Environmental health risks, property values and neighborhood composition*

Jules van Binsbergen[1], João F. Cocco[2], Marco Grotteria[2], and S. Lakshmi Naaraayanan[3]

[1]University of Pennsylvania, CEPR and NBER.
[2]London Business School and CEPR.
[3]London Business School.

March 14, 2024

Abstract

We quantify the impact of carcinogenic risk exposure on property values and neighborhood composition. Following a plant's first reported carcinogen emission, we observe a 0.8% to 2% decline in property values for houses closer to the plant relative to houses farther away. Combining our estimates with cancer hazard ratios from epidemiological studies implies a value of statistical life ranging from \$2.6M to \$6.4M. Our analysis reveals a shift towards a higher presence of minority and higher credit-risk households in houses closer to the plant. This evidence of changes in neighborhood composition in the wake of a change in perceived environmental health risk informs the debate on environmental justice and health inequities.

Keywords: Environmental health risks, Inequality, Real estate prices, Household finance.
JEL Classification: G11, G50, Q51.

*We thank Jamie Coen, Jefferson Duarte, Francisco Gomes, Byoung-Hyoun Hwang, Amit Kumar, Kasper Nielsen, Henri Servaes, Daniel Wolfenzon, Anthony Lee Zhang, and participants at the TCU Neeley Finance Conference, EPA TRI National Conference, 17^{th} North American Meeting of the Urban Economics Association (Toronto), CEPR Endless Summer Conference on Financial Intermediation and Corporate Finance, Society for the Advancement of Economic Theory, Université de Namur, King's College, London Business School, Nanyang Business School, and Peking University for helpful comments and suggestions. Corresponding authors: julesv@wharton.upenn.edu (van Binsbergen), jcocco@london.edu (Cocco), mgrotteria@london.edu (Grotteria), and lnaaraayanan@london.edu (Naaraayanan).

1 Introduction

In the United States every year 1.9 million people are newly diagnosed with and 609 thousand people die of cancer, with a disproportionate burden on people from minority groups. Cancer is the second leading cause of mortality in the United States with spending on cancer care exceeding $200 billion in 2020.[1] Given the importance of environmental factors on carcinogenic risk exposure, the past few decades have seen a marked increase in the number of industrial chemicals that are scientifically identified as causing cancer in humans. This confronts policymakers with important trade-offs between stimulating industrial activity and protecting the health and well-being of citizens. A key input in this trade-off is the value that citizens attribute to environmental health risk, which cannot be directly observed and is generally hard to estimate.

In this paper, we quantify the dollar value of changes in perceived environmental health risk using house values and by identifying a set of plants that already existed and that started reporting the emission of carcinogenic toxins. We rely on two sets of plant-year *events*: first, the initial reporting of carcinogenic toxin emissions by plants; second, within a subset of these plants, the identification of those emitting a newly recognized human carcinogen in its year of reclassification, without prior emissions of any other carcinogens. Both of these events help us estimate the value of environmental risk exposure as we hold constant plant location choices.

The toxic events identified might stem from increased production, leading to higher plant employment and local economic growth. Therefore, the empirical strategy should isolate the impact of environmental health risks on property values from local economic benefits. 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 events on house values within the vicinity of the plant, namely those within a 3-mile ring of the plant and those in a ring of between three and five miles from the same plant. The former set of houses is our treated group, and the latter is the control group. The identifying assumption is that environmental health risk will be higher for properties closest to the toxic plant, whereas all properties within the 5-mile radius will benefit from the same local economic activity. The choice of a 3-mile ring is informed by the empirical evidence on the estimated cancer risk decay rate as a function of distance from the toxic plants, but we also provide estimates for smaller treated rings.

[1] A cross-country comparison of cancer-related spending and mortality rates can be found here.

Our data source for housing values is Corelogic. We restrict the sample to properties that were transacted *both* in the year before or in the year of the event and in the year after the event. This allows us to a) keep the same composition of properties before and after the event, and b) use property fixed effects and estimate *within-property* changes to housing values, which is important since properties are heterogeneous across many unobservable dimensions. We also include plant times sale-year fixed effects to control for changes in economic conditions in the 5-mile ring around the plant. 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 the value households assign to changes in environmental health risks as captured by changes in house values. We estimate effects on the value of properties in the treated group relative to the control group of between -0.8% and -1.5%.

We use the same plant-year events and methodology to estimate their impact on model-estimated cancer scores from the Environmental Protection Agency (EPA). We find a relative increase of 1 unit in cancer score in the treated areas compared to the control group. To translate the higher cancer score into an increased probability of getting cancer, we use the estimates of Boyle, Ward, Cerhan, Rothman, and Wheeler (2023) that relate EPA's cancer scores to hazard ratios. Combining our estimates with cancer hazard ratios from the epidemiological studies implies a value of $40,341 for one additional year of life, or if we multiply it by 80 years, $3.2 million as the value of one statistical life. This compares to the estimates by the EPA of $7.6 million as the value of one statistical life.

One potential explanation for this difference may be individuals' limited attention towards cancer risk. To explore the role of event salience, we exploit the widely-advertised *re-classification* of carcinogenicity of industrial chemicals by the National Toxicology Program's Report on Carcinogens (RoC). We assess the effects on properties near the subset of plants that emitted non-carcinogenic chemicals, which were later re-classified as carcinogens —a change independent of plant actions. In this case, the estimated property value decline doubles to approximately 2%, corresponding to a value of statistical life of $6.4 million.

An additional contribution of this paper is to study changes in neighborhood composition in the wake of a change in perceived environmental health risk. We examine a) the heterogeneous response of properties based on their price, b) the ethnicity of buyers, and c) the creditworthiness of buyers. To study the heterogeneous response of properties based on their price, we divide 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 reporting events for each of these two groups. We find that the differences in house values between treatment and control groups are more significant, both statistically and economically, for above- than for below-median-price properties.[2] These results provide evidence of an income channel through which sorting may occur to the extent that the more expensive properties are owned by households with higher incomes that value carcinogenic risk exposure more.

Next, we focus on the ethnicity of buyers, which requires data on the characteristics of individuals living in the vicinity of the plants. The Corelogic data includes the name of the sellers and buyers in property transactions. We use the algorithm classification of Sood (2017) and Laohaprapanon, Sood, and Naji (2022) to predict the race and ethnicity of those involved in each property transaction. The estimates show a relative increase of 0.6 percentage points in minority (Hispanic or African-American) home buyers in the vicinity of the treated plant in the year after the events, compared to the control group and to the pre-event period. Finally, for a subset of home buyers in our sample, we are able to obtain their FICO score. The estimates show that the FICO score in the post-event treated group is 4.5 points lower than that of home buyers in the control group, and when compared to the pre-event period. These results provide the first evidence of *granular* changes in neighborhood composition around the plants towards minorities and higher credit-risk households after the toxic events.

The paper is organized as follows. Section 2 discusses our contribution relative to the existing literature. Section 3 describes the data and the identification. Section 4 shows the estimated house value effects as well as several tests that show the robustness of our results, including longer time windows and robustness to measurement error and outliers. Section 5 studies neighborhood composition. In section 6 we develop a stylized heterogeneous agent model that replicates the estimated effects, and sheds light on the mechanisms behind the empirical estimates. The last section concludes.

2 Related literature

Our paper is related to several strands of literature. First, the extant literature on agglomeration argues for spillovers and their propagation through firm networks to the local economy in

[2]However, not all estimated differences are statistically significant, which may be due to the lower number of observations. We require three transactions for the same property so as to sort it into an above and below median price using and observation other than the the two that we use to measure price changes.

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, Moretti, 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).[3] 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.

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

3 Data and methodology

3.1 Data sources

Toxics Release Inventory. Our first main data source is the Toxics Release Inventory (TRI) data from the EPA. Firms that satisfy the following criteria must report their emissions to the EPA: (i) the number of employees (at least 10); (ii) the industry sector where the plant operates (some NAICS codes are covered); (iii) the manufacture, production, or use of TRI listed chemicals; and (iv) the plant exceeds at least one of the thresholds for a chemical or a chemical category. When these four criteria are met, the plant must report its emissions of chemicals covered by the EPA. The list of TRI chemicals includes those that are carcinogenic, on which we focus, and others.

The timing of the data reporting is as follows. From January to June, the plants 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.[4] Our paper uses variation introduced by plants' reporting for the *first time* the release of *known human carcinogens* into the environment.[5] That is, a given plant may have previously reported the emission of non-carcinogenic chemicals; the event is the first time that the plant reports exceeding the threshold of a known carcinogen.

Our algorithm for selecting these plants is as follows. The starting year of our sample is 2000. Plants that already satisfied our event definition in the year 2000 are excluded since we do not know whether this is the first year in which they did so. For each subsequent calendar year, we construct an indicator for those plants and the year in which they first report emitting pollutants that are known human carcinogens at the time of the report. More precisely, the treated plants are those with flags for the emission of carcinogenic toxins classified as such under the Clean Air Act and as a carcinogen by the Occupational Safety and Health Administration (OSHA).

A difficulty is there are instances when a plant reports emitting a pollutant that was not a known carcinogen at the time, but was classified as such at a later date. And when this re-classification happens, there is back-filling of the data. To determine whether chemicals were considered known or reasonably anticipated human carcinogens at the time of the report, we

[4] Further details are available here.

[5] 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.

utilize the National Toxicology Program's *Report on Carcinogens (RoC)*. For those plants that were emitting a carcinogen not known as such at the time of the reporting, we take the event to be the year in which the chemical became known as a carcinogen (according to the *RoC*).

Therefore, our sample of events includes two types of plant-year events: (i) plants that first reported the emission of carcinogens known as such at the time of the reporting; and (ii) plants that were emitting a carcinogen in the year that the chemical became known as such (and were not previously emitting any other carcinogenic chemical). It comprises a total of 14,787 unique plants.

National Establishment Time-Series. A potential confounder for our estimates comes from those plants that both open and meet the EPA's reporting criteria in the same year, making it unclear whether observed effects stem from the plant's opening or from the reporting of carcinogenic toxins. The opening of new plants could affect nearby property values through factors like aesthetics. To address this issue, we use a second data source, the National Establishment Time-Series (NETS) dataset provided by Walls & Associates and Dun and Bradstreet, which Rossi-Hansberg, Sarte, and Trachter (2021) show compares favorably with Census data in terms of quality and coverage.

We exclude from our sample those plants that opened and met the EPA's reporting criteria in the same year (3,226 plants). This leaves us with 11,561 plants that first reported carcinogen emissions in years distinct from their opening. Figure 1 shows these plants are dispersed throughout the US with a larger concentration in densely populated areas.

[Insert Figure 1 here]

Panel A of Table 1 shows the industry coverage of the 11,561 unique plants. More precisely, it shows the two-digit NAICS industry codes with the corresponding fraction of toxic plants belonging to that industry. The largest proportion of plants are in manufacturing (88%).[6] Panel B reports the 10 most frequent chemicals associated with the 11,561 reporting events: lead is by far the most frequent toxin, in roughly 47% of the plant-events, followed by nickel (17%). In the same table, we present the carcinogenicity classification for each chemical, as reported by the International Agency for Research on Cancer (IARC) and the National Toxicology Program (NTP). These two classifications alongside Occupational Safety and Health Administration

[6]The top three manufacturing sub-sectors are: fabricated metal products (15% of the total event sample); nonmetallic mineral products (14%); and chemicals (9%).

(OSHA) 29 CFR part 1910 Subpart Z, form the basis for the EPA to determine chemical carcinogenicity.

[Insert Table 1 here]

It is important to clarify that in our sample of 11,561 plant-events, 4,589 (approximately 40%) met our event criteria within the same year they initially reported emissions to the EPA. This implies that 60% of our sample consists of plants that had already reported non-carcinogenic emissions, which allows us to capture the impact of newly reported carcinogenic emissions. For the remaining 40%, our estimates will capture the *combined* effect of reporting to the EPA and of reporting emission of carcinogenic chemicals.

A further advantage of the NETS data is that it includes annual revenues and employment information at the plant level (from 1990 to 2020). It allows us to study whether there is a relationship between emissions and plant activity. For instance, plants may be meeting the TRI reporting 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 investigate this channel and take it into account in our identification strategy.

Corelogic Deed & Tax Records. To measure housing values of residential properties, we use the Corelogic Deed & Tax record data on housing transactions between 2000 and 2020. We restrict the sample to single-family residences, residential condominiums, duplexes, and apartments. For our granular analysis, the property's exact location is of utmost importance. Therefore, we exclude observations with missing block-level latitude and longitude data.[7] For each of the 11,561 plants, we merge the location information from TRI data with the property transactions data to calculate the distance between each plant and all residential properties using the Vincenty (1975)'s formula. Furthermore, we exclude those observations with missing information on the sale amount or year in which the property was built. Finally, we only keep transactions in which Corelogic records that the buyer purchased the property in cash or via a mortgage, thus excluding non-arm's length inter-family transactions or investor-recorded purchases.

[7]Block-level geographic coordinates specify the north-south and east-west position of a point based on the United States Postal Service address data for each parcel. These coordinates capture the most accurate property location instead of parcel-level centroid geographic coordinates which take into account the land area when computing the property location.

Figure 2 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. There is a wide geographical dispersion of transactions but, as expected, there tends to be more transactions on the East and West Coast, and those bordering the Great Lakes. Comparing Figures 1 and 2, we see that there is a significant geographical overlap between the two, with a larger incidence of plant-events in counties with more real estate transactions.

[Insert Figure 2 here]

Despite the geographical overlap, it still is the case that among the 11,561 plant events that we previously identified, there are 1,140 for which there are no property transactions in the 5-mile radius of the plant. This leaves us with 10,421 plant events. Furthermore, there are several instances of plants located close to each other, with events happening in different or even the same year. In these cases, and to avoid including observations with multiple treatments, we include only property observations corresponding to the first event and linked to the plant closest to it. This reduces the number of plant-events to 7,801.

To reduce the impact of outliers, we eliminate transactions at the extreme ends of the price spectrum, specifically those below \$30,000 (5^{th} percentile of the distribution) and above \$700,000 (95^{th} percentile), and focus on repeated property transactions occurring within one year before and after the plant event, which leaves us with 6,405 plant-events for the analysis of house price effects.

RSEI Geographic Microdata. The EPA uses the TRI data to construct Risk-Screening Environmental Indicators (RSEI) that measure 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 the EPA's model estimates to examine whether our events are associated with changes in cancer risk in the vicinity of the plants and how these changes vary with distance. We use this variation for our identification strategy. It is important to note that the EPA cancer risk estimates use the same data that we use to identify our events. Therefore, they should not be seen as independent evidence of their importance. Furthermore, the EPA only provides cancer risk scores for a subset of the plants.

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 and media as a product of the estimated dosage released by that specific toxic plant, the toxic concentrations, and the potentially exposed population. Moreover, the data includes toxic concentrations of chemicals for release by unit and media, and the number of people in the grid cell who are potentially exposed. We scale the RSEI cancer scores by the exposed population, so as to obtain at the chemical-grid cell-media level a measure of the estimated dosage multiplied by the toxicity of the chemical. We sum the so computed cancer-risk scores at the grid level.

As an illustrative example, Figure 3 plots the heatmap by grid cells (810m $\times$ 810m grids) for the RSEI cancer scores surrounding Brenner Tank LLC, Fond du Lac, WI 54935, in 2002, one of the plant-years included in our event sample. The darker-colored grid cells show a higher cancer risk in the vicinity of the plant. The rings show locations within a 3- and 5-mile radius of the plant.

[Insert Figure 3 here]

Figure 4 presents a binscatter plot of the relation between cancer risk score and distance for 3,539 of plant event years included in our sample (up to a 5-mile radius), for which we are able to obtain cancer risk scores. The data indicate cancer risk is highest in the vicinity of the plant, and declines with distance, without significant changes beyond 3 miles from the plant. In the next subsection, we use this and other evidence to define ring sizes.

[Insert Figure 4 here]

3.2 Identification

One of the main challenges is identifying the effects of environmental health risks on house values separately from other effects. For instance, it may be the case that the toxic events that we identify are the result of higher production levels, with plant-level employment effects, and enhanced economic activity in the area surrounding the plant.

