# The effects of environmental health risks on housing values and minorities*

Jules van Binsbergen† João F. Cocco† Marco Grotteria‡
S. Lakshmi Naaraayanan§

June 25, 2023

**Abstract**

Our study estimates the effects of environmental health risks on housing values using repeated property sales in a narrow time window when plants first report releasing carcinogenic toxins and in narrow geographical areas surrounding the plants. More expensive properties experience a decline in value of around 15%, while less expensive ones benefit from an increase in value. Additionally, the number of employees in the same toxic plants increases by two percentage points after the environmental incident. Our results suggest that the willingness of households to pay to avoid such plants is offset by an increase in industrial activity with greater benefits for those who purchase lower-priced houses in the area.

*Keywords*: Environmental health risks, Household finance, Real estate prices.

*JEL Classification*: G11, G50, Q51.

*We thank Francisco Gomes, Kasper Nielsen, Henri Servaes, Daniel Wolfenzon, and seminar participants at London Business School and Peking University for helpful comments and suggestions.

†University of Pennsylvania, CEPR and NBER. Email: julesv@wharton.upenn.edu

†London Business School and CEPR. Email: jcocco@london.edu.

‡London Business School. Email: mgrotteria@london.edu.

§London Business School. Email: lnaaraayanan@london.edu

# 1 Introduction

One of the key challenges in measuring the costs of environmental health risk is that the location choices of firms and households are endogenous. These choices lie behind the well-established observation that economically disadvantaged households are disproportionately more exposed to environmental harms (Office, 1983). Companies may intentionally select socioeconomically disadvantaged areas to establish new sites, which implies pre-existing socio-economic disparities.[1] Households may also choose to live in areas with lower environmental quality, for instance, driven by financial constraints, which widens environmental inequality among households (Kermani and Wong, 2021).

In this paper, we estimate the dollar value of environmental health risk by identifying a set of plants that already existed and that started reporting the emission of carcinogenic toxins. This helps us disentangle between the different mechanisms at play as we hold constant plant location choices. Our sample comprises 11,143 unique toxic plants operating throughout the United States over the last two decades. We focus on plants when they first report to the EPA the emission of carcinogenic toxins (environmental *event*).

We follow the recent literature (e.g., Currie, Davis, Greenstone, and Walker, 2015, Diamond and McQuade, 2019) in addressing this identification challenge by comparing the effects of the event on house values within the immediate vicinity of the plant, namely those within a 3-mile ring of the facility and those in a ring of between three and five miles from the same facility. The former set of houses is our treated group, and the latter is the control group. The idea is that emissions will have more of an impact on properties closest to the toxic plant, whereas all properties within the 5-mile radius will benefit from local economic activity effects. We provide estimates for smaller treated rings.

Our data source for housing values is Corelogic. We provide two sets of baseline estimates. First, we compare transaction prices in the treated and control groups before and after the event year controlling for plant and year times county fixed effects. Note that this empirical specification effectively holds constant plant siting decisions and compares changes in house

[1] This happens, for instance, when these companies believe they are less likely to face opposition from local communities in disadvantaged areas or when they can have better access to a particular type of workforce (Diamond and Gaubert, 2022).

prices around the first reporting year after controlling for regional shocks to the housing market. We estimate negative effects on the value of properties in the treated group relative to the control group of between -6% and -12%, increasing (in absolute value) as we decrease the size of the treated ring. The small ring size, the short time window, and the stringent fixed effects that we include in the empirical specification all help us better identify environmental pollution costs as captured by changes in house values.

In the second, we restrict the sample to properties that were transacted *both* in the year before and in the year after the event. Such a stringent restriction implies that for these tests we are left with a selected sample of properties. At the same time, it allows us to use property fixed effects and estimate *within property* changes to housing values. This is important since houses are heterogeneous across many unobservable dimensions. The (absolute) values of the estimates are significantly smaller, ranging between -1.4% and -1.7% (as we decrease the size of the treated group ring). Naturally, properties that are transacted twice in the space of three years are a very selected group, but we show that the estimates are robust to considering longer event windows (from year -3 before the event to year +1).

The relatively small coefficient estimates might suggest that the events have limited effects, or at least that their price effects are concentrated on properties with certain (unobservable) characteristics and are absorbed by the property fixed effects. However, we show that these average effects hide considerable heterogeneity. We divide the properties into those with a below and an above median price based on the ex-ante price distribution, and then estimate the effects of the event for each of these two groups. Properties in the above median group experience a decline of between -10% and -20% after the event relative to those in the below median group, depending on the size of the treated ring and the length of the event window.

Moreover, our estimates show increases in the value of houses in the below median group after the event, consistent with the hypothesis that the event generate positive economic benefits. Consistent with this, we show that employment and sales in the treated facilities increase in the year after the event.

These results are important since they provide evidence of a channel through which sorting may occur. To the extent that ex-ante the more expensive properties are owned by households with higher incomes, the significantly larger house price declines in the treated group show that they are willing to sell their houses at a significant loss in order to move. The higher employment in the facility after the event shows that the event has economic benefits for the local area.

Finally, we use the granularity of the data to study buying and selling by minorities. Specifically, we use names of buyers and sellers to understand whether new information about carcinogenic emissions affects the identity of buying and selling in a three-mile ring around the plant compared to a three-to-five-mile ring in the event window (-1,+1). We use the algorithm proposed by Laohaprapanon, Sood, and Naji (2022), which exploits data from the US census to predict race and ethnicity based on individual buyers' and sellers' last names. Our analysis suggests a significant and greater fraction of minorities buying and selling in the immediate vicinity of toxic plants relative to the control group and after accounting for time-invariant plant characteristics. These results provide the first evidence of *granular* changes in neighborhood composition in a short window emanating from house prices.

We perform several robustness tests to show that our baseline and heterogeneity results hold up under a variety of different specifications. First, we examine and show robustness to the concern that our estimates are sensitive to the size of treatment rings based on the distance between the toxic plant and the property. To do so, we examine the sensitivity of the estimates to the size of rings (1, 1.25, 2, 2.5, and 3 miles) and find that estimates are relatively stable both in economic magnitude and statistical significance. Second, we show robustness to longer time horizons by expanding the event window to include transacted properties within three years around the first reporting year of carcinogenic emissions. Third, we address the concern that changes in local economic conditions likely affect the estimates by including plant $\times$ year fixed effects in our empirical specification. Lastly, we rule out concerns about measurement error and outliers driving the estimates. Across all these tests, we find a robust and negative effect on housing values around the reporting of carcinogenic emissions, with more expensive properties experiencing a more significant negative effect on their prices.

Our paper is related to several strands of literature. First, extant literature on agglomeration argues for spillovers and their propagation through firm networks to the local economy in the form of input sharing, labor market pooling, and knowledge externalities (Giroud, Lenzu, Maingi, and Mueller, 2021, Bloom, Brynjolfsson, Foster, Jarmin, Patnaik, Saporta-Eksten, and Van Reenen, 2019, Neumark and Simpson, 2015, Enrico, 2011, Greenstone, Hornbeck, and Moretti, 2010). Unlike their focus on positive externalities, we aim to quantify the (net) impact after accounting for negative externalities, as captured by housing values.

Second, we build on the large literature that uses changes in house prices to estimate the willingness-to-pay for households and benefits from local environmental quality improvements (Chay and Greenstone, 2005, Greenstone and Gallagher, 2008, Bayer, Keohane, and Timmins, 2009, Currie, Davis, Greenstone, and Walker, 2015, Ito and Zhang, 2020). Relative to this literature, our findings suggest that changes in housing values reflect new information on plants reporting carcinogenic emissions, holding constant plant siting decisions. Importantly, these effects demonstrate substantial heterogeneity and vary by the value of the houses. To the extent that the value of houses is correlated with income and, more broadly, socioeconomic characteristics of the households, our results suggest that pollution externalities likely impact the long-run neighborhood composition through housing values (Banzhaf, Ma, and Timmins, 2019).

Lastly, our paper is related to a growing literature that studies the impact of climate risk on the value of real estate assets (Bernstein, Gustafson, and Lewis, 2019, Ortega and Taspinar, 2018, Baldauf, Garlappi, and Yannelis, 2020, Murfin and Spiegel, 2020, Giglio, Maggiori, Krishna, Stroebel, and Weber, 2021, among others) and the mortgages used to finance them (Issler, Stanton, Vergara, and Wallace, 2020, Gete and Tsouderou, 2021, Keys and Mulder, 2020).[2] Moreover, recent research has focused on the role of energy efficiency investments and the effects of the regulatory intervention to mitigate climate risk (e.g., Clara, Cocco, Naaraayanan, and Sharma, 2022, Fuerst, McAllister, Nanda, and Wyatt, 2015). In contrast to this literature, we study the (net) impact of environmental pollution as captured by housing values and our findings suggest that the willingness of individuals to pay to avoid toxic plants is offset by an increase in industrial activity, with greater benefits for individuals who purchase lower-priced houses in the area.

The paper is organized as follows. Section 2 describes the data and the identification. Section 3 shows the estimated effects on housing values. Section 4 focuses on heterogeneous effects on property values and it also includes evidence on buying and selling by minorities. The last section concludes.

[2] See also Giglio, Kelly, and Stroebel (2021) for a review of the literature on climate finance.

# 2 Data and methodology

## 2.1 Data sources

**Corelogic Deed & Tax Records.** For residential properties, we use the Corelogic Deed & Tax record data on housing transactions and property features. The sample covers transactions of US residential properties between 2000 and 2020. We restrict the sample to single-family residences, residential condominiums, duplexes, and apartments. For our very granular analysis, the property's exact location is of utmost importance. Therefore, we exclude observations with missing block-level latitude and longitude data. Furthermore, we exclude those with missing information on the sale amount or year in which the property was built. We only keep transactions in which Corelogic recorded the buyer purchased the property in cash or via a mortgage, thus excluding non-arm's length inter-family transactions or investor-recorded purchases.

Figure 1 shows the geographical dispersion of real estate transactions in our data. More precisely, we calculate the number of transactions by county over the sample period, and based on these we sort counties into into five ordered bins, from the ones with the least to the most transactions. As expected, there tend to be more transactions in counties located on the East and West Coasts and those bordering the Great Lakes.

[Insert Figure 1 here]

**Toxics Release Inventory.** Our second main data source is the Toxics Release Inventory (TRI) data of the Environmental Protection Agency (EPA). Firms that satisfy several criteria must report their emissions to the EPA. The criteria for reporting are: (i) the number of employees (at least 10); (ii) the industry sector where the facility operates (some NAICS codes are covered); (iii) the manufacture, production, or use of TRI listed chemicals; and (iv) the facility exceeds at least one of the thresholds for a chemical or a chemical category. When these four criteria are met, the facility must report its emissions of several chemicals to the EPA. The TRI data includes information on the latitude and longitude of each plant. We use it to merge the plant and property transactions data and to calculate the distance between each residential property and plant using the Vincenty (1975)'s formula.

The timing of the reporting of the TRI data is as follows. From January to June, the facilities prepare and submit reporting forms for the previous calendar year. In mid-July a preliminary dataset becomes available, and after some ongoing processing and data analysis, a complete national dataset becomes available in October. We use these complete datasets in our analysis covering the period 2000 to 2020.[3]

Our paper uses variation introduced by facilities reporting for the first time to the EPA the release of carcinogen toxins into the environment.[4]

Our algorithm for selecting these facilities is as follows. The starting year of our sample is 2000. For each subsequent calendar year, we construct an indicator for those facilities and the year in which they first report emitting carcinogen air pollutants. More precisely, the treated facilities are those with flags for the emission of harmful pollutants classified as such under the Clean Air Act and as a carcinogen by the Occupational Safety and Health Administration (OSHA). Plants that already satisfied these criteria in the year of 2000 are excluded since we do not know whether this is the first year in which they did so. And our identification relies on comparing house prices before and after the *first* reporting year in a narrow geographical area around the plants. This comprises a total of 14,732 unique facilities.

**National Establishment Time-Series.** A potential confounder for our estimates comes from those facilities that open in the same year in which they first satisfy our emissions criteria for inclusion. In these cases, we do not know whether the effects arise due to the opening of the facilities themselves or the emissions, the effects of which we aim to estimate. To address this issue, we use a third data source, the NETS dataset provided by Walls & Associates and Dun and Bradstreet. Rossi-Hansberg, Sarte, and Trachter (2021) show that NETS data compares favorably with Census data in terms of quality and coverage.

The NETS data includes the year of establishment of individual lines of business. We exclude from our sample those facilities that opened in the same year as the year in which they report violating the emissions criteria. Our final treated sample comprises 11,143 unique facilities submitting for the first time a report of carcinogen air pollutants in a year different

[3]Further details are available here.

[4]There may be less measurement error in the reporting flag than in the estimated amounts of emissions which reflect differences in companies' estimation methodologies both over time and in the cross-section.

from the opening year.

It is important to clarify that a large proportion (of roughly 92%) of these facilities report for the first time in the same year that they first report emitting carcinogenic air pollutants. In other words, our sample only includes 8% of facilities that previously reported emitting harmful pollutants that were not classified as carcinogenic, and then in the event year report the emission of carcinogenic pollutants. This means that our estimates will mostly capture the *joint* effects of a new reporting of harmful pollutants and of pollutants that are carcinogenic in nature.

A further advantage of the NETS data is that it includes annual revenues and employment information at the facility level (from 1990 to 2020). It allows us to study whether there is a relationship between emissions and plant activity.

Figure 2 shows the location of the plants included in our analysis. Comparing it to the figure for real estate transactions, we see that there is a significant geographical overlap between the two. Many plants are located in counties with many real estate transactions.

[Insert Figure 2 here]

The fact that treated plants are those that first meet the emissions criteria means that there is sample selection that we need to consider both in the analysis and in the interpretation of the results. For instance, plants might be meeting the criteria for the first time because of increased production levels with positive effects on the wages of plant workers and the local economy, which in turn has implications for local house prices. We will directly test this hypothesis.

**Air quality monitor.** The EPA manages the Air Quality System (AQS) data providing information on ambient air quality across the US. The AQS data are collected by a network of over 10,000 monitoring stations located throughout the United States and measuring various pollutants, including ozone, particulate matter, carbon monoxide, nitrogen oxides, sulfur dioxide, and lead. These data are collected on an hourly or daily basis, depending on the pollutant being measured. Moreover, the data are publicly available, permitting individuals and communities access to information on air quality in their local area.

In our analysis, we focus on hazardous air pollutants (HAP), also known as toxic air pollutants, 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." We extract readings from all air monitoring stations that are within a five-mile radius of the plant.

**RSEI Geographic Microdata.** We rely on EPA's model estimates to examine granular changes in cancer risk in the immediate vicinity of the plant. The EPA models toxic releases from the TRI program through the Risk-Screening Environmental Indicators (RSEI) to evaluate and prioritize potential risks to human health and the environment. The data draw on information from the TRI program on chemical releases into air, water, and soil and model their potential location-based health impacts on the population exposed to these chemicals.

We use data at the most granular spatial unit – grid cells of dimension 810m $\times$ 810m. We observe the cancer risk score for each grid cell and plant, a unitless measure computed for each chemical using the estimated dosage released by that specific toxic plant, the toxic concentrations, and potentially exposed populations. Moreover, the data also includes the risk-related non-cancer scores, toxic concentration of chemicals for release by unit and media, and the number of people in the grid cell who are potentially exposed.

## 2.2 Identification

One of the main challenges is identifying the effects of environmental health risks on house values separately from other effects, such as those arising from local economic activity. Facilities with higher production levels are more likely to emit more pollution and to employ more workers. In other words, we need to separate the pollution and local economic activity effects. The latter may have positive effects on house values. We address this identification challenge by exploiting the location of plants and properties at a very granular level.

For each plant, we identify the houses within its immediate vicinity, namely those within a 3-mile ring of the facility and those in a ring of between three and five miles from the same facility. The former set of houses is our treated group, and the latter is the control group. The idea is that emissions will have more of an impact on properties located closest to the toxic plant, whereas all properties located within the 5-mile radius will benefit from local economic activity effects. Several papers in the literature use this "ring" method for identification purposes (Butts, 2022, Diamond and McQuade, 2019, Ganduri and Maturana, 2022, LaPoint, 2022).

### 2.2.1 Defining ring sizes and cancer risk

Our choice of a benchmark ring size of 3 miles for the treated group is based on three observations. First, the environmental justice and medical research literature have used radii ranging from 100 yards (Sheppard, Leitner, McMaster, and Tian, 1999) to 3 miles (Perlin, Wong, and Sexton, 2001, Mohai and Saha, 2006).[5] Second, we rely on data from air monitoring stations on hazardous air pollutants collected by the EPA. Specifically, we focus on the six most carcinogenic pollutants and examine heterogeneity in their emissions as a function of the distance between the plant and air monitors within a 3-mile radius. The concentration of carcinogenic toxins in the area surrounding the plant is fairly high up until a 2.5-3 miles distance (depending on the chemical).

This relation can be visualized in Figure 3 plotting the fitted values and 95% confidence intervals for each of the six pollutants.[6] For most chemicals, the monitor readings show that pollution is highest up to a 1-mile distance from the plant. Given this, we also estimate smaller rings of treated properties (keeping the control group the same, i.e., within 3 to 5 miles of the plant). A smaller treated ring means fewer housing transactions but better identification.

