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.

June 4, 2024

Abstract

We quantify the impact of perceived environmental health risk on property values and neighborhood composition using a widely-advertised re-classification of chemical carcinogenicity. Focusing on plants that were previously emitting the chemicals, we estimate a 2% decline in property values for houses closer to the plant relative to those farther away. Our analysis also reveals a shift towards a higher presence of minority households in houses closer to the plant. This evidence of granular 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, Itay Goldstein, Todd Gormley, Byoung-Hyoun Hwang, Amit Kumar, Ryan Lewis, Holger Mueller, Kasper Nielsen, Henri Servaes, Huan Tang, Daniel Wolfenzon, Anthony Lee Zhang, and participants at the CEPR Endless Summer Conference on Financial Intermediation and Corporate Finance, EIEF Rome Junior Finance Conference, EPA TRI National Conference, LUISS Finance Workshop, 17^{th} North American Meeting of the Urban Economics Association (Toronto), Society for the Advancement of Economic Theory, TCU Neeley Finance Conference, University of Kentucky Finance Conference, EDHEC Business School, King's College, London Business School, Nanyang Business School, Nova SBE, Peking University, Université de Namur, and University of Washington (Seattle) 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 individuals are newly diagnosed with cancer and 609 thousand die of it, with minority groups disproportionately affected. Cancer is the second leading cause of mortality in the US with spending on cancer care exceeding $200 billion in 2020.[1] A crucial question for academics and policymakers is the value individuals place on reducing environmental health risks, specifically cancer risk, a value that is not directly observable and challenging to estimate. In this paper, we use house prices to quantify the dollar amount households attribute to changes in perceived environmental health risk. Using transaction-level house price data combined with carcinogenic chemical re-classification events we find that, around these events, prices within a 3-mile radius of a toxic plant exhibit a relative 1-2% drop compared to prices between 3 and 5 miles.

Our empirical strategy is designed to isolate the effect of environmental health risks from other factors such as economic factors, on house prices. To this end, we focus on already-existing plants that reported emitting carcinogenic chemicals for the first time due to the widely advertised national re-classification of chemical carcinogenicity by the National Toxicology Program's Report on Carcinogens (*RoC*). These re-classification events are salient information events (Gormley and Matsa, 2011), which we show to be unrelated to changes in plant-level employment and sales for the affected plants. Moreover, following the recent literature (e.g., Currie, Davis, Greenstone, and Walker, 2015, Diamond and McQuade, 2019), we compare price changes for properties within a 3-mile radius of the plants (treated group) to price changes for properties within a 3-to-5-mile radius (control group) around these re-classification events. Our identifying assumption is that, while, after the event, perceived environmental health risks are higher for properties closest to the toxic plants, all properties within a 5-mile radius are similarly affected by other factors influencing local house prices.

We restrict the sample to properties that were transacted both in the year before or in the year of the re-classification event and in the year after. 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 other time-varying changes in the 5-mile ring around the plant. The *RoC* events, the small ring size, the short time window, and the stringent fixed effects that we

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

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 show that the above-mentioned estimates (1-2%) are robust to longer event windows (from year -2 to year +2 around the event).

We then consider a set of plant-year events that include *all* existing plants that reported emitting carcinogenic toxins for the first time to the Environmental Protection Agency (EPA) under the Toxics Release Inventory (TRI) program. These reporting events consist of both the re-classification of chemicals as well as reporting in response to changes to the plants' production decisions. The latter are likely to be less salient than the RoC events. In this case, the estimated decrease in property values halves to approximately 1%.

Another 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 above-median-price properties in the area are the ones most affected by the re-classification events, experiencing a house price drop of over 4%.[2] Overall, we find that the drop between house price changes in the treated relative to the control group is larger for above-median price properties. To the extent that the more expensive properties are owned by households with higher incomes, these results provide evidence of an income channel through which sorting may occur.

Next, we focus on the ethnicity of buyers, which requires data on the characteristics of individuals living in the vicinity of the plants. Our dataset includes the names of the buyers in the property transactions. This allows us to use the classification algorithm of Sood (2017) and Laohaprapanon, Sood, and Naji (2022) to predict the race and ethnicity of those involved in each transaction. The estimates show a relative increase of 1 percentage point 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 the pre-event period. Finally, for a subset of home buyers in our sample, we observe their FICO score. The estimates show that the

[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 an observation other than the two that we use to measure price changes.

FICO scores in the post-event treated group are 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 carcinogenic plants towards minorities and higher credit-risk households.

Finally, we present a stylized model to rationalize our findings. We assume agents have different income levels and freely choose where to reside. There are two regions with differing levels of carcinogenic risk, and the residential choice leads to differences in agents' life expectancy. The model replicates the empirical results and sheds light on the mechanisms behind our estimates: households sort into areas of different environmental quality on the basis of their income. The model also sheds light on the importance of information rigidities and how they interact with income. We abstract from the possibility that high-income households are better informed than low-income households about the higher carcinogenic risk in the area close to the toxic plant and show that despite this assumption high-income households already exhibit a stronger response to the news. This optimal behavior leads to a migration pattern that mirrors the one documented in the empirical analysis. Our theoretical framework does illustrate that while differential salience of the events may certainly exist across income groups, the model does not require such a feature to fit the observed pattern in the data.

The rest of 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 robustness tests. 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 contributes to existing works that have documented changes in property values coming from changes in air quality (Bayer, Keohane, and Timmins, 2009, Chay and Greenstone, 2005, Smith and Huang, 1995), lead remediation (Gazze, 2021, Billings and Schnepel, 2017), hazardous waste remediation (Gamper-Rabindran and Timmins, 2013, Greenstone and Gallagher, 2008), toxic plant openings and closings (Currie, Davis, Greenstone, and Walker, 2015) and power plant openings (Davis, 2011). Our dataset and institutional setting allow us to identify the impact of carcinogenic risk on property values.

A contribution of our paper is the use of the re-classification of carcinogenicity of industrial chemicals by the *RoC* to estimate effects on house values. This is motivated by recent work by Gormley and Matsa (2011) who study firm-level responses to these events in the context of corporate liability risk management. Households may have imperfect information about hazardous emissions (Hausman and Stolper, 2021) or they may not possess the ability to fully process this complex information (Bui and Mayer, 2003). Using the widely advertised *RoC*-related events we find estimates that are twice as large as when we include the wider set of events in which plants first report the emission of carcinogenic toxins. This finding suggests the importance of information and salience for measuring the value that households attribute to environmental health risks.

Our findings also relate to a large literature providing evidence of information rigidities for households and professionals in a variety of settings (Mankiw and Reis, 2002, Carroll, 2003, Coibion and Gorodnichenko, 2012, 2015) as well as to a literature documenting the importance of information frictions regarding 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).

Finally, our paper contributes to the literature on environmental justice (Banzhaf, Ma, and Timmins, 2019, Burda and Harding, 2014, Shapiro and Walker, 2021). Prior work has documented disproportionate pollution exposure for low-income and disadvantaged households (Agyeman, Schlosberg, Craven, and Matthews, 2016, Mohai, Pellow, and Roberts, 2009, Tessum, Paolella, Chambliss, Apte, Hill, and Marshall, 2021) and neighborhood demographic changes in response to air quality changes (Gamper-Rabindran and Timmins, 2013, Banzhaf and Walsh, 2008, Currie, 2011). We add to this literature by showing a shift towards a higher presence of minority and higher credit-risk households in houses closer to carcinogenic plants. Our model rationalizes this finding; even when low- and high-income households are equally informed, high-income households respond more strongly to the news leading to a migration pattern mirroring the empirical observations.

3 Data and methodology

3.1 Data sources

Toxics Release Inventory. Our first data source is the TRI data from the EPA. Firms that satisfy the following criteria must report their emissions to the EPA: (i) the number of employees has to be 10 or larger; (ii) the industry sector in which the plant operates has a covered NAICS code; (iii) firms manufacture with, or use TRI-listed chemicals in the production process; 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 the carcinogenic ones on which we focus.[3]

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 from 2000 to 2020.[4] To determine which plant reports the emission of carcinogenic toxins for the first time due to the re-classification of chemical carcinogenicity at the national level, we proceed in two steps. First, we select plants that report for the first time the emission of carcinogenic toxins to the EPA. Then, among these plants, we select the ones that were already using non-carcinogenic chemicals, which were later re-classified as carcinogens by the NTP's *RoC*.

We first identify plant-years that reported for the first time the emission of carcinogenic toxins to the EPA. Plants that already satisfied our event definition in the year 2000 (the start date of our sample) are excluded, since we do not know whether this is the first year in which criteria were satisfied. For each subsequent calendar year, we construct an indicator variable 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 included in the Clean Air Act.

Our paper uses variation introduced by plants' reporting for the *first time* their release of

[3]The carcinogenicity classification for each chemical follows three agencies: 1) the International Agency for Research on Cancer (IARC); 2) the National Toxicology Program (NTP); 3) Occupational Safety and Health Administration (OSHA) 29 CFR part 1910 Subpart Z.

[4]Further details are available here.

known human carcinogens into the environment.[5] We emphasize the word *known* since the *RoC* re-classification events focus on 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). We define the event-year as the one in which the chemical became known as a carcinogen according to the RoC.

The overall set of plants that report emitting carcinogenic toxins for the first time during the sample period consists of 14,787 unique plants. This set includes both plants that because of their production decisions started reporting to the EPA (the event year is the reporting year) and those that were previously emitting chemicals that were later on re-classified as carcinogenic (the event year is the re-classification year).

National Establishment Time-Series. A second data source we use is the National Establishment Time-Series (NETS) dataset provided by Walls & Associates and Dun and Bradstreet, which, as Rossi-Hansberg, Sarte, and Trachter (2021) show, compares favorably with Census data in terms of quality and coverage. The NETS data includes annual revenues and employment information at the plant level (from 1990 to 2020). When merged with TRI data, it allows us to test for the potential relation between the events and plants' economic activities.

The NETS data has an additional use in the analysis of the first-time reporting of carcinogenic toxins sample. A potential confounder in this sample 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. The NETS data includes the year of the plant opening, which we use to exclude 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. Appendix Figure D.1 plots the location of these plants, distinguishing between *RoC* and non-RoC events. There is significant geographical overlap between the two. Appendix Table D.1 reports the industry coverage and the most frequent chemicals for our sample of 11,561 plant-year events (lead is responsible for 47% of the plant-events followed by nickel with

[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. 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.

17%, whereas the largest proportion of plants are in manufacturing, 88%).[6]

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.[7] The other 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.

Corelogic Deed & Tax Records. To measure 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.[8] For each of the 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.

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 events, we include only property observations corresponding to the first event and we link them to the plant closest to it. This reduces the number of plant-events to 7,801.

Finally, 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

[6]This larger sample includes plants that 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 use the NETS data to investigate this hypothesis.

[7]For this 40% of plant-events, our estimates will capture the *combined* effect of reporting to the EPA and of reporting the emission of carcinogenic chemicals.

[8]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.

above \$700,000 (95^{th} percentile), and focus on repeated property transactions, which leaves us with 6,405 plant-events for the analysis of house price effects. In the Appendix, we show the robustness of our results to a less stringent definition of outliers and consider dropping only houses with a value above \$2.57 million (99^{th} percentile of the distribution), which leads to larger price drop estimates.

Identifying changes in the carcinogenic status of chemicals. We hand-collect information on changes in the carcinogenic status of chemicals using the NTP's *RoC*, following Gormley and Matsa (2011). The report is published by the US 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 then. 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 change to the status or addition of chemicals to the report are indicative of the scientific consensus that these chemicals are likely to be carcinogenic.

We focus on reports published during the sample period that include the years 2002, 2004, 2011, 2014, and 2016.[9] In our analyses, we use chemicals that are newly identified as carcinogens, or are reasonably anticipated to be human carcinogens ($RAHC$) as defined by the NTP. Based on these criteria, we match these chemicals to plants that use them (as we know from the TRI reports). We require these firms to already report the usage of these chemicals in the year before the change in their status and to continue reporting their usage in the year after. This reduces the possibility of changes in the plant activity that may affect house prices directly. The Chemical Abstract Service (CAS) Registry Number unique to chemicals allows us to link the information across the two datasets.[10] We merge this list with the plant data mentioned above and find a total of 1,585 unique plant-events impacted by the re-classification.

[9] In Appendix Figure D.3 we present an excerpt from the Washington Post discussing the case of Styrene, a toxic chemical in our dataset.

[10] The Comprehensive Environmental Response, Compensation, and Liability Act of 1980 (CERCLA, 1980) enacted by the US 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).

