Environmental Health Risks and Housing Values:
Evidence from 1600 Toxic Plant Openings and Closings

Janet Currie   Lucas Davis   Michael Greenstone   Reed Walker[1]

Abstract

Regulatory oversight of toxic emissions from industrial plants and understanding about these emissions' impacts are in their infancy. Applying a research design based on the openings and closings of 1,600 industrial plants to rich data on housing markets and infant health, we find that: toxic air emissions affect air quality only within 1 mile of the plant; plant openings lead to 11% declines in housing values within 0.5 mile or a loss of about $4.25 million for these households; and a plant's operation is associated with a roughly 4% increase in the probability of low birth weight within 1 mile.

1. Introduction

Industrial plants that emit toxics are ubiquitous in the United States today, and many lie in close proximity to major population centers. These plants emit nearly 4 billion pounds of toxic pollutants in the United States annually, including 80,000 different chemical compounds (GAO 2010).[2] Whereas criteria air pollutants like particulate matter have been regulated for decades, regulation of other airborne toxics remains in its infancy. The unveiling of the Mercury and Air Toxics Standards in December 2011 represents the first time the U.S. government has enforced limits on mercury and other toxic chemicals and most of the chemicals emitted have never undergone any form of toxicity testing (U.S. Department of Health and Human Services 2010).[3]

Toxic emissions are one of the reasons why siting industrial plants is so controversial. Policy makers must balance the negative externalities associated with industrial plants with their potential to create jobs, increase local economic activity, and lead to positive economic spillovers (Greenstone, Hornbeck, and Moretti, 2010). While negative externalities often generate intense local opposition (e.g. "not in my backyard" or NIMBY movements), there is also frequently intense competition among communities to entice industrial plants to locate within their jurisdictions. If siting decisions are to be made efficiently, it is crucial that policymakers have reliable measures of the different costs and benefits. The existing evidence on toxic emissions is particularly sparse, despite the fact that toxics are widely believed to cause cancer, birth defects, and damage to the brain and reproductive systems (U.S. Centers for Disease Control and Prevention 2009).

This paper represents a first step towards understanding the external costs of industrial plants that emit toxics in the United States in terms of both individuals' willingness to pay to avoid toxic industrial facilities and population health. In order to address this question, we have assembled an extraordinarily rich dataset on the location and economic activity of industrial plants in five large U.S. states. Our analysis focuses, in particular, on plants that report toxic emissions to the U.S. Environmental Protection Agency's *Toxic Release Inventory*. We link information on these "toxic" plants with administrative data that provides detailed information on the near-universe of housing

---

[2] Source: http://www.gao.gov/highrisk/risks/safety-security/epa_and_toxic_chemicals.php. Accessed on March 19th, 2012.
[3] The Environmental Protection Agency characterizes their risk assessments as "not completely accurate" because "scientists don't have enough information on actual exposure and on how toxic air pollutants harm human cells. The exposure assessment often relies on computer models when the amount of pollutant getting from the source(s) to people can't be easily measured. Dose-response relationships often rely on assumptions about the effects of pollutants on cells for converting results of animal experiments at high doses to human exposures at low doses." Source: http://www.epa.gov/ttn/atw/3_90_024.html (accessed on March 20, 2012).

transactions and birth outcomes in these states. All three datasets provide geographic coordinates, so we are able to perform the analysis with an unusually high degree of spatial detail.

Since the previous literature offers little guidance about how far toxic air pollution emissions travel, our first contribution is to measure the relationship between toxic emissions and air quality. Using data from pollution monitoring stations and a difference-in-difference estimator, we document that on average there are significantly higher levels of ambient toxic pollution within one mile of operating plants but no significant effect at further distances. On average, each birth in our sample lies within one mile of 1.27 toxic plants, so our results imply that the total amount of exposure could be substantial. Moreover, these results motivate our baseline research design which compares households within one mile of a plant to households one to two miles away.

The findings on the distance that toxic air emissions travel guide our research design, which is based on the sharp changes in local amenities that result from more than 1,600 toxic plants openings and closings.[4] Our estimates are based on comparing housing prices and birth outcomes within 0.5 miles or 1 mile of plants with these same outcomes measured 1-2 miles away from plants, after adjustment for all unobserved time-varying factors that are common within 2 miles of the plants.[5] Further, the estimates are based on millions of births and hundreds of thousands of housing transactions.

This research design reveals that housing prices within 0.5 miles of a toxic plant's site decrease by about 11% after a plant opens, relative to the period before the plant was constructed. This decline implies an aggregate loss in housing values of approximately $4.25 million for the average plant opening. Housing prices are largely unaffected by a plant closing, relative to the period when the plant was operating, implying that toxic plants continue to negatively affect housing prices after they cease operations. Potential explanations for a plant's lasting effect include persistent visual disamenities, concerns about local contamination of soil, and an expectation that the plant will reopen.

Many toxic pollutants are colorless, odorless, and not well monitored, making them less salient than other negative externalities. Thus, it is valuable to contrast housing prices with health

---

[4] Our approach is inspired by pioneering studies by C. Arden Pope and collaborators who examined the health effects of opening and closing the Geneva steel mill near Provo, Utah in the late 1980s (Pope 1989; Ransom and Pope 1992; Pope, Schwartz, and Ransom 1992). These studies have been influential largely because the resulting sharp changes in airborne particulates over a short period of time make the empirical analyses transparent and highly credible.

[5]There have been attempts to study the health and housing price responses of toxic emissions at the county level (Agarwala, Banternghansa, and Bui, 2010; Bui and Mayer, 2003; Currie and Schmieder, 2009), but counties are too large due to the short transport distances of most airborne toxic pollutants (see Figure 1).

outcomes, which should immediately respond to changes in plant activity. We find that the incidence of low birth weight increases by roughly 3% within one mile of operating toxic plants, with comparable magnitudes between 0 and 0.5 miles and 0.5 and 1 miles. Using an estimate of the monetary costs of low birth weight, these estimates imply an external cost of $4,500 annually per plant. The estimated health impacts also appear to be highly localized, with no impact beyond one mile.

We believe our study is the first large-scale empirical analysis of the external costs of toxic plants.[6] This large scale allows us to begin to characterize the heterogeneity of effects across plants. In additional results, we stratify plants by size, the amount and toxicity of emissions, and local demographic characteristics and find that the housing price and health impacts are experienced broadly across different types of plants. There is some evidence that housing price responses are stronger in lower income communities, whereas the estimated health effects are relatively uniform across plant and community types.

The rest of the paper proceeds as follows: Section 2 presents an analytical framework which helps motivate the empirical analysis. Section 3 discusses the data, and Section 4 discusses the research design. Sections 5 and 6 outline the econometric specifications and results for housing values and infant health respectively. Finally, Section 7 interprets the results and 8 concludes.

2. Conceptual Framework for the Incidence of Toxic Plant Openings

To motivate our empirical strategy, we outline a partial equilibrium model of housing incidence in the context of toxic plant externalities.[7] A local economy consists of a continuum of agents of measure one (denoted L) who choose to live in one of two locations $g \in \{N, F\}$; some choose to live near a plant ($g = N$) and others choose to live further away from a plant ($g = F$), but in the same local labor market. Toxic plant activity is assumed to generate local economic benefits for both sets of residents in the form of wage income, w. Wages are assumed to be an exogenous function of local productivity and are the same across groups. Residents in each location enjoy location-specific amenities net of any housing costs, A_g, associated with their location. Lastly, each resident i has some idiosyncratic preference for both locations, ϵ_{ig}, representing heterogeneity in the

---

[6] Studies of individual plants include the studies by C. Arden Pope mentioned above, as well as Blomquist (1974), Nelson (1981), and Kiel and McClain (1995). For studies of multiple plants see, e.g., Bui and Mayer (2003) and Davis (2011).

[7] The results from this partial equilibrium exercise generalize into a model of general equilibrium of the sort found in Kline (2010) and Moretti (2011). These models are themselves generalizations of the canonical models of Rosen (1974) and Roback (1982).

valuation of local amenities. The ϵ_{ig}'s are independently and identically distributed across individuals and assumed to possess a continuous multivariate distribution with mean zero.

An individual seeks to maximize utility by choosing over locations

$$U_{ig} = \max\{v_N + \epsilon_{iN}, v_F + \epsilon_{iF}\},$$

where v_g represents mean utility in location g. Individuals will locate in whichever community yields the highest utility. Without heterogeneity in locational preferences, all individuals will locate in the community that offers the highest amenities. With heterogeneity in tastes, individuals in location N will have $v_N - v_F > \epsilon_{iF} - \epsilon_{iN}$. Define the distribution function $\eta_i \equiv \epsilon_{iF} - \epsilon_{iN}$ by $G(.)$. Then, $L_N \equiv \Pr(\eta_i < v_N - v_F)$ is the measure of individuals in location N.

Write the total welfare of workers in location N and F as

$$V = \mathrm{E}[\max\{v_N + \epsilon_{iN}, v_F + \epsilon_{iF}\}]$$

and consider a positive economic shock stemming from a toxic plant opening in the community. We model this shock as a marginal improvement in productivity in the local community, which is assumed to increase wages in both the near and far locations equally. The plant opening, however, creates a negative externality for residents living near the plant through, for example, air pollution and related health effects.

Taking the derivative of workers' welfare with respect to the economic shock associated with a plant opening yields the expression:

(1) $$\frac{dV}{d\theta} = L_F \cdot \frac{\partial w}{\partial \theta} + L_N \cdot \left[\frac{\partial w}{\partial \theta} + \frac{\partial A_N}{\partial \theta}\right] = L \cdot \frac{\partial w}{\partial \theta} + L_N \cdot \frac{\partial A_N}{\partial \theta}$$

where $d\theta$ represents the marginal effect of a plant opening and $\frac{dV}{dv_g} = L_g$.[8] Equation (1) suggests the incidence of the plant opening may be summarized by two terms. The first term is the total wage effect associated with the plant opening. Since in our empirical application, all residents near or far live within two miles of a plant, we assume that the wage effects are similar for both nearby residents and those a little further from a plant. The second term consists of the non-wage changes in amenities associated with a plant opening for residents near the plant. Since negative plant externalities in the form of noise or air pollution are highly localized, these costs will only accrue to the residents living near the plant.

After the plant opening some "marginal" residents who initially lived near the plant are better off moving further away. However, since workers are assumed to be optimizing with respect to

---

[8] The relationship $\frac{dV}{dv_g} = L_g$ follows directly from assuming that preference heterogeneity is drawn from a Type I Extreme Value distribution (Train 2003). However, this relationship also holds independent of the distribution of the taste heterogeneity. See Busso, Gregory, and Kline (2013).

location decisions, a simple envelope result suggests that workers who switch locations in response to a change in local amenities experienced large gains in private utility by doing so.[9] Therefore, the incidence of the plant opening may be approximated simply by the change in prices experienced by the immobile population.[10]

This paper aims to estimate the local disamenities of toxic plant operation, $\frac{\partial A_N}{\partial \theta}$, *holding all other factors fixed*. We do this by comparing residents near a plant to those within the same local labor market who live slightly further away. Since, by assumption, both groups are affected similarly by the productivity shock, the difference-in-differences estimate will approximate $\frac{\partial A_N}{\partial \theta}$. By explicitly controlling for the first component of equation (1) in this way, our estimates will reflect the gross external costs/benefits of a toxic plant opening or closing rather than the net external costs/benefits after accounting for any local economic gains associated with toxic plant production.

3. Data Sources and Summary Statistics

3.1. The Toxic Release Inventory Data

We identify toxic plants using the Toxic Release Inventory (TRI), a publicly-available database established and maintained by the U.S. Environmental Protection Agency (EPA).[11] The TRI was established by the Emergency Planning, Community Right to Know Act (EPCRA) in 1986, in response to the Bhopal disaster and a series of smaller spills of dangerous chemicals at American Union Carbide plants. Bhopal added urgency to the claim that communities had a "right to know" about hazardous chemicals that were being used or produced in their midst. EPCRA requires manufacturing plants (those in Standard Industrial Classifications 2000 to 3999) with more than 10 full-time employees that either use or produce more than threshold amounts of listed toxic substances to report releases to the EPA.[12]

---

[9] Although the change in amenities induces changes in behavior, these behavioral responses cannot have a first-order effect on private welfare; if they did, agents would not be optimizing. Alternatively, in this model the marginal migrant is indifferent between location 1 and location 2. Thus, any marginal shift in amenities in location 1 cannot make the agent much better off given the pre-intervention indifference between the two locations. Of course, plant openings and closings might not be marginal changes.

[10] In the case of non-marginal changes in productivity or local amenities, the envelope theorem no longer holds, and taste-based sorting may also have first-order implications for welfare. However, in the case of localized disamenities such as a single plant, Bartik (1987) and Palmquist (1992) show that the slope of the hedonic price function is an approximate measure of the willingness to pay for a non-marginal change. See Greenstone and Gallagher (2008) for a more complete discussion of non-marginal changes in the context of environmental amenities.Equilibrium sorting models may also yield insight into the welfare effects of non-marginal changes in the context of environmental disamenities. See Kuminoff, Smith, and Timmins (2013) for a recent review.

[11] See EPA (2007) and EPA (2011b) for detailed descriptions of the TRI.

[12] Currently, facilities are required to report if they manufactured or processed more than 25,000 pounds of a listed chemical or "otherwise used" 10,000 pounds of a listed chemical. For persistent bio-accumulative toxins, the

The toxic emissions measures in the TRI have been widely criticized (Marchi and Hamilton, 2006; Koehler and Spengler, 2007; Bennear, 2008). The emissions data are self-reported, and believed to contain substantial measurement error.[13] Moreover, coverage has expanded over time to include additional industries and chemicals, making comparisons of total emissions levels over time extremely misleading.[14] Finally, because of the minimum thresholds for reporting, plants may go in and out of reporting even if they are continually emitting toxic chemicals. This feature of the TRI introduces additional measurement error, and also makes the TRI poorly suited for identifying plant openings and closings.

The TRI is extremely useful, however, for identifying which U.S. industrial plants emit toxic pollutants. The approach we adopt in this paper is to ignore the self-reported magnitudes and instead exploit variation introduced by plant openings and closings. Using the publicly-available TRI data, we create a list of all U.S. "toxic" plants by keeping every plant that ever reported toxic emissions to the TRI in any year. This method sidesteps the problems introduced by changes in reporting requirements because plants end up being classified as "toxic" plants, even if, for example, they are in industries which were not included in the early years of the TRI. We then link this list of toxic plants to establishment-level data from the U.S. Census Bureau to determine the years in which each plant opened (and closed, if applicable).

3.2. The Longitudinal Business Database

We determine the exact years in which plants open and close using the U.S. Census Bureau's Longitudinal Business Database (LBD). Started in 1975, the LBD is a longitudinal, establishment-level database of the universe of establishments in the United States.[15] The LBD has been used widely by economists, for example, in studying plant-level employment dynamics (Davis, Faberman, Haltiwanger, Jarmin, and Miranda, 2010), and is by far the most accurate existing record of U.S. plant activity.

thresholds are lower. These thresholds have changed periodically over the life of the program. For example, in 1998, EPA added the receipt or disposal of chemical waste to the definition of "otherwise used".

