number of CBSA residents

t_i = number of residents in tract i

k = number of tracts in CBSA

P_m = proportion of CBSA residents of racial/ethnic group m

p_{im} = proportion of tract i's residents of racial/ethnic group m

$P_{mn} = \sum_{i=1}^{k} \sum_{j=1}^{k} [(t_i p_{im} t_j p_{jn} c_{ij})/TP_m TP_n]$

$c_{ij} = e^{-dij}$

d_{ij} = distance between tract i and tract j

APPENDIX C

ADDITIONAL TABLES FOR CHAPTER 3

Table C.1. Regressions with birth outcomes and pollutants used in Currie and Schmeider (2009), non-moving mother fixed effects, top 1% of obs by exposure removed

	Gest (wks)	Bwt (g)	PTB	LBW	VLBW	Infant Death
All, aggregate	0.0114	1.6215	-0.1606	-0.1091	-0.0417	-0.0126
p-val	(0.1199)	(0.2556)	(0.0918)	(0.1348)	(0.1922)	(0.6201)
moms w/ var	*223,099*	*261,994*	*40,865*	*23,840*	*4,375*	*3,040*
All, tox-wt agg	0.0009	0.1348	0.0196	-0.0081	-0.0135	-0.0105
p-val	(0.7539)	(0.8042)	(0.5929)	(0.7769)	(0.2835)	(0.2762)
moms w/ var	*223,151*	*262,002*	*40,882*	*23,875*	*4,392*	*3,047*
Developmental, agg	0.0631	2.6769	-0.9522	-0.5999	-0.3678	-0.1009
p-val	(0.0853)	(0.7117)	(0.0444)	(0.1043)	(0.0169)	(0.4261)
moms w/ var	*220,593*	*259,003*	*40,366*	*23,577*	*4,335*	*3,009*
Developmental, tox-wt agg	0.0151	-2.7349	0.0196	-0.0574	-0.0674	-0.0217
p-val	(0.1657)	(0.1905)	(0.8906)	(0.6022)	(0.1501)	(0.5956)
moms w/ var	*220,685*	*259,106*	*40,402*	*23,638*	*4,362*	*3,027*
VOCs, agg	0.0836	4.4581	-1.3368	-0.4910	-0.2412	-0.1525
p-val	(0.0799)	(0.6277)	(0.0288)	(0.2982)	(0.2405)	(0.3924)
moms w/ var	*214,646*	*252,090*	*39,188*	*22,896*	*4,215*	*2,907*
VOCs, tox-wt agg	0.2405	-14.8697	-2.9337	-2.3536	-0.5524	0.3864
p-val	(0.0235)	(0.4689)	(0.0278)	(0.0232)	(0.2272)	(0.2512)
moms w/ var	*214,201*	*251,559*	*39,165*	*22,928*	*4,223*	*2,919*
Toluene	0.1014	16.5978	-1.7316	-0.9165	-0.5483	-0.4362
p-val	(0.1084)	(0.1767)	(0.0376)	(0.1497)	(0.0415)	(0.0641)
moms w/ var	*209,974*	*246,588*	*38,341*	*22,386*	*4,112*	*2,849*
Epichlorohydrin	4.803	-1,732.179	-36.491	-64.619	-1.722	14.773
p-val	(0.3165)	(0.0630)	(0.5468)	(0.1681)	(0.9331)	(0.4106)
moms w/ var	*104,786*	*122,782*	*18,903*	*11,206*	*2,102*	*1,454*
Metals, agg	9.07	3,181.74	3.21	-50.59	-25.51	1.54
p-val	(0.0788)	(0.0010)	(0.9612)	(0.3249)	(0.2430)	(0.9272)
moms w/ var	*212,084*	*248,932*	*38,819*	*22,701*	*4,195*	*2,901*
Metals, tox-wt agg	0.04983	7.2116	0.2526	-0.0034	-0.0681	0.0591
p-val	(0.0926)	(0.1933)	(0.5112)	(0.9907)	(0.5951)	(0.5693)
moms w/ var	*212,028*	*248,843*	*38,790*	*22,698*	*4,191*	*2,898*
Lead	12.50	4,536.87	6.78	-93.34	-35.70	-1.01
p-val	(0.0423)	(0.0001)	(0.9296)	(0.1204)	(0.1687)	(0.9619)
moms w/ var	*209,613*	*246,014*	*38,335*	*22,409*	*4,137*	*2,863*
Cadmium	337.9	56,897.9	-3,644.3	-1,089.8	-630.8	149.8
p-val	(0.1155)	(0.1695)	(0.1740)	(0.6078)	(0.4759)	(0.8218)
moms w/ var	*75,894*	*88,519*	*13,562*	*7,789*	*1,419*	*995*

Source: Author's calculations based on 1) birth data from the Texas Center for Health Statistics (1994-2003) merged with 2) km^2-grid-cell-level RSEI chemical concentrations and 3) tract-level data from the 2000 U.S. Census.
Notes: Dependent variable listed in column headings. Sample excludes top 1% of all infants by exposure from full CBSA sample. Regressions use variation within mothers who did not move between births (non-moving mother fixed effects). Standard errors adjusted for heteroskedasticity and clustered at mother level. "Moms w/ var" indicates the number of mothers in the sample who had between-birth variation in both chemical exposure and infant health outcome. Coefficients on PTB, LBW, VLBW and infant death multiplied by 100.

Table C.2. Regressions with birth outcomes and pollutants used in Currie and Schmeider (2009), non-moving mother fixed effects

	Gest (wks)	Bwt (g)	PTB	LBW	VLBW	Infant Death
All, aggregate	0.0045	0.4063	-0.0412	-0.0264	-0.0181	0.0057
p-val	(0.0239)	(0.2403)	(0.0724)	(0.1298)	(0.0352)	(0.6235)
moms w/ var	*225,728*	*265,060*	*41,346*	*24,208*	*4,433*	*3,102*
All, tox-wt agg	0.0011	0.0474	-0.0027	-0.0023	-0.0031	0.0023
p-val	(0.0156)	(0.5454)	(0.6567)	(0.6475)	(0.0502)	(0.5434)
moms w/ var	*225,728*	*265,060*	*41,346*	*24,208*	*4,433*	*3,102*
Developmental, agg	0.0128	2.2451	-0.1524	-0.0873	-0.0814	-0.0119
p-val	(0.1164)	(0.1262)	(0.0829)	(0.2173)	(0.0328)	(0.6278)
moms w/ var	*223,124*	*261,980*	*40,837*	*23,928*	*4,387*	*3,067*
Developmental, tox-wt agg	0.0019	-0.0693	0.0246	0.0096	-0.0058	-0.0061
p-val	(0.3015)	(0.8161)	(0.2253)	(0.5686)	(0.2375)	(0.1187)
moms w/ var	*223,124*	*261,980*	*40,837*	*23,928*	*4,387*	*3,067*
VOCs, agg	0.0053	2.1216	-0.1395	-0.0941	-0.0589	-0.0045
p-val	(0.5700)	(0.1999)	(0.1450)	(0.2196)	(0.1311)	(0.8593)
moms w/ var	*217,106*	*255,011*	*39,656*	*23,258*	*4,276*	*2,976*
VOCs, tox-wt agg	0.0058	0.0832	0.0024	0.0071	-0.0068	-0.0056
p-val	(0.0320)	(0.8297)	(0.8656)	(0.8042)	(0.0897)	(0.1053)
moms w/ var	*217,105*	*255,010*	*39,655*	*23,258*	*4,276*	*2,976*
Toluene	-0.0011	2.2329	-0.0976	-0.0903	-0.0467	-0.0044
p-val	(0.9156)	(0.2415)	(0.3555)	(0.2844)	(0.3036)	(0.8772)
moms w/ var	*212,396*	*249,448*	*38,805*	*22,743*	*4,182*	*2,922*
Epichlorohydrin	0.105	2.020	0.172	0.187	-0.105	-0.120
p-val	(0.0423)	(0.7760)	(0.5010)	(0.7243)	(0.1132)	(0.0839)
moms w/ var	*107,775*	*126,322*	*19,390*	*11,515*	*2,156*	*1,507*
Metals, agg	1.89	536.07	-0.14	-12.03	-2.38	-3.89
p-val	(0.0524)	(0.0016)	(0.9919)	(0.1802)	(0.4494)	(0.1731)
moms w/ var	*214,319*	*251,596*	*39,196*	*22,988*	*4,225*	*2,939*
Metals, tox-wt agg	0.00028	0.2664	0.0456	0.0132	-0.0032	-0.0125
p-val	(0.9269)	(0.6074)	(0.2865)	(0.5610)	(0.7255)	(0.1384)
moms w/ var	*214,255*	*251,507*	*39,186*	*22,982*	*4,225*	*2,937*
Lead	3.02	810.25	-7.60	-24.29	-3.22	-4.65
p-val	(0.0219)	(0.0001)	(0.6781)	(0.0459)	(0.4491)	(0.2412)
moms w/ var	*212,263*	*249,189*	*38,799*	*22,747*	*4,176*	*2,903*
Cadmium	-26.4	-1,016.8	239.0	17.0	41.0	-75.4
p-val	(0.1821)	(0.6200)	(0.1969)	(0.8234)	(0.2400)	(0.1367)
moms w/ var	*85,856*	*100,566*	*15,164*	*8,769*	*1,614*	*1,115*

Source: Author's calculations based on 1) birth data from the Texas Center for Health Statistics (1994-2003) merged with 2) km^2-grid-cell-level RSEI chemical concentrations and 3) tract-level data from the 2000 U.S. Census.
Notes: Dependent variable listed in column headings. Regressions use variation within mothers who did not move between births (non-moving mother fixed effects). Standard errors adjusted for heteroskedasticity and clustered at mother level. "Moms w/ var" indicates the number of mothers in the sample who had between-birth variation in both chemical exposure and infant health outcome. Coefficients on PTB, LBW, VLBW and infant death multiplied by 100.

Table C.3. Regressions with birth outcomes and pollutants used in Currie and Schmeider (2009), CBSA fixed effects, top 1% of obs by exposure removed

	Gest (wks)	Bwt (g)	PTB	LBW	VLBW	Infant Death
All, aggregate	-0.00049	1.170	-0.0088	-0.0286	-0.0073	0.00041
p-val	(0.9171)	(0.4127)	(0.8698)	(0.3252)	(0.3742)	(0.8573)
All, tox-wt agg	0.00058	0.9256	-0.0128	-0.0253	-0.0034	-0.0024
p-val	(0.6535)	(0.1022)	(0.4863)	(0.0138)	(0.1667)	(0.0007)
Developmental, agg	0.0104	5.339	-0.115	-0.103	-0.0297	-0.0061
p-val	(0.5380)	(0.5089)	(0.6234)	(0.6027)	(0.4691)	(0.6808)
Developmental, tox-wt agg	0.0142	6.076	-0.138	-0.171	-0.0355	-0.0014
p-val	(0.0000)	(0.0001)	(0.0514)	(0.0002)	(0.0057)	(0.8589)
VOCs, agg	0.0196	7.198	-0.226	-0.218	-0.026	-0.017
p-val	(0.2692)	(0.4215)	(0.3771)	(0.2333)	(0.5058)	(0.2092)
VOCs, tox-wt agg	0.0278	26.95	-0.648	-0.566	-0.0286	0.00086
p-val	(0.1977)	(0.1060)	(0.0147)	(0.0277)	(0.7291)	(0.9729)
Toluene	0.0316	4.642	-0.240	-0.215	-0.0226	-0.0421
p-val	(0.3365)	(0.7087)	(0.5975)	(0.4833)	(0.7717)	(0.0538)
Epichlorohydrin	-1.421	819.9	-23.74	-33.73	-7.996	2.209
p-val	(0.2209)	(0.0161)	(0.0249)	(0.0018)	(0.0000)	(0.0966)
Metals, agg	3.116	860.9	-22.28	-22.44	-1.854	3.125
p-val	(0.4212)	(0.4864)	(0.3379)	(0.4589)	(0.7881)	(0.4813)
Metals, tox-wt agg	0.022	2.650	-0.213	-0.0135	0.0123	0.0265
p-val	(0.2505)	(0.6429)	(0.0888)	(0.9363)	(0.7765)	(0.2587)
Lead	2.588	380.2	0.086	-10.58	-1.379	4.074
p-val	(0.6412)	(0.8134)	(0.9981)	(0.7955)	(0.8883)	(0.5260)
Cadmium	212.3	41,395	-1,201	-1,359	79.53	-207.8
p-val	(0.2351)	(0.3880)	(0.2593)	(0.0864)	(0.5710)	(0.0154)

Source: Author's calculations based on 1) birth data from the Texas Center for Health Statistics (1994-2003) merged with 2) km^2-grid-cell-level RSEI chemical concentrations and 3) tract-level data from the 2000 U.S. Census.
Notes: Dependent variable listed in column headings. Sample excludes top 1% of all infants by exposure. Regressions include CBSA fixed effects. Standard errors adjusted for heteroskedasticity and clustered at CBSA level. Coefficients on PTB, LBW, VLBW and infant death multiplied by 100.