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 (media) and model their potential location-based health impacts on the population exposed to these chemicals. We use these indicators to measure cancer risk in the vicinity of the plants (intensive margin) and how it varies with distance. This evidence informs the definition of the ring size in our empirical strategy.

We use data at the most granular spatial unit - grid cells of dimension 810m $\times$ 810m. The data includes toxic concentrations of chemicals for release by spatial unit and media (air, water, or soil), and the number of people in the grid cell who are potentially exposed. We use two measures of cancer risk. First, we rely on model estimates of cancer risk by the EPA, namely, *RSEI cancer scores*. They offer 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. The RSEI cancer scores do not directly translate to the probability of a lifetime cancer diagnosis.[11] 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.[12] We sum the so-computed-cancer-risk scores at the grid level.

As an illustrative example, Figure 1 plots the heatmap by grid cells (810m $\times$ 810m grids) for the computed cancer risk scores surrounding Arch Wood Protection Inc., Conley, GA 30288, in 2004, 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.

Figure 2a presents a binscatter plot of the relation between the cancer-risk scores and distance (up to a 5-mile radius) for *RoC* plant-event-years for which we are able to compute cancer-risk scores. The data indicate cancer risk is highest in the immediate vicinity of the plant and declines with distance, without significant changes beyond 3 miles from the plant.

[11] To the best of our knowledge, the only paper linking RSEI scores to cancer risk is Boyle, Ward, Cerhan, Rothman, and Wheeler (2023), which however focuses on a specific type of cancer —Non-Hodgkins Lymphoma —for 4 cities in the US.

[12] It is important to scale the RSEI cancer scores by exposed population who live in a particular grid to obtain for each chemical-grid cell-media the product of concentration times toxicity weight for cancer effects. Not scaling will lead to lower RSEI cancer values as we shrink the ring closer to the plant: this is the mechanical result of fewer individuals living in treated areas that are smaller.

Second, for a subset of chemicals, we observe the oral cancer slope factor. This factor quantifies the incremental lifetime cancer risk from oral exposure to a particular chemical and provides an estimate for the rate at which cancer risk increases with the dosage in the low-dose region for carcinogens. We compute the lifetime probability of a cancer diagnosis by combining these TRI-reported oral slope factors (Q^*) with chemical air releases. For each chemical, we use the following relationship:

$$\text{Lifetime Cancer Probability}_c = \text{Concentration}_c \times \frac{\text{IR} \times \text{ED}}{\text{BW} \times \text{AT}} \times Q^*_c, \tag{1}$$

where Lifetime Cancer Probability$_c$ is the additional lifetime cancer probability due to exposure to chemical c, Concentration$_c$ is the concentration of the chemical in the contaminated environmental medium (air) to which the person is exposed (measured in μg/m^3) multiplied by 1000 for consistency of measurement units, IR is the intake rate of contaminated air (measured in m^3/day), ED is the exposure duration (measured in number of days), BW is the body weight of the exposed person (measured in kilograms), AT is the average length of time over which the average dose is calculated (measured in days), and Q^*_c is the oral slope factor expressed as risk per mg/kg-day.

We follow the EPA's assumptions of people breathing $20m^3$ of air every day. Relative to the EPA's assumption of a 70-kilogram body weight, we make a more conservative assumption of the average person weighing 84 kilograms. We follow Health Canada (2004), Li, Iwayemi, Li, Komives, and Chakrabarti (2022) and calculate an upper-bound to lifetime cancer risk under continuous exposure to contaminated air in a particular grid (ED/AT equal to 1). Since we are able to compute this measure at the chemical-grid level, we aggregate the estimated values of lifetime cancer probabilities at the grid-level.

Figure 2b repeats the binscatter plot using the lifetime cancer probability from air emissions for RoC plant-event-years included in our sample that release any of the chemicals for which the EPA provides the estimates of the oral cancer slope factor. Again, the data indicate cancer risk is highest in the immediate 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, that are central to our empirical strategy.

3.2 Empirical strategy

Our empirical strategy is designed to isolate the effect of environmental health risks from other factors, such as economic factors, on house prices. To achieve this, we proceed in two steps. First, we use chemicals that are newly identified as carcinogens by the NTP's *RoC*. To hold constant plant production decisions, we require plants to already report the usage of these chemicals in the year before the change in their status, and to continue reporting their usage in the year after. Consistently, we show that these events are unrelated to changes in employment and sales at the affected plants. We then introduce the ring methodology, which compares housing transactions close and farther away from these plants and uses this comparison to identify environmental-health-risk effects on house values. The identifying assumption is that, while, after the event, perceived environmental health risks are higher for properties closest to the toxic plants, all properties within a 5-mile radius are similarly affected by other factors influencing local house prices.

3.2.1 No change in economic activity around the *RoC* events

We use the NETS data to analyse changes in plant-level sales and employment around the *RoC* events, using a matched sample empirical approach. Specifically, for each plant j in our sample, we find control plants within the same state, and industry in the same event year. We assign the same event-year as a pseudo event-year to the matched control plant. We allow for our effects to be estimated within each pair of treated and control plants.

Let n be the 6-digit NAICS code, s the state and t the year. We estimate changes in plant-level employment and sales within a 5-year window around the event (from year -2 to year +2) using the following empirical specification:

$$\log(y)_{jnst} = \alpha + \beta_{\text{Post}\times\text{Treated}} \times \text{Post}_{jt} \times \text{Treated}_{jt} + \gamma_j + \gamma_{nt} + \gamma_{st} + \epsilon_{jt}, \tag{2}$$

where the dependent variable y can be either plant-level employment or sales for plant j, and the variable Post_{jt} takes the value of 1 for observations after the event year, and zero otherwise. We include plant fixed effects so that the estimates capture within-plant changes in employment. We also control for time-varying industry and state characteristics through industry $\times$ year and state $\times$ year fixed effects. Thus, our estimates measure changes in economic activity associated with the plant, beyond those reflected in industry and state fixed effects.

Table 1 shows the estimates for plant-level employment and sales. The estimated coefficients are not significantly different from zero: changes in the carcinogenic status of chemicals are not associated with increased production levels at the plants emitting those chemicals.

3.2.2 Ring methodology

Even though the *RoC* events are not associated with changes in plant-level economic activity, there may be other sources of local economic activity influencing the evolution of house prices in the same year as the events. We address this identification challenge by exploiting the location of plants and properties at a 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 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 and the *RoC* information event have more of an impact on properties located closest to the toxic plant, whereas all properties located within the 5-mile radius are exposed to the same changes in local economic activity (or other factors affecting house values). Several recent papers use this "ring" method for identification purposes (Butts, 2022, Diamond and McQuade, 2019, Ganduri and Maturana, 2022, LaPoint, 2022).

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).[13] Second, Figure 2 plots the relationship between the computed probability of cancer diagnosis (cancer-risk score) and distance from the plant. The figure shows a higher incidence of such risks manifesting within the 3-mile ring compared to the 3-to-5-mile ring 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 keeping the control group the same. Third, in Online Appendix B, we show that air quality is significantly worse near the plants up to 1.5 miles.

Next, we test whether, around the re-classification events, plants mitigate/modify their emissions of carcinogenic chemicals. 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 a) the lifetime probability

[13]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.

of a cancer diagnosis, and b) the RSEI cancer score scaled by the exposed population.

In this context, it is important to note that the *RoC* events are simply information events. After all, the plants were producing the carcinogenic chemicals before the re-classification and continued their production afterward. Furthermore, after the re-classification the cancer scores are updated and back-filled, so that the data will reflect actual changes in cancer risk with the benefit of hindsight.

We estimate changes in cancer risk in the year after the *RoC* events relative to the year before. To avoid including observations with multiple events, we include only grid-cell observations corresponding to the first event-year and link them 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 in 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 centroid 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 centroid between 3 and 5 miles of the same plant. The dependent variable in the regressions is the level of cancer risk at the 810m×810m grid level. The empirical specification is as follows:

$$\text{Cancer Risk}_{ljt} = \alpha + \beta_{\text{Post}\times\text{Distance}} \times \text{Post}_{lt} \times \mathbb{1}_{lj}^{\text{Distance}_{lj}<X\text{miles}} + \gamma_{jt} + \gamma_l + \epsilon_{ljt}. \tag{3}$$

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 characteristics of the plant. Standard errors are double-clustered at the plant and year levels.

Table 2 reports the estimated relative increase in lifetime cancer probability ($\beta_{\text{Post}\times\text{Distance}}$). The estimated coefficients are positive, but economically very small (between 0.2 and 0.3 basis points) and mostly statistically insignificant. This is consistent with firms not adjusting their production plans around the re-classification events.

3.2.3 Identifying environmental-health-risk effects on house values

Next, we present the empirical specification that we use to estimate house price effects. Let j be 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:

$$\log(\text{Sale amount})_{ijt} = \alpha + \beta_{\text{Post}\times\text{Distance}} \times \text{Post}_{it} \times \mathbb{1}_{ij}^{\text{Distance}_{ij}<X\text{miles}} + \gamma_{jt} + \gamma_i + \epsilon_{ijt}, \quad (4)$$

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).[14] 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.

The above specification includes plant×property-sale-year fixed effects (γ_{jt}) that control for time-varying economic 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. The parameter of interest $\beta_{\text{Post} \times \text{Distance}}$ measures the *within*-property relative changes in the sale amount from one year before to one year after.

The use of a repeated-sale approach over a narrow time window helps with the identification of house price effects, but introduces selection in the properties included in the sample. Specifically, the sample includes properties that are transacted at least twice in a window of three years, including at least once in year +1. This sample of properties 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.[15] Moreover, for this subset of properties, increases in prices may at least partially reflect the value from associated improvements. While our data does not allow us to measure such improvements, we try to (at least partially) address the issue by dropping properties that are transacted twice in the same year. To further address selection, we also provide estimates for a wider window of five years around the event (from year -2 to year +2) and find similar results.

[14] We first focus on transactions that took place in the calendar year before the event (year -1), in the calendar year of the reporting event (year 0), and in the year after (year +1), but later on we expand the event window to [-2,+2].

[15] Appendix Figure D.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.

4 Environmental-health-risk effects on house values

4.1 Baseline estimates

Panel A of Table 3 presents our baseline results. The different columns present 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% 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 1 mile from the reporting plant. The estimated decline in house values increases to 2.3% 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 regression 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 2%.[16]

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 bootstrapped samples 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 bootstrapped sample to estimate equation (4) for a treated ring of 1 mile. Figure 3a plots the empirical distribution of the estimated house price effects in the bootstrapped samples. The vertical line denotes the estimates reported from panel A of column 5 in Table 3. The p-value for the null hypothesis from this empirical cumulative distribution is 0.001, confirming the significance of our estimate.

The estimates of Table 3 panel A are for a tight window around the event (from year -1 to year +1), which helps with the identification of the effects of the re-classification on house prices, but also reduces the number of properties in our regression sample. 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 event window, property fixed effects control for all the property features that remain constant in these 3 years.

To assess the sensitivity to our choice of the event window, in Table 3 panel B we repeat the analysis expanding the window to [-2,+2]. That is, we now use properties that are trans-

[16] In Appendix Table D.3 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 significant decrease in the number of transactions after the event in the treated relative to the control group.

acted at least once before the event and once in the two years after.[17] The estimated effects are statistically significant, albeit slightly smaller in economic magnitude than those in panel A. Interestingly, unlike the [-1,+1] event window, the relative house value declines are now monotonically larger for smaller ring sizes.

We further investigate the importance of distance from the plant for the estimates by constructing separate dummy variables for properties in the treated group based on their distance. Specifically, $\mathbb{1}^{\text{Distance}_{ij}\in(m_0,m_1]}$ are dummy variables that take a value of one if property i is between m_0 and m_1 miles, and zero otherwise. The control group is the same as before (properties between 3 and 5 miles of the plant). The empirical specification is as follows:

$$\begin{aligned}\log(\text{Sale amount})_{ijt} = \alpha &+ \beta_1 \times \text{Post}_{it} \times \mathbb{1}_{ij}^{\text{Distance}_{ij}\in(0,1]} + \beta_2 \times \text{Post}_{it} \times \mathbb{1}_{ij}^{\text{Distance}_{ij}\in(1,2]} \\ &+ \beta_3 \times \text{Post}_{it} \times \mathbb{1}_{ij}^{\text{Distance}_{ij}\in(2,3]} + \gamma_i + \gamma_{jt} + \epsilon_{ijt},\end{aligned} \tag{5}$$

where, as before, we include property and plant × sale-year fixed effects and standard errors are clustered at the property level. Figure 4 plots the coefficient estimates (β_1, β_2, β_3) and the corresponding 95% confidence intervals, for event windows [-1,+1] and [-2,+2]. Both figures show that the effects are driven by properties closer to the plants, with the effects for those between 2 and 3 miles not significantly different from the control group.