[Insert Figure 3 here]

Third, we use EPA's modeled estimate of cancer risk as a function of distance from the plant and find a higher incidence of such risks manifesting within the three-mile ring when compared to 3 and 5 miles from the facility. Figure 4 shows, for example, the heatmap by grid cells (810m $\times$ 810m grids) for the RSEI cancer scores surrounding Schuff Steel Company in Stockton, CA 95206, in 2011.

[Insert Figure 4 here]

[5]For instance, Whitworth, Symanski, and Coker (2008) have shown that children who resided within a distance of 2 miles from the Houston ship channel were at a 56 percent increased risk of developing acute lymphocytic leukemia when compared to children living more than 10 miles from the channel.

[6]The fitted values are generated using median regressions of emissions on a fifth-order polynomial of the distance between the monitors and the operating toxic plant. Confidence intervals are estimated using bootstrap replications with replacement. We keep the size of the resampled data to be same as the original data to capture the sampling uncertainty of the original sample.

We also test the relation between cancer risk and distance from a facility more formally using a regression analysis. For a plant $j$ and a distance bin $b$ (representing the distance from the plant), we compare the incidence of cancer risk, as measured by the RSEI cancer score, the year before the first time a plant reports the emission of carcinogenic toxins and the year after (excluding the year of the emission event, year 0) by estimating the following equation:

$$\begin{aligned}\text{RSEI cancer}_{jbt} = \alpha + \beta_{Distance} \times \mathbb{1}_{jb}^{\text{Distance}_{jb}<X\text{miles}} + \\ \beta_{\text{Post}\times\text{Distance}} \times \text{Post}_{bt} \times \mathbb{1}_{jb}^{\text{Distance}_{jb}<X\text{miles}} + \gamma_{jt} + \epsilon_{jbt},\end{aligned} \tag{1}$$

where $\text{Post}_{bt}$ takes the value of one for if the RSEI cancer score was computed for the year after the event year (and zero otherwise) and $\mathbb{1}_{jb}^{Distance_{jb}<Xmiles}$ takes the value one if RSEI score refers to an area within $X$ miles of plant $j$, with $X = 3$ in the baseline regressions (and zero if it is between 3 and 5 miles).[7] The above equation includes plant-by-year fixed effects to control for time-varying characteristics of the plant and macroeconomic conditions in the area where the plant is located.

Table 1 reports the results. Compared to the control group, We find a higher cancer risk in the treated area closer to the plant and a very large increase in cancer risk in the year after the emissions event in the same area. The fact the the estimated coefficients decrease as we decrease the size of the treated ring may at first come as a surprise. However, this is simply due to the fact the very granular RSEI cancer score measures take into account the size of the population affected which decreases as we consider narrow geographical areas around the plant. We test the robustness of this relation using a Poisson pseudo-maximum likelihood regression, given the issues with OLS when there are a lot of zeroes in the data (Cohn, Liu, and Wardlaw, 2022). Results are indeed robust and reported in Table A.1.[8]

[Insert Table 1 here]

[7] For each plant, we aggregate the RSEI cancer scores at the grid-level to distance bins, allowing us to compare changes in cancer risk for the same plant as a function of distance.

[8] We have also checked the robustness of our estimates to excluding plants that do not report any carcinogenic emission in a particular year. Table A.2 finds that the results are robust and consistent with Table 1.

### 2.2.2 Identifying house value effects

Having shown how carcinogenic risk changes in the area closer to a plant after the year in which it first reports the emission of carcinogenic toxins to the EPA, we now move to our baseline regression to evaluate changes in property transactions and the impact on house values. We focus on transactions that took place in the calendar year before the event (year -1), in the calendar year of the emissions event (year 0), and in the year after (year +1). We let $i$ denote the property, $j$ the toxic establishment matched to property $i$, $c$ the county where the property is located as identified by Federal Information Processing Standard (FIPS), and $t$ the year of the property transaction. The equation that we estimate is:

$$\begin{aligned}\log(\text{Sale amount})_{ijct} = {} & \alpha + \beta_{Post} \times \text{Post}_{it} + \beta_{\text{Post}\times\text{Distance}} \times \text{Post}_{it} \times \mathbb{1}_{ij}^{Distance_{ij}<X miles} + \\ & \gamma_j + \gamma_{ct} + \epsilon_{ijct},\end{aligned} \quad (2)$$

where $Post_{it}$, is an indicator variable taking a value of one if property $i$ is sold in the year $t$ after the event year and zero otherwise. We define $\mathbb{1}_{ij}^{\text{Distance}_{ij}<X\text{miles}}$ to take a value of one if property $i$ is within $X$ miles from a plant $j$, with $X = 3$ in the baseline regressions, and zero for properties between 3 and 5 miles of the same plant. The above equation includes establishment fixed effects that control for time-invariant characteristics of the establishment and year times county fixed effects that control for time-varying macroeconomic conditions in the county where the property is located.

In addition to the above equation, we use the sample of repeated transactions of the *same* properties to estimate:

$$\begin{aligned}\log(\text{Sale amount})_{ijct} = {} & \alpha + \beta_{Post} \times \text{Post}_{it} + \beta_{\text{Post}\times\text{Distance}} \times \text{Post}_{it} \times \mathbb{1}_{ij}^{Distance_{ij}<X miles} + \\ & \gamma_i + \gamma_{ct} + \epsilon_{ijct},\end{aligned} \quad (3)$$

where we now include property fixed effects ($\gamma_i$). This specification controls for time-invariant property characteristics and provides within-property estimates of the effects of air pollution on property values. But it also means that properties need to be transacted at least twice in the space of three years, including at least once in year +1, to be included in the regressions.[9]

[9]For properties that are transacted twice in year +1, in the regressions we use the value from the last transaction.

Therefore, the use of repeated sales over a narrow time window helps for identification, but it does mean that there is sample selection in the properties included.

Note that the empirical specification is akin to a difference-in-difference-in-differences (DiDiD) strategy that exploits variation in the first reporting year of harmful carcinogenic pollutants by the toxic plant and distance of the property to the plant. Thus, at any point in time, the treated properties are those within $X$ miles of plant $j$, with $X = 3$ in the baseline regressions, and control properties are those between 3 and 5 miles of a toxic plant. The parameter of interest is $\beta_{\text{Post} \times \text{Distance}}$, which measures the *within* property changes in the sale amount from one year before to one year after, conditional on the set of fixed effects.

The sample of properties that are transacted at least twice in three years may over-sample transactions by flippers, i.e. buyers who acquire run down properties, fix them up, and then sell them in a short period of time. This means that that increases in property prices may at least partially reflect the value arising from property improvements. While our data does not allow us to measure such improvements, we try to address the issue by performing robustness, including providing estimates for wider windows, dropping from the sample properties with the largest changes in value between transactions, and dropping from the sample properties that were transacted multiple times in the same calendar year.

Due to the geographical concentration of some of the facilities in our sample, there are instances in which properties are located within a 5-mile radius of a treated establishment, and in a later year, the same property is also located in a 5-mile radius of a different treated establishment. In these instances, we include in the regressions only the property observations corresponding to the first event. This avoids having multiple observations for the same property, potentially in both treatment and control groups. Furthermore, we match properties with the closest establishment when a property is located within 5 miles of multiple establishments that first satisfy the reporting criteria in the same year.

## 2.3 Summary statistics

We provide summary statistics on the toxic plants in our sample and on the counties in which they are located. Panel A of Table 2 shows the industry coverage of the 11,143 unique facilities in the event year. More precisely, it shows the three-digit NAICS industry codes with the corresponding fraction of toxic plants belonging to that industry. The largest proportion is in fabricated metal products. Panel B reports the offending chemicals: lead is by far the worst offender, in roughly 53% of the plants, followed by nickel (17%).

[Insert Table 2 here]

Table 3 shows the top ten carcinogenic chemicals ranked by toxicity. Toxicity is calculated as the product of air emissions times inhalation toxicity weight summed over all toxic plants in our sample. The third column of the table reports the corresponding frequency of plants responsible for the chemicals releases. The final column reports the inhalation unit risk expressed as the upper bound excess risk of developing cancer over a person's lifetime that can be attributed to ongoing exposure to a substance at a level of 1 gram per cubic meter. While the inhalation risk from asbestos is highest at 170000, the frequency of plants emitting is relatively small equal 0.06 percent. Chromium and chromium compounds have the second highest inhalation risk and they are fairly frequent among the plants in our sample.

[Insert Table 3 here]

Table 4 compares the average values for several characteristics across counties with treated plants and counties that were never treated (i.e. counties that do not include any of the 11,143 treated plants). More precisely, we compute the average value of the county characteristics in the year of treatment and compare them to average value in the same year for the never treated counties.

Counties with treated plants tend to have larger gross domestic product (GDP) but lower per capita GDP, although the difference in the latter is not statistically significant. Treated plants tend to be located in counties with lower unemployment rate, but the difference relative to the never treated group is only marginally significant. Treated counties tend to experience lower house price growth. They also tend to have a larger proportion of Black population.

[Insert Table 4 here]

# 3 Baseline results

## 3.1 All property transactions

The dependent variable in the regressions is the natural logarithm of the sale amount of the property. Table 5 panel A shows the estimation results when we include all property transactions that took place during the event window (and not only repeated transactions). The regressions include plant fixed effects (in addition to year times county fixed effects). Below the estimated coefficients we report robust standard errors clustered at the county level.

The estimated coefficient on the interaction term in column (1) is negative and statistically significant. It shows that the prices of properties within the 3-mile radius decrease by 6.3% relative to those between 3 and 5 miles. In the remaining columns of the table, we decrease the size of the ring of the treated group from 3 to 2 miles and then up to 1 mile from the reporting facility. The estimated declines in the value of treated properties relative to the control group increase as we restrict properties in the treated group to those located nearer to the plant. The relative decline is largest and equal to 12.4% for treated properties within one mile.

The estimated coefficients on the post variable in Panel A of Table 5 are positive and statistically significant, with estimates ranging between 2.2% and 0.8%. This means that property prices around the facility (in a 5-mile radius) tend to increase in the year after treatment, which could be due to improved economic activity from the plant that benefits the local area. We will test this hypothesis more directly below.

[Insert Table 5 here]

The estimates in Panel A of Table 5 refer to the full sample of house transactions surround unique facilities located in the different US states. A question of interest is whether the results are driven by a few states or event years, or whether they hold more generally across states and over time. To address this question, we split our data into sub-samples, first by state and then by event year, and estimate separate regressions for each sub-sample, including establishment fixed effects as before and focusing on the 1-mile treated group ring.

Figure 5 plots the estimated coefficients on the Post $\times$ $\mathbb{1}^{Distance<1\ mile}$ variable for the different states. For the ones shown in grey, there are not enough data to perform the analysis. For almost all states the estimated coefficients are negative, indicating that our results are not driven by a small subset of them. For the five states that are the focus of Currie, Davis, Greenstone, and Walker (2015) (i.e., Texas, New Jersey, Pennsylvania, Michigan, and Florida) the estimated coefficients are lower than -10%.

[Insert Figure 5 here]

Figure 6 plots the estimated coefficients on the Post $\times$ $\mathbb{1}^{Distance<1\ mile}$ variable by event year. The estimates are obtained by splitting the sample of plants by the year in which they first reported the emission of harmful carcinogen pollutants. For most years the estimated coefficients are statistically negative (the standard errors are clustered at the county level) and economically meaningful. Furthermore, the figure shows that the results are not driven by a small number of years.

[Insert Figure 6 here]

## 3.2 Repeated sales

We now focus on the sample of repeated housing transactions, i.e. those properties that were transacted at least once in the year before or in the event year and once in the year after. Appendix Figure A.1 compares the distributions of sale prices for all transactions (those corresponding to the roughly 7.5 million observations in column (1) of Panel A of Table 5) and of the repeated transactions (roughly 1.1 million observations). The shapes of the distributions are similar, but the sample of repeated transactions has a slightly larger proportion of properties transacted at below the median prices.

Panel B of Table 5 shows the estimation results for the sample to repeated transactions, The estimated coefficients of the interaction term share a similar pattern. Prices of properties within the 3-mile radius decrease by 6.6% relative to those between 3 and 5 miles. When decreasing the size of the ring of the treated group from 3 to 2 miles and then up to 1 mile from the reporting facility, the estimated declines in the value of treated properties relative to the control group again increase. The relative decline is largest and equal to 11.2% for treated properties within one mile of the plant.

The major difference relative to Panel A are the positive and economically larger estimated coefficients for the Post variable, ranging between 7.9% and 6.8%. Therefore, properties in the control group increase in value in the year after treatment. This can be driven by the joint effect of improved economic activity in the local area around the plant in the year after treatment and sample selection in the houses that are repeatedly transacted during the three years event window.

We can also control for local economic conditions more narrowly defined using plant-by-year fixed effects (instead of county-by-year fixed effects). In this case we cannot include the Post variable in the regressions. Appendix Table A.3 reports the results. The estimation coefficients on interaction term for the different treated rings size are very similar to the ones in Panel B of Table 5.

Sample selection may play an important role in explaining the differences in estimated coefficients on the post variable between Panels A and B of Table 5. After all, the sample of properties with multiple transactions over a relatively short three years time window may over-sample properties that were bought "cheap" because they were run down and then sold after renovations. In Appendix Figure A.1 we plot the cumulative distributions of house sale amount of the properties included in the regressions of Panel A (all properties) and in Panel B (repeated). The distribution of repeated transactions has a slightly larger mass on below-median values (the median is about 12 for both samples), consistent with the hypothesis that a larger proportion of these properties were bought for relatively lower values,

The advantage of using the repeated sample is that we are able to include property fixed effects in the regressions and provide within property estimate, i.e., the effect on transaction price post event year (year +1) compared to the price of the same property on or before the event (in years -1 or 0). Table 6 shows the results. The estimated coefficients on the post variable are positive but now not statistically significant. The estimated coefficients on the interaction term are negative, so that the price of properties closer to the plant decrease relative to those in the control group, but the (absolute) values of the estimated coefficients are an order of magnitude smaller than before. For instance, for treated properties in a one-mile radius, the change is -1.6% compared to -12.4% in Panel A of Table 5 where we include plant fixed effects.

[Insert Table 6 here]

These results show that unobservable property characteristics matter for the magnitude of the estimated effects, likely driven by differences in the properties that are included in the treatment and control groups.

With such a short time event window, property fixed effects should control for all the property features that remain constant in these 3 years and address all the unchanging variations in the data (the $R^2$ increases from 43% in Table 5 to 86% in Table 6). Adding property-fixed effects lowers our estimates from 11% in column 5 of Table 5 panel B to 1.6%.

For the diff-in-diff methodology to be valid, the parallel trends assumption must be satisfied. Figure 7 plots the estimated coefficient on Post $\times$ $\mathbb{1}^{Distance<3miles}$ in event time and the corresponding 95% confidence intervals. The estimates are normalized to time zero and the standard errors are clustered at the county level. Prior to the event, the estimated coefficients are not significantly different from zero, neither economically nor statistically, satisfying the parallel trends assumption. After the event, the estimated coefficient becomes negative and statistically significant.

[Insert Figure 7 here]

### 3.3 Robustness

The previous estimates are for a tight (-1,1) time event window, which helps with identification, but also reduces the number of properties included in the estimation. Moreover, for our within-property estimates, we need properties to be transacted at least twice in the event window, once before (in years -1 or 0) and once after treatment (year +1), which reduces the sample even further.

In the first five columns of Table 7 we expand the event window to (-2,1). That is, we now use properties that are transacted at least once in the two years prior to the event and once in the year after. For all properties that are transacted more than two times in the two years prior to the event, we include all the observed transactions in the regressions. In the different columns, as before, we vary the size of the ring of treated properties. The estimated coefficients vary between -1.2% for a treated property ring of 3 miles and -1.4% for a ring of 1.25 and 1 mile. These values are similar to the ones reported in Table 6.

In the last five columns of Table 7 we increase the width of the event window to (-3,1) with the caveat that for a wider window, the effects of the toxic plant event may be confounded by other events. However, we observe estimated effects that are still negative and similar in economic magnitude to our previous estimates.

[Insert Table 7 here]

In a second robustness exercise we restrict the sample to those events for which we have at least one hundred housing transactions. This means an average of twenty five transactions in the treated and control groups in the pre/post periods. Appendix Table A.4 shows that the results are robust with estimated house price effects in the treated group of -1.4% (for the 3 miles ring). In Appendix Table A.5 we show results for regressions with house prices defined in levels instead of in logs. The estimates imply house price declines of between roughly two and a half and six and a half thousand dollars depending on the size of the treated ring.

# 4 Heterogeneous effects on property prices

## 4.1 Baseline results

The previous results showed that a plant that starts reporting the emission of carcinogenic toxins into the environment has negative effects on the price of nearby houses, but that the price effects are on average relatively modest, or around -1.5% for the treated group relative to the control group.

We now evaluate whether there are heterogeneous effects that depend on how expensive the properties are. For all the properties included in the repeated transactions sample, we create a dummy variable equal to one if the property was transacted in the years before the event for a price above the median property price by plant-year and zero otherwise. We then estimate regressions where we interact the new dummy with the post variable and the indicator for the distance from the facility.