Table C.4. Gestation regressions with alternate pollution variables

Pollution Variable	N	Bivar	p-val	Multivar	p-val	CBSA FE	p-val	Mom FE	p-val	moms w/ var	No Move Mom FE	p-val	moms w/ var
Toluene (μg/m^3)													
A) Fug+stack, yr wts	2,868,306	-0.00248	(0.3077)	0.00039	(0.8818)	-0.00047	(0.8217)	0.00126	(0.6930)	444,809	-0.00111	(0.9156)	212,439
B) Fug+stack concep yr	2,868,306	-0.00184	(0.4156)	0.00098	(0.7241)	0.00013	(0.9525)	0.00214	(0.4867)	441,517	0.00914	(0.3908)	210,451
C) Fug+stack, birth yr	2,868,306	-0.00304	(0.2437)	-0.00024	(0.9264)	-0.00107	(0.6176)	0.00000	(0.9991)	439,205	-0.00612	(0.4836)	209,228
Like (A), top 2.5 pct rm	2,796,587	0.09021	(0.2962)	0.20243	(0.0084)	0.07208	(0.1121)	0.13192	(0.0051)	428,821	0.27378	(0.0053)	206,360
Like (A), top 1.0 pct rm	2,839,622	0.03390	(0.4544)	0.09845	(0.0257)	0.03159	(0.3365)	0.07628	(0.0112)	438,517	0.10143	(0.1084)	210,102
Like (A), fugitive only	2,868,306	-0.00232	(0.3075)	-0.00021	(0.9251)	-0.00047	(0.8096)	0.00143	(0.6631)	435,055	0.00057	(0.9589)	206,418
Epichlorohydrin (μg/m^3)													
A) Fug+stack, yr wts	2,868,306	0.035	(0.0000)	0.041	(0.0000)	0.028	(0.0000)	0.048	(0.1808)	229,767	0.105	(0.0423)	107,761
B) Fug+stack concep yr	2,868,306	0.035	(0.0000)	0.040	(0.0000)	0.028	(0.0000)	0.045	(0.1602)	196,158	0.100	(0.0700)	91,860
C) Fug+stack, birth yr	2,868,306	0.038	(0.0000)	0.044	(0.0000)	0.030	(0.0000)	0.049	(0.1967)	210,574	0.106	(0.0312)	98,370
Like (A), top 2.5 pct rm	2,796,599	11.225	(0.0789)	16.085	(0.0159)	-4.874	(0.0301)	-0.945	(0.8521)	214,092	7.051	(0.3894)	100,974
Like (A), top 1.0 pct rm	2,839,622	6.292	(0.0569)	9.330	(0.0105)	-1.421	(0.2209)	0.734	(0.8038)	223,041	4.803	(0.3165)	104,859
Like (A), fugitive only	2,868,306	0.032	(0.0000)	0.037	(0.0000)	0.026	(0.0000)	0.046	(0.2151)	147,424	0.102	(0.0535)	68,963
Lead (μg/m^3)													
A) Fug+stack, yr wts	2,868,306	-0.215	(0.6853)	0.490	(0.3535)	0.909	(0.1251)	2.108	(0.0066)	445,734	3.021	(0.0219)	212,315
B) Fug+stack concep yr	2,868,306	-0.145	(0.7611)	0.463	(0.3332)	0.959	(0.0782)	2.111	(0.0043)	436,690	3.150	(0.0147)	206,238
C) Fug+stack, birth yr	2,868,306	-0.419	(0.4542)	0.335	(0.5274)	0.610	(0.3601)	1.611	(0.0390)	440,223	2.384	(0.0576)	208,281
Like (A), top 2.5 pct rm	2,796,591	-0.712	(0.9415)	4.743	(0.5763)	-1.060	(0.9165)	9.531	(0.1104)	430,915	17.679	(0.0913)	205,796
Like (A), top 1.0 pct rm	2,839,622	-1.748	(0.7538)	3.308	(0.5300)	2.588	(0.6412)	8.391	(0.0175)	439,975	12.502	(0.0423)	209,688
Like (A), fugitive only	2,868,306	-0.037	(0.9656)	1.079	(0.1917)	0.380	(0.5051)	1.148	(0.3545)	417,203	2.893	(0.0944)	196,239
Cadmium (μg/m^3)													
A) Fug+stack, yr wts	2,868,306	5.23	(0.0001)	7.23	(0.0011)	7.01	(0.0012)	-4.54	(0.5146)	190,529	-26.36	(0.1821)	85,907
B) Fug+stack concep yr	2,868,306	5.30	(0.0000)	6.87	(0.0016)	7.76	(0.0000)	-1.37	(0.8225)	167,489	-20.27	(0.1799)	67,125
C) Fug+stack, birth yr	2,868,306	4.59	(0.0006)	7.73	(0.0000)	5.62	(0.0060)	-12.77	(0.0990)	158,787	-56.77	(0.0275)	62,995
Like (A), top 2.5 pct rm	2,796,598	126.41	(0.6947)	522.37	(0.1238)	-131.84	(0.7776)	-246.90	(0.3799)	169,526	-157.32	(0.7524)	72,572
Like (A), top 1.0 pct rm	2,839,623	78.50	(0.4934)	280.93	(0.0603)	212.28	(0.2351)	171.83	(0.1556)	177,567	337.87	(0.1155)	75,890
Like (A), fugitive only	2,868,306	5.81	(0.0000)	9.33	(0.0000)	5.40	(0.0000)	-14.77	(0.0629)	121,230	-65.19	(0.0267)	45,009

Source: Author's calculations based on 1) birth data from the Texas Center for Health Statistics (1994-2003) merged with 2) km^2-grid-cell-level RSEI chemical concentrations and 3) tract-level data from the 2000 U.S. Census.
Notes: Dependent variable is gestation in weeks. OLS regressions used for models in columns (1) and (2); within -CBSA, -mother and -non-moving mother regressions used in columns (3), (4) and (5), respectively. Standard errors adjusted for heteroskedasticity and clustered at CBSA level in columns (1)-(3) and at mother level in columns (4) and (5). Numbers shown here are estimated coefficients and p-values from the pollution variable only. "Moms w/ var" indicates the number of mothers in the sample who had variation in both chemical exposure and length of gestation.

Table C.5. Birthweight regressions with alternate pollution variables

Pollution Variable	N	Bivar	p-val	Multivar	p-val	CBSA FE	p-val	Mom FE	p-val	moms w/ var	No Move Mom FE	p-val	moms w/ var
Toluene													
A) Fug+stack, yr wts	2,868,306	-3.42	(0.1061)	-0.81	(0.4755)	-1.19	(0.1785)	0.23	(0.6870)	520,686	2.23	(0.2415)	249,451
B) Fug+stack concep yr	2,868,306	-3.26	(0.0996)	-0.78	(0.4784)	-1.14	(0.1851)	0.16	(0.7754)	516,807	3.01	(0.0989)	247,110
C) Fug+stack, birth yr	2,868,306	-3.48	(0.1041)	-0.90	(0.4216)	-1.29	(0.1437)	0.16	(0.7803)	514,126	0.95	(0.5844)	245,686
Like (A), top 2.5 pct rm	2,796,587	-48.29	(0.1225)	33.91	(0.0876)	17.24	(0.2123)	12.98	(0.1463)	501,842	18.40	(0.3322)	242,303
Like (A), top 1.0 pct rm	2,839,622	-37.63	(0.0311)	14.01	(0.3193)	4.64	(0.7087)	8.74	(0.1245)	513,262	16.60	(0.1767)	246,708
Like (A), fugitive only	2,868,306	-3.21	(0.1219)	-0.95	(0.4011)	-1.27	(0.1406)	0.23	(0.6831)	509,198	2.62	(0.1840)	242,310
Epichlorohydrin													
A) Fug+stack, yr wts	2,868,306	3.12	(0.0943)	7.56	(0.0000)	5.44	(0.0000)	1.51	(0.8021)	268,374	2.02	(0.7760)	126,263
B) Fug+stack concep yr	2,868,306	3.81	(0.0250)	8.13	(0.0000)	6.23	(0.0000)	1.68	(0.7453)	228,759	0.13	(0.9840)	107,423
C) Fug+stack, birth yr	2,868,306	3.48	(0.0807)	8.23	(0.0000)	5.95	(0.0000)	2.23	(0.7180)	245,843	4.95	(0.4949)	115,188
Like (A), top 2.5 pct rm	2,796,599	-1,410.89	(0.6652)	2,244.35	(0.0959)	278.91	(0.6402)	-1,744.67	(0.0702)	249,813	-1,948.22	(0.2080)	118,189
Like (A), top 1.0 pct rm	2,839,622	-885.46	(0.6054)	1,829.55	(0.0128)	819.90	(0.0161)	-1,383.56	(0.0137)	260,485	-1,732.18	(0.0630)	122,835
Like (A), fugitive only	2,868,306	1.61	(0.3023)	6.07	(0.0000)	4.16	(0.0000)	0.93	(0.8776)	171,744	1.77	(0.8034)	80,531
Lead													
A) Fug+stack, yr wts	2,868,306	-1,150	(0.0001)	-418	(0.0546)	148	(0.5393)	256	(0.0680)	521,592	810	(0.0001)	249,196
B) Fug+stack concep yr	2,868,306	-1,035	(0.0003)	-387	(0.0465)	185	(0.3867)	218	(0.1167)	510,894	795	(0.0003)	241,959
C) Fug+stack, birth yr	2,868,306	-1,219	(0.0000)	-454	(0.0381)	48	(0.8460)	177	(0.1932)	514,960	596	(0.0021)	244,325
Like (A), top 2.5 pct rm	2,796,591	-13,243	(0.0036)	-6,448	(0.1132)	-545	(0.7889)	474	(0.6779)	504,001	3,693	(0.0717)	241,518
Like (A), top 1.0 pct rm	2,839,622	-8,656	(0.0016)	-3,846	(0.1308)	380	(0.8134)	1,487	(0.0275)	514,737	4,537	(0.0001)	246,067
Like (A), fugitive only	2,868,306	-1,199	(0.0004)	-53	(0.8793)	-89	(0.7357)	158	(0.4136)	487,947	753	(0.0006)	230,113
Cadmium													
A) Fug+stack, yr wts	2,868,306	-2,165	(0.0087)	1,198	(0.2449)	2,646	(0.0040)	-1,880	(0.2045)	222,622	-1,017	(0.6200)	100,589
B) Fug+stack concep yr	2,868,306	-1,914	(0.0013)	696	(0.4970)	2,388	(0.0000)	-1,659	(0.2072)	194,935	-1,691	(0.3606)	77,894
C) Fug+stack, birth yr	2,868,306	-2,317	(0.0140)	1,642	(0.0131)	2,330	(0.0073)	-2,567	(0.1114)	184,835	-626	(0.8436)	73,049
Like (A), top 2.5 pct rm	2,796,598	-402,091	(0.0097)	-74,480	(0.4788)	-12,953	(0.9242)	8,697	(0.8706)	197,611	143,566	(0.1397)	84,661
Like (A), top 1.0 pct rm	2,839,623	-213,520	(0.0001)	-42,656	(0.4607)	41,395	(0.3880)	27,586	(0.2318)	207,039	56,898	(0.1695)	88,509
Like (A), fugitive only	2,868,306	-1,834	(0.0032)	2,276	(0.0000)	1,932	(0.0000)	-3,173	(0.0550)	140,383	-3,045	(0.3695)	51,596

Source: Author's calculations based on 1) birth data from the Texas Center for Health Statistics (1994-2003) merged with 2) km^2-grid-cell-level RSEI chemical concentrations and 3) tract-level data from the 2000 U.S. Census.
Notes: Dependent variable is birthweight in grams. OLS regressions used for models in columns (1) and (2); within -CBSA, -mother and -non-moving mother regressions used in columns (3), (4) and (5), respectively. Standard errors adjusted for heteroskedasticity and clustered at CBSA level in columns (1)-(3) and at mother level in columns (4) and (5). Numbers shown here are estimated coefficients and p-values from the pollution variable only. "Moms w/ var" indicates the number of mothers in the sample who had variation in both chemical exposure and child birthweight.