Next, we explore the changes in house values year-by-year. Figure 5 plots the coefficients on event-time dummies and their corresponding 95% confidence intervals for the treated and control groups separately. We focus on a window of 7 years around the *RoC* events and, given the just-mentioned results from Figure 4, define properties that are transacted within a 2-mile radius of a toxic plant as the treated group. Properties transacted within 3-to-5 miles of the plant form our control group. The estimates are normalized to time zero, relative to the event year. Further, the empirical specification includes property fixed effects to control for time-invariant property characteristics. Standard errors are clustered at the property level.

As evident from the figure, there is a significant increase in house prices in our window of 7 years around the *RoC* events for both the treated and control groups.[18] Specifically, we see that, relative to time zero, before the event, the estimated coefficients are negative and increase over time for both treated and control groups. Importantly, the difference between the

[17] 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.

[18] We observe substantial increases in house prices, also because a large fraction of the *RoC* events are concentrated in the year 2004, a period that saw significant house price growth in the US.

groups is neither economically nor statistically significant. At the same time, after the event, the estimated coefficients increase for both groups but less so for properties in the treated group. However, the difference between the two groups (approximately 3%) is economically and statistically significant.

In the appendix, we present estimates from additional robustness exercises for our main specification in Table 3. 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 D.4 shows that the results are robust with estimated house price effects of similar magnitude and around 2%. In the second robustness exercise, we show results when using house prices in levels rather than in logs. Estimates are presented in Appendix Table D.5. We observe house price declines between 5 and 13 thousand dollars, relative to an average price of around $210,000, depending on the size of the treated ring and the event window. The larger magnitude of the estimates compared to the benchmark values in Table 3 is due to the large skewness of the distribution of house price levels. Lastly, in Appendix Table D.6 we show that our baseline estimates are robust and larger in magnitude when we drop only the top 1% of the price distribution, namely, properties whose price exceeds $2.57 million.

4.2 First-time reporting of carcinogenic toxins

In this section we present results for the sample that includes all the events for the first-time reporting of carcinogenic toxins. This is a larger sample than the one focusing on the reclassification events. The additional events in this larger sample are likely to be less salient and coincide with changes in plant-level economic activity.[19]

Table 4 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, the decline in the value of treated properties is around 1%. This decline is roughly half of the previous estimates focused on the RoC events. Figure 3b plots the results of a placebo test with randomized distance between properties and plants, similar to the one that we have previously described. The figure shows that the estimated house price change in

[19] Table D.2 shows the results for the larger sample of first time reporting of carcinogenic toxins. 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 in general there may be plant-level changes in economic activity associated with pollution reporting events.

the treated group of -1% is significantly lower than the values obtained for the samples with randomized distance. The vertical line denotes the estimates reported from panel A of column 5 in Table 4. The p-value for the null hypothesis from this empirical cumulative distribution is 0.002, confirming the significance of our estimate.

In Appendix C, we provide a back-of-the-envelope calculation that uses our estimates of house price drops and increases in cancer probability for this broader sample to estimate the value of 1 additional year of statistical life. Our estimates imply a value ranging from $28k to $87k per additional year of statistical life.

5 Neighborhood composition

We study changes in neighborhood composition in the narrow geographical areas around the carcinogenic plants. To do so, we examine a) the heterogeneous response of properties based on their price, b) the ethnicity of buyers, and c) the creditworthiness of buyers.

5.1 Heterogeneity by property prices

We begin by focusing on the heterogeneous effects depending on how expensive the properties are. To the extent that the more expensive properties are owned by households with higher incomes, such an analysis is informative of an income channel through which sorting occurs.

In our empirical approach, we define expensive properties as follows. To avoid using the same price observations for separating properties and estimating house price effects, for each event-plant j, 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).[20] 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.

We then estimate regressions of changes in housing values around two years before and after the event year by interacting the *Post* dummy with the indicator for the distance from the plant and the indicator for the above-median price level. Specifically, we estimate the following empirical specification:

[20]This choice has implications in terms of the number of observations available and the properties that are selected.

$$
\begin{aligned}
\log(\text{Sale amount})_{ijt} = \alpha &+ \beta_{Distance} \times \mathbb{1}_{ij}^{\text{Distance}_{ij}<X\text{miles}} + \beta_{Post} \times \text{Post}_{it} + \beta_{Above} \times \text{Above}_{ij} \\
&+ \beta_{\text{Above}\times\text{Distance}} \times \text{Above}_{ij} \times \mathbb{1}_{ij}^{\text{Distance}_{ij}<X\text{miles}} + \beta_{\text{Post}\times\text{Distance}} \\
&\times \text{Post}_{it} \times \mathbb{1}_{ij}^{\text{Distance}_{ij}<X\text{miles}} + \beta_{\text{Post}\times Above} \times \text{Post}_{it} \times \text{Above}_{ij} \\
&+ \beta_{\text{Post}\times\text{Distance}\times\text{Above}} \times \text{Post}_{it} \times \mathbb{1}_{ij}^{\text{Distance}_{ij}<X\text{miles}} \times \text{Above}_{ij} + \gamma_i + \gamma_{jt} + \epsilon_{ijt}.
\end{aligned} \tag{6}
$$

Panel A of Table 5 shows the results for the *RoC* sample. Panel B of Table 5 shows the results for the sample of first-time reporting of carcinogenic toxins. From the table, we see that above-median-price (expensive) properties are the ones experiencing a drop of 4−6% and among the expensive properties, those in the vicinity of a toxic plant are estimated to have a larger house price drop compared to expensive properties that are farther away from the toxic plant. Moreover, the bottom two rows of panel B show that differences between the treatment and control groups are more significant, both statistically and economically, for above- compared to below-median-price properties. The drop in estimated house values for above-median-price properties in the treatment relative to the control ring varies between 70 basis points (for a treated ring size of 3 miles) and 107 basis points (for a 1-mile ring size). Taken together, if more expensive (above-median-price) properties are purchased by higher-income households, these results suggest that these households value environmental health risks more significantly (or are better informed about them).

5.2 Changes in minorities and household creditworthiness

Next, we aim to shed light on granular changes in neighborhood composition by focusing on the ethnicity of the buyers. The existing census data is too coarse for this purpose. However, the Corelogic data includes the names of the buyers in property transactions. To classify race and ethnicity using names, we rely on the algorithm proposed by Sood (2017) and Laohaprapanon, Sood, and Naji (2022). They 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.[21]

More precisely, we first use their algorithm to predict the likelihood that the buyer is of Hispanic or African-American origin. We then construct a dummy variable, $\mathbb{1}$(Minority), that

[21] 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 Hispanics treated as one group.

takes the value one when the individual is predicted to be of Hispanic or African-American origin (and zero otherwise). We focus on Hispanics and African-Americans since the algorithm can identify individuals belonging to these groups with a higher degree of precision. However, within these two groups, Hispanics can be better identified than African-Americans.

Out of all buyers, 14.61% and 4.05% are classified as Hispanic and African-American origin, respectively. As a comparison, we have obtained US homeownership data by race and ethnicity from the 2020 US Census: 10.5% of homeowners are identified as Hispanic and 7.9% as African-American.[22] 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. A potential reason is that the areas where the events take place are different than the US as a whole. In our regression analysis, we take the differences in race and ethnicity of buyers within the same area and compare their incidence before and after the event.

The dependent variable $\mathbb{1}(\text{Minority})_{ijt}$ is the minority indicator variable referring to the buyer of property i near plant j in year t. The equation we estimate is:

$$\mathbb{1}(\text{Minority})_{ijt} = \alpha + \beta_{\text{Post}\times\text{Distance}} \times \text{Post}_{it} \times \mathbb{1}_{ij}^{\text{Distance}_{ij} < X \text{miles}} + \gamma_{jt} + \gamma_i + \epsilon_{ijt}, \tag{7}$$

where the set of fixed effects is similar to before. The event window is [-2,+2] to maximize the number of observations available for the analysis. Our focus is on the identity of buyers in the post-event period and in the treated areas, relative to the pre-event period and the control group, as measured by the estimated $\beta_{\text{Post}\times\text{Distance}}$. 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 the identity of the seller in the post-event transaction. In other words, we compare buyers and sellers in the post-event transactions relative to before and as a function of distance to the toxic plant.

Panel A of Table 6 examines changes in buyer ethnicity for the *RoC* sample. The estimated positive coefficient on the Post $\times$ $\mathbb{1}^{\text{Distance}< \text{ X miles}}$ in column (1) shows a relative increase of 0.9 percentage points in minority home buyers in the vicinity of the treated plant in the period after the re-classification of the chemicals 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. Panel B reports the estimates for the sample of first-time reporting of carcinogenic toxins. The estimated changes are slightly smaller, consistent with the events being overall less

[22] The US 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.

salient, but they still are positive and statistically significant (up to 1.25 miles).

Finally, for a small proportion of the buyers in our sample, we observe their FICO scores from the CoreLogic Loan Level Market Analytics data.[23] The average FICO score of buyers 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.[24] The Experian data on FICO scores include both homeowners and renters, whereas our data is for house purchasers only, who tend to have higher FICO scores. Despite this, the average FICO score in our data is lower, reflecting the fact that the sample of individuals who live near the carcinogenic plants is not representative of the US population.[25]

We estimate (7) using the level of the FICO score of home buyers as the dependent variable. In Appendix Table D.7, we present the results. In Panel A we show the estimates for FICO score changes around the RoC events. The coefficients are almost always negative, but not statistically significant. The lack of statistical significance may be due to the small number of borrowers for which we observe credit scores. In Panel B we report estimates for the sample of first-time reporting of carcinogenic toxins for which the sample is larger. The estimates are more significant both statistically and economically. They increase in economic magnitude and reach -8 points as we decrease the size of the treated ring to 1.25 miles.

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

6 A stylized model

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

[23]It contains detailed loan data, including loan origination, loan performance, and FICO scores for buyers who purchase their properties with a mortgage in the US.

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

[25]It would be interesting to study the effects of the event on rents, which capture the value of using the properties and are likely to reflect the demand for it by the lower-income segment of the population. Unfortunately, we were not able to obtain data on rents with the granularity needed for the identification.

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 farther away. The two regions differ in the amenities that they have, providing agents with utility A_N and A_F, respectively. Furthermore, the two areas differ in the exposure to cancer risk. Let $P(c \mid r)$ represent the subjective probability of dying of cancer for agents residing in region r (with $r = N, F$), with $P(c \mid N) > P(c \mid F)$ for all agents without loss of generality.

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:

$$U_{i|r} = \log(x_i) + \gamma \log(h_{ir}) + A_r, \tag{8}$$

where γ denotes the weight of housing in the utility function. Agents choose the region they live in, as well as the quantity of non-durable consumption and 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:

$$V_{it} = \max_{x_{it}, h_{it}, r} U_{i|r} + \tilde{\varepsilon}_{ir} + (1 - P(c \mid r))V_{i,t+1} \tag{9}$$

$$s.t. \quad w_i = x_i + p_r h_{ir}, \tag{10}$$

taking as given the wage and the price of housing services in each of the regions. To simplify, we abstract from discounting other than through the effects of the survival probabilities. The random term $\tilde{\varepsilon}_{ir} \equiv P(c \mid r)\varepsilon_{ir}$ represents worker i's idiosyncratic preferences for region r. We scale the random draw ε by $P(c \mid r)$ so that the results are not a direct result of the preference shock specification. We assume that $\varepsilon_{iF} - \varepsilon_{iN} \sim U[-s, s]$ and identically and independently distributed across types.

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:

$$x_i^* = \frac{w_i}{1+\gamma} \tag{11}$$

$$h_{ir}^* = \frac{w_i \gamma}{1+\gamma}\frac{1}{p_r}. \tag{12}$$

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 is

$$V_{ir} = \frac{\tilde{U}_i - \gamma \log(p_r) + A_r}{P(c \mid r)} + \varepsilon_{ir}. \tag{13}$$

Agent i chooses N rather than F if and only if:

$$\frac{\tilde{U}_i - \gamma \log(p_N) + A_N}{P(c \mid N)} - \frac{\tilde{U}_i - \gamma \log(p_F) + A_F}{P(c \mid F)} > \varepsilon_{iF} - \varepsilon_{iN}. \tag{14}$$

To close our stylized model, we assume that agents who die are replaced by new agents of the same type. We also assume that in each region housing supply is fixed and equal to $\overline{H}$, which we normalize to 1. In equilibrium, the price of housing services adjusts so that in each region demand is equal to supply:

$$\int h_{iN}^* d\Phi_N(i) = \int h_{iF}^* d\Phi_F(i) = 1 \tag{15}$$

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

We model the information event as an increase δ in the probability of dying of cancer if agents choose to reside in region N. So the new probability in region N becomes $P(c \mid N) + \delta$. We capture information rigidities by assuming that not all agents are informed about this increase. A fraction λ of households of each type is uninformed and believes the probability didn't change, whereas a fraction $1-\lambda$ obtains the new information about the state of cancer in the area and correctly perceives the new probability to be $P(c \mid N) + \delta$. Information is independently and identically distributed across worker types. We exclude that households can learn over time. We solve for the new equilibrium steady state and compare it to the previous one.