[13] The EPCRA explicitly states that plants need not engage in efforts to measure their emissions. The EPA provides guidance about possible estimation methodologies, but plants estimate their emissions themselves, and estimating methodologies vary between plants and over time. In addition, EPA enforcement of TRI reporting has typically taken the form of ensuring compliance rather than accuracy (de Marchi and Hamilton, 2006).

[14] Federal facilities were added in 1994. Mining, electric utilities, hazardous waste treatment and disposal facilities, chemical wholesale distributors, and other additional industrial sectors were added in 1998. Treatment of persistent bio-accumulative toxins was changed in 2000. By the EPAs own admission, the TRI is not well suited for describing changes in total amounts of toxic releases over time (U.S. EPA 2011b).

[15] For more information about the LBD see Davis, Haltiwanger, and Schuh (1998) and Jarmin and Miranda (2002).

These data must be accessed at a census research data center under authorization from the Census Bureau. In addition to the year of opening and closing (if applicable) for each plant these data report mean annual employment and mean annual total salaries.[16] We merge the LBD with a second restricted access Census database called the Standard Statistical Establishment List (SSEL), which contains plant names and addresses for all plants in the LBD. Finally, we merge the LBD/SSEL dataset with the EPA's TRI database via a name and address-matching algorithm.[17]

3.3. Housing Values

The housing data for this project includes housing transactions in five large states (Texas, New Jersey, Pennsylvania, Michigan, and Florida). These data report the date, price, mortgage amount, and address of all property sales for these five states from approximately 1998 to 2005.[18] The data also include the exact street address of the property, which allows us to link the housing data with plant level data from the TRI based on the latitude and longitude of the geocoded address (described in more detail below). The main limitation of the housing data is that it contains very little information pertaining to housing unit characteristics.[19] These data include both residential and commercial real estate transactions; we focus only on single-family, residential properties. To limit the influence of outliers and focus on "arms length" transactions, we exclude properties that sold for less than $25,000 or more than $10 million. All housing prices have been adjusted to year 2000 dollars.

3.4. Vital Statistics Data

Data on infant health comes from vital statistics natality and mortality data for the same five large states: Texas, New Jersey, Pennsylvania, Michigan, and Florida, from 1990 to 2002. Together, these states accounted for 10.9 million births between 1990 and 2002, approximately 37 percent of all U.S. births. The substantial advantage of these restricted-access data is their geographic detail, including the residential address of the mother. This precision is crucial in our context because the health consequences of toxic plants are highly localized.

These data include detailed information about the universe of births and infant deaths in each state. We focus, in particular, on whether the infant is low birth weight defined as birth weight less than 2,500 grams. Low birth weight is not uncommon, affecting about seven percent of the births in

---

[16] The year of a plant opening is left-censored for those plants that were operating on or before 1975.

[17] See Walker (2013) for further details pertaining to the match algorithm.

[18] The transaction records are public due to state information disclosure acts, but the raw data are often housed in PDF images on county websites making them inaccessible for computational analysis on a large scale. We used an external data provider who compiled the information from the county registrar websites into a single dataset. Data availability and temporal coverage varies by county but is fairly consistent between 1998-2005, the years of our housing analysis.

[19] For example, we observe square footage of the housing unit for less than half of the transactions.

our sample. Low birth weight is also one of the most widely used overall indicators of infant health, in part because it has been shown to predict adult wellbeing.[20] Other birth outcomes that we examine include a continuous measure of birth weight, very low birth weight (defined as birth weight less than 1,500 grams), prematurity (defined as gestation less than 37 weeks), congenital abnormalities, and infant mortality (death in the first year).[21] Focusing on infant health is advantageous, relative to adult outcomes, because infants do not have a long unobserved health history, reducing concerns about time lags between exposure and outcomes.

In addition to these health outcomes, the vital statistics data include a number of important maternal characteristics including age, education, race, and smoking behavior. In the empirical analyses below we control explicitly for these factors, as well as for month of birth, birth order, and gender of child. In all analyses we exclude multiple births since they are likely to have poor birth outcomes for reasons that have little to do with environmental pollution. We also test whether plant openings and closings have affected these characteristics directly, either by changing the composition of neighborhoods near plants and/or by changing fertility.

The fact that the LBD data is annual, while births are reported monthly raises the question of how to appropriately structure the empirical models for infant health outcomes. We focus the analysis on a data file comprised of births in November, December, January, and February. Births in November and December are merged to LBD data from the same calendar year, while births from January and February are merged to LBD data from the preceding calendar year. The idea is that a baby born January 1, 2002 has not been exposed to any of the toxic plant activity for calendar 2002, but was exposed to toxic emissions in 9 out of 12 months of 2001. Similarly, a baby born in November 2001 was exposed to toxic emissions for 9 out of 12 months of 2001. This restriction has the additional advantage of limiting the extent to which seasonality in plant activity or birth outcomes affects our findings. The robustness of the results to alternative timing assumptions is explored in the subsequent analysis.

3.5. Data Linkages and Aggregation

We link plants in the TRI and LBD to the housing and vital statistics, based on the latitude and longitude of the plants, houses, and mother's residence. Specifically, we first create a large

---

[20] Black, Devereux, and Salvanes (2007) use twin and sibling fixed effects models on data for all Norwegian births over a long time period to show that birth weight has a significant effect on height and IQ at age 18, earnings, and education. Using U.S. data from California, Currie and Moretti (2007) find that mothers who were low birth weight have less education at the time they give birth and are more likely to live in a high poverty zip code. They are also more likely to have low birth weight children.

[21] These are all outcomes that have been previously examined in the environment-infant health literature (e.g., Chay and Greenstone 2003; Currie, Neidell, and Schmieder, 2009; Currie, Greenstone, and Moretti, 2011; and Currie and Walker, 2011).

dataset consisting of all pairwise combinations of plants and outcome variables (i.e. births and/or housing transactions). We keep outcome and explanatory variables within two miles of a plant. This means that any house or birth observation within two miles of more than one plant will contribute one observation for each plant-outcome pair. For the primary specifications, we collapse the outcome measures into various distance bins surrounding plants in a given year to minimize the computational burden of working with the universe of birth and housing transactions crossed with plants. That is, for each plant-year, we construct the mean of the outcome variable and key covariates for outcomes that occurred within 0 to 0.5, 0.5 to 1.0, 0 to 1.0, and 1.0 to 2.0 miles of a plant. In addition to easing the computational burden, the collapsing of the data accounts for issues pertaining to inference when the identifying variation occurs at a more aggregate level. In supplementary specifications, we analyze subsamples using the underlying microdata.

3.6. Summary Statistics

Panel A of Table 1 presents summary statistics for the sample of plants that form the basis for our analysis. The three columns reflect the sample characteristics for plants that were always open, newly opened, and newly closed within our sample frame respectively. A plant can appear in both columns (2) and (3), and we have about 1,600 total plants that either open or close. In practice, the plants in our sample tend to be long-lived, with a median age of around 17 years.[22] For continuously operating plants, the mean value of plant equipment and structures is $22 million, and mean annual salary and wages is $11.7 million.[23] Mean salary and wages is lower for plants that opened or closed. The table also reports mean annual toxic emissions, which exceeds 17,000 pounds in all three columns. These are the self-reported measures of airborne toxic emissions from the TRI, and are averaged over all non-missing observations (i.e. if a plant does not report to the TRI during a particular year in which we know the plant is operating, we treat this as missing rather than zero).

Panel B of Table 1 describes community characteristics near plants that either opened or closed during our sample period. Statistics are reported separately by distance to the plant and observations are restricted to the 2 years after a plant opening or the 2 years before a plant closing. Note that a house or birth can be close to more than one plant, and so the same house or birth can appear in more than one column. Within columns, we have restricted houses and births so that they appear only once in this panel, implicitly giving equal weight to each birth and housing outcome.

---

[22] Plant age in the LBD is left-censored in 1975 (the first year the plants are observed in the sample). Therefore, the median age of the plants in our sample is likely to be a bit larger.

[23] The capital stock measures come from the Annual Survey of Manufacturers, and are computed using a modified perpetual inventory method (Mohr and Gilbert, 1996). Since the ASM is a sample and oversamples large establishments, these statistics are not available for all plant years and reflect statistics for larger plants.

Both housing values and maternal characteristics tend to improve with distance from the plant. The average housing value is $124,424 within a half mile of a plant compared to $132,227 for houses between one and two miles away. Similarly, average maternal education rises from 11.93 to 12.22 over the same distance. Rather than rely on equality of levels, our difference-in-differences style identification strategy relies on the assumption that trends in the unobserved determinants of the outcomes are evolving equally in the 0-1 (or 0-0.5 and 0.5-1.0) and 1-2 mile distance from the plant categories. The subsequent analysis provides graphical evidence supporting the validity of this assumption.

4. The Transport of Airborne Toxics as the Basis of a Research Design

Our difference-in-differences strategy compares houses and births in areas "near" a toxic plant to those in areas slightly farther away. While this is a simple idea conceptually, there is little guidance in the literature about how near a household must be to a plant for proximity to affect either housing prices or birth outcomes (or alternatively, about how far toxic emissions are transported). Hence we characterize this relationship empirically. This evidence is of significant independent interest and an important contribution of our paper.

Our approach uses data from monitoring stations about ambient levels of hazardous air pollution. While the EPA has been monitoring criteria air pollutants for four decades, they have only recently begun monitoring hazardous air pollutants (HAPs).[24] The first year of data availability was 1998 and monitors have been gradually added over time. As of 2005, the last year of our sample, there were 84 pollutants being monitored across the 5 states we examine. We investigate the ways in which plant operating status maps into local ambient hazardous air pollution in two separate ways. First, we take the 8 most monitored pollutants in our data and examine pollutant-by-pollutant heterogeneity in emissions transport as a function of plant operating status and distance between a plant and a monitor. Second, we combine all pollutants into a single summary measures by standardizing each pollutant to have mean zero and standard deviation of one.[25]

We matched the monitoring station data to our data on toxic plants using latitude and longitude, keeping monitor-plant pairs in which the plant had ever reported releasing the monitored

---

[24] Hazardous air pollutants, also known as toxic air pollutants, are defined by the EPA as "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." In contrast, criteria air pollutants, are the more commonly found air pollutants that are regulated according to the EPA's National Ambient Air Quality Standards (NAAQS), such as particulate matter.
[25] Note that some pollutants are more toxic or hazardous than others. For the purposes of this particular econometric exercise, we are simply trying to understand if *any* detectible relationship exists between toxic plant activity and ambient levels of hazardous air pollutants, irrespective of the toxicity of a given pollutant.

pollutant and in which the monitor was less than 4 miles away from the plant. We then estimate the following linear regression model,

(2) $Poll_{jmt} = \beta_0 + \beta_1 1[Plant\ Operating]_{jt} +$

$(1[Plant\ Operating]_{jt} \cdot Distance_{jm})'\beta_d + \eta_{jm} + \tau_t + \varepsilon_{jmt}$

where the dependent variable is one of the pollution measures described in the previous paragraph for monitor m linked to plant j in year t. The regression includes an indicator variable for whether a plant is operating in a given year, and the interaction between the indicator and a quartic polynomial in the distance between the plant and the monitor.[26, 27] We also include monitor-plant pair fixed effects, η_{jm}, which are collinear with the main effect of the distance polynomial. The inclusion of these fixed effects ensures that identification comes from plant openings and closings. Lastly, we include year fixed effects, τ_t, to control for overall trends in ambient pollution concentrations. The standard errors are two-way clustered on both monitor and plant.

Figure 1 plots the marginal effect of an operating plant on hazardous air pollution as a function of distance from the plant for eight of the most widely monitored pollutants.[28] Each panel of Figure 1 presents the pollutant specific distance gradient, showing how the marginal effect of plant operation fades with distance. Each pollutant has been standardized by subtracting the pollutant-specific mean and dividing by the standard deviation so that the distance gradient may be interpreted as standard deviations from the mean value. Below each graph is a histogram showing the number of monitors in 0.1 mile increments. While there is some heterogeneity across pollutants, plant operation is typically associated with higher pollution levels that fall exponentially with distance from the plant. In most cases, pollution is only detectable within 1 mile of a plant.

Figure 2 plots the standardized pollution measure pooling over all 84 pollutants in our sample. Average levels of ambient hazardous air pollution are one standard deviation higher immediately adjacent to an operating plant, and decline exponentially with distance, reaching zero at roughly one mile from a plant. Most previous analyses of the economic impacts of toxic emissions have used county-level data, making it impossible to measure these highly-localized impacts. An important exception is Banzhaf and Walsh (2008), who use block-level aggregates from the 1990 and

---

[26] We have also examined different functional forms for distance and the results are similar. Models using more flexible distance specifications, such as replacing a continuous distance measure with dummy variables for different distance bins yield similar results, but the models are less precisely estimated.

[27] The LBD provides information on the first year and last year that a plant is observed in the data. We define $1[Plant\ Operating]_{jt} = 1$ if year t is greater than or equal to the first year the plant is observed in the data and less than or equal to the last year the plant is observed in the data.

[28] Specifically, we are plotting B_1 + B_d * distance. Appendix Table A1 presents the table form regression estimates from the 8 separate models.

2000 censuses for urban areas in California to examine localized changes in average household income.

Documenting this relationship between toxic plant activity and ambient levels of hazardous air pollution helps to motivate our empirical specification. There are several ways for an industrial plant to affect housing values and human health including aesthetics, congestion, and noise. Toxic emissions may be among the channels that have the most distant effects, and the evidence suggests that on average emissions do not reach further than one mile.[29] This finding underscores the importance of performing the analyses which follow using spatial data at a high level of resolution. In most analyses below, we define "near" as within 0.5 or 1 mile of a plant and "far" as one to two miles away. That is, houses and households between one and two miles are used as comparison groups. We also present results using alternative distances. As discussed above, the underlying assumption is that the comparison groups are close enough to experiences the wage and productivity effects of the plant. A second assumption is that outcomes in the near and far areas are evolving with similar trends. Under these assumptions, differences in the impact of plant operations reflect the effects of the local disamenities of plant operation.

5. Housing Values

5.1. Housing Values: Empirical Strategy

We begin our investigation of the effects of toxic plants on housing values by fitting the following econometric model:

(3) $Y_{jdt} = \beta_0 + \beta_1 1[Plant\ Operating]_{jt} + \beta_2 1[Near]_{jd} +$

$\beta_3 1[Plant\ Operating]_{jt} * 1[Near]_{jd} + \eta_{jd} + \tau_t + \beta_4 (X1990_{jd} * T_t) + \varepsilon_{jdt}$

where Y_{jdt} denotes the natural log of average housing values near plant site j, within distance group d, in year t. For each plant j, there are two observations per year. In each plant-year, one observation consists of average housing prices "near" a plant (i.e. within 0.5, 0.5 to 1.0, or 1 mile of the plant). The second observation per plant-year consists of average house prices for houses within 1-2 miles of the plant; this second group provides a counterfactual for housing prices near the plant. The availability of these two groups allows for a difference in differences-style estimator.