Table C.6. Preterm birth (PTB) regressions with alternate pollution variables

Pollution Variable	N	Bivar	p-val	Multivar	p-val	CBSA FE	p-val	Mom FE	p-val	moms w/ var	No Move Mom FE	p-val	moms w/ var
Toluene													
A) Fug+stack, yr wts	2,868,306	0.0600	(0.2285)	0.0027	(0.9177)	0.0061	(0.7792)	-0.0262	(0.4982)	83,603	-0.0976	(0.3555)	38,747
B) Fug+stack concep yr	2,868,306	0.0501	(0.2230)	-0.0051	(0.8431)	-0.0024	(0.9198)	-0.0429	(0.2513)	83,004	-0.2548	(0.0212)	38,355
C) Fug+stack, birth yr	2,868,306	0.0630	(0.2459)	0.0064	(0.8171)	0.0102	(0.6504)	-0.0181	(0.6412)	82,625	-0.0062	(0.9499)	38,199
Like (A), top 2.5 pct rm	2,796,587	1.8355	(0.0356)	-0.5109	(0.3780)	-0.3232	(0.4777)	-0.8248	(0.1669)	80,636	-3.6184	(0.0044)	37,604
Like (A), top 1.0 pct rm	2,839,622	1.0282	(0.0312)	-0.3333	(0.4148)	-0.2401	(0.5975)	-0.6108	(0.1055)	82,402	-1.7316	(0.0376)	38,295
Like (A), fugitive only	2,868,306	0.0491	(0.2753)	0.0044	(0.8589)	0.0042	(0.8379)	-0.0236	(0.5502)	82,001	-0.0661	(0.5397)	37,780
Epichlorohydrin													
A) Fug+stack, yr wts	2,868,306	0.18	(0.0002)	0.064	(0.0240)	0.036	(0.0087)	0.064	(0.7861)	43,003	0.17	(0.5010)	19,311
B) Fug+stack concep yr	2,868,306	0.12	(0.0047)	0.017	(0.5095)	-0.0088	(0.4689)	-0.12	(0.6390)	37,179	-0.16	(0.6949)	16,621
C) Fug+stack, birth yr	2,868,306	0.16	(0.0010)	0.038	(0.2168)	0.0076	(0.6008)	0.14	(0.5729)	39,497	0.24	(0.3886)	17,655
Like (A), top 2.5 pct rm	2,796,599	173.80	(0.0195)	54.59	(0.2863)	-1.85	(0.9239)	-24.08	(0.7037)	40,169	-135.31	(0.1889)	18,160
Like (A), top 1.0 pct rm	2,839,622	89.44	(0.0253)	10.50	(0.7078)	-23.74	(0.0249)	12.71	(0.7310)	41,845	-36.49	(0.5468)	18,844
Like (A), fugitive only	2,868,306	0.19	(0.0000)	0.080	(0.0016)	0.054	(0.0001)	0.10	(0.6760)	28,025	0.22	(0.3870)	12,590
Lead													
A) Fug+stack, yr wts	2,868,306	11.6	(0.3799)	-8.2	(0.2745)	-9.0	(0.0746)	-15.2	(0.1548)	83,877	-7.6	(0.6781)	38,740
B) Fug+stack concep yr	2,868,306	8.1	(0.5111)	-9.5	(0.1506)	-9.9	(0.0143)	-17.3	(0.0599)	82,064	-18.7	(0.2219)	37,482
C) Fug+stack, birth yr	2,868,306	14.7	(0.2673)	-5.7	(0.5092)	-6.8	(0.3017)	-13.6	(0.2298)	82,849	-7.0	(0.7192)	37,976
Like (A), top 2.5 pct rm	2,796,591	224.8	(0.1483)	46.2	(0.5782)	26.4	(0.5783)	-24.1	(0.7498)	81,130	67.3	(0.6183)	37,557
Like (A), top 1.0 pct rm	2,839,622	141.8	(0.1464)	10.6	(0.8486)	0.1	(0.9981)	-67.8	(0.1277)	82,850	6.8	(0.9296)	38,283
Like (A), fugitive only	2,868,306	31.9	(0.0001)	1.3	(0.8974)	-4.5	(0.6198)	-6.8	(0.6849)	78,708	-11.2	(0.6617)	35,888
Cadmium													
A) Fug+stack, yr wts	2,868,306	33.6	(0.4577)	-67.7	(0.0054)	-71.0	(0.0000)	45.0	(0.6617)	35,033	239.0	(0.1969)	15,108
B) Fug+stack concep yr	2,868,306	3.8	(0.9469)	-76.5	(0.0015)	-76.6	(0.0000)	22.5	(0.7987)	31,360	225.2	(0.1256)	12,209
C) Fug+stack, birth yr	2,868,306	55.1	(0.1234)	-61.5	(0.0363)	-69.9	(0.0004)	120.7	(0.3023)	29,855	422.4	(0.1508)	11,545
Like (A), top 2.5 pct rm	2,796,598	14,267.6	(0.0193)	3,509.4	(0.3714)	3,010.6	(0.2578)	2,113.8	(0.5583)	31,355	-186.8	(0.9770)	12,882
Like (A), top 1.0 pct rm	2,839,623	4,724.3	(0.2061)	-671.9	(0.7081)	-1,200.9	(0.2593)	-3,925.9	(0.0114)	32,914	-3,644.3	(0.1740)	13,503
Like (A), fugitive only	2,868,306	66.2	(0.0002)	-54.6	(0.0009)	-68.9	(0.0000)	135.5	(0.2597)	23,444	549.3	(0.0964)	8,618

Source: Author's calculations based on 1) birth data from the Texas Center for Health Statistics (1994-2003) merged with 2) km^2-grid-cell-level RSEI chemical concentrations and 3) tract-level data from the 2000 U.S. Census.
Notes: Dependent variable is occurance of preterm birth (< 37 weeks gestation). Linear probability model OLS regressions used for models in columns (1) and (2); within -CBSA, -mother and -non-moving mother regressions used in columns (3), (4) and (5), respectively. Standard errors adjusted for heteroskedasticity and clustered at CBSA level in columns (1)-(3) and at mother level in columns (4) and (5). Numbers shown here are estimated coefficients and p-values from the pollution variable only. All coefficients multipled by 100. "Moms w/ var" indicates the number of mothers in the sample who had variation in both chemical exposure and child PTB.

Table C.7. Very low birthweight (VLBW) regressions with alternate pollution variables

Pollution Variable	N	Bivar	p-val	Multivar	p-val	CBSA FE	p-val	Mom FE	p-val	moms w/ var	No Move Mom FE	p-val	moms w/ var
Toluene													
A) Fug+stack, yr wts	2,868,306	0.025	(0.0044)	0.018	(0.0110)	0.010	(0.0644)	0.0018	(0.9103)	9,314	-0.047	(0.3036)	4,189
B) Fug+stack concep yr	2,868,306	0.023	(0.0029)	0.017	(0.0109)	0.0090	(0.0827)	0.0040	(0.7942)	9,230	-0.039	(0.4024)	4,125
C) Fug+stack, birth yr	2,868,306	0.025	(0.0076)	0.018	(0.0110)	0.010	(0.0404)	-0.0029	(0.8452)	9,184	-0.058	(0.1048)	4,112
Like (A), top 2.5 pct rm	2,796,587	0.367	(0.0023)	0.088	(0.6215)	-0.137	(0.1286)	-0.520	(0.0088)	8,899	-0.446	(0.2892)	4,042
Like (A), top 1.0 pct rm	2,839,622	0.270	(0.0010)	0.111	(0.3403)	-0.023	(0.7717)	-0.239	(0.0668)	9,127	-0.548	(0.0415)	4,119
Like (A), fugitive only	2,868,306	0.025	(0.0052)	0.020	(0.0029)	0.012	(0.0041)	0.002	(0.8778)	9,148	-0.042	(0.3694)	4,101
Epichlorohydrin													
A) Fug+stack, yr wts	2,868,306	-0.077	(0.0000)	-0.078	(0.0000)	-0.095	(0.0000)	-0.057	(0.1730)	4,916	-0.105	(0.1132)	2,161
B) Fug+stack concep yr	2,868,306	-0.068	(0.0000)	-0.069	(0.0000)	-0.084	(0.0000)	-0.049	(0.1968)	4,288	-0.116	(0.0828)	1,852
C) Fug+stack, birth yr	2,868,306	-0.082	(0.0000)	-0.083	(0.0000)	-0.101	(0.0000)	-0.051	(0.2341)	4,571	-0.083	(0.1635)	2,004
Like (A), top 2.5 pct rm	2,796,599	40.419	(0.0000)	22.285	(0.1552)	-9.166	(0.0081)	6.094	(0.7830)	4,616	13.769	(0.6925)	2,034
Like (A), top 1.0 pct rm	2,839,622	19.216	(0.0000)	8.814	(0.2901)	-7.996	(0.0000)	-18.610	(0.1344)	4,802	-1.722	(0.9331)	2,114
Like (A), fugitive only	2,868,306	-0.075	(0.0000)	-0.076	(0.0000)	-0.092	(0.0000)	-0.044	(0.2657)	3,251	-0.095	(0.1326)	1,426
Lead													
A) Fug+stack, yr wts	2,868,306	0.96	(0.7490)	-0.88	(0.7166)	-1.89	(0.1733)	-3.02	(0.2920)	9,373	-3.22	(0.4491)	4,186
B) Fug+stack concep yr	2,868,306	0.29	(0.9190)	-1.28	(0.5653)	-2.02	(0.1004)	-3.50	(0.2090)	9,163	-4.46	(0.3179)	4,033
C) Fug+stack, birth yr	2,868,306	1.66	(0.5830)	-0.28	(0.9129)	-1.51	(0.3300)	-2.49	(0.3641)	9,236	-2.80	(0.4532)	4,079
Like (A), top 2.5 pct rm	2,796,591	49.04	(0.1612)	33.13	(0.2440)	11.27	(0.4998)	31.40	(0.2209)	9,067	-13.31	(0.7728)	4,048
Like (A), top 1.0 pct rm	2,839,622	22.86	(0.2733)	11.03	(0.5423)	-1.38	(0.8883)	-9.18	(0.5364)	9,261	-35.70	(0.1687)	4,136
Like (A), fugitive only	2,868,306	4.81	(0.1496)	1.43	(0.6788)	-2.24	(0.3176)	0.31	(0.9482)	8,790	-4.50	(0.3404)	3,883
Cadmium													
A) Fug+stack, yr wts	2,868,306	-1.30	(0.8981)	-13.25	(0.0574)	-14.19	(0.0000)	9.00	(0.6332)	3,840	40.96	(0.2400)	1,609
B) Fug+stack concep yr	2,868,306	-4.27	(0.7138)	-12.98	(0.0457)	-12.26	(0.0000)	10.41	(0.5730)	3,425	56.69	(0.1427)	1,278
C) Fug+stack, birth yr	2,868,306	4.46	(0.4499)	-9.81	(0.0900)	-13.46	(0.0000)	20.54	(0.3591)	3,306	23.44	(0.6685)	1,240
Like (A), top 2.5 pct rm	2,796,598	1,745.27	(0.0661)	384.71	(0.6681)	-462.98	(0.2539)	-249.64	(0.8309)	3,401	-4,599.57	(0.0254)	1,331
Like (A), top 1.0 pct rm	2,839,623	794.38	(0.2314)	200.34	(0.6111)	79.53	(0.5710)	69.00	(0.8921)	3,572	-630.78	(0.4759)	1,405
Like (A), fugitive only	2,868,306	8.12	(0.0000)	-7.56	(0.1216)	-14.13	(0.0000)	27.77	(0.2550)	2,625	66.45	(0.2672)	929

Source: Author's calculations based on 1) birth data from the Texas Center for Health Statistics (1994-2003) merged with 2) km^2-grid-cell-level RSEI chemical concentrations and 3) tract-level data from the 2000 U.S. Census.
Notes: Dependent variable is occurance of very low birthweight (< 1500 grams). Linear probability model OLS regressions used for models in columns (1) and (2); within -CBSA, -mother and -non-moving mother regressions used in columns (3), (4) and (5), respectively. Standard errors adjusted for heteroskedasticity and clustered at CBSA level in columns (1)-(3) and at mother level in columns (4) and (5). Numbers shown here are estimated coefficients and p-values from the pollution variable only. All coefficients multipled by 100. "Moms w/ var" indicates the number of mothers in the sample who had variation in both chemical exposure and child VLBW.