6.2 Calibration

Define the set of parameters Θ as

$$\Theta = \{\overline{w}, \underline{w}, \gamma, s, A_F, A_N, P(c \mid F), P(c \mid N), \delta\}. \tag{16}$$

The model is calibrated at an annual frequency. Table 7 lists the calibrated parameters.

For the values of high and low income, $\{\overline{w}, \underline{w}\}$, we use the FICO scores in the mortgage origination data for the years before the reporting events and divide individuals between above and below median. The above(below)-median FICO scores have a median value of 760 (655) points. To map this into income, we use the Fannie Mae origination sample from 2000Q1 to 2021Q3 and estimate a linear relationship between FICO and income. We obtain an estimate for $\overline{w}$ of \$51k and for $\underline{w}$ of \$38k.

We assume that $P(c \mid F)$ is equal to the 2021 unconditional annual mortality rate in the US of 1.05%. To estimate the $P(c \mid N)$ and δ, we use the values for the back-of-the-envelope calculation for the first-time-reporting sample used in Appendix C, that is 41.19% and 41.45% lifetime cancer probability in the pre- and post-event, respectively. We convert these values into annual probabilities using a 70-year lifetime and assume that half of those individuals impacted by cancer die.

We set the parameter $\gamma = 0.3477$ so that the model matches the fraction of income that individuals spend on housing from 2022 Bureau of Labor Statistics (BLS) data (0.258). The parameter determining the support of the idiosyncratic preferences s is calibrated to match the pre-event proportion of individuals in F who were high-income. The value of amenities in region F helps us capture the ratio of average prices in the control and treated areas in the pre-event, that is, 1.08. We normalize A_N to 0. The model also matches the pre-event proportion of individuals in N who were high-income (48.1% in the data and 48.35% in the model).

6.3 Comparative statics

Our comparative statics involves varying the degree of information rigidity λ. Figure 6 shows the relative house price change as we increase the proportion of informed households. Compared to the baseline pre-event scenario, an increase in the fraction of informed individuals leads to larger relative drop in house prices in the N area. More specifically, given that housing services in each region are in fixed supply, as a larger fraction of agents become aware of the increased cancer risk, house prices in F go up, whereas house prices in N decline. Agents opting to reside in N are implicitly choosing a shorter expected lifespan. The response in Figure 6 can be attributed to the agents' preference in the model for a longer lifespan.

We have assumed that the fraction λ of households who are uninformed is the same in each income group. Despite this assumption, the model predicts a larger decline in the demand among high-income than low-income households. This response prompts a migration pattern, depicted in Figure 7. Notably, for a larger proportion of informed individuals, the proportion of high-income individuals residing in F is larger than before the event. It follows that the average FICO score declines more 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 lower than observed in actual data. Specifically, the maximum relative drop in FICO score is -1.47 for λ equal to 0, 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. The model overall matches the patterns documented in the empirical section shedding light on the importance of the income channel as well as of information rigidities.

7 Conclusion

We offer novel empirical evidence on the value that households attribute to environmental health risks by focusing on plants that emitted chemicals initially deemed to be non-carcinogenic, which were later re-classified as carcinogens by the National Toxicology Program's Report on Carcinogens. The re-classification is a significant information event that, as we show, is not related to the plants' economic activity in the year of re-classification. Focusing on a narrow window around the time of re-classification, we estimate a decline in property values of around -2% for houses closer to the plant relative to houses farther from the plant. The decline is roughly double the one we estimate for a larger sample that includes plants that first report the emission of known carcinogenic toxins to the EPA. The latter set of events is less salient, so that our results reveal the importance of information for the measurement of the value that households attribute to environmental health risks.

Finally, our paper studies the implications of these toxic events on neighborhood composition, revealing meaningful heterogeneity in the estimates as well as a shift towards a higher presence of minority and higher credit-risk households in affected areas. We show that more expensive properties are more sensitive to the reporting event. Moreover, we document an inflow of minority and higher credit-risk households in affected areas, thereby providing important evidence on the social dynamics at play in the wake of a change in the perceived environmental hazards and contributing to the ongoing discussion on environmental justice.

References

Agyeman, J., D. Schlosberg, L. Craven, and C. Matthews (2016). Trends and directions in environmental justice: from inequity to everyday life, community, and just sustainabilities. *Annual Review of Environment and Resources 41*, 321–340.

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, H. S. and R. P. Walsh (2008). Do people vote with their feet? an empirical test of tiebout's mechanism. *American Economic Review 98*(3), 843–863.

Banzhaf, S., L. Ma, and C. Timmins (2019, February). 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.

Billings, S. B. and K. T. Schnepel (2017). The value of a healthy home: Lead paint remediation and housing values. *Journal of Public Economics 153*, 69–81.

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.

Bui, L. T. and C. J. Mayer (2003). Regulation and capitalization of environmental amenities: evidence from the toxic release inventory in massachusetts. *Review of Economics and Statistics 85*(3), 693–708.

Burda, M. and M. Harding (2014). Environmental justice: Evidence from superfund cleanup durations. *Journal of Economic Behavior & Organization 107*, 380–401.

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

Carroll, C. D. (2003). Macroeconomic expectations of households and professional forecasters. *Quarterly Journal of Economics 118*(1), 269–298.

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

Coibion, O. and Y. Gorodnichenko (2012). What can survey forecasts tell us about information rigidities? *Journal of Political Economy 120*(1), 116–159.

Coibion, O. and Y. Gorodnichenko (2015). Information rigidity and the expectations formation process: A simple framework and new facts. *American Economic Review 105*(8), 2644–2678.

Currie, J. (2011, May). Inequality at birth: Some causes and consequences. *American Economic Review 101*(3), 1–22.

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.

Davis, L. W. (2011). The effect of power plants on local housing values and rents. *Review of Economics and Statistics 93*(4), 1391–1402.

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.

Gamper-Rabindran, S. and C. Timmins (2013). Does cleanup of hazardous waste sites raise housing values? evidence of spatially localized benefits. *Journal of Environmental Economics and Management 65*(3), 345–360.

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

Gazze, L. (2021). The price and allocation effects of targeted mandates: evidence from lead hazards. *Journal of Urban Economics 123*, 103345.

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.

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

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

Hausman, C. and S. Stolper (2021). Inequality, information failures, and air pollution. *Journal of Environmental Economics and Management 110*, 102552.

Health Canada (2004). Federal contaminated site risk assessment in canada, part i: Guidance on human health preliminary quanttitative risk assessment (PQRA). *Minister of Health*.

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)*.

Li, L., A. Iwayemi, F. Li, T. Komives, and T. Chakrabarti (2022). Persistent organic pollutants: Contaminated site investigation and management toolkit.

Mankiw, N. G. and R. Reis (2002). Sticky information versus sticky prices: a proposal to replace the new keynesian phillips curve. *Quarterly Journal of Economics 117*(4), 1295–1328.

Mohai, P., D. Pellow, and J. T. Roberts (2009). Environmental justice. *Annual Review of Environment and Resources 34*(Volume 34, 2009), 405–430.

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

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

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 and 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.

Shapiro, J. S. and R. Walker (2021, May). Where is pollution moving? environmental markets and environmental justice. *AEA Papers and Proceedings 111*, 410–14.

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).

Smith, V. K. and J.-C. Huang (1995). Can markets value air quality? a meta-analysis of hedonic property value models. *Journal of political economy 103*(1), 209–227.

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

Tessum, C. W., D. A. Paolella, S. E. Chambliss, J. S. Apte, J. D. Hill, and J. D. Marshall (2021). $PM_{2.5}$ polluters disproportionately and systemically affect people of color in the united states. *Science Advances 7*(18), eabf4491.

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.

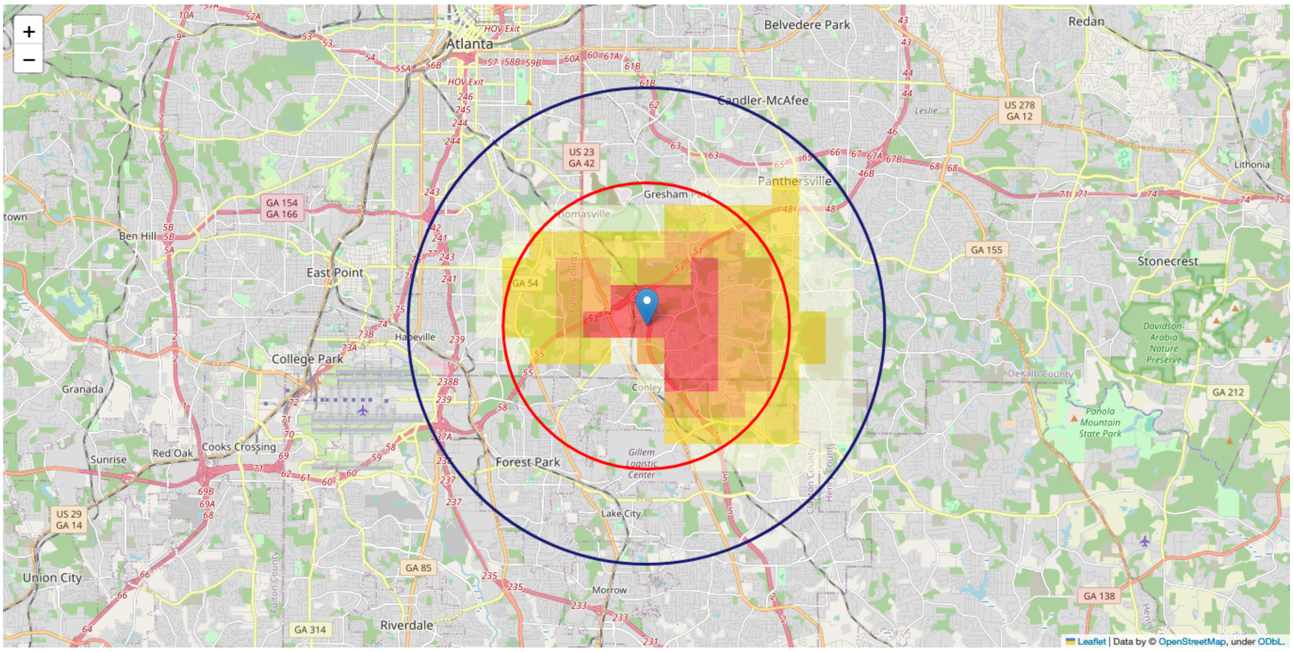
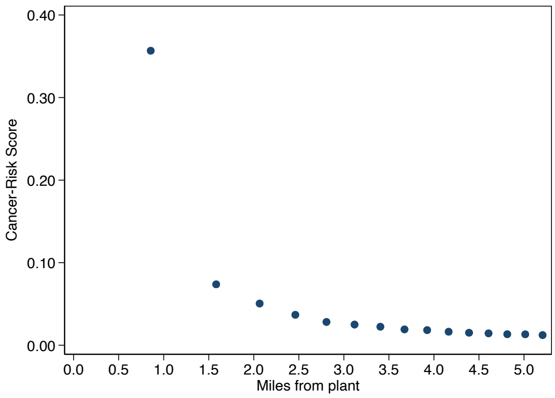
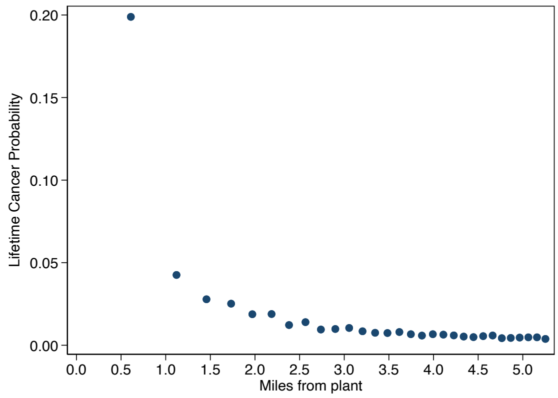


Figure 1: Heatmap of RSEI Cancer score by grid cells for Arch Wood Protection Inc., Conley, GA.