We use the NETS data to provide evidence of this channel. More precisely, for each plant j in our sample, we find a control plant within the same state, industry in the same event year (see Online Appendix A for more details). Let n be the 6-digit NAICS code, and s the state.

We estimate changes in plant-level employment and sales within two years of the event using the following empirical specification:

$$\log(y)_{jnst} = \alpha + \beta_{\text{Post}\times\text{Treated}} \times \text{Post}_{jt} \times \text{Treated}_{jt} + \gamma_j + \gamma_{nt} + \gamma_{st} + \epsilon_{jt}, \quad (1)$$

where the dependent variable y can be either plant-level employment or sales for plant j, the variable Post_{jt} takes the value of 1 for employment observations t after the event year, i.e. year +1 and year +2 observations, and zero otherwise (year 0 and year -1 observations, where year 0 denotes the event year). We include plant fixed effects so that the estimates capture within-plant changes in employment. We also control for time-varying industry and State shocks through Industry × year and State × year fixed effects. Thus, our estimates measure changes in economic activity associated with the plant, beyond that reflected in coarser Industry and State fixed effects. We estimate similar regressions for the log of plant sales as the dependent variable.

Table 2 shows the results. Plant-level employment and sales are on average 5% higher after the event than before relative to the control group. This confirms the hypothesis that there are indeed plant-level changes in economic activity associated with our events, and that these changes are plant-specific, in that they are not captured by the Industry and State fixed effects included in the regressions.[8]

[Insert Table 2 here]

To separate the effects of local economic activity and pollution on house values, we exploit 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 around the plant and those in a ring of between 3 and 5 miles around the same plant. The former set of houses is our treated group, and the latter is the control group. The idea is that emissions have more of an impact on properties located closest to the toxic plant, whereas all properties located within the 5-mile radius benefit from the local economic activity effects. Several recent papers use this "ring" method for identification purposes (Butts, 2022, Diamond and McQuade, 2019, Ganduri and Maturana, 2022, LaPoint, 2022).

[8] Appendix Table B.1 shows results for county-level employment, GDP, average wages, and population, which are consistent with general improvement in economic activity.

3.2.1 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).[9] Second, Figure 4 plots the relation between cancer risk and distance from the plant and shows a higher incidence of such risks manifesting within the 3-mile ring compared to between 3 and 5 miles from the plant in the year of the event. The same figure shows that, within the 3-mile radius, cancer risk declines with distance, so we will consider smaller ring sizes for the treated group while maintaining the control group the same.

Third, we study the relation between changes in cancer risk around our events and distance using regression analysis. For each plant j and grid-cell location l within 5 miles of the plant, we compute the incidence of cancer risk, as measured by the RSEI cancer score scaled by the exposed population. The dependent variable in the regressions is the level of the cancer scores at the 810m×810m grid level.

We want to estimate changes in cancer risk in the year after a plant's first report of carcinogenic emissions to the EPA's TRI program (year +1), relative to before (years 0 and -1 observations). To avoid including observations with multiple treatments, we include only grid-cell observations corresponding to the first event-year and linked to the plant closest to it. Therefore, we define an indicator variable $Post_{lt}$ that takes the value of one if the cancer score corresponds to a year t after the event year and zero otherwise. For distance, we define five treatment rings based on the grid's distance from the toxic plant. Specifically, $\mathbb{1}_{lj}^{\text{Distance}_{lj}<X\text{miles}}$ takes a value of one if the mid-point of grid l is within X miles of plant j, where X is 3, 2, 1.5, 1.25, or 1 mile, and zero for grids with mid-points between 3 and 5 miles of the same plant. The empirical specification is as follows:

$$\begin{aligned}\text{Cancer Score}_{ljt} = \alpha + \beta_{Distance} \times \mathbb{1}_{jl}^{\text{Distance}_{jl}<X\text{miles}} + \beta_{\text{Post}\times\text{Distance}} \\ \times \text{Post}_{lt} \times \mathbb{1}_{lj}^{\text{Distance}_{lj}<X\text{miles}} + \gamma_{jt} + \gamma_l + \epsilon_{ljt}.\end{aligned} \tag{2}$$

The above equation includes plant × year fixed effects (γ_{jt}) and grid fixed effects (γ_l). Therefore, we estimate within-grid changes in cancer risk around our events controlling for time-varying

[9] 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.

characteristics of the plant and macroeconomic conditions in the area where the plant is located. Standard errors are double-clustered at the plant and year levels.

The estimated coefficients in Table 3 show that, after the events, there are larger increases in cancer risk in the treated areas compared to those in the control group, and that the increases are larger for locations closer to the plant. This can be seen from the estimated coefficients on the interaction terms $\text{Post}_{lt} \times \mathbb{1}_{lj}^{\text{Distance}_{lj} < X\text{miles}}$ which are statistically significant and become economically larger as we reduce the radius of the treated ring from 3 to 1 mile (columns 1 to 5), while keeping the control ring at the base level of between 3 and 5 miles.

It is important to clarify that the RSEI cancer score measure that we use as dependent variable in the regressions is scaled by the number of individuals who live in the ring. When we estimate regressions without this re-scaling, that take into account the size of the population affected, the estimated coefficients mechanically decrease in magnitude as we decrease the size of the treated ring. This is simply the result of fewer individuals living in treated areas that are smaller.

[Insert Table 3 here]

3.2.2 Identifying house value effects

The previous results showed how cancer risk changes in the area around plants after they first report the emission of carcinogenic toxins. We now present the specification that we use to estimate house value effects. We let i denote the property, j the toxic establishment matched to property i, and t the year of the property transaction. The dependent variable is the logarithm of the property sale amount and the equation that we estimate is:

$$\log(\text{Sale amount})_{ijt} = \alpha + \beta_{\text{Post}\times\text{Distance}} \times \text{Post}_{it} \times \mathbb{1}_{ij}^{\text{Distance}_{ij} < X\text{miles}} + \gamma_{jt} + \gamma_i + \epsilon_{ijt}, \quad (3)$$

where $Post_{it}$ is an indicator variable taking a value of one for transactions of property i that take place in year t after the event year (year +1) and zero otherwise (transactions in years 0 and -1).[10] The indicator $\mathbb{1}_{ij}^{\text{Distance}_{ij} < X\text{miles}}$ takes a value of one if property i is within X miles of plant j, where X is 3, 2, 1.5, 1.25, or 1 mile, and zero for properties located between 3 and 5 miles of the same plant.

[10]We first 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), but later on we expand the event window to [-2,+2].

The above specification includes plant times property-sale-year fixed effects (γ_{jt}) that control for time-varying macroeconomic conditions in the area around the plant where the property is located and property fixed effects (γ_i) that control for time-invariant property characteristics. It provides within-property estimates of the effects of carcinogenic emissions on the value of properties located at various distances from the plants, controlling for time-varying local economic conditions.

Note that the empirical specification 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 use of repeated sales over a narrow time window helps with identification, but it means that there is sample selection in the properties included. More specifically, 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. 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 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 (at least partially) address the issue by dropping from the sample properties that are transacted twice in the same year.

In addition, Appendix Figure B.2 compares the distributions of sale prices for all transactions and for repeated transactions (roughly 17% of the total sample). The shapes of the distributions are similar, but the sample of repeated transactions has a slightly larger proportion of properties transacted below the median prices. In order to further address selection, we provide estimates for a wider window (of between -2 and +2) and when dropping from the sample properties with the largest changes in value between transactions.

Due to the geographical concentration of some of the plants 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. Properties located within 5 miles of multiple establishments that first satisfy the reporting criteria in the same year (same event year) are matched to the closest establishment. These choices avoid having multiple observations for the same property, potentially in both treatment and control groups.

4 House value effects

4.1 Baseline estimates

Table 4 panel A shows the results. The different columns provide the estimates for different treatment ring sizes. The estimated coefficient on the interaction term between post and distance in column (1) is negative and statistically significant: prices of properties within the 3-mile radius decrease by 1.2% relative to those between 3 and 5 miles after the event. 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 plant. The estimated decline in house values increases to 1.5% for a treated ring radius of 2 miles, before declining slightly as we decrease ring size further. Naturally, as we do so the number of transactions included in the regressions declines. The main result from the table is that the estimated declines in the value of treated properties relative to the control group are significant and relatively stable at around 1%. For an average house value of $200,000, the estimates imply a drop in price of $2,000.[11]

[Insert Table 4 here]

The estimates in Table 4 include events and house transactions surrounding unique plants located in the different US states. A question of interest is whether the results are driven by a few states, or whether they hold more generally across states. To address this question, we split our data into sub-samples by state and estimate separate regressions for each sub-sample focusing on the 3-mile treated group ring. Figure 5 plots the estimated coefficients on the Post $\times$ $\mathbb{1}^{Distance<3\ mile}$ variable for the different states. For the ones shown in grey, there is not enough data to perform the analysis. For most states the estimated coefficients are negative, indicating that our results are not driven by a small subset of them. (The figure plots the values of the estimated coefficients, some of which are not statistically significant.)

[11] In the Appendix Table B.2 we estimate regressions similar to those that we estimate for house prices, but with the number of property transactions as the dependent variable. The results show a decrease in the number of transactions after the event in the treated relative to the control group of 16 (for the 1-mile radius).

[Insert Figure 5 here]

The estimates of Table 4 panel A are for a tight [-1,1] time event window, which helps with the identification of the effects of the emission of carcinogenic toxins on house prices, but also reduces the number of properties in our regression sample. In order to estimate within-property changes, we need properties to be transacted at least twice in the event window, once before (in years -1 or 0) and once after the event (year +1). With such a narrow 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.

On the other hand, requiring properties to be transacted at least twice in such a short event window reduces the sample and may select specific types of properties. Therefore, in Table 4 panel B we repeat the analysis expanding the event window to [-2,+2]. That is, we now use properties that are transacted at least once before the event (in years -2, -1, or 0) and once in the two years after (years +1 and +2).[12] Table 4 panel B shows that estimated effects are similar to those in the previous table both in terms of statistical significance and economic magnitude. However, these results come with the caveat that for a wider window the effects of the toxic plant event may be confounded by other events.

For the diff-in-diff methodology to be valid, the parallel trends assumption must be satisfied. Figure 6 plots the estimated coefficient on the interaction term Post $\times$ $\mathbb{1}^{Distance<1miles}$ in event time and the corresponding 95% confidence intervals. The estimates are normalized to time zero and the standard errors are clustered at the property level. Before 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 6 here]

In the appendix, we show estimates for two additional robustness exercises. In the first, we restrict the sample to those events for which we have at least two hundred housing transactions per toxic plant. This means an average of at least fifty transactions in both the treated and control groups in the pre/post periods. Appendix Table B.3 shows that the results are robust with estimated house price effects in the treated group of -1.1% (for the 3-mile ring). In the

[12] For all properties that are transacted more than two times in the two years before the event, we include all the observed transactions in the regressions.

second, in Appendix Table B.4 we show results when using house prices in levels rather than in logs. The estimates imply house price declines of between roughly five and seven and a half thousand dollars depending on the size of the treated ring. The larger magnitude of the estimates compared to the benchmark values in Table 4 is due to the large skewness of the distribution of house price levels (the skewness value is 1.26).

4.1.1 A back-of-the-envelope calculation

We can use our results to estimate the value of an additional 1 year of statistical life. Assume that a person lives for a maximum number of n years. Furthermore, assume a baseline death rate of 1,043.8 deaths per 100,000 population, which is equal to the US average death rate in 2021. Out of all the death cases, 17.47% were due to cancer, which amounts to an annual death rate of 0.1824% due to cancer. Let c denote the probability of dying of cancer and p probability of dying due to reasons other than cancer. The expected number of years of life can be computed from:

$$E[\#years] = \sum_{j=1}^{n}(1-p-c)^{j-1}(p+c)j + (1-p-c)^{n}n. \tag{4}$$

We use the estimates in Table 3 that show an increase in cancer risk after the event for the treated group. In particular, we use the estimate of an increase in cancer score of 0.7 points. In order to translate this value into an increased probability of dying of cancer, we use the estimates of Boyle et al. (2023) that relate cancer risk score to death probabilities. Boyle et al. (2023) estimate the log of odds ratio as a linear function of the RSEI levels and their results imply an annual increase in the odds ratio of $exp(\hat{\beta})$ =1.52% per a unit increase in the RSEI level. Defining $\log(p_0/(1-p_0)) = \alpha$, where p_0 is the unconditional probability of dying of cancer before the event equal to 0.1824%, and $\log(p_1/(1-p_1)) = \alpha + \hat{\beta} \times 0.7$, where p_1 is the probability of dying of cancer after an increase in RSEI of 0.7, we obtain $p_1/p_0 = 1.0106$. Putting the two together implies that those living in the area near toxic plant N face a higher probability of dying of cancer $\bar{c} = 1.0106c$ after the event. This means an increase in the annual probability of dying from cancer from 0.1824% to 0.1843%.[13] Assuming that p remains unchanged after

[13]Naturally, there is uncertainty in these estimates, so below we perform calculations for other values.

the toxic event, the number of years that a person can expect to live is:

$$E[\#years|N] = \sum_{j=1}^{n} (1-p-\bar{c})^{j-1}(p+\bar{c})j + (1-p-\bar{c})^n n. \tag{5}$$

We compute the value of 1 additional year of statistical life by dividing the differential house price effect in the treated group by the difference in the expected number of years of life:

$$\text{Value of an additional 1 year of life} = \frac{\$2,000}{E[\#years] - E[\#years|N]}. \tag{6}$$

Figure 8 plots the value of 1 year of life computed for different levels of the increase in c (i.e. for $\bar{c}/c$). Naturally, the smaller the increase in the probability of death as a result of the event the larger the value of an additional year of life.[14] The dashed vertical line represents our estimate. If we multiply it by 80, we obtain \$3,227,241 as the value of life.

[Insert Figure 8 here]

4.1.2 Placebo

In order to provide evidence of the significance of our estimates, we carry out a placebo test of 1,000 bootstrap estimates of $\beta_{\text{Post} \times \text{Distance}}$. We generate a bootstrap sample by randomly drawing the distance between a property and its nearest toxic plant from a uniform distribution of between 0 and 5 miles. We then use this boostrap sample to estimate equation (8). Figure 7a plots the distribution of the estimated house price effects in these samples with randomized distance. The vertical line represents our baseline estimated value. It is significantly lower than the bootstrap values, confirming the significance of our estimate. We will come back to the right-hand panel of the figure below.

4.2 Economic mechanisms

In this section, we provide two additional pieces of evidence that provide insights on the mechanism behind the estimated house price effects. The first focuses on the nature of the information event, while the second explores heterogeneity in house price effects.

[14] In the limit, when there is no change in the probability of death, the value of an additional year of life is \$2,000 divided by zero or infinity.

4.2.1 Changes in the carcinogenic status of chemicals

Our events sample includes plants that were emitting chemicals that were not classified as carcinogens, but whose classification based on National Toxicology Program's Report on Carcinogens (RoC) subsequently changed. In this case, we take the event year to be the year of re-classification as a carcinogen. An advantage of using these events to identify the effects is that the re-classification of the chemicals does not depend on plants' decisions. In this section, we report results for the sample of plants that were producing these chemicals before the re-classification and continued their production afterward. There are 2,701 such plants in our sample. When we exclude multiple treatments this number is reduced to 1,585, which we use in this section.

In Panel A of Table 5 we show the estimates for changes in plant-level employment and sales around the events. The estimated coefficients are not significantly different from zero. This reflects the fact that the changes in the carcinogenic status of chemicals are not necessarily associated with increased production levels at the plants emitting those chemicals.

[Insert Table 5 here]

Panel B shows the house price effects, with statistically significant declines in the price of properties in the treated group compared to those in the control group. In terms of economic magnitude, and although the exact values depend on the size of the treated ring, the decline in the value of treated properties is around 2%. This decline is twice as large as the previous estimates that included all events. A possible explanation is that the change in the carcinogenic status of chemicals, and plants nearby currently emitting those chemicals, is a more salient event than when plants start emitting an already known carcinogen. This shows that the nature of the information event matters.