Table 8 shows the results. The estimated negative coefficient on the Post $\times$ $\mathbb{1}_{\text{Distance} < \text{ X miles}}$ variable in the first column shows that the value of those houses located within three miles of the facility increased in the year after the treatment by 10.6% compared by those properties in the control group. The estimated coefficient in the bottom row shows that there is an additional -29.6% effect for properties above the median of the ex-ante price distribution. This implies an overall decline of about 20% for the more expensive properties located within three miles of the treated plants.

[Insert Table 8 here]

In the remaining columns of the Table 8, we vary, as before, the size of the ring of the treated group. The decline in the value of expensive properties relative to the control group is fairly stable across the different specifications. These results are important since they provide a house value channel through which different households may sort into neighborhoods with varying degrees of exposure to pollution.

## 4.2 Robustness

As before, we repeat the estimation for the repeated sample using a longer time horizon, i.e. event windows (-2,1) and (-3,1). The sample requirements is that the same property is transacted at least once before the event date 0 and at least once after the event date. For this expanded event windows, it means at least two transactions in four and five years respectively.

Table 9 shows the results. Focusing for instance on the first column, we see that the value of above median properties within 3 miles of the plant decreases by an average of 24.8% relative to those in the control group, and that they experience an overall decline in value of 14.3% (adding the estimated coefficients in column (1) of the table). Therefore, the main conclusions are similar.

[Insert Table 9 here]

In the Appendix we carry out several robustness exercises. If more expensive properties, i.e., those above the median concentrate geographically, most of them can be in the treated or control groups. When we run the regressions the estimates will be based on a small number of observations in the other group. This could possibly affect the estimated coefficients. In Appendix Table A.6 we show that our results are robust to defining above/below median within ring. The results are also robust to regressions in which we restrict the sample to those events for which there are at least one hundred house sales (Appendix Table A.7).

Our estimates are within property. However, we do not observe property improvements that have been made by the owners between the transactions, which could be the source of some of the observed price increases. For this to affect our estimates, it would to be the case that such improvements are more likely to take place for the lower price properties (which may be likely), but also that such properties are relatively more likely to be located in the treated ring (which may be less likely).

In order to address this concern, we estimate regressions in which we drop from the sample the properties for which we observe the largest changes in prices, both positive and negative. Some of these large negative house price declines are sold by financial institutions, suggesting that they result from foreclosures. In Appendix Table A.8 we shows results the cases in which we drop the top/bottom 5% (and then 10%) of the observations. Although naturally the estimated values change, the main conclusions remain the same. The results in this section are evidence of distributional impact of the event.

## 4.3 Economic benefits

The previous tables estimated a positive price effect for below median values in the post period. A potential explanation is that the polluting event is positively correlated with enhanced economic activity by the offending facility with economic benefits for the surrounding area, valued more by purchasers of lower value properties. In order to investigate this hypothesis we have obtained employment and sales data for the plants in our sample in the years surrounding the event. The regression that we estimate:

$$\log(\text{employment})_{jt} = \alpha + \beta_{Post} \times \text{Post}_{jt} + \gamma_j + \gamma_t + \epsilon_{jt}, \tag{4}$$

where $j$ denotes the treated facility, $t$ time, and $\text{Post}_{jt}$ takes the value of one for treated facilities in the year after the event year (and zero otherwise, i.e. on or the year before the event year). The equation is within facility, meaning that it captures changes in employment at the facility in the year after the event compared to just before the event. The equation also includes year fixed effects that control for aggregate economic conditions. We estimate similar regressions for (log) sales.

The estimates in column (1) of Table 10 show that employment at the plant increases by roughly 2% after the event year compared to just before. We also find a positive 1.4% increase in sales, although the estimate coefficient is not statistically different from zero. These estimates show that the polluting event is positively correlated with an event that has economic benefits for those in the local area, namely an increase in employment.

[Insert Table 10 here]

## 4.4 Buying and selling by minorities

Next, we use the granularity of the data to study buying and selling by minorities. Specifically, we use names of buyers and sellers to understand whether new information about carcinogenic emissions affects the identity of buying and selling in a three-mile ring around the plant compared to a three-to-five-mile ring in the event window (-1,+1). We use the algorithm proposed by Laohaprapanon, Sood, and Naji (2022), which exploits data from the US census to predict race and ethnicity based on individual buyers' and sellers' last names. Using this classification algorithm, for each property transaction, we can predict the ethnicity of buyers and sellers in our sample, focusing on Hispanic due to the higher accuracy of the algorithm.[10]

For the test, we define $\mathbb{1}$(Hispanic) to take the value of one if the predicted probability that an individual is of Hispanic ethnicity is greater than 85%. Panel A of Table 11 examines changes in buyer ethnicity using the predicted probabilities. We control for establishment and county times year fixed effects in these regressions. The estimated positive coefficient on the Post $\times$ $\mathbb{1}_{\text{Distance}< \text{ X miles}}$ shows a relative increase of 1.2 percentage point in the Hispanic home buyers in the immediate vicinity of the treated plant in the year after the plant first reports the emission of carcinogenic toxins compared to the control group. The negative coefficient on Post indicates a drop in the fraction of buyers predicted to be Hispanics in the control group after the event.

[Insert Table 11 here]

In Panel B of Table 11, we focus on the predicted ethnicity of home sellers. The empirical specification is similar to before with the exception that the dependent variable, $\mathbb{1}$(Hispanic), takes the value of one if the predicted probability that an individual seller is of Hispanic ethnicity is greater than 85%. The results show a relative increase of 0.8 percentage point of Hispanic sellers in the immediate vicinity of the treated plant compared to the control group, which instead experiences a drop in the fraction of sellers predicted to be Hispanic.

We check the sensitivity of the estimates in the use of the name classification algorithm and assuage two specific concerns. First, in Table A.9, we increase the minimum predicted probability above which an individual is classified as "Hispanic" to 90%. The results are

[10] For this classification, we use the last names of all sellers and buyers who are individuals. In our sample, for buyers (sellers), we can predict race and ethnicity for 79% (60.2%) of all transactions.

qualitatively similar. Overall, these results provide the first evidence of *granular* changes in neighborhood composition in a short window emanating from house prices around toxic plants.

# 5 Conclusion

This paper investigates the impact of environmental health risk on the valuation of residential properties in the United States. Focusing on a narrow window around the time in which a facility first reports the emission of carcinogenic toxins to the EPA, we show that the values of houses closer to the facility drops by about 2 percentage points more than the values of houses farther from the plant. These average effects, albeit small, mask considerable heterogeneity. Specifically, expensive properties, defined as those above the median price before the reporting event, experience a significant decline in their values between -10% and -20%. In contrast, cheaper properties experience an increase in value by 7% to 10%.

One potential explanation for the heterogeneity is that the polluting event is positively related with enhanced economic activity at the carcinogen-emitting plant and the concomitant economic benefits for the surrounding area are valued more by purchasers of lower-value properties. Consistent with this hypothesis, we find employment at the plant increases by roughly 2% after the event year compared to just before.

Our work sheds light on the trade-off between enhanced economic activity emanating from plants and an increase in environmental health risks. The setting we focus on allows us to hold constant plant siting decisions and isolate costs associated with plants that are newly reported to emit carcinogenic pollutants. Our results imply that changes in house values for expensive properties reflect significant changes in neighborhood composition and have implications for long-run environmental inequality for households.

# References


Baldauf, M., L. Garlappi, and C. Yannelis (2020). Does climate change affect real estate prices? only if you believe in it. *Review of Financial Studies 33*, 1256–1295.

Banzhaf, S., L. Ma, and C. Timmins (2019). Environmental justice: The economics of race, place, and pollution. *Journal of Economic Perspectives 33*(1), 185–208.

Bayer, P., N. Keohane, and C. Timmins (2009). Migration and hedonic valuation: The case of air quality. *Journal of Environmental Economics and Management 58*(1), 1–14.

Bernstein, A., M. T. Gustafson, and R. Lewis (2019). Disaster on the horizon: The price effect of sea level rise. *Journal of Financial Economics 134*(2), 253–272.

Bloom, N., E. Brynjolfsson, L. Foster, R. Jarmin, M. Patnaik, I. Saporta-Eksten, and J. Van Reenen (2019). What drives differences in management practices? *American Economic Review 109*(5), 1648–1683.

Butts, K. (2022). Jue insight: Difference-in-differences with geocoded microdata. *Journal of Urban Economics*, 103493.

Chay, K. Y. and M. Greenstone (2005). Does air quality matter? evidence from the housing market. *Journal of political Economy 113*(2), 376–424.

Clara, N., J. F. Cocco, S. L. Naaraayanan, and V. Sharma (2022). Investments that make our homes greener: The role of regulation. *Working Paper*.

Cohn, J. B., Z. Liu, and M. I. Wardlaw (2022). Count (and count-like) data in finance. *Journal of Financial Economics 146*(2), 529–551.

Currie, J., L. Davis, M. Greenstone, and R. Walker (2015). Environmental health risks and housing values: Evidence from 1,600 toxic plant openings and closings. *American Economic Review 105*(2), 678–709.

Diamond, R. and C. Gaubert (2022). Spatial sorting and inequality. *Annual Review of Economics 14*, 795–819.

Diamond, R. and T. McQuade (2019). Who wants affordable housing in their backyard? an equilibrium analysis of low-income property development. *Journal of Political Economy 127*(3), 1063–1117.

Enrico, M. (2011). Local labor markets. In *Handbook of labor economics*, Volume 4, pp. 1237–1313. Elsevier.

Fuerst, F., P. McAllister, A. Nanda, and P. Wyatt (2015). Does energy efficiency matter to home buyers? an investigation of epc ratings and transaction prices in england. *Energy Economics 48*, 145–151.

Ganduri, R. and G. Maturana (2022). Do property rehabs affect neighboring property prices? *Working Paper*.

Gete, P. and A. Tsouderou (2021). Climate risk and mortgage markets: Evidence from hurricanes harvey and irma. *Working Paper, IE Business School*.

Giglio, S., B. Kelly, and J. Stroebel (2021). Climate finance. *Annual Review of Financial Economics 13*, 15–36.

Giglio, S., M. Maggiori, R. Krishna, J. Stroebel, and A. Weber (2021). Climate change and long-run discount rates: Evidence from real estate. *Review of Financial Studies 34*(8), 3527–3571.

Giroud, X., S. Lenzu, Q. Maingi, and H. Mueller (2021). Propagation and amplification of local productivity spillovers. *Working Paper*.

Greenstone, M. and J. Gallagher (2008). Does hazardous waste matter? evidence from the housing market and the superfund program. *The Quarterly Journal of Economics 123*(3), 951–1003.

Greenstone, M., R. Hornbeck, and E. Moretti (2010). Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings. *Journal of political economy 118*(3), 536–598.

Issler, P., R. Stanton, C. Vergara, and N. Wallace (2020). Mortgage markets with climate-change risk. evidence from wildfires in california. *Working Paper, UC Berkeley*.

Ito, K. and S. Zhang (2020). Willingness to pay for clean air: Evidence from air purifier markets in china. *Journal of Political Economy 128*(5), 1627–1672.

Kermani, A. and F. Wong (2021). Racial disparities in housing returns. *Working Paper*.

Keys, B. J. and P. Mulder (2020). Neglected no more: Housing markets, mortgage lending, and sea level rise. *Working Paper*.

Laohaprapanon, S., G. Sood, and B. Naji (2022). ethnicolr: Predict race and ethnicity from name (version: 0.9.1).

LaPoint, C. (2022). Property tax sales, private capital, and gentrification in the us. *Private Capital, and Gentrification in the US (September 15, 2022)*.

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

Murfin, J. and M. Spiegel (2020). Is the risk of sea level rise capitalized in residential real estate? *Review of Financial Studies 33*, 1217–1255.

Neumark, D. and H. Simpson (2015). Place-based policies, in 'handbook of regional and urban economics', vol. 5.

Office, U. G. A. (1983). Siting of hazardous waste landfills and their correlation with racial and economic status of surrounding communities. *RCED-83-168*.

Ortega, F. and S. Taspinar (2018). Rising sea levels and sinking property values: Hurricane sandy and new york's housing market. *Journal of Urban Economics 106*, 81–100.

Perlin, S. A., D. Wong, and K. Sexton (2001). Residential proximity to industrial sources of air pollution: interrelationships among race, poverty, and age. *Journal of the Air & Waste Management Association 51*(3), 406–421.

Rossi-Hansberg, E., P.-D. Sarte, and N. Trachter (2021). Diverging trends in national and local concentration. *NBER Macroeconomics Annual 35*, 115–150.

Sheppard, E., H. Leitner, R. B. McMaster, and H. Tian (1999). Gis-based measures of environmental equity: exploring their sensitivity and significance. *Journal of Exposure Analysis & Environmental Epidemiology 9*(1).

Vincenty, T. (1975). Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations. *Survey review 23*(176), 88–93.

Whitworth, K. W., E. Symanski, and A. L. Coker (2008). Childhood lymphohematopoietic cancer incidence and hazardous air pollutants in southeast texas, 1995–2004. *Environmental Health Perspectives 116*(11), 1576–1580.

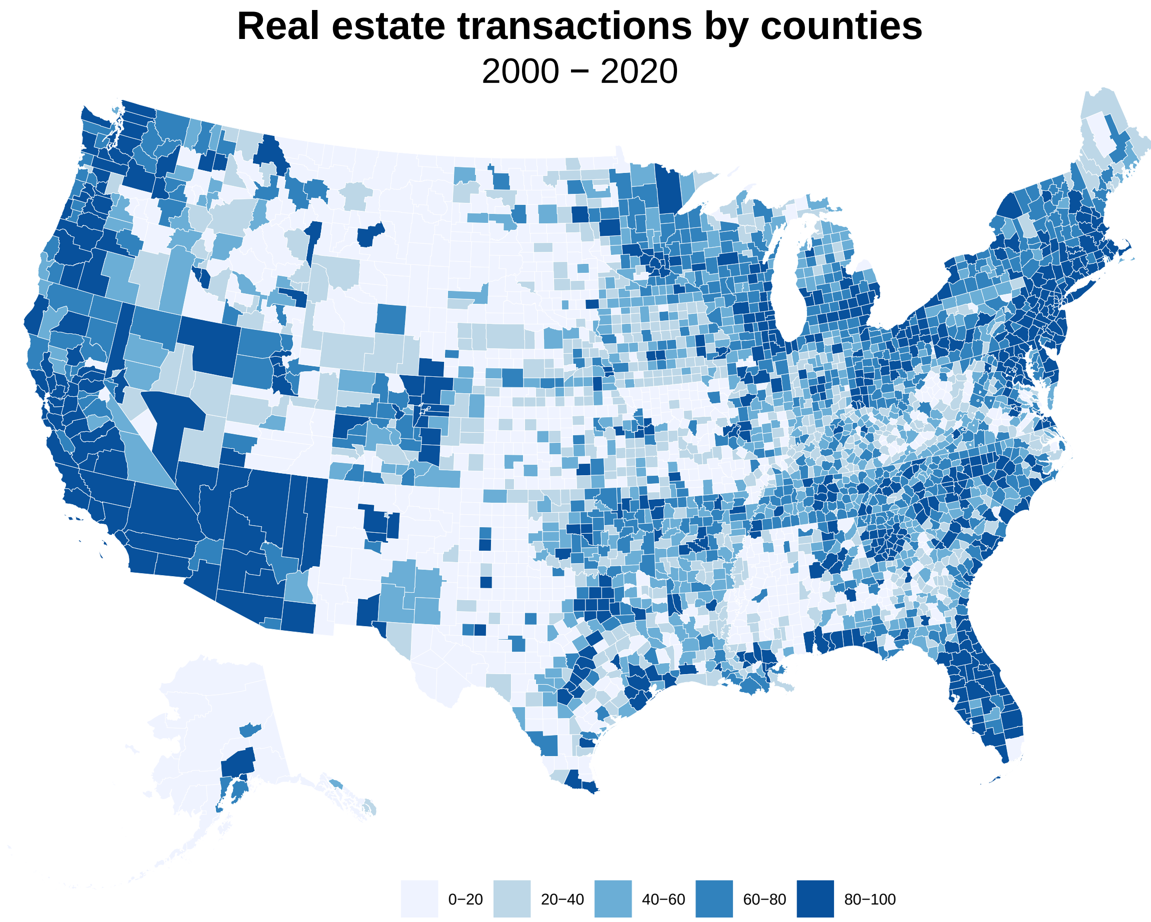


**Figure 1: Number of real estate transactions by county.**
*Notes*: The data are from the Corelogic Deed & Tax record data from 2000 and 2020. We calculate the number of housing transactions by county, and based on these we sort counties into 5 ordered bins, from the ones with the least to the most transactions.