Notes: The figure shows the heatmap by grid cells (810m × 810m grids) for the RSEI cancer scores aggregated for toxic chemicals released by Arch Wood Protection Inc., Conley, GA 30288, in 2004. 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 1.866; the orange area represents values between 1.08 and 1.866; the yellow area represents values between 0.673 and 1.08; the light-yellow area between 0.496 and 0.673; values below 0.496 are not shown. The figure has been produced using the Folium Python Library and Leaflet maps (http://leafletjs.com/).

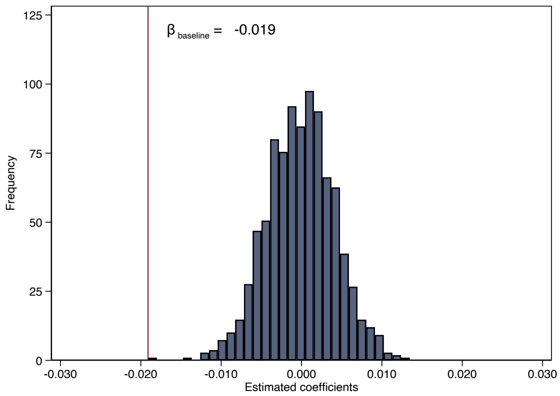


(a) Modeled cancer risk.

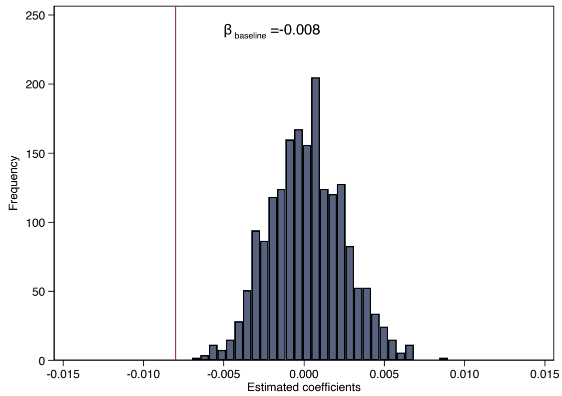


(b) Lifetime cancer risk.

Figure 2: The effect of toxic plants on lifetime cancer risk and modeled cancer risk.
Notes: Panel A shows the binscatter plot for modeled cancer risk for 810m × 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 × 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. Panel B shows the binscatter plot for lifetime cancer probability for 810m × 810m grid cells as a function of their distance from a toxic plant. For the subset of chemicals, we observe the oral cancer slope factor, that quantifies the incremental lifetime risk of cancer from oral exposure to a chemical and provides an estimate for the rate at which cancer risk increases with dose in the low-dose region for carcinogens. We compute the lifetime probability of a cancer diagnosis by combining these TRI-reported oral slope factors (Q^*) with chemical air releases. We follow the EPA's assumption of people breathing $20m^3$ of air every day, and assume the average person in the area weighs 84 kilograms. Please see the text for more details.

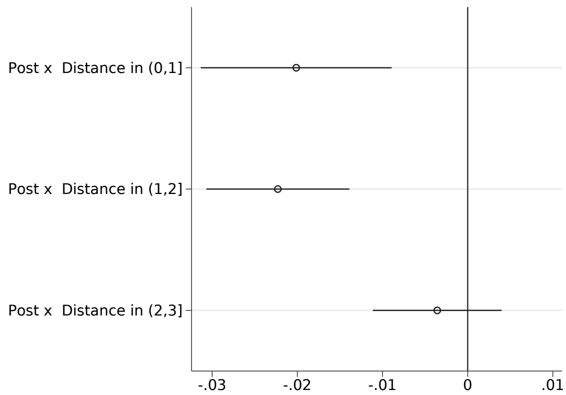


(a) Change in carcinogenic status of chemical

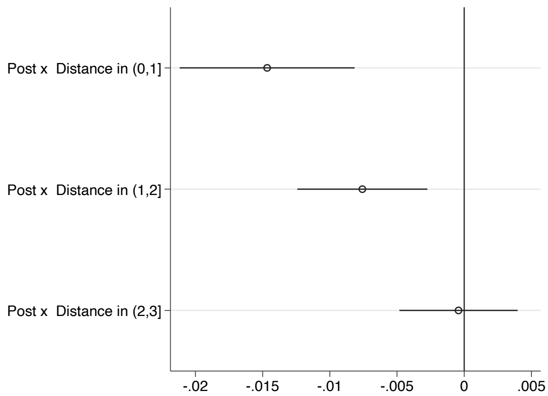


(b) First-time reporting of carcinogenic toxins

Figure 3: 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 (4). 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 (4) in the bootstrap sample. Panel (a) utilizes 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. Panel (b) focuses on changes within one year of a plant's first report of carcinogenic emissions to the EPA's TRI program. In panel (a), the vertical line shows the treatment effect from column 5 in Table 3 panel A while in panel (b), the vertical line shows the treatment effect from column 5 in Table 4 panel A.



(a) Window [-1,+1]



(b) Window [-2,+2]

Figure 4: Empirical estimates as a function of distance.
This figure reports the coefficient estimates (β_1, β_2, β_3) and the corresponding 95% confidence intervals. Specifically, the dependent variable is the natural logarithm of the sale amount of a property, *Log (sale amount)*. The independent variables are indicator variables taking a value of one if property i is sold in the year t after the event year and zero otherwise. We define five rings based on the property's distance from the nearest toxic plant. Specifically, $\mathbb{1}^{\text{Distance}_{ij}}$ takes a value of one if property i is between 0 and 1 mile, 1 to 2 miles, 2 to 3 miles, and zero for properties between 3 and 5 miles of the same plant (control ring). The empirical specification is as follows:

$$\log(\text{Sale amount})_{ijt} = \alpha + \beta_1 \times \text{Post}_{it} \times \mathbb{1}_{ij}^{\text{Distance}_{ij} \in (0,1]\text{miles}} + \beta_2 \times \text{Post}_{it} \times \mathbb{1}_{ij}^{\text{Distance}_{ij} \in (1,2]\text{miles}}$$
$$+ \beta_3 \times \text{Post}_{it} \times \mathbb{1}_{ij}^{\text{Distance}_{ij} \in (2,3]\text{miles}} + \gamma_i + \gamma_{jt} + \epsilon_{ijt}.$$

All regressions include property and plant $\times$ sale-year fixed effects. Standard errors are clustered at the property level.

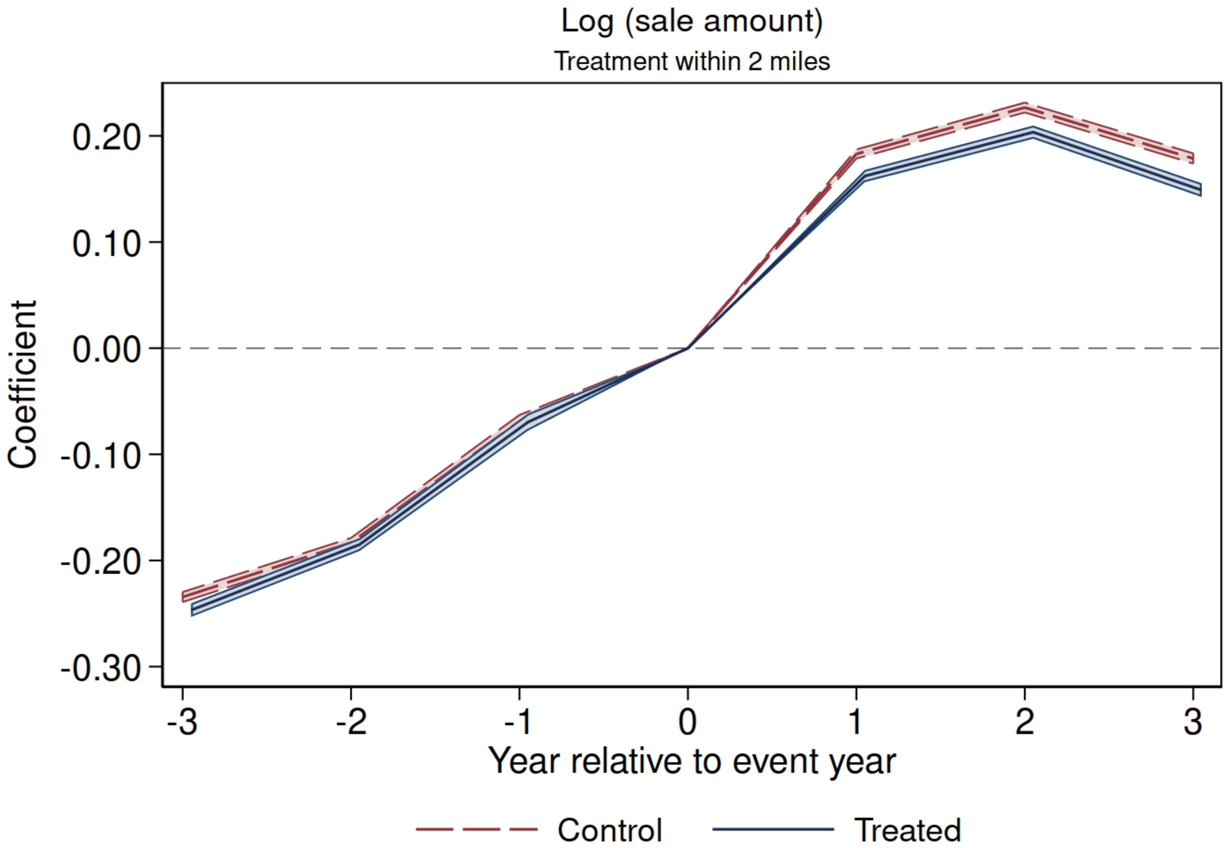


Figure 5: Changes in housing values around the events.
Notes: The figure plots the coefficient on μ_k and the corresponding 95% confidence interval for the treated and control groups separately. We define properties that are transacted within a 2-mile radius of a toxic plant as the treated group while properties transacted within 3-to-5 miles of a toxic plant form our control group. Specifically, for each group, we estimate the following dynamic equation:

$$\text{Log (sale amount)}_{ijt} = \alpha + \sum_{k=-3}^{-1} \mu_k + \sum_{k=1}^{3} \mu_k + \gamma_i + \epsilon_{ijt}$$

We plot, μ_k, which are coefficients on event-time dummies 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 event year. We utilize the National Toxicology Program's Report on Carcinogens (*RoC*) to determine when chemicals were identified as known or reasonably anticipated human carcinogens and use the year of the *RoC* as the event year. The regression includes property (γ_i) fixed effects. Standard errors are clustered at the property level.

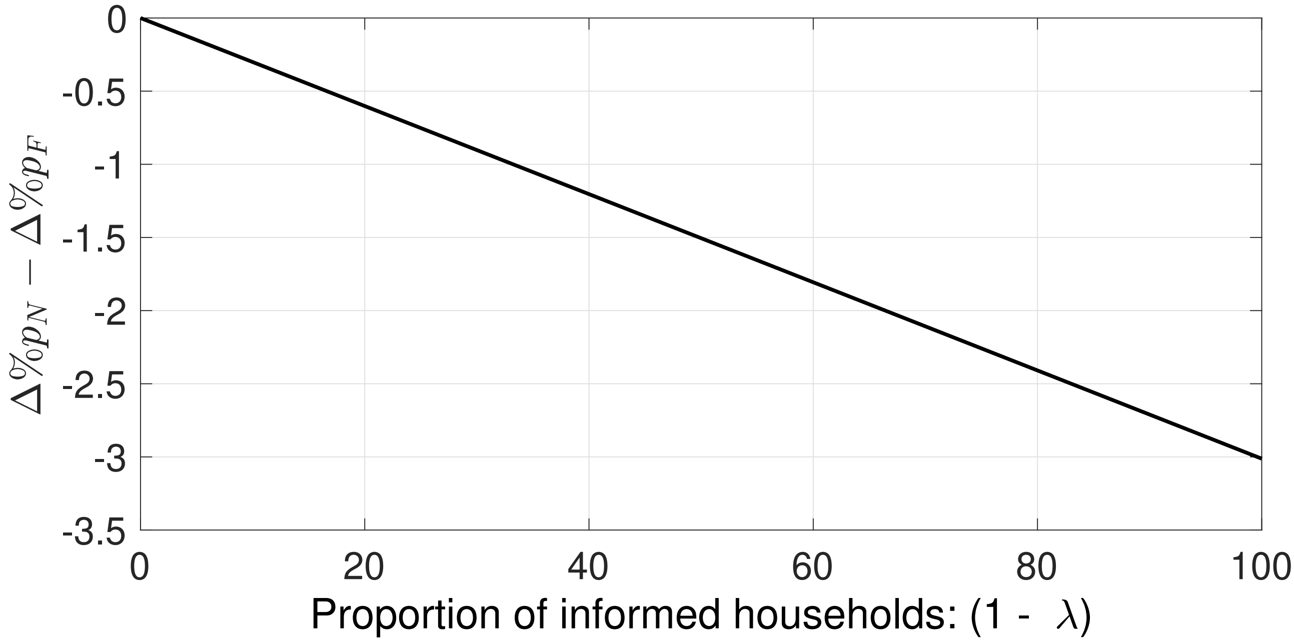


Figure 6: Model-implied price changes.
Notes: Our comparative static involves varying the degree of information rigidity λ. The figure shows model-implied price changes as we vary the fraction of informed households. The model is described in Section 6.