In terms of the back-of-the-envelope calculations that we have previously done, the 2% drop corresponds to an average house price decline of $4,000 and a value of one additional year of life of $80,681 or a value of life of $6,454,482. Figure 7b plots the results of a placebo test with randomized distance between properties and plants, similar to the one that we have previously described for the full sample, but for the restricted sample of RoC events. The figure shows that the estimated house price change in the treated group of -0.019 is significantly lower than the values obtained for the samples with randomized distance.

4.2.2 Heterogeneous effects on property prices

We now evaluate whether there are heterogeneous effects that depend on how expensive the properties are. For each event-plant j, first, we separate properties into above-median and below-median values based on their sale value between years -5 and -3 (relative to event year 0). This naturally requires that we observe a transaction for the property in this period. We create a dummy variable $Above_{ij}$ equal to one if the property i was transacted in [-5,-3] years before the event for a price above the median property price by plant-year and zero otherwise. The reason why we divide properties into above- and below-median using the [-5,-3] event window is so that we do not use the same price observation for separating them and for estimating the house price effects. But naturally, this choice has implications in terms of the number of observations available and the properties that are selected.

We then estimate regressions of changes in housing values around one year before and after the event year by interacting the *Post* dummy with the indicator for the distance from the plant and for the above median. The equation that we estimate is:

$$\begin{aligned}\log(\text{Sale amount})_{ijt} = \alpha &+ \beta_{Distance} \times \mathbb{1}_{ij}^{\text{Distance}_{ij}<X\text{miles}} + \beta_{Post} \times \text{Post}_{it} + \beta_{Above} \times \text{Above}_{ij} \\ &+ \beta_{\text{Above}\times\text{Distance}} \times \text{Above}_{ij} \times \mathbb{1}_{ij}^{\text{Distance}_{ij}<X\text{miles}} + \beta_{\text{Post}\times\text{Distance}} \\ &\times \text{Post}_{it} \times \mathbb{1}_{ij}^{\text{Distance}_{ij}<X\text{miles}} + \beta_{\text{Post}\times Above} \times \text{Post}_{it} \times \text{Above}_{ij} \\ &+ \beta_{\text{Post}\times\text{Distance}\times\text{Above}} \times \text{Post}_{it} \times \mathbb{1}_{ij}^{\text{Distance}_{ij}<X\text{miles}} \times \text{Above}_{ij} + \gamma_i + \gamma_{jt} + \epsilon_{ijt}.\end{aligned} \tag{7}$$

Table 6 shows the results. The bottom two rows show that differences between treated and control are more significant, both statistically and economically, for above than for below median properties. The estimated estimated drops for above median properties vary between 41 basis points (for a treated ring size of 3 miles) and 96 basis points (for a 1.25-mile ring size). Not all estimated differences are statistically significant, which may be due to the lower number of observations.

[Insert Table 6 here]

In order to address this, in Table 7, we provide estimates for a [-2,+2] event window. As before, the properties are divided into an above and below median value using years [-5,-3] transactions. The house price changes are more significant than before, both statistically and economically, with declines ranging between 70 basis points and 141 basis points for treated above-median properties. If more expensive (above median) properties are purchased by higher-income households, these results suggest that higher-income households either value environmental cancer risks more significantly (or they are better informed about them).

5 Neighborhood composition

We are interested in studying the implications of the toxic events for neighborhood composition in the narrow geographical areas around the plants, as it may shed light on the economic mechanism and on the heterogeneous response of households to the event. This requires that we are able to identify the characteristics of individuals living in the vicinity of the plants.

5.1 Identifying minorities

The existing census data is too coarse for this purpose. However, the Corelogic data includes the name of the sellers and buyers in property transactions. Furthermore, Sood (2017) and Laohaprapanon, Sood, and Naji (2022) use the 2017 Florida Voter Registration data as training and testing data and propose a classification algorithm for predicting the race and ethnicity of individuals from their first and last name. The Florida Voter Registration data has information on nearly 15 million voters along with their self-reported race, and treats race and ethnicity as one dimension with Hispanic treated as one group. We use their classification algorithm to predict the race and ethnicity of those involved in each property transaction in our data.

More precisely, we first use their algorithm to predict the likelihood that the individual is of Hispanic or African-American origin. We then construct a dummy variable $\mathbb{1}$(Minority) that takes the value one when the individual is predicted to be of Hispanic or African-American origin (and zero otherwise). We focus on Hispanic and African-American since the algorithm can identify individuals belonging to these groups with a higher degree of precision. But even within these two groups, Hispanic can be better identified than African-American (due to the nature of their names).

Using this methodology, we classify buyers of properties within a 5-mile ring of a toxic plant and 1 year of the toxic event. In our sample, out of all buyers, 14.61% and 4.05% are classified as Hispanic and African-American, respectively. As a comparison, we have obtained US homeownership data by race and ethnicity from the 2020 U.S. Census: 10.5% of homeowners are identified as Hispanic and 7.9% as African-American.[15] Therefore, our sample of homebuyers has a higher proportion of Hispanics and a lower proportion of African-American than the proportion of houses owned by these two groups as a whole. There could be two reasons for the difference. First, the areas where the events take place are different than the US average. In our regression analysis, we control for this by comparing the race and ethnicity of buyers before and after the event and focusing on the differences. Second, as explained, the algorithm tends to under-predict African-Americans.

For a small proportion of the buyers in our sample, we are able to obtain their FICO score from the LLMA data, that we use to further characterize them. The average FICO score in our sample is 703, which is below the average FICO score in the US population of 715 (Experian), and the average FICO score of 746 in the Fannie Mae Single-Family Loan Acquisition and Performance Data.[16] The FICO scores from Experian include both homeowners and renters, whereas our data is for house purchasers only, who tend to have higher FICO scores. In spite of this, the average FICO score in our data is lower, reflecting the fact that the sample of individuals who live near the offending plants is not representative of the US population.[17]

5.2 Results

The methodology that we use is similar to before. The dependent variable is the minority indicator variable (or the FICO score of the buyer) and the equation that we estimate is:

$$\log(\mathrm{y})_{ijt} = \alpha + \beta_{\mathrm{Post}\times\mathrm{Distance}} \times \mathrm{Post}_{it} \times \mathbb{1}_{ij}^{\mathrm{Distance}_{ij}<X\mathrm{miles}} + \gamma_{jt} + \gamma_i + \epsilon_{ijt}, \tag{8}$$

where the dependent variables and the set of fixed effects are similar to before. The event window is [-1,+1]. Our focus is on the identity of buyers in the post-event year and in the treated areas, relative to the pre-event and the control group, as measured by the estimated $\beta_{\mathrm{Post}\times\mathrm{Distance}}$. We focus on the buyers since the seller in the post-event property transaction is the buyer in the pre-event transaction of the same property, so our estimates already speak to

[15] The U.S. Census Bureau released an interactive map illustrating 2020 Census data about homeownership by the age, race, and ethnicity of the householder which can be found here.

[16] Data from the first quarter of 2000 to the third quarter of 2021.

[17] It would be interesting to study the effects of the event on rents, which capture the value of using the space, and are likely to reflect the demand for space by a lower income segment of the population than those covered in our data. Unfortunately, we were not able to obtain data on rents with the granularity needed for the identification.

the identity of the seller in the post-event transaction. In other words, what we are essentially doing is comparing buyers and sellers in the post-event transactions.

Panel A of Table 8 examines changes in buyer ethnicity. The estimated positive coefficient on the Post $\times \mathbb{1}^{\text{Distance}< \text{ X miles}}$ in column (1) shows a relative increase of 0.6 percentage points in minority home buyers in the vicinity of the treated plant in the year after the plant first reports the emission of carcinogenic toxins compared to the control group. The estimated coefficients remain at this level and are statistically significant as we decrease the treated ring size up to 1.25 miles.

[Insert Table 8 here]

In Panel B of Table 8, we focus on the FICO score of home buyers. The estimated coefficient in column (1) shows that the FICO score in the post-event treated group is 4.5 points lower than that of home buyers in the control group, and when compared to the pre-event period. The difference is statistically significant and it increases in economic magnitude and reaches -8 points as we decrease the size of the treated ring to 1.25 miles. For the smallest treated ring size, of 1 mile, the estimated change is still negative but not statistically significant, possibly due to the relatively small number of observations.

Overall, the results in this section provide the first evidence of *granular* changes in neighborhood composition around the plants towards minorities and higher credit-risk households after the toxic events.

6 A stylized model

In this section, we develop a stylized heterogeneous agent model that replicates the estimated effects. The goal is to provide evidence of the economic mechanisms behind the empirical estimates using comparative statics with respect to survival probabilities.

6.1 Model setup

There is a continuum of agents who supply one unit of labour inelastically and two types of workers with low and high wages $\underline{w}$ and $\overline{w}$, respectively. Workers are equally distributed across the two types. There are two regions: N is near to a toxic plant and F is further away. The two regions differ only in their exposure to cancer risk: $P(c \mid r)$ represents the probability of dying of cancer for agents living in region r (with $r = N, F$). For simplicity, we assume that $P(c \mid F)$ is equal to zero and $P(c \mid N)$ is greater than 0.[18]

Agent i living in region r derives utility from the consumption of a freely-tradable homogeneous good (x_i) and housing services (h_i) according to:

$$U_{i|r} = \log(x_i) + \gamma \log(h_{ir}) + \epsilon_{ir}, \tag{9}$$

where γ denotes the weight of housing in the utility function and ϵ_{ir} is a random term that represents worker i's idiosyncratic preferences for region r. We assume that $\epsilon_{iF} - \epsilon_{iN} \sim U[-s, s]$ and identically and independently distributed across types.[19]

Agents choose the region that they live in, the quantity of non-durable consumption and of housing services. The price of non-durable consumption is normalized to 1, and p_r denotes the per-unit price of housing services. Each agent i solves the following optimization problem:

$$V_{it} = \max_{x_{it}, h_{it}, r} U_{i|r} + (1 - P(c \mid r))\beta V_{i,t+1} \tag{10}$$

$$s.t. \quad w_i = x_i + p_r h_{ir} \tag{11}$$

taking as given the wage and the price of housing services in each of the regions. The constant elasticity structure implies that agents spend constant shares $\frac{1}{1+\gamma}$ and $\frac{\gamma}{1+\gamma}$ of their income on the tradable good and housing, respectively, so the demand functions are given by:

$$x_i^* = \frac{w_i}{1+\gamma} \tag{12}$$

$$h_{ir}^* = \frac{w_i \gamma}{1+\gamma} \frac{1}{p_r}. \tag{13}$$

Define $\tilde{U}_i = \log\left(\frac{w_i}{1+\gamma}\right) + \gamma \log\left(\frac{w_i \gamma}{1+\gamma}\right)$. The value for agent i of choosing location r,

$$V_{ir} = \frac{\tilde{U}_i - \gamma \log(p_r)}{1 - (1 - P(c \mid r))\beta} + \epsilon_{ir}. \tag{14}$$

[18] It would be straightforward to allow the two regions to also differ along other dimensions that the survival probability.

[19] Assuming that the two areas exhibit different levels of amenities would add two additional parameters to calibrate but would not affect the overall conclusions from the model.

Therefore, agent i chooses N rather than F if ad only if:

$$\frac{\tilde{U}_i - \gamma \log(p_N)}{1-(1-P(c \mid N))\beta} - \frac{\tilde{U}_i - \gamma \log(p_F)}{1-\beta} > \epsilon_{iF} - \epsilon_{iN}. \tag{15}$$

In order to close our stylized model, we assume that agents who die are replaced by new agents of the same type. We also assume that housing supply is fixed in each of the regions N and F and equal to H_N and H_F. In equilibrium the price of housing services in each region adjusts so that demand is equal to supply:

$$\int h^*_{iN} d\Phi_N(i) = H_N \tag{16}$$

$$\int h^*_{iF} d\Phi_F(i) = H_F \tag{17}$$

where the integral is taken over the demand of agents living the region.

6.2 Calibration and numerical example

The model includes 6 parameters, listed in Table 9. We use our data, parameter values from the literature, and model-implied quantities to choose their baseline values. We think of area F as the control area in our empirical analysis (i.e., between 3 and 5 miles from the toxic plants) and area N as the treated area near the plants.

In order to parameterize high and low income, and the proportion of workers living in each of the areas, in a first step, we use the FICO scores in the mortgage origination data. For those years before the events, i.e., years -1 and 0, and for those living within the 5-mile radius, we use the distribution of FICO scores to divide individuals into above and below median. The (median) values for individuals in each of these two groups are 760 and 655 points, respectively. Since our data does not include income information, we use the Fannie Mae origination sample from 2000Q1 to 2021Q3 to estimate a linear relationship between FICO and income. We then use this regression model to predict the income of individuals with 760 and 655 points, respectively, which gives the \$51,000 and \$38,000 reported in Table 9. Using the median value for FICO score calculated from all individuals in the 5-mile radius, we calculate the proportions of those above/below the median value living in each of the regions.

There are several estimates in the literature for the time discount factor; we use β equal to

0.96 from Gourinchas and Parker (2002). We set the parameter $\gamma = 0.4$ so that the model yields a reasonable value for the fraction of income that individuals spend on housing (0.29). The parameter determining the support of the idiosyncratic preferences s is calibrated to match the proportion of individuals in N who are high-income, 48%. Finally, the housing supply values H_F and H_N are set so the the model captures (a) the proportion of individuals in the control area before the event who are above-median FICO score, i.e., 52%; (b) the ratio of average prices in the control and treated are in the data, that is, 1.08. In particular, the calibrated model predicts that, *ceteris paribus*, a higher probability of dying in N is associated to lower prices of housing services in N relative to F.

Our comparative statics involves varying the probability of cancer mortality in region N, $P(c|N)$, from a benchmark figure of 0.1824%, which represents the annual cancer death rate in the United States. Figure 9 illustrates the variation in housing prices within the F and N regions as the value of $P(c|N)$ increases. Compared to a baseline scenario where $P(c|N)$ is 0.1824%, an increase in $P(c|N)$ leads to a rise in the price of housing services in the F area, while prices in the N area decline. Agents opting to reside in the N area are implicitly choosing a shorter expected lifespan and the response in Figure 9 can be attributed to the agents' preference in the model for a longer lifespan. An increase in the annual probability of dying from cancer $P(c|N)$ from 0.1824% to 0.1843%, that is, the same value used for the back-of-the-envelope calculations in Section 4.1.1, is associated to a relative decline of housing services in N relative to F of -0.5%. This adjustment is slightly smaller than the estimates presented in Table 4.

As the annual cancer mortality rate, $P(c|N)$, in region N increases, we observe a decline in demand among both high- and low-income groups. This dynamic prompts a migration pattern within the model, depicted in Figure 10. Notably, high-income individuals exhibit a more rapid migration response, leading to a decrease in the average FICO score in region N compared to region F. Assuming high-income individuals have a FICO score of 760 and low-income individuals have a score of 655—consistent with the values used for model calibration—the resulting relative decrease in FICO scores is considerably less than observed in actual data. Specifically, the decrease is -0.12 for a $P(c|N)$ increase to 0.1843% and -1 for an increase to 0.2043%, which is less pronounced than the changes reported in the data. One possible reason is that the model only has migration across regions, but not inflows from and to other areas. However, qualitatively the model matches the overall patterns documented in the empirical section.

7 Conclusion

We offer novel empirical evidence on the adverse effects of increased cancer mortality on property values and community demographics. Focusing on a narrow window around the time in which a plant first reports the emission of known carcinogenic toxins to the EPA, we show a) an increase in RSEI cancer scores associated with the proximity to the plant, and b) a decline in property values ranging from -0.8% to -2% for houses closer to the plant relative to houses farther from the plant. A back-of-the-envelope calculation implies a value of statistical life ranging from \$2.6M to \$6.4M.