The variable $1[Plant\ Operating]_{jt}$ is an indicator equal to one if a toxic plant j is operating in year t and zero otherwise. It is equal to 1 for both distance groups associated with a plant. The

---

[29] A recent literature also finds that other forms of housing externalities are very localized (See for example, Linden and Rockoff (2008), Harding, Rosenblatt, and Yao (2009), Rossi-Hansberg, Sarte, and Owens (2010), and Campbell, Giglio, and Pathak (2011)).

indicator $1[Near]_{jd}$ is equal to one for observations from the near category, regardless of whether the plant is currently operating. Equation (3) also includes plant-by-distance fixed effects η_{jd} to control for all time-invariant determinants of house prices in a plant-by-distance group, which in practice is collinear with the indicator $1[Near]_{jd}$. Additional controls include 1990 census tract characteristics, $X1990_{jd}$, interacted with quadratic time-trends T_t.[30]

Equation (3) also includes time fixed effects, τ_t, to flexibly account for trends in housing values over time. We report specifications that include either state-by-year fixed effects to account for state-level trends in housing prices or plant-by-year fixed effects to account for highly localized trends. The former specification requires the estimation of 25 fixed effects, while the latter adds approximately 10,000 fixed effects, one for each plant-year.

The parameter of interest in Equation (3) is β_3, the coefficient on the interaction term: $1[Plant\ Operating]_{jt} * 1[Near]_{jd}$. It captures the differential impact of an open plant on locations "near" the plant, relative to those one to two miles away. Given that our models include plant-by-distance fixed effects, η_{jd}, β_3 is identified by changes in the operating status of a plant (i.e. plant openings and closings). The model with plant-by-year and plant-by-distance fixed effects provides an average of the estimates that would be derived from the roughly 1,600 case studies of plant openings and closing that underlie this analysis. Specifically, β_3 is identified by within-year differences in the change in house prices among houses "near" and 1-2 miles from toxic plant openings and closings.

We also estimate a "repeat-sales" model with individual-level, rather than grouped, data. The advantage of this model is that our housing value data contain few housing characteristics, so the estimates of β_3 from equation (3) may confound willingness to pay to avoid a toxic plant with changes in the composition or type of house sold. To distinguish between these two possibilities we focus on a sample of houses that sold more than once between 1998-2005, allowing us to difference out the unobserved time invariant qualities of a house.

We use several versions of the following first differenced specification,

(4) $\Delta Y_{ijt,t-\alpha} = \beta_1 \Delta 1[Plant\ Operating]_{jt,t-\alpha} + \beta_2 \Delta 1[Near]_{ij} +$

$\beta_3 \Delta (1[Plant\ Operating]_{jt,t-\alpha} * 1[Near]_{ij}) + \Delta\tau_{t,t-\alpha} + \beta_4 \Delta (X1990_{jd} * T_{t,t-\alpha}) + \Delta\varepsilon_{jdt,t-\alpha}$.

where $\Delta Y_{ijt,t-\alpha}$ denotes the difference in ln(house price) between sales of house i, near plant site j, in years t and $t - \alpha$. Notice, that the time between sales varies across houses so α takes different values

---

[30] Census tract characteristics were mapped to plant radii using ArcGIS, where the radius characteristics consist of the area weighted averages of census tracts that intersect the distance circle/radius. Results are similar with and without these controls.

across houses. Since houses are in fixed locations, there is no variation in $\Delta 1[Near]_{ij}$ and it is infeasible to obtain estimates of β_2.

The coefficient of interest remains β_3, which captures the variation in housing prices when there is a change in plant operating status for houses "near" sites, relative to the change in housing prices among houses 1-2 miles from the site. It is important to recognize that β_3 does not compare the operating period to either the period before a plant opened or to the period after it closed. Rather, it compares the operating period to a weighted average of periods before the plant opened and periods after the plant closed that is specific to this sample, so that its external validity may be limited.

Because of these important issues of interpretation, we also estimate an alternative version of equation (4) that allows us to separately identify the effects of plant openings and plant closings. For these models, the variable $1[Plant\ Operating]_{jt}$ is replaced by two separate indicators $1[Plant\ Opened]_{jt}$ and $1[Plant\ Closed]_{jt}$. The variable $1[Plant\ Opened]_{jt}$ is an indicator equal to zero before the plant opens, and equal to one in all years after the plant opens, even if the plant subsequently closed. The variable $1[Plant\ Closed]_{jt}$ is an indicator variable equal to zero before the plant opens and while it is operating, and then equal to one for all years after the plant closes.[31] These indicators are then interacted with $1[Near]_{jd}$.

The result is that the $1[Plant\ Opened]_{jt}$ interaction measures the effect on housing prices in near locations, relative to the 1-2 mile locations, during the period that the plant is operating, relative to the period before it opened. Because of the way that the indicators are defined, the interaction with $1[Plant\ Closed]_{jt}$ tests for an additional effect on housing prices in near locations, relative to 1-2 mile locations, after the plant has closed, relative to the period when it was operating; so, the coefficient associated with this interaction provides a direct test of whether plant closings affect housing prices, relative to the period that the plant was operating. We also report on tests of the hypothesis that the parameters associated with the two interactions are equal and of opposite sign, which would be the case if a plant's closing completely reversed the effect of its opening.

Note that housing values reflect both current and expected future amenities. In our setting, these expectations are likely to include valuations of local air pollution, visual disamenities, traffic-related to plant activity, and soil and water pollution, as well as expectations about how long the plant will operate and whether it will reopen if it closes. These expectations are, of course,

---

[31] Formally, we define $1[Plant\ Closed]_{jt} = 1$ if year *t* is greater than the last year the plant is observed in the LBD and $1[Plant\ Opened]_{jt} = 1$ if year *t* is greater than or equal to the first year the plant is observed in the LBD.

unobservable (see, e.g., Bishop, 2012), but it is nevertheless important to keep in mind that housing values reflect the present discounted value of the entire stream of amenities associated with a particular location when interpreting the estimates.

5.2. Housing Values: Results

We first present event study graphs that motivate the regression analyses which follow. These graphs are derived from the estimation of versions of equation (3) that include plant-by-year fixed effects and allow the coefficients on $1[Plant\ Opened]_{jt} * 1[Near]_{jd}$ and $1[Plant\ Closed]_{jt} * 1[Near]_{jd}$ to vary with event time; here, year zero is the year that the plant's operating status changes (i.e., the year of the plant opening or closing). The figures plot these coefficients and their 95% confidence intervals.[32] They provide an opportunity to judge the validity of the difference in a differences-style approach that is based on the assumption of similar trends in advance of the opening or closing.

Figure 3 plots event study coefficients from two separate regressions. The left panel of Figure 3 plots event study coefficients for years before/after a plant opening, and the right panel plots event time coefficients before/after a plant closing. The plotted coefficients represent the time path of housing values within 0-1 miles from a plant, relative to 1-2 miles from a plant, conditional on plant-by-distance and plant*year fixed effects. Both panels support the validity of the design as there is little evidence of differential trends in housing prices between houses 0-1 and 1-2 miles from the plant in the years preceding the opening or the closing. There is clear evidence that plant openings lead to housing price declines in the year that the plant opens. The plant-closing figure provides less decisive evidence, although on average prices rise slightly after the year of a closing.

Table 2 reports baseline estimates for the effect of toxic plants on housing values. Panel (A) shows least squares estimates from various versions of Equation (3), in each case reporting the coefficient and standard error associated with the interaction of $1[Plant\ Operating]_{jt} * Near_{jd}$. We estimate these models on a balanced panel of plant-by-distance-by-year observations, excluding a subset of plants for which no housing values occurred in a specific distance-by-year cell.[33] Panels (B)

---

[32] The available housing price data only allow for the estimation of the coefficients for event years -3 through +5 for plant openings and -5 through +5 for plant closings since plant openings are concentrated in the earlier part of our sample.

[33] Results using an unbalanced panel are similar. Models estimated using plant-by-year fixed effects are estimated in two steps. The first step demeans all regression model variables by plant-by-year. The second step then estimates the model on the remaining covariates using the demeaned data. Given all the fixed effects in these models, it is not surprising that they explain a lot of the variation in housing prices. The r-squareds are around .7 and .9 for models with and without the repeat sales, respectively.

and (C) report estimates of Equation (4), where Panel (B) reports the coefficient and standard error associated with the interaction of $1[Plant\ Operating]_{jt} * Near_{jd}$, and Panel (C) allows the effects of openings and closings to differ.

In all regressions the comparison group is homes located between one and two miles from the plant, whereas the definition of "near" changes across regressions, as indicated by the column headings. The odd-numbered columns report estimates from specifications that include state-by-year fixed effects and the even-numbered columns report estimates from specifications that use plant-by-year fixed effects (or county by year fixed effects in the repeat sales analysis).[34]

The estimates in columns (1) and (2) of Panel (A) show that an operating toxic plant within a half mile is associated with a 2 to 3% percent decrease in housing values. The point estimates in columns (3) and (4) are smaller in magnitude, suggesting that the effects of plant operations on housing values tend to fade with distance. For example, the point estimate in column 3 suggests that the effect of an operating plant falls to one percent in the half mile to one mile range. The standard errors are large enough, however, that their 95% confidence intervals overlap the 95% confidence intervals of the estimates in columns (1) and (2). Hence, in columns (5) and (6) we compare the entire zero to one mile area with the one to two mile zone.[35] Not surprisingly given the previous estimates, the overall impact on housing values within one mile is about -1.5%.

The last two columns of Table 2 report estimates from specifications that restrict observations to within two years of a change in plant operation. In the short-run, prices will do a better job of capturing the full welfare effects because supply is relatively inelastic over short periods of time; over the longer run, the full welfare effects are captures by adjustments in prices and quantities (which are unobservable in our data). This restriction attenuates the point estimates, but the 95% confidence intervals overlap those associated with the estimates in columns (5) and (6).

Panels (B) and (C) present the repeat sales estimates from fitting equation (4). For the most part, the estimates in Panel (B) are similar to those found in Panel (A), albeit somewhat smaller in absolute magnitude. The differences between the two panels are consistent with the interpretation that some of the estimated impacts in Panel (A) are driven by less expensive houses selling near to a plant whenever a plant is operating. The disparities between the results in Panels (A) and (B) are also consistent with greater attenuation due to measurement error in a first difference setting. However,

---

[34] We ran into computational challenges when estimating the full set of plant-by-year fixed effects in the first difference setting, and thus we rely on county-by-year fixed effects as a compromise. This being said, estimates using Equation (3) with county-by-year or plant-by-year fixed effects are almost identical.
[35] The column (6) specification is the difference-in-differences analogue to the event-time regression plotted in Figure 2.

the 95% confidence intervals overlap across all estimates, and thus we are not able to make strong conclusions about the difference in magnitudes.

Panel (C) presents parameter estimates associated with $1[Plant\ Opened]_{jt} * 1[Near]_{jd}$ and $1[Plant\ Closed]_{jt} * 1[Near]_{jd}$. Within 0.5 miles, a plant's operation is associated with a 10%-11% decline in housing prices; these estimates are economically large and statistically significant. There is little evidence of an effect on housing prices between 0.5 and 1.0 miles from the plant. As Figure 3 foreshadowed, plant closings appear to modestly increase housing prices, but this effect is small economically (less than 2%, even less than 0.5 miles from a plant) and statistically indistinguishable from zero.

The final row reports the results from a test that the opening and closing coefficients are equal and opposite in sign. This null hypothesis can be rejected in the 0-0.5 mile range. One possible interpretation is that households expect closed plants to reopen. However, we measure closings using the last year that a plant is observed in the LBD. Consequently, our data generally pick up permanent (not temporary) plant closures, though home buyers and sellers may not realize this at the time of the closure.[36]

Thus far we have concentrated on the average effect of plant openings and closings. We next explore heterogeneity in our baseline estimates by stratifying plants by observable characteristics. Since the housing price impacts are almost entirely concentrated within 0.5 miles of a plant, we focus on housing values within this range.

We group plants into whether the median value of a particular variable (taken over all years of plant operation) is above or below the population median (taken over the plant-level medians). The plant characteristics we explore are plant employment, payroll, stack emissions, fugitive emissions, and total emissions, as well as the mean and maximum toxicity of the chemicals that are released. Plants in the TRI report both stack and fugitive emissions. Stack emissions occur during the normal course of plant operations, and are emitted via a smoke stack or some other form of venting equipment which is, in many cases, fitted with pollution abatement equipment. Because stacks are often extremely high, these emissions tend to be dispersed over a wide geographic area. Fugitive

---

[36] We also tested whether plant openings and closings affect the volume of housing transactions. We used the baseline housing regression approach (aggregated at the plant-distance-year level), but replaced mean log(sales price) with the number of houses sold (in logs). While the housing price regressions weight cells by the number of houses sold, we excluded regression weights from this volume regression so as to not weight observations by the outcome variable. The results suggest that the number of transactions decreases when there is an operating toxic plant nearby, especially within 0.5 miles after plants open. It is difficult to draw definitive conclusions, however, because most of the estimates are not statistically significant.

emissions are those that escape from a plant unexpectedly, generally without being treated. These emissions may be more likely to be manifest to households in the form of noxious odors or residues. The toxicity measures were calculated using the EPA's Risk-Screening Environmental Indicators.[37] We also stratify plants based on the characteristics of the nearby communities (i.e. within 2 miles), including the fraction of the population that is college educated, the fraction of the population that is Caucasian, the median housing value surrounding a plant, and median income.

Table 3 reports the results of this exploration. We focus on the baseline first-differences specification, augmenting Equation (4) to include an additional interaction term for whether or not a plant is above the median for each of the above listed characteristics. We then estimate the full three-way interaction, allowing for all lower order interaction terms. The estimates indicate that the housing results are fairly homogenous across various plant types (columns (1)-(7)) but that the negative impacts appear to be concentrated in relatively disadvantaged communities (columns (8)-(11)). If households were aware of the toxicity measures and they were valued (negatively) by households, then one might have expected to see relative toxicity reflected in housing price differentials. A possible explanation for the absence of such a pattern is that households have imperfect information. Given the lack of scientific evidence about the health effects of exposure, such ignorance would not be surprising.

The online appendix presents estimates from several additional specifications. Appendix Table A2 examines the sensitivity of the baseline estimates to varying sets of controls. The qualitative findings are unchanged across several different approaches. Appendix Table A3 presents estimates of Equation (3) that use a comparison group of two to four miles from a plant instead of one to two miles, and the results are similar to the baseline results in Table 2. This is reassuring because it suggests that the results are not driven by patterns in housing prices in the one to two mile zone. Appendix Table A4 presents regressions identical to the baseline estimates of Equation (3) except that each regression is estimated using only observations from a single distance bandwidth (e.g. 0 to 0.5 miles, 0.5 to 1 miles, 1 to 1.5 miles, 1.5 to 2.0 miles, etc…) for each plant. Identification in these models comes from differential timing of openings and closings across plants. Estimates from this specification corroborate our baseline findings and choice of comparison group; the effects of plant operating status are highly localized, and there seems to be little negative effect of plant openings in areas more than one mile away from a plant.