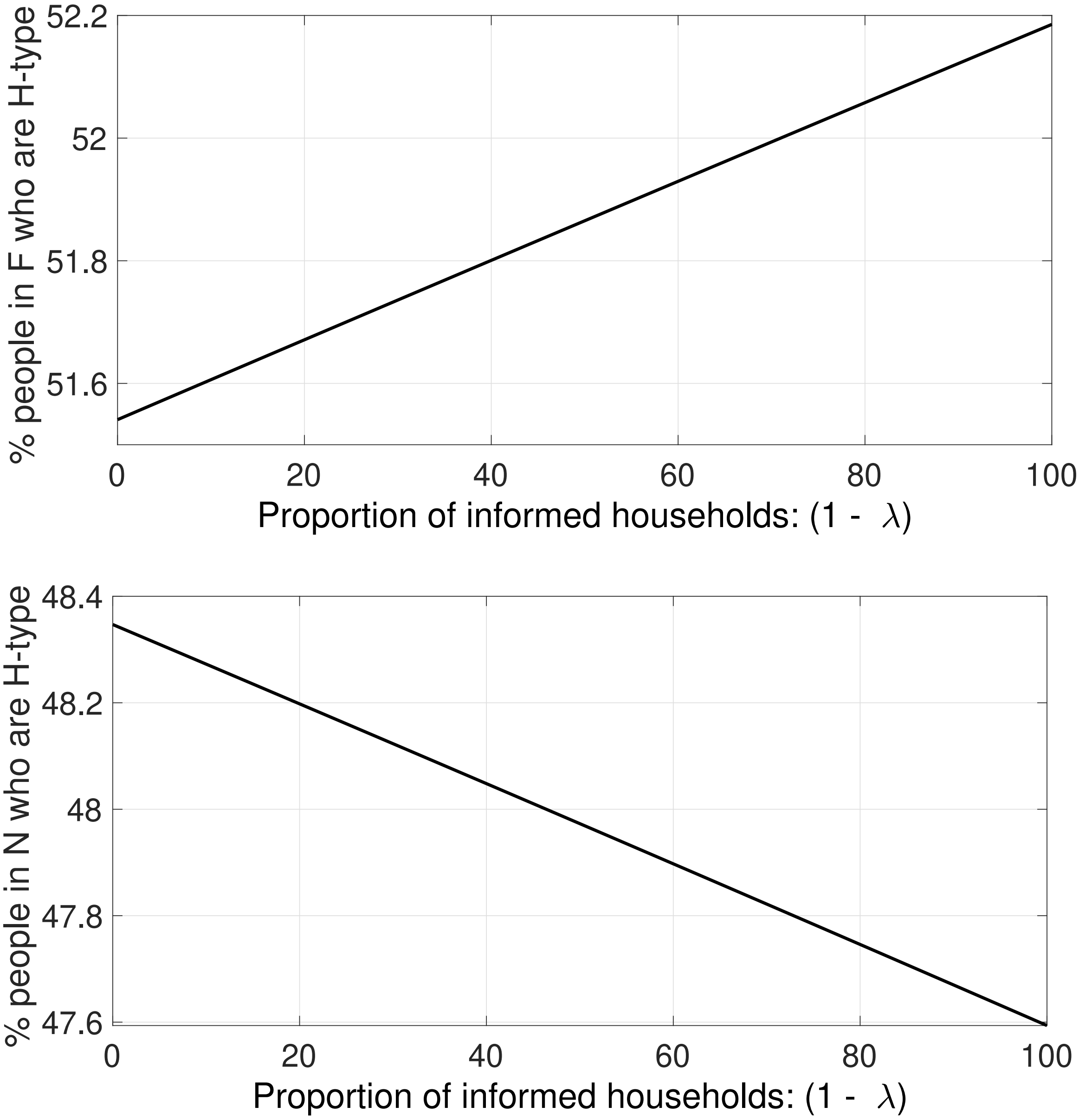


Figure 7: Model-implied migration.
Notes: Our comparative static involves varying the degree of information rigidity λ. The figure shows the model-implied proportion of high-income households choosing to locate near the toxic plant and the proportion of high-income households choosing to locate far from the toxic plant as we vary the fraction of informed households. The model is described in Section 6.

Table 1: Employment and sales around changes in the carcinogenic status of chemicals, Plant-level evidence using matched sample

This table presents regression estimates of changes in plant-level employment and sales within a two year window around the *RoC* events. Specifically, 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* classification is defined as 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 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, while $Treated_{it}$, is an indicator variable taking a value of one for treated plants after the event year and zero for matched control plants. Specifically, for each plant j in our sample, we find control plants within the same state, and 6-digit NAICS industry in the same event year. Further, we assign the event-year as a pseudo event-year to the matched control plant. The empirical specification is as follows:

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

All regressions include plant, industry $\times$ event-year, and state $\times$ event-year fixed effects. 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)
Post $\times$ Treated	-0.013	-0.012
	(0.011)	(0.012)
Plant fixed effects	Yes	Yes
Industry $\times$ event-year fixed effects	Yes	Yes
State $\times$ event-year fixed effects	Yes	Yes
R^2	0.99	0.99
Observations	108,628	108,256

Table 2: Changes in estimated lifetime cancer risk probability

This table presents regression estimates of changes in lifetime cancer probability in a one year window around the *RoC* events. Specifically, 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* classification is defined as 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 level of the lifetime cancer probability at a given location (grid-cells of dimension 800m × 800m) l within 5 miles of the plant j in percentage points. The independent variable, $Post_{lt}$, is an indicator variable taking a value of one if the cancer probability is computed in the year t after the event year and zero otherwise. We define five treatment rings based on the distance of the property from a toxic plant. Specifically, $\mathbb{1}_{lj}^{\text{Distance}_{lj}<X\text{miles}}$ takes a value of one if the cancer probability is within X miles from a plant j, where X is 3, 2, 1.5, 1.25, or 1 mile (columns 1 to 5), and zero for properties between 3 and 5 miles of the same plant. The empirical specification is as follows:

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

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

Dependent variable:	Lifetime cancer probability				
Treatment (Distance in miles)	3	2	1.5	1.25	1
Post × $\mathbb{1}_{Distance<Xmiles}$	0.002***	0.003**	0.003	0.003	0.003
	(0.001)	(0.001)	(0.002)	(0.003)	(0.005)
Grid-cell fixed effects	Yes	Yes	Yes	Yes	Yes
Plant × year fixed effects	Yes	Yes	Yes	Yes	Yes
R^2	0.99	0.99	0.99	0.99	0.99
Observations	253,496	206,734	189,860	183,050	177,504

Table 3: House prices around changes in carcinogenic status of chemicals

This table presents regression estimates of changes in house price within 5 miles of the toxic plant around *RoC* events. The analysis focuses on a window of one year (Panel A) or two years (Panel B) around a *RoC* event. Specifically, 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* classification is defined as 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 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
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

	Panel B: Event window [−2,+2]				
Dependent variable:	Log(sale amount)				
Treatment (Distance in miles)	3	2	1.5	1.25	1
Post × $\mathbb{1}_{Distance<Xmiles}$	-0.003**	-0.009***	-0.012***	-0.012***	-0.015***
	(0.002)	(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	0.87	0.87	0.87	0.87	0.87
Observations	1,218,451	926,936	797,427	740,518	692,721

Table 4: Changes in house prices around the first-time reporting of carcinogenic toxins

This table presents regression estimates of changes in house prices within 5 miles of the toxic plant around the first-time a plant reports carcinogenic emissions to the EPA's TRI program. The analysis focuses on a window of one year (Panel A) or two years (Panel B) around the event. 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.

	Panel A: Event window [−1,+1]				
Dependent variable:	Log(sale amount)				
Treatment (Distance in miles)	3	2	1.5	1.25	1
Post $\times$ $\mathbb{1}_{Distance<Xmiles}$	-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
Post $\times$ $\mathbb{1}_{Distance<Xmiles}$	-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 neighbourhood composition: Heterogeneity by property price

This table presents regression estimates of changes in house prices within 5 miles of the toxic plant around 2 years of an event. In panel A, 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* re-classification defines the event year. The sample is restricted to plants that were producing these chemicals before their re-classification in the *RoC* and continued their production afterward. In panel B, we focus on plants that report for the first-time that they emit carcinogenic toxins in the EPA's TRI program (event-year). Across both panels, 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 event 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.

	Panel A: Changes in carcinogenic status of chemicals				
Dependent variable:	Log(sale amount)				
Treatment (Distance in miles)	3	2	1.5	1.25	1
Below median					
Treated	-0.0063**	-0.0095**	-0.0132***	-0.008	-0.011
	(0.0031)	(0.0038)	(0.0047)	(0.0054)	(0.0067)
Above median					
Control	-0.0491***	-0.0494***	-0.0496***	-0.0498***	-0.0493***
	(0.0032)	(0.0032)	(0.0033)	(0.0033)	(0.0033)
Treated	-0.0509***	-0.0534***	-0.0616***	-0.0685***	-0.0644***
	(0.0037)	(0.0049)	(0.0062)	(0.0072)	(0.0092)
Difference: Treated minus Control					
Below median	-0.0063**	-0.0095**	-0.0132***	-0.008	-0.011
	(0.0031)	(0.0038)	(0.0047)	(0.0054)	(0.0067)
Above median	-0.0018	-0.004	-0.0121*	-0.0187***	-0.0151
	(0.0037)	(0.0049)	(0.0062)	(0.0072)	(0.0092)
Property fixed effects	Yes	Yes	Yes	Yes	Yes
Sale-year × plant fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	265,294	201,347	172,372	160,007	149,850

	Panel B: First-time reporting of carcinogenic toxins				
Dependent variable:	Log(sale amount)				
Treatment (Distance in miles)	3	2	1.5	1.25	1
Below median					
Treated	-0.002	-0.0043*	-0.0054*	-0.00001	0.00001
	(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*	-0.00001	0.00001
	(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 $\times$ plant fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	903,333	695,451	609,632	574,151	544,358

Table 6: Changes in neighbourhood composition: Ethnicity

This table presents regression estimates for the impact of toxic plants on the neighborhood composition within 5 miles of the toxic plant around 2 years of an event. In panel A, 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 re-classification defines the event year. The sample is restricted to plants that were producing these chemicals before their re-classification in the RoC and continued their production afterward. In panel B, we focus on plants that report for the first-time that they emit carcinogenic toxins in the EPA's TRI program (event-year). Across both panels, 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. The independent variable, $Post_{it}$, is an indicator variable taking a value of one if property i was sold in the year following the event 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:

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

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: Changes in carcinogenic status of chemicals				
Dependent variable:	$\mathbb{1}_{Minority}$				
Treatment (Distance in miles)	3	2	1.5	1.25	1
Post × $\mathbb{1}_{Distance<Xmiles}$	0.008***	0.011***	0.008**	0.009**	0.009**
	(0.002)	(0.002)	(0.003)	(0.003)	(0.004)
Property fixed effects	Yes	Yes	Yes	Yes	Yes
Sale-year× plant fixed effects	Yes	Yes	Yes	Yes	Yes
R^2	0.64	0.64	0.64	0.64	0.64
Observations	879,371	666,855	574,629	534,106	499,962

	Panel B: First-time reporting of carcinogenic toxins				
Dependent variable:	$\mathbb{1}_{Minority}$				
Treatment (Distance in miles)	3	2	1.5	1.25	1
Post × $\mathbb{1}_{Distance<Xmiles}$	0.005***	0.006***	0.008***	0.009***	0.008***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
Property fixed effects	Yes	Yes	Yes	Yes	Yes
Sale-year× plant fixed effects	Yes	Yes	Yes	Yes	Yes
R^2	0.64	0.64	0.64	0.64	0.64
Observations	3,142,449	2,415,779	2,118,721	1,994,740	1,890,403

Table 7: Calibrated or externally fixed parameters

For each parameter, the table reports the parameter symbol, numerical value, 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.