The estimated declines in property values mask meaningful heterogeneity. Specifically, we find that more expensive properties are more sensitive to the reporting event and that the differences in house values between treatment and control groups are more significant, both statistically and economically, for above- than for below-median-price properties. Finally, our analysis extends to the implications of these toxic events on neighborhood composition, revealing a shift towards a higher presence of minority and higher credit-risk households in affected areas. This shift provides important evidence on the social dynamics at play in the wake of a change in the perceived environmental hazards, further contributing to the ongoing discussion on environmental justice.

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.

Boyle, J., M. H. Ward, J. R. Cerhan, N. Rothman, and D. C. Wheeler (2023). Modeling historic environmental pollutant exposures and non-hodgkin lymphoma risk. *Environmental Research 224*, 115506.

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*.

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 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.

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*.

Gormley, T. A. and D. A. Matsa (2011). Growing out of trouble? corporate responses to liability risk. *The Review of Financial Studies 24*(8), 2781–2821.

Gourinchas, P.-O. and J. A. Parker (2002). Consumption over the life cycle. *Econometrica 70*(1), 47–89.

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.

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.

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

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.

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).

Sood, G. (2017). Florida Voter Registration Data (2017 and 2022).

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.

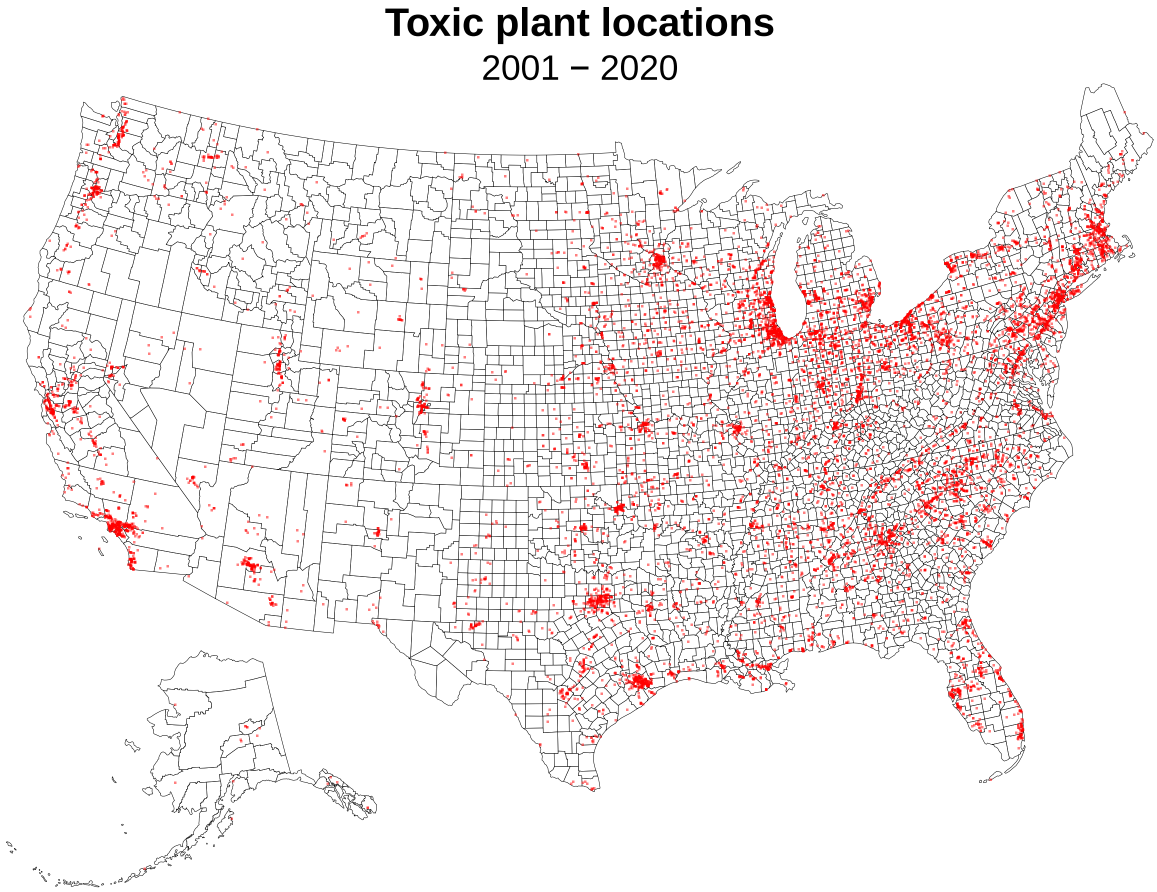


Figure 1: Location of the reporting plants.
Notes: The figure shows the location of plants that report the emission of harmful carcinogen pollutants for the first time during the sample period. We exclude all plants for which the first reporting year is the same as the opening year. The data are from the EPA's TRI program between 2001 and 2020.

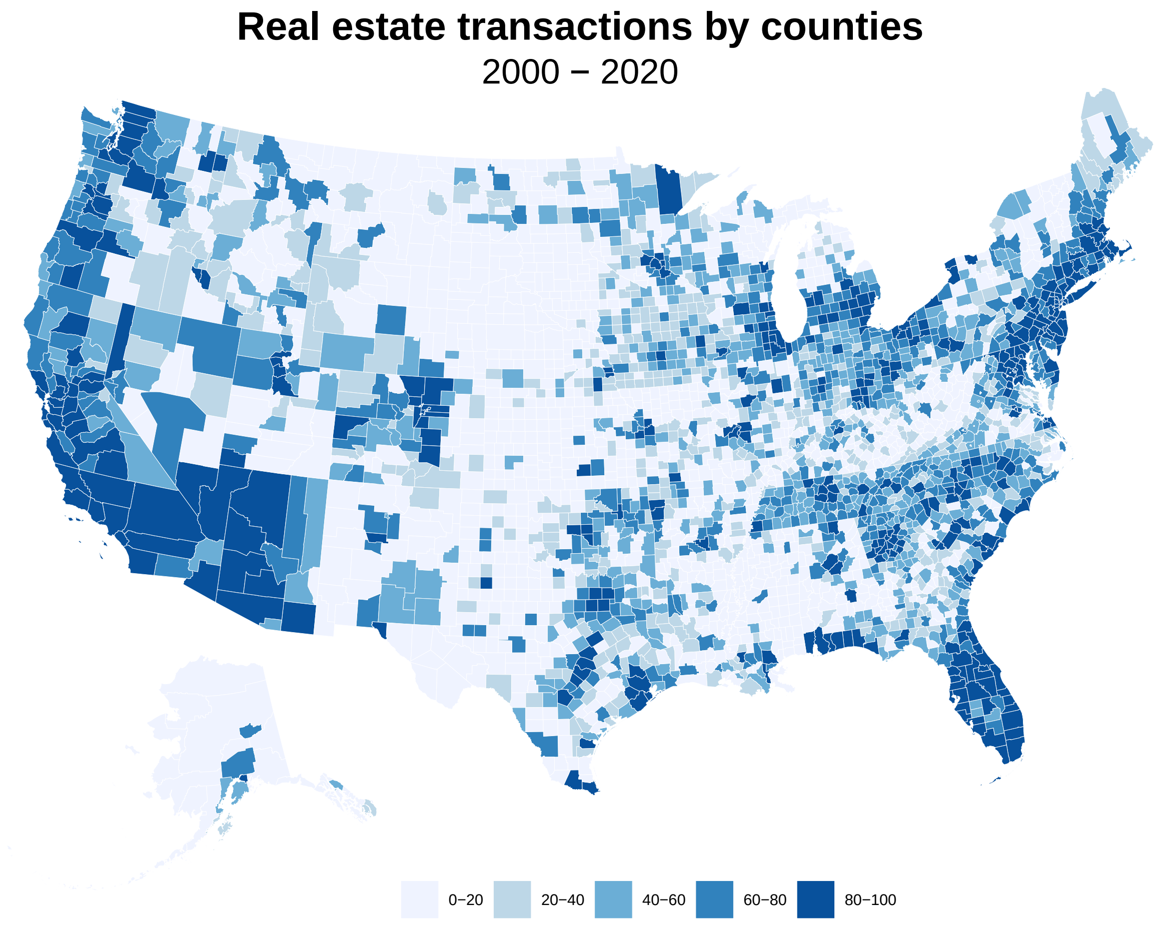


Figure 2: Number of real estate transactions by county.
Notes: The data are from the Corelogic Deed & Tax record data from 2000 and 2020. We calculate the number of housing transactions by county, and based on these we sort counties into 5 ordered bins, from the ones with the least to the most transactions.

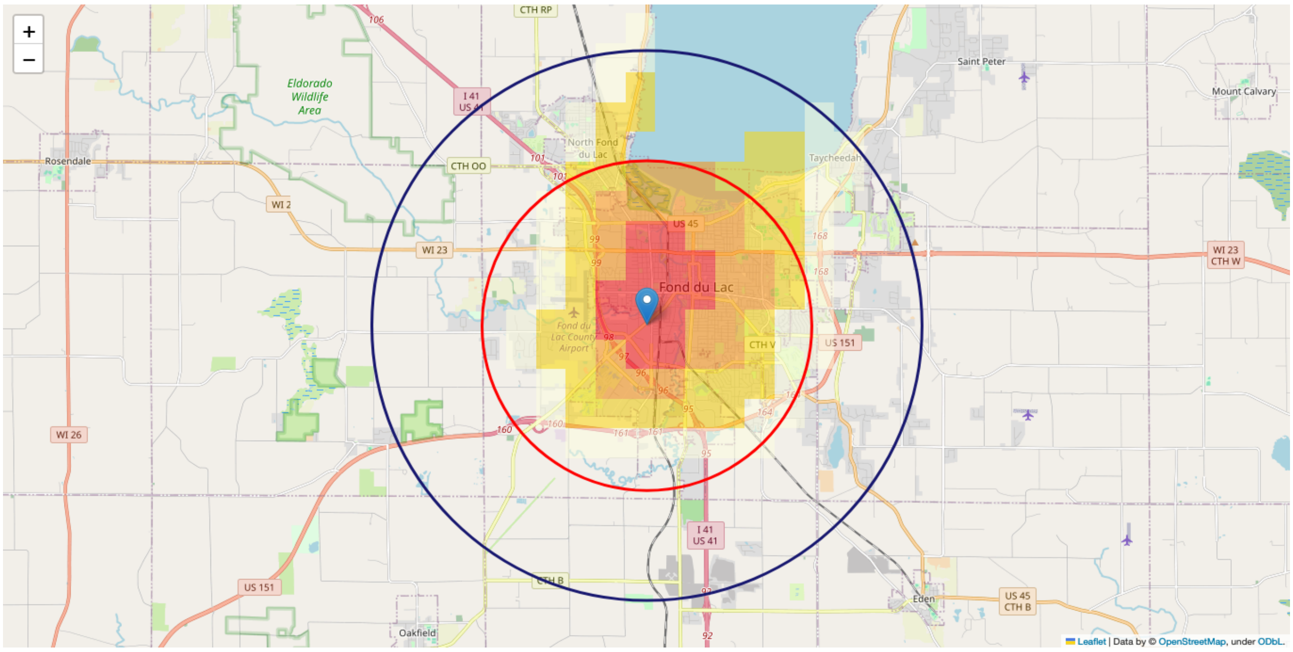


Figure 3: Heatmap of RSEI Cancer score by grid cells for Brenner Tank LLC, Fond du Lac, WI.

Notes: The figure shows the heatmap by grid cells (810m × 810m grids) for the RSEI cancer scores aggregated for toxic chemicals released by Brenner Tank LLC, Fond du Lac, WI 54935 , in 2002. 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 plant, and darker-colored grid cells show a higher cancer risk. The red circle defines an area with a three-mile radius of the plant, whereas the blue circle defines an area with a five-mile radius of the plant. The red area around the plant represents cancer score values above 6; the orange area represents values between 2.50 and 6; the yellow area represents values between 1.35 and 2.5; the light-yellow area between 1 and 1.35; values below 1 are not shown. The figure has been produced using the Folium Python Library and Leaflet maps (http://leafletjs.com/).

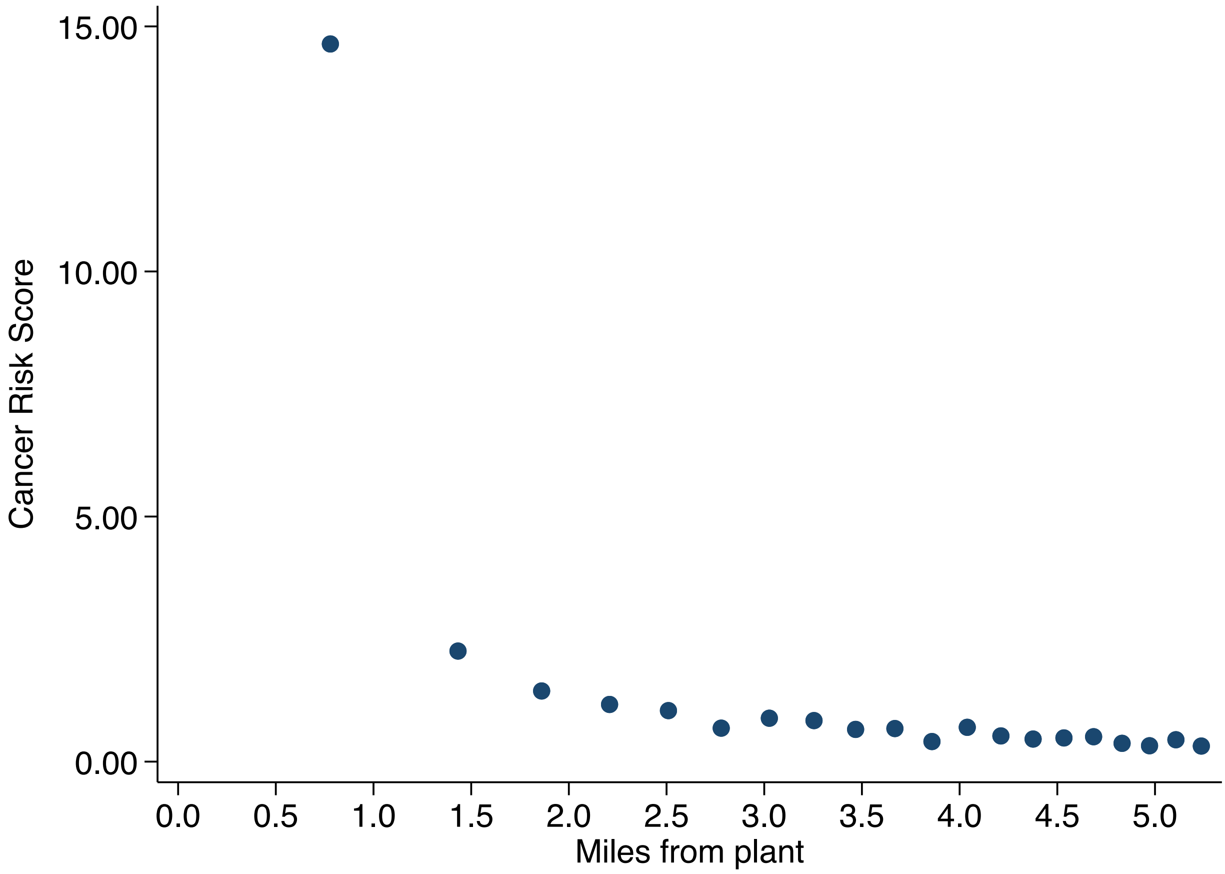


Figure 4: The effect of toxic plants on modeled cancer risk.
Notes: The figure shows the binscatter plot for modeled cancer risk for 810m $\times$ 810m grid cells as a function of their distance from a toxic plant. Specifically, we obtained disaggregated geographic microdata from the Risk-Screening Environmental Indicators (RSEI) from the EPA. The EPA models the impact of chemical releases from toxic plants on grid cells (810m $\times$ 810m grids) using estimated dosage, its toxic concentrations, and potentially exposed populations and provides a unitless score (RSEI Cancer score) that captures the effect of chemical releases on cancer. We scale the level of the RSEI cancer scores by the number of exposed population for each grid cell and compute cancer risk at the grid level. Please see the text for more details.