---

[37] Surprisingly little is known about the relative toxicity of different chemicals. Although animal testing is broadly used for evaluating the toxicity of chemical compounds, these studies are of limited relevance for evaluating which chemicals are likely to be most damaging for human health.

6. Infant Health

6.1. Infant Health: Empirical Strategy

The empirical strategy for examining infant health outcomes is very similar to the approach used for housing values. Again, our main focus is on comparing outcomes "near" a plant with outcomes one to two miles away. We estimate models of the form:

(5) $\quad Z_{jdt} = \alpha_0 + \alpha_1 1[Plant\ Operating]_{jt} + \alpha_2 1[Near]_{jd} +$
$\qquad \alpha_3 1[Plant\ Operating]_{jt} * 1[Near]_{jd} + \eta_{jd} + \tau_t + \beta_4 X1990_{jd} * T_t + \epsilon_{jdt}$

where Z_{jdt} denotes the average incidence of low birthweight or another measure of infant health near plant site j, within distance group d, in year t. As before, the specification includes plant-by-distance fixed effects, η_{jd}, year fixed effects τ_t (which in practice are state-by-year or plant-by-year fixed effects), and census controls, $X1990_{jd}$, interacted with quadratic time-trends T_t.

As in the housing equations, the coefficient of interest, now denoted α_3, is the differential impact of an operating plant within one mile. We again explore a version of this specification that replaces the $1[Plant\ Operating]_{jt}$ variable with the $1[Plant\ Opened]_{jt}$ and $1[Plant\ Closed]_{jt}$ variables. For this richer specification, we again test whether the coefficients on the interactions of these variables with $1[Near]_{jd}$ are equal and opposite in sign. If air toxic emissions are the channel for any infant health effects, then the plausibility of this null is stronger than in the housing price regressions where plant closings may be perceived as temporary and visual disamenities could remain after a closure.

The vital statistics data include a rich set of mother's characteristics that can be used to control for possible changes in the composition of mothers. However, the identifying variation in our models comes at a much higher level of aggregation; hence, in order to avoid overstating the precision of our estimates and to limit the computational burden of our most stringent specifications we control for mother's characteristics using a two-step, group-level estimator (Baker and Fortin, 2001; Donald and Lang, 2007). In the first step, we estimate the relationship between low birth weight (Z_{jdt}) and plant-by-distance by year indicators (g_{jdt}), after controlling for mother's characteristics (m_{it}):

(6) $\qquad\qquad\qquad Z_{jdt} = m'_{it}\theta + g_{jdt} + \xi_{jdt}.$

The vector m_{it} controls for maternal characteristics including indicators for: age categories (19-24, 25-34, and 35+), education categories (<12, high school, some college, and college or more), race

(African American or Hispanic), smoking during pregnancy, month of birth, birth order, and gender of child.[38]

The estimated $\widehat{g_{jdt}}$ provides group-level, residualized averages of each specific birth outcome after controlling for the observable characteristics of the mother. These averages are used as the dependent variable in equation (5), instead of Z_{jdt}. In this second step, the equation is weighted by the group-level cell size.[39,40]

6.2. Infant Health: Results

We start by presenting event study graphs for the incidence of low birth weight (i.e., an infant born weighing less than 5.5 pounds or 2500 grams) based on a version of equation (5). The plotted estimates and 95th percentile confidence intervals correspond to the interaction of event-time indicators with $1[Plant\ Opened]_{jt} * 1[Near]_{jd}$ and $1[Plant\ Closed]_{jt} * 1[Near]_{jd}$. The specification includes plant-by-distance and plant-by-year fixed effects, as well as the census controls interacted with a quadratic time trend. The birth data cover a longer period than the housing prices data and we can estimate the parameters of interest for all event years from five years before an opening/closing through five years after an opening/closing.

Figure 4 suggests that operating plants raise the incidence of low birth weight. There is little evidence of differential trends in the adjusted incidence of low birth weight between mothers living 0-1 and 1-2 miles away during the years leading up to plant openings or closings, which supports the validity of the design. After plant openings, there is a relative increase in the incidence of low birth weight among mothers living within one mile of a plant. After plant closings, there is some evidence of an opposite effect. Specifically, the incidence of low birth weight within one mile decreases modestly relative to what is observed between one and two miles although the decline is less sharp than in the plant opening panel.

---

[38] For a small number of observations there is missing data for one or more of these control variables and we include indicator variables for missing data for each variable.
[39] To limit the computational burden of estimating the first stage of the full sample, the first stage is estimated separately by state. Alternative group-level weights include the inverse of the sampling error on the estimated fixed effects, but since we are estimating state by state, the estimated standard errors are likely to be inefficient (although the group level estimates are still consistent) making this weighting mechanism less attractive. Donald and Lang (2007) present an alternative feasible GLS specification where the weights come from the group level residual and the variance of the group effect. Since all of these weights are proportional and highly correlated, the choice of weights has little effect on the results. We follow Angrist and Lavy (2009), who weight by the group cell size. These models have r-squareds of about .3.
[40] We obtain similar results from group-level models that convert micro-level covariates into indicator variables and take means within cells.

Table 4 presents regression estimates, and is structured similarly to Panels (B) and (C) of Table 2 which reports the housing price results. We focus on the Panel B results, which have a clearer counterfactual and greater external validity. Further, due to the finding that toxic air emissions travel roughly 1 mile on average, we concentrate on the 0-1 mile results.

The final four columns suggest that an operating toxic plant increases the incidence of low birth weight by 0.0024 – 0.0037 percentage points or 3.3% - 5.1%. The effects among infants born to mothers in the 0-0.5 mile and 0.5-1 mile ranges are nearly identical. It is also interesting that the larger estimates come from the restricted sample that only includes births within 2 years of a change in operating status.

The results are less conclusive on the question of whether a plant closing reverses the negative effects of a plant's operation on the incidence of low birth weight. On the one hand, all of the point estimates suggest that low birth weight declines after a plant closing. This decline, however, is only statistically significant at the 95% level of confidence in column (8), though this specification is perhaps the most reliable one. The null that the coefficients are equal and of opposite sign cannot be rejected in any of the specifications.

We probed the robustness of these results in several ways. The results are qualitatively similar when we vary the set of controls used in our baseline regressions (see Appendix Table A5), and when we use a comparison group of births that occur two to four miles from a plant, rather than one to two miles (see Appendix Table A3). The results are also similar when we estimate the regressions separately by distance group (see Appendix Table A4). These alternate specifications corroborate the main results, again indicating that the effects of plant operating status are highly localized, and providing additional empirical support for the choice of comparison group.

We also tested for changes in the composition of mothers giving birth in Appendix Table A6. Documenting this type of compositional change is of significant independent interest (see for example, Cameron and McConnaha, 2006; Banzhaf and Walsh, 2008; and Currie, 2011). Overall, impacts of plant openings and closings on mothers' characteristics are small and generally statistically insignificant, suggesting that the low birth weight estimates are not driven by changes in the composition of mothers who live near plants. If anything, toxic plants appear to be associated with a small *increase* in the socio-economic status of mothers; if the regressions fail to adequately adjust for these changes, then the measured health effects may modestly understate the true effects.

When assigning plant events to birth outcomes, there is some ambiguity as to whether the plant event occurred before or after a birth because we observe plant operating status just once a year in the LBD. In Appendix Table A7 we investigate the sensitivity of our results to alternative

approaches to timing. First, we created a dataset of plant-month-year observations and then calculated the expected exposure of a birth to a plant opening or closing, assuming that plant openings/closings are uniformly distributed throughout the year. This assumption allows us to calculate the expected fraction of in utero exposure to a plant opening/closing based on month of birth. We use this expected fraction as a continuous measure of treatment for the birth-months immediately surrounding the plant. The exact details of this fractional assignment model are more thoroughly discussed in the online appendix. We also explored models that exclude ambiguous birth months altogether and simply look at pre-post outcomes for births in which the pregnancy was clearly on either side of openings and closings. This specification has the added benefit of being somewhat agnostic about the prenatal production technology, whereas the continuous assignment mechanism above imposes additive linearity across months in utero. Estimates from these alternative specifications are largely consistent with our baseline findings.

Table 5 examines plant heterogeneity, stratifying plants as was done in the housing regressions (i.e. Table 3) using the version of equation (5) that includes plant-by-year fixed effects. There is little evidence of heterogeneity across these cuts of the data, except that there are no effects on low birth weight in areas with above median housing values. It is possible that richer households are better able to take compensatory measures to protect themselves.

6.3. Alternative Measures of Infant Health: Results

This section presents estimates for alternative measures of infant health. We begin by examining the influence of toxic plant activity on the birthweight distribution. We create nine separate indicators for births falling within independent 500 gram birthweight intervals, and aggregate these outcomes to the plant-by-distance bin by year level. We use these binned averages as the dependent variable when estimating nine different versions of equation (5); the resulting estimates on of the parameter associated with $1[Plant\ Operating]_{jt} * 1[Near]_{jd}$ is plotted in Figure 5. Each bar corresponds to results from a separate regression, comparing birth outcomes for mothers less than one mile from a plant to those of mothers living one to two miles from a plant, so that these models are comparable to those presented in columns (5) and (6) of Table 4. Figure 5 suggests that when a plant is operating the birthweight distribution is skewed to the left, increasing the likelihood of births below 2500 grams. Appendix Table A8 reports on the regression results that underlie this figures, as well as results that replace the $1[Plant\ Operating]_{jt}$ variable with the $1[Plant\ Opened]_{jt}$ and $1[Plant\ Closed]_{jt}$ variables.

Table 6 reports estimates of equation (5) using additional measures of infant health as the dependent variables. These estimates support the hypothesis that toxic plants damage infant health; birth weight decreases and the incidence of prematurity increases. The other birth outcomes are not individually statistically different from zero although this is perhaps unsurprising given that many of these outcomes, such as the incidence of very low birthweight (i.e., an infant born weighing less than 3.3 pounds or 1500 grams) and infant deaths, are an order of magnitude rarer than low birth weight and prematurity.

In light of this issue of precision, the last two columns show models using a summary index measure of infant health as the dependent variable. We first convert each birth outcome measure so that they all move in the same direction (i.e. an increase is undesirable) and then subtract the mean and divide by the standard deviation of each outcome. We construct our summary measure by taking the mean over the standardized outcomes, weighting by the inverse covariance matrix of the transformed outcomes in order to ensure that outcomes that are highly correlated with each other receive less weight than those that are uncorrelated, and thus represent new information, receive more weight (Hochberg, 1988; Kling, Liebman, and Katz, 2007; Anderson 2008).[41] An operating plant has a small but statistically significant positive effect on the index, increasing the probability of a bad health outcome by 0.016-0.017 standard deviations.

7. Interpretation

The estimates in Table 2 indicate that the opening of a toxic plant reduces housing values by roughly 11% within 0.5 miles of an operating toxic plant and this effect appears to persist even after the plant ceases operations. As with all of our estimates, this effect is measured relative to homes 1 to 2 miles away. Since the mean housing value within 0.5 miles of a plant is $125,927, this decrement corresponds to about $14,000 for the average house. In our sample, the value of the housing stock within 0.5 miles of a toxic plant is $38.5 million. Multiplying this figure by 11 percent yields a decline in local housing values of about $4.25 million per plant. Although non-negligible, these housing prices changes are small compared to the capital cost of new industrial plants; for example, a typical natural gas power plant (620MW) costs about $570 million to build.[42]

---

[41] Alternatively, we have created summary index measures that weight each outcome variable equally, as in Kling, Liebman, and Katz (2007), with little appreciable effect on our results.

[42] U.S. Department of Energy (DOE), Energy Information Administration. 2013. *Updated Capital Cost Estimates for Utility Scale Electricity Generation Plants.* April. U.S. Department of Energy. http://www.eia.gov/forecasts/capitalcost/

However, this is certainly an incomplete measure of these plants' welfare consequences. For example, it misses the effects of increased emissions of criteria pollutants and any impacts on non-residential property (which could even be positive if there are spillovers in production efficiency)[43] Moreover under our imposed assumption that the economic benefits of plant production accrue equally to homes within two miles from the plant, this estimate reflects an upper bound on the net costs associated with toxic plants. As we have emphasized throughout, these plants have positive as well as negative externalities, bringing jobs to local communities and potentially raising wages and housing prices over a wide area.

An appealing feature of the analysis is that it provides estimates of the effect of toxic plant openings on *both* housing prices and on an important health outcome. The estimates from the housing value analysis can hence be compared with a valuation of the low birth weight impacts. The point estimate in Table 4, column (6) implies that an operating toxic plant within one mile reduces the incidence of low birth weight by 0.0024 percentage points or 3.1%. There is an average of 67 births within one mile of each toxic plant per year. Thus, the estimate implies that there are approximately 0.16 additional low birth weight births per toxic plant per year. Using estimates in the literature, this corresponds to about $5,600 in decreased lifetime earnings per toxic plant per year.[44] Of course, this calculation requires several assumptions, so it should be interpreted cautiously. While there may well be other health consequences of exposure to toxic plants, this measure is small compared to the estimated value of losses in the housing market. Further, the finding that housing prices remain depressed after the plant has closed and air toxic emissions have ceased suggests that willingness to pay is comprised of more than health effects in this setting.

8. Conclusion

Toxic emissions are widely believed to cause birth defects, cancer, and other severe health impacts, yet there is little evidence about their effects on humans. Governments have only recently

---

[43] The $4.5 million welfare figure does not capture changes to the value of industrial, commercial, or undeveloped property. While some industrial uses may not be substantially affected by toxic plant proximity, commercial property and, perhaps more importantly, the price of undeveloped land may be affected. Moreover, it is important to emphasize that this measure reflects only the highly-localized externalities from industrial plants. Industrial plants also emit criteria pollutants, such as particulates, nitrogen oxides, and sulfur dioxide, which impose costs over a much broader geographic area and are therefore not captured in these estimates.

[44] Black, Devereux, and Salvanes (2007) estimate that each 1% decrease in birth weight decreases expected earnings by about 0.13%. Based on our analysis of the distribution of birth weight, the impact appears to be more births 1000-2000 grams, compared to about 3,200 grams for the average birth, for a back-of-the-envelope average reduction of about 50%. So a low birth weight birth would be associated with 6.5% lower lifetime earnings. Isen, Rossin-Slater, and Walker (2013) calculate that the mean present value of lifetime earnings at age zero in the U.S. population is $542,000 (2000$) using a real discount rate of 3% (i.e., a 5% discount rate with 2% wage growth), so this is equivalent to $35,320 per low birth weight birth.

begun to regulate these emissions. In many respects, this state of affairs resembles the situation that prevailed more than four decades ago when the Clean Air Act compelled the EPA to begin to regulate airborne particulate matter and other criteria air pollutants. This paper represents a first step towards understanding the local external effects of toxic plant production on the health and well-being of local residents.