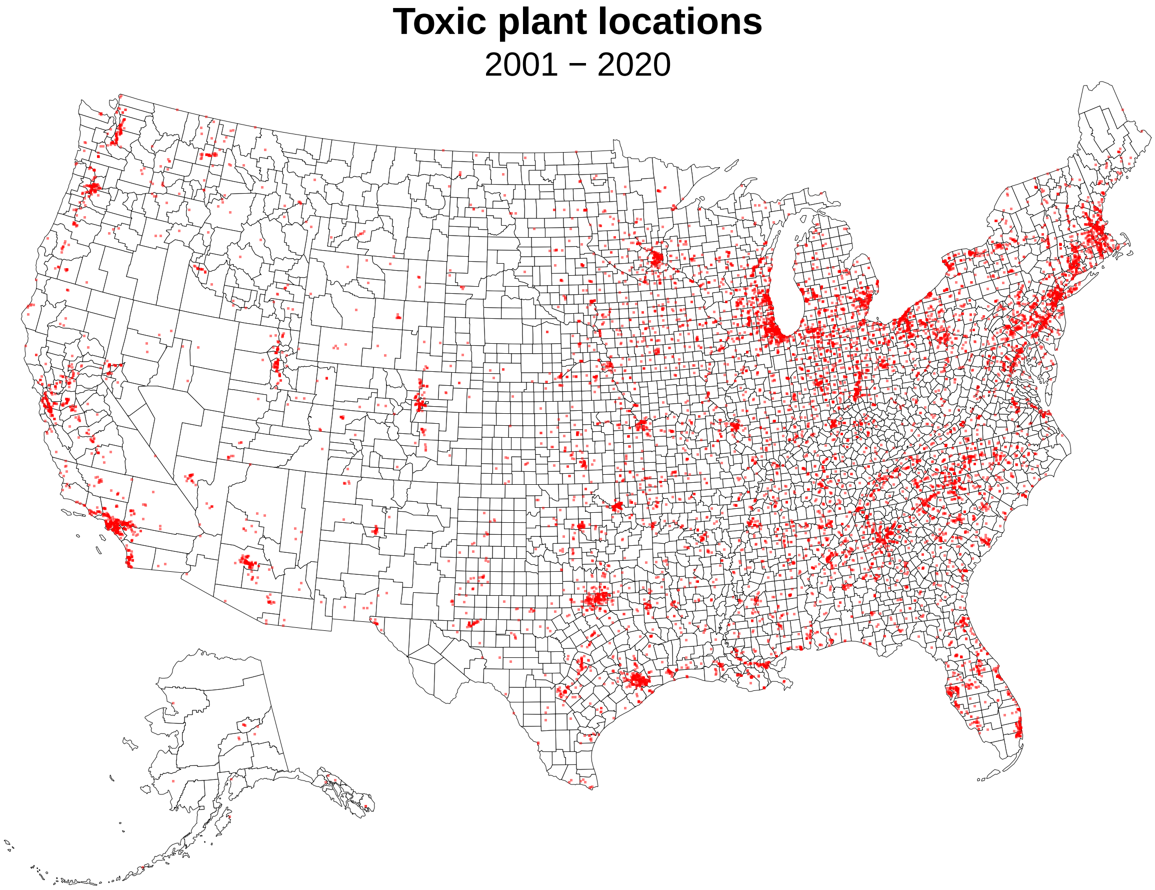


**Figure 2: Location of the reporting facilities.**
*Notes*: The figure shows the location of plants that report the emission of harmful carcinogen pollutants for the first time during the sample period. We exclude all plants for which the first reporting year is the same as the opening year. The data are from the EPA's TRI program between 2001 and 2020.

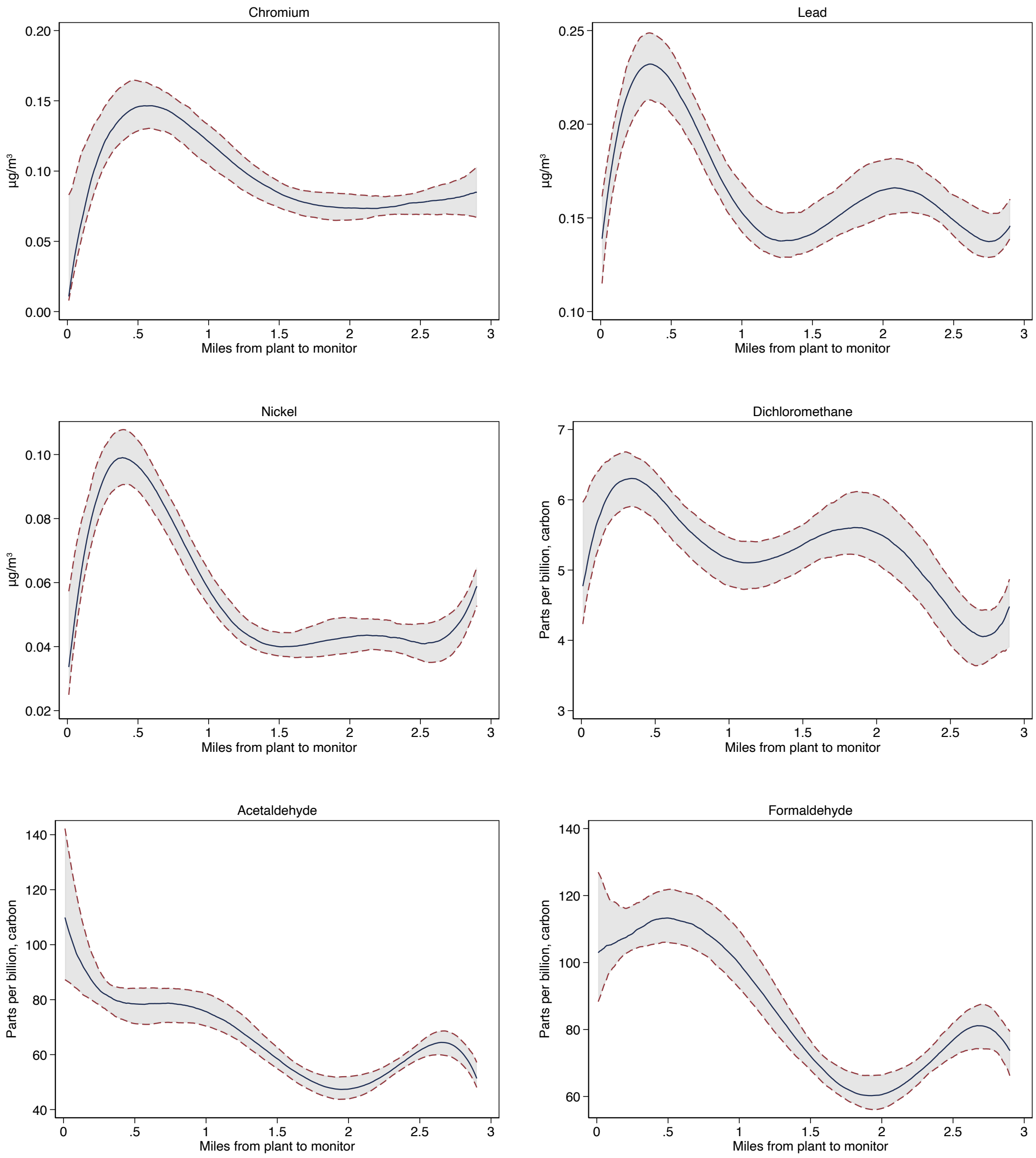


**Figure 3: The effect of toxic plant on carcinogenic air pollution.**

*Notes*: The figure plots the fitted values and 95% confidence intervals from 6 separate median regressions of the concentration of toxic air pollutants in the air on a fifth-order polynomial of the distance from the monitor to an operating toxic plant. Confidence intervals are estimated via 500 bootstrap replications with replacement. The size of the resampled data is the same as the size of the original data, to capture the sampling uncertainty of the original sample. The unit of observation is a monitor-plant-year triad.

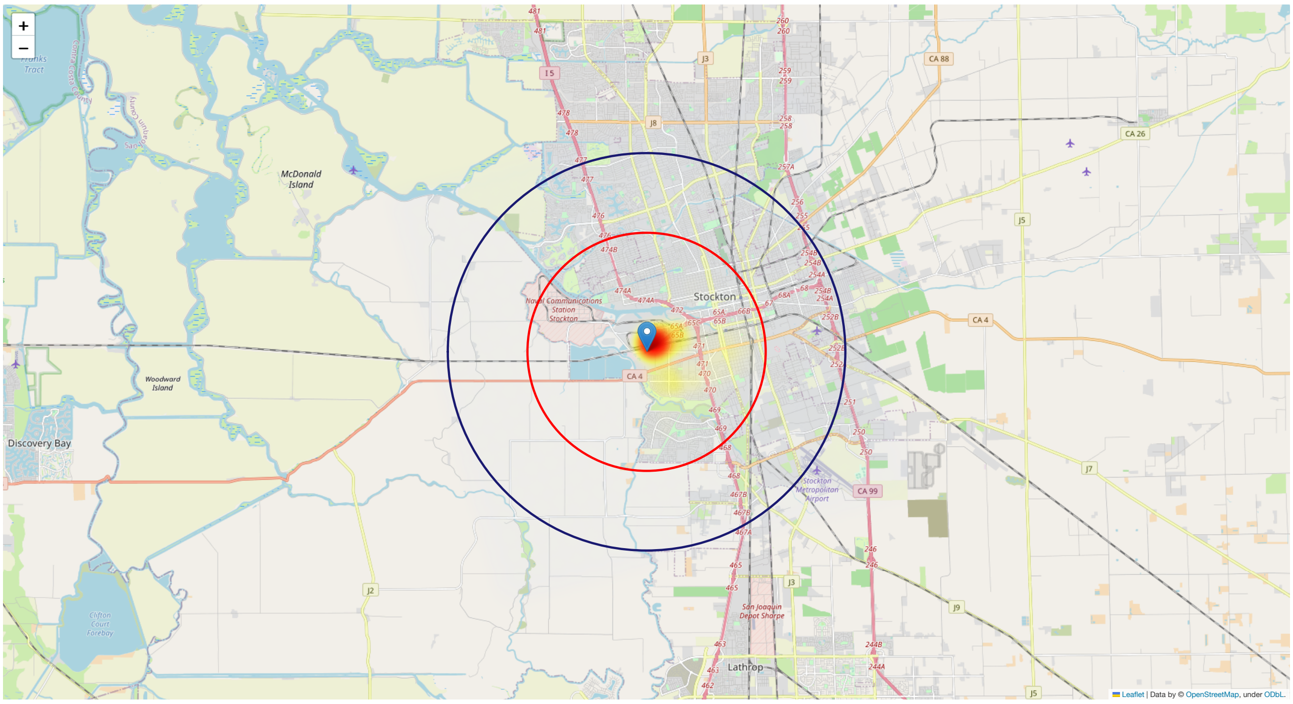


**Figure 4: Heatmap of RSEI Cancer score by grid cells for Schuff Steel Company, Stockton CA.**

*Notes*: The figure shows the heatmap by grid cells (810m $\times$ 810m grids) for the RSEI cancer scores aggregated for toxic chemicals released by Schuff Steel Company in Stockton, CA 95206, in 2011. We obtained disaggregated geographic microdata from the Risk-Screening Environmental Indicators (RSEI). These EPA models the impact of chemical releases from toxic plants on grid cells using estimated dosage, its toxic concentrations, and potentially exposed populations and provides a unitless score (RSEI Cancer score) to capture the effect of chemical releases on cancer. Please see the text for more details. The light blue marker identifies the facility, and darker-colored grid cells show a higher cancer risk. The red circle defines an area with a three-mile radius of the facility, whereas the blue circle defines an area with a five-mile radius of the facility. The figure has been produced using the Folium Python Library and Leaflet maps (http://leafletjs.com/).

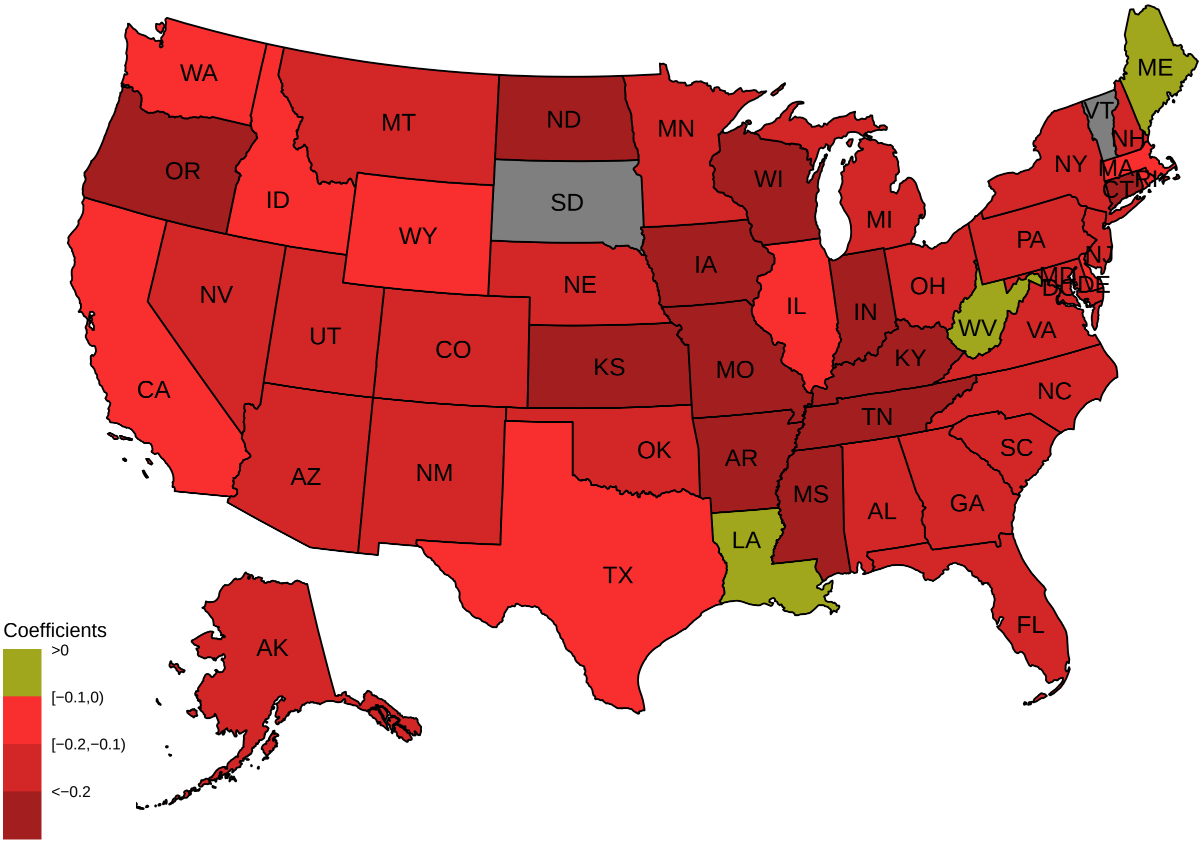


**Figure 5: Estimated housing value effects by state.**
*Notes*: The figure plots the estimated coefficients of the term Post $\times$ $\mathbb{1}^{Distance<1\ mile}$ by state. The estimates are obtained by splitting the sample of plants by the state in which they are located. States shown in grey do not have enough data for the analysis. The empirical specification includes plant and year $\times$ county fixed effects.

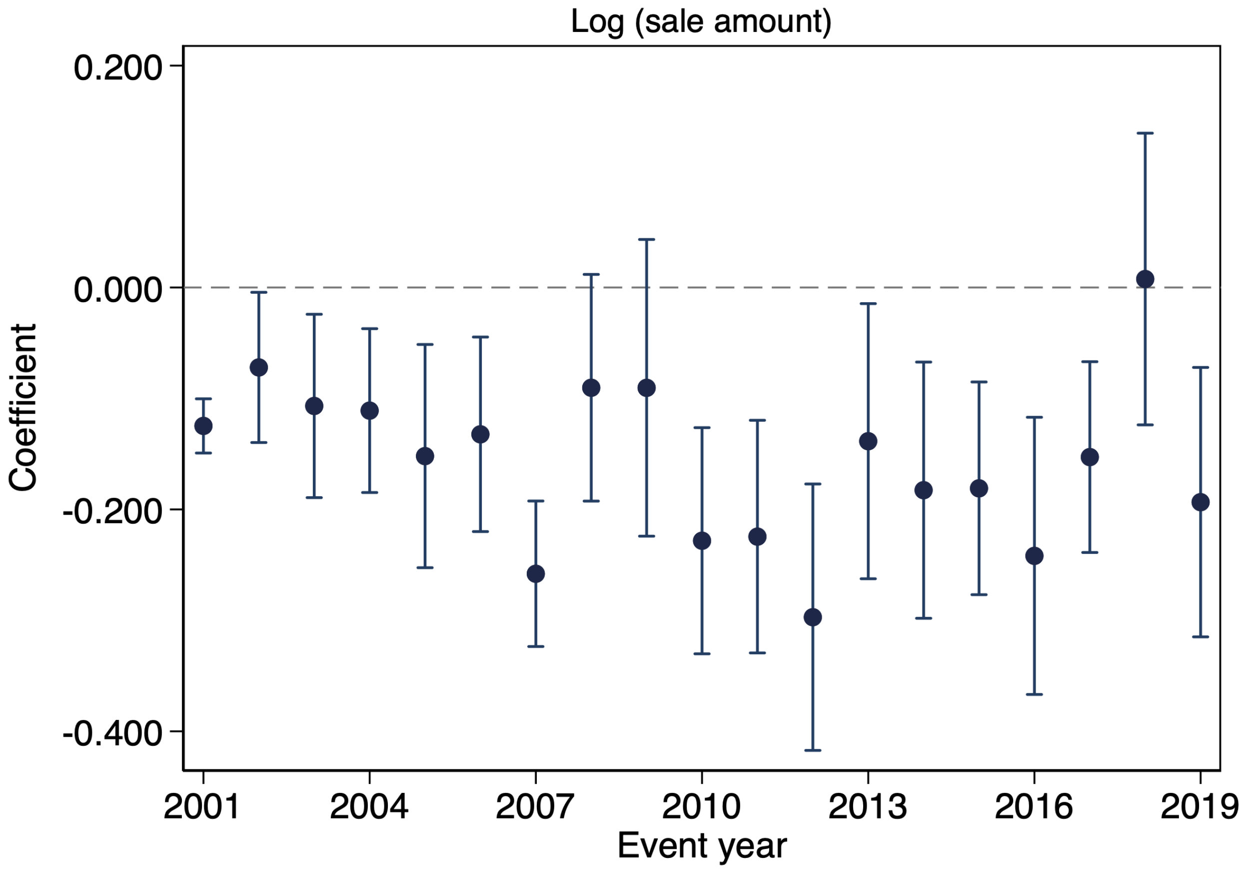


**Figure 6: Estimated housing value effects by event year.**
*Notes*: The figure plots the estimated coefficients of the term Post $\times$ $\mathbb{1}^{Distance<1\ mile}$ by event year. The estimates are obtained by splitting the sample of plants by the year in which they first reported the emission of harmful carcinogen pollutants during the sample period. The empirical specification includes plant and year $\times$ county fixed effects. Standard errors are clustered at the county level.