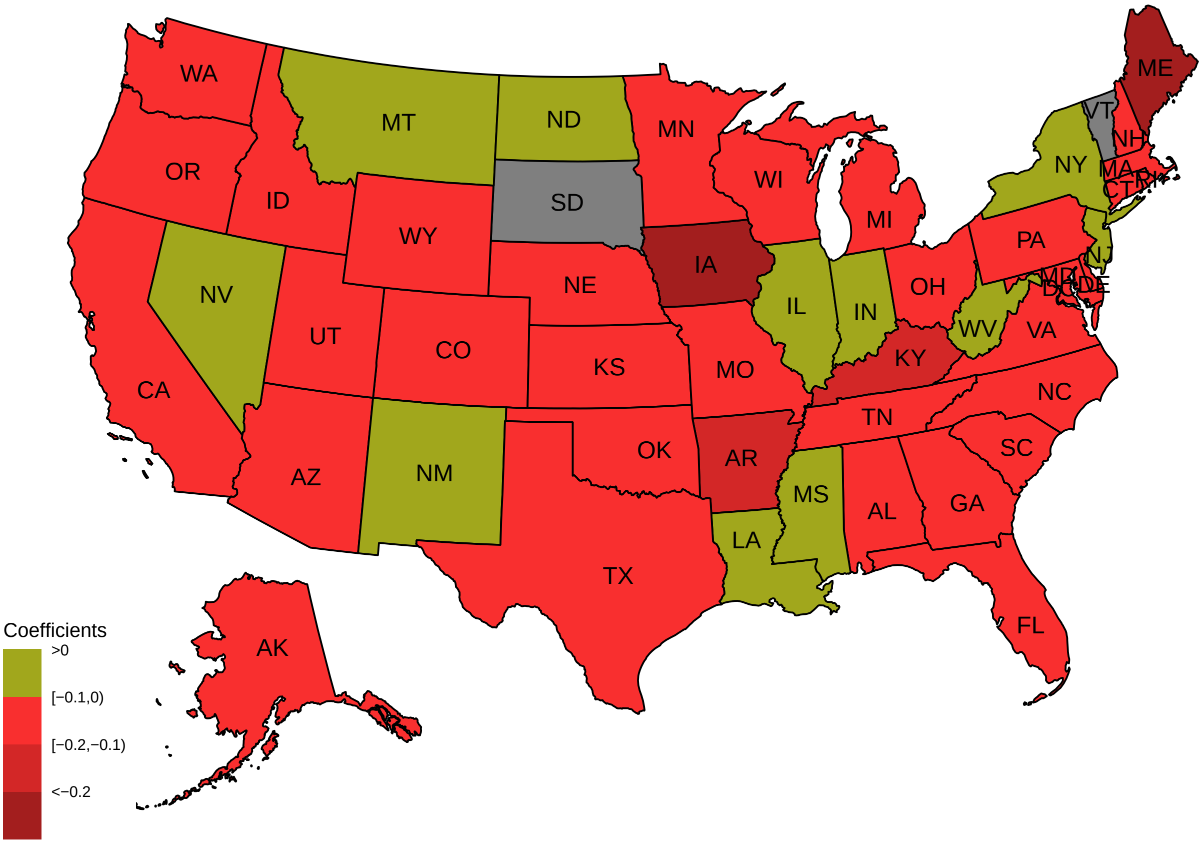


Figure 5: Estimated housing value effects by state.
Notes: The figure presents regression estimates $\beta_{\text{Post}\times\text{Distance}}$ by state for the impact of toxic plants on the price of houses within 5 miles of the toxic plant. For each state where the property is located, we estimate the following specification:

$$\log(\text{Sale amount})_{ijt} = \alpha + \beta_{\text{Post}\times\text{Distance}} \times \text{Post}_{it} \times \mathbb{1}_{ij}^{\text{Distance}_{ij}<3\text{miles}} + \gamma_i + \gamma_{jt} + \epsilon_{ijt}.$$

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 the treatment rings based on the property's distance from the nearest toxic plant. Specifically, $\mathbb{1}_{ij}^{\text{Distance}_{ij}<3\text{ miles}}$ takes a value of one if property i is within 3 miles from a plant j, and zero for properties between 3 and 5 miles of the same plant (control ring). All regressions include property and plant $\times$ sale-year fixed effects. The analysis focuses on changes within one year of a plant's first report of carcinogenic emissions to the EPA's TRI program.

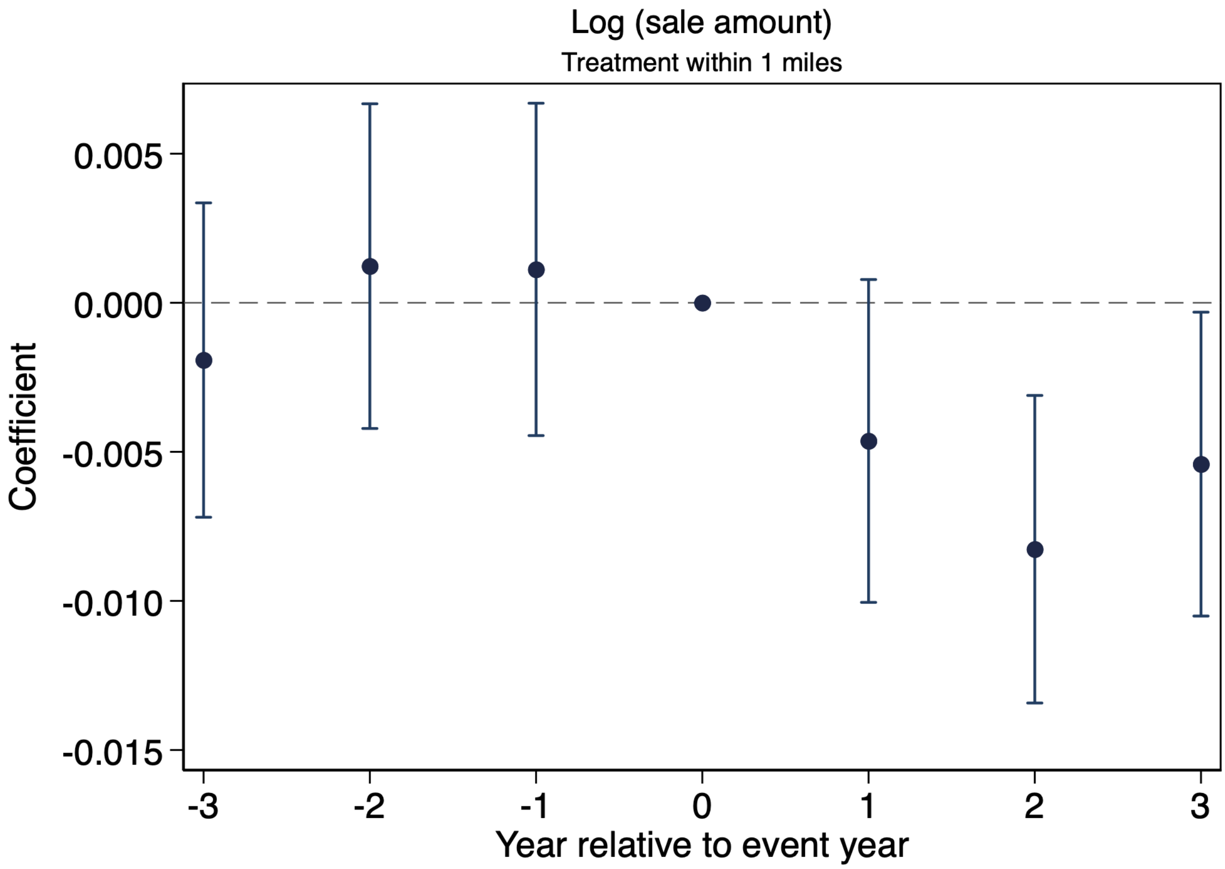
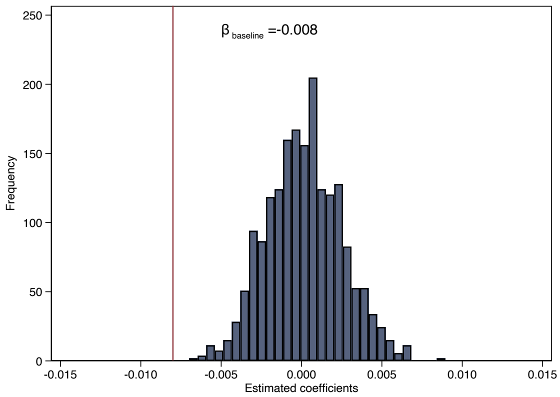


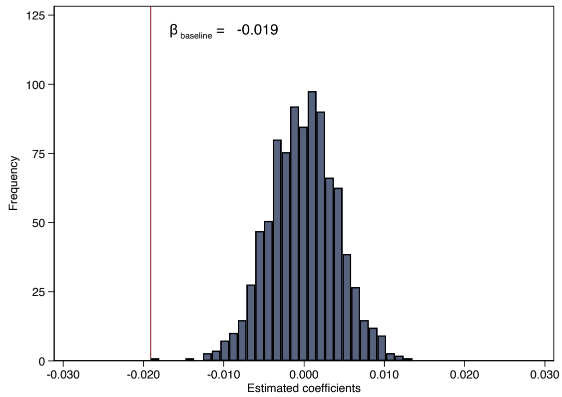
Figure 6: Changes in housing values around the events.
Notes: The figure plots the coefficient on $\text{Post}_{it} \times \mathbb{1}_{ij}^{\text{Distance}_{ij}<1\text{miles}}$ in event time and the corresponding 95% confidence interval. Specifically, we estimate the following dynamic Difference-in-Differences equation:

$$\text{Log (sale amount)}_{ijt} = \alpha + \sum_{k=-3}^{-1} \beta_k \times \mathbb{1}_{ij}^{\text{Distance}<1\text{miles}} + \sum_{k=1}^{3} \beta_k \times \mathbb{1}_{ij}^{\text{Distance}<1\text{miles}} + \gamma_i + \gamma_{jt} + \epsilon_{ijt}$$

We plot coefficients, β_k, on relative differences between properties within 1 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 [-3,3] years relative to the first year the plant reports emitting carcinogenic toxins in the EPA's TRI program. The regression includes property (γ_i) and plant $\times$ sale-year (γ_{jt}) fixed effects. Standard errors are clustered at the property level.



(a) All events



(b) Change in carcinogenic status of chemical

Figure 7: Distribution of bootstrap estimates.
Notes: The figure plots the distribution of 1,000 bootstrap estimates of the $\beta_{\text{Post}\times\text{Distance}}$ from Equation (8). In each bootstrap sample, we randomly draw the distance between a property and its nearest toxic plant from a uniform distribution from 0 to 5 miles and re-estimate Equation (8) in the bootstrap sample. Panel A focuses on changes within one year of a plant's first report of carcinogenic emissions to the EPA's TRI program. Panel B utilize the National Toxicology Program's Report on Carcinogens (RoC) to determine when chemicals were identified as known or reasonably anticipated human carcinogens and uses the year of the RoC as the event year. In panel A, the vertical line shows the treatment effect from column 5 in Table 4 panel A while in panel B, the vertical line shows the treatment effect from column 5 in Table 5 Panel B.

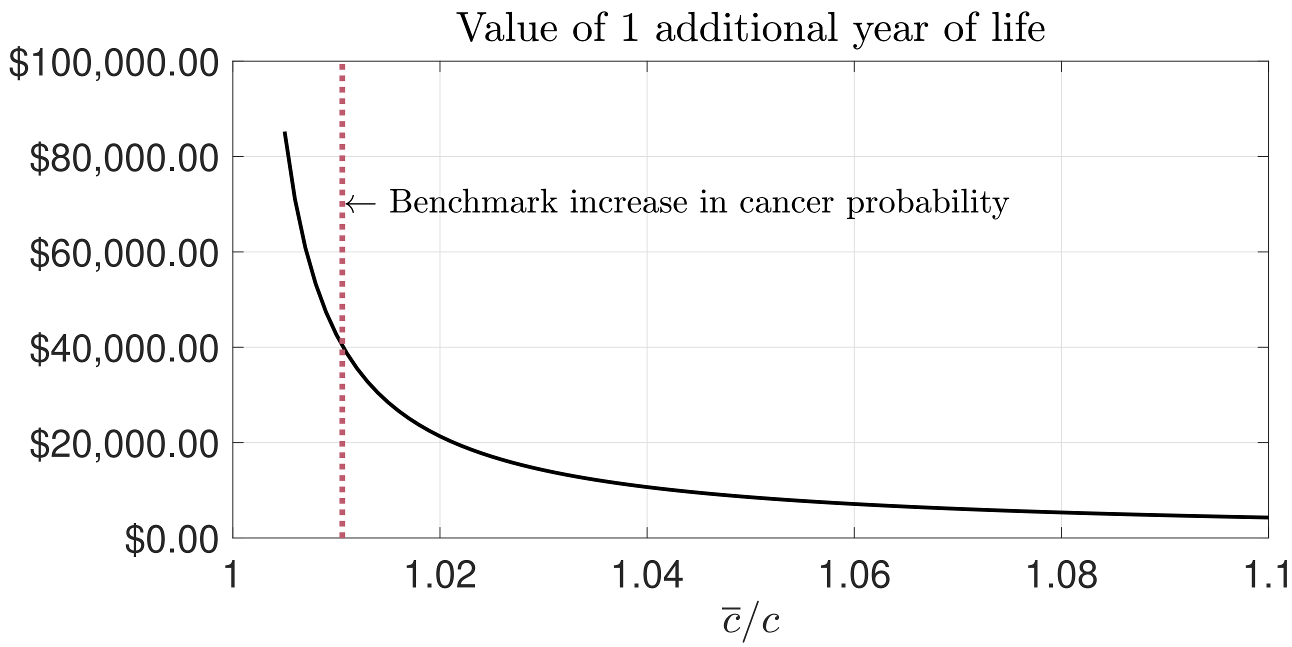


Figure 8: Dollar value of 1 additional year of life.
Notes: The figure shows the dollar value of 1 additional year of life implied by our empirical estimates for different probabilities c of dying of cancer in a given year. Calculations are described in Section 4.1.1.