The application of a research design based on more than 1,600 plant openings and closings matched to extraordinarily detailed, geo-coded data yields three primary findings. First, on average, toxic air pollutants affect ambient air quality only within 1 mile of the plants, suggesting that health effects from these emissions should be concentrated in this range. The highly localized range differs substantially from particulate matter emissions which can affect ambient air quality several hundred miles away from their source. Second, the opening of a plant that emits these pollutants leads to a roughly 11% decline in housing prices within 0.5 miles, or a loss of about $4.25 million. Housing prices are largely unaffected by a plant closing, implying that toxic plants continue to negatively affect housing prices after they cease operations. Third, the incidence of low birth weight increases by roughly 3% within one mile of an operating toxic plant, with comparable magnitudes between 0 and 0.5 miles and 0.5 and 1 miles.

These results underscore opportunities for further research in several areas. We interpret the estimated effects of low birth weight to be a rejection of the null hypothesis that there are no health effects from toxic air emissions. This finding opens the door to seeking creative approaches to testing for longer run health effects on children and adults. It is also possible that toxic air emissions cause households to engage in costly behaviors to protect themselves and documenting these costs would be a contribution (see e.g. Deschenes, Greenstone, and Shapiro 2012).

This paper also raises broader questions around the determinants of housing prices. As computing power increases and more detailed data are accessible, it will be possible to assess the degree to which housing markets fully capture the present discounted value of all present and expected future amenities associated with a particular location. A related and important question is the degree to which health effects are capitalized into housing prices. Finally, we believe that a better understanding of belief formation around local amenities and how these beliefs interact with willingness to pay in the context of local housing markets is a critical area for future research.

References

Agarwal, N, C. Banternghansa, and L. Bui (2010): "Toxic Exposure in America: Estimating Fetal and Infant Health Outcomes from 14 Years of TRI Reporting," *Journal of Health Economics*, 29(4), 2010: 557-74.

Anderson, M. (2008): "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects," *Journal of the American Statistical Association*, 103(484), 1481–1495.

Angrist, J., and V. Lavy (2009): "The Effects of High Stakes High School Achievement Awards: Evidence from a Randomized Trial," *American Economic Review*, 99(4), 1384–1414.

Baker, M., and N. Fortin (2001): "Occupational Gender Composition and Wages in Canada, 1987–1988," *Canadian Journal of Economics*, 34(2), 345–376.

Banzhaf, H. S., and R. Walsh (2008): "Do People Vote with their Feet?: An Empirical Test of Environmental Gentrification," *American Economic Review*, 98(3), 843–863.

Bartik, T. (1987): "Estimating Hedonic Demand Parameters with Single Market Data: The Problems Caused by Unobserved Tastes," *Review of Economics and Statistics*, 69(1), 178–180.

Bennear, Lori. (2008): "What Do We Really Know: The Effect Of Reporting Thresholds On Inference Using Environmental Right-To-Know Data." *Regulation and Governance,* 2(3), 293-315.

Bishop, Kelly C. "A Dynamic Model of Location Choice and Hedonic Valuation," Working Paper, 2012.

Black, S., P. Devereux, and K. Salvanes (2007): "From the Cradle to the Labor Market? The Effect of Birth Weight on Adult Outcomes," *Quarterly Journal of Economics*, 122(1), 409–439.

Blomquist, G.C. (1974): "The Effect of Electric Utility Power Plant Location on Area Property Value," *Land Economics,* 50: 97-100.

Bui, Linda and Christopher Mayer (2003): "Regulation and Capitalization of Environmental Amenities: Evidence From the Toxic Release Inventory in Massachusetts." *Review of Economics and Statistics*, 85(3), 693-708.

Busso, M., J. Gregory, and P. Kline (2013): "Assessing the Incidence and Efficiency of a Prominent Place Based Policy," *American Economic Review*, 103(2): 897-947.

Cameron, Trudy Ann and Ian T. McConnaha (2006): "Evidence of Environmental Migration." *Land Economics*, 82(2):273-290.

Campbell, J., S. Giglio, and P. Pathak (2011): "Forced Sales and House Prices." *American Economic Review*, 101(5): 2109-2131.

Chay, K. Y., and M. Greenstone (2003): "The Impact of Air Pollution on Infant Mortality: Evidence from Geographic Variation in Pollution Shocks Induced by a Recession," *Quarterly Journal of*

Economics, 118(3), 1121–1167.

Currie, J. (2011): "Inequality at Birth: Some Causes and Consequences," *American Economic Review: Papers and Proceedings*, 101(3), 1-22.

Currie, J., M. Greenstone, and E. Moretti (2011): "Superfund Cleanups and Infant Health," *American Economic Review: Papers and Proceedings*, 101(3), 435-41.

Currie, J.M. and E. Moretti (2007): "Biology as Destiny? Short and Long-Run Determinants of Intergenerational Transmission of Birth Weight," *The Journal of Labor Economics*, 25(2), April, 231-264,

Currie, J., M. Neidell, and J. F. Schmieder (2009): "Air Pollution and Infant Health: Lessons from New Jersey," *Journal of Health Economics*, 28(3), 688–703.

Currie, J., and R. Walker (2011): "Traffic Congestion and Infant Health: Evidence from E-ZPass," *American Economic Journal: Applied Economics*, 3(1), 65-90.

Davis, L. W. (2011): "The Effect of Power Plants on Local Housing Values and Rents," *Review of Economics and Statistics*, 93(4), 1391-1402.

Davis, S. J., J. C. Haltiwanger, and S. Schuh (1998): Job Creation and Destruction. MIT Press.

Davis, S.J., R. Jason Faberman, John Haltiwanger, Ron Jarmin, Javier Miranda (2010): "Business Volatility, Job Destruction, and Unemployment," *American Economic Journal: Macroeconomics*, 2(2), 259-287.

Deschenes, Olivier, Michael Greenstone, and Joseph Shapiro (2012): "Defensive Investments and the Demand for Air Quality: Evidence from the NOx Budget Program and Ozone Reductions," Mimeo.

Donald, S., and K. Lang (2007): "Inference with Difference-in-Differences and Other Panel Data," *Review of Economics and Statistics*, 89(2), 221–233.

Greenstone, M., and J. Gallagher (2008): "Does Hazardous Waste Matter? Evidence from the Housing Market and the Superfund Program," *Quarterly Journal of Economics*, 123(3), 951–1003.

Greenstone, M., R. Hornbeck, and E. Moretti (2010): "Identifying Agglomeration Spillovers: Evidence from Million Dollar Plants," *Journal of Political Economy*, 118, 536–98.

Harding, J., E. Rosenblatt, and V. Yao (2009): "The Contagion Effect of Foreclosed Properties," *Journal of Urban Economics*, 66(3), 164–178.

Hochberg, Y. (1988): "A Sharper Bonferroni Procedure for Multiple Tests of Significance," *Biometrika*, 75(4), 800–802.

Jarmin, R. S., and J. Miranda (2002): "The Longitudinal Business Database," Discussion paper, U.S. Census Bureau.

Kline, P. (2010): "Place Based Policies, Heterogeneity, and Agglomeration," *American Economic Review: Papers and Proceedings*, 100(2), 383–387.

Kling, J. R., J. B. Liebman, and L. F. Katz (2007): "Experimental Analysis of Neighborhood Effects," *Econometrica*, 75(1), 83–119.

Kiel, K.A. and K.T. McClain. (1995): "House Prices During Siting Decision Stages: The Case of an Incinerator From Rumor Through Operation" *The Journal of Environmental Economics and Management*, 28(2): 241-255.

Koehler, D. A., and J. D. Spengler (2007): "The Toxics Release Inventory: Fact or Fiction? A Case Study of the Primary Aluminum Industry," *Journal of Environmental Management*, 85(2), 296–307.

Kuminoff, N., V. Smith, and C. Timmins (2013): "The New Economics of Equilibrium Sorting and Policy Evaluation Using Housing Markets," *Journal of Economic Literature*, 51(4), 1007-1062.

Linden, Leigh and Jonah Rockoff (2008): "Estimates of the Impact of Crime Risk on Property Values from Megan's Laws," *American Economic Review*, 98(3), 1103-1127.

de Marchi, S., and J. Hamilton (2006): "Assessing the Accuracy of Self-Reported Data: An Evaluation of the Toxics Release Inventory," *Journal of Risk and Uncertainty*, 32(1), 57–76.

Mohr, Michael F. and Charles E. Gilbert. (1996): "Capital Stock Estimates for Manufacturing Industries: Methods and Data," Board of Governors of the Federal Reserve System.

Moretti, E. (2011): "Local Labor Markets," Handbook of Labor Economics, 4, 1237–1313.

Nelson, J. P. (1981): "Three Mile Island and Residential Property Values: Empirical Analysis and Policy Implications," *Land Economics* 57:3, 363-372.

Palmquist, R. (1992): "Valuing Localized Externalities," *Journal of Urban Economics*, 31(1), 59–68.

Pope, A. C. (1989): "Respiratory Disease Associated with Community Air Pollution and a Steel Mill, Utah Valley," *American Journal of Public Health*, 79(5), 623–628.

Pope, A. C., J. Schwartz, and M. R. Ransom (1992): "Daily Mortality and PM10 Pollution in Utah Valley," *Archives of Environmental Health*, 47(3), 211–217.

Pope, Jaren C. 2008a. "Do Seller Disclosures Affect Property Values? Buyer Information and the Hedonic Model," *Land Economics*, 84(4), 551-572.

Pope, Jaren C. 2008b. "Buyer Information and the Hedonic: The Impact of a Seller Disclosure on the Implicit Price for Airport Noise," *Journal of Urban Economics*, 63, 498-516.

Ransom, M. R., and A. C. Pope (1992): "Elementary School Absences and PM10 Pollution in Utah Valley," *Environmental Research*, 58(1-2), 204–219.

Roback, J. (1982): "Wages, Rents, and the Quality of Life," *Journal of Political Economy*, 90(6), 1257–1278.

Rosen, S. (1974): "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," *Journal of Political Economy*, 82(1), 34-55 .

Rossi-Hansberg, E., P. Sarte, and R. Owens III (2010): "Housing Externalities," *Journal of Political Economy*, 118(3), 485-535.

U.S. Centers for Disease Control and Prevention (2009): "Fourth National Report on Human Exposure to Environmental Chemicals," Discussion paper, Department of Health and Human Services.

U.S. Department of Health and Human Services (2010): "Reducing Environmental Cancer Risk: What We Can Do Now," National Institute of Health: National Cancer Institute.

U.S. Environmental Protection Agency (2007): "Toxics Release Inventory Public Data Release Report," Discussion paper, Office of Pollution Prevention and Toxics.

U.S. Environmental Protection Agency (2009): "TSCA Statute, Regulations, and Enforcement," Discussion paper, Office of Compliance and Enforcement.

U.S. Environmental Protection Agency (2011a): "National Emission Standards for Hazardous Air Pollutants," Discussion paper, Office of Pollution Prevention and Toxics.

U.S. Environmental Protection Agency (2011b): "The Toxics Release Inventory and Factors to Consider When Using TRI Data," Discussion paper, Office of Pollution Prevention and Toxics.

Walker, Reed (2013): "The Transitional Costs of Sectoral Reallocation: Evidence from the Clean Air Act and the Workforce," *Quarterly Journal of Economics*, 128(4), 1787-1835.

1 Figures and Tables for Paper

Figure 1: The Effect of Toxic Plants on Ambient Hazardous Air Pollution
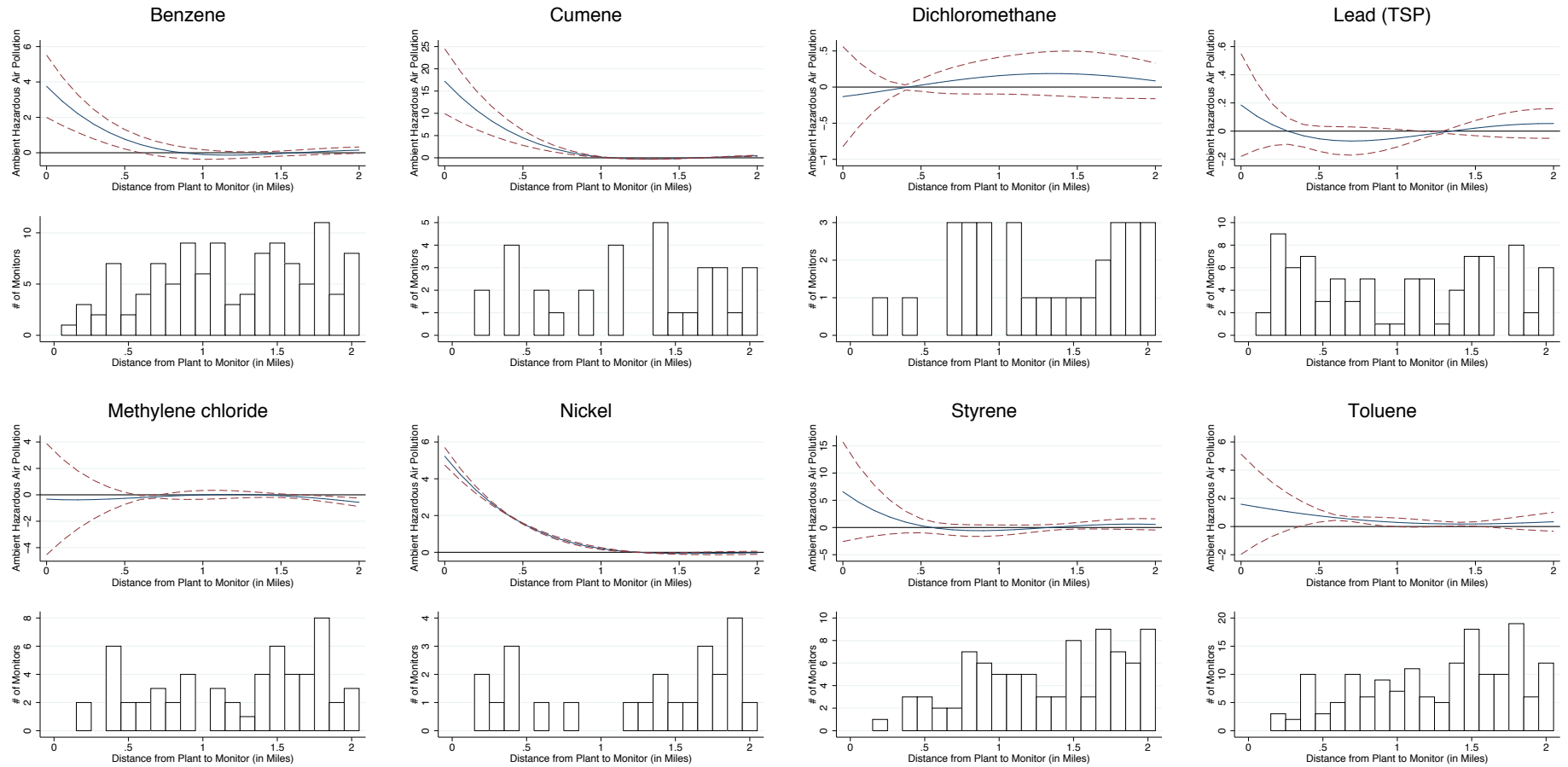
NOTE: This figure plots marginal effects and 95th percentile confidence intervals from 8 separate regressions of a single form of ambient hazardous pollution on a quartic in distance to the nearest operating toxic plant. The unit of observation is the monitor-plant pair and all regressions include monitor-plant fixed effects so the distance gradient is identified using plant openings and closings. In the regression sample, each pollutant has been standardized to be mean 0 and standard deviation 1. The distance gradient can therefore be interpreted as standard deviations from the mean value. Standard errors for the regression are two-way clustered on plant and monitor, and the pointwise standard errors in the figure are calculated using the delta method. Below each pollutant specific graph is a histogram, representing the number of monitors at various distance bins from the plants in the sample.