Table C.8. Infant death regressions with alternate pollution variables

Pollution Variable	N	Bivar	p-val	Multivar	p-val	CBSA FE	p-val	Mom FE	p-val	moms w/ var	No Move Mom FE	p-val	moms w/ var
Toluene													
A) Fug+stack, yr wts	2,868,306	0.0145	(0.0000)	0.0108	(0.0009)	0.0062	(0.0344)	0.0127	(0.0932)	5,787	-0.0044	(0.8772)	2,900
B) Fug+stack concep yr	2,868,306	0.0140	(0.0000)	0.0105	(0.0004)	0.0061	(0.0201)	0.0153	(0.0501)	5,733	0.0095	(0.7727)	2,857
C) Fug+stack, birth yr	2,868,306	0.0147	(0.0000)	0.0110	(0.0002)	0.0063	(0.0148)	0.0102	(0.1728)	5,712	-0.0239	(0.2178)	2,857
Like (A), top 2.5 pct rm	2,796,587	0.1191	(0.0722)	0.0010	(0.9858)	-0.0473	(0.2499)	-0.3517	(0.0203)	5,579	-0.4554	(0.1727)	2,810
Like (A), top 1.0 pct rm	2,839,622	0.0764	(0.0236)	0.0019	(0.9644)	-0.0421	(0.0538)	-0.2437	(0.0123)	5,685	-0.4362	(0.0641)	2,852
Like (A), fugitive only	2,868,306	0.0153	(0.0000)	0.0121	(0.0000)	0.0072	(0.0009)	0.0142	(0.0584)	5,678	0.0083	(0.7877)	2,842
Epichlorohydrin													
A) Fug+stack, yr wts	2,868,306	0.078	(0.0000)	0.072	(0.0000)	0.072	(0.0000)	-0.128	(0.0939)	3,018	-0.120	(0.0839)	1,492
B) Fug+stack concep yr	2,868,306	0.065	(0.0000)	0.059	(0.0000)	0.059	(0.0000)	-0.092	(0.1610)	2,625	-0.067	(0.1861)	1,293
C) Fug+stack, birth yr	2,868,306	0.079	(0.0000)	0.072	(0.0000)	0.072	(0.0000)	-0.178	(0.0893)	2,761	-0.213	(0.1719)	1,360
Like (A), top 2.5 pct rm	2,796,599	1.806	(0.6361)	-2.555	(0.5684)	-3.451	(0.1159)	-1.736	(0.9187)	2,826	27.626	(0.3362)	1,412
Like (A), top 1.0 pct rm	2,839,622	4.951	(0.0194)	1.712	(0.4733)	2.209	(0.0966)	5.608	(0.5744)	2,944	14.773	(0.4106)	1,462
Like (A), fugitive only	2,868,306	0.085	(0.0000)	0.079	(0.0000)	0.078	(0.0000)	-0.133	(0.0894)	1,886	-0.108	(0.0959)	932
Lead													
A) Fug+stack, yr wts	2,868,306	1.46	(0.4600)	0.46	(0.7460)	0.33	(0.7369)	0.68	(0.8048)	5,798	-4.65	(0.2412)	2,889
B) Fug+stack concep yr	2,868,306	1.26	(0.4943)	0.39	(0.7708)	0.36	(0.6883)	0.67	(0.8026)	5,658	-5.24	(0.1773)	2,781
C) Fug+stack, birth yr	2,868,306	1.63	(0.3999)	0.60	(0.6814)	0.34	(0.7441)	0.22	(0.9348)	5,714	-4.75	(0.1937)	2,831
Like (A), top 2.5 pct rm	2,796,591	27.13	(0.1057)	17.23	(0.2879)	11.39	(0.4037)	20.37	(0.3057)	5,631	38.07	(0.3040)	2,815
Like (A), top 1.0 pct rm	2,839,622	13.95	(0.2027)	7.06	(0.4646)	4.07	(0.5260)	4.66	(0.6910)	5,738	-1.01	(0.9619)	2,864
Like (A), fugitive only	2,868,306	3.41	(0.1090)	1.81	(0.3530)	1.02	(0.5895)	1.33	(0.7774)	5,438	-5.20	(0.2395)	2,686
Cadmium													
A) Fug+stack, yr wts	2,868,306	4.38	(0.6448)	-0.63	(0.9349)	1.40	(0.7986)	5.94	(0.7173)	2,323	-75.42	(0.1367)	1,109
B) Fug+stack concep yr	2,868,306	2.65	(0.8062)	-1.32	(0.8744)	0.97	(0.8848)	9.99	(0.5251)	2,081	-37.68	(0.3527)	910
C) Fug+stack, birth yr	2,868,306	8.28	(0.1350)	2.50	(0.5916)	3.60	(0.1905)	16.04	(0.4243)	1,970	-144.92	(0.0890)	848
Like (A), top 2.5 pct rm	2,796,598	-218.53	(0.3362)	-739.97	(0.0086)	-747.36	(0.1029)	-40.06	(0.9671)	2,102	991.22	(0.5632)	973
Like (A), top 1.0 pct rm	2,839,623	-28.32	(0.8887)	-309.60	(0.0038)	-207.81	(0.0154)	-53.29	(0.8900)	2,193	149.83	(0.8218)	1,007
Like (A), fugitive only	2,868,306	10.47	(0.0000)	4.32	(0.0473)	4.00	(0.0040)	7.43	(0.7029)	1,541	-151.80	(0.1431)	636

Source: Author's calculations based on 1) birth data from the Texas Center for Health Statistics (1994-2003) merged with 2) km^2-grid-cell-level RSEI chemical concentrations and 3) tract-level data from the 2000 U.S. Census.
Notes: Dependent variable is occurance of infant death (death before 1 year of age). Linear probability model OLS regressions used for models in columns (1) and (2); within -CBSA, -mother and -non-moving mother regressions used in columns (3), (4) and (5), respectively. Standard errors adjusted for heteroskedasticity and clustered at CBSA level in columns (1)-(3) and at mother level in columns (4) and (5). Numbers shown here are estimated coefficients and p-values from the pollution variable only. All coefficients multipled by 100. "Moms w/ var" indicates the number of mothers in the sample who had variation in both chemical exposure and infant death.

Table C.9. Gestation, Birthweight and PTB regressions with polynomial pollution variables, non-moving mother fixed effects

	Gest (wks)			Bwt (g)			PTB		
	conc	$conc^2$	$conc^3$	*conc*	$conc^2$	$conc^3$	*conc*	$conc^2$	$conc^3$
Toluene									
moms w/ var	*212,383*	.	.	*249,397*	.	.	*38,691*	.	.
Linear	-0.0011	.	.	2.23	.	.	-0.098		
p-val	(0.9156)	.	.	(0.2415)	.	.	(0.3555)	.	.
Quadratic	0.0022	-0.000043	.	1.41	0.011	.	-0.34	0.0031	
p-val	(0.9063)	(0.8028)	.	(0.6812)	(0.7624)	.	(0.1095)	(0.1156)	.
Cubic	0.0064	-0.00021	0.0000012	0.072	0.066	-0.00039	-0.38	0.0048	-0.000011
p-val	(0.7856)	(0.7164)	(0.7524)	(0.9869)	(0.5487)	(0.5924)	(0.1789)	(0.5134)	(0.8191)
Epichlorohydrin									
moms w/ var	*107,813*	.	.	*126,323*	.	.	*19,295*	.	.
Linear	0.105	.	.	2.02	.	.	0.17		
p-val	(0.0423)	.	.	(0.7760)	.	.	(0.5010)	.	.
Quadratic	0.112	-0.00038	.	7.16	-0.26	.	0.61	-0.022	
p-val	(0.4623)	(0.9591)	.	(0.8212)	(0.8486)	.	(0.7215)	(0.7654)	.
Cubic	0.035	0.017	-0.00065	30.7	-5.6	0.20	1.41	-0.201	0.0067
p-val	(0.8811)	(0.6158)	(0.5860)	(0.5214)	(0.3937)	(0.3595)	(0.6074)	(0.5571)	(0.5389)
Lead									
moms w/ var	*212,255*	.	.	*249,148*	.	.	*38,710*	.	.
Linear	3.02	.	.	810.2	.	.	-7.6		
p-val	(0.0219)	.	.	(0.0001)	.	.	(0.6781)	.	.
Quadratic	3.02	0.00026	.	1,301	-1,697	.	-10.5	10.1	
p-val	(0.1625)	(1.0000)	.	(0.0008)	(0.0302)	.	(0.6852)	(0.8749)	.
Cubic	5.71	-27.0	39.1	1,654	-5,243	5,138	-22.0	125.5	-167.1
p-val	(0.0332)	(0.0858)	(0.0666)	(0.0013)	(0.0374)	(0.0718)	(0.5018)	(0.5519)	(0.5179)
Cadmium									
moms w/ var	*85,958*	.	.	*100,630*	.	.	*15,121*	.	.
Linear	-26.4	.	.	-1,017	.	.	239.0		
p-val	(0.1821)	.	.	(0.6200)	.	.	(0.1969)	.	.
Quadratic	-55.8	479.1	.	-2,638	26,370	.	448.6	-3,408	
p-val	(0.0237)	(0.0089)	.	(0.4517)	(0.3211)	.	(0.1354)	(0.1289)	.
Cubic	-23.9	-1,290	12,202	-1,986	-9,785	249,351	545.1	-8,759	36,905
p-val	(0.5273)	(0.4598)	(0.3022)	(0.7471)	(0.9700)	(0.8846)	(0.2545)	(0.6590)	(0.7796)

Source: Author's calculations based on 1) birth data from the Texas Center for Health Statistics (1994-2003) merged with 2) km^2-grid-cell-level RSEI chemical concentrations and 3) tract-level data from the 2000 U.S. Census.
Notes: Dependent variable listed in column headings. Regressions use only variation within mothers who did not move between births (non-moving mother fixed effects). Standard errors adjusted for heteroskedasticity and clustered at mother level. "Linear", "quadratic" and "cubic" refer to models in which the relationship between birth outcome and pollution is allowed to have the shape indicated. "Moms w/ var" indicates the number of mothers in the sample who had between-birth variation in both chemical exposure and infant health outcome. Coefficients in PTB regressions multiplied by 100.

Table C.10. LBW, VLBW and infant death regressions with polynomial pollution variables, non-moving mother fixed effects

	LBW			VLBW			Infant Death		
	$conc$	$conc^2$	$conc^3$	$conc$	$conc^2$	$conc^3$	$conc$	$conc^2$	$conc^3$
Toluene									
moms w/ var	*22,710*	.	.	*4,174*	.	.	*2,886*	.	.
Linear	-0.090			-0.047			-0.0044	.	.
p-val	(0.2844)	.	.	(0.3036)	.	.	(0.8772)	.	.
Quadratic	-0.38	0.0037		-0.11	0.00080		-0.041	0.00047	.
p-val	(0.0369)	(0.0416)	.	(0.1985)	(0.2434)	.	(0.4575)	(0.2140)	.
Cubic	-0.42	0.0054	-0.000012	-0.18	0.0039	-0.000022	-0.12	0.0038	-0.000023
p-val	(0.0839)	(0.3716)	(0.7796)	(0.0804)	(0.1059)	(0.1165)	(0.0417)	(0.0078)	(0.0135)
Epichlorohydrin									
moms w/ var	*11,468*	.	.	*2,145*	.	.	*1,488*	.	.
Linear	0.19			-0.10			-0.12	.	.
p-val	(0.7243)	.	.	(0.1132)	.	.	(0.0839)	.	.
Quadratic	-0.35	0.027		-0.78	0.034		-0.42	0.015	.
p-val	(0.8991)	(0.8108)	.	(0.0633)	(0.0654)	.	(0.3236)	(0.4424)	.
Cubic	-4.21	0.89	-0.033	-1.53	0.20	-0.0063	-1.71	0.30	-0.011
p-val	(0.1857)	(0.0748)	(0.0788)	(0.0514)	(0.0507)	(0.0523)	(0.0651)	(0.1113)	(0.1391)
Lead									
moms w/ var	*22,723*	.	.	*4,182*	.	.	*2,887*	.	.
Linear	-24.3			-3.22			-4.6	.	.
p-val	(0.0459)	.	.	(0.4491)	.	.	(0.2412)	.	.
Quadratic	-24.0	-0.94		-5.36	7.38		-8.4	13.0	.
p-val	(0.2285)	(0.9839)	.	(0.5446)	(0.6492)	.	(0.2909)	(0.3648)	.
Cubic	-25.9	18.0	-27.4	-5.22	6.04	1.94	-8.7	16.1	-4.5
p-val	(0.3130)	(0.8971)	(0.8702)	(0.6597)	(0.9071)	(0.9721)	(0.3909)	(0.7013)	(0.9205)
Cadmium									
moms w/ var	*8,734*	.	.	*1,611*	.	.	*1,106*	.	.
Linear	17.0			41.0			-75.4	.	.
p-val	(0.8234)	.	.	(0.2400)	.	.	(0.1367)	.	.
Quadratic	68.6	-839.7		96.2	-898.7		-148.2	1,184	.
p-val	(0.6601)	(0.5293)	.	(0.1514)	(0.1375)	.	(0.0987)	(0.1106)	.
Cubic	220.2	-9,245	57,972	240.6	-8,904	55,213	-268.8	7,870	-46,112
p-val	(0.5123)	(0.4273)	(0.4189)	(0.1176)	(0.1056)	(0.1055)	(0.0824)	(0.0862)	(0.0888)

Source: Author's calculations based on 1) birth data from the Texas Center for Health Statistics (1994-2003) merged with 2) km^2-grid-cell-level RSEI chemical concentrations and 3) tract-level data from the 2000 U.S. Census.
Notes: Dependent variable listed in column headings. Linear probability model regressions using only variation within mothers who did not move between births (non-moving mother fixed effects). Standard errors adjusted for heteroskedasticity and clustered at mother level. "Linear", "quadratic" and "cubic" refer to models in which the relationship between birth outcome and pollution is allowed to have the shape indicated. "Moms w/ var" indicates the number of mothers in the sample who had between-birth variation in both chemical exposure and infant health outcome. Coefficients multiplied by 100.