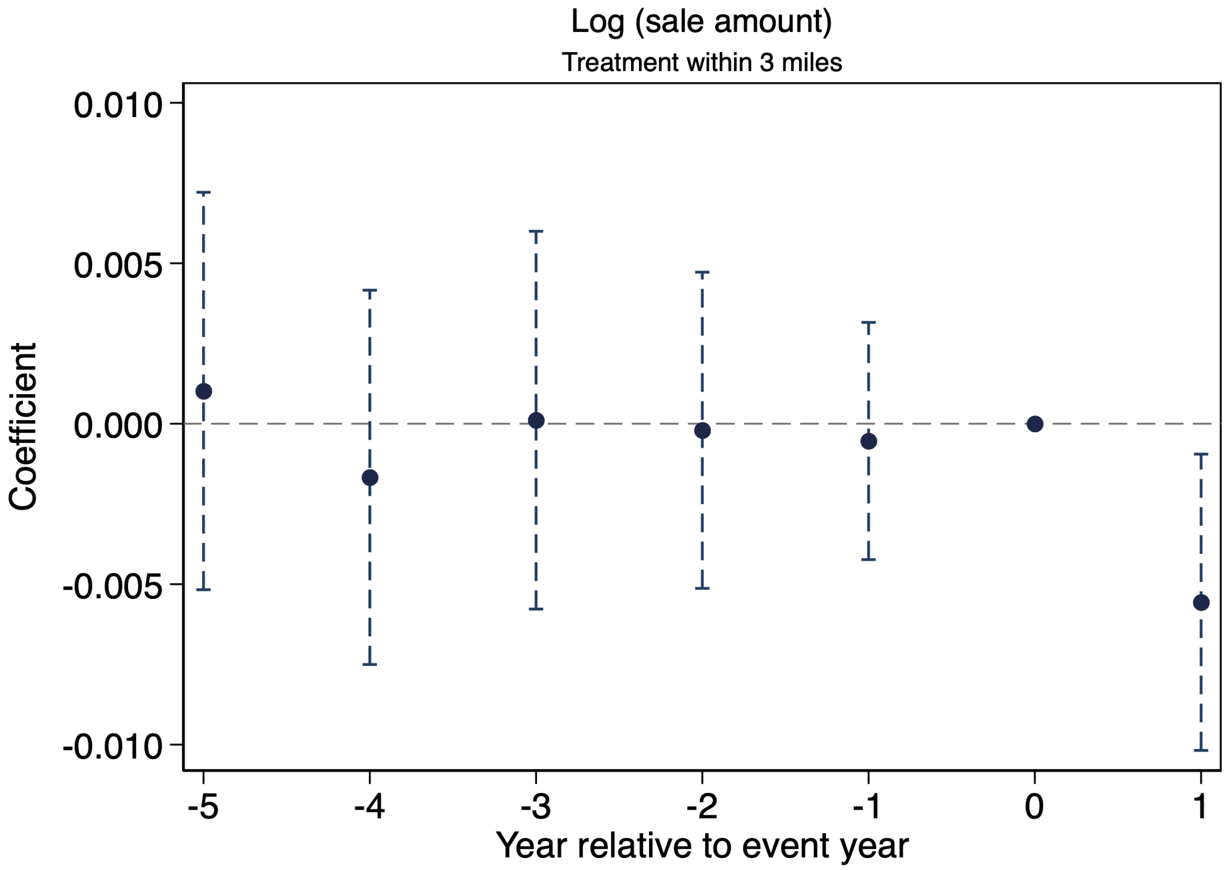


**Figure 7: Estimated housing value effects around first year of reporting carcinogenic emissions.**

*Notes*: The figure plots the coefficient on $\text{Post}_{it} \times \mathbb{1}_{ij}^{\text{Distance}_{ij}<\text{3miles}}$ in event time and the corresponding 95% confidence interval. Specifically, we estimate the following dynamic Difference-in-Differences equation:

$$\text{Log (sale amount)}_{ijct} = \sum_{k=-1}^{-5} \alpha_k + \mu_1 + \sum_{k=-1}^{-5} \beta_k \times \mathbb{1}_{ij}^{\text{Distance}<\text{3miles}} + \mu_1 \times \mathbb{1}_{ij}^{\text{Distance}<\text{3miles}} + \gamma_i + \omega_{jt} + \epsilon_{ijct}$$

We plot coefficients, $(\beta_k, \mu_1)$, on relative differences between properties within 3 miles ("treated") and between 3 and 5 miles ("control") of a plant, and normalized to time zero. The sample is restricted to repeated transactions within 5 miles of a plant. The estimates are for an event window of (-5,1) years relative to the first year the plant reports emitting carcinogenic toxins in the EPA's TRI program. The regression includes property $(\gamma_i)$ and county $\times$ sale-year $(\omega_{jt})$ fixed effects. Standard errors are clustered at the county level.

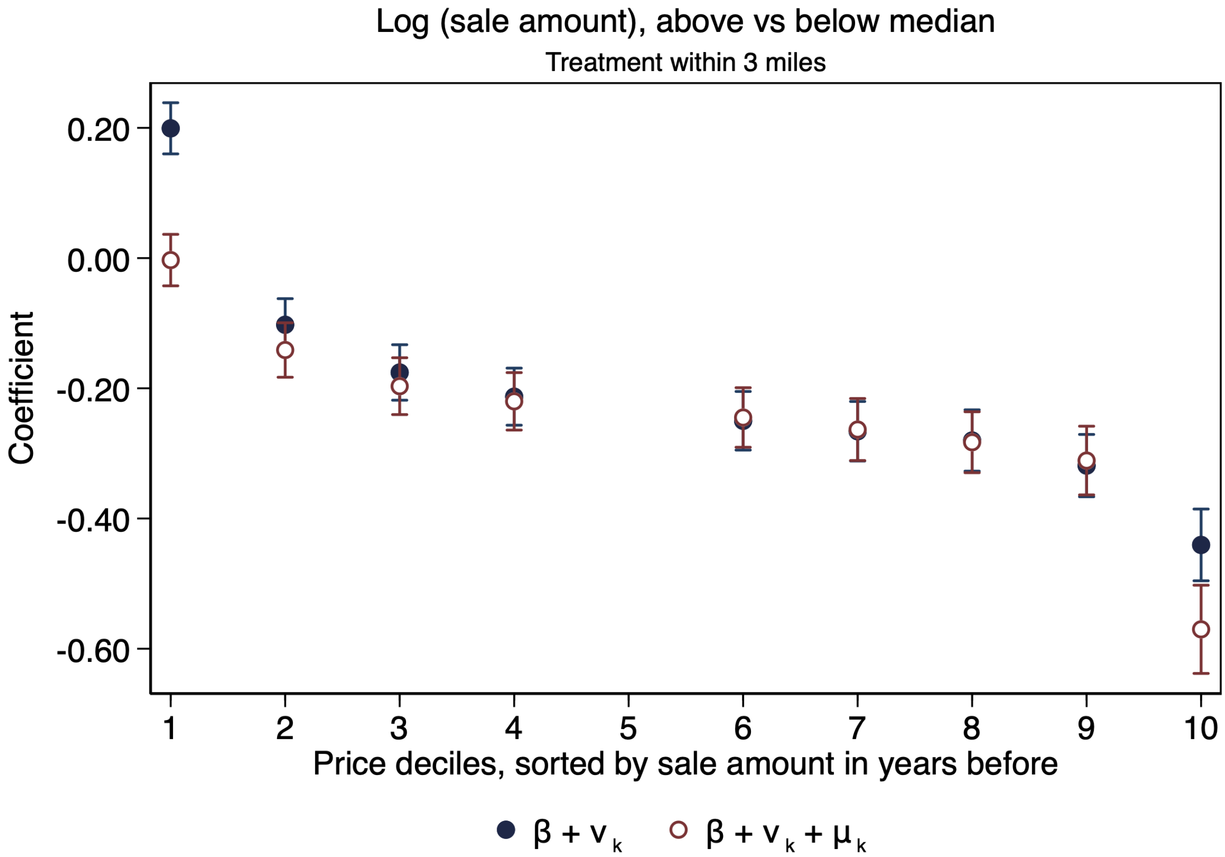


**Figure 8: Heterogeneity in treatment effects by sale amount.**
*Notes*: The figure plots the coefficient on $\mathbb{1}_i^{\text{Decile}} \times \text{Post}_{it} \times \mathbb{1}_{ij}^{\text{Distance}_{ij}<3\text{miles}} \times \text{Above}$ by price decile and the corresponding 95% confidence interval. Specifically, we estimate the following dynamic Difference-in-Differences equation:

$$\text{Log (sale amount)}_{ijct} = \beta \times \text{Post}_{it} + \sum_{k=1}^{10} \nu_k \times \mathbb{1}_i^{\text{Decile}=k} \times \text{Post}_{it} \times \mathbb{1}_{ij}^{\text{Distance}_{ij}<3\text{miles}} + \sum_{k=1}^{10} \mu_k \times \mathbb{1}_i^{\text{Decile }=k} \times \text{Post}_{it} \times \mathbb{1}_{ij}^{\text{Distance}_{ij}<3\text{miles}} \times \textit{Above} + \gamma_i + \omega_{jt} + \epsilon_{ijct}$$

Blue solid circles display the sum of coefficients, $\beta+\nu_k$, for below median properties in each price decile while red hollow circle display the sum of coefficients, $\beta + \nu_k + \mu_k$, for above median properties. The coefficients are relative to decile 5. The sample is restricted to repeated transactions within 5 miles of a plant. The estimates are for an event window of (-1,1) years relative to the first year the plant reports emitting carcinogenic toxins in the EPA's TRI program. The regression includes property ($\gamma_i$) and county $\times$ sale-year ($\omega_{jt}$) fixed effects. Standard errors are clustered at the county level.

**Table 1:** Changes in RSEI cancer scores

*Notes:* This table presents regression estimates on changes in RSEI cancer scores within one year around the first time a toxic plant reports emitting carcinogenic toxins in the EPA's TRI program (event year). The dependent variable is the level of the RSEI cancer scores computed for a given location $l$ within 5 miles of the plant. The independent variable, $Post_{lt}$, is an indicator variable taking a value of one if the RSEI cancer score is computed in the year $t$ after the event year and zero otherwise. We define five treatment rings based on the distance of the property from a toxic plant. Specifically, $\mathbb{1}_{lj}^{\text{Distance}_{lj}<X\text{miles}}$ takes a value of one if the RSEI cancer score is within $X$ miles from a plant $j$, where $X$ is 3, 2, 1.5, 1.25, or 1 mile (columns 1 to 5), and zero for properties between 3 and 5 miles of the same plant. The empirical specification is as follows:

$$\text{RSEI cancer}_{ljt} = \alpha + \beta_{Distance} \times \mathbb{1}_{jb}^{\text{Distance}_{jb}<X\text{miles}} + \beta_{\text{Post}\times\text{Distance}} \times \text{Post}_{lt} \times \mathbb{1}_{lj}^{\text{Distance}_{lj}<X\text{miles}} + \gamma_{jt} + \epsilon_{ljt}.$$

All regressions include plant $\times$ year fixed effects. Standard errors are double-clustered at the plant and year level *c*. $^{**}$, $^{*}$, $^{*}$ denote significance at the 1%, 5%, and 10% level, respectively.

| Dependent variable: | RSEI cancer score | | | | |
|---|---|---|---|---|---|
| Treatment (Distance in miles) | 3 | 2 | 1.5 | 1.25 | 1 |
| | (1) | (2) | (3) | (4) | (5) |
| $\mathbb{1}_{Distance<Xmiles}$ | 5.916*** | 5.065*** | 3.979*** | 2.717*** | 1.366*** |
| | (0.853) | (0.724) | (0.547) | (0.352) | (0.165) |
| Post $\times$ $\mathbb{1}_{Distance<Xmiles}$ | 26.028*** | 22.465*** | 17.745*** | 12.525*** | 6.650*** |
| | (4.638) | (3.983) | (3.126) | (2.165) | (1.055) |
| Plant $\times$ year fixed effects | Yes | Yes | Yes | Yes | Yes |
| $R^2$ | 0.60 | 0.62 | 0.64 | 0.67 | 0.74 |
| Observations | 44,000 | 44,000 | 44,000 | 44,000 | 44,000 |

**Table 2:** Summary statistics: Industry coverage and chemical usage

*Notes:* This table presents the frequency of toxic plants emitting carcinogen chemicals by industry and chemical usage. Panel A reports the top ten industries by three-digit North American Industry Classification System (NAICS) industry while panel B reports the ten most common carcinogenic chemicals emitted by toxic plants in our sample.

**Panel A: Industry coverage**

| Three-digit NAICS | Industry description | Percent of plants |
|---|---|---|
| 332 | Fabricated metal products | 12.82 |
| 327 | Nonmetallic mineral products | 9.69 |
| 334 | Computer & electronic products | 9.00 |
| 336 | Transportation equipment | 7.45 |
| 325 | Chemicals | 7.11 |
| 333 | Machinery, except electrical | 6.95 |
| 331 | Primary metal manufacturing | 6.00 |
| 423 | Merchant Wholesalers, Durable Goods | 4.53 |
| 424 | Merchant wholesalers non durable goods | 4.01 |
| 335 | Electrical Equipment, Appliances & Components | 3.53 |

**Panel B: Chemical usage**

| Chemical name | Percent of plants |
|---|---|
| Lead | 52.63 |
| Nickel | 16.94 |
| Chromium | 13.87 |
| Manganese | 12.92 |
| Copper | 12.43 |
| Xylene (mixed isomers) | 9.02 |
| Toluene | 8.25 |
| Polycyclic aromatic compounds | 7.77 |
| Zinc compounds | 6.67 |
| Lead compounds | 6.16 |

**Table 3:** Carcinogenic chemicals

*Notes:* This table presents top ten carcinogenic chemicals ranked by toxicity. Toxicity is defined as the product of air emissions times inhalation toxicity weight summed across all toxic plants in our sample. We also report the corresponding frequency of plants responsible for these chemical releases (column 3). Inhalation unit risk is expressed as the upper-bound excess risk of developing cancer over a person's lifetime that can be attributed to ongoing exposure to a substance at a level of 1 gram per cubic meter (column 4).

| CAS Registry Number | Chemical | Percent of plants | Inhalation unit risk |
|---|---|---|---|
| 7440-47-3 | Chromium | 13.87 | 43000 |
| N090 | Chromium compounds | 3.75 | 43000 |
| 7440-48-4 | Cobalt | 1.68 | 17000 |
| 7440-02-0 | Nickel | 16.94 | 930 |
| 75-21-8 | Ethylene oxide | 0.26 | 11000 |
| 1332-21-4 | Asbestos (friable) | 0.06 | 170000 |
| N096 | Cobalt compounds | 0.64 | 17000 |
| N590 | Polycyclic aromatic compounds | 7.77 | 390 |
| 126-99-8 | Chloroprene | 0.03 | 1100 |
| N495 | Nickel compounds | 4.59 | 930 |

**Table 4:** Comparison of county characteristics: Ever treated vs Never treated

*Notes:* This table presents the average values for several characteristics across counties. "Ever treated" are counties with at least one of the 11,143 toxic plants emitting carcinogen toxins under the EPA's TRI program in our sample. "Never treated" are counties with none of these 11,143 toxic plants. Standard errors are double-clustered by county and year. ***, **, * denote significance at the 1%, 5%, and 10% level, respectively..

| | Ever treated | Never treated | Difference |
|---|---|---|---|
| | (1) | (2) | (1) - (2) |
| Real GDP per capita (US $1,000) | 42.88 | 63.04 | -20.16 |
| Log (real GDP — US $1,000) | 14.49 | 13.03 | 1.45*** |
| Log (Property taxes — US $1,000) | 9.34 | 8.97 | 0.37 |
| Log (annual payroll — US $1,000) | 13.12 | 11.26 | 1.86*** |
| Log (# establishments) | 7.05 | 5.64 | 1.41*** |
| Unemployment rate (%) | 6.23 | 6.43 | -0.19* |
| Annual change in house price (%) | 2.60 | 2.83 | -0.23** |
| Female population (%) | 50.40 | 49.58 | 0.82*** |
| Black population (%) | 10.23 | 8.64 | 1.59*** |
| Hispanic population (%) | 7.27 | 9.74 | -2.47*** |

**Table 5:** Changes in house prices around first year of reporting carcinogenic emissions

*Notes:* This table presents regression estimates on changes in house prices within one year around the first time a toxic plant reports emitting carcinogenic toxins in the EPA's TRI program (event year). Panel A shows results from all property transactions, while Panel B only includes properties that have been sold multiple times during the event window. The dependent variable is the natural logarithm of the sale amount of a property, *Log (sale amount)*. The independent variable, $Post_{it}$, is an indicator variable taking a value of one if property $i$ is sold in the year $t$ after the event year and zero otherwise. We define five treatment rings based on the distance of the property from a toxic plant. Specifically, $\mathbb{1}_{ij}^{\text{Distance}_{ij}<X\text{miles}}$ takes a value of one if property $i$ is within $X$ miles from a plant $j$, where $X$ is 3, 2, 1.5, 1.25, or 1 mile (columns 1 to 5), and zero for properties between 3 and 5 miles of the same plant. The empirical specification is as follows:

$$\log(\text{Sale amount})_{ijct} = \alpha + \beta_{Post} \times \text{Post}_{it} + \beta_{\text{Post}\times\text{Distance}} \times \text{Post}_{it} \times \mathbb{1}_{ij}^{\text{Distance}_{ij}<X\text{miles}} + \gamma_j + \gamma_{ct} + \epsilon_{ijct}.$$

All regressions include plant and county $\times$ sale-year fixed effects. Standard errors are clustered at the county level $c$. ***, **, * denote significance at the 1%, 5%, and 10% level, respectively.