Figure 2: The Effect of Toxic Plants on Ambient Hazardous Air Pollution, All Pollutants
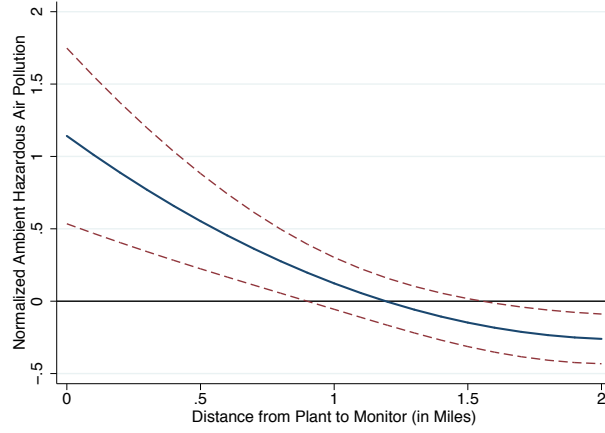

NOTE: This figure plots marginal effects and 95th percentile confidence intervals from a regression of ambient hazardous pollution on a quartic in distance to the nearest operating toxic plant. The unit of observation is the monitor-plant pair and the regression includes monitor-plant fixed effects so the distance gradient is identified using plant openings and closings. In the regression sample, pollutants are pooled, standardizing each pollutant to be mean 0 and standard deviation 1. The distance gradient can therefore be interpreted as standard deviations from the mean value. Standard errors for the regression are two-way clustered on plant and monitor, and the pointwise standard errors in the figure are calculated using the delta method.

Figure 3: Event Study: The Effect of Toxic Plant Openings and Closings on Local Housing Values
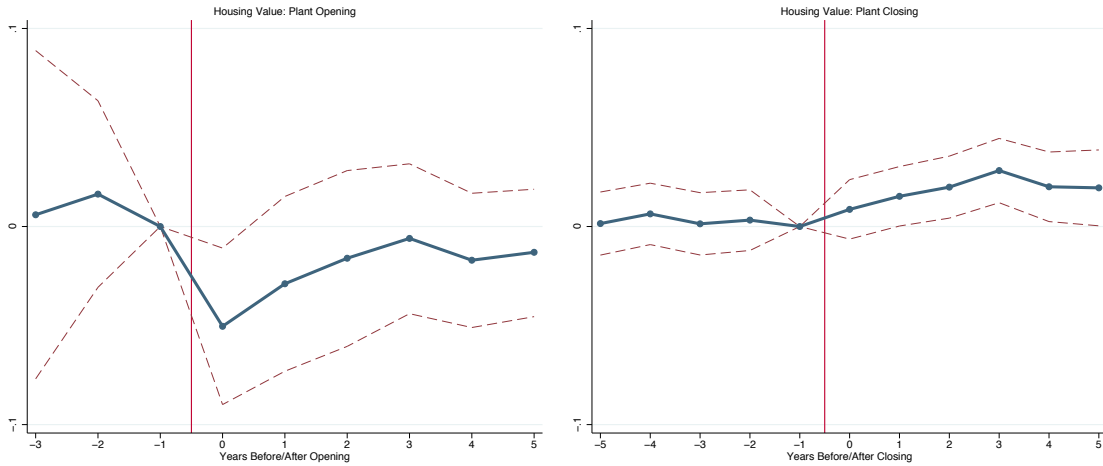
NOTE: These are event study plots created by regressing log housing sale price for a plant by distance by year cell on a full set of event time indicators interacted with an indicator for "near", plant by distance fixed effects, plant by year fixed effects, and census controls (interacted with quadratic trends), weighting by the group-level cell size. Reported are the coefficients for event-time, which plot the time path of housing values "near" relative to "far" before and after a plant opening or closing. "Near" is defined as less than 0-1 miles between a plant and a house, and "far" is defined as 1-2 miles between a house and plant. The dashed lines represent 95% confidence intervals, where standard errors are computed using two-way cluster-robust standard errors, clustering on plant and year. Time is normalized relative to the year that the plant's operating status changes ($\tau = 0$), and the coefficients are normalized to zero in the year prior to a change in operating status ($\tau = -1$). The coefficients corresponding to four or more years before a plant opening are not identified due to the lack of openings in the second half of our sample period and the lack of housing data prior to 1998.

Figure 4: Event Study: The Effect of Toxic Plant Openings and Closings on the Incidence of Low Birthweight
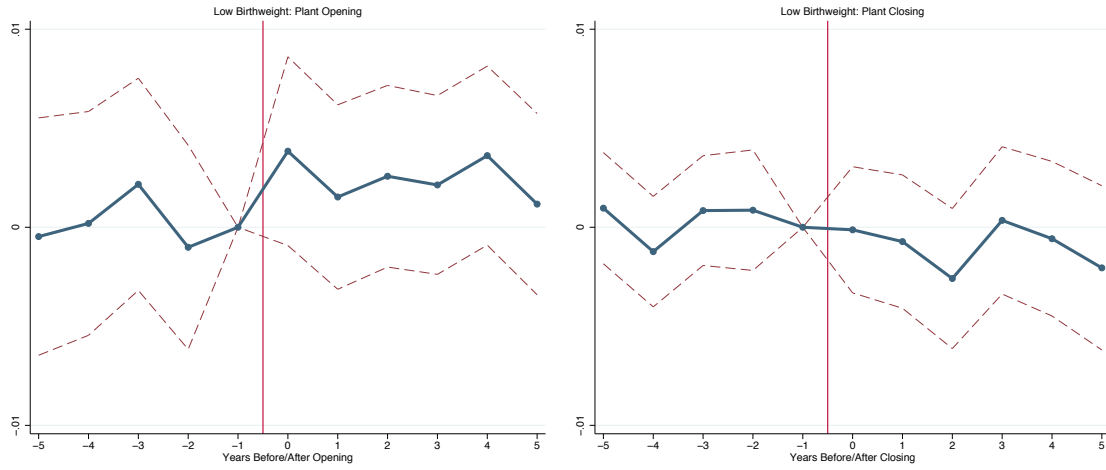
NOTE: These are event study plots created by regressing the incidence of low birthweight for a plant by distance by year cell on a full set of event time indicators interacted with an indicator for "near", plant by distance fixed effects, plant by year fixed effects, and census controls (interacted with quadratic trends), weighting by the group-level cell size. The dependent variable in the regression is the residualized mean incidence of low birthweight for a plant by distance by year, adjusted for micro-level covariates in a first stage. Reported are the coefficients for event-time, which plot the time path of low birthweight "near" relative to "far" before and after a plant opening or closing. "Near" is defined as less than 1 miles between a plant and a house, and "far" is defined as 1-2 miles between a house and plant. The dashed lines represent 95% confidence intervals, where standard errors are computed using two-way cluster-robust standard errors, clustering on plant and year. Time is normalized relative to the year that the plant's operating status changes ($\tau = 0$), and the coefficients are normalized to zero in the year prior to a change in operating status ($\tau = -1$).

Figure 5: Effect of Plant Operation on the Distribution of Birthweight 0-1 Miles from a Plant
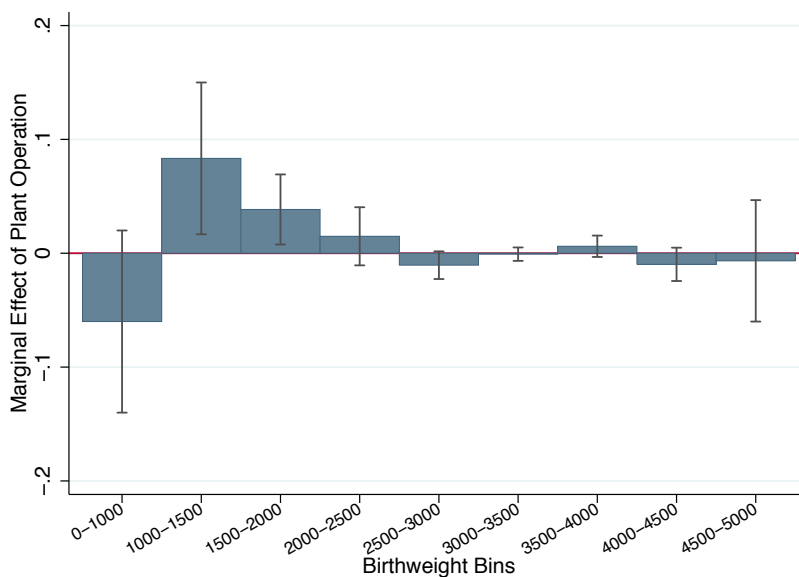
NOTE: This figure reports regression coefficients from 9 separate regressions. The dependent variable in each regression is an indicator variable for whether a birth falls in a particular birthweight range as indicated on the x-axis, and the data have been aggregated to plant by distance by year cells. The estimates reflect the effect of plant operation on "near" relative to "far" birth outcomes, expressed as a percentage of the mean incidence in each birthweight range. All regression estimates control for census tract characteristics (interacted with quadratic trends) and regressions are weighted by the group-level cell size. Multiple births are dropped from regressions. Standard errors are two-way clustered by plant and year, and reported confidence intervals reflect 2 standard errors above and below the estimate.

Table 1: Characteristics of Toxic Plants and the Surrounding Community

Panel A: Plant Characteristics by Opening and Closing Status

	(1)	(2)	(3)
	Open Continuously 1990-2002	*Opened Between 1990 and 2002*	*Closed Between 1990 and 2002*
Number of Plants	1846	689	1062
Average Plant Employment (Total Workers)	224	90	114
Average Plant Age (Years)	18.6	2.0	16.2
Mean Value of Plant Equipment (in millions)	$15.8	$15.4	$14.9
Mean Value of Plant Structures (in millions)	$6.2	$5.8	$5.1
Mean Annual Salary and Wages (in millions)	$11.7	$5.5	$6.2
Mean Annual Toxic Emissions (in pounds)	22016	23303	17919

Panel B: Community Characteristics by Distance, d, from Plants that Opened or Closed 1990-2002

	(1)	(2)	(3)	(4)
	$0 < d \leq 0.5$	$0.5 < d \leq 1$	$0 < d \leq 1$	$1 < d \leq 2$
Housing Characteristics:				
Mean Housing Value	$124,424	$126,492	$125,927	$132,227
Aggregate Housing Value (in millions)	$38.56	$60.00	$98.57	$174.80
Birth and Maternal Characteristics:				
Mother's Education	11.93	12.08	12.05	12.22
Mother's Age	26.33	26.50	26.46	26.70
Proportion Teenage Mother	0.15	0.15	0.15	0.15
Proportion Smoker	0.14	0.13	0.13	0.13
Proportion African American	0.23	0.25	0.25	0.26
Proportion Hispanic	0.32	0.30	0.31	0.29
Proportion White/Caucasian	0.72	0.71	0.71	0.70

NOTES: Panel A describes plants in Florida, Michigan, New Jersey, Pennsylvania, and Texas that reported to the Toxic Release Inventory at least one year between 1990 and 2002. In calculating plant characteristics in columns (2) and (3), the sample is restricted to observations in the 2 years after a plant opening or 2 years before a plant closing, and a single plant can appear in both columns. Plant age is right censored, as the the year a plant opened is not available for plants opened before 1975 in the Longitudinal Business Database. The value of plant equipment, structures, and salary and wages come from the NBER Productivity Database microdata and is only available for a subset of our data that matches the NBER Productivity Database in a given year. The value of plant equipment and structures is constructed using the perpetual inventory method from investment data (Mohr and Gilbert, 1996). All dollar amounts are in 2000 dollars. Panel B statistics describe community characteristics surrounding toxic plants that either opened or closed between 1990 and 2002. Housing sales and births may appear in multiple columns if they are within 2 miles of more than one plant opening or closing, but within each column a house or birth appears only once.

Table 2: The Effect of Toxic Plants on Local Housing Values

	0-0.5 Miles		0.5-1 Miles		0-1 Miles		0-1 Miles (+/- 2 years)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Estimated Effect of Plant Operation								
1(Plant Operating)×Near	-0.030***	-0.022***	-0.010**	-0.012***	-0.015***	-0.014***	-0.009**	-0.010***
	(0.007)	(0.006)	(0.005)	(0.004)	(0.005)	(0.004)	(0.004)	(0.003)
N (plant-distance-year cells)	34736	34736	34736	34736	34736	34736	30492	30492
Plant by Distance-Bin FE	X	X	X	X	X	X	X	X
State by Year FE	X		X		X		X	
Plant by Year FE		X		X		X		X
Panel B: First Difference - Estimated Effect of Plant Operation								
1(Plant Operating)×Near	-0.020**	-0.014**	-0.008*	-0.003	-0.010**	-0.005	-0.005	-0.002
	(0.010)	(0.007)	(0.004)	(0.004)	(0.005)	(0.004)	(0.006)	(0.005)
N	1114248	1114248	1305780	1305780	1375751	1375751	1196000	1196000
Panel C: First Difference - Estimated Effect of Plant Openings and Closings								
1(Plant Opening)×Near	-0.096***	-0.107***	-0.007	-0.008	-0.020	-0.022	-0.030	-0.038
	(0.036)	(0.034)	(0.023)	(0.020)	(0.022)	(0.019)	(0.028)	(0.025)
1(Plant Closing)×Near	0.017	0.010	0.008	0.003	0.010*	0.005	0.005	0.001
	(0.011)	(0.009)	(0.005)	(0.004)	(0.006)	(0.005)	(0.007)	(0.005)
H_0: Opening=-Closing (p-value)	0.051	0.013	0.968	0.827	0.688	0.438	0.402	0.164
N	1114248	1114248	1305780	1305780	1375751	1375751	1196000	1196000
State by Year FE	X		X		X		X	
County by Year FE		X		X		X		X

NOTES: This table reports regression coefficients from 24 separate regressions, 8 per panel, from a sample of 2171 plants. The dependent variable in all regressions is housing values (in logs). Both the regression sample and the indicator variable "Near" change as one moves across the columns, indicated by the column headings. For example, the specification in columns (1) and (2) examines how group-level average housing values within 0.5 miles of a plant (i.e. "Near") respond to plant operating status, relative to the comparison group. The comparison group in all columns is homes between 1 and 2 miles from a plant. In columns (7) and (8), the sample removes observations more than two years before and after changes in plant activity. In panel A, the data have been aggregated to plant by distance by year cells and regressions are weighted by the group-level cell size. Panel B reports the same estimates as Panel A using the set of houses we observe selling more than once in our sample and estimating in first differences. Panel C estimates of the asymmetric effect of plant openings/closings using the first difference specification, including p-values from tests that the two coefficients are equal, but of opposite sign. All specifications control for census tract characteristics (interacted with quadratic trends). Standard errors two-way clustered by plant and year are in parentheses.