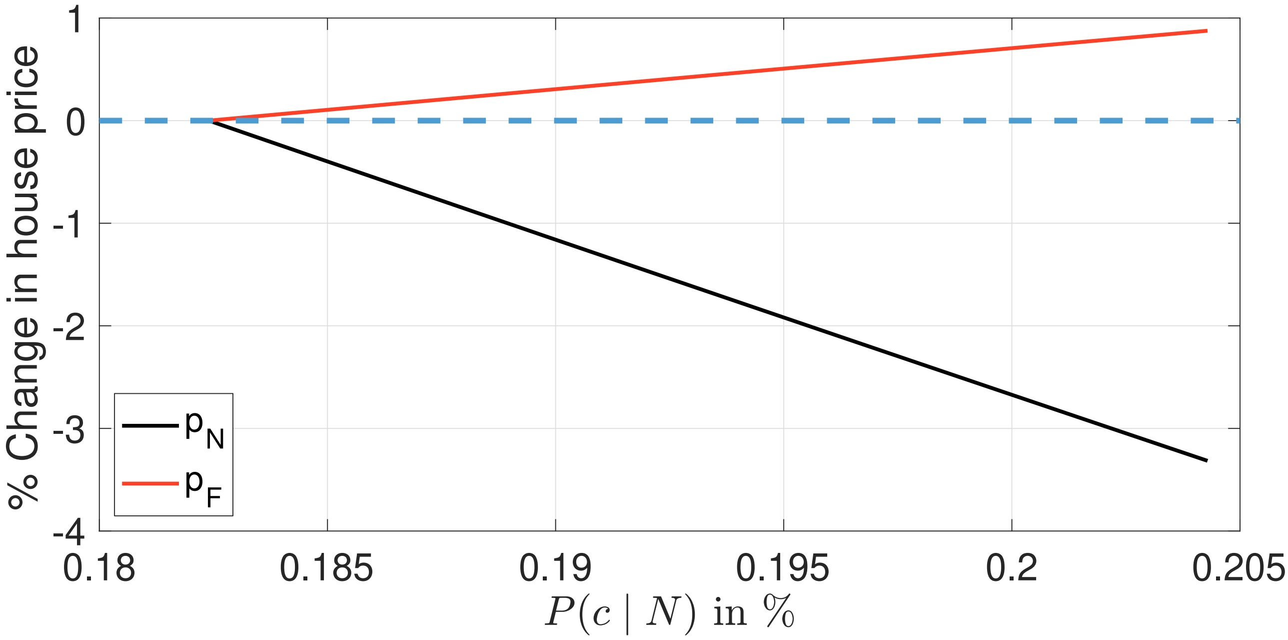
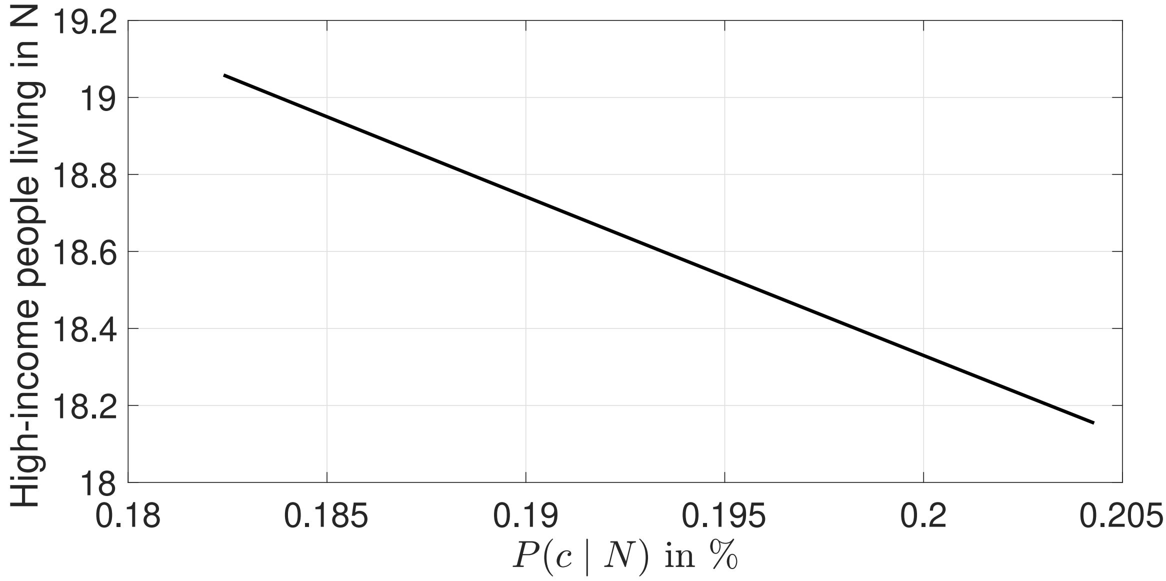


Figure 9: Model-implied price changes.
Notes: Our comparative static involves increasing the probability of cancer mortality in region N, $P(c|N)$, from a benchmark figure of 0.1824%. The figure shows model-implied price changes for different levels of cancer mortality. The model is described in Section 6.



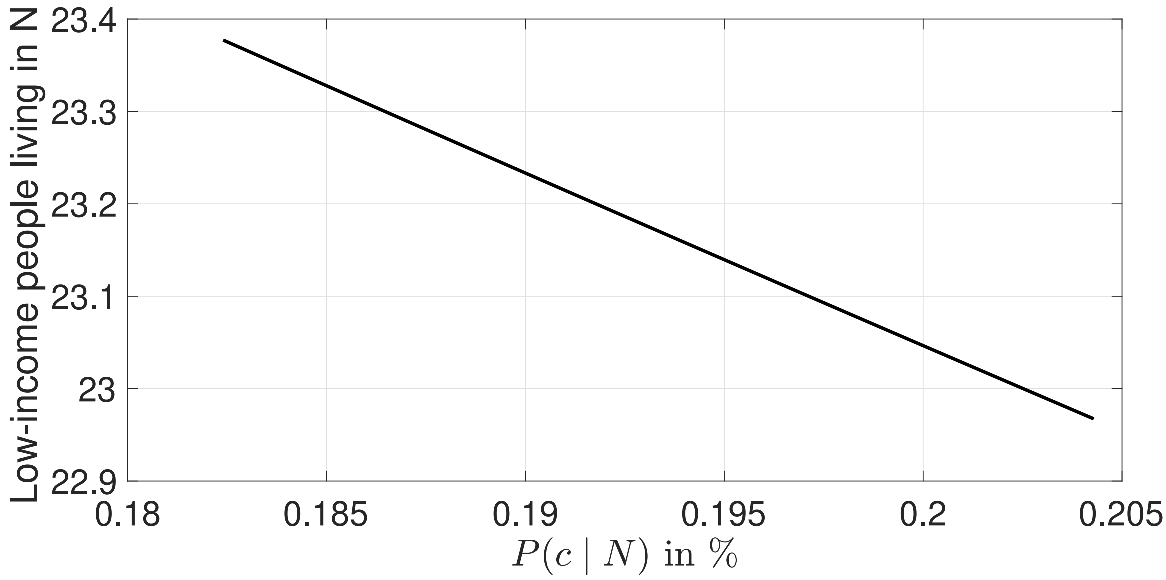


Figure 10: Model-implied migration.
Notes: Our comparative static involves increasing the probability of cancer mortality in region N, $P(c|N)$, from a benchmark figure of 0.1824%. The figure shows the model-implied proportion of high-income households and low-income households choosing to locate near the toxic plant for different levels of cancer mortality. The model is described in Section 6.

Table 1: Summary Statistics

Notes: This table presents the frequency of toxic plants emitting carcinogen chemicals by industry and chemical toxicity. Panel A reports the industry distribution of toxic plants in our sample by their primary two-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. In panel B, we report the distribution of toxic chemicals for the set of plant-events in our sample. We present the toxicity classification as reported by International Agency for Research on Cancer (IARC) and National Toxicology Program (NTP). Specifically, IARC uses the following classification scheme: 1 —The chemical is carcinogenic to humans; 2A —the chemical is probably carcinogenic to humans; 2B —the chemical is possibly carcinogenic to humans. NTP uses the following classification scheme: K —The chemical is known to be a human carcinogen; RA —The chemical is reasonably anticipated to be a human carcinogen. Lastly, – denotes missing classification. These classifications obtained from the November 2019 update of the "Toxics Release Inventory (TRI) Basis of OSHA Carcinogens."

Panel A: Industry distribution of toxic plants

Two-digit NAICS	Industry description	Fraction total(%)
11	Agriculture, Forestry, Fishing and Hunting	0.06
21	Mining, Quarrying, and Oil and Gas Extraction	0.95
22	Utilities	1.32
23	Construction	0.15
31-33	Manufacturing	87.89
42	Wholesale Trade	4.85
44-45	Retail Trade	0.11
48-49	Transportation and Warehousing	0.09
51	Information	0.01
54	Professional, Scientific, and Technical Services	0.15
56	Administrative & Support and Waste Management & Remediation Services	0.92
61	Educational Services	0.01
62	Health Care and Social Assistance	0.04
71	Arts, Entertainment, and Recreation	0.01
81	Other Services (except Public Administration)	0.05
92	Public Administration	3.38

Panel B: Distribution of toxic chemicals at plants

Chemical	CAS Number	NTP classification	IARC classification	Fraction total(%)
Lead	7439-92-1	RA	2A	47.15
Nickel	7440-02-0	RA	2B	16.65
Polycyclic aromatic compounds	N590	P	2B	8.90
Styrene	100-42-5	RA	2B	6.97
Ethylbenzene	100-41-4	-	2B	6.42
Naphthalene	91-20-3	RA	2B	5.25
Nickel compounds	N495	RA	1	4.77
Benzene	71-43-2	K	1	4.11
Formaldehyde	50-00-0	K	1	2.59
Methyl isobutyl ketone	108-10-1	-	2B	2.10

Table 2: Effects on employment and sales, Plant-level evidence using matched sample

Notes: This table presents regression estimates on changes in plant-level employment and sales within two years 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. $Treated_{jt}$, is an indicator variable taking a value of one for the set of plant-events and zero, for a matched set of control plant-events. See Online Appendix A for more details on matching. All regressions include plant, industry × year, and state × year fixed effects. We allow these fixed effects to be estimated within each pair of treated and control plant-events. Standard errors are double clustered at the plant and year level. Standard errors are clustered at the plant level. ***, **, * denote significance at the 1%, 5%, and 10% level, respectively.

Dependent variable:	Log (employment)	Log (sales)
	(1)	(2)
Post × Treated	0.051***	0.052***
	(0.008)	(0.008)
Plant fixed effects	Yes	Yes
Industry × year fixed effects	Yes	Yes
State × year fixed effects	Yes	Yes
R^2	0.98	0.98
Observations	906,720	905,500

Table 3: Changes in cancer risk scores around the events

Notes: This table presents regression estimates for the impact of toxic plants on nearby RSEI cancer scores, computed for each 810m x 810m grid, within 5 miles from the toxic plant. The analysis focuses on changes within one year of a plant's first report of carcinogenic emissions to the EPA's TRI program. The dependent variable is the level of the RSEI cancer scores scaled by the number of people 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 property's distance from a toxic plant. Specifically, $\mathbb{1}_{lj}^{\text{Distance}_{lj}<X\text{miles}}$ takes a value of one if the grid l 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 grids between 3 and 5 miles of the same plant. The empirical specification is as follows:

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

All regressions include plant $\times$ year fixed effects and grid 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 scaled by number of people				
Treatment (Distance in miles)	3	2	1.5	1.25	1
Post $\times$ $\mathbb{1}_{Distance<Xmiles}$	0.175***	0.324***	0.505***	0.672***	0.956***
	(0.024)	(0.046)	(0.081)	(0.116)	(0.180)
Property fixed effects	Yes	Yes	Yes	Yes	Yes
Year$\times$ Plant fixed effects	Yes	Yes	Yes	Yes	Yes
R^2	0.86	0.87	0.87	0.87	0.87
Observations	1,439,426	1,185,567	1,095,294	1,059,642	1,030,113

Table 4: Changes in house prices around a plant's first report of carcinogenic emissions

This table presents regression estimates for the impact of toxic plants on the price of houses within 5 miles of the toxic plant. The analysis focuses on changes within one year (Panel A) or within two years (Panel B) of a plant's first report of carcinogenic emissions to the EPA's TRI program. 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 property's distance from the nearest 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). The empirical specification is as follows:

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

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

	Panel A: Event window [-1,+1]				
Dependent variable:	Log(sale amount)				
Treatment (Distance in miles)	3	2	1.5	1.25	1
	(1)	(2)	(3)	(4)	(5)
Post × $\mathbb{1}_{Distance<Xmiles}$	-0.012***	-0.015***	-0.012***	-0.010***	-0.008**
	(0.001)	(0.002)	(0.002)	(0.003)	(0.003)
Property fixed effects	Yes	Yes	Yes	Yes	Yes
Sale-year× plant fixed effects	Yes	Yes	Yes	Yes	Yes
R^2	0.86	0.86	0.86	0.86	0.86
Observations	1,506,933	1,161,063	1,017,934	958,329	909,049

	Panel B: Event window [-2,+2]				
Dependent variable:	Log(sale amount)				
Treatment (Distance in miles)	3	2	1.5	1.25	1
	(1)	(2)	(3)	(4)	(5)
Post × $\mathbb{1}_{Distance<Xmiles}$	-0.007***	-0.010***	-0.009***	-0.009***	-0.006***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
Property fixed effects	Yes	Yes	Yes	Yes	Yes
Sale-year × plant fixed effects	Yes	Yes	Yes	Yes	Yes
R^2	0.87	0.87	0.87	0.87	0.87
Observations	4,431,190	3,418,925	3,001,801	2,828,846	2,683,027

Table 5: Changes in the carcinogenic status of chemicals

This table presents regression estimates on changes in plant-level employment and sales (Panel A) and on the impact of toxic plants on the price of houses within 5 miles of the toxic plant (Panel B). We utilize the National Toxicology Program's Report on Carcinogens (RoC) to determine when chemicals were identified as known or reasonably anticipated human carcinogens. The year of the RoC is the event year. The sample is restricted to plants that were producing these chemicals before their classification in the RoC and continued their production afterward. The dependent variable is the natural logarithm of the plant employment and sales (Panel A) and the natural logarithm of the sale amount of a property (Panel B). The independent variable in Panel A is $Post_t$, i.e., an indicator variable taking a value if year t is after the event year and zero otherwise. The independent variable, $Post_{it}$, in Panel B 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. The independent variable, $Post_{it}$, is an indicator variable taking a value of one for all years after the event year and zero otherwise. $Treated_{jt}$, is an indicator variable taking a value of one for the set of plant-events and zero, for a matched set of control plant-events. See Online Appendix A for more details on matching. Regressions include plant, industry × year, and state × year fixed effects. We allow these fixed effects to be estimated within each pair of treated and control plant-events. Standard errors are double clustered at the plant and year level. In Panel B, we define five treatment rings based on the property's distance from the nearest 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). The empirical specification is as follows:

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

Standard errors are double clustered at the plant and year level in Panel A and are clustered at the property level in Panel B. ***, **, * denote significance at the 1%, 5%, and 10% level, respectively.

	Panel A: Plant-level employment and sales	
Dependent variable:	Log (employment)	Log (sales)
	(1)	(2)
Post × Treated	-0.013	-0.012
	(0.011)	(0.012)
Plant fixed effects	Yes	Yes
Industry × year fixed effects	Yes	Yes
State × year fixed effects	Yes	Yes
R^2	0.99	0.99
Observations	108,628	108,256

	Panel B: Changes in house prices				
Dependent variable:	Log(sale amount)				
Treatment (Distance in miles)	3	2	1.5	1.25	1
	(1)	(2)	(3)	(4)	(5)
Post × $\mathbb{1}_{Distance<Xmiles}$	-0.010***	-0.023***	-0.022***	-0.019***	-0.019***
	(0.003)	(0.003)	(0.004)	(0.005)	(0.006)
Property fixed effects	Yes	Yes	Yes	Yes	Yes
Sale-year× plant fixed effects	Yes	Yes	Yes	Yes	Yes
R^2	0.87	0.87	0.87	0.87	0.87
Observations	443,565	337,679	290,089	268,741	251,589

Table 6: Heterogeneity by price: Above median

Notes: This table presents regression estimates for the impact of toxic plants on the price of houses within 5 miles of the toxic plant. The analysis focuses on changes within one year of a plant's first report of carcinogenic emissions to the EPA's TRI program. 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 was sold in the year following the plant's emission report. We define five treatment rings based on the property's distance 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 separate higher and lower-priced properties: we include an indicator variable $Above_i$ for whether property i sale price was above the median value, calculated from sales 3 to 5 years before the emissions report for properties around a plant. The empirical specification is as follows:

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

All regressions include property and plant × sale-year fixed effects. Standard errors are clustered at the property level. ***, **, * 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
Below median					
Treated	-0.0072**	-0.0101***	-0.0119**	0	0.0039
	(0.003)	(0.0039)	(0.0048)	(0.0056)	(0.007)
Above median					
Control	-0.0367***	-0.0365***	-0.0369***	-0.037***	-0.0372***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Treated	-0.044***	-0.049***	-0.0481***	-0.0485***	-0.0462***
	(0.0035)	(0.0046)	(0.0059)	(0.0071)	(0.0091)
Difference: Treated minus Control					
Below median	-0.0072**	-0.0101***	-0.0119**	0	0.0039
	(0.003)	(0.0039)	(0.0048)	(0.0056)	(0.007)
Above median	-0.0073**	-0.0125***	-0.0113*	-0.0115	-0.009
	(0.0035)	(0.0047)	(0.0059)	(0.0071)	(0.0091)
Property fixed effects	Yes	Yes	Yes	Yes	Yes
Sale-year × plant fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	316,499	242,398	211,692	199,020	188,630

Table 7: Heterogeneity by price: Above median, event window [-2,+2]