**Panel A: All transactions**

| Dependent variable: | Log(sale amount) | | | | |
|---|---|---|---|---|---|
| Treatment (Distance in miles) | 3 | 2 | 1.5 | 1.25 | 1 |
| | (1) | (2) | (3) | (4) | (5) |
| Post | 0.022*** | 0.019*** | 0.015*** | 0.011*** | 0.008** |
| | (0.004) | (0.004) | (0.004) | (0.004) | (0.004) |
| Post $\times$ $\mathbb{1}^{\text{Distance}<X\text{miles}}$ | -0.063*** | -0.085*** | -0.101*** | -0.111*** | -0.124*** |
| | (0.005) | (0.007) | (0.008) | (0.008) | (0.009) |
| Plant fixed effects | Yes | Yes | Yes | Yes | Yes |
| Year $\times$ county fixed effects | Yes | Yes | Yes | Yes | Yes |
| $R^2$ | 0.43 | 0.43 | 0.43 | 0.43 | 0.43 |
| Observations | 7,542,012 | 5,744,154 | 4,998,638 | 4,688,460 | 4,424,724 |

**Panel B: Properties with repeated transactions**

| Dependent variable: | Log(sale amount) | | | | |
|---|---|---|---|---|---|
| Treatment (Distance in miles) | 3 | 2 | 1.5 | 1.25 | 1 |
| | (1) | (2) | (3) | (4) | (5) |
| Post | 0.079*** | 0.077*** | 0.075*** | 0.072*** | 0.068*** |
| | (0.009) | (0.010) | (0.010) | (0.010) | (0.010) |
| Post $\times$ $\mathbb{1}^{\text{Distance}<X\text{miles}}$ | -0.066*** | -0.084*** | -0.093*** | -0.103*** | -0.112*** |
| | (0.006) | (0.008) | (0.009) | (0.010) | (0.011) |
| Plant fixed effects | Yes | Yes | Yes | Yes | Yes |
| Year $\times$ county fixed effects | Yes | Yes | Yes | Yes | Yes |
| $R^2$ | 0.39 | 0.39 | 0.39 | 0.39 | 0.39 |
| Observations | 1,085,693 | 829,738 | 724,260 | 680,180 | 642,095 |

**Table 6:** Changes in house prices, repeated sales approach

*Notes:* This table presents regression estimates on changes in house prices within one year around the first year a toxic plant reports emitting carcinogenic pollutants in the EPA's TRI program (event year). The dependent variable is the natural logarithm of the sale amount of a property, *Log (sale amount)*. The independent variable, $Post_{it}$, is an indicator variable taking a value of one if property $i$ is sold in the year $t$ after the event year and zero otherwise. We define five treatment rings based on the distance of the property from a toxic plant. Specifically, $\mathbb{1}_{ij}^{\text{Distance}_{ij}<X\text{miles}}$ takes a value of one if property $i$ is within $X$ miles from a plant $j$, where $X$ is 3, 2, 1.5, 1.25, or 1 mile (columns 1 to 5), and zero for properties between 3 and 5 miles of the same plant. The empirical specification is as follows:

$$\log(\text{Sale amount})_{ijct} = \alpha + \beta_{Post} \times \text{Post}_{it} + \beta_{\text{Post}\times\text{Distance}} \times \text{Post}_{it} \times \mathbb{1}_{ij}^{\text{Distance}_{ij}<X\text{miles}} + \gamma_i + \gamma_{ct} + \epsilon_{ijct}.$$

All regressions include property and county × sale-year fixed effects. Standard errors are clustered at the county level $c$. ***, **, * denote significance at the 1%, 5%, and 10% level, respectively.

| Dependent variable: | Log(sale amount) | | | | |
|---|---|---|---|---|---|
| Treatment (Distance in miles) | 3<br>(1) | 2<br>(2) | 1.5<br>(3) | 1.25<br>(4) | 1<br>(5) |
| Post | 0.009 | 0.007 | 0.009 | 0.009 | 0.007 |
| | (0.009) | (0.009) | (0.009) | (0.010) | (0.009) |
| Post × $\mathbb{1}^{\text{Distance}<X\text{miles}}$ | -0.014*** | -0.015** | -0.016** | -0.017* | -0.016* |
| | (0.005) | (0.006) | (0.008) | (0.009) | (0.010) |
| Property fixed effects | Yes | Yes | Yes | Yes | Yes |
| Year × county fixed effects | Yes | Yes | Yes | Yes | Yes |
| $R^2$ | 0.86 | 0.86 | 0.86 | 0.86 | 0.86 |
| Observations | 1,085,693 | 829,738 | 724,260 | 680,180 | 642,095 |

**Table 7:** Robustness to longer time horizons

*Notes:* This table presents robustness to longer time horizons of the estimates of changes in house prices around the first time a toxic plant reports emitting carcinogenic toxins in the EPA's TRI program (event year). Columns 1 through 5 use an event window of (-2,1) years, while columns 6 through 10 use an event window of (-3,1) years. The dependent variable is the natural logarithm of the sale amount of a property, *Log (sale amount)*. The independent variable, $Post_{it}$, is an indicator variable taking a value of one if property $i$ is sold in the year $t$ after the event year and zero otherwise. We define five treatment rings based on the distance of the property from a toxic plant. Specifically, $\mathbb{1}_{ij}^{\text{Distance}_{ij}<X\text{miles}}$ takes a value of one if property $i$ is within $X$ miles from a plant $j$, where $X$ is 3, 2, 1.5, 1.25, or 1 mile (columns 1 to 5), and zero for properties between 3 and 5 miles of the same plant. The empirical specification is as follows:

$$\log(\text{Sale amount})_{ijct} = \alpha + \beta_{Post} \times \text{Post}_{it} + \beta_{\text{Post}\times\text{Distance}} \times \text{Post}_{it} \times \mathbb{1}_{ij}^{\text{Distance}_{ij}<X\text{miles}} + \gamma_i + \gamma_{ct} + \epsilon_{ijct}.$$

All regressions include property and county $\times$ sale-year fixed effects. Standard errors are clustered at the county level $c$. ***, **, * denote significance at the 1%, 5%, and 10% level, respectively.

| Dependent variable: | Log(sale amount) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Event window: | (-2,1) | | | | | (-3,1) | | | | |
| Treatment (Distance in miles) | 3 | 2 | 1.5 | 1.25 | 1 | 3 | 2 | 1.5 | 1.25 | 1 |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Post | 0.014*** | 0.014*** | 0.017*** | 0.016*** | 0.016*** | 0.016*** | 0.015*** | 0.016*** | 0.015*** | 0.016*** |
| | (0.005) | (0.005) | (0.005) | (0.006) | (0.006) | (0.004) | (0.004) | (0.005) | (0.005) | (0.005) |
| Post $\times$ $\mathbb{1}^{\text{Distance}<X\text{miles}}$ | -0.012*** | -0.013** | -0.013** | -0.014** | -0.014* | -0.011*** | -0.011** | -0.012** | -0.013** | -0.014* |
| | (0.004) | (0.005) | (0.006) | (0.007) | (0.008) | (0.004) | (0.005) | (0.006) | (0.007) | (0.007) |
| Property fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Year $\times$ county fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| $R^2$ | 0.87 | 0.86 | 0.86 | 0.86 | 0.86 | 0.87 | 0.87 | 0.87 | 0.87 | 0.87 |
| Observations | 2,024,515 | 1,546,964 | 1,348,404 | 1,266,398 | 1,195,207 | 3,127,197 | 2,387,778 | 2,080,655 | 1,954,800 | 1,846,326 |

**Table 8:** Heterogeneity by price: Above median

*Notes:* This table presents regression estimates on changes in house prices within one year around the first time a toxic plant reports emitting carcinogenic toxins in the EPA's TRI program (event year) separating more and less expensive properties. The dependent variable is the natural logarithm of the sale amount of a property, *Log (sale amount)*. The independent variable, $Post_{it}$, is an indicator variable taking a value of one if property $i$ is sold in the year $t$ after the event year and zero otherwise. We define five treatment rings based on the distance of the property from a toxic plant. Specifically, $\mathbb{1}_{ij}^{\text{Distance}_{ij}<X\text{miles}}$ takes a value of one if property $i$ is within $X$ miles from a plant $j$, where $X$ is 3, 2, 1.5, 1.25, or 1 mile (columns 1 to 5), and zero for properties between 3 and 5 miles of the same plant (control ring). We interact our treatment variable $\text{Post}_{it} \times \mathbb{1}_{ij}^{\text{Distance}_{ij}<X\text{miles}}$ with an indicator for whether the property was transacted for an amount above the median value of properties surrounding the plant in the years before the event, where the median is computed using all properties in the treatment and control rings surrounding a plant. The empirical specification is as follows:

$$\begin{aligned}\log(\text{Sale amount})_{ijct} = \alpha + \beta_{Post} \times \text{Post}_{it} + \beta_{\text{Post}\times\text{Distance}} \times \text{Post}_{it} \times \mathbb{1}_{ij}^{\text{Distance}_{ij}<X\text{miles}} \\ + \beta_{\text{Post}\times\text{Distance}\times Above} \times \text{Post}_{it} \times \mathbb{1}_{ij}^{\text{Distance}_{ij}<X\text{miles}} \times Above_{ij} + \gamma_i + \gamma_{ct} + \epsilon_{ijct}.\end{aligned}$$

All regressions include property and county × sale-year fixed effects. Standard errors are clustered at the county level $c$. ***, **, * denote significance at the 1%, 5%, and 10% level, respectively.

| Dependent variable: | Log(sale amount) | | | | |
|---|---|---|---|---|---|
| Treatment (Distance in miles) | 3 | 2 | 1.5 | 1.25 | 1 |
| | (1) | (2) | (3) | (4) | (5) |
| Post | 0.004 | 0.004 | 0.007 | 0.007 | 0.006 |
| | (0.009) | (0.009) | (0.009) | (0.010) | (0.009) |
| Post × $\mathbb{1}^{\text{Distance}<X\text{miles}}$ | 0.106*** | 0.103*** | 0.098*** | 0.094*** | 0.091*** |
| | (0.006) | (0.008) | (0.009) | (0.010) | (0.010) |
| Post × $\mathbb{1}^{\text{Distance}<X\text{miles}}$ × Above | -0.296*** | -0.303*** | -0.300*** | -0.301*** | -0.299*** |
| | (0.014) | (0.016) | (0.017) | (0.019) | (0.020) |
| Property fixed effects | Yes | Yes | Yes | Yes | Yes |
| Year × county fixed effects | Yes | Yes | Yes | Yes | Yes |
| $R^2$ | 0.87 | 0.86 | 0.86 | 0.86 | 0.86 |
| Observations | 1,085,693 | 829,738 | 724,260 | 680,180 | 642,095 |

**Table 9:** Heterogeneity by price and longer time horizon: Above vs. Below median. Robustness to longer time horizons

*Notes:* This table presents robustness to longer time horizons of the estimates of changes in house prices around the first time a toxic plant reports emitting carcinogenic toxins in the EPA's TRI program (event year). Columns 1 through 5 use an event window of (-2,1) years, while columns 6 through 10 use an event window of (-3,1) years. The dependent variable is the natural logarithm of the sale amount of a property, *Log (sale amount)*. The independent variable, $Post_{it}$, is an indicator variable taking a value of one if property $i$ is sold in the year $t$ after the event year and zero otherwise.We define five treatment rings based on the distance of the property from a toxic plant. Specifically, $\mathbb{1}_{ij}^{\text{Distance}_{ij}<X\text{miles}}$ takes a value of one if property $i$ is within $X$ miles from a plant $j$, where $X$ is 3, 2, 1.5, 1.25, or 1 mile (columns 1 to 5), and zero for properties between 3 and 5 miles of the same plant (control ring). We interact our treatment variable $\text{Post}_{it} \times \mathbb{1}_{ij}^{\text{Distance}_{ij}<X\text{miles}}$ with an indicator for whether the property was transacted for an amount above the median value of properties surrounding the plant in the years before the event, where the median is computed using all properties in the treatment and control rings surrounding a plant. The empirical specification is as follows:

$$\begin{aligned}\log(\text{Sale amount})_{ijct} = \alpha &+ \beta_{Post} \times \text{Post}_{it} + \beta_{\text{Post}\times\text{Distance}} \times \text{Post}_{it} \times \mathbb{1}_{ij}^{\text{Distance}_{ij}<X\text{miles}} \\ &+ \beta_{\text{Post}\times\text{Distance}\times Above} \times \text{Post}_{it} \times \mathbb{1}_{ij}^{\text{Distance}_{ij}<X\text{miles}} \times Above_{ij} + \gamma_i + \gamma_{ct} + \epsilon_{ijct}.\end{aligned}$$

All regressions include property and county × sale-year fixed effects. Standard errors are clustered at the county level *c*. ***, **, * denote significance at the 1%, 5%, and 10% level, respectively.

| Dependent variable: | Log(sale amount) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Event window: | (-2,1) | | | | | (-3,1) | | | | |
| Treatment (Distance in miles) | 3 | 2 | 1.5 | 1.25 | 1 | 3 | 2 | 1.5 | 1.25 | 1 |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Post | 0.012** | 0.013** | 0.016*** | 0.015*** | 0.016*** | 0.015*** | 0.015*** | 0.015*** | 0.014*** | 0.015*** |
| | (0.005) | (0.005) | (0.005) | (0.006) | (0.006) | (0.004) | (0.004) | (0.004) | (0.005) | (0.005) |
| Post × $\mathbb{1}_{Distance<Xmiles}$ | 0.093*** | 0.089*** | 0.086*** | 0.083*** | 0.079*** | 0.084*** | 0.081*** | 0.078*** | 0.074*** | 0.070*** |
| | (0.006) | (0.007) | (0.008) | (0.008) | (0.009) | (0.005) | (0.006) | (0.007) | (0.008) | (0.008) |
| Post × $\mathbb{1}_{Distance<Xmiles}$ × Above | -0.248*** | -0.252*** | -0.252*** | -0.252*** | -0.247*** | -0.219*** | -0.222*** | -0.222*** | -0.220*** | -0.216*** |
| | (0.010) | (0.012) | (0.012) | (0.014) | (0.014) | (0.009) | (0.010) | (0.010) | (0.011) | (0.012) |
| Property fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Year × county fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| $R^2$ | 0.87 | 0.87 | 0.86 | 0.86 | 0.86 | 0.87 | 0.87 | 0.87 | 0.87 | 0.87 |
| Observations | 2,024,515 | 1,546,964 | 1,348,404 | 1,266,398 | 1,195,207 | 3,127,197 | 2,387,778 | 2,080,655 | 1,954,800 | 1,846,326 |

**Table 10:** Changes in plant-level employment and sales

*Notes:* This table presents regression estimates on changes in plant-level employment and sales within one year around the first year a toxic plant reports emitting carcinogenic pollutants in the EPA's TRI program (event year). The dependent variable in column 1 (column 2) is the natural logarithm of employment (sales). The independent variable, $Post_{it}$, is an indicator variable taking a value of one for all years after the event year and zero otherwise. All regressions include plant and year fixed effects. Standard errors are clustered at the county level $c$. ***, **, * denote significance at the 1%, 5%, and 10% level, respectively.