Table 3: The Effect of Toxic Plants on Local Housing Values: Above/Below Median (0-0.5 Miles), First-Difference

	(1) Employment	(2) Payroll	(3) Fugitive Emissions	(4) Stack Emissions	(5) Total Emissions	(6) Mean Toxicity	(7) Max Toxicity	(8) Fraction College	(9) Fraction White	(10) Housing Value	(11) Median Income
1(Plant Operating)×1(<0.5 Miles)	-0.005	-0.002	-0.009	-0.008	-0.009	-0.010	-0.003	-0.014**	-0.016**	-0.012	-0.024***
	(0.009)	(0.009)	(0.006)	(0.006)	(0.006)	(0.006)	(0.008)	(0.007)	(0.006)	(0.007)	(0.006)
1(Plant Op.)×1(<0.5 Miles)×1(Above Median)	-0.012	-0.017	-0.010	-0.010	-0.008	-0.006	-0.016	0.007	0.014	0.002	0.032***
	(0.012)	(0.012)	(0.015)	(0.014)	(0.014)	(0.014)	(0.013)	(0.014)	(0.012)	(0.012)	(0.009)
N	1140399	1140399	1140399	1140399	1140399	1140399	1140399	1140399	1140399	1140399	1140399

NOTES: This table reports regression coefficients from 11 separate regressions, where we interact our treatment variable 1(Plant Operating)×1(<0.5 Miles) with an indicator for whether the plant/community is above or below the median characteristic indicated in the column heading. The median indicator is equal to 1 if the plant-level median of the column variable (taken over plant operating years) is above or below the sample median value (taken over median plant values). The dependent variable in all regressions is housing values (in logs). All regressions control for county by year fixed effects and census tract characteristics (interacted with quadratic trends), and the model is estimated in first differences. Standard errors are two-way clustered by plant and year.

∞

Table 4: The Effect of Toxic Plants on Low Birthweight

	0-0.5 Miles		0.5-1 Miles		0-1 Miles		0-1 Miles (+/- 2 years)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Estimated Effect of Plant Operation								
1(Plant Operating)×Near	0.0010	0.0012	0.0014**	0.0015**	0.0013**	0.0014**	0.0021**	0.0026***
	(0.0010)	(0.0012)	(0.0006)	(0.0006)	(0.0006)	(0.0007)	(0.0009)	(0.0009)
N	88958	88958	88958	88958	88958	88958	63324	63324
Plant Count	3438	3438	3438	3438	3438	3438	3438	3438
Panel B: Estimated Effect of Plant Openings and Closings								
1(Plant Opened)×Near	0.0025	0.0022	0.0024***	0.0027***	0.0024**	0.0024***	0.0031*	0.0037**
	(0.0019)	(0.0018)	(0.0009)	(0.0010)	(0.0009)	(0.0008)	(0.0017)	(0.0017)
1(Plant Closed)×Near	-0.0002	-0.0007	-0.0009	-0.0009	-0.0007	-0.0009	-0.0016	-0.0021*
	(0.0016)	(0.0016)	(0.0009)	(0.0010)	(0.0009)	(0.0009)	(0.0012)	(0.0013)
H_0: Opening=-Closing (p-value)	0.44	0.56	0.32	0.28	0.22	0.24	0.51	0.48
N	88958	88958	88958	88958	88958	88958	63324	63324
Plant Count	3438	3438	3438	3438	3438	3438	3438	3438
Plant×Distance-Bin FE	X	X	X	X	X	X	X	X
State×Year FE	X		X		X		X	
Plant×Year FE		X		X		X		X

NOTES: This table reports regression coefficients from 16 separate regressions, 8 per panel. The dependent variable in all regressions is the mean incidence of low birthweight where the data have been aggregated to plant by distance by year cells. Cell level averages have been adjusted for maternal characteristics including age, education, race, and smoking behavior, as well as for month of birth, birth order, and gender of child. See text for details. The mean incidence of low birthweight in our sample is 0.07. Both the regression sample and the indicator variable "Near" change as one moves across the columns, indicated by the column headings. For example, the specification in columns (1) and (2) examines how group-level average birth outcomes within 0.5 miles of a plant (i.e. "Near") respond to plant operating status, relative to the comparison group. The comparison group in all columns is births between 1 and 2 miles from a plant. In columns (7) and (8), the sample removes observations more than two years before and after changes in plant activity. Panel A estimates the effect of plant operating status on local birth outcomes, where 1(Plant Operating) is a indicator variable equal to one for plants that have opened and/or have not yet closed. Panel B estimates the asymmetric effect of plant openings/closings. Panel B reports p-values from tests that the two coefficients are equal in magnitude but of opposite sign. All columns control for census tract characteristics (interacted with quadratic trends) and regressions are weighted by the group-level cell size. Multiple births are dropped from regressions. Standard errors are two-way clustered by plant and year.

Table 5: The Effect of Toxic Plants on Low Birthweight: Above/Below Median (0-1 Miles)

	(1) Employment	(2) Payroll	(3) Fugitive Emissions	(4) Stack Emissions	(5) Total Emissions	(6) Mean Toxicity	(7) Max Toxicity	(8) Fraction College	(9) Fraction White	(10) Housing Value	(11) Median Income
1(Plant Operating)×1(<1 Mile)	0.0016**	0.0012	0.0010	0.0008	0.0013	0.0013*	0.0006	0.0007	0.0010*	0.0028***	0.0014*
	(0.0007)	(0.0007)	(0.0007)	(0.0007)	(0.0009)	(0.0007)	(0.0007)	(0.0007)	(0.0005)	(0.0010)	(0.0007)
1(Plant Op.)×1(<1 Mile)×1(Above Median)	-0.0010	0.0003	0.0007	0.0010	0.0000	-0.0001	0.0020	0.0017	0.0009	-0.0028*	-0.0003
	(0.0013)	(0.0014)	(0.0013)	(0.0011)	(0.0020)	(0.0017)	(0.0013)	(0.0017)	(0.0010)	(0.0017)	(0.0013)
N	88958	88958	88958	88958	88958	88958	88958	88958	88958	88958	88958

NOTES: This table reports regression coefficients from 11 separate regressions, where we interact our treatment variable 1(Plant Operating)×1(<1 Mile) with an indicator for whether the plant/community is above or below the median characteristic indicated in the column heading. The median indicator is equal to 1 if the plant-level median of the column variable (taken over plant operating years) is above or below the sample median value (taken over median plant values). The dependent variable in all regressions is the mean incidence of low birthweight, where the data have been aggregated to plant by distance by year cells. Cell level averages have been adjusted for micro covariates, and all regressions control for plant by distance and plant by year fixed effects, as well as census tract characteristics (interacted with quadratic trends). Regressions are weighted by the group-level cell size. Standard errors are two-way clustered by plant and year.

Table 6: The Effect of Toxic Plants on Alternative Measures of Infant Health

	(1) Birthweight (grams, in logs)	(2) Birthweight (grams, in logs)	(3) Very Low Birthweight	(4) Very Low Birthweight	(5) Premature	(6) Premature	(7) Congenital Anomaly	(8) Congenital Anomaly	(9) Infant Death	(10) Infant Death	(11) Summary Index	(12) Summary Index
1(Plant Op)× 1(<1 Mile)	-0.0011***	-0.0012***	0.0001	0.0001	0.0009**	0.0009**	-0.0003	-0.0005	0.0001	0.0000	0.0165**	0.0158**
	(0.0003)	(0.0004)	(0.0002)	(0.0001)	(0.0004)	(0.0005)	(0.0009)	(0.0010)	(0.0001)	(0.0001)	(0.0066)	(0.0071)
N	88958	88958	88958	88958	88404	88404	88212	88212	89388	89388	89388	89388
Mean	8.09	8.09	0.0113	0.0113	0.0845	0.0845	0.0679	0.0679	0.0068	0.0068	-0.0004	-0.0004
Plant×Dist-Bin FE	X	X	X	X	X	X	X	X	X	X	X	X
State×Year FE	X		X		X		X		X		X	
Plant×Year FE		X		X		X		X		X		X

NOTES: This table reports regression coefficients from 12 separate regressions on a sample of 3438 plants. The dependent variable is listed in the column heading, and data have been aggregated to plant by distance by year cells. Cell level averages have been adjusted for maternal characteristics. See text for details. The comparison group in all columns is births between 1 and 2 miles from a plant. Columns (11) and (12) present results from a summary index measure of outcomes to address concerns pertaining to inference with a large number of outcomes. Outcomes for the summary index measure include those listed above, in addition to low birthweight, and outcomes are standardized to be mean 0 and standard deviation 1. Before combining, birthweight is multiplied by −1 so that an increase in the summary index measure reflects an increase in adverse health outcomes. All regressions control for tract characteristics (interacted with quadratic trend). Multiple births are dropped from regressions. Regressions are weighted by the group-level cell size. Standard errors are two-way clustered by plant and year.

A   Appendix Tables and Figures: Not For Publication

Table A1: Effect of Plant Operating Status on Ambient Air Pollution: Pollutant Specific Coefficients

	(1)	(2)	(3)	(4)	(5) Methylene	(6)	(7)	(8)
	Benzene	Cumene	Dichloromethane	Lead (TSP)	Chloride	Nickel	Styrene	Toluene
Plant Open$\times$Dist	-9.1502***	-37.4841***	-0.8801	-0.6627	0.2579	-10.6628***	-21.0906	-1.9633
	(2.3810)	(8.5595)	(0.7433)	(6.0252)	(1.3544)	(0.8827)	(14.0195)	(4.9933)
Plant Open$\times$Dist^2	7.5063***	28.2352***	0.9816	2.2785	0.2332	7.8313***	20.9231	0.5489
	(2.2812)	(7.3815)	(0.8171)	(5.6212)	(1.5556)	(0.9525)	(13.3794)	(4.2769)
Plant Open$\times$Dist^3	-2.5056***	-8.7304***	-0.3856	-1.5783	-0.2477	-2.4739***	-7.8927	0.1807
	(0.8857)	(2.7271)	(0.3201)	(2.1001)	(0.6738)	(0.3816)	(4.8879)	(1.3857)
Plant Open$\times$Dist^4	0.2944**	0.9455**	0.0492	0.2873	0.0469	0.2837***	0.9772	-0.0600
	(0.1194)	(0.3630)	(0.0408)	(0.2700)	(0.0951)	(0.0509)	(0.5901)	(0.1524)
Monitor Count	49	23	51	32	21	11	51	67
N	1106	358	1077	532	315	324	1000	1970

NOTES: This table reports regression coefficients from 8 separate regressions. The unit of observation is the monitor-plant pair and the dependent variable in all regressions is ambient hazardous pollution, standardized so each pollutant has mean 0 and standard deviation is 1. All regressions include monitor-pair fixed effect and an indicator variable for whether each plant is operating 1(Plant Operating). Standard errors are two-way clustered on plant and monitor.

Table A2: The Effect of Toxic Plants on Housing Values: Model Sensitivity to Alternative Controls

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Baseline					**First Difference**			
1(Plant Operating)×1(<0.5 Miles)	-0.027***	-0.032***	-0.030***	-0.025***	-0.025***	-0.034***	-0.020*	-0.020**	-0.014**
	(0.008)	(0.007)	(0.007)	(0.007)	(0.008)	(0.008)	(0.010)	(0.010)	(0.007)
N	34736	34736	34736	34736	34736	1114248	1114248	1114248	1114248
Plant×Distance-Bin FE	X	X	X	X	X				
Census Tract Quadratic Trends		X	X	X	X		X	X	X
State×Year FE			X					X	
County×Year FE				X					X
Plant×Year FE					X				

NOTES: This table reports regression coefficients from 9 separate regressions. The dependent variable in all regressions is housing values (in logs). In columns (1)-(5), the data have been aggregated to plant by distance by year cells, and regressions are weighted by the group-level cell size. In columns (6)-(9), we estimate the model on the set of houses for which we observe 2+ sales in our sample period. Each regression reports estimates of the effect of plant operating status on local housing values, where 1(Plant Operating) is a indicator variable equal to one for plants that have opened and/or have not yet closed. Standard errors two-way clustered by plant and year are in parentheses.

Table A3: The Effect of Toxic Plants on Local Housing Values and Low Birthweight: Using 2-4 Mile Radius Comparison Group

	0-0.5 Miles		0.5-1 Miles		0-1 Miles		0-1 Miles (+/- 2 years)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Estimated Effect of Plant Operation on Housing Values								
1(Plant Operating)×1(< 1 Mile)	-0.032***	-0.026***	-0.013**	-0.017***	-0.018***	-0.019***	-0.011*	-0.012***
	(0.009)	(0.008)	(0.007)	(0.006)	(0.007)	(0.006)	(0.006)	(0.005)
N	34736	34736	34736	34736	34736	34736	30492	30492
PlantCount	2171	2171	2171	2171	2171	2171	2171	2171
Panel B: Estimated Effect of Plant Operation on Low Birthweight								
1(Plant Operating)×1(< 1 Mile)	0.0001	-0.0000	0.0009*	0.0011**	0.0007	0.0008	0.0017**	0.0020***
	(0.0011)	(0.0011)	(0.0006)	(0.0005)	(0.0006)	(0.0005)	(0.0008)	(0.0008)
N	88922	88922	88922	88922	88922	88922	63301	62984
PlantCount	3438	3438	3438	3438	3438	3438	3438	3438
Plant×Distance-Bin FE	X	X	X	X	X	X	X	X
State×Year FE	X		X		X		X	
Plant×Year FE		X		X		X		X

NOTES: This table reports regression estimates from specifications that are identical to Table 2 (Panel A) and Table 4 (Panel A) except the comparison group is observations between 2 and 4 miles from a plant.

Table A4: The Effect of Toxic Plants on Local Housing Values and Low Birthweight, Distance Radius Specification

	(1) 0-0.5	(2) 0-0.5	(3) 0.5-1	(4) 0.5-1	(5) 1-1.5	(6) 1-1.5	(7) 1.5-2	(8) 1.5-2	(9) 0-1	(10) 0-1	(11) 1-2	(12) 1-2
	Panel A: Estimated Effect of Plant Operation on Housing Values											
1(Plant Operating)	-0.018**	-0.018**	-0.011	-0.013*	-0.002	-0.006	0.004	-0.000	-0.015**	-0.016**	0.000	-0.003
	(0.009)	(0.008)	(0.007)	(0.007)	(0.006)	(0.007)	(0.006)	(0.006)	(0.007)	(0.007)	(0.006)	(0.006)
N	17336	17336	17368	17368	17368	17368	15976	15976	17368	17368	17368	17368
PlantCount	2167	2167	2171	2171	2171	2171	1997	1997	2171	2171	2171	2171
	Panel B: Estimated Effect of Plant Operation on Low Birthweight											
1(Plant Operating)	0.0004	0.0003	0.0012**	0.0012**	-0.0012*	-0.0012*	-0.0003	-0.0002	0.0009	0.0009	-0.0003	-0.0003
	(0.0011)	(0.0012)	(0.0005)	(0.0005)	(0.0007)	(0.0007)	(0.0004)	(0.0005)	(0.0006)	(0.0006)	(0.0003)	(0.0003)
N	42077	42077	42140	42140	42114	42114	40556	40556	74890	74890	75002	75002
PlantCount	3254	3254	3256	3256	3256	3256	3136	3136	5787	5787	5794	5794
Plant×Distance-Bin FE	X	X	X	X	X	X	X	X	X	X	X	X
Year FE	X		X		X		X		X		X	
State×Year FE		X		X		X		X		X		X

NOTES: This table reports regression estimates from specifications that are identical to Table 2 (Panel A) and Table 4 (Panel A) except that each regression is estimated using only observations from the distance range indicated in the column headings. For example, the regressions described in columns (1) and (2) include only observations from within 0.5 miles of a plant.