Notes: This table presents regression estimates for the impact of toxic plants on the price of houses within 5 miles of the toxic plant. The analysis focuses on changes within two years of a plant's first report of carcinogenic emissions to the EPA's TRI program. 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 was sold in the year following the plant's emission report. We define five treatment rings based on the property's distance 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 separate higher and lower-priced properties: we include an indicator variable $Above_i$ for whether property i sale price was above the median value, calculated from sales 3 to 5 years before the emissions report for properties around a plant. The empirical specification is as follows:

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

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

Treatment (Distance in miles)	3	2	1.5	1.25	1
Below median					
Treated	-0.002	-0.0043*	-0.0054*	-9e-04	9e-04
	(0.0018)	(0.0022)	(0.0028)	(0.0032)	(0.004)
Above median					
Control	-0.0346***	-0.0349***	-0.0352***	-0.0353***	-0.035***
	(0.0017)	(0.0017)	(0.0017)	(0.0017)	(0.0018)
Treated	-0.0416***	-0.0444***	-0.0451***	-0.0494***	-0.0457***
	(0.002)	(0.0027)	(0.0034)	(0.0041)	(0.0053)
Difference: Treated minus Control					
Below median	-0.002	-0.0043*	-0.0054*	-9e-04	9e-04
	(0.0018)	(0.0022)	(0.0028)	(0.0032)	(0.004)
Above median	-0.007***	-0.0095***	-0.0099***	-0.0141***	-0.0107**
	(0.002)	(0.0027)	(0.0035)	(0.0041)	(0.0053)
Property fixed effects	Yes	Yes	Yes	Yes	Yes
Sale-year × plant fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	903,333	695,451	609,632	574,151	544,358

Table 8: Changes in neighbourhood composition

This table presents regression estimates for the impact of toxic plants on the neighborhood composition within 5 miles of the toxic plant. The analysis focuses on changes within one year of a plant's first report of carcinogenic emissions to the EPA's TRI program. The dependent variable y is a dummy variable $\mathbb{1}(Minority)_{it}$ taking a value of 1 if property i in year t was purchased by a buyer of Hispanic or African-American heritage (Panel A) or the FICO Score of the buyer when available (Panel B). 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 property's distance from the nearest 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). The empirical specification is as follows:

$$y_{ijt} = \alpha + \beta_{\text{Post}\times\text{Distance}} \times \text{Post}_{it} \times \mathbb{1}_{ij}^{\text{Distance}_{ij}<X\text{miles}} + \gamma_i + \gamma_{jt} + \epsilon_{ijt}.$$

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

	Panel A: Changes in ethnic minorities				
Dependent variable:	$\mathbb{1}(Minority)$				
Treatment (Distance in miles)	3	2	1.5	1.25	1
Post $\times$ $\mathbb{1}_{Distance<Xmiles}$	0.006***	0.007***	0.006**	0.007*	0.003
	(0.002)	(0.002)	(0.003)	(0.003)	(0.004)
Property fixed effects	Yes	Yes	Yes	Yes	Yes
Sale-year$\times$ plant fixed effects	Yes	Yes	Yes	Yes	Yes
R^2	0.65	0.65	0.65	0.65	0.65
Observations	938,995	719,847	630,468	592,684	561,496

	Panel B: Changes in FICO Score				
Dependent variable:	FICO Score				
Treatment (Distance in miles)	3	2	1.5	1.25	1
Post $\times$ $\mathbb{1}_{Distance<Xmiles}$	-4.515***	-4.857**	-7.394***	-8.054**	-5.859
	(1.659)	(2.253)	(2.832)	(3.395)	(4.149)
Property fixed effects	Yes	Yes	Yes	Yes	Yes
Sale-year$\times$ plant fixed effects	Yes	Yes	Yes	Yes	Yes
R^2	0.81	0.81	0.82	0.82	0.82
Observations	76,721	58,842	51,727	48,930	46,719

Table 9: Calibrated or externally fixed parameters

Parameter	Value	Description	Target moment or source	Data	Model
Income					
$\overline{w}$	51	High income value	1st-stage estimation		
$\underline{w}$	38	Low income value	1st-stage estimation		
Preference					
β	0.93	Discount factor	Gourinchas and Parker (2002)		
γ	0.4	Housing weight	Spending on housing services (BLS, 2022)	0.258	0.286
s	2	Idiosyncratic preferences	% of individuals in N who are high-income	48.1%	44.9%
Housing supply					
H_F	5	Supply of housing in F	% of individuals in F who are high-income	50.8%	51.4%
H_N	1.5	Supply of housing in N	Ratio of average house prices in F vs N	1.081	1.135

Notes: For each parameter, the table reports the parameter symbol, numerical value, a description, the target moment or source, and if a target moment is used to calibrate the parameter, the table also shows the data and model-implied moment.

Online Appendix

Table of Contents

A Additional data description

A.1 Data construction

Plant selection. The starting point of our analyses are the EPA's TRI Basic Plus data files which collectively contain all of the data reported by plants on the TRI Reporting Form R and Form A Certification Statement. We begin with the file for the reporting year of 2000 and retain plant observations that are marked as carcinogens and those that fall under the purview of the Clean Air Act. We then iterate over other reporting years 2001 to 2020, ensuring that each year we retain new plant additions, relative to the previous reporting years. We combine all plants across the years and exclude plants with missing geo-location information and those from the year 2000. This leaves us with a sample of 14,787 unique toxic plant-events. To ensure that we capture plants that were reporting to the TRI in years other than the one they were established, we merge these toxic plants with the NETS database which contains detailed information on the first year of establishment. Applying this filter yields a sample of 11,561 plant-events.

Merge with Corelogic. Next, we merge location information on toxic plants to the universe of residential property transactions in the continental United States as reported by the Corelogic Deeds & Tax Records. We begin with the property information and restrict our attention to single-family residences, condominiums, duplexes, and apartments. We then exclude properties with missing information on their location (block-level latitude and longitude), year of sale, and the year in which the property was built. Subsequently, we retain transactions where the buyer purchased the property using cash or a mortgage. Subsequently, we merge the housing transactions with plants by iterating over them and calculating distances (in miles) between each property and plant. Once the distances are calculated, we retain properties within a 5-mile radius of each plant and ensure that there are at least five housing transactions around each plant.

Lastly, when a property is associated with multiple plant-events we retain the first year among the associated events and the closest toxic plant. We also exclude houses with multiple transactions in the same year. Further, to mitigate the outsized effects of very small houses (e.g., mobile homes) and very large houses (e.g., mansions), we trim the data to remove transactions at the top and bottom 5% of the empirical distribution of the sale amount.

Merge with RSEI Geographic Microdata. To measure changes in cancer risk around each plant, we rely on the modeled-estimate provided by the EPA. Specifically, it uses the TRI data to construct Risk-Screening Environmental Indicators (RSEI) that measure potential risks to human health and the environment. The data are drawn 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 obtain data at the most granular spatial unit - grid cells of dimension 810m $\times$ 810m. For each grid cell, we observe the cancer risk score associated with a plant. The score is a unitless measure computed for each chemical and media as a product of the estimated dosage released by that specific toxic plant, the toxic concentrations, and the potentially exposed population. For each grid-cell, we retain observations associated with on-site chemical releases and all media. We then compute the cancer-risk score as the ratio of RSEI cancer risk score divided by the potentially exposed population, and then aggregate at the grid-cell level. Lastly, when a grid-cell is associated with multiple plant-events we retain the first year among the associated events and the closest toxic plant.

Identifying changes in the carcinogenic status of chemicals. We hand-collected information on changes in the carcinogenic status of chemicals using the National Toxicology Program's (NTP) *Report on Carcinogens (RoC)*, following Gormley and Matsa (2011). The report is published by the U.S. Department of Health and Human Services under a mandate by the Congress introduced in 1978. The first two reports were published in 1980 and 1981, and the subsequent reports have been updated approximately biannually since. These reports provide information on scientific discoveries related to chemical-specific carcinogenicity and the associated timing. Each report provides information on new carcinogens and changes in the status of chemicals identified in previous reports. Therefore, any changes to the status or addition of chemicals to the report are indicative of the scientific consensus that these chemicals are likely to be carcinogens.

We focus on reports published during the sample period that include the years 2001, 2002, 2004, 2011, and 2016. In our analyses, we use chemicals that are either (i) newly identified as carcinogens, or (ii) that changed status from unknown to reasonably anticipated to be human carcinogens (RAHC). Based on these criteria, we identify 23 specific chemicals and match them to plants that report using these chemicals from the TRI database and were already reporting the usage of these chemicals in the 2 years before the change in their status. The Chemical Abstract Service (CAS) Registry Number unique to chemicals allows us to cleanly link the information across the two datasets.[20]

Plant-level matching for employment and sales. For each plant-event observed in our dataset, we construct a control plant-event in the following manner. Specifically, in each event-year, control plants are drawn from the distribution of toxic plants observed in the NETS database reporting to the TRI program. Subsequently, we limit to plants that operate in the same state and industry (NAICS six-digit) as treated plants. We then assign the same event-year as a pseudo event-year to the matched control plant. In total, for 8,881 set of treated plant-events, we obtain 20,218 control plant-events. We repeat the exercise for the subset of plants that experience a change in carcinogenicity of the chemicals that they use obtained from the *RoC*. For this subset of events, we obtain 7,545 control plant-events for 1,266 treated plant-events. Note that in our estimations, we allow for our effects to be estimated within each pair of treated and control plants.

A.2 Number of events

Our dataset encompasses 11,561 plant events, out of which 10,421 involved transactions executed with cash or through mortgage agreements for houses located within a 5-mile radius of the plants. For each of these houses, we calculated the first year a plant within a 5-mile radius of the house reported emitting human carcinogens to the EPA and designated this as the event-year. This process resulted in a reduction of the events to 8,204 plant-events. We filtered our dataset by retaining only the plant closest to the transacted properties, ultimately narrowing it down to 7,801 plant-events. To reduce the impact of outliers, we eliminated transactions at the extreme ends of the price spectrum, specifically those below $30,000 (5th percentile of the distribution) or above $700,000 (95th percentile of the distribution), which resulted in a slight reduction to 7,768 plant-events. Lastly, we applied temporal constraints to our dataset, focusing on transactions occurring within one year before or after the plant event, which left us with 6,405 plant-events. When considering a larger window of two years before and after the

[20]The Comprehensive Environmental Response, Compensation, and Liability Act of 1980 (CERCLA, 1980) enacted by the Congress aids in the identification of the designated hazardous substances by providing the CAS Registry Numbers. These are unique numeric identifiers associated with designated hazardous substances and are provided for the convenience of the regulated community and the public (see, 40 CFR 302.4).

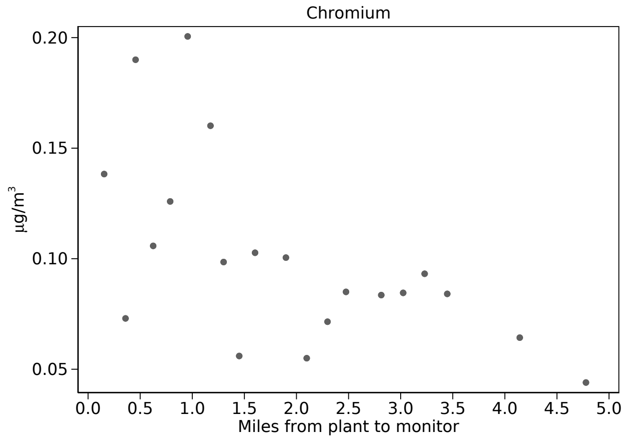
event, our sample size slightly increases to 6,696 plant-events.

A.3 Air pollution measured from monitors

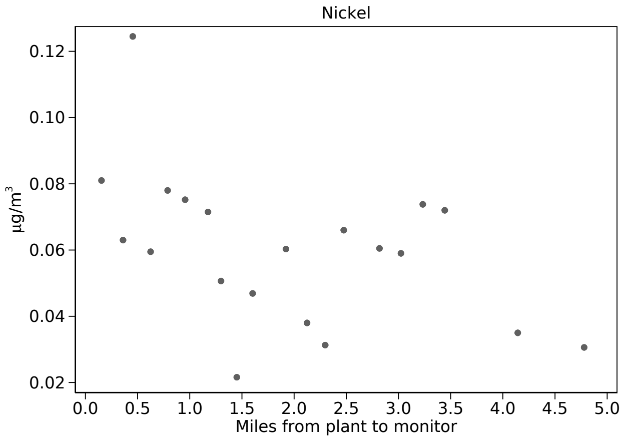
In the main paper we use measures of cancer risk near the plants from the EPA. These measures use as an input to the calculations the TRI data. Therefore, they are an independent observation of the higher cancer risk near the plants. In this section we provide direct evidence of cancer risk near the plants using data on ambient air quality across the US from the the Air Quality System (AQS).

The AQS data are collected by the EPA using a network of over 10,000 monitoring stations located throughout the United States. They measure various pollutants, including ozone, particulate matter, carbon monoxide, nitrogen oxides, sulfur dioxide, and lead. The data are collected on an hourly or daily basis, depending on the pollutant being measured. Moreover, the data are publicly available, providing information on local area air quality to households. We extract readings from all 266 air monitoring stations that are within a five-mile radius of the 11,561 plants included in our sample. We use readings for all years around the events. When interpreting the data, it is important to keep in mind that the location of the monitors and plants is likely to be endogenous.

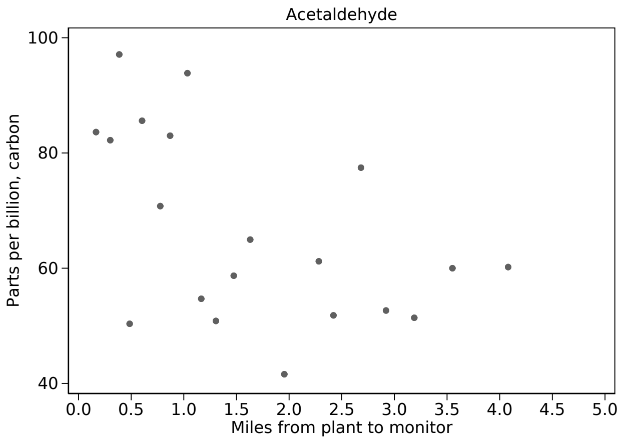
Figure A.1 presents a binscatter plot of the relation between distance from the plant and air monitor (within a 5-mile radius) and the concentration of four predominant carcinogenic pollutants. Although the data are noisy, as it is based on only 266 air monitors, it shows a more pronounced concentration of toxins up to a distance of 1 to 1.5 miles from the plant, varying by pollutant. This evidence on the importance of the distance from the plant for air quality is consistent with that for cancer risk scores.



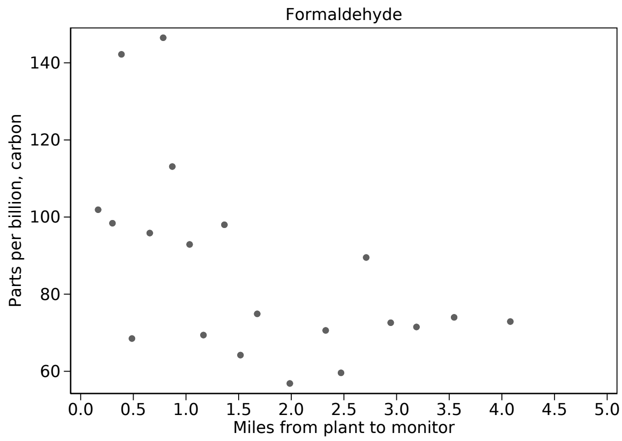
(a) Chromium



(b) Nickel



(c) Acetaldehyde



(d) Formaldehyde

Figure A.1: The effect of toxic plants on hazardous air pollution.
Notes: The figure shows the scatter plot for the concentration of 4 toxic air pollutants as a function of distance between the monitor and the operating toxic plant. Hazardous air pollutants, also known as toxic air pollutants, are defined by the EPA as "pollutants that are known or suspected to cause cancer or other serious health effects, such as reproductive effects or birth defects, or adverse environmental effects."