Table C.11. Gestation, birthweight and PTB regressions with polynomial pollution variables, CBSA fixed effects, top 1% of obs by exposure removed

	Gest (wks)			Bwt (g)			PTB		
	conc	$conc^2$	$conc^3$	*conc*	$conc^2$	$conc^3$	*conc*	$conc^2$	$conc^3$
Toluene									
Linear	0.0316	.	.	4.64	.	.	-0.240		
p-val	(0.3365)	.	.	(0.7087)	.	.	(0.5975)	.	.
Quadratic	0.109	-0.103	.	29.02	-32.12	.	-0.611	0.489	
p-val	(0.2643)	(0.3978)	.	(0.0993)	(0.0851)	.	(0.3331)	(0.4631)	.
Cubic	0.124	-0.15	0.0000012	3.04	59.33	-66.48	0.259	-2.57	2.23
p-val	(0.6203)	(0.8178)	(0.9283)	(0.8764)	(0.3230)	(0.0881)	(0.8281)	(0.4271)	(0.2912)
Epichlorohydrin									
Linear	-1.42	.	.	819.9	.	.	-23.74		
p-val	(0.2209)	.	.	(0.0161)	.	.	(0.0249)	.	.
Quadratic	-10.60	1,046	.	-941.9	200,832	.	25.39	-5,600	
p-val	(0.0019)	(0.0002)	.	(0.2902)	(0.0032)	.	(0.3946)	(0.0196)	.
Cubic	-20.66	4,016	-193,839	-6,301	1.782E+06	-1.032E+08	190.21	-54,243	3.175E+06
p-val	(0.0019)	(0.0028)	(0.0072)	(0.0000)	(0.0000)	(0.0000)	(0.0007)	(0.0000)	(0.0000)
Lead									
Linear	2.59	.	.	380.2	.	.	0.0859		
p-val	(0.6412)	.	.	(0.8134)	.	.	(0.9981)	.	.
Quadratic	-3.50	889	.	-895.0	186,026	.	32.15	-4,678	
p-val	(0.7072)	(0.2403)	.	(0.6949)	(0.3014)	.	(0.4755)	(0.4463)	.
Cubic	-11.36	3,918	-246,870	-6,455.4	2.329E+06	-1.746E+08	88.55	-26,418	1.771E+06
p-val	(0.4210)	(0.2522)	(0.2986)	(0.0502)	(0.0520)	(0.0544)	(0.3671)	(0.3537)	(0.4284)
Cadmium									
Linear	212.3	.	.	41,395	.	.	-1,201		
p-val	(0.2351)	.	.	(0.3880)	.	.	(0.2593)	.	.
Quadratic	-83.86	1.120E+06	.	-19,951	2.320E+08	.	3,004	-1.590E+07	
p-val	(0.8893)	(0.5018)	.	(0.8901)	(0.5386)	.	(0.3844)	(0.1100)	.
Cubic	-925.2	9.700E+06	-1.816E+10	-244,937	2.526E+09	-4.856E+12	9,794	-8.515E+07	1.465E+11
p-val	(0.2530)	(0.0210)	(0.0019)	(0.3447)	(0.1246)	(0.0808)	(0.0032)	(0.0000)	(0.0000)

Source: Author's calculations based on 1) birth data from the Texas Center for Health Statistics (1994-2003) merged with 2) km^2-grid-cell-level RSEI chemical concentrations and 3) tract-level data from the 2000 U.S. Census.
Notes: Dependent variable listed in column headings. Sample excludes top 1% of all infants by exposure. Regressions include CBSA fixed effects. Standard errors adjusted for heteroskedasticity and clustered at CBSA level. "Linear", "quadratic" and "cubic" refer to models in which the relationship between birth outcome and pollution is allowed to have the shape indicated. Coefficients in PTB regressions multiplied by 100.

Table C.12. LBW, VLBW and infant death regressions with polynomial pollution variables, CBSA fixed effects, top 1% of obs by exposure removed

	LBW			VLBW			Infant Death		
	conc	$conc^2$	$conc^3$	*conc*	$conc^2$	$conc^3$	*conc*	$conc^2$	$conc^3$
Toluene									
Linear	-0.215			-0.0226			-0.0421	.	.
p-val	(0.4833)	.	.	(0.7717)	.	.	(0.0538)	.	.
Quadratic	-1.10	1.17		-0.197	0.230		-0.106	0.0845	.
p-val	(0.0685)	(0.0369)	.	(0.1778)	(0.0722)	.	(0.1478)	(0.3903)	.
Cubic	-1.08	1.08	0.061161	-0.413	0.989	-0.552	-0.136	0.189	-0.0762
p-val	(0.3282)	(0.7056)	(0.9725)	(0.1656)	(0.1641)	(0.2191)	(0.4908)	(0.7498)	(0.8451)
Epichlorohydrin									
Linear	-33.73			-8.00			2.21	.	.
p-val	(0.0018)	.	.	(0.0000)	.	.	(0.0966)	.	.
Quadratic	-22.18	-1,316		-0.112	-898.7		-3.43	642.6	.
p-val	(0.4000)	(0.4738)	.	(0.9835)	(0.0794)	.	(0.3135)	(0.0088)	.
Cubic	81.72	-31,979	2.001E+06	0.639	-1,120	14,466	-5.77	1,335	-45,193
p-val	(0.0167)	(0.0000)	(0.0000)	(0.9598)	(0.6857)	(0.9232)	(0.2859)	(0.1897)	(0.4142)
Lead									
Linear	-10.58			-1.38			4.07	.	.
p-val	(0.7955)	.	.	(0.8883)	.	.	(0.5260)	.	.
Quadratic	-78.10	9,850		3.73	-745.0		8.63	-664.5	.
p-val	(0.2418)	(0.0487)	.	(0.8678)	(0.7165)	.	(0.6441)	(0.7401)	.
Cubic	-64.41	4,572	4.300E+05	1.42	146.1	-72,603	4.93	762.1	-116,241
p-val	(0.4793)	(0.8890)	(0.8688)	(0.9734)	(0.9911)	(0.9424)	(0.8665)	(0.9100)	(0.8017)
Cadmium									
Linear	-1,359			79.53			-207.8	.	.
p-val	(0.0864)	.	.	(0.5710)	.	.	(0.0154)	.	.
Quadratic	-1,477	446,951		-555.5	2.402E+06		-628.0	1.589E+06	.
p-val	(0.4715)	(0.9255)	.	(0.1453)	(0.0216)	.	(0.2242)	(0.3554)	.
Cubic	586.6	-2.060E+07	4.455E+10	391.5	-7.256E+06	2.044E+10	-1,425.5	9.722E+06	-1.721E+10
p-val	(0.8991)	(0.5212)	(0.4513)	(0.3695)	(0.0018)	(0.0001)	(0.0884)	(0.0680)	(0.0322)

Source: Author's calculations based on 1) birth data from the Texas Center for Health Statistics (1994-2003) merged with 2) km^2-grid-cell-level RSEI chemical concentrations and 3) tract-level data from the 2000 U.S. Census.
Notes: Dependent variable listed in column headings. Sample excludes top 1% of all infants by exposure. Regressions include CBSA fixed effects. Standard errors adjusted for heteroskedasticity and clustered at CBSA level. "Linear", "quadratic" and "cubic" refer to models in which the relationship between birth outcome and pollution is allowed to have the shape indicated. Coefficients multiplied by 100.

Table C.13: Summary Statistics for RSEI Chemicals Affecting Texas Infants, 1994-2003

Chemical Name	Dev/ Rep Tox?	RSEI Tox Weight	Number Exposures	Min	Median	Mean (Units: $\mu g/m^3$)	Max	SD	Rank Correlation b/w RSEI 2002 and NATA 2002	
									All Source	Major Source
Acetaldehyde		200	1,534,340	0	5.885E-08	5.699E-03	39.9158	1.021E-01	-0.307	0.752
Acetamide		14	672,908	0	0	5.097E-07	0.0113	2.935E-05	0.310	0.329
Acetonitrile		30	1,219,999	0	0	2.369E-03	7.0255	1.621E-02	0.874	0.910
Acrolein		90000	768,291	0	0	1.569E-04	1.0489	3.212E-03	0.248	0.788
Acrylamide		9300	875,940	0	0	8.030E-06	0.0293	1.606E-04	0.893	0.893
Acrylic acid		1800	1,467,994	0	7.283E-08	1.917E-03	1.1353	1.004E-02	0.897	0.896
Acrylonitrile		900	1,534,394	0	1.439E-06	1.842E-03	2.5534	1.819E-02	0.785	0.858
Allyl alcohol		100	1,014,119	0	0	2.105E-04	1.2889	1.757E-03		
Allyl chloride		1800	682,998	0	0	1.900E-04	4.2750	8.585E-03	0.937	0.946
Aluminum (fume or dust)		360	1,365,654	0	0	1.090E-03	1.4043	1.771E-02		
Aniline		1800	927,811	0	0	5.053E-04	2.4681	7.020E-03	0.915	0.915
o-Anisidine		9000	559,664	0	0	5.286E-06	0.0027	4.001E-05		
Anthracene		1.7	1,528,871	0	9.225E-09	2.730E-05	0.3877	8.108E-04		
Antimony and antimony compounds		9000	1,909,699	0	2.055E-05	3.101E-04	1.8160	6.357E-03	0.922	0.922
Arsenic and arsenic compounds	x	60000	1,673,080	0	2.027E-07	9.820E-05	0.6818	1.502E-03	0.233	0.091
Asbestos (friable)		1000000	880,284	0	0	4.631E-06	0.0129	1.267E-04		
Barium and barium compounds		2.5	1,793,352	0	5.740E-06	2.457E-04	25.8346	2.039E-02		
Benzene	x	60	2,136,321	0	1.892E-04	3.638E-02	20.8744	2.306E-01	0.298	0.841
Benzidine		480000	7,612	0	0	2.610E-07	0.0066	1.507E-05		
Benzotrichloride		26000	204,209	0	0	4.865E-06	0.1344	3.092E-04		
Benzyl chloride		350	1,026,360	0	0	1.976E-05	0.1695	6.348E-04	0.074	-0.037
Beryllium and beryllium compounds		17000	77,821	0	0	2.808E-09	0.0001	1.449E-07		
Biphenyl		10	1,399,860	0	0	3.330E-04	1.5509	6.610E-03	0.756	0.827
Bis(2-chloroethyl)ether		2400	720,688	0	0	3.955E-05	0.2322	9.306E-04	0.925	0.921
Bromomethane (Methyl bromide)	x	360	926,922	0	0	2.064E-04	2.4889	1.010E-02	0.476	0.489
1,3-Butadiene	x	900	1,710,131	0	6.334E-08	1.867E-02	35.3387	2.100E-01	0.313	0.776
tert-Butyl alcohol		5	1,400,013	0	0	4.486E-03	5.6556	2.795E-02		
Butyl acrylate		1800	1,555,510	0	9.446E-06	1.024E-03	1.0506	4.518E-03		
n-Butyl alcohol		5	2,284,478	0	2.318E-03	2.212E-02	34.7774	1.832E-01		
sec-Butyl alcohol		5	1,512,225	0	6.828E-08	3.311E-04	3.5068	8.668E-03		
1,2-Butylene oxide		90	871,689	0	0	7.013E-05	1.1818	2.512E-03	0.875	0.875
CFC-11 (trichlorofluoromethane)		2.6	924,872	0	0	6.461E-04	11.7103	2.785E-02		
CFC-12 (dichlorodifluoromethane)		2.5	1,027,430	0	0	4.267E-03	38.7278	1.091E-01		
Cadmium and cadmium compounds	x	90000	1,015,413	0	0	3.187E-05	0.1284	5.250E-04	0.745	0.759
Captan		3.8	176,992	0	0	8.049E-08	0.0015	5.483E-06		
Carbaryl	x	50	293,014	0	0	8.656E-06	0.0465	1.567E-04		
Carbon disulfide	x	2.6	1,014,373	0	0	3.796E-03	37.8805	7.655E-02	0.572	0.871
Carbon tetrachloride		110	857,270	0	0	3.422E-03	6.3731	5.284E-02	0.435	0.898
Carbonyl sulfide		150	952,253	0	0	1.541E-03	7.7560	1.944E-02	0.823	0.858
Catechol		18	994,268	0	0	3.518E-05	0.1180	5.855E-04	0.469	0.469
Chlorine		9000	2,164,067	0	6.019E-05	8.523E-03	20.7178	9.283E-02	0.866	0.866
Chlorine dioxide		9000	545,274	0	0	4.360E-04	1.2993	1.147E-02		
Chloroacetic acid		250	752,736	0	0	2.423E-06	0.0232	6.283E-05	0.777	0.777
Chlorobenzene		1.8	1,351,493	0	0	2.486E-03	2.3791	1.435E-02	0.640	0.852
Chloroethane (Ethyl chloride)		0.18	1,346,635	0	0	5.929E-04	3.1463	9.065E-03	0.324	0.881
Chloroform	x	160	1,601,799	0	3.182E-07	1.451E-02	28.0339	3.448E-01	0.068	0.568
Chloromethane	x	20	1,250,345	0	0	8.699E-03	77.4992	1.381E-01	0.759	0.874
Chlorophenols		330	292,043	0	0	3.595E-06	0.0171	5.685E-05		
Chloroprene		1800	661,971	0	0	1.595E-03	4.8201	2.267E-02	0.568	0.835
Chlorothalonil		25	1,039,217	0	0	2.835E-04	1.2288	6.857E-03		
Chromium and chromium compounds	x	86000	2,637,600	0	3.055E-05	1.937E-04	0.1717	1.618E-03	0.704	0.738
Cobalt and cobalt compounds		34000	1,480,288	0	8.090E-08	6.813E-05	0.1272	8.617E-04	0.716	0.716
Copper and copper compounds		750	2,578,732	0	2.360E-04	4.406E-03	6.8909	4.069E-02		
o-Cresol		3	1,001,355	0	0	7.112E-05	0.2104	7.571E-04		
p-Cresol		3	752,121	0	0	1.045E-04	2.8703	2.691E-03		

Continued on Next Page...