Table A5: The Effect of Toxic Plants on Low Birthweight: Model Sensitivity to Alternative Controls

	(1)	(2)	(3)	(4)	(5)
	Estimated Effect of Plant Operation				
1(Plant Operating)×1(< 1 Mile)	0.0016***	0.0016***	0.0014**	0.0013**	0.0014**
	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0007)
N	88958	88958	88958	88958	88958
PlantCount	3438	3438	3438	3438	3438
Plant×Distance-Bin FE	X	X	X	X	X
Maternal Characteristics		X	X	X	X
Census Tract Quadratic Trends			X	X	X
State×Year FE				X	
Plant×Year FE					X

NOTES: This table reports regression coefficients from 9 separate regressions aimed at evaluating the sensitivity of the birth outcome estimates to alternative specifications. The dependent variable in all columns is the mean incidence of low birthweight where the data have been aggregated to plant by distance by year cells. Cell level averages have been adjusted for maternal characteristics including age, education, race, and smoking behavior, as well as for month of birth, birth order, and gender of child. See text for details. The mean incidence of low birthweight in our sample is 0.07. The comparison group in all columns is births between 1 and 2 miles from a plant. All columns control for census tract characteristics (interacted with quadratic trends) and regressions are weighted by the group-level cell size. Multiple births are dropped from regressions. Standard errors are two-way clustered by plant and year.

Table A6: The Effect of Toxic Plants on Maternal Characteristics

	(1) African American	(2) African American	(3) Hispanic	(4) Hispanic	(5) Mother's Education	(6) Mother's Education	(7) Teenage Mother	(8) Teenage Mother
1(Plant Operating)×1(<1 Mile)	-0.007***	-0.006***	-0.002	-0.001	0.008	0.001	-0.001	-0.000
	(0.002)	(0.002)	(0.002)	(0.002)	(0.010)	(0.009)	(0.001)	(0.001)
N	88062	88062	87438	87438	82608	82608	89302	89302
PlantCount	3438	3438	3438	3438	3438	3438	3438	3438
Mean	0.201	0.201	0.219	0.219	12.555	12.555	0.137	0.137

	(9) Smoker	(10) Smoker	(11) White College	(12) White College	(13) Predictive Index	(14) Predictive Index	(15) Fertility	(16) Fertility
1(Plant Operating)×1(<1 Mile)	-0.001	0.002	0.002**	0.001	0.001***	0.001***	0.002	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.005)	(0.004)
N	88077	88077	86093	86093	89388	89388	89388	89388
PlantCount	3438	3438	3438	3438	3438	3438	3438	3438
Mean	0.149	0.149	0.151	0.151	8.084	8.084	4.233	4.233

Plant×Distance-Bin FE	X	X	X	X	X	X	X	X
State×Year FE	X		X		X		X	
Plant×Year FE		X		X		X		X

NOTES: This table reports regression coefficients from 16 separate regressions, 8 per panel. The dependent variable is listed in the column heading, and data have been aggregated to plant by distance by year cells. Cell level averages have been adjusted for maternal characteristics. See text for details. The comparison group in all columns is births between 1 and 2 miles from a plant. The dependent variable "Predictive Index" is created by first running a regression of birthweight on observable characteristics of the mother. The predicted values from this regression are used as a summary index of demographic changes. Fertility is measured as the log number of births in a cell. Regressions are weighted by the group-level cell size (with the exception of "Fertility"). All regressions control for tract characteristics (excluding the dependent variable of interest), interacted with quadratic trends. Multiple births are dropped from regressions. Regressions are weighted by the group-level cell size. Standard errors two-way clustered by plant and year.

Table A7: The Effect of Toxic Plants on Low Birthweight: Alternative Timing Assumptions

	(1)	(2)	(3)	(4)
	Fractional Assignment		Outer Donut	
	0-1 (+/- 2 years)	0-1 (+/- 2 years)	0-1 (+/- 2 years)	0-1 (+/- 2 years)
Panel A: Estimated Effect of Plant Operation				
1(Plant Operating)×1(< 1 Mile)	0.0010	0.0011*	0.0005	0.0010
	(0.0007)	(0.0006)	(0.0007)	(0.0007)
N	753329	753329	695179	695179
PlantCount	3438	3438	3438	3438
Panel B: Estimated Effect of Plant Openings and Closings				
1(Plant Opened)×1(< 1 Mile)	0.0017**	0.0022***	0.0018*	0.0025**
	(0.0009)	(0.0008)	(0.0010)	(0.0011)
1(Plant Closed)×1(< 1 Mile)	0.0000	-0.0003	0.0001	-0.0003
	(0.0008)	(0.0008)	(0.0009)	(0.0009)
H_0: Opening=-Closing (p-value)	0.087	0.056	0.103	0.097
N	753329	753329	695179	695179
Plant Count	3438	3438	3438	3438
Plant×Distance-Bin FE	X	X	X	X
State×Year FE	X		X	
Plant×Year FE		X		X

NOTES: This table reports regression coefficients from 8 separate regressions, 4 per panel. The dependent variable in all regressions is the mean incidence of low birthweight where the data have been aggregated to plant by distance by month-year cells. Cell level averages have been adjusted for maternal characteristics including age, education, race, and smoking behavior, as well as for month of birth, birth order, and gender of child. Columns (1) and (2) assign plant operating status "fractionally" based on the expected number of months a plant was open during the child's gestational period $\in [0,1]$. Columns (3) and (4) focus only on births in which there is no ambiguity as to whether the plant was opened or closed (i.e. dropping births just around the plant opening/closing event). In all columns we focus on the sample +/- 2 years from the plant opening event. Panel A estimates the effect of plant operating status on local birth outcomes, where 1(Plant Operating) is a indicator variable equal to one for plants that have opened and/or have not yet closed. Panel B estimates the asymmetric effect of plant openings/closings. Panel B reports p-values from tests that the two coefficients are equal in magnitude but of opposite sign. All columns control for census tract characteristics (interacted with quadratic trends) and regressions are weighted by the group-level cell size. Multiple births are dropped from regressions. Standard errors are two-way clustered by plant and year.

Table A8: The Effect of Toxic Plants on Birthweight, Additional Evidence

	(1) $bw \leq 1000$	(2) $bw > 1000$ $bw \leq 1500$	(3) $bw > 1500$ $bw \leq 2000$	(4) $bw > 2000$ $bw \leq 2500$	(5) $bw > 2500$ $bw \leq 3000$	(6) $bw > 3000$ $bw \leq 3500$	(7) $bw > 3500$ $bw \leq 4000$	(8) $bw > 4000$ $bw \leq 4500$	(9) $bw > 4500$ $bw \leq 5000$
Panel A: Estimated Effect of Plant Operation									
1(Plant Operating)×1(< 1 Mile)	-0.0003	0.0005**	0.0005**	0.0007	-0.0019	-0.0003	0.0017	-0.0008	-0.0001
	(0.0002)	(0.0002)	(0.0002)	(0.0006)	(0.0011)	(0.0011)	(0.0013)	(0.0006)	(0.0004)
Mean	0.005	0.006	0.013	0.047	0.181	0.374	0.277	0.082	0.015
Panel B: Estimated Effect of Plant Openings and Closings									
1(Plant Opening)×1(< 1 Mile)	-0.0001	0.0003	0.0011***	0.0013	-0.0006	0.0007	0.0001	-0.0016	-0.0011**
	(0.0004)	(0.0003)	(0.0004)	(0.0008)	(0.0025)	(0.0023)	(0.0026)	(0.0011)	(0.0005)
1(Plant Closing)×1(< 1 Mile)	0.0004	-0.0006**	-0.0001	-0.0004	0.0025*	0.0008	-0.0026*	0.0004	-0.0004
	(0.0003)	(0.0003)	(0.0003)	(0.0007)	(0.0013)	(0.0011)	(0.0014)	(0.0005)	(0.0005)
H_0: Opening=-Closing (p-value)	0.616	0.463	0.080	0.310	0.510	0.555	0.398	0.280	0.004
Mean	0.005	0.006	0.013	0.047	0.181	0.374	0.277	0.082	0.015

NOTES: This table reports regression coefficients from 18 separate regressions, 9 per panel, on a regression of 88958 plant-distance-year cells and 3438 plants. The dependent variable in each regression is an indicator variable for whether a birth falls in a particular birthweight range as indicated by column headings, and the data have been aggregated to plant by distance by year cells. The regression sample changes as one moves across the columns, indicated by the column headings. For example, the specification in column (1) examines the relative likelihood of a birth being below 1000 grams within 1 mile of a plant responds to plant operating status, relative to the comparison group. The comparison group in all columns is births between 1 and 2 miles from a plant. Panel A estimates the effect of plant operating status on local birth outcomes, where 1(Plant Operating) is a indicator variable equal to one for plants that have opened and/or have not yet closed. Panel B estimates the asymmetric effect of plant openings/closings. Panel B reports p-values from tests that the two coefficients are equal in magnitude but of opposite sign. All columns control for plant×distance-bin and plant×year fixed effects, census tract characteristics (interacted with quadratic trends), and regressions are weighted by the group-level cell size. Multiple births are dropped from regressions. Standard errors are two-way clustered by plant and year.

Supplementary Online Appendix

Alternative Timing of Assigning Plant Activity to Birth Outcomes

Below we describe alternative methods of assigning birth outcomes to plant events due to the ambiguity within a year for when a plant opens/closes. In the body of the paper, we restrict our birth outcomes to births between November and December of year *t-1* and January and February of year *t*. This sample restriction reflects the fact that we do not observe *when* during the year a plant opens or closes. Census reporting occurs in March so we have a 12-month window during which each opening/closing could have occurred. By focusing on births towards the end of the year (or the beginning of the following year), we increase the likelihood of that birth being exposed to these events. The idea is that a baby born January 1, 2002 has not been exposed to any of the toxic plant activity for calendar 2002, but was exposed to toxic emissions in 9 out of 12 months of 2001. Similarly, a baby born in November 2001 was exposed to toxic emissions for 9 out of 12 months of 2001. In contrast, timing is much more difficult for births in the middle of the year. A June birth, for example, may or may not have been exposed to a change in plant operations depending on *when* during the year the change occurred. Restricting our sample to these four months also has the additional advantage of limiting the extent to which seasonality in plant activity or birth outcomes affects our findings.

We nonetheless investigated the sensitivity of our estimates to additional sample selection criteria in two separate ways. First, we created a dataset of plant-month-year observations for all births around a given plant in a given month-year. We then calculated the expected exposure of a birth to a plant opening or closing assuming that plant openings/closings are uniformly distributed by month throughout a year. This allows us to calculate the expected fraction of in utero exposure to a plant opening/closing based on month of birth. We use this expected fraction as a continuous measure of treatment for the birth-months immediately surrounding the plant. For birth-months long before/after the event, they are unambiguously treated or not-treated given their "distance" from the ambiguously timed event. A summary of this assignment mechanism is described in the tables below, where the last line of each table represents the expected fraction of gestation exposed to a plant opening/closing conditional on month of birth. This expected fraction makes up our treatment variable.

		Month of Birth for Year 2007 Opening				
		Apr-06	May-06	…	Oct-07	Nov-07
Month of Opening for Year 2007 Opening	Apr-06	1	2	…	9	9
	May-06	0	1	…	9	9
	Jun-06	0	0	…	9	9
	Jul-06	0	0	…	9	9

31

	Aug-06	0	0	…	9	9
	Sep-06	0	0	…	9	9
	Oct-06	0	0	…	9	9
	Nov-06	0	0	…	9	9
	Dec-06	0	0	…	9	9
	Jan-07	0	0	…	9	9
	Feb-07	0	0	…	9	9
	Mar-07	0	0	…	8	9
	Expected Number of Months Gestational Exposure	0.08	0.25	…	8.92	9.00
	Expected Fraction of Gestation	0.01	0.03	…	0.99	1.00

		Month of Birth for Year 2007 Closing				
		Apr-06	May-06	…	Oct-07	Nov-07
Month of Closing for Year 2007 Closing	Apr-06	9	8	…	0	0
	May-06	9	9	…	0	0
	Jun-06	9	9	…	0	0
	Jul-06	9	9	…	0	0
	Aug-06	9	9	…	0	0
	Sep-06	9	9	…	0	0
	Oct-06	9	9	…	0	0
	Nov-06	9	9	…	0	0
	Dec-06	9	9	…	0	0

	Jan-07	9	9	… 0	0
	Feb-07	9	9	… 1	0
	Mar-07	9	9	… 2	1
	Expected Number of Months Gestational Exposure	9	8.917	… 0.250	0.083
	Expected Fraction of Gestation	1	0.991	… 0.028	0.009

We also explored models that exclude ambiguous birth months altogether and simply look at pre-post outcomes for births/gestation clearly on either side of openings and closings. This has the added benefit of being somewhat agnostic about the prenatal production technology, whereas the continuous assignment mechanism above imposes some structure in the form of additive linearity across months in utero. In both this specification and in the "fractional apportionment" specification, we restricted the sample to an "event window" that limits the sample to 2 years before and after an event (i.e. as in columns (7) and (8) of the baseline table to avoid putting too much weight on observations far before and after the event. Results from both exercises are presented in Appendix Table A7.

A Note on Geocoding

Our housing transaction data provides information on the street address of the housing parcel which we use to geocode our data. We used Yahoo Geocoding API to match street address data with latitude and longitude via a Python routine. The Yahoo API provides much better geocoding capabilities in the case of minor errors in street addresses relative to some more commonly used geocoding routines (e.g. ArcGIS). For example, if one were to type in a single street address without a city identifier into Yahoo, and that address was fairly unique (e.g. 2342 Bristol Lane, KY), Yahoo would more than likely be able to figure out where the exact location of the house was without city information. Using the Yahoo Geocoding API, we were able to match over 98% of our transactions to latitude and longitude. We geocoded the birth data using ArcGIS v9.3 geocoding engine combined with the US Census Bureau's Tiger/Line reference layer. For each state, we audited the match using external web services such as Google Maps, and we found that the ArcGIS match was accurate for nearly all of the randomly sampled birth addresses.