| Dependent variable: | Log (employment) | Log (sales) |
|---|---|---|
| | (1) | (2) |
| Post | 0.019** | 0.014 |
| | (0.009) | (0.010) |
| Plant fixed effects | Yes | Yes |
| Year fixed effects | Yes | Yes |
| $R^2$ | 0.97 | 0.97 |
| Observations | 30,162 | 29,633 |

**Table 11:** Changes in fraction of hispanic home buyers, with plant fixed effects

*Notes:* This table presents regression estimates of the house purchases and sales by Hispanic individuals within one year around the first time a toxic plant reports emitting carcinogenic toxins in the EPA's TRI program (event year). In Panel A (Panel B) the dependent variable, $\mathbb{1}(\textit{Hispanic})$, is an indicator variable taking the value of one if the predicted probability that an individual buyer (seller) is of Hispanic ethnicity is greater than 85%. The independent variable, $Post_{it}$, is an indicator variable taking a value of one if property $i$ is sold in the year $t$ after the event year and zero otherwise. We define five treatment rings based on the distance of the property from a toxic plant. Specifically, $\mathbb{1}_{ij}^{\text{Distance}_{ij}<X\text{miles}}$ takes a value of one if property $i$ is within $X$ miles from a plant $j$, where $X$ is 3, 2, 1.5, 1.25, or 1 mile (columns 1 to 5), and zero for properties between 3 and 5 miles of the same plant. The empirical specification is as follows:

$$\mathbb{1}(\text{Hispanic})_{ijct} = \alpha + \beta_{Post} \times \text{Post}_{it} + \beta_{\text{Post}\times\text{Distance}} \times \text{Post}_{it} \times \mathbb{1}_{ij}^{\text{Distance}_{ij}<X\text{miles}} + \gamma_j + \gamma_{ct} + \epsilon_{ijct}.$$

All regressions include plant and county × sale-year fixed effects. Standard errors are clustered at the county level $c$. ***, **, * denote significance at the 1%, 5%, and 10% level, respectively.

| **Panel A: Buyers** | | | | | |
|---|---|---|---|---|---|
| Dependent variable: | $\mathbb{1}$(Hispanic) | | | | |
| Treatment (Distance in miles) | 3 | 2 | 1.5 | 1.25 | 1 |
| | (1) | (2) | (3) | (4) | (5) |
| Post | -0.005*** | -0.005*** | -0.004** | -0.003* | -0.002 |
| | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) |
| Post × $\mathbb{1}_{Distance<Xmiles}$ | 0.012*** | 0.016*** | 0.018*** | 0.019*** | 0.022*** |
| | (0.002) | (0.003) | (0.003) | (0.003) | (0.004) |
| Plant fixed effects | Yes | Yes | Yes | Yes | Yes |
| Year× county fixed effects | Yes | Yes | Yes | Yes | Yes |
| $R^2$ | 0.14 | 0.14 | 0.14 | 0.14 | 0.13 |
| Observations | 6,177,760 | 4,701,805 | 4,088,040 | 3,832,287 | 3,615,276 |

| **Panel B: Sellers** | | | | | |
|---|---|---|---|---|---|
| Dependent variable: | $\mathbb{1}$(Hispanic) | | | | |
| Treatment (Distance in miles) | 3 | 2 | 1.5 | 1.25 | 1 |
| | (1) | (2) | (3) | (4) | (5) |
| Post | -0.004*** | -0.003** | -0.003** | -0.002** | -0.002* |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Post × $\mathbb{1}_{Distance<Xmiles}$ | 0.008*** | 0.010*** | 0.012*** | 0.014*** | 0.015*** |
| | (0.002) | (0.002) | (0.003) | (0.003) | (0.004) |
| Plant fixed effects | Yes | Yes | Yes | Yes | Yes |
| Year− county fixed effects | Yes | Yes | Yes | Yes | Yes |
| $R^2$ | 0.10 | 0.10 | 0.10 | 0.10 | 0.10 |
| Observations | 4,795,888 | 3,640,031 | 3,158,636 | 2,959,388 | 2,788,211 |

# A Additional results

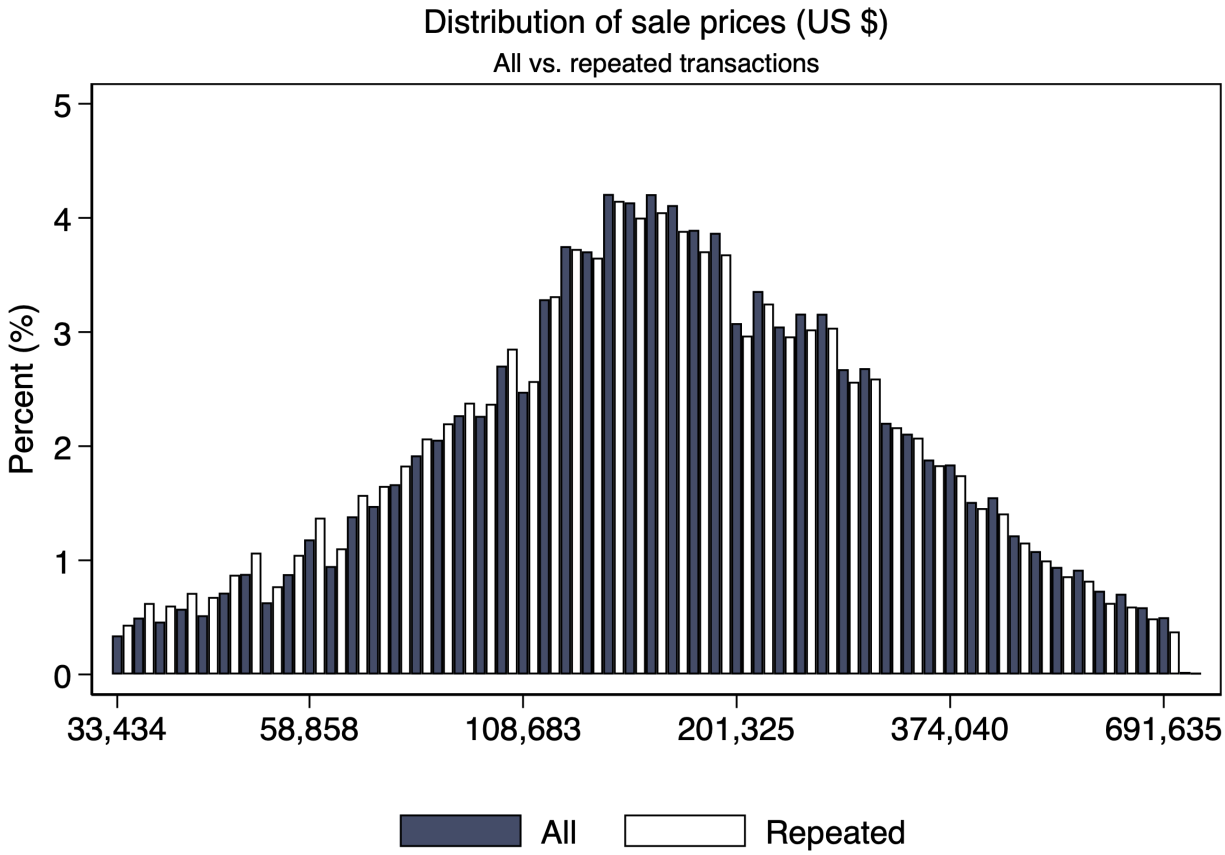


**Figure A.1: Empirical distribution of sale prices.**
*Notes*: The figure shows the empirical distribution of sale prices for all transactions in our sample (blue solid bars) as in column (1) of panel A of Table 5 and the sample of repeated transactions (white hollow bars) as in column (1) of panel B of Table 5. The sample is restricted to sale prices one year around the first year a plant reports emitting carcinogenic pollutants.

**Table A.1:** Changes in RSEI cancer scores, Poisson estimation

*Notes:* This table presents the Poisson pseudo-maximum likelihood regression estimates on changes in RSEI cancer scores within one year around the first time a toxic plant reports emitting carcinogenic toxins in the EPA's TRI program (event year). The dependent variable is the level of the RSEI cancer scores computed for a given location $l$ within 5 miles of the plant. The independent variable, $Post_{lt}$, is an indicator variable taking a value of one if the RSEI cancer score is computed in the year $t$ after the event year and zero otherwise. We define five treatment rings based on the distance of the property from a toxic plant. Specifically, $\mathbb{1}_{lj}^{\text{Distance}_{lj}<X\text{miles}}$ takes a value of one if the RSEI cancer score is within $X$ miles from a plant $j$, where $X$ is 3, 2, 1.5, 1.25, or 1 mile (columns 1 to 5), and zero for properties between 3 and 5 miles of the same plant. The empirical specification is as follows:

$$\text{RSEI cancer}_{ljt} = \exp\{\alpha + \beta_{Distance} \times \mathbb{1}_{jb}^{\text{Distance}_{jb}<X\text{miles}} + \beta_{\text{Post}\times\text{Distance}} \times \text{Post}_{lt} \times \mathbb{1}_{lj}^{\text{Distance}_{lj}<X\text{miles}} + \gamma_{jt} + \epsilon_{ljt}\}.$$

All regressions include plant × year fixed effects. Standard errors are double-clustered at the plant and year level. $^{**}$, $^{*}$, $^{*}$ denote significance at the 1%, 5%, and 10% level, respectively.

| Dependent variable: | RSEI cancer score | | | | |
|---|---|---|---|---|---|
| Treatment (Distance in miles) | 3 | 2 | 1.5 | 1.25 | 1 |
| | (1) | (2) | (3) | (4) | (5) |
| $\mathbb{1}_{Distance<Xmiles}$ | 2.449$^{***}$ | 2.308$^{***}$ | 2.093$^{***}$ | 1.767$^{***}$ | 1.236$^{***}$ |
| | (0.082) | (0.088) | (0.093) | (0.098) | (0.099) |
| Post × $\mathbb{1}_{Distance<Xmiles}$ | 0.131 | 0.135 | 0.136 | 0.152 | 0.165$^{*}$ |
| | (0.080) | (0.085) | (0.088) | (0.095) | (0.094) |
| Plant × year fixed effects | Yes | Yes | Yes | Yes | Yes |
| Pseudo-$R^2$ | 0.97 | 0.97 | 0.95 | 0.93 | 0.88 |
| Observations | 4,208 | 4,208 | 4,208 | 4,208 | 4,208 |

**Table A.2:** Changes in RSEI cancer scores

*Notes:* This table presents regression estimates on changes in RSEI cancer scores within one year around the first time a toxic plant reports emitting carcinogenic toxins in the EPA's TRI program (event year). The dependent variable is the level of the RSEI cancer scores computed for a given location $l$ within 5 miles of the plant. The independent variable, $Post_{lt}$, is an indicator variable taking a value of one if the RSEI cancer score is computed in the year $t$ after the event year and zero otherwise. We define five treatment rings based on the distance of the property from a toxic plant. Specifically, $\mathbb{1}_{lj}^{\text{Distance}_{lj}<X\text{miles}}$ takes a value of one if the RSEI cancer score is within $X$ miles from a plant $j$, where $X$ is 3, 2, 1.5, 1.25, or 1 mile (columns 1 to 5), and zero for properties between 3 and 5 miles of the same plant. The empirical specification is as follows:

$$\text{RSEI cancer}_{ljt} = \alpha + \beta_{Distance} \times \mathbb{1}_{jb}^{\text{Distance}_{jb}<X\text{miles}} + \beta_{\text{Post}\times\text{Distance}} \times \text{Post}_{lt} \times \mathbb{1}_{lj}^{\text{Distance}_{lj}<X\text{miles}} + \gamma_{jt} + \epsilon_{ljt}.$$

All regressions include plant × year fixed effects. Standard errors are double-clustered at the plant and year level $c$. $^{**}$, $^{*}$, $^{*}$ denote significance at the 1%, 5%, and 10% level, respectively.

| Dependent variable: | RSEI cancer score | | | | |
|---|---|---|---|---|---|
| Treatment (Distance in miles) | 3 | 2 | 1.5 | 1.25 | 1 |
| | (1) | (2) | (3) | (4) | (5) |
| $\mathbb{1}_{Distance<Xmiles}$ | 30.201$^{***}$ | 26.422$^{***}$ | 21.205$^{***}$ | 14.838$^{***}$ | 8.066$^{***}$ |
| | (4.028) | (3.461) | (2.660) | (1.815) | (1.009) |
| Post × $\mathbb{1}_{Distance<Xmiles}$ | 33.359$^{***}$ | 29.439$^{***}$ | 23.592$^{***}$ | 17.524$^{***}$ | 10.160$^{***}$ |
| | (9.981) | (8.694) | (6.883) | (4.892) | (2.537) |
| Plant × year fixed effects | Yes | Yes | Yes | Yes | Yes |
| $R^2$ | 0.59 | 0.60 | 0.61 | 0.64 | 0.69 |
| Observations | 27,374 | 27,374 | 27,374 | 27,374 | 27,374 |

**Table A.3:** Changes in house prices, with plant-year fixed effects

*Notes:* This table presents regression estimates on changes in house prices within one year around the first time a toxic plant reports emitting carcinogenic toxins in the EPA's TRI program (event year) using the sample of properties that have been sold multiple times during the event window. The dependent variable is the natural logarithm of the sale amount of a property, *Log (sale amount)*. The independent variable, $Post_{it}$, is an indicator variable taking a value of one if property $i$ is sold in the year $t$ after the event year and zero otherwise. We define five treatment rings based on the distance of the property from a toxic plant. Specifically, $\mathbb{1}_{ij}^{\text{Distance}_{ij}<X\text{miles}}$ takes a value of one if property $i$ is within $X$ miles from a plant $j$, where $X$ is 3, 2, 1.5, 1.25, or 1 mile (columns 1 to 5), and zero for properties between 3 and 5 miles of the same plant. The empirical specification is as follows:

$$\log(\text{Sale amount})_{ijct} = \alpha + \beta_{\text{Post}\times\text{Distance}} \times \text{Post}_{it} \times \mathbb{1}_{ij}^{\text{Distance}_{ij}<X\text{miles}} + \gamma_{jt} + \gamma_c + \epsilon_{ijct}.$$

All regressions include plant × sale-year and county fixed effects. Standard errors are clustered at the county level $c$. $^{***}$, $^{**}$, $^{*}$ denote significance at the 1%, 5%, and 10% level, respectively.

| Dependent variable: | Log(sale amount) | | | | |
|---|---|---|---|---|---|
| Treatment (Distance in miles) | 3 | 2 | 1.5 | 1.25 | 1 |
| | (1) | (2) | (3) | (4) | (5) |
| Post × $\mathbb{1}^{\text{Distance}<X\text{miles}}$ | -0.068*** | -0.089*** | -0.101*** | -0.112*** | -0.122*** |
| | (0.006) | (0.008) | (0.009) | (0.010) | (0.012) |
| Plant×year fixed effects | Yes | Yes | Yes | Yes | Yes |
| County fixed effects | Yes | Yes | Yes | Yes | Yes |
| $R^2$ | 0.39 | 0.39 | 0.39 | 0.39 | 0.39 |
| Observations | 1,085,203 | 829,243 | 723,758 | 679,667 | 641,571 |

**Table A.4:** Changes in house prices, greater than 100 observations

*Notes:* This table presents regression estimates on changes in house prices within one year around the first time a toxic plant reports emitting carcinogenic toxins in the EPA's TRI program (event year). The dependent variable is the natural logarithm of the sale amount of a property, *Log (sale amount)*. The independent variable, $Post_{it}$, is an indicator variable taking a value of one if property $i$ is sold in the year $t$ after the event year and zero otherwise. We define five treatment rings based on the distance of the property from a toxic plant. Specifically, $\mathbb{1}_{ij}^{\text{Distance}_{ij} < X \text{miles}}$ takes a value of one if property $i$ is within $X$ miles from a plant $j$, where $X$ is 3, 2, 1.5, 1.25, or 1 mile (columns 1 to 5), and zero for properties between 3 and 5 miles of the same plant. The sample is restricted to transaction with more than 100 observations. The empirical specification is as follows:

$$\log(\text{Sale amount})_{ijct} = \alpha + \beta_{Post} \times \text{Post}_{it} + \beta_{\text{Post}\times\text{Distance}} \times \text{Post}_{it} \times \mathbb{1}_{ij}^{\text{Distance}_{ij} < X \text{miles}} + \gamma_i + \gamma_{ct} + \epsilon_{ijct}.$$

All regressions include property and county $\times$ sale-year fixed effects. Standard errors are clustered at the county level $c$. ***, **, * denote significance at the 1%, 5%, and 10% level, respectively.

| Dependent variable: | Log(sale amount) | | | | |
|---|---|---|---|---|---|
| Treatment (Distance in miles) | 3 | 2 | 1.5 | 1.25 | 1 |
| | (1) | (2) | (3) | (4) | (5) |
| Post | 0.010 | 0.006 | 0.009 | 0.009 | 0.005 |
| | (0.011) | (0.012) | (0.012) | (0.013) | (0.013) |
| Post $\times$ $\mathbb{1}_{Distance<Xmiles}$ | -0.014*** | -0.015** | -0.015* | -0.015 | -0.013 |
| | (0.005) | (0.006) | (0.008) | (0.010) | (0.010) |
| Property fixed effects | Yes | Yes | Yes | Yes | Yes |
| Year$\times$ county fixed effects | Yes | Yes | Yes | Yes | Yes |
| $R^2$ | 0.86 | 0.86 | 0.86 | 0.86 | 0.86 |
| Observations | 950,321 | 717,668 | 622,391 | 583,362 | 549,856 |