Table C.13 – Continued

Chemical Name	Dev/ Rep Tox?	RSEI Tox Weight	Number Exposures	Min	Median	Mean (Units: $\mu g/m^3$)	Max	SD	Rank Correlation b/w RSEI 2002 and NATA 2002 All Source	Rank Correlation b/w RSEI 2002 and NATA 2002 Major Source
m-Cresol		3	783,768	0	0	1.214E-04	1.0243	1.742E-03		
Cresol (mixed isomers)		3	1,699,094	0	2.280E-07	1.110E-03	3.6914	1.677E-02	0.102	0.650
Cumene		4.5	1,787,622	0	8.340E-06	4.929E-03	8.9789	3.849E-02	0.576	0.780
Cumene hydroperoxide		330	726,363	0	0	1.047E-04	0.5721	2.523E-03		
Cupferron		450	102,798	0	0	2.271E-09	0.0001	1.529E-07		
Cyanide compounds		600	1,261,040	0	0	7.038E-04	25.5780	3.184E-02	0.187	0.394
Cyclohexane		0.3	1,959,708	0	5.937E-06	1.841E-02	31.9437	1.415E-01		
2,4-D ((2,4-dichlorophenoxy)acetic acid)		50	567,709	0	0	1.904E-06	0.0088	3.381E-05	0.175	0.414
Decabromodiphenyl ether		50	1,061,431	0	0	1.725E-05	0.1243	4.005E-04		
Di(2-ethylhexyl) phthalate	x	26	1,697,002	0	7.912E-07	3.961E-05	0.4023	1.086E-03	-0.148	0.098
4,4'-Diaminodiphenylether		280	95,671	0	0	3.333E-09	0.0000	8.246E-08		
2,4-Diaminotoluene		7900	704,346	0	0	3.308E-05	0.1250	1.051E-03	0.794	0.794
Diaminotoluene (mixed isomers)		46000	747,993	0	0	6.660E-05	0.6415	2.449E-03		
1,2-Dibromoethane	x	4300	761,825	0	0	4.298E-05	0.0287	2.399E-04	0.855	0.544
Dibutyl phthalate	x	5	903,857	0	0	1.385E-05	0.0356	1.771E-04	-0.145	0.593
1,2-Dichlorobenzene		5.6	1,337,392	0	0	4.064E-04	1.8016	7.765E-03		
1,4-Dichlorobenzene		7.9	738,143	0	0	3.346E-05	0.1051	7.264E-04	0.323	0.889
1,3-Dichlorobenzene		9	397,547	0	0	6.788E-08	0.0004	1.568E-06		
Dichlorobenzene (mixed isomers)		9	746,184	0	0	2.888E-06	0.0382	8.185E-05		
Dichlorobromomethane		120	290	0	0	1.175E-09	0.0000	1.338E-07		
1,2-Dichloroethane		190	1,214,209	0	0	2.963E-03	27.7677	7.007E-02	0.856	0.868
1,2-Dichloroethylene		56	850,247	0	0	4.922E-05	0.8891	2.023E-03		
Dichloromethane		3.4	1,993,360	0	1.448E-03	1.616E-02	58.4616	2.618E-01	0.163	0.426
2,4-Dichlorophenol		170	177,868	0	0	1.058E-05	0.1041	2.261E-04		
1,2-Dichloropropane		450	674,397	0	0	3.483E-04	14.3539	2.765E-02	-0.029	0.481
1,3-Dichloropropylene		90	546,787	0	0	9.907E-05	3.3749	7.543E-03	0.096	0.610
Dichlorvos		3600	217,267	0	0	1.457E-06	0.0138	5.901E-05		
Dicofol		420	532,795	0	0	1.913E-06	0.0078	3.619E-05		
Diethanolamine		600	1,713,143	0	1.333E-06	1.607E-03	13.6567	3.251E-02	0.689	0.689
Diethyl sulfate		2400	864,334	0	0	1.196E-04	0.8432	4.130E-03	0.844	0.844
1,1-Dimethyl Hydrazine		7100	594,194	0	0	3.095E-08	0.0002	8.615E-07	0.264	0.264
2,4-Dimethylphenol		25	827,800	0	0	4.227E-05	0.2205	6.363E-04		
4,6-Dinitro-o-cresol		3800	700,982	0	0	3.213E-06	0.0041	3.699E-05		
2,4-Dinitrophenol		250	309,175	0	0	4.462E-08	0.0004	1.294E-06		
2,4-Dinitrotoluene	x	640	172,991	0	0	6.785E-07	0.0120	4.117E-05		
2,6-Dinitrotoluene	x	1400	172,991	0	0	1.699E-07	0.0030	1.026E-05		
Dioxane		55	853,311	0	0	1.979E-04	0.9541	3.145E-03	0.862	0.862
Epichlorohydrin	x	1800	1,101,668	0	0	1.658E-03	25.2777	7.767E-02	0.919	0.944
2-Ethoxyethanol	x	9	1,029,041	0	0	1.011E-04	0.9793	1.864E-03		
Ethyl acrylate		38	1,502,043	0	1.103E-06	3.905E-04	0.2272	1.975E-03	0.864	0.864
Ethylbenzene		1.8	2,287,245	0	1.502E-03	1.506E-02	20.3518	8.832E-02	0.318	0.802
Ethylene		0.29	1,087,214	0	0	1.532E-01	222.4212	8.324E-01		
Ethylene glycol		4.5	2,325,127	0	2.572E-04	1.318E-02	32.8758	1.110E-01	0.193	0.640
Ethylene oxide	x	630	1,596,458	0	6.095E-08	2.653E-03	15.3360	2.949E-02	0.697	0.816
Fluometuron		91	234,097	0	0	9.855E-08	0.0008	3.857E-06	0.495	0.663
Formaldehyde		600	2,057,384	0	8.082E-04	6.716E-03	10.8607	5.305E-02	0.051	0.570
Freon 113		2.6	1,560,857	0	1.465E-06	1.811E-03	6.1030	2.861E-02		
Glycol ethers		90	2,390,475	0	6.528E-03	2.530E-02	38.6635	1.922E-01	0.766	0.766
Hexachloro-1,3-butadiene		16	872,714	0	0	3.530E-05	0.3133	1.058E-03	-0.259	-0.238
Hexachlorocyclopentadiene		9000	645,960	0	0	1.125E-05	0.0657	3.116E-04	-0.134	-0.134
Hexachloroethane		2.9	863,399	0	0	2.675E-05	0.2742	9.947E-04	0.882	0.882
Hydrazine		35000	632,980	0	0	3.155E-05	0.0829	4.350E-04	0.851	0.865
Hydrogen cyanide		600	761,634	0	0	2.177E-03	2.2461	1.850E-02	0.445	0.696
Hydrogen fluoride		130	2,365,954	0	6.273E-04	6.210E-03	14.0500	6.549E-02	0.609	0.757
Hydroquinone		13	1,185,936	0	0	1.104E-05	0.0324	1.869E-04	0.518	0.518
Isobutyraldehyde		51	1,071,977	0	0	7.249E-04	1.8009	1.213E-02		

Continued on Next Page...

Table C.13 – Continued

Chemical Name	Dev/ Rep Tox?	RSEI Tox Weight	Number Exposures	Min	Median	Mean (Units: $\mu g/m^3$)	Max	SD	Rank Correlation b/w RSEI 2002 and NATA 2002 All Source	Rank Correlation b/w RSEI 2002 and NATA 2002 Major Source
4,4'-Isopropylidenediphenol		10	1,318,282	0	0	7.092E-04	4.3930	1.174E-02		
Lead and lead compounds	x	8800	2,555,049	0	1.521E-04	7.216E-04	0.7150	4.262E-03	0.570	0.610
Maleic anhydride		2600	1,369,544	0	0	5.792E-04	1.7673	5.122E-03	0.923	0.923
Maneb		120	164,097	0	0	8.136E-08	0.0015	6.676E-06		
Manganese and manganese compounds		36000	2,391,758	0	1.293E-04	7.663E-04	2.1505	7.518E-03	0.760	0.788
Methanol		0.45	2,430,795	0	7.201E-03	1.119E-01	55.9614	7.168E-01	0.474	0.862
2-Methoxyethanol	x	90	928,170	0	0	1.867E-04	2.1521	8.779E-03		
Methyl acrylate		17	1,439,798	0	2.137E-10	2.445E-04	3.0898	4.267E-03		
Methyl ethyl ketone		0.36	2,532,247	0	9.609E-03	9.046E-02	282.7252	1.709E+00		
Methyl hydrazine		2200	449,675	0	0	6.224E-08	0.0012	3.693E-06	-0.136	-0.136
Methyl iodide		580	668,272	0	0	2.311E-04	0.8749	4.486E-03	0.885	0.916
Methyl isobutyl ketone		0.6	2,115,223	0	1.313E-03	1.233E-02	80.5669	1.609E-01	0.308	0.760
Methyl isocyanate		1800	563,004	0	0	1.637E-06	0.0389	3.241E-05	0.398	0.398
Methyl methacrylate		2.6	1,585,508	0	1.115E-05	1.881E-03	1.8368	9.281E-03	0.819	0.815
Methyl tert-butyl ether		0.6	1,668,372	0	5.855E-07	1.236E-02	10.6917	8.359E-02	0.472	0.815
4,4'-Methylenebis(2-chloroaniline)		3100	17,662	0	0	3.558E-06	0.2224	4.119E-04		
4,4'-Methylenebis(N,N-dimethylbenzenamine)		92	183,643	0	0	8.538E-09	0.0003	5.118E-07		
4,4'-Methylenedianiline		3300	853,714	0	0	7.162E-06	0.1739	3.366E-04	0.255	0.255
Molybdenum trioxide		7500	709,856	0	0	6.620E-05	0.0266	3.724E-04		
Naphthalene		600	2,031,105	0	1.524E-04	4.870E-03	23.4470	6.635E-02	0.487	0.621
Nickel and nickel compounds		20000	2,593,404	0	6.952E-05	3.895E-04	0.6368	3.530E-03	0.739	0.876
Nitric acid		140	2,347,899	0	2.690E-04	9.356E-04	0.7176	4.360E-03		
Nitrilotriacetic acid		40	111,985	0	0	1.145E-08	0.0002	3.247E-07		
Nitrobenzene	x	1100	693,460	0	0	2.723E-04	1.3162	3.731E-03	0.950	0.954
4-Nitrophenol		390	306,152	0	0	2.074E-07	0.0013	2.872E-06	-0.012	0.749
2-Nitropropane		19000	77,938	0	0	1.418E-07	0.0068	1.268E-05		
N-Nitrosodiphenylamine		9.8	149,335	0	0	1.158E-06	0.0192	6.646E-05		
p-Nitrosodiphenylamine		4.5	104,260	0	0	1.120E-09	0.0000	1.128E-07		
Pentachlorophenol		330	575,206	0	0	5.993E-06	0.1507	2.763E-04	0.133	-0.549
p-Phenylenediamine		2.6	157,958	0	0	7.890E-07	0.0141	4.833E-05	0.884	0.884
Phthalic anhydride		90	1,524,328	0	2.040E-07	1.217E-03	5.3497	2.482E-02	0.879	0.879
Propylene (Propene)		0.6	1,696,381	0	1.715E-05	1.067E-01	117.5850	6.095E-01	0.552	0.951
Propylene oxide		60	1,061,177	0	0	3.670E-03	14.0254	3.958E-02	0.873	0.901
Propyleneimine		300000	757,291	0	0	2.019E-06	0.0082	1.474E-05	0.901	0.901
Pyridine		500	795,991	0	0	1.303E-04	0.1399	6.822E-04		
Quinoline		6000	573,978	0	0	4.882E-06	0.0154	8.081E-05	0.689	0.260
Quintozene		52	242,716	0	0	6.878E-06	0.0477	1.677E-04		
Selenium and selenium compounds		90	145,151	0	0	9.460E-05	0.1603	1.256E-03	0.584	0.616
Styrene		1.8	2,616,546	0	8.840E-03	4.921E-02	83.2517	3.289E-01	0.667	0.845
1,1,2,2-Tetrachloroethane		41	917,623	0	0	1.073E-04	2.1807	5.452E-03	0.447	0.750
Tetrachloroethylene (Perchloroethylene)		42	1,929,541	0	2.914E-04	7.488E-03	12.3859	1.087E-01	0.385	0.728
Tetrachlorvinphos		12	135,570	0	0	7.862E-08	0.0041	7.977E-06		
Thiourea		2000	185,328	0	0	2.760E-08	0.0001	3.861E-07		
Toluene	x	0.36	2,592,325	0	2.456E-02	1.094E-01	134.8416	1.005E+00	0.388	0.869
Toluene-2,4-diisocyanate		26000	1,200,275	0	0	5.359E-06	0.0125	7.797E-05	-0.636	0.938
Toluene-2,6-diisocyanate		26000	545,726	0	0	3.259E-06	0.0186	7.239E-05		
Toluenediisocyanate		26000	1,691,555	0	3.904E-06	1.026E-04	0.3288	1.594E-03		
o-Toluidine		480	729,750	0	0	8.286E-05	0.2388	2.487E-03	0.949	0.949
Trichlorfon		250	345,637	0	0	1.469E-07	0.0017	5.760E-06		
1,2,4-Trichlorobenzene		9	854,584	0	0	1.756E-04	0.5200	4.197E-03	0.322	0.167
1,1,1-Trichloroethane		1.8	1,325,138	0	0	1.146E-02	52.0573	2.015E-01	0.154	0.272
1,1,2-Trichloroethane		11	1,047,714	0	0	3.398E-04	2.4078	8.278E-03	0.698	0.884
Trichloroethylene		14	2,170,844	0	1.127E-03	1.439E-02	19.0134	1.353E-01	0.345	0.632
1,2,4-Trimethylbenzene		300	2,167,796	0	7.638E-04	7.873E-03	29.4611	9.350E-02	-0.373	-0.218
Vanadium and vanadium compounds		71	788,105	0	0	4.592E-05	0.0647	5.764E-04		
Vinyl acetate		9	1,688,157	0	7.579E-05	1.590E-02	34.5199	1.489E-01	0.894	0.892

Continued on Next Page...