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					
γ	0.348	Housing weight	Spending on housing services (BLS, 2022)	0.258	0.258
s	4	Idiosyncratic preferences	% of individuals in F who are high-income	50.8%	51.5%
Region characteristics					
A_F	0.004	Utility from amenities in F	Ratio of average house prices in F vs N	1.0828	1.0828
$P(c\|F)$	1.045%	Mortality rate in F	US unconditional annual mortality rate (2021)	1.045%	1.045%
$P(c\|N)$	1.052%	Additional cancer mortality in N	Additional lifetime cancer probability – pre-period, 1 miles	0.65%	0.65%
δ	0.0031%	Additional cancer mortality in N	Additional lifetime cancer probability – post-period, 1 miles	0.26%	0.26%

Online Appendix

Table of Contents

A 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 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). We show the robustness of our results to a less stringent definition of outliers and consider dropping only houses with a value above \$2.57 million (99th percentile of the distribution).

Merge with RSEI Geographic Microdata. To measure changes in cancer risk around each plant, we rely on estimates provided by the EPA. Specifically, they use 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.

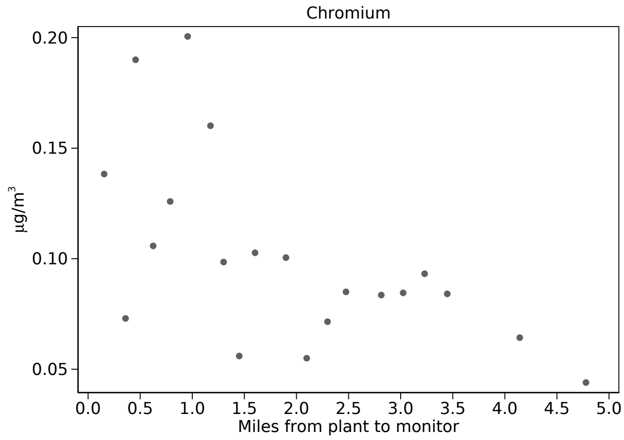
We obtain data at the most granular spatial unit - grid cells of dimension 810m $\times$ 810m. For each grid cell, we observe the cancer risk score associated with a plant. The score is a unitless measure computed for each chemical and media as a product of the estimated dosage released by that specific 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.

B Air pollution measured from monitors

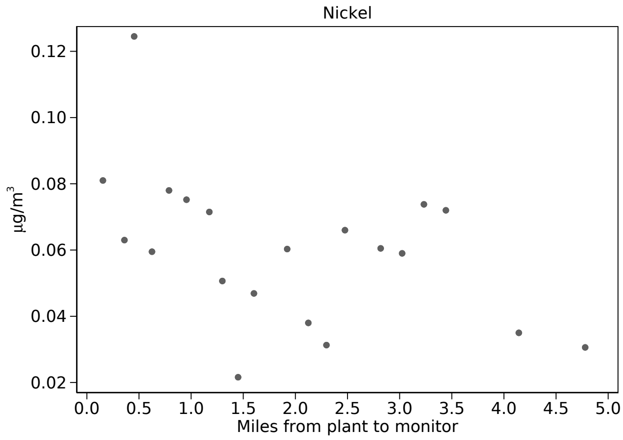
In the main paper, we use measures of cancer risk near the plants from the EPA. These measures are used as input to the calculations of the TRI data. In this section, we provide direct evidence of cancer risk near the plants using data on ambient air quality across the US from the the Air Quality System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

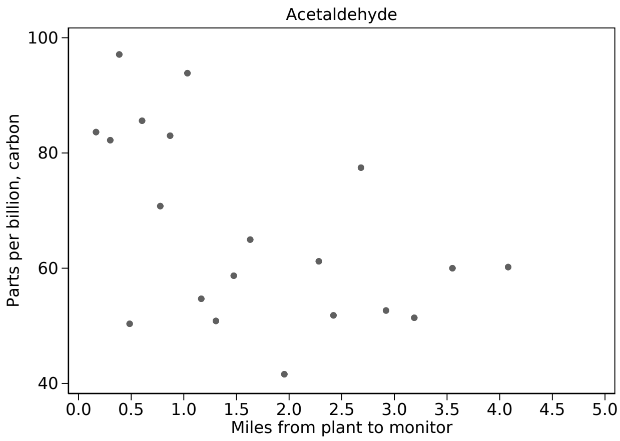
Figure B.1 presents a binscatter plot of the relation between distance from the plant and air monitor (within a 5-mile radius) and the concentration of four predominant carcinogenic pollutants. Although the data are noisy, as it is based on only 266 air monitors, it shows a more pronounced concentration 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.



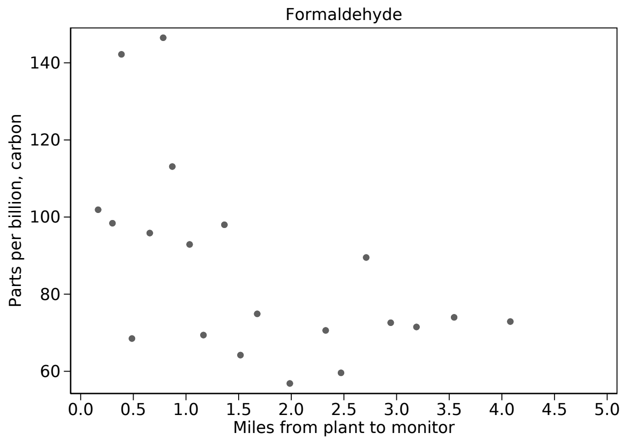
(a) Chromium



(b) Nickel



(c) Acetaldehyde



(d) Formaldehyde

Figure B.1: The effect of toxic plants on hazardous air pollution.
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."

C A back-of-the-envelope calculation

We can use our results to estimate the value of an additional 1 year of statistical life. Our starting point is the mortality rate by age p_t as reported in Table I–10 of the National Vital Statistics Report, Vol. 72, No. 10, September 22, 2023 (data as of 2020). These values represent the base-level mortality rates for our computation.

Regarding cancer, from the American Cancer Society, we observe that 41.6 out of 100 males and 39.6 out of 100 females develop cancer over the course of their lifetime (from birth to death): we use a straight average of the two values (i.e., 40.6).[26] To calculate the likelihood of developing cancer before and after the event in the treated relative to control areas, we augment the base lifetime cancer risk of 40.6% with the increased risk associated with residing continuously within 1 mile of the toxic plant. The resulting total risk is 41.19% for the pre-event (an increase of 59 basis points), and 41.45% for the period after the disclosure (an increase of 84.6 basis points). These risks are then translated into yearly probabilities over a lifetime of 70 years, called c_{pre} and c_{post}. The duration aligns with the EPA's default "lifetime" assumption of 70 years for cancer risk assessments. We assume that with probability 50% an individual who gets cancer dies. The expected lifespan is determined by the formula:

$$E[\#years|i] = \sum_{t=1}^{n} (1 - p_t - c_i)^{t-1} \times (p_t + c_i) \times t, \tag{C.1}$$

where i is either pre or $post$ and n is 105.[27] For an average house value of \$200,000, our estimates imply a drop in price of \$2,000.

We determine the value of 1 additional year of statistical life calculating the ratio of the relative change in housing prices for the treated area to the difference in expected lifetime:

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

Figure C.1 plots the value of 1 year of life computed for different levels of p_{post}/p_{pre}.[28] Our estimates imply a value of \$28,312 per additional year of life corresponding to the 1% drop in

[26] Access the source here. The figures are based on estimates as of January 18, 2024.

[27] We assume no change in the cancer probability for the control areas. Moreover, we assume that in the last year, everybody dies.

[28] In scenarios where the risk of death does not change, the value of an extra year of life approaches infinity, as the denominator in the calculation becomes zero.

house values. On the other hand, for RoC events the drop in house price is \$4,000 implying an estimate of \$56,624 per additional year of life. These values are sensitive to the conditional mortality rate upon having cancer. Assuming that 30% of individuals die of cancer upon diagnosis, our estimates imply a \$43,681 value per additional year of life for all events and a \$87,362 per additional year of life for RoC events.

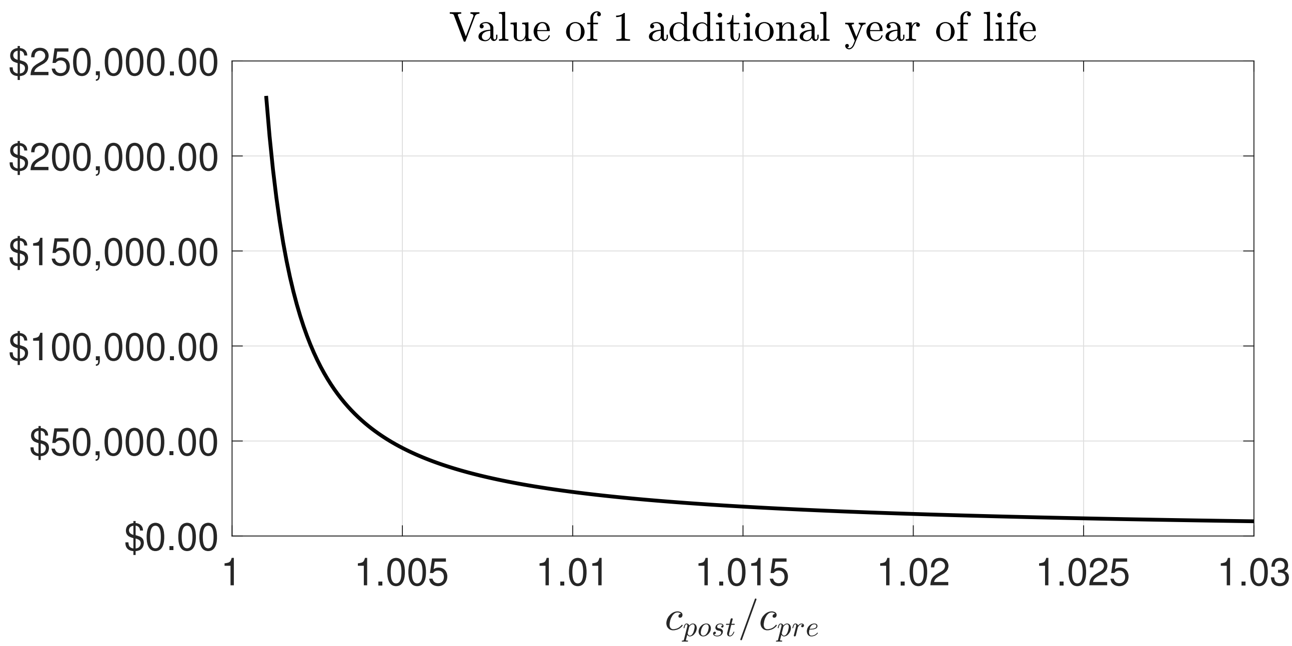


Figure C.1: Dollar value of 1 additional year of life.
The figure shows the dollar value of 1 additional year of life implied by our empirical estimates for different probabilities c of dying of cancer in a given year. Calculations are described in Section C.