**Table A.5:** Changes in house prices in level, with property fixed effects

*Notes:* This table presents regression estimates on changes in house prices within one year around the first year a toxic plant reports emitting carcinogenic pollutants in the EPA's TRI program (event year). The dependent variable is the sale amount in US$ of a property, *Sale amount*. The independent variable, $Post_{it}$, is an indicator variable taking a value of one if property $i$ is sold in the year $t$ after the event year and zero otherwise. We define five treatment rings based on the distance of the property from a toxic plant. Specifically, $\mathbb{1}_{ij}^{\text{Distance}_{ij}<X\text{miles}}$ takes a value of one if property $i$ is within $X$ miles from a plant $j$, where $X$ is 3, 2, 1.5, 1.25, or 1 mile (columns 1 to 5), and zero for properties between 3 and 5 miles of the same plant. The empirical specification is as follows:

$$\text{Sale amount}_{ijct} = \alpha + \beta_{Post} \times \text{Post}_{it} + \beta_{\text{Post}\times\text{Distance}} \times \text{Post}_{it} \times \mathbb{1}_{ij}^{\text{Distance}_{ij}<X\text{miles}} + \gamma_i + \gamma_{ct} + \epsilon_{ijct}.$$

All regressions include property and county × sale-year fixed effects. Standard errors are clustered at the county level $c$. ***, **, * denote significance at the 1%, 5%, and 10% level, respectively.

| Dependent variable: | Sale amount | | | | |
|---|---|---|---|---|---|
| Treatment (Distance in miles) | 3 | 2 | 1.5 | 1.25 | 1 |
| | (1) | (2) | (3) | (4) | (5) |
| Post | 1747.970 | 1313.927 | 1583.286 | 1382.705 | 695.509 |
| | (1906.569) | (2006.493) | (2054.050) | (2082.697) | (2036.329) |
| Post × $\mathbb{1}_{Distance<Xmiles}$ | -4225.660*** | -5771.382*** | -6026.167*** | -7063.518*** | -7245.870*** |
| | (966.320) | (1272.229) | (1650.031) | (1808.607) | (1868.643) |
| Property fixed effects | Yes | Yes | Yes | Yes | Yes |
| Year× county fixed effects | Yes | Yes | Yes | Yes | Yes |
| $R^2$ | 0.87 | 0.87 | 0.87 | 0.87 | 0.87 |
| Observations | 1,085,693 | 829,738 | 724,260 | 680,180 | 642,095 |

**Table A.6:** Heterogeneity by price, above and below median defined within ring

*Notes:* This table presents regression estimates on changes in house prices within one year around the first time a toxic plant reports emitting carcinogenic toxins in the EPA's TRI program (event year) separating more and less expensive properties. The dependent variable is the natural logarithm of the sale amount of a property, *Log (sale amount)*. The independent variable, $Post_{it}$, is an indicator variable taking a value of one if property $i$ is sold in the year $t$ after the event year and zero otherwise. We define five treatment rings based on the distance of the property from a toxic plant. Specifically, $\mathbb{1}_{ij}^{\text{Distance}_{ij}<X\text{miles}}$ takes a value of one if property $i$ is within $X$ miles from a plant $j$, where $X$ is 3, 2, 1.5, 1.25, or 1 mile (columns 1 to 5), and zero for properties between 3 and 5 miles of the same plant (control ring) We interact our treatment variable $\text{Post}_{it} \times \mathbb{1}_{ij}^{\text{Distance}_{ij}<X\text{miles}}$ with an indicator for whether the property was transacted for an amount above the median value of properties surrounding the plant in the years before the event, where the median is computed separately for treatment and control rings. The empirical specification is as follows:

$$\begin{aligned}\log(\text{Sale amount})_{ijct} &= \alpha + \beta_{Post} \times \text{Post}_{it} + \beta_{\text{Post}\times\text{Distance}} \times \text{Post}_{it} \times \mathbb{1}_{ij}^{\text{Distance}_{ij}<X\text{miles}} \\ &+ \beta_{\text{Post}\times\text{Distance}\times Above} \times \text{Post}_{it} \times \mathbb{1}_{ij}^{\text{Distance}_{ij}<X\text{miles}} \times Above_{ij} + \gamma_i + \gamma_{ct} + \epsilon_{ijct}.\end{aligned}$$

All regressions include property and county × sale-year fixed effects. Standard errors are clustered at the county level $c$. ***, **, * denote significance at the 1%, 5%, and 10% level, respectively.

| Dependent variable: | Log(sale amount) | | | | |
|---|---|---|---|---|---|
| Treatment (Distance in miles) | 3 | 2 | 1.5 | 1.25 | 1 |
| | (1) | (2) | (3) | (4) | (5) |
| Post | 0.004 | 0.005 | 0.008 | 0.008 | 0.007 |
| | (0.009) | (0.009) | (0.009) | (0.010) | (0.009) |
| Post × $\mathbb{1}^{\text{Distance}<X\text{miles}}$ | 0.115*** | 0.117*** | 0.116*** | 0.112*** | 0.109*** |
| | (0.007) | (0.009) | (0.009) | (0.010) | (0.011) |
| Post × $\mathbb{1}^{\text{Distance}<X\text{miles}}$ × Above | -0.300*** | -0.307*** | -0.307*** | -0.303*** | -0.297*** |
| | (0.013) | (0.015) | (0.016) | (0.017) | (0.018) |
| Property fixed effects | Yes | Yes | Yes | Yes | Yes |
| Year × county fixed effects | Yes | Yes | Yes | Yes | Yes |
| $R^2$ | 0.87 | 0.86 | 0.86 | 0.86 | 0.86 |
| Observations | 1,085,693 | 829,738 | 724,260 | 680,180 | 642,095 |

**Table A.7:** Heterogeneity by price: Above median – Robustness, greater than 100 observations

This table presents regression estimates on changes in house prices within one year around the first year a toxic plant reports emitting carcinogenic pollutants in the EPA's TRI program (event year) separating more and less expensive properties. The dependent variable is the natural logarithm of the sale amount of a property, *Log (sale amount)*. The independent variable, $Post_{it}$, is an indicator variable taking a value of one if property $i$ is sold in the year $t$ after the event year and zero otherwise. We define five treatment rings based on the distance of the property from a toxic plant. Specifically, $\mathbb{1}_{ij}^{\text{Distance}_{ij}<X\text{miles}}$ takes a value of one if property $i$ is within $X$ miles from a plant $j$, where $X$ is 3, 2, 1.5, 1.25, or 1 mile (columns 1 to 5), and zero for properties between 3 and 5 miles of the same plant (control ring). We interact our treatment variable $\text{Post}_{it} \times \mathbb{1}_{ij}^{\text{Distance}_{ij}<X\text{miles}}$ with an indicator for whether the property was transacted for an amount above the median value of properties surrounding the plant in the years before the event, where the median is computed using all properties in the treatment and control rings surrounding a plant. The sample is restricted to transaction with more than 100 observations. The empirical specification is as follows:

$$\log(\text{Sale amount})_{ijct} = \alpha + \beta_{Post} \times \text{Post}_{it} + \beta_{\text{Post}\times\text{Distance}} \times \text{Post}_{it} \times \mathbb{1}_{ij}^{\text{Distance}_{ij}<X\text{miles}} + \beta_{\text{Post}\times\text{Distance}\times Above} \times \text{Post}_{it} \times \mathbb{1}_{ij}^{\text{Distance}_{ij}<X\text{miles}} \times Above_{ij} + \gamma_i + \gamma_{ct} + \epsilon_{ijct}.$$

All regressions include property and county × sale-year fixed effects. Standard errors are clustered at the county level *c*. ***, **, * denote significance at the 1%, 5%, and 10% level, respectively.

| Dependent variable: | Log(sale amount) | | | | |
|---|---|---|---|---|---|
| Treatment (Distance in miles) | 3 | 2 | 1.5 | 1.25 | 1 |
| | (1) | (2) | (3) | (4) | (5) |
| Post | 0.005 | 0.003 | 0.006 | 0.006 | 0.004 |
| | (0.011) | (0.012) | (0.012) | (0.013) | (0.012) |
| Post × $\mathbb{1}_{Distance<Xmiles}$ | 0.105*** | 0.102*** | 0.099*** | 0.096*** | 0.093*** |
| | (0.007) | (0.009) | (0.010) | (0.011) | (0.011) |
| Post × Above × $\mathbb{1}_{Distance<Xmiles}$ | -0.292*** | -0.298*** | -0.297*** | -0.297*** | -0.294*** |
| | (0.014) | (0.017) | (0.018) | (0.020) | (0.022) |
| Property fixed effects | Yes | Yes | Yes | Yes | Yes |
| Year × county fixed effects | Yes | Yes | Yes | Yes | Yes |
| $R^2$ | 0.87 | 0.86 | 0.86 | 0.86 | 0.86 |
| Observations | 950,321 | 717,668 | 622,391 | 583,362 | 549,856 |

**Table A.8:** Robustness to dropping large price changes between consecutive transactions

*Notes:* This table presents regression estimates on changes in house prices within one year around the first time a toxic plant reports emitting carcinogenic toxins in the EPA's TRI program (event year) separating more and less expensive properties. The dependent variable is the natural logarithm of the sale amount of a property, *Log (sale amount)*. The independent variable, $Post_{it}$, is an indicator variable taking a value of one if property $i$ is sold in the year $t$ after the event year and zero otherwise. We define five treatment rings based on the distance of the property from a toxic plant. Specifically, $\mathbb{1}_{ij}^{\text{Distance}_{ij}<X\text{miles}}$ takes a value of one if property $i$ is within $X$ miles from a plant $j$, where $X$ is 3, 2, 1.5, 1.25, or 1 mile (columns 1 to 5), and zero for properties between 3 and 5 miles of the same plant (control ring). We interact our treatment variable $\text{Post}_{it} \times \mathbb{1}_{ij}^{\text{Distance}_{ij}<X\text{miles}}$ with an indicator for whether the property was transacted for an amount above the median value of properties surrounding the plant in the years before the event, where the median is computed using all properties in the treatment and control rings surrounding a plant. The empirical specification is as follows:

$$\log(\text{Sale amount})_{ijct} = \alpha + \beta_{Post} \times \text{Post}_{it} + \beta_{\text{Post}\times\text{Distance}} \times \text{Post}_{it} \times \mathbb{1}_{ij}^{\text{Distance}_{ij}<X\text{miles}} + \beta_{\text{Post}\times\text{Distance}\times Above} \times \text{Post}_{it} \times \mathbb{1}_{ij}^{\text{Distance}_{ij}<X\text{miles}} \times Above_{ij} + \gamma_i + \gamma_{ct} + \epsilon_{ijct}.$$

All regressions include property and county × sale-year fixed effects. Standard errors are clustered at the county level *c*. ***, **, * denote significance at the 1%, 5%, and 10% level, respectively.

**Panel A: Drop 10% of observations by price changes – bottom 5% and top 5%**

| Dependent variable: | Log(sale amount) | | | | |
|---|---|---|---|---|---|
| Treatment (Distance in miles) | 3 | 2 | 1.5 | 1.25 | 1 |
| | (1) | (2) | (3) | (4) | (5) |
| Post | -0.001 | 0.001 | 0.002 | 0.003 | 0.003 |
| | (0.004) | (0.004) | (0.004) | (0.004) | (0.004) |
| Post × $\mathbb{1}^{\text{Distance}<X\text{miles}}$ | 0.071*** | 0.070*** | 0.068*** | 0.069*** | 0.071*** |
| | (0.003) | (0.004) | (0.005) | (0.006) | (0.006) |
| Post × $\mathbb{1}^{\text{Distance}<X\text{miles}}$ × Above | -0.167*** | -0.172*** | -0.172*** | -0.174*** | -0.177*** |
| | (0.006) | (0.007) | (0.008) | (0.009) | (0.009) |
| Property fixed effects | Yes | Yes | Yes | Yes | Yes |
| Year × county fixed effects | Yes | Yes | Yes | Yes | Yes |
| $R^2$ | 0.94 | 0.94 | 0.94 | 0.94 | 0.94 |
| Observations | 977,926 | 745,405 | 649,851 | 609,778 | 575,283 |

**Panel B: Drop 20% of observations by price changes – bottom 10% and top 10%**

| Dependent variable: | Log(sale amount) | | | | |
|---|---|---|---|---|---|
| Treatment (Distance in miles) | 3 | 2 | 1.5 | 1.25 | 1 |
| | (1) | (2) | (3) | (4) | (5) |
| Post | -0.004 | -0.002 | -0.001 | -0.000 | -0.000 |
| | (0.003) | (0.003) | (0.004) | (0.004) | (0.004) |
| Post × $\mathbb{1}^{\text{Distance}<X\text{miles}}$ | 0.055*** | 0.054*** | 0.055*** | 0.055*** | 0.056*** |
| | (0.003) | (0.003) | (0.004) | (0.004) | (0.005) |
| Post × $\mathbb{1}^{\text{Distance}<X\text{miles}}$ × Above | -0.122*** | -0.126*** | -0.128*** | -0.129*** | -0.131*** |
| | (0.004) | (0.005) | (0.006) | (0.006) | (0.007) |
| Property fixed effects | Yes | Yes | Yes | Yes | Yes |
| Year × county fixed effects | Yes | Yes | Yes | Yes | Yes |
| $R^2$ | 0.96 | 0.96 | 0.96 | 0.96 | 0.96 |
| Observations | 882,582 | 671,732 | 585,016 | 548,666 | 517,483 |

**Table A.9:** Robustness: Changes in fraction of hispanic home buyers and sellers

*Notes:* This table presents regression estimates of the house purchases and sales by Hispanic individuals within one year around the first time a toxic plant reports emitting carcinogenic toxins in the EPA's TRI program (event year). In Panel A (Panel B) the dependent variable, $\mathbb{1}(\textit{Hispanic})$, is an indicator variable taking the value of one if the predicted probability that an individual buyer (seller) is of Hispanic ethnicity is greater than 90%. The independent variable, $Post_{it}$, is an indicator variable taking a value of one if property $i$ is sold in the year $t$ after the event year and zero otherwise. We define five treatment rings based on the distance of the property from a toxic plant. Specifically, $\mathbb{1}_{ij}^{\text{Distance}_{ij}<X\text{miles}}$ takes a value of one if property $i$ is within $X$ miles from a plant $j$, where $X$ is 3, 2, 1.5, 1.25, or 1 mile (columns 1 to 5), and zero for properties between 3 and 5 miles of the same plant. The empirical specification is as follows:

$$\mathbb{1}(\text{Hispanic})_{ijct} = \alpha + \beta_{Post} \times \text{Post}_{it} + \beta_{\text{Post}\times\text{Distance}} \times \text{Post}_{it} \times \mathbb{1}_{ij}^{\text{Distance}_{ij}<X\text{miles}} + \gamma_j + \gamma_{ct} + \epsilon_{ijct}.$$

All regressions include plant and county × sale-year fixed effects. Standard errors are clustered at the county level $c$. ***, **, * denote significance at the 1%, 5%, and 10% level, respectively.

| Panel A: Buyers | | | | | |
|---|---|---|---|---|---|
| Dependent variable: | $\mathbb{1}$(Hispanic) | | | | |
| Treatment (Distance in miles) | 3 | 2 | 1.5 | 1.25 | 1 |
| | (1) | (2) | (3) | (4) | (5) |
| Post | -0.007*** | -0.007*** | -0.006*** | -0.005** | -0.003* |
| | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) |
| Post × $\mathbb{1}_{Distance<Xmiles}$ | 0.016*** | 0.021*** | 0.024*** | 0.025*** | 0.028*** |
| | (0.003) | (0.003) | (0.004) | (0.004) | (0.005) |
| Plant fixed effects | Yes | Yes | Yes | Yes | Yes |
| Year × county fixed effects | Yes | Yes | Yes | Yes | Yes |
| $R^2$ | 0.17 | 0.17 | 0.17 | 0.17 | 0.17 |
| Observations | 6,177,760 | 4,701,805 | 4,088,040 | 3,832,287 | 3,615,276 |

| Panel B: Sellers | | | | | |
|---|---|---|---|---|---|
| Dependent variable: | $\mathbb{1}$(Hispanic) | | | | |
| Treatment (Distance in miles) | 3 | 2 | 1.5 | 1.25 | 1 |
| | (1) | (2) | (3) | (4) | (5) |
| Post | -0.005*** | -0.005*** | -0.004*** | -0.004*** | -0.003** |
| | (0.002) | (0.002) | (0.001) | (0.001) | (0.001) |
| Post × $\mathbb{1}_{Distance<Xmiles}$ | 0.010*** | 0.014*** | 0.016*** | 0.018*** | 0.019*** |
| | (0.002) | (0.003) | (0.003) | (0.004) | (0.005) |
| Plant fixed effects | Yes | Yes | Yes | Yes | Yes |
| Year × county fixed effects | Yes | Yes | Yes | Yes | Yes |
| $R^2$ | 0.13 | 0.13 | 0.13 | 0.12 | 0.12 |
| Observations | 4,795,888 | 3,640,031 | 3,158,636 | 2,959,388 | 2,788,211 |