Table C.13 – Continued

Chemical Name	Dev/ Rep Tox?	RSEI Tox Weight	Number Exposures	Min	Median	Mean (Units: $\mu g/m^3$)	Max	SD	Rank Correlation b/w RSEI 2002 and NATA 2002 All Source	Rank Correlation b/w RSEI 2002 and NATA 2002 Major Source
Vinyl chloride		63	891,739	0	0	3.036E-03	6.6457	2.496E-02	0.868	0.936
Vinylidene chloride (1,1-dichloroethylene)		9	731,431	0	0	7.081E-05	1.7884	4.697E-03	0.236	0.479
o-Xylene		18	1,467,065	0	1.105E-08	4.260E-03	9.0078	4.366E-02		
p-Xylene		18	1,290,753	0	0	2.903E-03	23.7696	4.606E-02		
m-Xylene		18	1,465,203	0	1.085E-09	3.754E-03	26.6661	5.099E-02		
Xylene (mixed isomers)		18	2,652,070	0	2.066E-02	9.335E-02	92.1253	6.793E-01	0.415	0.817
Zinc and zinc compounds		51	2,464,804	0	6.243E-04	1.207E-02	30.7570	2.712E-01		

Source: Author's calculations based on 1) birth data from the Texas Center for Health Statistics (1994-2003) merged with 2) km^2-grid-cell-level RSEI chemical concentrations.
Notes: The dataset from which summary statistics (min, median, mean, max, sd) were calculated includes 2,868,306 births from the 1994-2003 Texas "CBSA" sample merged with RSEI chemical concentration measures. The concentration measures are weighted averages of the annual concentrations for the year(s) in which the child was in utero. The unit for all chemicals is ($\mu g/m^3$). Only chemicals reported to the TRI during each year between 1994-2003 are on the list. Chemicals marked as developmental or reproductive toxicants are so designated based on inclusion on the list of such toxicants produced by the California EPA in accordance with Proposition 65 [23]. Toxicity weights are those assigned to TRI chemicals in RSEI model 2.15 [49]. The NATA-RSEI correlation measures use RSEI and NATA data from the year 2002 only. RSEI tract-average concentrations were found by merging all 2002 births with 2002 km^2-grid-cell-level concentration values and averaging these concentration values across tracts. Spearman rank correlation coefficients were then calculated between RSEI 2002 tract-average concentrations and NATA 2002 tract concentrations from 1) all sources and 2) major ("point") sources only.

BIBLIOGRAPHY

[1] Acevedo-Garcia, D. Zip code-level risk factors for tuberculosis: neighborhood environment and residential segregation in New Jersey, 1985-1992. *American Journal of Public Health 91*, 5 (2001), 734.

[2] Acevedo-Garcia, D., Soobader, M.J., and Berkman, L.F. The differential effect of foreign-born status on low birth weight by race/ethnicity and education. *Pediatrics 115*, 1 (2005), e20.

[3] Agarwal, N., Banternghansa, C., and Bui, L. Toxic exposure in America: Estimating fetal and infant health outcomes from 14 years of TRI reporting. *Journal of health economics 29*, 4 (2010), 557–574.

[4] Ananth, C.V., Misra, D.P., Demissie, K., and Smulian, J.C. Rates of preterm delivery among Black women and White women in the United States over two decades: an age-period-cohort analysis. *American Journal of Epidemiology 154*, 7 (2001), 657.

[5] Anderton, D.L., Anderson, A.B., Oakes, J.M., and Fraser, M.R. Environmental equity: the demographics of dumping. *Demography 31*, 2 (1994), 229–248.

[6] Arnold, G.L., Kirby, R.S., Langendoerfer, S., and Wilkins-Haug, L. Toluene embryopathy: clinical delineation and developmental follow-up. *Pediatrics 93*, 2 (1994), 216.

[7] Ash, M., Boyce, J., Chang, G., Scoggins, J., and Pastor, M. *Justice in the Air: Tracking Toxic Pollution from America's Industries and Companies to Our States, Cities, and Neighborhoods*. Political Economy Research Institute, University of Massachusetts at Amherst, 2009.

[8] Ash, M., Boyce, J.K., Chang, G., and Scharber, H. Is Environmental Justice Good for White Folks? *Political Economy Research Institute Working Papers* (2010).

[9] Ash, M., and Fetter, T.R. Who Lives on the Wrong Side of the Environmental Tracks? Evidence from the EPA's Risk-Screening Environmental Indicators Model. *Social Science Quarterly 85*, 2 (2004), 441–462.

[10] Banzhaf, H.S., and Walsh, R.P. Do people vote with their feet? An empirical test of Tiebout's mechanism. *The American Economic Review 98*, 3 (2008), 843–863.

[11] Baum, C.F. *An introduction to modern econometrics using Stata*. Stata Corp, 2006.

[12] Been, V., and Gupta, F. Coming to the Nuisance or Going to the Barrios—A Longitudinal Analysis of Environmental Justice Claims. *Ecology LQ 24* (1997), 1–56.

[13] Bellinger, D.C. Teratogen update: lead and pregnancy. *Birth Defects Research Part A: Clinical and Molecular Teratology 73*, 6 (2005), 409–420.

[14] Boone, C.G. Environmental justice as process and new avenues for research. *Environmental Justice 1*, 3 (2008), 149–154.

[15] Borders, A.E.B., Grobman, W.A., Amsden, L.B., and Holl, J.L. Chronic stress and low birth weight neonates in a low-income population of women. *Obstetrics & Gynecology 109*, 2, Part 1 (2007), 331.

[16] Bowen, W. An analytical review of environmental justice research: What do we really know? *Environmental Management 29*, 1 (2002), 3–15.

[17] Boyce, J.K. Let them eat risk? Wealth, rights and disaster vulnerability. *Disasters 24*, 3 (2000), 254–261.

[18] Breslau, N., DelDotto, J.E., Brown, G.G., Kumar, S., Ezhuthachan, S., Hufnagle, K.G., and Peterson, E.L. A gradient relationship between low birth weight and IQ at age 6 years. *Archives of Pediatrics and Adolescent Medicine 148*, 4 (1994), 377.

[19] Brooks, N., and Sethi, R. The distribution of pollution: Community characteristics and exposure to air toxics. *Journal of Environmental Economics and Management 32*, 2 (1997), 233–250.

[20] Brooks-Gunn, J., Klebanov, P.K., and Duncan, G.J. Ethnic differences in children's intelligence test scores: Role of economic deprivation, home environment, and maternal characteristics. *Child Development 67*, 2 (1996), 396–408.

[21] Buescher, P.A., and Mittal, M. Racial disparities in birth outcomes increase with maternal age: recent data from North Carolina. *North Carolina medical journal 67*, 1 (2005), 16.

[22] California EPA. OEHHA/ARB Approved Chronic Reference Exposure Levels and Target Organs. `http://www.arb.ca.gov/toxics/healthval/chronic.pdf`, 2009.

[23] California Office of Environmental Health Hazard Assessment. Chemicals Known to the State to Cause Cancer or Reproductive Toxicity (Prop 65). Tech. rep., Sept. 2009.

[24] Callaghan, W.M., MacDorman, M.F., Rasmussen, S.A., Qin, C., and Lackritz, E.M. The contribution of preterm birth to infant mortality rates in the United States. *Pediatrics 118*, 4 (2006), 1566.

[25] Canfield, M.A., Ramadhani, T.A., Langlois, P.H., and Waller, D.K. Residential mobility patterns and exposure misclassification in epidemiologic studies of birth defects. *Journal of Exposure Science and Environmental Epidemiology 16*, 6 (2006), 538–543.

[26] Carman, A.E. *Ambient Air Toxic Releases and Adverse Pregnancy Outcomes in Allegheny County, Pennsylvania*. PhD thesis, 2009.

[27] Centers for Disease Control. Why not change the blood lead level of concern? `http://www.cdc.gov/nceh/lead/policy/changeBLL.htm`, 2009.

[28] Chakraborty, J., and Armstrong, M.P. Assessing the impact of airborne toxic releases on populations with special needs. *The Professional Geographer 53*, 1 (2001), 119–131.

[29] Chakraborty, J., and Maantay, J.A. Proximity Analysis for Exposure Assessment in Environmental Health Justice Research. *Geospatial Analysis of Environmental Health* (2011), 111–138.

[30] Chay, K.Y., and Greenstone, M. The Impact of Air Pollution on Infant Mortality: Evidence from Geographic Variation in Pollution Shocks Induced by a Recession. *The Quarterly Journal of Economics 118*, 3 (2003), 1121–1167.

[31] Clements, A.L., Flatt, V.B., Fraser, M.P., Hamilton, W.J., Ledvina, P.S., Mathur, S.K., Tamhane, A., and Ward, J.B. *The Control of Air Toxics: Toxicology Motivation and Houston Implications*. Rice University, Houston, 2006.

[32] Clinton, W.J. Executive Order 12898 (Environmental Justice). `http://www.epa.gov/fedreg/eo/eo12898.htm`, Feb. 1994.

[33] Collins, C.A., and Williams, D.R. Segregation and mortality: The deadly effects of racism? *Sociological Forum 14*, 3 (1999), 495–523.

[34] Collins Jr, J.W., Simon, D.M., Jackson, T.A., and Drolet, A. Advancing maternal age and infant birth weight among urban African Americans: the effect of neighborhood poverty. *Ethnicity & disease 16*, 1 (2006), 180.

[35] Currie, J. Healthy, wealthy, and wise: Socioeconomic status, poor health in childhood, and human capital development. *Journal of Economic Literature 47*, 1 (2009), 87–122.

[36] Currie, J. Inequality at birth: Some causes and consequences. *National Bureau of Economic Research* (2011).

[37] Currie, J., and Hyson, R. Is the impact of health shocks cushioned by socioeconomic status? The case of low birthweight. *National Bureau of Economic Research* (1999).

[38] Currie, J., Neidell, M., and Schmieder, J.F. Air pollution and infant health: Lessons from New Jersey. *Journal of Health Economics 28*, 3 (2009), 688–703.

[39] Currie, J., and Schmieder, J.F. Fetal Exposures to Toxic Releases and Infant Health. *The American Economic Review 99*, 2 (2009), 177.

[40] Currie, J., and Walker, R. Traffic congestion and infant health: Evidence from E-Zpass, 2009.

[41] David, R.J., and Collins Jr, J.W. Differing birth weight among infants of US-born blacks, African-born blacks, and US-born whites. *New England Journal of Medicine 337*, 17 (1997), 1209–1214.

[42] de Barros, R.P., and Ferreira, F.H.G. *Measuring inequality of opportunities in Latin America and the Caribbean*. World Bank Publications, 2009.

[43] Downey, L. Single mother families and industrial pollution in metropolitan America. *Sociological Spectrum 25*, 6 (2005), 651–675.

[44] Downey, L. US metropolitan-area variation in environmental inequality outcomes. *Urban Studies 44*, 5-6 (2007), 953.

[45] Downey, L., and Hawkins, B. Single-Mother Families and Air Pollution: A National Study. *Social science quarterly 89*, 2 (2008), 523–536.

[46] Ekwo, E.E., and Moawad, A. Maternal age and preterm births in a black population. *Paediatric and perinatal epidemiology 14*, 2 (2000), 145–151.

[47] Environmental Defense Fund. *Toxic ignorance: The continuing absence of basic health testing for top-selling chemicals in the United States*. 1997.

[48] EPA. Risk-Screening Environmental Indicators Geographic Microdata, 1993-2003.