B Additional results

REGISTER FILE PHOTO

In addition to selling corn, farmers could sell the plant remnants, called stover, to ethanol plants to be burned in the production of the fuel. But that idea could lose merit if the Environmental Protection Agency considers the emissions to be a greenhouse gas.

How it works

The Environmental Protection Agency's greenhouse gas regulations would sweep all of the ethanol industry into the paperwork and fee requirements of the Clean Air Act because of the way it counts carbon emissions. A plant that produces 100 million gallons of ethanol a year would emit about 325,000 tons of carbon dioxide. The agency will start regulating sources of more than 100,000 tons of greenhouse gases.

Permits in Iowa

Fifteen Iowa ethanol plants already have air pollution permits:

- Archer Daniels Midland, Clinton and Cedar Rapids
- Cargill, Eddyville
- Corn LP, Goldfield
- Big River Resources, West Burlington
- Global Ethanol, Lakota
- Grain Processing Corp., Muscatine
- Golden Grain, Mason City
- Homeland Energy Solutions, Lawler
- Lincolnway Energy, Nevada
- Little Sioux Corn Processors, Marcus
- Penford Products, Cedar Rapids
- Platinum Ethanol, Arthur
- Poet Biorefining, Coon Rapids
- Southwest Iowa Renewable Energy, Council Bluffs

EMISSIONS

FROM PAGE 1D

Act because of other pollutants such as nitrogen oxide that they emit.

What the greenhouse-gas regulations will do is sweep the rest of the state's plants under the Clean Air Act, and require them to start filing reports on emissions and paying fees on their pollutants as well, said Marnie Stein of the DNR's Air Quality Bureau.

The fees would likely average $5,600 to $11,200 a year, she said.

"Those fees could be quite costly for some ethanol plants," said Geoff Cooper, who follows regulatory issues for the Renewable Fuels Association, a Washington trade group.

No one questions that fermenting corn or burning biomass for electricity puts carbon dioxide into the atmosphere.

But most greenhouse gas calculations, including those of the United Nations' climate panel, don't include such emissions, known as biogenic, because they come from biological sources.

The reason: The carbon released from corn or biomass will eventually be returned to earth as crops and other biomass sources are replanted. Plants take carbon dioxide out of the air as they grow.

Coal, on the other hand, is dug from the earth and never replaced, so its carbon is lost into the atmosphere when the coal is burned.

In a bit of irony, some analysts believe the rules could actually discourage ethanol plants from cutting their use of fossil fuels.

Here's how: The environmental agency is expected to eventually go beyond demanding paperwork and fees from greenhouse-gas sources and start requiring them to reduce emissions.

An easy way for ethanol producers to do that would be to stop running their plants with coal or natural gas and instead burn corn stalks and other sources of biomass. But that won't work if the agency continues to count emissions from burning biomass as greenhouse gases, said Nathaniel Baer, who follows energy policy for the Iowa Environmental Council.

"If the biomass is considered carbon neutral, then it seems like it would create a viable compliance option down the road for reducing emissions" from ethanol plants, Baer said.

The environmental agency announced this summer that it was taking a second look at the emissions issue and asked for industry and public comment.

In the meantime, the DNR is taking steps to implement the regulations in Iowa and has no choice but to mirror what the environmental agency does, Stein said. If the state tried to exempt the emissions from corn fermentation, as the ethanol industry wants, the agency would under federal law override the state rules, she said.

If the agency reverses itself on ethanol and biomass emissions, the state will follow suit later, she said.

The Renewable Fuels Association wants the state to exempt the emissions now without waiting on the environmental agency, noting that even if the agency doesn't change its rules they could be struck down by the courts.

Not all producers are worried about the regulations.

Green Plains Renewable Energy is experimenting at its Shenandoah plant with reducing its carbon dioxide emissions by using the gas to grow algae that could be used for fuel or feed. Green Plains hopes to eventually offer the technology to ethanol plants and other companies that want to reduce carbon dioxide emissions, a spokesman said.

(a) Des Moines

A town wants answers

The town of LaPlace, Louisiana, lies along the Mississippi River, a stone's throw from Lake Pontchartrain and the Maurepas Swamp. It sits in the heart of an area that's become known by locals as "Cancer Alley," a vast industrial stretch between New Orleans and Baton Rouge where dozens of petrochemical plants dot the landscape.

One sign posted by a local advocacy group warns the town's 29,000 residents that they are "more likely to get cancer due to chloroprene air emissions." The warning refers to the Denka Performance Elastomer plant at the edge of town, where a vast network of pipes and valves stand testament to industry.

The facility looms over Fifth Ward Elementary School, where children run around the playground oblivious to the toxic emissions in the air.

CNN

The Denka plant is the nation's sole producer of neoprene, the synthetic rubber used in everything from wetsuits to automotive hoses. It's also at the center of a battle over its emissions.

The plant, formerly operated by DuPont, employs more than 200 workers and has been in this spot for nearly 50 years. The facility plays a vital manufacturing role as the nation's only producer of neoprene, a synthetic rubber that's found in everything from gaskets and hoses to fishing waders and wetsuits. But it also emits 99% of the nation's chloroprene pollution, according to the EPA. Chloroprene is the main chemical used in the production of neoprene.

In 2010, the EPA determined that chloroprene is "likely carcinogenic to humans," meaning studies showed it likely causes cancer in people. The EPA has not set a legal limit for chloroprene emissions. But according to a May 2016 memo, federal regulators said the "upper limit of acceptability" for cancer risk was an annual average of 0.2 micrograms of chloroprene per cubic meter. Anything more than that would increase the risk of developing cancer, the EPA determined.

RELATED ARTICLE

Pollution linked to 9 million deaths worldwide in 2015, study says

Residents say they were largely unaware of the 2010 EPA finding. But in December 2015, the EPA updated its National Air Toxics Assessment map, which showed an elevated risk of cancer around the plant – prompting Denka to enter into an agreement with the state of Louisiana to voluntarily reduce chloroprene emissions by 85%.

(b) CNN

Figure B.1: Local and national news coverage of two events in our sample. This figure shows an excerpt from the Des Moines Register from Des Moines, Iowa in 2010 and from the CNN on Denka Performance elastometer plant in 2017.

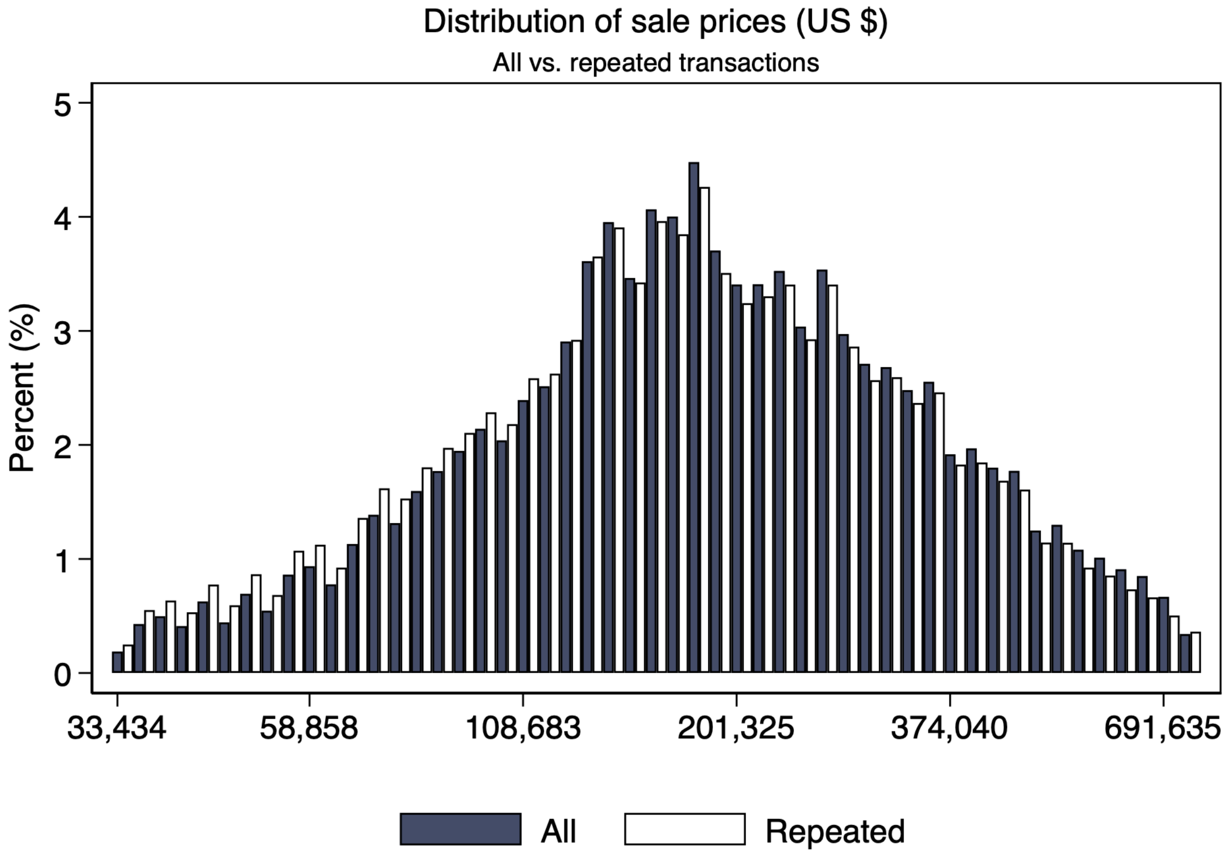


Figure B.2: Empirical distribution of sale prices.
Notes: The figure shows the empirical distribution of sale prices for all transactions in our sample (blue solid bars) and the sample of repeated transactions (white hollow bars). The sample is restricted to sale prices one year around the first year a toxic plant reports emitting carcinogenic pollutants.

Table B.1: Changes in county-level economic activity around events

Notes: This table presents regression estimates of changes in county-level outcomes 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 population (employment) while in column 3 (column 4) it is the nominal wages (personal income). In column 5, the dependent variable is the gross domestic product. The independent variable, $Post_{it}$, is an indicator variable taking a value of one for all years after the event year and zero otherwise while $\mathbb{1}_{Treated}$, is an indicator variable taking a value of one for the counties with atleast one plant reporting a new carcinogenic toxin and zero otherwise. All regressions include county 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:	Population	Employment	Wages	Income	GDP
	(1)	(2)	(3)	(4)	(5)
Post	-0.000***	-0.000*	-0.000	0.000***	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Post × $\mathbb{1}_{Treated}$	0.014***	0.007***	-0.008***	0.005***	0.010***
	(0.001)	(0.002)	(0.001)	(0.002)	(0.003)
County fixed effects	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes
R^2	1.00	0.99	0.93	0.99	0.97
Observations	120,158	131,396	131,396	120,158	117,191

Table B.2: Number of transactions around the events

Notes: This table presents regression estimates for the impact of toxic plants on the number of house transactions within 5 miles of the toxic plant. The analysis focuses on changes within one year of a plant's first report of carcinogenic emissions to the EPA's TRI program. The dependent variable is the number of transactions, *# transactions*, computed separately for the treated area and the control area around the plant. Specifically, we define 5 different treated areas, i.e., within 3, 2, 1.5, 1.25, or 1 mile (columns 1 to 5), and one control area between 3 and 5 miles of the same plant. The independent variable, $Post_t$, is an indicator variable taking a value of one if the year t is after the event year and zero otherwise. The dummy variable $Treated_i$ takes a value of one if the area i is a treated area and 0 otherwise. The empirical specification is as follows:

$$\#\text{ transactions}_{it} = \alpha + \beta_{\text{Post}\times Treated_i} \times Treated_i \times \text{Post}_t + \gamma_i + \gamma_{jt} + \epsilon_{ijt}.$$

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

Dependent variable:	# transactions				
Treatment (Distance in miles)	3	2	1.5	1.25	1
	(1)	(2)	(3)	(4)	(5)
Post × Treated	-0.126	-9.031***	-12.578***	-14.514***	-16.486***
	(1.537)	(2.416)	(2.888)	(3.104)	(3.275)
Sale-year× plant fixed effects	Yes	Yes	Yes	Yes	Yes
R^2	0.90	0.73	0.64	0.61	0.59
Observations	35,624	35,624	35,624	35,624	35,624

Table B.3: Robustness, more than 199 observations per toxic plant

This table presents regression estimates for the impact of toxic plants on the price of houses within 5 miles of the toxic plant. The analysis focuses on changes within one year of a plant's first report of carcinogenic emissions to the EPA's TRI program. We restrict the sample to toxic plants for which we have more than 199 observations in the 3-year window [-1,+1]. 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 property's distance from the nearest 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). The empirical specification is as follows:

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

All regressions include property and plant $\times$ sale-year fixed effects. Standard errors are clustered at the property level. ***, **, * 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 $\times$ $\mathbb{1}_{Distance<Xmiles}$	-0.011***	-0.014***	-0.011***	-0.009***	-0.006
	(0.001)	(0.002)	(0.002)	(0.003)	(0.004)
Property fixed effects	Yes	Yes	Yes	Yes	Yes
Sale-year $\times$ plant fixed effect	Yes	Yes	Yes	Yes	Yes
R^2	0.86	0.86	0.86	0.86	0.86
Observations	1,346,031	1,027,637	897,361	843,979	800,023

Table B.4: Changes in house prices, with property fixed effects

This table presents regression estimates for the impact of toxic plants on the price of houses within 5 miles of the toxic plant. The analysis focuses on changes within one year of a plant's first report of carcinogenic emissions to the EPA's TRI program. The dependent variable is the sale amount 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 property's distance from the nearest 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). The empirical specification is as follows:

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

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

Dependent variable:	Sale amount (US$)				
Treatment (Distance in miles)	3 (1)	2 (2)	1.5 (3)	1.25 (4)	1 (5)
Post × $\mathbb{1}_{Distance<Xmiles}$	-4811.581*** (278.625)	-6863.670*** (364.377)	-7266.981*** (449.406)	-7390.588*** (522.977)	-7197.827*** (645.557)
Property fixed effects	Yes	Yes	Yes	Yes	Yes
Sale-year × plant fixed effects	Yes	Yes	Yes	Yes	Yes
R^2	0.87	0.87	0.87	0.87	0.87
Observations	1,506,933	1,161,063	1,017,934	958,329	909,049

Table B.5: Heterogeneity by price: Above median – around a plant's emission of newly identified carcinogens

Notes: This table presents regression estimates for the impact of toxic plants on the price of houses within 5 miles of the toxic plant. We utilize the National Toxicology Program's Report on Carcinogens (RoC) to determine when chemicals were identified as known or reasonably anticipated human carcinogens. The year of the RoC is the event year. The sample is restricted to plants that were producing these chemicals before their classification in the RoC and continued their production afterward. 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 was sold in the year following the plant's emission report. We define five treatment rings based on the property's distance 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 separate higher and lower-priced properties: we include an indicator variable $Above_i$ for whether property i sale price was above the median value, calculated from sales 3 to 5 years before the emissions report for properties around a plant. The empirical specification is as follows:

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

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

Treatment (Distance in miles)	3	2	1.5	1.25	1
Below median					
Treated	-0.0074	-0.0075	-0.0167**	-0.0097	-0.0103
	(0.0051)	(0.0064)	(0.0077)	(0.009)	(0.0112)
Above median					
Control	-0.0437***	-0.042***	-0.0423***	-0.0428***	-0.0429***
	(0.0053)	(0.0053)	(0.0054)	(0.0054)	(0.0054)
Treated	-0.0381***	-0.0409***	-0.0495***	-0.0473***	-0.0521***
	(0.0061)	(0.0079)	(0.0098)	(0.0116)	(0.0153)
Difference: Treated minus Control					
Below median	-0.0074	-0.0075	-0.0167**	-0.0097	-0.0103
	(0.0051)	(0.0064)	(0.0077)	(0.009)	(0.0112)
Above median	0.0057	0.001	-0.0072	-0.0044	-0.0092
	(0.0061)	(0.0079)	(0.0098)	(0.0116)	(0.0153)
Property fixed effects	Yes	Yes	Yes	Yes	Yes
Sale-year × plant fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	98,792	74,703	63,978	59,304	55,572