D Additional results

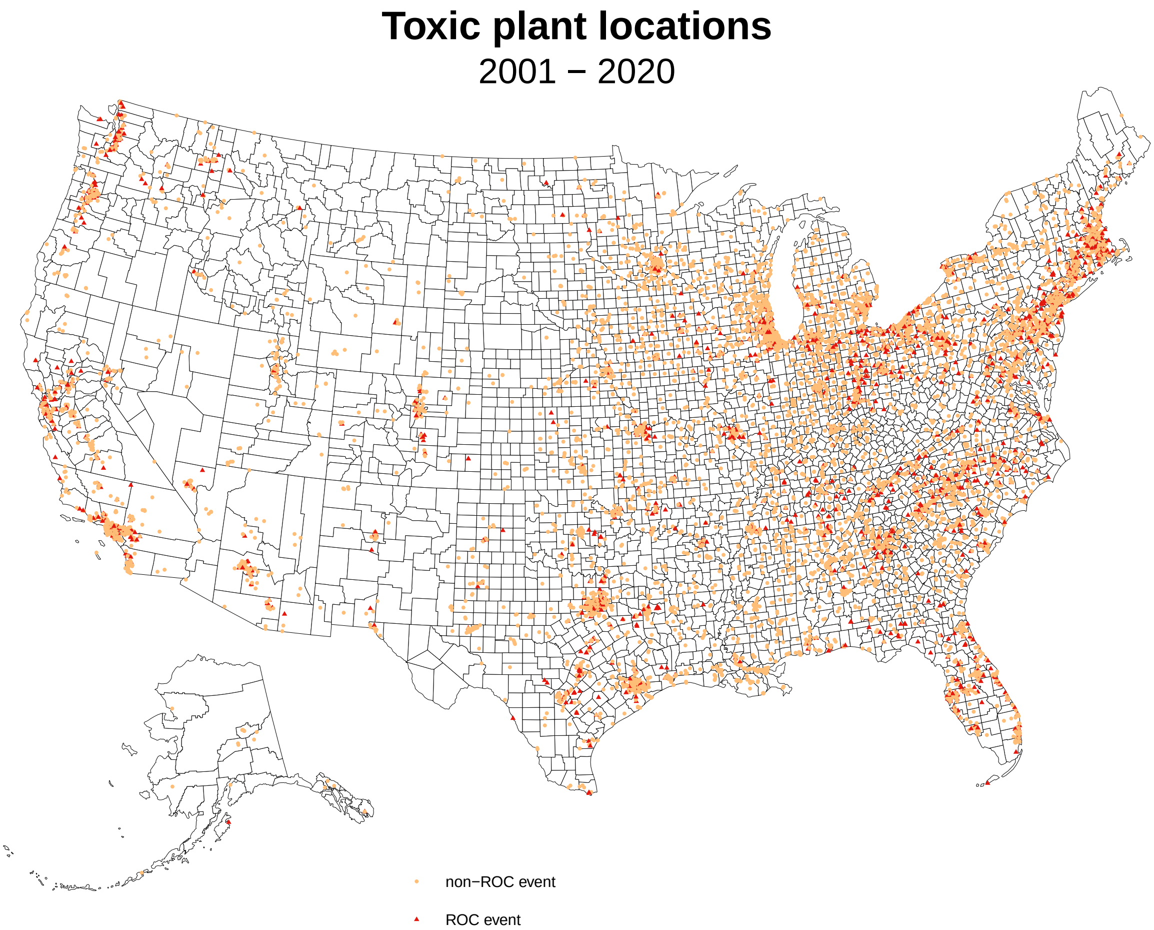


Figure D.1: Location of the reporting plants.
The figure shows the location of plants in our sample. Further, we distinguish events in which plants report the emission of carcinogenic toxins because of a *RoC* re-classification (*ROC events*) from events in which plants report the emission of carcinogenic toxins for the first-time (*non-ROC events*). 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.

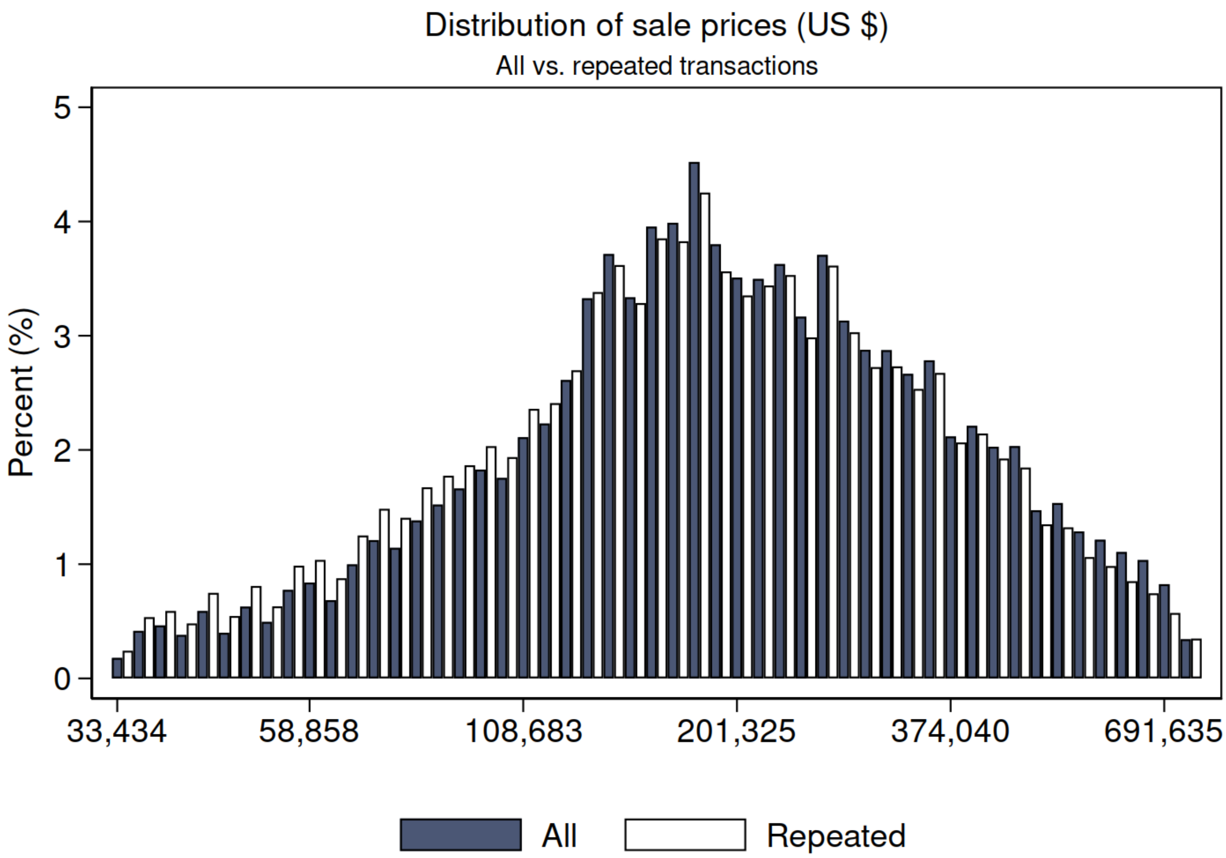


Figure D.2: Empirical distribution of sale prices.
The figure shows the empirical distribution of sale prices for all property transactions (blue solid bars) and repeated property transactions (white hollow bars) in the sample of *RoC* events.

The Washington Post

Democracy Dies in Darkness

Formaldehyde, styrene among substances deemed carcinogens or likely to cause cancer

By Rob Stein

June 10, 2011 at 7:42 p.m. EDT

Share Comment Save

Styrene, which is used to make those ubiquitous white foam coffee cups, food containers and many other products, is probably a human carcinogen, the federal government declared Friday.

The declaration came in the government's latest update of its official list of known or possible carcinogens. It categorized formaldehyde, a chemical widely used to make many products, and a family of substances found in some herbal remedies as known carcinogens.

Figure D.3: News coverage of the change in carcinogenic status of chemicals. The figure shows an excerpt from Washington Post from 2011 describing an update in the carcinogenic status of Styrene.

Table D.1: Summary statistics, first-time reporting of carcinogenic toxins

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

Notes: This table presents regression estimates of 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, while $Treated_{it}$, is an indicator variable taking a value of one for treated plants after the event year and zero for matched control plants. Specifically, for each plant j in our sample, we find control plants within the same state, and 6-digit NAICS industry in the same event year. Further, we assign the event-year as a pseudo event-year to the matched control plant. The empirical specification is as follows:

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

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

Dependent variable:	Log (employment)	Log (sales)
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 D.3: Number of transactions around changes in carcinogenic status of chemicals

This table presents regression estimates for the impact of toxic plants on the number of house transactions within 5 miles of the toxic plant within one year around an RoC event. Specifically, 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 number of transactions, *# transactions*, computed separately for the treated and the control rings around the toxic 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 treatement-ring and 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 × $\mathbb{1}_{Distance<Xmiles}$	0.425	-7.649	-11.722*	-13.195**	-15.933**
	(3.097)	(4.861)	(6.005)	(6.514)	(6.882)
Treatment ring fixed effects	Yes	Yes	Yes	Yes	Yes
Sale-year× plant fixed effects	Yes	Yes	Yes	Yes	Yes
R^2	0.87	0.73	0.66	0.63	0.61
Observations	8,714	8,714	8,714	8,714	8,714

Table D.4: Robustness, more than 199 observations per toxic plant

This table presents regression estimates of changes in house prices within 5 miles of the toxic plant in a one-year window around an *RoC* event. Specifically, 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* classification is defined as 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. 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 × 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
Post × $\mathbb{1}_{Distance<Xmiles}$	-0.007***	-0.021***	-0.021***	-0.016***	-0.014**
	(0.003)	(0.004)	(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	398,095	301,232	258,154	239,025	223,669

Table D.5: Robustness, changes in house prices in levels ($ amount)

This table presents regression estimates of changes in price of houses of within 5 miles of the toxic plant around an *RoC* event. The analysis focuses on a window of one year (Panel A) or two years (Panel B) around a *RoC* event. Specifically, 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* classification is defined as 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 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.

	Panel A: Event window [−1,+1]				
Dependent variable:	Sale amount ($)				
Treatment (Distance in miles)	3	2	1.5	1.25	1
Post × $\mathbb{1}_{Distance<Xmiles}$	-6,373.1***	-11,581.7***	-13,286.1***	-12,641.5***	-12,552.0***
	(532.921)	(683.515)	(818.220)	(937.109)	(1123.646)
Property fixed effects	Yes	Yes	Yes	Yes	Yes
Sale-year× plant fixed effects	Yes	Yes	Yes	Yes	Yes
R^2	0.88	0.88	0.88	0.88	0.88
Observations	443,565	337,679	290,089	268,741	251,589

	Panel B: Event window [−2,+2]				
Dependent variable:	Sale amount ($)				
Treatment (Distance in miles)	3	2	1.5	1.25	1
Post × $\mathbb{1}_{Distance<Xmiles}$	-5,285.2***	-8,901.9***	-10,858.4***	-10,813.2***	-11,980.2***
	(312.553)	(398.275)	(483.381)	(557.935)	(668.830)
Property fixed effects	Yes	Yes	Yes	Yes	Yes
Sale-year× plant fixed effects	Yes	Yes	Yes	Yes	Yes
R^2	0.88	0.88	0.88	0.88	0.88
Observations	1,218,451	926,936	797,427	740,518	692,721

Table D.6: Robustness, dropping properties in the top 1% of price distribution

This table presents regression estimates of changes in price of houses of within 5 miles of the toxic plant around an *RoC* event. The analysis focuses on a window of one year (Panel A) or two years (Panel B) around a *RoC* event. Specifically, 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* classification is defined as 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. Additionally, we drop properties in the top 1% of the price distribution, namely, properties whose price exceeds \$2.57 million. 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
Post × $\mathbb{1}_{Distance<Xmiles}$	-0.045***	-0.068***	-0.060***	-0.045***	-0.028***
	(0.003)	(0.004)	(0.005)	(0.006)	(0.007)
Property fixed effects	Yes	Yes	Yes	Yes	Yes
Sale-year × plant fixed effects	Yes	Yes	Yes	Yes	Yes
R^2	0.84	0.85	0.85	0.86	0.86
Observations	488,956	372,323	318,894	295,507	276,750

	Panel B: Event window [−2,+2]				
Dependent variable:	Log(sale amount)				
Treatment (Distance in miles)	3	2	1.5	1.25	1
Post × $\mathbb{1}_{Distance<Xmiles}$	-0.022***	-0.036***	-0.027***	-0.018***	-0.009**
	(0.002)	(0.002)	(0.003)	(0.003)	(0.004)
Property fixed effects	Yes	Yes	Yes	Yes	Yes
Sale-year × plant fixed effects	Yes	Yes	Yes	Yes	Yes
R^2	0.85	0.86	0.86	0.86	0.86
Observations	1,326,702	1,010,340	868,699	807,606	756,372

Table D.7: Changes in neighbourhood composition: Creditworthiness

This table presents regression estimates for the impact of toxic plants on the neighborhood composition within 5 miles of the toxic plant in the 3 years around the events. In panel A, 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* re-classification defines the event year. The sample is restricted to plants that were producing these chemicals before their re-classification in the *RoC* and continued their production afterward. In panel B, we focus on plants that report for the first-time that they emit carcinogenic toxins in the EPA's TRI program (event-year). Across both panels, the dependent variable is the *FICO Score* of the buyer when available. The independent variable, $Post_{it}$, is an indicator variable taking a value of one if property i was sold in the year following the event 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:

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

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 carcinogenic status of chemicals				
Dependent variable:	FICO Score				
Treatment (Distance in miles)	3	2	1.5	1.25	1
Post $\times$ $\mathbb{1}_{Distance<Xmiles}$	-3.226	-4.948	-3.774	-2.503	2.169
	(2.761)	(3.755)	(4.701)	(5.765)	(7.161)
Property fixed effects	Yes	Yes	Yes	Yes	Yes
Sale-year $\times$ plant fixed effects	Yes	Yes	Yes	Yes	Yes
R^2	0.65	0.66	0.67	0.67	0.67
Observations	15,409	11,440	9,811	9,068	8,521

	Panel B: First-time reporting of carcinogenic toxins				
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