[49] EPA. User's Manual for RSEI Model Version 2.1, Technical Appendix A: Listing of All Toxicity Weights for TRI Chemicals and Chemical Categories. `http://www.epa.gov/oppt/rsei/pubs/tech_app_a.pdf`, 2002.

[50] EPA. EPAs Risk-Screening Environmental Indicators (RSEI) Methodology (Version 2.1.5). `http://www.epa.gov/oppt/rsei/pubs/method_oct2007.pdf`, Oct. 2007.

[51] EPA. Technical Appendices for RSEI Version 2.1.5. `http://www.epa.gov/oppt/rsei/pubs/archive/index.html`, Oct. 2007.

[52] EPA. About Air Toxics. http://www.epa.gov/oar/toxicair/newtoxics.html, 2009.

[53] EPA. TSCA Inventory Reset. http://www.epa.gov/champ/pubs/hpv/tsca.html, 2009.

[54] EPA. What is the Toxics Release Inventory Program? http://www.epa.gov/TRI/triprogram/whatis.htm, 2009.

[55] EPA. About the EPA Administrator. http://www.epa.gov/aboutepa/administrator.html, 2011.

[56] EPA. Assessing Outdoor Air Near Schools. http://www.epa.gov/schoolair/index.html, 2011.

[57] EPA. Integrated Risk Information System (IRIS). http://www.epa.gov/IRIS/, 2011.

[58] Evans, G.W., and Kantrowitz, E. Socioeconomic status and health: the potential role of environmental risk exposure. *Annual Review of Public Health 23*, 1 (2002), 303–331.

[59] Franzini, L., Ribble, J.C., and Keddie, A.M. Understanding the Hispanic paradox. *Ethnicity & disease 11*, 3 (2001), 496.

[60] Geronimus, A.T. The weathering hypothesis and the health of African-American women and infants: evidence and speculations. *Ethnicity & Disease 2*, 3 (1992), 207.

[61] Geronimus, A.T. Black/white differences in the relationship of maternal age to birthweight: a population-based test of the weathering hypothesis. *Social Science & Medicine 42*, 4 (1996), 589–597.

[62] Geronimus, A.T., Bound, J., and Neidert, L.J. On the validity of using census geocode characteristics to proxy individual socioeconomic characteristics. *Journal of the American Statistical Association 91*, 434 (1996), 529–537.

[63] Grandjean, P, and Landrigan, P. Developmental neurotoxicity of industrial chemicals. *The Lancet 368*, 9553 (Dec. 2006), 2167–2178.

[64] Henson, M.C., and Chedrese, P.J. Endocrine disruption by cadmium, a common environmental toxicant with paradoxical effects on reproduction. *Experimental Biology and Medicine 229*, 5 (2004), 383.

[65] Hogue, C.J.R., and Bremner, J.D. Stress model for research into preterm delivery among black women. *American Journal of Obstetrics and Gynecology 192*, 5 (2005), S47–S55.

[66] Holzman, C., Eyster, J., Kleyn, M., Messer, L.C., Kaufman, J.S., Laraia, B.A., O'Campo, P., Burke, J.G., Culhane, J., and Elo, I.T. Maternal weathering and risk of preterm delivery. *American journal of public health 99*, 10 (2009), 1864.

[67] Hunter, L.M. The spatial association between US immigrant residential concentration and environmental hazards. *International Migration Review 34*, 2 (2000), 460–488.

[68] Kids Count Data Center. Texas Infant Health Outcome Data. `http://datacenter.kidscount.org/data/bystate/StateLanding.aspx?state=TX`, 2009.

[69] Kotelchuck, M. An evaluation of the Kessner Adequacy of Prenatal Care Index and a proposed Adequacy of Prenatal Care Utilization Index. *American Journal of Public Health 84*, 9 (1994), 1414.

[70] Landrigan, P.J., Kimmel, C.A., Correa, A., and Eskenazi, B. Children's health and the environment: Public health issues and challenges for risk assessment. *Environmental health perspectives 112*, 2 (2004), 257.

[71] Lerner, S. *Sacrifice Zones: The Front Lines of Toxic Chemical Exposure in the United States*. The MIT Press, 2010.

[72] Lindstrom, K., Winbladh, B., Haglund, B., and Hjern, A. Preterm infants as young adults: a Swedish national cohort study. *Pediatrics 120*, 1 (2007), 70.

[73] Love, C., David, R.J., Rankin, K.M., and Collins, J.W. Exploring weathering: effects of lifelong economic environment and maternal age on low birth weight, small for gestational age, and preterm birth in African-American and white women. *American journal of epidemiology 172*, 2 (2010), 127.

[74] Maantay, J. Mapping environmental injustices: Pitfalls and potential of geographic information systems in assessing environmental health and equity. *Environmental Health Perspectives 110*, Suppl 2 (2002), 161.

[75] MacDorman, M.F., Mathews, TJ, and for Health Statistics (US), National Center. Recent trends in infant mortality in the United States.

[76] Maisonet, M., Correa, A., Misra, D., and Jaakkola, J. J.K. A review of the literature on the effects of ambient air pollution on fetal growth. *Environmental Research 95*, 1 (2004), 106–115.

[77] Massey, D.S., and Denton, N.A. The dimensions of residential segregation. *Social forces 67*, 2 (1988), 281–315.

[78] Mathews, TJ, and MacDorman, M.F. Infant mortality statistics from the 2006 period linked birth/infant death data set. *National vital statistics reports: from the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System 58*, 17 (2010), 1.

[79] Messer, L.C., Laraia, B.A., Kaufman, J.S., Eyster, J., Holzman, C., Culhane, J., Elo, I., Burke, J.G., and Ocampo, P. The development of a standardized neighborhood deprivation index. *Journal of Urban Health 83*, 6 (2006), 1041–1062.

[80] Mohai, P., and Saha, R. Reassessing racial and socioeconomic disparities in environmental justice research. *Demography 43*, 2 (2006), 383–399.

[81] Moore, C.F. *Children and Pollution: Why Scientists Disagree*. Oxford University Press, USA, 2009.

[82] Morello-Frosch, R., and Jesdale, B.M. Separate and unequal: Residential segregation and estimated cancer risks associated with ambient air toxics in US metropolitan areas. *Environmental Health Perspectives 114*, 3 (2006), 386.

[83] Morello-Frosch, R., and Lopez, R. The riskscape and the color line: Examining the role of segregation in environmental health disparities. *Environmental research 102*, 2 (2006), 181–196.

[84] Morello-Frosch, R., Zuk, M., Jerrett, M., Shamasunder, B., and Kyle, A.D. Understanding The Cumulative Impacts Of Inequalities In Environmental Health: Implications For Policy. *Health Affairs 30*, 5 (2011), 879.

[85] Nishijo, M., Nakagawa, H., Honda, R., Tanebe, K., Saito, S., Teranishi, H., and Tawara, K. Effects of maternal exposure to cadmium on pregnancy outcome and breast milk. *Occupational and environmental medicine 59*, 6 (2002), 394.

[86] Oakes, J.M., Anderton, D.L., and Anderson, A.B. A Longitudinal Analysis of Environmental Equity in Communities with Hazardous Waste Facilities. *Social Science Research 25*, 2 (1996), 125–148.

[87] Osypuk, T.L., and Acevedo-Garcia, D. Are racial disparities in preterm birth larger in hypersegregated areas? *American journal of epidemiology 167*, 11 (2008), 1295.

[88] Pastor, M., Sadd, J., and Hipp, J. Which Came First? Toxic Facilities, Minority Move-in, and Environmental Justice. *Journal of Urban Affairs 23*, 1 (2001), 1–21.

[89] Pastor, M., Sadd, J.L., and Morello-Frosch, R. Reading, writing, and toxics: children's health, academic performance, and environmental justice in Los Angeles. *Environment and Planning C 22*, 2 (2004), 271–290.

[90] Petronis, Arturas. Epigenetics as a unifying principle in the aetiology of complex traits and diseases. *Nature 465* (June 2010), 721–727.

[91] Polednak, A.P. Trends in US urban black infant mortality, by degree of residential segregation. *American Journal of Public Health 86*, 5 (1996), 723.

[92] Rauh, V.A., Andrews, H.F., and Garfinkel, R.S. The contribution of maternal age to racial disparities in birthweight: a multilevel perspective. *American journal of public health 91*, 11 (2001), 1815.

[93] Rauh, V.A., Garfinkel, R., Perera, F.P., Andrews, H.F., Hoepner, L., Barr, D.B., Whitehead, R., Tang, D., and Whyatt, R.W. Impact of prenatal chlorpyrifos exposure on neurodevelopment in the first 3 years of life among inner-city children. *Pediatrics* (2006), peds–2006.

[94] Reyes, J. The Impact of Prenatal Lead Exposure on Infant Health. *Department of Economics, Amherst College, working paper* (2003).

[95] Rich-Edwards, J.W., Buka, S.L., Brennan, R.T., and Earls, F. Diverging associations of maternal age with low birthweight for black and white mothers. *International journal of epidemiology 32*, 1 (2003), 83.

[96] Right-to-Know Network. The Right-to-Know Network Website. `http://www.rtknet.org/db/tri/`, 2011.

[97] Ringquist, E.J. Assessing evidence of environmental inequities: A meta-analysis. *Journal of Policy Analysis and Management 24*, 2 (2005), 223–247.

[98] Schempf, A.H., Branum, A.M., Lukacs, S.L., and Schoendorf, K.C. Maternal age and parity-associated risks of preterm birth: differences by race/ethnicity. *Paediatric and perinatal epidemiology 21*, 1 (2007), 34–43.

[99] Sexton, K., and Hattis, D. Assessing cumulative health risks from exposure to environmental mixtures: three fundamental questions. *Environmental health perspectives 115*, 5 (2007), 825.

[100] Shin, I.S., Park, N.H., Lee, J.C., Kim, K.H., Moon, C., Kim, S.H., Shin, D.H., Park, S.C., Kim, H.Y., and Kim, J.C. One-generation reproductive toxicity study of epichlorohydrin in Sprague-Dawley rats. *Drug and chemical toxicology 33*, 3 (2010), 291–301.

[101] Slama, R., Darrow, LA, Parker, JD, Woodruff, TJ, Strickland, M., Nieuwenhuijsen, M., Glinianaia, S., Hoggatt, K.J., Kannan, S., Hurley, F., et al. Atmospheric pollution and human reproduction: report of the Munich International Workshop. *Environmental Health Perspectives 116*, 6 (2008), 791–798.

[102] Stevens-Simon, C. The weathering hypothesis. *American Journal of Public Health 92*, 4 (2002), 507.

[103] Stillerman, K.P., Mattison, D.R., Giudice, L.C., and Woodruff, T.J. Environmental Exposures and Adverse Pregnancy Outcomes: A Review of the Science. *Reproductive Sciences 15*, 7 (Sept. 2008), 631–650.

[104] Subramanian, S.V., and Almeida-Filho, N. A glossary for health inequalities. *Journal of Epidemiology and Community Health 56* (2002), 647–652.

[105] Suh, H. H., Bahadori, T., Vallarino, J., and Spengler, J. D. Criteria air pollutants and toxic air pollutants. *Environmental Health Perspectives 108*, Suppl 4 (2000), 625.

[106] Texas Department of State Health Services, Center for Health Statistics. Texas birth records, 1994-2003.

[107] Thompson, J., and Bannigan, J. Cadmium: Toxic effects on the reproductive system and the embryo. *Reproductive Toxicology 25*, 3 (2008), 304–315.

[108] UNICEF. Childinfo.org: Statistics by Area - Low birthweight - Country data. `http://www.childinfo.org/low\_birthweight\_profiles.php`, Apr. 2011.

[109] United Church of Christ Commission for Racial Justice. Race in the United States: A National Report on the Racial and Socio-Economic Characteristics of Communities with Hazardous Waste Sites.

[110] United States Census Bureau. Census 2000 Summary File 3-Texas, 2002.

[111] US Census Bureau. Census Tracts and Block Numbering Areas. `http://www.census.gov/geo/www/cen\_tract.html`, Apr. 2000.

[112] USA Today. The Smokestack Effect: Toxic Air and Americas Schools.

[113] Wake Forest University School of Medicine. Statistical Power Calculators. `http://power.phs.wfubmc.edu/index.cfm?calc=osf`, 2011.

[114] Wildsmith, E.M. Testing the weathering hypothesis among Mexican-origin women. *Ethnicity and Disease 12*, 4 (2002), 470–479.

[115] Wilkins-Haug, L., and Gabow, P.A. Toluene abuse during pregnancy: obstetric complications and perinatal outcomes. *Obstetrics & Gynecology 77*, 4 (1991), 504.

[116] Williams, D.R., and Collins, C. Racial residential segregation: a fundamental cause of racial disparities in health. *Public Health Reports 116*, 5 (2001), 404–416.

[117] Wolverton, A. Effects of socio-economic and input-related factors on polluting plants' location decisions. *The BE Journal of Economic Analysis & Policy 9*, 1 (2009), 1–30.

[118] Wooldridge, J.M. *Econometric analysis of cross section and panel data*. The MIT press, 2002.